Why are Donated Kidneys Rejected While the Transplant Waiting List is Long and Growing?

Kurt E. Schnier[♦]
School of Social Sciences, Humanities and Arts
University of California, Merced
kschnier@ucmerced.edu

James C. Cox
Experimental Economics Center and Department of Economics
Georgia State University

Vjollca Sadiraj
Experimental Economics Center and Department of Economics
Georgia State University

Nicole Turgeon
Department of Surgery
Emory University School of Medicine

Preliminary Draft Do Not Cite or Circulate without Permission of the Authors
Draft Date: October 8, 2014

Abstract:
We ask why deceased donor kidneys are rejected while the transplant waiting list is long and growing. We investigate the possible causes of the low acceptance rate for deceased donor kidneys including low kidney quality, patient-donor mismatches, and regulatory review triggers. Among the specific questions we address are whether there are transplant surgeon/center behavioral reference point effects from median kidney donor quality or threats from regulatory review criteria that contribute to low acceptance rates. Our results provide evidence for both types of behavioral reference points. Donor quality reference points consistently arise in our analysis, whereas on *average* the regulatory reference points have no effect. However, this is driven by substantial heterogeneity in transplant center responses to the regulation.

JEL Codes: I10, I18, D03
Keywords: Transplantation, Organ Utilization, Reference-Dependence

---

♦ Corresponding Author: Professor of Economics; School of Social Sciences, Humanities and Arts; University of California, Merced; 5200 North Lake Road; Merced, CA 95343. Phone 209-205-6461; Fax 209-228-4007. The authors would like to thank Andrew Balthrop, William Martin, Colleen McIntyre and Ravi Ruhil for their assistance in constructing the Scientific Registry of Transplant Recipients (SRTR) data set. The corresponding author would also to thank Spencer Banzhaf for his suggestions and assistance with this research. The authors also wish to acknowledge funding support provided by NIH (NIA -1RC4AG039071-01).

I. Introduction

In 2011 the United States Renal Data System (USRDS) estimated that the current incidence of End Stage Renal Disease, that which requires organ transplantation, was approximately 357 per one million people. The annual cost of care for these patients is estimated at $49.3 billion.[1] Critical to the care of patients with ESRD is the supply of organs, obtained from both deceased and living donors, and the utilization of these scarce organs. Currently, there are over 100,000 candidates waiting for a kidney transplants in the United States yet in 2013 there were only 16,895 kidney transplants conducted, 11,163 of which obtained from deceased donors and the remaining 5,732 from living donors (OPTN 2014). Furthermore, in 2013 there were 7,524 kidney transplant candidates removed from the waiting list due to death or a deteriorated health status that prevented transplantation and an additional 36,395 candidates added to the waiting list (OPTN 2014). Although patients suffering from ESRD can be maintained for extended periods of time with dialysis there is a growing need to better understand the behavioral aspects of transplantation so as to facilitate the treatment of ESRD patients. Both the economics (Kessler and Roth 2012; Deck and Kimbrough 2012; Li et al. 2013) and medical (Howard et al. 2007; Leichtman et al. 2008; Punch et al. 2007; Shafer et al. 2008; Sung et al. 2008) literatures have researched policies that may increase the supply of deceased donor organs; however, little research has empirically investigated the organ utilization decision.[2] Using a dataset on deceased donor kidney offers and acceptances, paired with detailed information on a patient's health status, donor quality and patient-donor matching characteristics, we investigate the organ utilization decision and the degree to which it is influenced by behavioral reference points.

Despite the ever-growing demand for kidney transplants, the acceptance rate of deceased donor organs is exceptionally low, approximately 2%. These exceptionally low acceptance rates are influenced by the complex decision task that transplant surgeons are faced with. Although the organ utilization question has been theoretically investigated (Howard 2002; Alagoz et al. 2004, 2007), our paper provides the first empirical investigation of the micro-behavioral decisions made by transplant surgeons when electing to accept or reject an organ for their patient. The central focus of our research is a better understanding of the behavioral heuristics utilized by transplant surgeons and to determine whether or not these heuristics possess behavioral reference points. Our results provide considerable support for the presence of behavioral reference points in both the donor and regulatory domain. However, the response to the regulatory induced reference point is highly heterogeneous which if ignored suggests the absence of a regulatory reference point.

[1] This information can be found on the USRDS website: http://www.usrds.org/.
[2] The economics research of Roth et al. (2004, 2005, 2007) on the paired-kidney exchange model has made a noticeable impact on the transplant community with the eventual creation of the New England Kidney Exchange.

A behavioral reference point exists if an agent's marginal utility for taking an action shifts at a distinct point, the reference point, in the choice space (Tversky and Kahneman 1991). The presence of behavioral reference points has been used to expand our knowledge of decision making in a number of different contexts including consumer brand choice (Hardie et al. 1993), labor supply (Camerer et al. 1997; Chang and Gross 2014; Farber 2005, 2008, Fehr and Goette 2007; Crawford and Meng 2011), patient treatment in health care (Rizzo and Zeckhauser 2003), real estate (Genesove and Mayer 2001), asymmetries between willingness to pay and willingness to accept (De Borger and Fosgarau 2008), the behavior of game show contestants (Post et al. 2008) and the behavior of professional athletes (Pope and Schweitzer 2011) to cite a few. There is no theory on the emergence of reference points; selection of behavioral reference points in the literature is often based on the outcomes observed in the data, or induced by the environment, status quo or focal points. Based on the previous outcomes a commonly used method to define reference points is the rational expectations approach proposed by Köszegi and Rabin (2006).[3] They proposed that reference points are formed based on an agent's rational expectations of an outcome, based on prior knowledge, to generate what they refer to as a "personal equilibrium" in which the agent correctly predicts the decision environment as well as their reference points and maximizes utility based on all contingencies relative to the reference point.

As will be discussed in more detail in the upcoming sections, transplant surgeons may form expectations for the quality of a donor organ and the impact that its utilization will have on their patients waiting for a transplant. These expectations determine one of the behavioral reference points used in our empirical analysis and can be defended by the existing medical literature on donor organ quality (Rao et al. 2009). Our second hypothesized reference point comes from the regulatory environment in which transplant surgeons operate.

The following section briefly discusses the decision environment within which transplant surgeons operate and the current regulatory environment. Section III discusses the data set used in our study and Section IV outlines the model. The econometric results are discussed in Section V and Section VI illustrates the high degree of transplant center heterogeneity in behavioral responses to the reference points. Section VII provides a series of robustness checks on our empirical findings and tests the validity of our hypothesized donor organ behavioral reference point. The final section concludes our discussion.

[3] Abeler et al. (2011) experimentally investigated using expectations as a behavioral reference point. By design they are able to exogenously determine the expected outcome of subjects within the experiment and determine whether or not the expected outcome serves as a behavioral reference point. Their experimental results provide support for using rational expectations in the spirit of Köszegi and Rabin (2006) as a behavioral reference point.

II. Transplantation Decision Environment

The deceased donor organ procurement and allocation process is managed by The United Network of Organ Sharing (UNOS), which is overseen by the U.S. Department of Health and Human Services' Health Resources and Services Administration (HRSA). There are 11 UNOS regions, which are further divided into 58 donor service areas (DSA), and the organs are allocated through the Organ Procurement Organizations (OPO) servicing each DSA. Once a potential organ donor is identified at a given hospital the OPO will contact the deceased's next of kin and inquire whether or not they are willing to have the deceased's organ be donated to a patient waiting to receive an organ. After obtaining consent from the family the OPO allocates organs through UNOS using a "match run" process. The match run is an ordinal list of potential patients that the OPO is willing to offer a specific organ to for transplantation. The match run contains information on a donor's medical history, comorbidity factors and other factors that influence the compatibility of a donor's organ for a specific patient (i.e., antigens and blood type). The suitability of a potential recipient is subject to the transplant center's selection criteria and is influenced by multiple factors including length of time on the waiting list and the histocompatibility of the organ. There exists a continuum of "suitability" measures and the decision to accept or reject an organ is left up to the professional judgment of the transplant surgeon responsible for the patient. Based on both donor and recipient information a transplant surgeon determines whether or not the organ offered is a suitable organ for her patient. If she elects to reject the organ for her patient the OPO then proceeds further down the list to the next potential recipient and the process is repeated. Since 2007 this process has been automated using DonorNet (Massie et al. 2009), a web-based offer system, and match runs are processed rapidly until a transplant surgeon accepts an organ.

There are four factors central to the decision to accept or reject a given organ: (1) donor quality, (2) patient status, (3) donor-patient histocompatibility and (4) the regulatory environment within which the transplant surgeon operates. Within the context of our empirical analysis we hypothesize that surgeon expectations regarding the quality of the donor organ as well as the regulatory environment provide behavioral reference points. The other factors, patient status and donor-patient histocompatibility, will be controlled for in our analysis. The data used to construct our measures of donor quality, patient status and donor-patient histocompatibility will be discussed in more detail in the upcoming data description section; however the regulatory environment within which surgeons operate may also have an impact on their utilization decisions.

The Centers for Medicare and Medicaid Services (CMS) monitor the performance of transplant centers to ensure that the transplants conducted meet the performance criteria established under their Conditions of Participation (CoP) that took effect on June 28, 2007. Centers that perform below the CoP standards set by CMS are subject to review and may, in extreme cases, be removed from the Medicare

program. All transplant centers are required to report patient and graft outcomes for the transplants they conduct. Using this information the Scientific Registry of Transplant Recipients (SRTR) reports risk-adjusted outcomes for each transplant center using a 2½ year rolling patient cohort that is updated and published every 6 months. This is the data that CMS uses to monitor a transplant center's CoP. The primary outcomes that CMS monitors are the 1-year graft and patient survival rates which are calculated for the total (deceased and living) donor pool.[4]

A transplant center will be non-compliant with the CMS standards (i.e., not meet a CoP) if three triggers are met. Contextualizing these triggers in the terms of patient survival rates, a CMS review will result if: (1) the observed number of patient deaths exceeds the expected number by three or more; (2) the observed number of patient deaths is more than 50% higher than the expected number; and (3) the one-sided *p-value* resulting from an exact Poisson test comparing the observed and expected patient deaths is less than 0.05. Furthermore, all of the expected outcomes calculated are risk-adjusted based on the patient and donor characteristics (Abecassis et al. 2008; Dickinson et al. 2006, 2008).[5] Any center that does not meet the CMS CoPs twice within a three year period goes through a CMS review and enters a Systems Improvement Agreement with CMS. Failure to meet the terms of the Systems Improvement Agreement may result in the loss of funding and transplant center status. For the purpose of our analysis we focus on a center's compliance with the three "triggers" set forth by CMS under the CoP. A byproduct of these triggers is that those transplant centers that did not meet the CMS CoP standards and activated the three triggers have reduced the volume of their transplants (Schold et al. 2012) as well as increased the waiting time for patients on the waiting list (Schnier et al. 2014).

The incentives induced by the CMS CoP, which penalizes transplant centers that perform transplants that result in worse than expected outcomes (Abecassis et al. 2008), may conflict with the policy of HRSA which advocates the maximum utilization of organs. The impact of this incentive structure is hypothesized to have contributed to a decrease in deceased donor transplants while the number of deceased donors increased in 2007 (Howard et al. 2009). Although the CMS review criteria are hardly the sole cause of the documented widening gap that exists between those patients waiting for a kidney transplant and those receiving a kidney transplant (Wolfe et al. 2010), it has played a role. Given that the demand for kidney transplants has risen 86% from 1999 through 2008 and continues to grow (Axelrod et al. 2009), it is imperative that we better understand the behavioral impacts of the CMS regulations.

[4] These statistics include only those patients who have received either a deceased or living donor organ transplant. A patient who passes away while on the waiting list is not counted in the patient survival statistics monitored by CMS.

[5] The risk adjustment factors used are the cold ischemia time, stroke as cause of death, donor age, donor diabetes status, donor hypertension status, kidney received on a pump, donor creatinine level, patient-donor BMI, patient PRA level, previous transplant recipient, Hepatitis C status, ECD kidney and DCD kidney.

II. Data Description

Our data come from UNOS and SRTR. The UNOS data contains information on all patients waiting to receive an organ as well as the donors who have provided organs. This information describes the patient's and donor's medical histories, comorbidity factors and factors that influence the histocompatibility of potential organ matches (e.g., antigens and blood type). UNOS also provided the match runs conducted during our time period of study. The match run data indicate the donor from which the organ was obtained, the wait listed patient the organ was offered to, the date on which the organ was offered, and the surgeon's decision to accept or reject the organ. Linking the match run data with the patient and donor information contained in the UNOS data set, we were able to construct a large set of patient-specific, donor-specific and patient-donor specific factors that may influence the surgeon's decision to accept or reject an organ.[6]

Our data set contains 5,298,155 match run offers that were made to 291,216 unique patients on the deceased donor organ waiting list at 266 transplant centers within the United States between July 1, 2003 and December 30, 2012.[7] These match run offers resulted in 94,806 deceased donor kidney transplants, approximately 96% of the total deceased donor transplants conducted during this time period, that were derived from 55,360 unique organ donors.[8] The mean acceptance rate in our data was 1.79%. Clearly, the decision to accept an organ is a low probability event. Transplant surgeons are highly selective, on average each donated organ is offered to 56 unique patients before being accepted for transplantation by a transplant surgeon. For each of the 266 transplant centers observed in our data we obtained the SRTR program specific reports from June 30, 2003 through December 31, 2012 (reports are issued every six months). These reports were used to determine whether a specific transplant center did or did not meet the CMS CoP triggers at the time the decision to accept or reject an organ for a patient was made.

From the UNOS data set we construct two variables that are used in our empirical modeling, the Kidney Donor Profile Index ($KDPI$) and the expected cold ischemic time if the organ is accepted for transplantation. Both of these variables are important considerations when electing to accept or reject an organ. The $KDPI$ is derived from the Kidney Donor Risk Index ($KDRI$), which combines a set of key

[6] A limitation of our model is that we only observe match-run data for those organs that were eventually accepted by a transplant surgeon. Currently, the match run data for discarded organs is not available and we have been informed by UNOS that the little which is recorded is unreliable.

[7] The match run data contains refusal codes used to indicate the reason why the organ was not accepted. We eliminated all refusal codes that indicate the organ was either bypassed by the patient or where the reasons for refusal were not solely based on donor quality or the suitability of the organ for the patient.

[8] Our match run sample does not contain all of the deceased donor transplants during this time period because we removed pediatric transplants from our data set and in a few instances we were unable to confidently match up the donor and patient data.

donor characteristics to quantify the risk of graft failure following transplantation (Rao et al. 2009). The variables utilized are the donor's age, height, weight, ethnicity, history of hypertension, history of diabetes, cause of death, creatinine level, Hepatitis C status and whether or not the organ was donated after cardiac death (DCD) (Rao et al. 2009). The $KDRI$ is scaled so that a value of one represents the median donor quality observed. This statistic is defined as the "reference donor" by UNOS. Values below one represent higher quality and a reduced probability of graft failure whereas values greater than 1 have a higher probability of failure (Rao et al. 2009). The $KDPI$ is a normalized measure of the $KDRI$ where the reference donor, $KDRI = 1$, is assigned a $KDPI = 0.5$ indicating that the organ has a risk of failure greater than 50% of the donor pool. This conversion from the $KDRI$ to the $KDPI$ is preferred in our analysis because the $KDPI$ follows a uniform distribution defined over zero and one with a median value of 0.5 and $KDRI$ follows a Poisson distribution defined over zero and four with a median value of one. This is illustrated by the $KDRI$ and $KDPI$ distributions in Figure 1. This implies that the same marginal change in the $KDRI$ reflects a different change in the organ quality depending on where in the distribution the improved quality is observed. This is not true if we use the $KDPI$ variable.

Currently, both the $KDRI$ and $KDPI$ are reported at the time the organ is offered to a transplant surgeon. This information was not provided to transplant surgeons during the time period of our study. However, the estimation of the $KDRI$ and $KDPI$ is based on historically observed organ quality that encompasses our study period. Therefore, $KDPI$ represent the transplant surgeon's rational expectations regarding the quality of a deceased donor organ, prior to knowing the true $KDPI$ value. This is consistent with the reference points proposed by Köszegi and Rabin (2006) and a $KDPI$ of 0.5 can serve as a rational reference point in our decision model. The $KDPI$ merely aggregates the donor quality measures into a single metric that can be used to represent the surgeon's perception of donor quality. The robustness of this assumption is tested in Section VII of this paper.

Cold ischemia time represents the time period between which an organ is severed from its blood supply and placed on ice to the time when it is warmed with a restored blood supply. At the time that an organ is offered to a patient the transplant surgeon forms expectations regarding the cold ischemia time that will be incurred before the organ is eventually transplanted. This expectation weighs heavily in the decision of whether or not to accept an organ. Within our data set we only observe the cold ischemia time at the time of transplantation. To ensure that the expected cold ischemia time is incorporated into the decision model we estimated the cold ischemia time for all transplanted organs and then used the parameter estimates from the regression to impute the expected cold ischemia time for all the unique match runs within the data. As explanatory variables in the model we used the sequence number in the match run (the order of the organ offer), the length of time between the organ being obtained from the

donor and the expected date of transplantation,[9] the donor's BMI and age, regional fixed effects, fixed effects for the donor's death circumstance, an indicator for whether or not the donor was a non-heart beating donor and an indicator variable for whether or not they were an extended criteria donor (ECD).[10] The dependent variable in our model is the log transformed cold ischemia time observed for all the transplants conducted between July 1, 2003 and December 30, 2012 at the transplant centers in our analysis.[11] The results from the cold ischemia time regression are reported in Table 1.

The regression results are consistent with our prior expectations. The cold ischemic time is strongly influenced by the expected time until transplantation and the sequence number in the match run (squared term). Furthermore, there exists a high degree of regional heterogeneity. This is expected given how the organ procurement and allocation process is conducted (spatial scale influences the cold ischemic time). Lastly, the cold ischemic time is longer for non-heart beating donors as well as ECD kidneys. Given the importance of the imputed cold ischemia time to the decision model, we conducted a Q-Q plot of the observed and predicted cold ischemia time distributions that is illustrated in Figure 2. The figure illustrates that our predicted cold ischemic time does possess a downward bias which is exacerbated at the higher quantiles of the actual cold ischemic time distribution, however it does a relatively good job predicting the cold ischemic time at the lower quantiles. More specifically, the model performs well when the cold ischemic time is between 10 and 20 hours, which covers approximately 47% of the observed cold ischemic time distribution. Aside from the cold ischemic time being unobserved when an organ in not accepted it is an endogenous variable, therefore we need to use some form of instrumentation to incorporate it into the model. Given that it is employed as a control variable in our regression model we are satisfied with using the predicted estimates of cold ischemic time.

Another key decision variable in the decision to accept or reject an organ is the degree of histocompatibility between the donor and patient, an integral component of the success of a transplant. Graft survival decreases with the number of HLA mismatches as it increases the level of immunosuppressant medications required to try a transplant and ensure its success. HLA mismatches are calculated based on their unique antigens, A, B and DR, with two potential loci for each antigen. The total number of possible mismatches is six, corresponding with each of the two loci on the three antigens. For transplants conducted after May 1, 2007 the number of HLA mismatches is recorded in the UNOS data

[9] Within our data set the mean difference between the date of transplantation and the date of the match run is 1.081 days. Since we are working with the logarithm of cold ischemia time, to deal with data recorded as 0s we added one day to the difference between the recovery date (date organ was acquired) and the match run date (date the organ was offered) in our imputations of cold ischemia time.

[10] An extended criteria donor (ECD) is a donor over the age of 60 or a donor between the age of 50 and 59 that meets one of the three criteria: (1) cause of death is stroke, (2) history of hypertension, and (3) creatinine level greater than 1.5. The $KDRI$, and hence the $KDPI$, contains all of this information.

[11] The cold ischemic time regression was run using 92.48% of the observed transplants. We removed a small sample of transplants because we did not have sufficient data to incorporate it into the model.

set, however prior to this time period we only observe whether or not the HLA mismatch is a zero mismatch. A zero mismatch will allow a wait-listed patient to move up the match run list because it is a perfectly histocompatible organ.

For the match runs observed prior to May 1, 2007 we constructed virtual HLA mismatches. To construct virtual HLA mismatches we determined the parent-split relationship for all the unique A, B and DR antigen loci and then assigned antigen loci to ensure that all parent-split relationships were preserved. As a check on our virtual mismatch algorithm we determined what percentage of the zero mismatches reported in the match run data set we accurately predicted using our virtual mismatch algorithm and we predicted approximately 95% of the zero mismatches recorded in the match run data set prior to May 1, 2007. Therefore we expect the estimation bias resulting from using our virtual mismatches in this time period to be minimal. To illustrate the role that the HLA mismatches have on the organ acceptance decision, Figure 3 illustrates the percentage of match runs, broken down by acceptance and rejection, that fall into one of the seven categories of HLA mismatches (zero to six). It is very clear that the acceptance rate is extremely high when a zero mismatch donor organ is offered to a patient and the distribution of HLA mismatches for accepted organs is skewed to the left relative to the distribution of HLA mismatches for rejected organs.

In addition to the data referenced above, the UNOS data set contains an exceptionally large number of patient, donor and patient-donor match data, which we utilize as covariate controls in the empirical analysis. On the patient side we observe their diagnosis code at listing, functional status at listing, length of time on the waiting list, ethnicity, BMI, age, gender, transplant history, level of panel reactive antibodies in their blood (PRA) at listing, albumin level, education, and number of other medical history variables such as diabetes, pregnancies, dialysis, hypertension, peripheral vascular disease, vascular access status, peritoneal access status, HPV, HCV, HBV, and COPD. On the donor side we observe the circumstances of their death as well as all the covariates that are used to calculate the $KDPI$. The patient-donor match data we observe are blood type matches and the HLA mismatches discussed earlier. All of these variables are used as covariates in our empirical modeling.

Our research hypotheses utilize two potential behavioral reference points and either a difference-in-difference or a triple difference estimation strategy. To illustrate some of the key variables contained in the data as well as to test for statistical differences in the covariates we provide two tables of descriptive statistics for a select number of key covariates. Table 2 contains a description of the data broken down by the pre- and post-CMS CoP time period as well as whether or not a center met the triggers set forth by CMS. To simplify the description of the data we define $Trig_{tc}$ to take a value of one if either the 1-year graft or 1-year patient survival rate CMS CoP triggers were activated. Table 3 contains a similar set of descriptive statistics based on whether or not the donor organ offered to a surgeon was below the

reference threshold level of donor quality ($KDPI_d \leq 0.5$) or above the threshold ($KDPI_d > 0.5$). In addition, to the means and standard deviations we estimate the normalized differences proposed by Imbens and Rubin (forthcoming) and discussed in Imbens and Wooldridge (2009).[12]

The unit of observation in both Table 2 and 3 is the organ offer, this is consistent with the empirical model we use to investigate the behavioral reference points. Therefore, the variables are weighted according to their occurrence within the data set. The descriptive statistics in Table 2 illustrate the average donor organ offered was obtained from a 41 year old donor with a BMI of around 27. The average $KDPI$ was between 0.62 and 0.66. Although the median donor has a value of 0.50, the higher average value indicates that a lower quality donor, higher $KDPI$, is offered to more patients before being accepted than a higher quality organ. The average cold ischemic time for an organ that is offered to a patient is a little over 3 hours and a little over 20% of the organs offered were ECD kidneys. The average patient offered an organ was nearly 50 years old with a BMI around 27. Furthermore, a little less than 5% of the patients being offered an organ had Type I diabetes, whereas approximately 25% have Type II diabetes. Lastly, between 75% and 83% of the organ offers are made to patients who were on dialysis at the time of listing. The normalized differences across the different partitions of the data indicate that there are no structural differences that would invalidate our empirical specifications.

The acceptance rates from the pre-CMS to post-CMS period decreased by 27% for centers that do not hit the CMS triggers and 30% for those that do hit the CMS triggers. However, comparing these two measures using the normalized differences generates statistics of 0.0303 and 0.0404 respectively, indicating that acceptance rates are not structurally different across these time periods. The distribution of $KDPI$ variable across the pre-CMS and post-CMS time periods also appears to illustrate the percentage of donor organ offers above the reference donor decreased from the pre-CMS to the post-CMS period. However, comparing the normalized differences across the pre-CMS and post-CMS time periods generates statistics of 0.0754 for centers that did not hit the CMS triggers and 0.0455 for those that did hit the CMS triggers, indicating these changes should not generate any empirical issues within our econometric modeling.

[12] The normalized difference is: $\frac{\bar{X}_1 - \bar{X}_0}{\sqrt{(S_1^2 + S_0^2)}}$, where $\bar{X}_j$ are the respective sample means and S_j^2 are the sample variances. The normalized difference in lieu of the *t-test* because with our exceptionally large sample the *t-test* would generate statistical differences whereas we are primarily interested in determining whether or not it is possible to determine the treatment effect using our covariates and linear regression (Imbens and Wooldridge 2009). Their rule of thumb is that a value greater than one quarter makes the linear regression methods utilized to test for treatment effects sensitive to specification.

The descriptive statistics when we break down the organ offers based on whether or not they are for an organ that is above the reference donor threshold, $KDPI_d \leq 0.5$, versus below it, $KDPI_d > 0.5$, generate a similar profile of patients that are offered organs and donors that provide them as observed in Table 3. Although the acceptance rate for organs above the reference donor is nearly twice that observed for those below the threshold, the normalized differences indicate that they are similar. We will leave this to further investigation in the empirical modeling section. The only covariates that do not meet the 25% threshold proposed by Imbens and Rubin (forthcoming) are the age of the donor and the ECD kidney status. This is precisely what we would have expected given that the $KDPI$ variable captures the risk of graft failure based on the quality of the donor organ. A donor's age and ECD status is given considerable weight in the $KDPI$ calculation. The structural differences in the $KDPI$ are as expected by construction of the table.

IV. Model

To provide a theoretical framework for our empirical model we build on the stylized model developed by Rizzo and Zeckhauser (2003). We assume that the probability of accepting the organ is a function of the transplant surgeon's difference in utility between accepting and rejecting an organ, *P(accept)=F(*ΔU*)*. Let the difference in utilities have an additively separable form[13]

$$\Delta U = V(h, d, m, s) + Z(d - r_d, s - r_s), \quad (1)$$

where: h is the patient's health status, d is the quality of the donor organ, m is the histocompatibility of the patient and donor match, s is the transplant center's regulatory status, r_d is the reference donor, and r_s is the reference regulatory status. The function Z incorporates reference points whereas the function V does not. The arguments $(h, d, m, s, d - r_d, s - r_s)$ are variables observed at the time the transplant surgeon makes the decision whether to accept an organ.

We assume that $V(h, d, m, s)$ is decreasing and concave in a patient's health status, $V_h < 0$ and $V_{h,h} < 0$, because a transplant surgeon derives more utility from accepting an organ for a patient with a lower health status than a healthier patient, *ceteris paribus*.[14] Given two identical offers for two patients, one with a lower health status (i.e., longer wait time) and the other with a higher health status (i.e., shorter

[13] In the theoretical literature the decision to accept or reject an organ has been modeled as dynamic choice model (Howard 2002; Alagoz et al. 2004, 2007). We have elected to utilize a static behavioral model of surgeon decision-making to simplify the implied state space in our econometric model. The econometric model would be intractable if we utilize the same number of state variables used in our empirical investigation.

[14] We use a standard notation for derivatives of functions. For any function F one has: F_x and $F_{x,x}$ are the first and second partial derivatives with respect to the variable x and $F_{x,y}$ and $F_{y,x}$ are second cross partial derivatives for variables x and y.

wait time), we assume the transplant surgeon is more likely to accept an organ for the lower heath status patient. Difference in utilities is assumed to be increasing and concave in both donor quality, $V_d > 0$ and $V_{d,d} < 0$, and patient histocompatibility match, $V_m > 0$ and $V_{m,m} < 0$, as well as improvements in their transplant center's regulatory status, $V_s > 0$ and $V_{s,s} < 0$.

The primary difference between our specification and that of Rizzo and Zeckhauser (2003) is that our function Z incorporates two behavioral reference points that may or may not interact with each other in the transplant surgeon's utility function. We assume that utility is increasing in donor quality, $Z_{d-r_d} > 0$ and the transplant center's regulatory status, $Z_{s-r_s} > 0$. Let S^- and S^+ denote two points in the neighborhood of the regulatory reference point, with S^- below and S^+ above the reference point. Similarly let $D^-$and D^+ denote two points in a neighborhood of the donor reference point, with D^- below and D^+ above the reference point. The variables $D^{+/-}$ and $S^{+/-}$ represent points that are either above or below the donor and regulatory reference point respectively. A transplant surgeon will exhibit a regulatory behavioral reference point if $Z_{s-r_s}(D^{+/-}, S^-) = \beta Z_{s-r_s}(D^{+/-}, S^+), \beta \neq 1$ and donor reference point if $Z_{d-r_d}(D^-, S^{+/-}) = \lambda Z_{d-r_d}(D^+, S^{+/-}), \lambda \neq 1$. Based on this model we have four primary research hypotheses.

Hypothesis 1: A transplant surgeon's organ acceptance marginal difference in utility has a regulatory induced reference point such that $\beta \neq 1$.

The current transplantation literature has illustrated that the CMS CoP criteria have resulted in a reduction in transplant volume over time (Schold et al. 2012) and an increase in waiting time for patients (Schnier et al. 2014) at centers that have not met the CMS CoP. It has also been hypothesized that these responses are due to the regulatory induced reference points (Schnier et al. 2013). To test our first research hypothesis we look at whether the CMP CoP provides a behavioral reference point that contemporaneously impacts the decision of a surgeon to accept or reject an organ.

Hypothesis 2: A transplant surgeon's organ acceptance marginal difference in utility has a donor quality reference point such that $\lambda \neq 1$.

Our prior is that a transplant surgeon's evaluation of an available donor organ is based, in part, on whether or not the quality of the organ is above or below the quality of a reference donor organ. Our assumption that a $KDPI$ of 0.5 represents the "reference donor" is consistent with the reference points suggested by Köszegi and Rabin (2006) as its represents the expected quality of a donated organ prior to

evaluating its quality characteristics. As mentioned earlier, we investigate the robustness of this assumption, and therefore the validity of testing this hypothesis, in our robustness check section.

Our final two hypotheses concern whether or not a transplant surgeon's behavioral responses to either of our hypothesized reference points is magnified by whether they are above or below the other behavioral reference point at the same time. For the regulatory induced reference point we can represent this as $Z_{s-r_s}(D^-, S^-) = \gamma Z_{s-r_s}(D^-, S^+), \gamma \neq 1$ and $|\gamma| > |\beta|$. This provides our third hypothesis.

Hypothesis 3: The regulatory induced reference point effect is magnified if the organ offered to a transplant surgeon is also below the donor regulatory reference point, $\gamma \neq 1$ *and* $|\gamma| > |\beta|$.

Our fourth hypothesis mirrors the third. For the donor induced reference point we will observe a stronger effect when also below the regulatory reference point: $Z_{d-r_d}(D^-, S^-) = \delta Z_{d-r_d}(D^+, S^-), \delta \neq 1$ and $|\delta| > |\lambda|$. This provides our final hypothesis.

Hypothesis 4: The donor quality reference point effect is magnified if the organ offer to a transplant surgeon is made when the surgeon's center is below the regulatory reference point, $\delta \neq 1$ *and* $|\delta| > |\lambda|$.

Our first hypothesis does not differentiate based on whether or not the organ being offered to the transplant surgeon is above or below the donor reference point, $D^{+/-}$, therefore we can test our first hypothesis utilizing a difference-in-difference estimation strategy. The first empirical specification we utilize is (Model 1), $P(accept) = F(\Delta U_{itdc})$ where

$$\Delta U_{itdc} = X'_{itdc}\beta + \alpha_1 KDPI_d + \alpha_2 Post_t + \alpha_3 Trig_{tc} + \alpha_4 (Post_t)(Trig_{tc}) + \mu_c + \rho_m + \varepsilon_{itdc}$$

The dependent variable in this regression is a binary indicator variable indicating whether the organ offer presented to patient i in time period t that was derived from donor d was accepted at transplant center c ($\Delta U_{itdc} > 0$). The observation matrix, X_{itdc}, contains the transplant center, patient and donor-patient characteristics. These variables consist of patient diagnostic code fixed effects, patient ethnicity, patient's current waiting time, patient comorbidity factors (i.e., diabetes, hypertension, COPD, drug use, hepatitis status, age, BMI, PRA, etc.), donor-patient match factors (i.e., cold ischemia time, total number of HLA mismatches, blood type matching) and the current number of patients on a transplant center's waiting list. The $KDPI$ is the kidney donor profile index, which we use to control for donor quality. The variable $Post_t$ is an indicator variable that takes a value of one if the organ offer was made after the first active CMS reporting period (June 30, 2007). The variable $Trig_{tc}$ is an indicator variable that takes a value of

one if transplant center c did not meet the three standards set by CMS in time period t. This variable is defined over both the pre-CMS and post-CMS time periods. The parameter μ_c is a transplant center fixed effect and ρ_m is a month-year dummy, indicating the month and year in which the organ offer was made, that is used to control for time varying characteristics at the monthly level within the transplant community.

By construction the treatment group contains surgeon organ utilization decisions that occur at centers that did not meet the CMS standard in the time period following the activation of the policy. The control group consists of all other surgeon organ utilization decisions, including decisions made at centers which do not meet the CMS standard in the pre-CMS regulation time period. Therefore, our treatment variable, α_4, captures the average treatment effect of a binding CMS regulation.

Our second behavioral hypothesis does not depend on whether or not the organ offer was made when a transplant surgeon's transplant center was above or below the regulatory reference point, $S^{+/-}$. Therefore we can investigate this hypothesis using the following specification (Model 2),

$$\Delta U_{itdc} = X'_{itdc}\beta + \alpha_1 KDPI_d + \alpha_2(KDPI_d > 0.5) + \alpha_3(KDPI_d)(KDPI_d > 0.5) + \mu_c + \rho_m + \varepsilon_{itdc},$$

where all variables are as previously described. In this model the treatment group represents organ offers that were made to a surgeon where the donor quality is lower than reference donor quality, $KDPI = 0.5$, and the control group represents all other organ offers. A limitation of Model 1 and 2 is that they do not account for the other potential behavioral reference point in their investigation of the treatment effects. This is addressed in our empirical tests of hypotheses 3 and 4. To empirically test our third hypothesis we estimate the following triple difference model (Model 3),

$$\begin{aligned}\Delta U_{itdc} = {} & X'_{itdc}\beta + \alpha_1 KDPI_d + \alpha_2(KDPI_d > 0.5) + \alpha_3 Post_t + \alpha_4 Trig_{tc} \\ & + \alpha_5(KDPI_d > 0.5)(Post_t) + \alpha_6(KDPI_d > 0.5)(Trig_{tc}) + \alpha_7(Post_t)(Trig_{tc}) \\ & + \alpha_8(KDPI_d > 0.5)(Post_t)(Trig_{tc}) + \mu_c + \rho_m + \varepsilon_{itdc}.\end{aligned}$$

The coefficient on α_8 captures the impact that being below both the regulatory reference point and the donor quality reference point has on a surgeon's decision to accept an organ. This is suitable for testing our third research hypothesis because it treats the donor reference point as a discrete event in the surgeon's choice space. Our fourth hypothesis is concerned with the marginal impact that donor quality has on surgeon decision-making when a surgeon is below both the regulatory reference point and the donor reference point. This can be investigated by expanding Model 3 to include a continuous measure of donor quality using the following specification (Model 4),

$$\Delta U_{itdc} = X'_{itdc}\beta + \alpha_1 KDPI_d + \alpha_2(KDPI_d > 0.5) + \alpha_3 Post_t + \alpha_4 Trig_{tc}$$
$$+\alpha_5(KDPI_d)(KDPI_d > 0.5) + \alpha_6(KDPI_d)(Post_t) + \alpha_7(KDPI_d)(Trig_{tc}) + \alpha_8(KDPI_d > 0.5)(Post_t)$$
$$+ \alpha_9(KDPI_d > 0.5)(Trig_{tc}) + \alpha_{10}(Post_t)(Trig_{tc}) + \alpha_{11}(KDPI_d)(KDPI_d > 0.5)(Post_t)$$
$$+ \alpha_{12}(KDPI_d)(KDPI_d > 0.5)(Trig_{tc}) + \alpha_{13}(KDPI_d)(Post_t)(Trig_{tc})$$
$$+ \alpha_{14}(KDPI_d > 0.5)(Post_t)(Trig_{tc})$$
$$+\alpha_{15}(KDPI_d)(KDPI_d > 0.5)(Post_t)(Trig_{tc}) + \mu_c + \rho_m + \varepsilon_{itdc}.$$

Using this specification the coefficient α_{15} will capture the marginal effect of $KDPI$ on surgeon's preferences when they are below both behavioral reference points.

Each of the models that use the $Trig_{tc}$ variable is defined in two different ways depending on which of the CMS CoP criteria is being used, either the 1-year graft survival rate or the 1-year patient survival rate. In addition, each of the empirical models is estimated using a linear probability model with the standard errors clustered at the transplant center level. The empirical results will be discussed in the following section.

V. Results

Prior to formally investigating our four research hypotheses, we will illustrate the patient and patient-donor characteristics that influence a transplant surgeon's decision to accept an organ. Table 4 contains the results for the patient and patient-donor characteristics contained in X_{itdc} resulting from our estimation of Model 1.[15] The parameter estimates for the covariates are predominately stable across the four different fixed effect specifications, with a few minor exceptions. Focusing on the parameter estimates that control for center and month-year fixed effects, there are a number of patient and patient-donor characteristics that influence the probability of accepting an organ. Factors that have a negative impact on the acceptance probability are the total number of antigen mismatches observed, a patient's BMI and a patient's albumin level at listing. Factors that have a positive effect on the probability of accepting an organ are whether or not the patient is in poor health at listing (Function Status 3)[16], if the patient: has been pregnant, is on dialysis at listing, had a previous kidney transplant, is HCV positive, is

[15] The parameter estimates observed in Table 4 are consistent across the four different models estimated. Therefore, we have elected to only illustrate the results for Model 1. The complete set of results from the other models can be obtained from the authors.

[16] Functional Status 1 indicates that the patient performs daily activities with no assistance. Functional Status 2 indicates that the patient performs daily activities with some assistance and Status 3 indicates that they require total assistance. The 4th status, and omitted variable, indicates whether or not the status was unobserved.

older, the PRA level at listing, the imputed cold ischemic time and the length of time waiting for an organ.

The importance of antigen mismatches and cold ischemic time were discussed earlier and they warrant further investigation. The negative effect of total mismatches indicates that increasing the number of antigen mismatches by one unit reduces the probability of acceptance by 0.23%. This result is consistent with the antigen mismatch distributions observed in Figure 3. The estimates of the cold ischemic time indicate a positive and diminishing marginal effect on the probability of accepting an organ for values below 39 hours but fall after that. Therefore, up to this time transplant surgeons appear to be comfortable with the cold ischemic time but beyond this time period it begins to adversely affect the organ acceptance decision.

The results from the test of our first research hypothesis are contained in Table 5 (Model 1). The parameter estimates illustrate that transplant surgeon's acceptance rates decrease with a poorer quality donor organ, a higher $KDPI$ value, which is congruent with our expectations. This behavioral phenomenon is consistent across all of the models that we estimate. Focusing on the impact of not being compliant with the standards set forth by CMS, we can clearly see that not being in compliance had no effect on a transplant surgeon's acceptance rate using the results from Model 1. What is apparent in all models is that transplant surgeons unilaterally reduced their acceptance rates immediately following the enactment of CMS CoP. This finding consistently arises when we control for both transplant center fixed effects as well as month-year time varying fixed effects, which suggest that the CMS CoP generated a behavioral shift in their acceptance rates that was independent of whether or not they were in compliance with the CMS CoP triggers or not. However some caution is warranted when we include the month-year fixed effects as this coefficient must be interpreted relative to the omitted month-year dummy variable; the omitted dummy was removed from the regression to prevent colinearity with $Post_t$. The statistically insignificant effect of the CMS CoP triggers in the post-CMS period is an *average* treatment effect; as will become more apparent in our robustness check section this effect is driven by the high degree of behavioral heterogeneity in response to the CMS CoP triggers.

The results from testing our second research hypothesis are contained in Table 6 (Model 2). The parameter estimates illustrate that a higher $KDPI$ value reduces the probability that an organ will be accepted and being below the reference donor, ($KDPI_d > 0.5$), generates a large structural shift in the probability by unilaterally decreasing the probability of acceptance. However, the marginal effect of donor quality on organ acceptance is positive and statistically significant when the donated organ is below the reference donor organ. This does not imply that the effect of poor donor quality is no longer negative, it indicates that it is less negative than when one is above the reference donor organ. The marginal effect of donor quality above the reference donor organ is -0.0246, whereas below the reference donor organ it

is -0.0053. This implies that a 0.10 unit increase in donor quality, moving toward a $KDPI$ of zero, will increase the probability of accepting the organ by 0.025% above the reference donor, but by only 0.005% below the reference donor. The probability estimates for Model 2 illustrate that the probability of accepting an organ is over 70% greater when the donor organ is of the highest quality, $KDPI = 0$, versus when it is from the median donor. Furthermore, the probability of accepting the lowest quality organ, $KDPI = 1$, is approximately 10% below that of the median donor. Clearly, the marginal value of organ quality is substantially larger above the reference donor organ than below it. These results clearly support the presence of a donor reference point in the behavioral heuristic utilized by transplant surgeons in their organ acceptance decisions.

The results from testing our first hypothesis illustrate that on *average* transplant surgeons do not possess a regulatory induced behavioral reference point. This test does not control for whether or not the organ offered to the surgeon is above or below the reference donor organ. Our third research hypothesis does account for the donor organ reference point that our second hypothesis confirms as a behavioral reference point. The parameter estimate that captures this is the triple difference estimation parameter, α_8, in Model 3. The results from this regression are contained in Table 7. For both of the CMS CoP models and across all the different fixed effect specifications, the sign on this coefficient is not statistically significant from zero. Therefore, on *average* the results do not support our third research hypothesis.

The results from estimating Model 4 are contained in Tables 8a and 8b. When we augment Model 3 to determine the impact that being below both behavioral reference points has on a transplant surgeon's marginal preferences over donor quality, our fourth research hypothesis, we observe a similar effect as the triple difference used to investigate our third research hypothesis (see Table 6). Our treatment variable of interest, α_{15}, is statistically insignificant across all of the model specifications and the two CMS CoP criteria. However, the statistically significant coefficients for $\alpha_1, \alpha_2, \alpha_3, \alpha_5, \alpha_6, \alpha_8$ and α_{11} suggest an interesting behavioral response. This behavioral response is illustrated in Figure 4 where we plot the probability of organ acceptance as a function of donor organ quality, $KDPI$, relative to the reference donor organ.

The results for Model 4 prior to the CMS CoPs and when the CMS CoP triggers are not being triggered (far left panel of Figure 4) illustrate that the highest quality organ possesses an acceptance rate that is approximately 125% greater than the reference donor and this probability falls off considerably as the quality of the donor falls. The results in Figure 4 indicate that the probability of accepting the lowest quality organ is only 1.26% lower than the reference donor organ. However, given the lack of statistical significance for a number of the parameters (see Tables 8a and 8b), we observe very wide confidence intervals for the predicted difference in acceptance probabilities relative to the reference donor in the post-CMS period and when the CMS CoP triggers are being triggered. This said, we do observe a similar

pattern of behavioral response to the $KDPI$ as the marginal utility of donor quality above the reference donor organ is greater than when it is below the reference donor organ.

Focusing on the results from the pooled data set it appears that on *average* when a center does not meet the CMS CoP in the post-regulation period the marginal effect is statistically insignificant from zero. Although the *average* results for the CMS CoP triggers do not support the presence of a regulatory reference point, the results from using the median donor quality as a behavioral reference point do support the presence of this reference point. Our results suggests that surgeon perceptions of donor organ quality put more weight on higher quality organs, but once an organ reaches a lower quality they do not respond as strongly to reductions in quality. In essence they do not differentiate nearly as much between an organ that has a $KDPI$ between 0.8 and 0.9 versus if it is between 0.1 and 0.2. The lower quality organs are in general treated more similarly on the margin than the higher quality organs. This would be consistent with an organ threshold model where surgeons target a certain quality of organ for a patient and then disregard any other organ that does not meet the threshold.

As referenced earlier some of the empirical findings are a construct of the pooled data structure. There exists a sizable degree of transplant center heterogeneity that is in fact driving a number of the *average* results observed. To address this in more detail we re-estimate a selection of the models at the transplant center level to illustrate the impact of transplant center heterogeneity. This is discussed in the following section.

VI. Transplant Center Heterogeneity

The estimates illustrated in Tables 4 through 8b represent the average behavioral responses to the patient, patient-donor and donor covariates. Therefore, in testing our behavioral hypotheses we cannot determine whether a particular transplant surgeon behaves differently than another when faced with the same information. Ideally, we would like to be able to utilize surgeon identifiers to investigate the degree of heterogeneity in our behavioral responses. However, surgeon identifiers cannot be obtained for our data set. In our data set we do observe the next best source of behavioral heterogeneity, the transplant center. Transplant centers provide care for a large number of patients and often possess center-level practices that the transplant team, the set of surgeons conducting transplants at the center, have elected to adopt. This transplant center specific behavior is captured by the center fixed effects in the regression models discussed above. However, for a number of the hypotheses investigated there exists a sizable degree of heterogeneity in the center-level behavior. To investigate this we re-ran models 1, 3 and 4 at the center level.[17] The results are contained in Figure 5, 6 and 7, focusing on the treatment coefficients of interest.

[17] We elected to not illustrate the transplant center level estimates for Model 2 because a majority of 54% of the parameter estimates are positive and statistically significant whereas only 19% of the center-level estimates were

The coefficient estimates have been ordered from lowest to highest with 90% confidence intervals calculated using the estimated standard errors.

In Model 1 the coefficient of interest was α_4 and it was statistically insignificant when all the data was pooled. The distribution of the center-level parameter estimates are contained in Table 5. Although a large percentage of the centers still possess a statistically insignificant coefficient for α_4 when investigating the CMS CoP for 1-year graft survival, approximately 62% of the center-level estimates of α_4 are statistically insignificant, there are some strong outliers in the distribution. 15% of the centers possess a statistically significant negative effect, with the largest negative point estimate being -1.62, and 23% possess a statistically significant positive effect with the largest positive point estimate being 1.06. The 1-year patient survival estimates indicate that approximately 73% of the centers possess statistically insignificant responses to the CMS CoP trigger in the post-regulation time period, while 11% have a negative effect (largest is -0.28) and 15% have a positive effect (largest is 0.06). The results from Model 1 confirm why we observe the statistically insignificant effect of the CMS CoP regulation but also illustrate there is a sizable degree of center heterogeneity.

The coefficient on α_8 captures the triple difference of being below both the regulatory and donor reference points. The distribution of center-level parameter estimates, conditional on observing a binding CMS CoP regulation, is illustrated in Figure 6. The degree of center-level heterogeneity observed increases when a center is below both the regulatory and donor reference points. 38% of the centers possess a statistically significant and negative response to the CMS 1-year graft survival triggers when they are below both reference points, 23% are positive and statistically significant and the remaining 39% are not statistically significant. The percentage of statistically significant and positive centers increases to 36% when we investigate the 1-year patient survival CMS triggers and the negative and statistically significant parameter estimates decreases to 24%. These results highlight that there is a subset of centers that behave more aggressively and conservatively when they are below both of our hypothesized reference points. The balance of these behavioral responses counteract each other and generate the statistically insignificant effect in our model investigating the average effect (see Table 7).

A final example of the high degree of heterogeneity that exists within the transplant community is illustrated in our plots of α_{15} for Model 4 when we estimate the model at the center level. This parameter is the triple differenced marginal effect on a surgeon's valuation of donor quality. Clearly, there is a high degree of heterogeneity. In the 1-year graft survival model 23% of the responses are negative, 47% are positive and the remainder statistically insignificant from zero. In the 1-year patient survival model 25% are negative, 54% are positive and the remainder are statistically insignificant from zero. The fact that on

---

negative and statistically significant. For the most part, these results just further confirm the results contained in Table 6.

average the treatment effect is statistically insignificant indicates that the negative responses, weighted by the respective number of organ offers made to these centers, is offset by the corresponding positive effect of the other centers, again weighted by the number of organ offers they receive. This further illustrates that at the transplant center level there appear to be two different strategies to deal with the CMS CoP triggers, which wash out when investigating the *average* treatment effect. One strategy is to become more selective with one's organ utilization criteria in an effort to increase the expected outcomes from the transplant and the other is to become more aggressive in an effort to increase the number of transplants conducted and reduce the impact that any one transplant has on a center's quality evaluations. Either of these strategies supports the presence of our hypothesized behavioral reference points and they further illustrate the high degree of heterogeneity in center-level practices.

VII. Robustness Checks

To investigate the robustness of our results we investigate modifications of Models 1 and 2. In our first robustness check we investigate the marginal impact of the different CMS CoP triggers on surgeon choices. As mentioned earlier, in order to not meet the CMS CoP standards a center must meet all three triggers simultaneously. However, our empirical specification may be too restrictive of a test that the CMS CoP triggers have on surgeon behavior. Centers may exert pressure on their surgeons to curb their utilization decisions when one or any two of the triggers are reached. Therefore, we may be estimating a behavioral response that is too low. To investigate this we ran five additional regressions of the CMS CoP triggers where we re-define the $Trig_{tc}$ to take a value of one if the following triggers are met: (1) trigger 1 (observed-expected > 3), (2) trigger 2 (observed/expected > 1.5), (3) triggers 1 and 2, (4) triggers 1 and 3 (*p-value* < 0.05), and (5) triggers 2 and 3. We do not run the model with trigger 3 separately as it is possible that the statistical significance in the differences could be indicating that a center is performing statistically better than expected, which is the polar opposite of the CMS CoP triggers objective. The results from these five regressions, as well as the corresponding complete trigger model (reproduced from Table 4), are illustrated in Table 9.

The parameter estimates generate very similar results to those observed in Table 5; it does not appear that on *average* any one of the triggers nor combinations of triggers generates a behavioral response. The only exception is the coefficient in Model (1) for 1-year patient survival CMS CoP. For this trigger, arising when the observed minus expected is greater than three, centers appear to be more aggressive with their utilization of organs. However, when we combine this trigger with either the second or third trigger it does not have any effect on the utilization rate. Again, these results are the *average* treatment effects observed. As was the case when all three triggers are met, see Table 4, these results may arise from the degree of heterogeneity in the population. To formally investigate this, Figures 8a and 8b

graphically illustrate the distribution of $\alpha_4's$ by trigger type when we estimate center-specific coefficients for α_4 for the 1-year patient graft survival and 1-year patient survival CMS CoP triggers respectively. The figures clearly illustrates that the *average* treatment effect is driven by the high degree of transplant center heterogeneity. For the 1-year patient graft survival triggers 35% possess a statistically significant and negative response to trigger 1, 29% possess a statistically positive response and the remaining are not statistically significant from zero. The other triggers all illustrate a similar pattern, where approximately a third of the centers possess a statistically significant and negative response, a third possess a statistically positive and significant response and the remaining third are not statistically significant from zero. The results from the 1-year patient survival models, contained in Figure 8b, illustrate a similar pattern of behavioral responses at the transplant center level with a slightly smaller amount of the distributional mass in the tails of the distribution and more mass focused in the null value range.

The CMS CoP regulations provide a concrete reference point for our analysis as it is dictated by an external regulatory body. The $KDPI$ reference point, on the other hand, is more subjective. Although a donor with a $KDPI$ value of 0.5 is referred to as the "reference donor" by UNOS and is constructed to be the median donor, this does not necessarily imply that this value is the "true" reference point as transplant surgeons may elect to focus on a different level as the reference point. To investigate the subjectivity of our assumption regarding the reference donor we estimated 100 different regressions by varying our definition of the reference donor, changing our hypothesized behavioral reference point, from 0.01 to 1.0 using intervals of 0.01. For each regression we then plotted out the probability of acceptance when the $KDPI$ increases from zero to one, again using intervals of 0.01. The scatter plot of these points are illustrated in Figure 10, along with a cubic polynomial fit to investigate whether or not our assumed reference point at the median $KDPI$ value (0.5) is valid or if the reference point arises somewhere else in the distribution of values.

The cubic polynomial fit in Figure 9 suggests that the median value (KDPI=0.5) separate the flatter part of the cubic fit from the steeper one. This provides some validation of our selection of the reference donor organ. The plot of data points illustrates a very similar pattern to the one in Figure 4. Transplant surgeons put a much higher weight on donor organ quality when the donor is above the reference donor threshold but then as one progresses beyond the reference donor the impact of donor quality on the acceptance decision is considerably smaller. These responses are consistent with a behavioral hypothesis that transplant surgeons possess threshold values for donor quality where if an organ falls below the threshold, in this case the reference donor, they treat the organ similarly to any organ of a lower quality. However, if the donor quality is perceived to be above the reference donor a considerable amount of weight is assigned to marginal changes in donor quality.

Combined, these robustness checks have illustrated that although *on average* transplant centers do not appear to treat the CMS CoP triggers as behavioral reference points, this is not true when we investigate this phenomenon at the center level. In the distribution of transplant centers there appears to be three different types of centers, those who respond negatively, those who respond positively and those that do not respond at all. Furthermore, for the most part, there appears to be a balance of these types across the distribution of transplant centers. The results for the donor organ reference point are substantially stronger and our robustness check illustrates that a transplant surgeon's behavioral responses when we alter the reference point generates a similar profile to that observed when we assume the reference donor is the median donor. This validates our selection of the median donor as a behavioral reference point.

VII. Conclusion

The amount of information available to a transplant surgeon is immense and it must be processed in a relatively short period of time when deciding whether or not an organ is suitable for their patient. To date very little empirical research has been conducted on the effect of behavioral reference points on the likelihood of a deceased donor organ being accepted by transplant surgeons. This paper reports the first empirical analysis of this decision process using a highly detailed data set on both the patient offered the organ and the donor providing the organ. Our results illustrate that transplant surgeon decisions are consistent with the presence of behavioral reference points. Although donor quality and a number of patient and patient-donor characteristics impact the decision to utilize an organ, this decision environment is characterized by two behavioral reference points. One reference point is based on the quality of the donor organ offered for transplantation, and the other is induced by the regulatory regime. We find that the median donor quality provides a significant reference point for transplant surgeons and *on average* the regulatory environment does not appear to provide a behavioral reference point. However, when we investigate the robustness of this average effect we find a number of centers that respond positively and other centers negatively to the CMS CoP triggers. These two effects, combined with a number of statistically insignificant responses, negate each other and generate a statistically insignificant average effect.

From a policy perspective our findings suggest that the response to the behavioral reference points exhibited by transplant surgeons may have a profound effect on the utilization of deceased donor organs. First, it is erroneous to assume that transplant centers do not respond to the CMS CoP regulations, as suggested by our average treatment effects, as there exists a high degree of transplant center heterogeneity. A number of centers behave more conservatively or aggressively in response to the CMS CoP triggers. This suggests that transplant surgeons follow one of three strategies when the CMS CoP

triggers are activated: (1) increase the number of acceptances, (2) decrease the number of acceptances, and (3) no change to their decision process. We leave it open for future research to ascertain the rationality of each of these strategies in the presence of the quality metrics set forth by CMS, but our results do illustrate that there is not a consensus on the best response to the regulations. Furthermore, these behavioral responses are exacerbated when a transplant surgeon is offered an organ below the reference donor at the time that their center is not compliant with the CMS CoP criteria.

The donor organ quality reference point results indicate that the marginal value of increased organ quality changes when the quality of the organ crosses the reference donor organ. Given that the outcomes measured to monitor the performance of a transplant center are based on risk adjusted donor organs, our results indicate that the risk adjustments used by SRTR may be incongruous with the objectives of increased deceased donor organ utilization. If the risk adjustments used in the transplant community accurately addressed the degree of heterogeneity in donor organ quality then reference dependent behavior would be offset by the risk adjustments. Those organs that are perceived of as a lower quality relative to the reference organ would be adjusted to make their marginal utility equal to the higher quality organs above the reference donor organ. If this were done it might generate an increase in the number of deceased donor organs that are utilized by transplant surgeons and increase the quality of life for many patients currently waiting for an organs. It is definitely too early to conclude on the robustness of this conjecture, but the possibility of it being true does warrant further investigation into the complex decision making of transplant surgeons.

References

Abecassis, M.M., Burke, R., Cosimi, A.B., Matas, A.J., Merion, R.M., Milman, D., Roberts, J.P. and G.B. Klintmalm. 2008. Transplant Center Regulations – A Mixed Blessing? An ASTS Council Viewpoint. *American Journal of Transplantation* 8: 2496-2502.

Abeler, J., Falk, A., Goette, L. and D. Huffman. 2011. Reference Points and Effort Provision. *American Economic Review* 101: 470-92.

Alagoz, O., L.M. Maillart, A.J. Schaefer, M.S. Roberts. 2007. Choosing among living-donor and cadaveric livers. *Management Science* 53(11): 1702-15.

Alagoz, O., L.M. Maillart, A.J. Schaefer, M.S. Roberts. 2004. The optimal timing of living-donor liver transplantation. *Management Science* 50(10): 1420-30.

Axelrod, D.A., Kalbfleisch, J.D., Sun R.J., Guidinger, M.K., Biswas, P, Levine, G.N. et al. 2009. Innovations in the Assessment of Transplant Center Performance: Implications for Quality Improvement. *American Journal of Transplantation* 9(4 Pt 2): 959-69.

Camerer, C., Babcock, L., Loewenstein, G. and R. Thaler. 1997. Labor Supply of New York Cabdrivers: One Day at a Time. *Quarterly Journal of Economics* 112(2): 407-441.

Chang, T. and T. Gross. 2014. How Many Pears Would a Pear Packer Pack if a Pear Packer Could Pack Pears at Quasi-Exogenously Varying Piece Rates? *Journal of Economic Behavior and Organization*. 99: 1-17.

Crawford, V.P. and J. Meng. 2011. New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income. *American Economic Review* 101(5): 1912-32.

De Borger, B. and M. Fosgerau. 2008. The Trade-Off Between Money and Travel Time: A Test of the Theory of Reference-Dependent Preferences. *Journal of Urban Economics* 64: 101-115.

Deck, C. and E.O. Kimbrough. 2012. Can Markets Save Lives? An Experimental Investigation of a Futures Market for Organs. *Working Paper, Department of Economics, Arkansas University*.

Dickinson, D.M., Arrington, C.J., Fant, G., Levine, N., Schaubel, D.E., Pruett, T.L., Roberts, M.S. and R.A. Wolfe. 2008. SRTR Program-Specific Reports on Outcomes: A Guide for the New Reader. *American Journal of Transplantation* 8: 1012-26.

Dickinson, D.M., Shearon, T.H., O'keefe, J.O., Wong, H. –H., Berg, C.L., Rosendale, J.D., Delmonico, F.L., Webb, R.L. and R.A. Wolfe. 2006. SRTR Center-Specific Reporting Tools: Posttransplant Outcomes. *American Journal of Transplantation* 6: 1198-1211.

Farber, H.S. 2005. Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers. *Journal of Political Economy* 113(1): 46-82.

Farber, H.S. 2008. Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers. *American Economic Review* 98(3): 1069-1082.

Fehr, E. and L. Goette. 2007. Do Workers Work More if Wages Are High? Evidence From a Randomized Field Experiment. *American Economic Review* 97(1): 298-317.

Genesove, D. and C. Mayer. 2001. Loss Aversion and Seller Behavior: Evidence from the Housing Market. *The Quarterly Journal of Economics* 116(4): 1233-60.

Hardie, B.G.S., Johnson, E.J. and P.S. Fader. 1993. Modeling Loss Aversion and Reference Dependence Effects on Brand Choice. *Marketing Science* 12(4): 378-394.

Howard, D. H. 2002. Why Do Transplant Surgeons Turn Down Organs? A Model of the Accept/Reject Decision. *Journal of Health Economics* 21:957-969.

Howard, D.H., Siminoff, L.A., McBride, V. and M. Lin. 2007. Does quality improvement work? Evaluation of the organ donation Breakthrough Collaborative. *Health Services Research* 42(6): 2160-2172.

Howard, R.J., Cornell, D.L. and J.D. Schold. 2009. CMS oversight, OPOs and transplant centers and the law of unintended consequences. *Clinical Transplantation* 23: 778-83.

Imbens, G.W. and D.R. Rubin. Forthcoming. *Causal Inference for Statistics, Social and Biomedical Sciences: An Introduction*. Cambridge University Press, Cambridge and New York.

Imbens, G.W. and J.M. Wooldridge. 2009. Recent Advancements in the Econometrics of Program Evaluation. *Journal of Economic Literature* 47(1): 5-86.

Kessler, J.B. and A.E. Roth. 2012. Organ Allocation Policy and the Decision to Donate. *American Economic Review* 102(5): 2018-47.

Köszegi, B. and M. Rabin. 2006. A Model of Reference-Dependent Preferences. *Quarterly Journal of Economics* 121(4): 1133-65.

Li, D., Hawley, Z. and K. Schnier. 2013. Increasing Organ Donation via Changes in the Default Choice or Allocation Rule. *Journal of Health Economics* 32(6): 1117-1129.

Leichtman, A.B., Cohen, D., Keith, D., O'Connor, K., Goldstein, M., McBride, V., Gould, C.J., Christensen, L.L. and V.B. Ashby. 2008. Kidney and pancreas transplantation in the United States, 1997-2006: The HRSA Breakthrough Collaboratives and the 58 DSA Challenge. *American Journal of Transplantation* 8(2): 946-57.

Massie, A.B., Zeger, S.L., Montgomery, R.A. and D.L. Segev. 2009. The Effects of DonorNet 2007 on Kidney Distribution Equity and Efficiency. *American Journal of Transplantation* 9(7): 1550-57.

Organ Procurement and Transplantation Network (OPTN). 2014. http://optn.transplant.hrsa.gov/latestData/step2.asp? Last accessed on August 13, 2014.

Pope, D.G. and M.E. Schweitzer. 2011. Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes. *American Economic Review* 101: 129-157.

Post, T., van den Assem, M.J., Baltussen, G. and R.H. Thaler. 2008. Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show. *American Economic Review* 98(1): 38-71.

Punch, J.D., Hayes, D.H., LaPorte, F.B., McBride, V. and M.S. Seely. 2007. Organ donation and utilization in the United States, 1996-2005. *American Journal of Transplantation* 7(2): 1327-1338.

Rao, P.S., Schaubel, D.E., Guidinger, M.K., Andreoni, K.A., Wolfe, R.A., Merion, R.M., Port, F.E. and R.S. Sung. 2009. A Comprehensive Risk Quantification Score for Deceased Donor Kidneys: The Kidney Donor Risk Index. *Transplantation* 88(2): 231-236.

Rizzo, J.A. and R.J. Zeckhauser. 2003. Reference Incomes, Loss Aversion, and Physician Behavior. *The Review of Economics and Statistics* 85(4): 909-22.

Roth, A.E., Sonmez, T. and M.U. Unver. 2004. Kidney Exchange. *Quarterly Journal of Economics* 119(2): 457-488.

Roth, A.E., Sonmez, T. and M.U. Unver. 2005. A Kidney Exchange Clearinghouse in New England. *American Economic Review, Papers and Proceedings* 95(2): 376-380.

Roth, A.E., Sonmez, T. and M.U. Unver. 2007. Efficient Kidney Exchange: Coincidence of Wants in Markets with Compability-Based Preferences. *American Economic Review* 97(3): 828-851.

Schnier, K.E., McIntyre, C., Ruhil, R., Sadiraj, V., Cox, J.C. and N. Turgeon. 2013. Transplantation at the Nexus of Behavioral Economics and Healthcare Reform. *American Journal of Transplantation* 13: 31-35.

Schnier, K.E., McIntyre, C., Ruhil, R., Sadiraj, V., Cox, J.C., Ouayogode, M., Pearson, T.C., Kirk, A.D. and N.A. Turgeon. 2014. The Association Between CMS CoP Review Criteria and Deceased Donor Kidney Transplant Waiting Times. *Working Paper, School of Social Sciences, Humanities and Arts, University of California-Merced.*

Schold, J., Buccini, L., Srinivas, T., Srinivas, R., Poggio, E., Flechner, S., Soria, C., Segev, D., Fung, J. and D. Goldfarb. 2012. The Association of Center Performance Evaluations and Kidney Transplant Volume in the United States. Forthcoming in *American Journal of Transplantation*.

Shafer, T.J., Wagner, D., Chessare, J., Schall, M.W., McBride, V., Zampiello, F.A., Perdue, J., O'Connor, K., Lin, M.J. and J. Burdick. 2008. US organ donation breakthrough collaborative increases organ donation. *Critical Care Nurse* 31(3): 190-210.

Sung, R.S., Galloway, J., Tuttle-Newhall, J.E., Mone, T., Laeng, R., Freise, C.E. and P.S. Rao. 2008. Organ donation and utilization in the United States, 1997-2006. *American Journal of Transplantation* 8(2): 922-34.

Tversky, A. and D. Kahneman. 1991. Loss Aversion in Riskless Choice: A Reference-Dependent Model. *Quarterly Journal of Economics* 106(4): 1039-61.

Wolfe, R.A., Roys, E.C. and R.M. Merion. 2010. Trends in Organ Donation and Transplantation in the United States, 1999-2008. *American Journal of Transplantation* 10(4 Pt. 2): 961-72.

Figure 1: Distribution of Kidney Donor Risk Index ($KDRI$) and Kidney Donor Profile Index ($KDPI$). Distributions are determined using the 55,360 unique donors in the data set and are not weighted according to their observance within the match run offer data set.

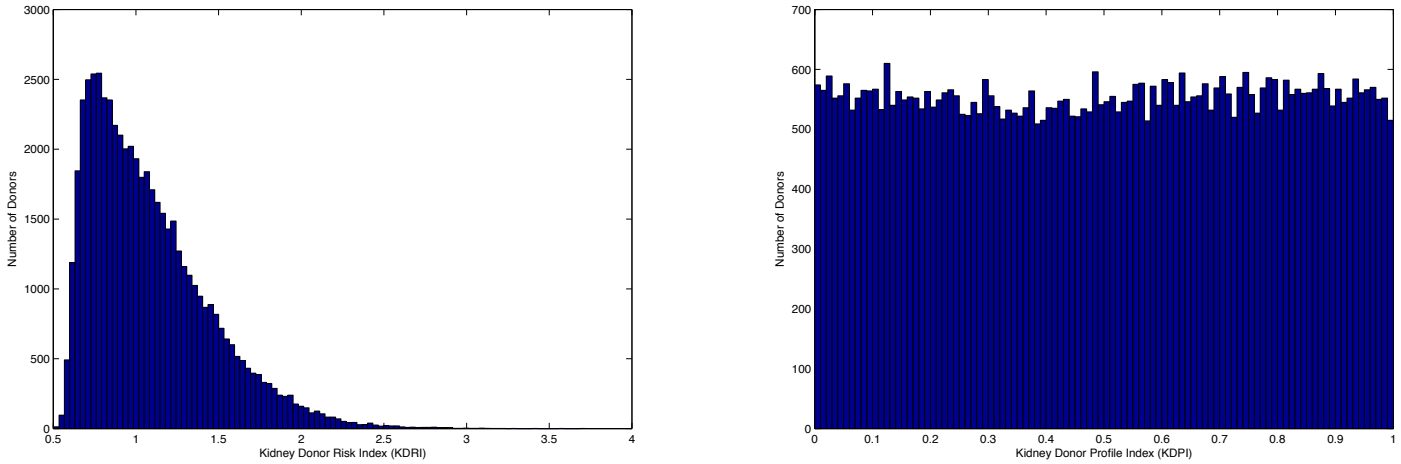

Figure 2: Q-Q Plot of Actual and Predicted Cold Ischemic Time. The X-axis represents that actual cold ischemic time observed for a transplant, whereas the predicted (Y-axis) is expected value using our regression model. The dashed line is a 45-degree line indicating equality in the distributions at the respective quantile comparison.

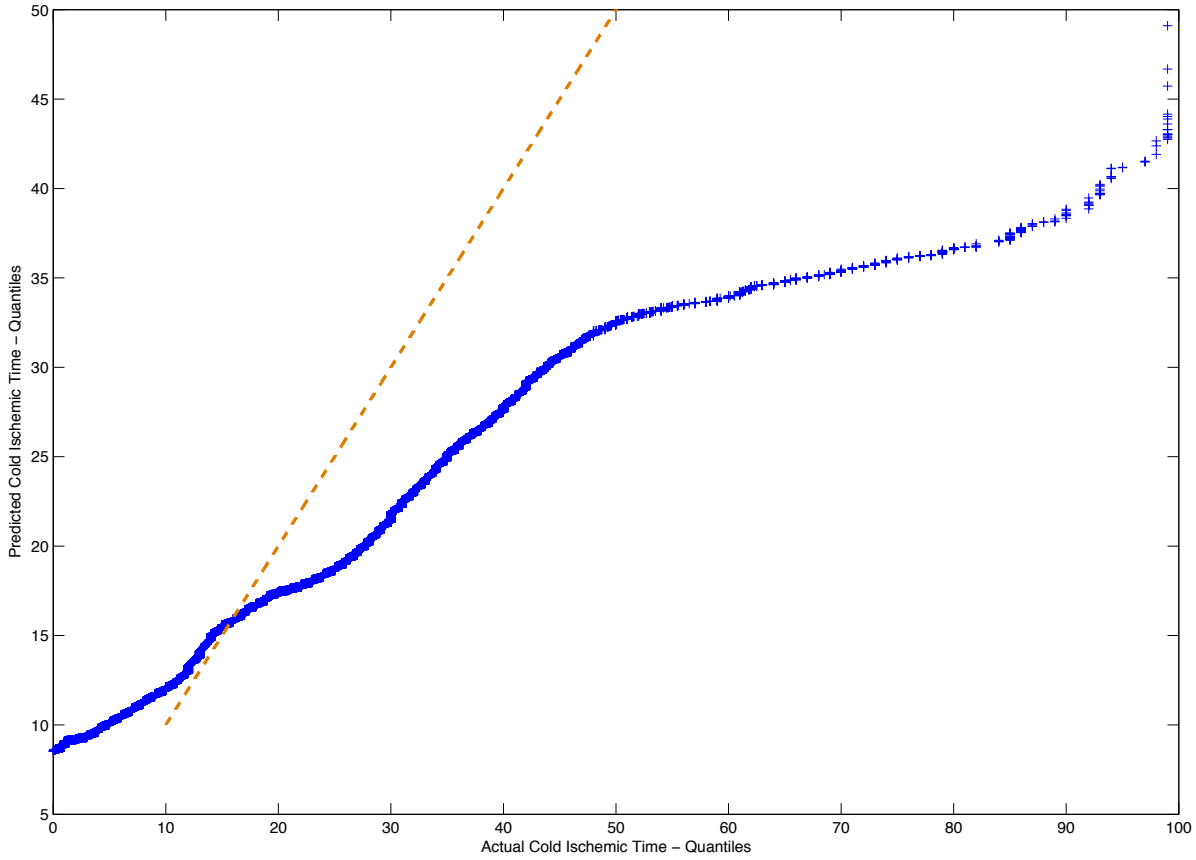

Figure 3: Histocompatibility plot by acceptance and rejection. Percentage of acceptances (blue) and rejections (orange) by number of HLA mismatches.

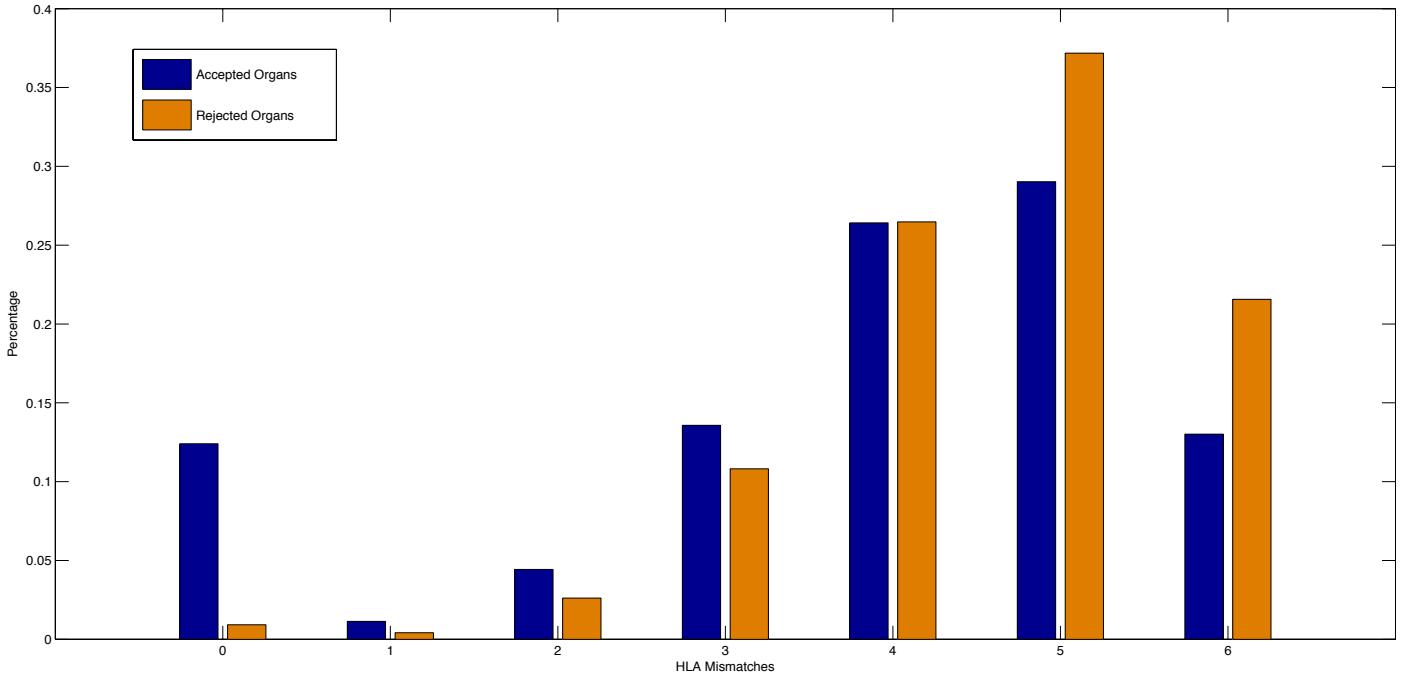

Figure 4: Graphical illustration of the marginal effects observed for Model 4 using the parameter estimates from the 1-year patient survival triggers.

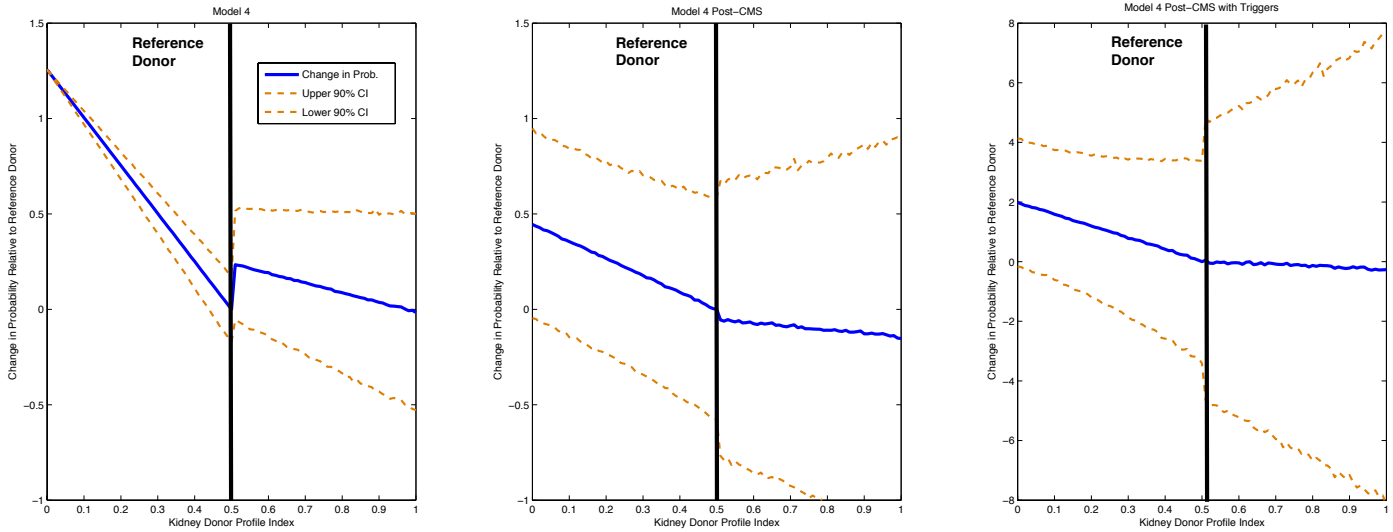

Figure 5: Plot of the transplant center coefficients for α_4 resulting from estimating Model 1 for each of the transplant center's data. The left panel represents the 1-year graft survival CMS CoP triggers and the right panel represents the 1-year patient survival CMS CoP triggers.

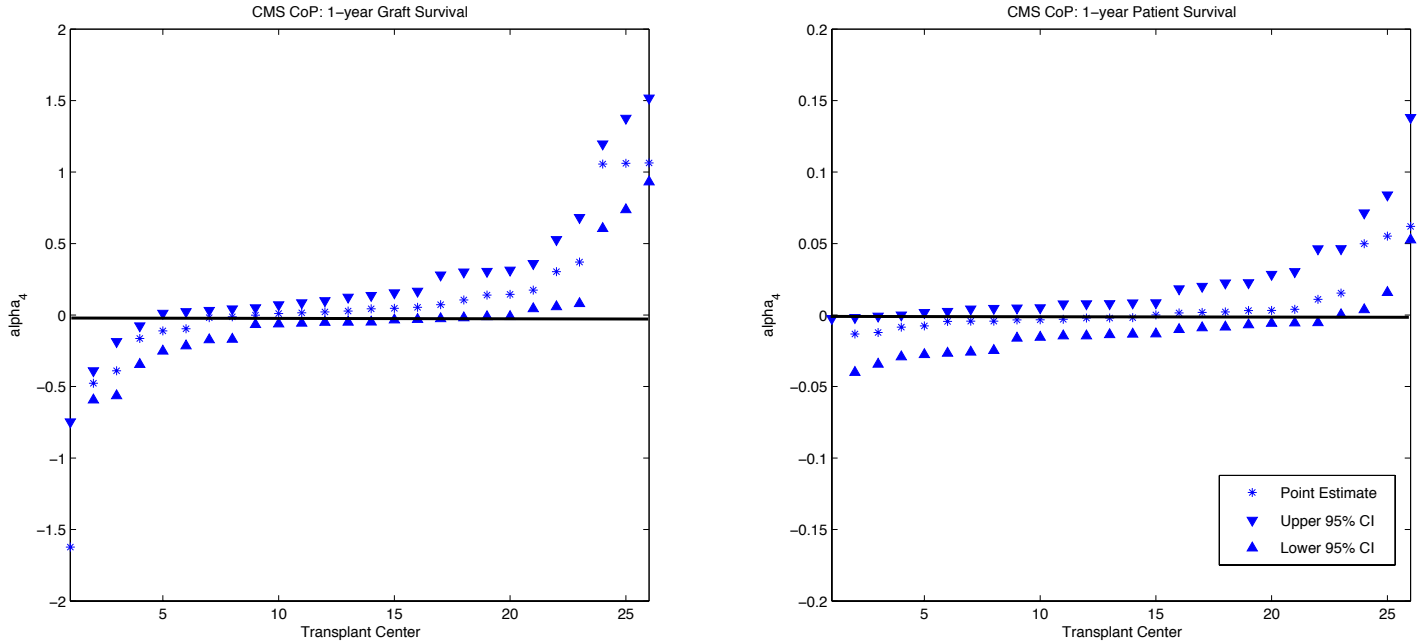

Figure 6: Plot of the transplant center coefficients for α_8 resulting from estimating Model 3 for each of the transplant center's data. The left panel represents the 1-year graft survival CMS CoP triggers and the right panel represents the 1-year patient survival CMS CoP triggers.

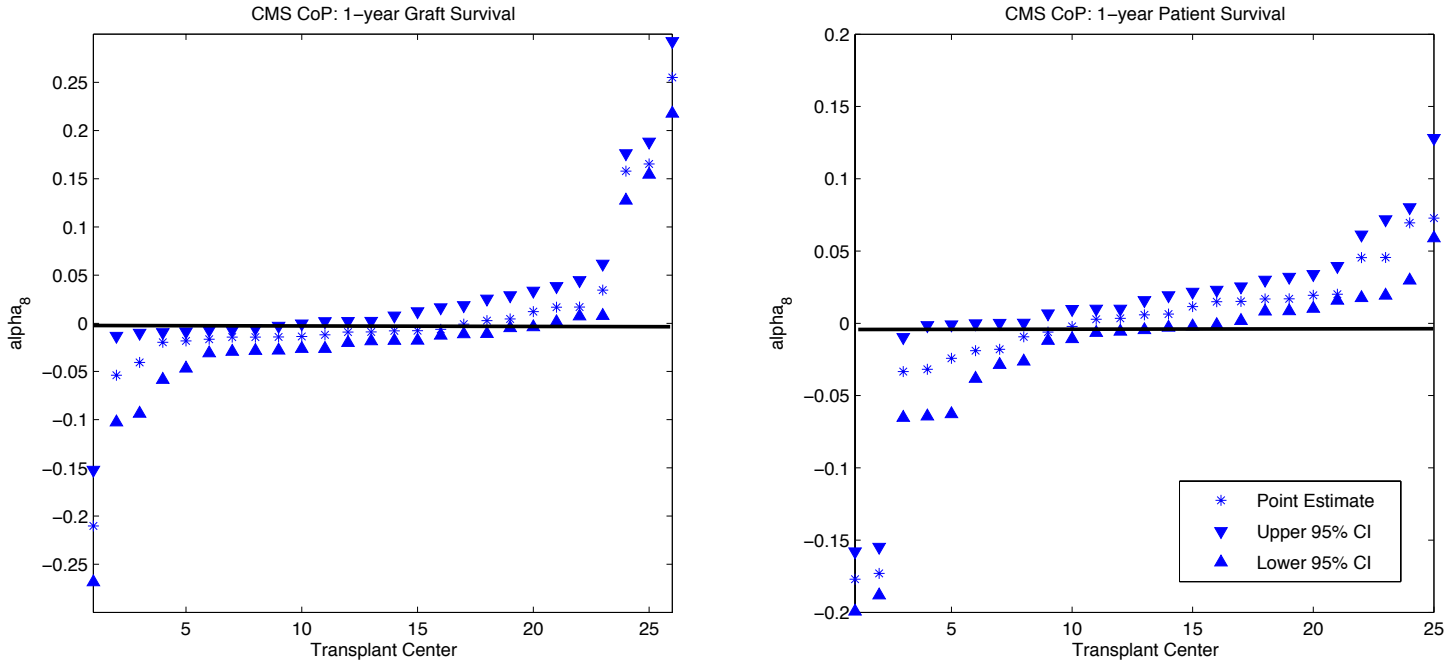

Figure 7: Plot of the transplant center coefficients for α_{15} resulting from estimating Model 4 for each of the transplant center's data. The left panel represents the 1-year graft survival CMS CoP triggers and the right panel represents the 1-year patient survival CMS CoP triggers.

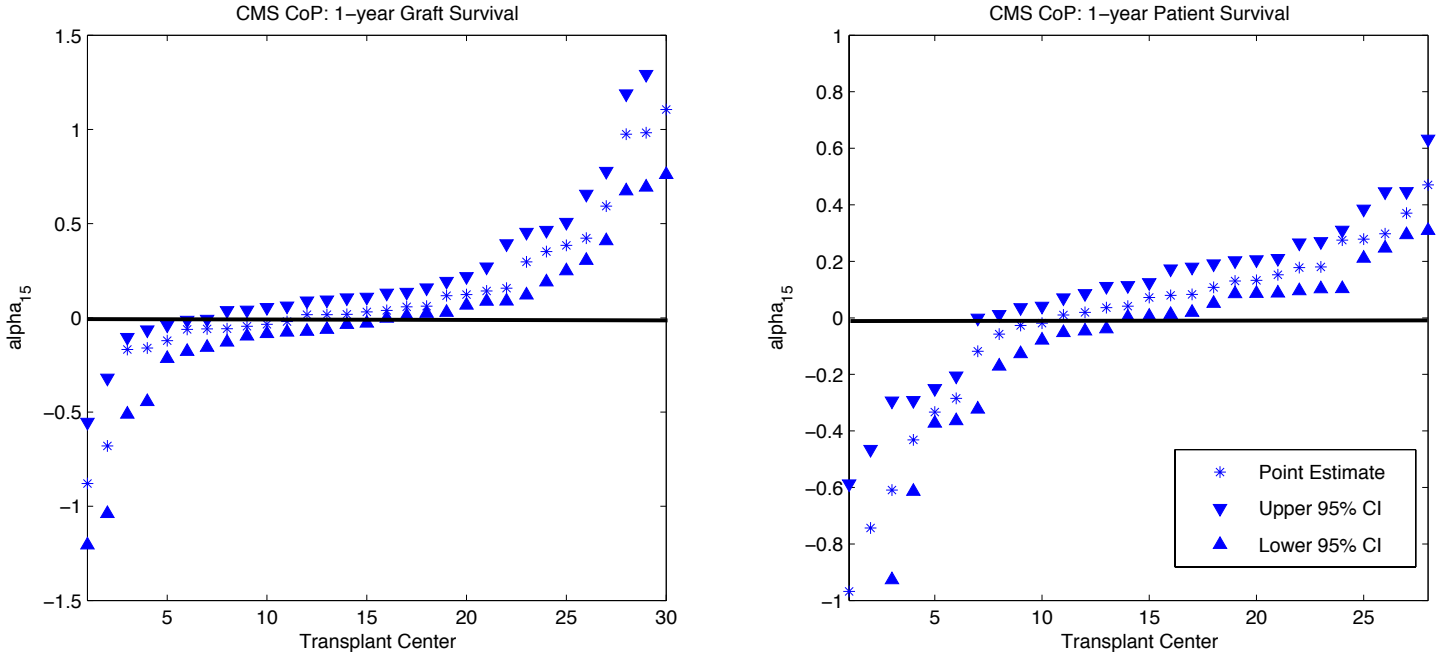

Figure 8a: Model 1 results broken down by triggers for the 1-year graft survival CMS CoP triggers. The triggers are (1) Observed-Expected>3; (2) Observed/Expected>1.5; and (3) p-value less than 0.05. Each panel indicated represents the center-level coefficient for α_4 with 90% confidence intervals plotted.

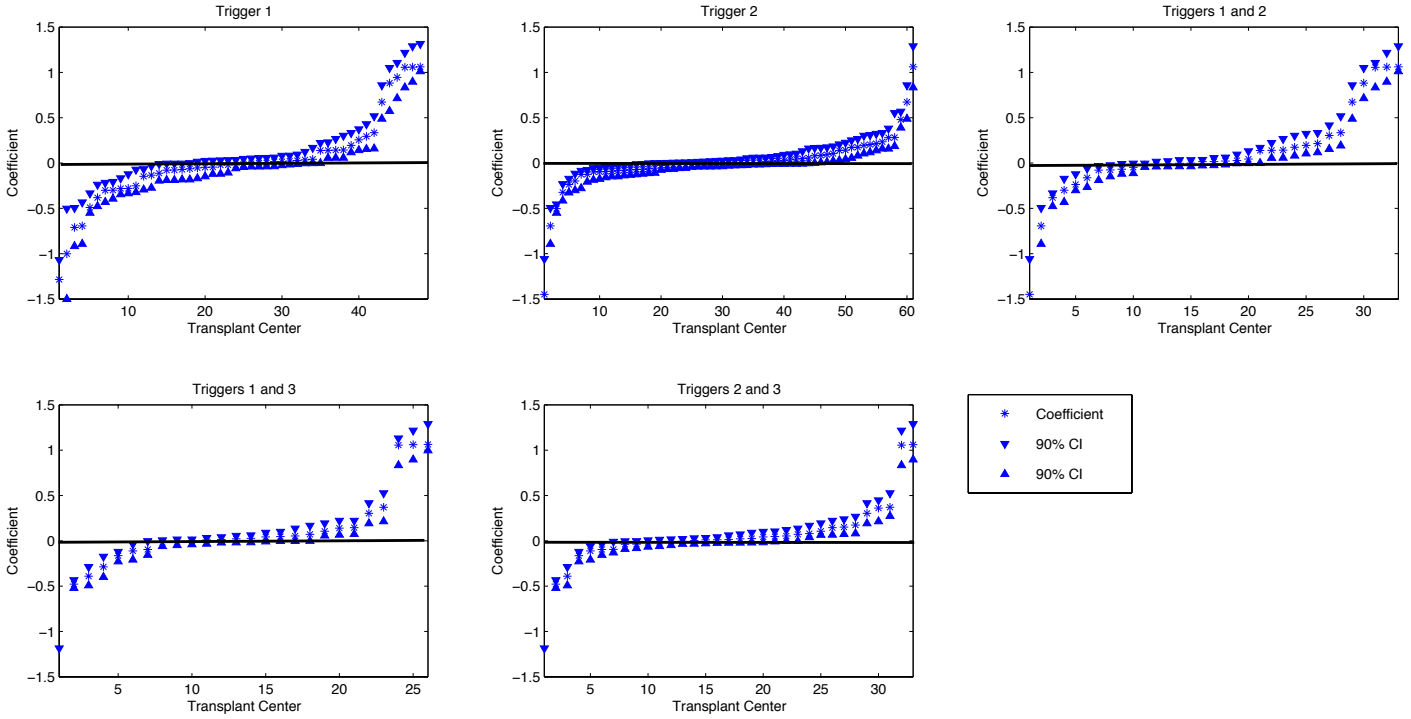

Figure 8b: Model 1 results broken down by triggers for the 1-year patient survival CMS CoP triggers. The triggers are (1) Observed-Expected>3; (2) Observed/Expected>1.5; and (3) p-value less than 0.05. Each panel indicated represents the center-level coefficient for α_4 with 90% confidence intervals plotted.

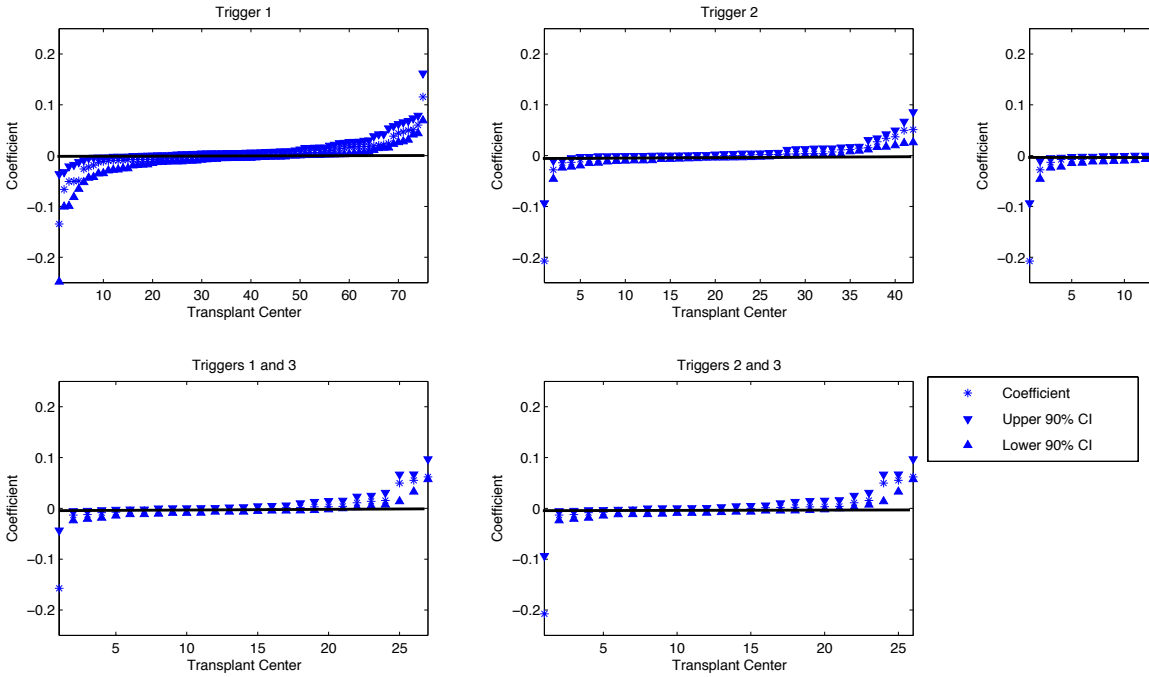

Figure 9: Plot of Model 3 varying the reference donor threshold from 0.01 to 1 using an interval of 0.01. The x-axis represents the $KDPI$ value and the y-axis the marginal effect of $KDPI$ on organ acceptance.

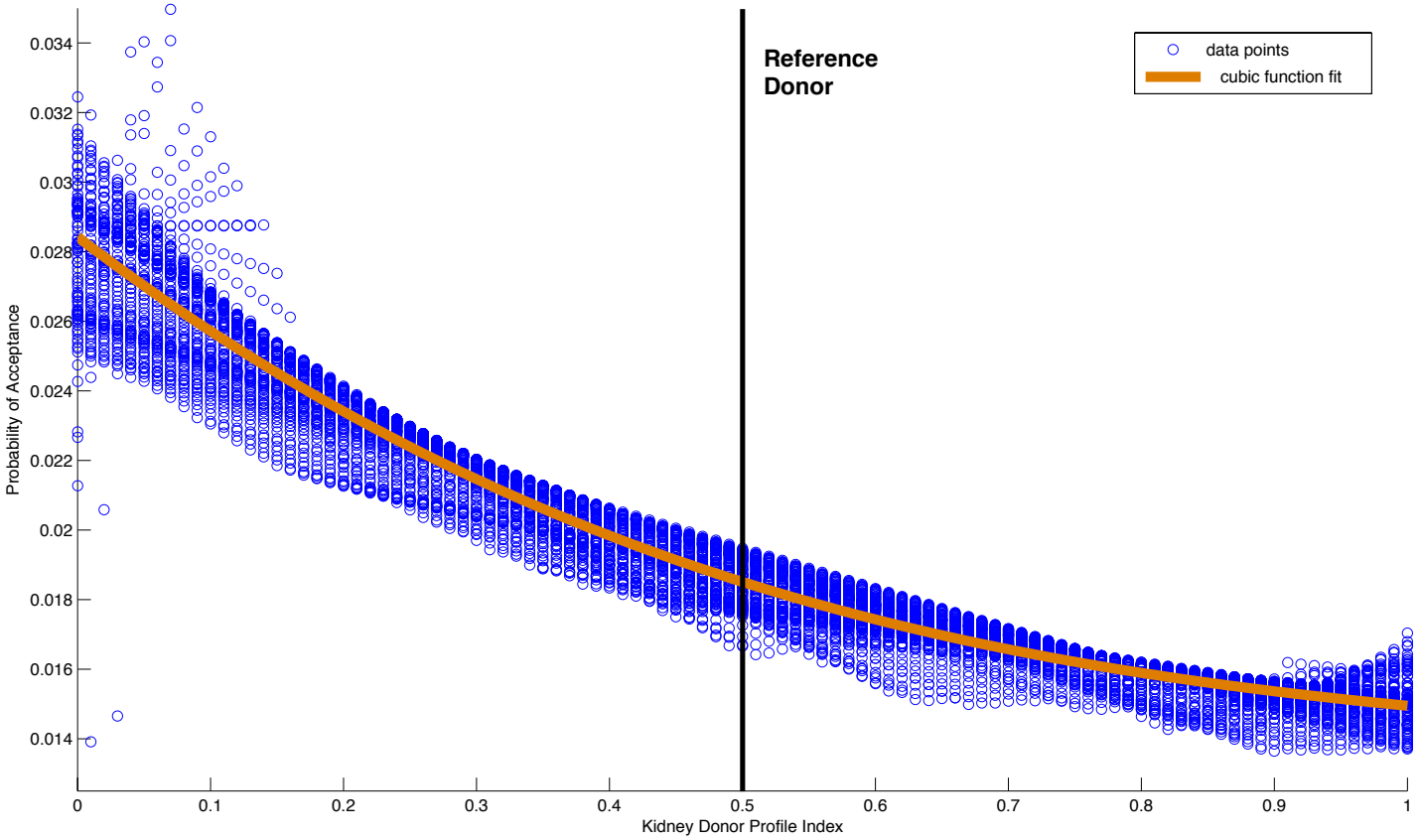

Table 1: Cold ischemia time regression. Dependent variable is the log of cold ischemia time reported in the UNOS data set for all transplants conducted between July 1, 2003 and December 30, 2012 for the transplant center within our analysis. The omitted HRSA region is Region 1 and the omitted death mechanism is natural causes.

Variable	Coefficient	Stand. Error	Variable	Coefficient	Stand. Error
Constant	2.4898***	0.2354	Region 3	0.2960***	0.0113
Log(Seq. Num.)	0.0172	0.0120	Region 4	0.2542***	0.0119
Log(Time Trans.)	0.4791***	0.0686	Region 5	0.1465***	0.0112
Log(Dnr. BMI)	-0.1847	0.1437	Region 6	0.1573***	0.0139
Log(Dnr. Age)	-0.0653*	0.0390	Region 7	0.2493***	0.0121
$(\text{Log(Seq. Num)})^2$	0.0074***	0.0004	Region 8	0.0551***	0.0123
$(\text{Log(Time Trans.)})^2$	0.3611***	0.0200	Region 9	0.3078***	0.0122
$(\text{Log(Dnr. BMI)})^2$	0.0322	0.0239	Region 10	0.1831***	0.0119
$(\text{Log(Dnr. Age)})^2$	0.0102***	0.0029	Region 11	0.2986***	0.0116
Log(Seq. Num)Log(Time Trans.)	-0.0467***	0.0024	Suicide Death	-0.0048	0.0065
Log(Seq. Num)Log(Dnr. BMI)	-0.0057	0.0039	Homicide Death	-0.0075	0.0077
Log(Seq. Num)Log(Dnr. Age)	0.0080***	0.0015	Child Abuse Death	0.0232	0.0222
Log(Time Trans.)Log(Dnr. BMI)	-0.0157	0.0223	Non-Motor Veh. Death	-0.0011	0.0063
Log(Time Trans.)Log(Dnr. Age)	-0.0293***	0.0082	Natural Causes Death	0.0017	0.0043
Log(Dnr. BMI)Log(Dnr. Age)	0.0021	0.0144	Non-Heart Beating Donor	0.0473***	0.0055
Region 2	0.1966***	0.0112	ECD Kidney	0.0265***	0.0057

Number of Observations	87,681	R-squared	0.1992

*** indicates statistically significant at the 99%; ** indicates statistically significant at the 95% level%; * indicates statistically significant at the 90% level.

Table 2: Descriptive Statistics broken down by the CMS CoP triggers. $Trig_t = 1$ implies that either the 1-year graft survival CMS CoPs or the 1-year patient survival triggers are met by a center. There are total of 1,673,350 observations in the Pre-CMS CoP period and 3,624,805 observations in the Post-CMS CoP period. The averages are reported in the respective columns with the standard deviations reported in parentheses. The normal difference is reported in the Norm. Diff. column.

	Pre-CMS CoP			Post-CMS CoP		
Variable	$Trig_t = 0$	$Trig_t = 1$	Norm. Diff.	$Trig_t = 0$	$Trig_t = 1$	Norm. Diff.
Acceptance Rate	0.0215	0.0287	-0.0325	0.0157	0.0199	-0.0227
	(0.1451)	(0.1669)		(0.1242)	(0.1397)	
Total HLA Mismatches	4.4434	4.4347	0.0053	4.6569	4.4612	0.1092
	(1.1675)	(1.1690)		(1.1076)	(1.4085)	
BMI Donor	26.6552	26.8593	-0.0201	27.5180	27.8729	-0.0324
	(7.0199)	(7.3153)		(7.6163)	(7.8850)	
Age Donor	41.6298	40.7173	0.0349	41.4406	40.9615	0.0199
	(18.5432)	(18.4197)		(16.8635)	(17.1311)	
KDPI	0.6604	0.6415	0.0510	0.6246	0.6170	0.0201
	(0.2587)	(0.2636)		(0.2688)	(0.2660)	
Cold Ischemic Time	3.1659	3.2537	-0.0238	3.1234	3.1632	-0.0124
	(2.4122)	(2.7866)		(2.2528)	(2.3019)	
ECD Donor	0.2467	0.2309	0.0262	0.2138	0.2268	-0.0221
	(0.4311)	(0.4214)		(0.4100)	(0.4187)	
BMI Patient	27.0333	27.2946	-0.0438	27.1028	27.2462	-0.0296
	(3.8384)	(4.5761)		(3.2492)	(3.5995)	
Age Patient	48.7244	47.9451	0.0416	51.0150	50.0180	0.0541
	(13.2683)	(13.2245)		(13.0203)	(13.0644)	
Patient Diabetes Type 1	0.0342	0.0266	0.0314	0.0574	0.0525	0.0150
	(0.1818)	(0.1609)		(0.2325)	(0.2231)	
Patient Diabetes Type 2	0.1103	0.1234	-0.0289	0.3034	0.2980	0.0084
	(0.3132)	(0.3288)		(0.4597)	(0.4574)	
Patient Dialysis	0.8038	0.8348	-0.0570	0.7547	0.7782	-0.0393
	(0.3971)	(0.3713)		(0.4303)	(0.4154)	
% Obs. $KDPI_d > 0.5$	0.7301	0.6963	0.0530	0.6813	0.6663	0.0227
	(0.4493)	(0.4599)		(0.4660)	(0.4715)	
% Obs. $Trig_t$			7.61%			7.03%

Table 3: Descriptive Statistics broken down by whether or not the organ offer was made when the donor was above the reference donor's organ quality ($KDPI_d \leq 0.5$) or above it ($KDPI_d > 0.5$). There are total of 1,614,785 observations with the $KDPI \leq 0.5$ and 3,683,370 observations with the $KDPI > 0.5$. The averages are reported in the respective columns with the standard deviations reported in parentheses. The normal difference is reported in the Norm. Diff. column.

Variable	$KDPI_d \leq 0.5$	$KDPI_d > 0.5$	Normalized IW
Acceptance Rate	0.0276 (0.1639)	0.0136 (0.1159)	0.0698
Total HLA Mismatches	4.5567 (1.1789)	4.5900 (1.1333)	-0.0203
BMI Donor	26.7044 (6.6816)	27.5143 (7.7705)	-0.0790
Age Donor	28.8702 (10.0317)	46.9727 (17.10153)	-0.9129
KDPI	0.2940 (0.1386)	0.7846 (0.1419)	-2.4729
Cold Ischemic Time	3.2202 (2.7326)	3.1061 (2.1081)	0.0330
ECD Donor	0.0000 (0.0122)	0.3228 (0.4675)	-0.6898
BMI Patient	27.0674 (3.4708)	27.1057 (3.4909)	-0.0078
Age Patient	49.2922 (13.4609)	50.6337 (12.9833)	-0.0717
Patient Diabetes Type 1	0.0487 (0.2151)	0.0501 (0.2181)	-0.0046
Patient Diabetes Type 2	0.2433 (0.4291)	0.2421 (0.4284)	0.0021
Patient Dialysis	0.7647 (0.4242)	0.7754 (0.4173)	-0.0179
Obs. $Trig_t$	0.0766 (0.2660)	0.0702 (0.2554)	0.0175
% Obs. $KDPI_d > 0.5$	69.52%		

Table 4: Estimation results for Model covariate controls, X_{itdc}, in Model 1 using the 1-year graft survival CMS CoP triggers. There are a total of 5,298,155 observations. The errors are clustered at the transplant center level. All regressions contain the following controls: patient diagnosis code, patient ethnicity, donor death circumstances, blood mismatches, and the other treatment variables in the regression. Treatment variables are contained in Table 5.

	KDPI Model			
Variable	(1)	(2)	(3)	(4)
Total Mismatches	-0.0023***	-0.0022***	-0.0025***	-0.0023***
Func. Stat. 1	0.0017*	0.0001	0.0012	-0.0002
Func. Stat. 2	0.0021*	0.0000	0.0013	-0.0004
Func. Stat. 3	0.0022*	0.0038***	0.0021*	0.0038***
Prev. Preg.	0.0008***	0.0004**	0.0010***	0.0004**
Male	0.0004*	-0.0002	0.0004**	-0.0001
BMI	-0.0003	-0.0001***	-0.0000	-0.0001***
On Dialysis	0.0013**	0.0012***	0.0014***	0.0013***
COPD	0.0006	0.0003	0.0006	0.0003
Hypertension	-0.0004	-0.0004	-0.0001	-0.0002
Diabetes Type 1	-0.0025	-0.0004	-0.0022	-0.0002
Diabetes Type 2	-0.0004	-0.0002	-0.0003	-0.0001
Albumin	-0.0014**	-0.0006***	-0.0014**	-0.0006***
Perip. Vasc.	0.0025	0.0010	0.0033	0.0014
Exh. Perip.	-0.0001	0.0002	0.0001	0.0003
Exh. Perit.	0.0009	-0.0006	0.0009	-0.0006
College Education	-0.0005	-0.0001	-0.0005	-0.0001
HPV	-0.0007	0.0004	-0.0005	0.0002
HBV	0.0002	0.0005	0.0006	0.0007
Prev. Tx.	0.0026**	0.0027**	0.0028**	0.0028***
Age	0.0000	0.0000***	0.0000	0.0000***
HCV	0.0031**	0.0028***	0.0033***	0.0028***
Initial PRA	0.0001***	0.0001***	0.0001***	0.0001***
Cold Ischemic Time	0.0839***	0.0845***	0.0840***	0.0845***
(Cold Isc. Time)^2	-0.0011***	-0.0011***	-0.0011***	-0.0011***
Wait Time	0.0001***	0.0001***	0.0000***	0.0000***
# on Center Waiting List	-0.0001***	-0.0000	-0.0000***	-0.0000
Center Fixed Effects	N	Y	N	Y
Mnth/Yr Fixed Effects	N	N	Y	Y
R^2	0.7654	0.7724	0.7660	0.7729

(*indicates statistically significant at 90% level; **indicates statistically significant at 95% level; ***indicates statistically significant at 99% level)

Table 5: Estimation results for Model 1. There are a total of 5,298,155 observations. The errors are clustered at the transplant center level. All regressions contain the following controls: patient diagnosis code, patient ethnicity, donor death circumstances, blood mismatches, patient functional status, previous pregnancies, patient gender, patient age, patient BMI, dialysis indicator, COPD indicator, hypertension indicator, patient diabetes, patient albumin levels, patient vascular capacity, patient education, patient HBV status, number of previous transplants, PRA levels at listing, cold ischemic time, patient waiting time and center's current waiting list.

CMS CoP : Patient Graft Survival				
Variable	(1)	(2)	(3)	(4)
$KDPI_d$	-0.01651***	-0.01106***	-0.01685***	-0.01137***
$Post_t$	-0.00163***	-0.00286***	0.00443	-0.00799**
$Trig_{tc}$	0.00169	-0.00011	0.00176	0.00004
$(Post_t)(Trig_{tc})$	-0.00197	-0.00025	-0.00212	-0.00031
Center Fixed Effects	N	Y	N	Y
Mnth/Yr Fixed Effects	N	N	Y	Y
R^2	0.7654	0.7724	0.7660	0.7729
CMS CoP : Patient Survival				
$KDPI_d$	-0.01652***	-0.01106***	-0.01686***	-0.01137***
$Post_t$	-0.00170***	-0.00289***	0.00430	-0.00803**
$Trig_{tc}$	-0.00078	-0.00188	-0.00130	-0.00224*
$(Post_t)(Trig_{tc})$	-0.00200	0.00015	-0.00052	0.00095
Center Fixed Effects	N	Y	N	Y
Mnth/Yr Fixed Effects	N	N	Y	Y
R^2	0.7654	0.7724	0.7660	0.7729

(*indicates statistically significant at 90% level; **indicates statistically significant at 95% level; ***indicates statistically significant at 99% level)

Table 6: Estimation results for Model 2. There are a total of 5,298,155 observations. The errors are clustered at the transplant center level. All regressions contain the following controls: patient diagnosis code, patient ethnicity, donor death circumstances, blood mismatches, patient functional status, previous pregnancies, patient gender, patient age, patient BMI, dialysis indicator, COPD indicator, hypertension indicator, patient diabetes, patient albumin levels, patient vascular capacity, patient education, patient HBV status, number of previous transplants, PRA levels at listing, cold ischemic time, patient waiting time and center's current waiting list.

	KDPI Model			
Variable	(1)	(2)	(3)	(4)
$KDPI_d$	-0.02542***	-0.02415***	-0.02598***	-0.02461***
$(KDPI_d > 0.5)$	-0.00366***	-0.00886***	-0.00364***	-0.00866***
$(KDPI_d)\ (KDPI_d > 0.5)$	0.01075***	0.01962***	0.01073***	0.01931***
Center Fixed Effects	N	Y	N	Y
Mnth/Yr Fixed Effects	N	N	Y	Y
R^2	0.7654	0.7725	0.7661	0.7730

(*indicates statistically significant at 90% level; **indicates statistically significant at 95% level; ***indicates statistically significant at 99% level)

Table 7: Model 3 results. There are a total of 5,298,155 observations. The errors are clustered at the transplant center level. All regressions contain the following controls: patient diagnosis code, patient ethnicity, donor death circumstances, blood mismatches, patient functional status, previous pregnancies, patient gender, patient age, patient BMI, dialysis indicator, COPD indicator, hypertension indicator, patient diabetes, patient albumin levels, patient vascular capacity, patient education, patient HBV status, number of previous transplants, PRA levels at listing, cold ischemic time, patient waiting time and center's current waiting list.

CMS CoP : Patient Graft Survival				
Variable	(1)	(2)	(3)	(4)
$KDPI_d$	-0.01822***	-0.01088***	-0.01851***	-0.01190***
$(KDPI_d>0.5)$	-0.00182***	-0.00247***	-0.00175**	-0.00240***
$Post_t$	-0.00469***	-0.00526***	0.00140	-0.01046***
$Trig_{tc}$	0.00288	0.00133	0.00303	0.00139
$(KDPI_d>0.5)(Post_t)$	0.00428***	0.00334***	0.00416***	0.00325***
$(KDPI_d>0.5)(Trig_{tc})$	-0.00180	-0.00215	-0.00191	-0.00212
$(Post_t)(Trig_{tc})$	-0.00248	-0.00166	-0.00285	-0.00177
$(KPRI_d>0.5)(Post_t)(Trig_{tc})$	0.00083	0.00218	0.00114	0.00223
Center Fixed Effects	N	Y	N	Y
Mnth/Yr Fixed Effects	N	N	Y	Y
R^2	0.7654	0.7725	0.7661	0.7730
CMS CoP : Patient Survival				
$KDPI_d$	-0.01825***	-0.01090***	-0.01853***	-0.01121***
$(KDPI_d>0.5)$	-0.00178***	-0.00245***	-0.00173**	-0.00237***
$Post_t$	-0.00479***	-0.00537***	0.00121	-0.01059***
$Trig_{tc}$	0.00226	0.00113	0.00175	0.00085
$(KDPI_d>0.5)(Post_t)$	0.00433***	0.00347***	0.00423***	0.00338***
$(KDPI_d>0.5)(Trig_{tc})$	-0.00457	-0.00446	-0.00459	-0.00459*
$(Post_t)(Trig_{tc})$	0.00009	0.00218	0.00140	0.00282
$(KPRI_d>0.5)(Post_t)(Trig_{tc})$	-0.00292	-0.00284	-0.00268	-0.00262
Center Fixed Effects	N	Y	N	Y
Mnth/Yr Fixed Effects	N	N	Y	Y
R^2	0.7654	0.7725	0.7661	0.7730

(*indicates statistically significant at 90% level; **indicates statistically significant at 95% level; ***indicates statistically significant at 99% level)

Table 8a: Model 4 results. There are a total of 5,298,155 observations. The errors are clustered at the transplant center level. CMS CoP for Patient Graft Survival. All regressions contain the following controls: patient diagnosis code, patient ethnicity, donor death circumstances, blood mismatches, patient functional status, previous pregnancies, patient gender, patient age, patient BMI, dialysis indicator, COPD indicator, hypertension indicator, patient diabetes, patient albumin levels, patient vascular capacity, patient education, patient HBV status, number of previous transplants, PRA levels at listing, cold ischemic time, patient waiting time and center's current waiting list.

CMS CoP : Patient Graft Survival				
Variable	(1)	(2)	(3)	(4)
$KDPI_d$	-0.05454***	-0.05042***	-0.05408***	-0.05024***
$(KDPI_d>0.5)$	-0.00929***	-0.01457***	-0.00969***	-0.01516***
$Post_t$	-0.01705***	-0.01657***	-0.01073***	-0.02220***
$Trig_{tc}$	0.00430	0.00321	0.00422	0.00317
$(KDPI_d)(KDPI_d>0.5)$	0.03157***	0.03943***	0.03182***	0.04008***
$(KDPI_d)(Post_t)$	0.04036***	0.03651***	0.03951***	0.03625***
$(KDPI_d)(Trig_{tc})$	-0.00493	-0.00655	-0.00409	-0.00604
$(KDPI_d>0.5)(Post_t)$	0.00652***	0.00726***	0.00752***	0.00870***
$(KDPI_d>0.5)(Trig_{tc})$	0.00009	-0.00397	0.00079	-0.00322
$(Post_t)(Trig_{tc})$	-0.00321	-0.00216	-0.00268	-0.00162
$(KDPI_d)(KDPI_d>0.5)(Post_t)$	-0.02748***	-0.02706***	-0.02849***	-0.02897***
$(KDPI_d)(KDPI_d>0.5)(Trig_{tc})$	0.00075	0.00649	-0.00082	0.00522
$(KDPI_d)(Trig_{tc})(Post_t)$	0.00258	0.00187	-0.00056	-0.00051
$(KDPI_d>0.5)(Trig_{tc})(Post_t)$	0.00447	0.00507	0.00254	0.00308
$(KDPI_d)(KDPI_d>0.5)(Post_t)(Trig_{tc})$	-0.00633	-0.00494	-0.00152	-0.00082
Center Fixed Effects	N	Y	N	Y
Mnth/Yr Fixed Effects	N	N	Y	Y
R^2	0.7656	0.7727	0.7663	0.7731

(*indicates statistically significant at 90% level; **indicates statistically significant at 95% level; ***indicates statistically significant at 99% level)

Table 8b: Model 4 results. There are a total of 5,298,155 observations. The errors are clustered at the transplant center level. CMS CoP for Total Patient Survival. All regressions contain the following controls: patient diagnosis code, patient ethnicity, donor death circumstances, blood mismatches, patient functional status, previous pregnancies, patient gender, patient age, patient BMI, dialysis indicator, COPD indicator, hypertension indicator, patient diabetes, patient albumin levels, patient vascular capacity, patient education, patient HBV status, number of previous transplants, PRA levels at listing, cold ischemic time, patient waiting time and center's current waiting list.

CMS CoP : Total Patient Survival				
Variable	(1)	(2)	(3)	(4)
$KDPI_d$	-0.05398***	-0.04974***	-0.05353***	-0.04956***
$(KDPI_d>0.5)$	-0.00912***	-0.01448***	-0.00951***	-0.01505***
$Post_t$	-0.01711***	-0.01655***	-0.01081***	-0.02218***
$Trig_{tc}$	0.00981	0.01085	0.00857	0.00985
$(KDPI_d)(KDPI_d>0.5)$	0.03105***	0.03892***	0.03127***	0.03956***
$(KDPI_d)(Post_t)$	0.04024***	0.03611***	0.03928***	0.03579***
$(KDPI_d)(Trig_{tc})$	-0.02618	-0.03368	-0.02376	-0.03127
$(KDPI_d>0.5)(Post_t)$	0.00691***	0.00751***	0.00778***	0.00885***
$(KDPI_d>0.5)(Trig_{tc})$	-0.00376	-0.00917	-0.00323	-0.00848
$(Post_t)(Trig_{tc})$	0.00210	0.00194	0.00418	0.00345
$(KDPI_d)(KDPI_d>0.5)(Post_t)$	-0.02784***	-0.02697***	-0.02860***	-0.02871***
$(KDPI_d)(KDPI_d>0.5)(Trig_{tc})$	0.01540	0.02719	0.01321	0.02467
$(KDPI_d)(Trig_{tc})(Post_t)$	-0.00641	0.00122	-0.00893	-0.00173
$(KDPI_d>0.5)(Trig_{tc})(Post_t)$	-0.11177	-0.00592	-0.01072	-0.00608
$(KDPI_d)(KDPI_d>0.5)(Post_t)(Trig_{tc})$	0.01455	0.00313	0.01578	0.00539
Center Fixed Effects	N	Y	N	Y
Mnth/Yr Fixed Effects	N	N	Y	Y
R^2	0.7656	0.7727	0.7663	0.7732

(*indicates statistically significant at 90% level; **indicates statistically significant at 95% level; ***indicates statistically significant at 99% level)

Table 9: Model 1 results broken down by triggers. The model specifications are: (1) The first CMS CoP trigger, Observed-Expected>3; (2) the second CMS CoP trigger, Observed/Expected>1.5; (3) Triggers 1 and 2; (4) Triggers 1 and 3, p-value less than 0.05; (5) Triggers 2 and 3. All regressions contain the following controls: patient diagnosis code, patient ethnicity, donor death circumstances, blood mismatches, patient functional status, previous pregnancies, patient gender, patient age, patient BMI, dialysis indicator, COPD indicator, hypertension indicator, patient diabetes, patient albumin levels, patient vascular capacity, patient education, patient HBV status, number of previous transplants, PRA levels at listing, cold ischemic time, patient waiting time and center's current waiting list.

CMS CoP : Patient Graft Survival						
Variable	(1)	(2)	(3)	(4)	(5)	(6)
$KDPI_d$	-0.01137***	-0.01137***	-0.01137***	-0.01137***	-0.01137***	-0.01137***
$Post_t$	-0.00798**	-0.00799**	-0.00799**	-0.00799**	-0.00806**	-0.00799**
$Trig_{tc}$	0.00008	0.00018	0.00012	-0.00010	-0.00088	-0.00004
$(Post_t)(Trig_{tc})$	-0.00027	-0.00005	-0.00015	-0.00022	0.00062	-0.00031
Center FE	Y	Y	Y	Y	Y	Y
Mnth/Yr FE	Y	Y	Y	Y	Y	Y
R^2	0.7729	0.7729	0.7729	0.7729	0.7729	0.7729
CMS CoP : Total Patient Survival						
Variable	(1)	(2)	(3)	(4)	(5)	(6)
$KDPI_d$	-0.01137***	-0.01137***	-0.01137***	-0.01137***	-0.01137***	-0.01137***
$Post_t$	-0.00816**	-0.00807**	-0.00807**	-0.00803**	-0.00802**	-0.00803**
$Trig_{tc}$	-0.00128	-0.00202*	-0.00260*	-0.00218*	-0.00199	-0.00224*
$(Post_t)(Trig_{tc})$	0.00314**	0.00121	0.00188	0.00096	0.00070	0.00095
Center FE	Y	Y	Y	Y	Y	Y
Mnth/Yr FE	Y	Y	Y	Y	Y	Y
R^2	0.7729	0.7729	0.7729	0.7729	0.7729	0.7729