THREE ESSAYS ON THE SUPPLY AND DELIVERY OF HEALTHCARE: EVIDENCE FROM KIDNEY DONATION AND TRANSPLANTATION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Matthew Joseph Sweeney

August 2011

THREE ESSAYS ON THE SUPPLY AND DELIVERY OF HEALTHCARE: EVIDENCE FROM KIDNEY DONATION AND TRANSPLANTATION

Matthew Joseph Sweeney, Ph. D.

Cornell University 2011

This dissertation focuses on the supply side of healthcare, using the specific case of kidney donation and transplantation to address three research questions. In the first chapter, I test whether kidney transplant candidates with better access to publicly available kidneys from deceased donors are less likely to opt for living kidney donation, or the "private supply." Identification comes from a discontinuous increase in the probability of receiving a publicly available kidney generated by the results of a blood test. I find that patients with better access to the public supply of kidneys are less likely to opt for living donation. The results indicate that policies aimed at increasing the number deceased donors will result in less than full crowd-out of living donation.

The second chapter examines the well-documented association between procedural volume and patient outcomes in the context of kidney transplantation. In particular, I test whether a volume effect exists in kidney transplantation, which would be consistent with the "practice makes perfect" hypothesis. Identification of the volume effect comes from plausibly exogenous supply shocks of kidneys within a year at a transplant center. The empirical results suggest that much of the observed volume-outcome relationship in kidney transplantation is due to between hospital differences in unobserved characteristics that are correlated with both transplant volume and patient outcomes. Concentration of transplants at higher volume transplant centers may reduce rates of short term patient mortality, but these gains would need to be

carefully weighed against any reductions in patient access to care that would result from regionalization of care.

The last chapter examines whether transplant centers experience “forgetting” during temporal breaks between kidney transplants. Identification relies on the randomness of arrivals of transplantable kidneys at transplant centers, which would create plausibly exogenous variation in the size of temporal breaks between transplants at a given transplant center. In addition, I test whether the level of experience immediately before a temporal break mitigates any deleterious effects that arise from the break. The estimated results suggest that there is little relationship between temporal breaks and transplant center productivity, as measured by the outcomes of patients transplanted immediately following a break in transplant activity.

BIOGRAPHICAL SKETCH

Matthew Joseph Sweeney grew up in Connecticut where he attended Canterbury School, and graduated as the valedictorian of his class in 1997. After high school, he attended the College of the Holy Cross in Worcester, Massachusetts and earned a bachelor's of arts degree with honors in economics. After graduating from Holy Cross in 2001, he went on to work for a health insurance company as an actuarial analyst. His experience working in the insurance industry sparked an interest in health policy and health economics, and he decided to return to school to earn his doctoral degree. He began his doctoral training in health policy at Cornell University in the Department of Policy Analysis and Management in 2005, and completed the Ph.D. program in 2011.

Matthew married Meg Guimond in 2007, and their daughter Caroline was born in April 2011. Matthew will be working for Mathematica Policy Research upon the completion of his doctoral degree.

ACKNOWLEDGMENTS

Even though only my name appears on this dissertation, the research presented here greatly benefited from the guidance and feedback from a number of faculty members at Cornell. First, I am indebted to Sean Nicholson for not only helping me steer this ship, but also taking me on as a research assistant early in my graduate training. That experience helped me learn firsthand what it means to be a good researcher in the field of health economics. I appreciate all of the time and patience he has exerted in helping me become a skilled researcher. John Cawley has shown me that it is not impossible to be both a well-respected researcher and a high-quality and influential teacher at the same time. I am grateful for the research opportunities that he has extended to me, and for providing me with an excellent model of what it means to be a great instructor in the classroom. I am very thankful to Kosali Simon for her feedback on my research, and for always putting in a kind word on my behalf while I was on the job market. I have spent many hours speaking with Jordan Matsudaira about my research, from the “big picture” questions, down to detailed Stata code. I appreciate all of the time he has spent thinking about my research, as well as the occasional round of golf.

I am also grateful for two of my undergraduate economics professors at Holy Cross: John Carter and Miles Cahill. They planted the seed in my mind that I would be successful in research and in graduate school. Coming to graduate school has been a wonderful experience, and I have John and Miles to thank for helping me make the decision to pursue a doctorate degree.

I am very fortunate to have such a loving and supportive family, and their thoughts, prayers, and laughs over the last six years have kept me motivated and inspired to complete my doctorate.

Finally, but by no means least, I will always be grateful for my beautiful and loving wife Meg. We came to Ithaca so that I could go to graduate school and fulfill my academic dreams, and I will forever be indebted to her for her love and patience over the last six years. She is my best friend, my source of comfort when I need it, the boost of confidence when I doubt myself,

my comedienne when I need a laugh, and even an occasional sounding board for research ideas. I am very lucky to have her in my life, and I look forward to many years together with our beautiful daughter Caroline.

TABLE OF CONTENTS

CHAPTER 1

Does the Public Provision of Kidneys Crowd Out Living Kidney Donation? A Regression Discontinuity Analysis

1.1 Introduction

The interaction of the public and private sectors in the provision of a good has received considerable interest from economists and public policy officials. For example, at the core of the ongoing health care debate in the United States is the appropriate role of the public sector in providing health care. Some argue that near universal coverage provided or subsidized by the federal government will simply crowd out private insurance for millions of Americans. Evidence from the economics literature supports this claim; expansions of public insurance programs have resulted in sizeable crowd out of private insurance (Cutler and Gruber 1996). In countries with universal coverage, however, private markets serve as effective substitutes or supplements for the public option (Besley, Hall, and Preston 1999 ; Gouveia 1997). In these countries, the private market allows individuals to circumvent waiting lists for care, and perhaps purchase a higher quality product.

In this paper, I explore the interaction of the public sector and the private sector in the supply of transplantable kidneys. The private supply is made up of kidneys that are donated to patients by living donors who are members of a patients' social network (family members or friends). The public supply comes from kidneys that are recovered from deceased donors and made available to patients registered on a transplant waitlist. From the patient's perspective, the key characteristic that distinguishes the two sources of supply is that a kidney from a living donor necessarily comes from someone within the patient's social network, whereas a kidney

from the public supply comes from an anonymous, deceased donor. While living donation results in better clinical outcomes and lower medical costs for the patient, it imposes financial and psychic costs on the patient and members of her social network. For example, living donors are not reimbursed for lost wages as a result of making the donation. In addition, the patient may incur search costs when trying to identify a potential living donor. Finally, the patient's social network may take on bargaining costs when deciding which individual will end up making the donation.

Theoretically, improving access to the pool of publicly available kidneys (kidneys from anonymous, deceased donors) should crowd out living donation and thus the private supply of kidneys. From the patient's perspective, a kidney from a deceased donor is less costly than a kidney from a living donor, since it does not carry with it the financial and psychic costs described above. Therefore, one might expect that every patient would opt for a publicly available kidney (i.e. "full" crowd-out). In reality, however, there are also costs associated with opting for a publicly available kidney. The median waiting time for a publicly available kidney is over 3 years in the United States due to excess demand. Most patients waiting for publicly provided kidney are on dialysis, which can reduce quality of life (Knoll and Nichol 2003). In addition, there is a non-trivial chance of death while waiting for a transplant; in 2009, approximately 5,000 patients died while on the waitlist for a publicly available kidney, and another 2,000 were removed from the waitlist because they became too sick to be considered a good transplant candidate.[1]

In this paper, I empirically test the hypothesis that patients with better access to the pool of publicly available kidneys will be less likely to opt for living donation. Identifying this crowd-

[1] Data taken from Organ Procurement and Transplantation Network (OPTN) data website, available here: http://optn.transplant.hrsa.gov/data/

out effect is challenging because there are many factors that influence whether a patient receives a kidney from a living donor that are unobserved to the researcher. I overcome these challenges by exploiting plausibly exogenous variation in a patient's ability to access the pool of publicly available kidneys. Patients are required to submit blood samples for testing when they register at a transplant center. Patients with a high enough score on a particular blood test are given priority access to the pool of publicly available kidneys. This special consideration is awarded in a discontinuous manner; only patients whose blood test score exceeds a particular cutoff have extra points added to their waitlist ranking. Therefore, relative to patients just below the blood test score cutoff, patients just above the cutoff have better access to the pool of publicly available kidneys. Consistent with the crowd-out hypothesis, patients just above the cutoff are 4.1 to 4.4 percentage points less likely to receive a privately supplied kidney from a living donor, as compared patients just below the cutoff. Tests of the identifying assumptions indicate that this estimation strategy yields an unbiased, reduced-form estimate of the crowd-out of living donation. Specifically, a 10 percentage point increase in the expected probability of receiving a publicly available kidney from the waitlist is predicted to reduce the probability that a patient will opt for living donation by 2.6 to 3.4 percentage points.

The implication of these findings is that policies aimed at increasing the number of publicly available kidneys will result in some crowd-out of living donation, although the crowd-out effect would be relatively small. This is a notable finding for two reasons. First, living donation is the most cost-effective treatment option for patients with kidney failure (Mullins et al. 2003). Second, Medicare is the primary insurer for patients with kidney failure.[2] In 2008, Medicare spent approximately $23 billion on patients with kidney failure, with $4.9 billion paid

[2] Any person with kidney failure is entitled to Medicare benefits. For patients with a source of insurance at the time of diagnosis, Medicare serves as the secondary insurer for the first 30 months of treatment. After 30 months, Medicare becomes the primary insurer.

for dialysis treatments (USRDS 2010). Therefore, from a societal point of view, crowd-out of living donation is troubling because it leads to increased Medicare expenditures as patients opt out of the cost-effective treatment (living donation) and decide to wait (on dialysis) for a kidney from the public supply to become available. However, the results presented here suggest that the crowd-out effect is small in magnitude, so that any policies aimed at increasing the number of deceased donor kidneys would be welfare improving, since they would likely create a net increase in the number of transplants performed, and thus the number of lives saved.

The remainder of this paper is organized as follows. The next section describes the kidney transplantation options available for patients, as well as the kidney allocation process in the United States. The third section sketches a conceptual framework that describes how the public provision of kidneys can crowd out the private supply of kidneys from living donors. I then describe the empirical strategy and the regression discontinuity design. The fifth section discusses the data used for estimation, and is followed by the empirical results. I then present the results from specification tests and tests of the identifying assumptions. The final section discusses the findings.

1.2 Kidney Transplantation in the United States

Patients in kidney failure have two treatment options available to them: dialysis, in which the blood is mechanically cleansed of impurities, and transplantation, in which a kidney from another person (either deceased or living) is surgically transplanted into the patient. Dialysis is not considered a long-term solution, because lengthened time on dialysis is correlated with progressive cardiovascular disease (Goodman and Danovitch 2005) and a lower quality of life for the patient while on dialysis: the typical dialysis regimen requires the patient to go to a

dialysis center 3 times a week for at least an hour at a time. In addition, dialysis has been shown to generate higher long-run financial costs, as compared to transplantation (Mullins et al. 2003).

Transplantable kidneys come from either deceased donors (the public supply) or from living donors (the private supply). With living donor kidney transplantation, a kidney from another living individual (usually a family member) is removed laparoscopically and transplanted immediately into the patient. The recovery time for the donor is typically 4 weeks, although it varies by individual. The donor's medical costs are paid by the patient's insurer. However, lost wages, lost home production, and travel costs are not reimbursed by insurers. The long-term prognosis for living donors is generally good, although there is some evidence that living donors are more likely to develop hypertension than non-donors (Davis 2009). In 2009, 38% of kidney transplants were performed using kidneys from living donors, and most living donors were biologically related to the patient.[3]

The supply of publicly available kidneys stems from deceased individuals who indicated a willingness to donate their organs after death. Organ donors have usually suffered some kind of trauma or stroke that has left them "brain dead." Organ procurement officials help coordinate the organ donation process, from obtaining consent from the deceased individual's family to procure the organs, to collecting data about the donor such as age, race, height, weight, circumstance of death, etc. Blood and tissue samples are collected and analyzed so that the donor's blood and antigen profiles can be entered into a central database. This data is fed into a computer algorithm to produce a list of potential transplant candidates for a particular kidney.

Kidneys recovered from deceased donors are allocated regionally. There are 58 organ procurement organizations (OPOs) that are assigned local service areas. Each OPO has at least

[3] Based on author's calculations using Organ Procurement and Transplantation (OPTN) data found here: http://optn.transplant.hrsa.gov/data/

one kidney transplant center in its service area. When kidneys are recovered from a deceased donor in a particular OPO service area, they are first offered to patients that are registered at transplant programs in that same OPO. Therefore, the supply of publicly available kidneys varies geographically, depending on the number of brain deaths that have occurred in the local area, and the number of families that consent to organ donation.

Each time a kidney is recovered from a deceased donor, a computer algorithm calculates a waitlist score for local patients on the waitlist who have a blood type that is compatible with the donor. The patient with the highest score will be offered the kidney first. If the patient (or more accurately, her physician) turns down the offer, then the kidney is offered to the patient with the next highest score. The length of time that a patient has been on the waitlist is the major component of the waitlist score. That is, patients with the longest wait times have the highest waitlist scores. Points can also be awarded based on the level of the immunological match to the recovered kidney, but this aspect of the allocation system has been de-emphasized because advancements in immunosuppressive treatments have made immunological matching less of a priority.

The policy also gives certain patient populations special consideration for deceased donor kidneys. For example, because kidney function is vital in the development of children (Al-Akash and Ettenger 2005), pediatric patients (age less than 18) receive extra points on their waitlist score, and recently have been given priority access to higher quality organs from deceased donors age 35 younger at time of death (the so-called “Share 35” policy, enacted in 2005). Another group of patients that receives special consideration are “sensitized” patients. A sensitized patient is one whose body has heightened levels of antibodies that will react against the antigens present on the cell surfaces of the transplanted kidney, which makes the patient’s

body more likely to reject the kidney. Common "sensitizing events" are childbirth, previous organ transplant, and receipt of a blood transfusion. The degree of sensitization is measured by the patient's panel reactive antibody (PRA) score. This score, which ranges from 0 to 100, is a rough measure of the likelihood that the patient's body will reject a kidney from the local donor pool. For example, a patient with a PRA score of 40 would be projected to reject 40% of the kidneys from the local donor pool because of her underlying antibody profile. When a deceased donor kidney becomes available, potential recipients must pass a "crossmatch" test, which tests whether the patient's immune system is likely to react against that particular kidney. Sensitized patients are more likely to fail this test because of their underlying antibody profile, and therefore can wait an extended period of time for a compatible deceased donor kidney to become available.

Recognizing that sensitized patients have a disadvantage in accessing the pool of publicly available kidneys, the allocation policy gives these patients priority access to a kidney they are compatible with (i.e. if the patient passes the crossmatch test) in the form of 4 extra points on their waitlist score. However, a patient can only qualify for the extra points if her PRA score is 80 or higher. Patients with a PRA score below 80 are not eligible for the extra waitlist points, even though they might be just as sensitized as patients with a PRA score just above the cutoff. In other words, the policy creates a discontinuous increase in the probability of receiving a publicly available kidney; patients with a PRA score just above the cutoff are more likely to receive a kidney from the waitlist than patients just below the PRA cutoff. This variation in the ability to access the pool of publicly available kidneys will serve as my identification strategy, which is discussed in more detail below.

1.3 Conceptual Framework

Consider a patient that is in kidney failure and in need of a kidney transplant, and her social network that is made up of individuals that are "emotionally related" to the patient. These individuals derive utility from the patient's health and well-being. Since each member of the patient's social network derives utility from the health of the patient, they will achieve higher utility if the patient receives a transplant.[4] If no one within the patient's social network donates a kidney, then it is assumed that the patient will wait for a publicly available kidney to become available. The expected costs and expected benefits associated with the two transplantation options will determine whether the patient receives a kidney from a living donor.

If the social network supplies a kidney to the patient privately, then the benefits of the improved health of the patient are realized immediately by everyone in the patient's social network. The transplant can be scheduled and performed as soon as a day is found that is convenient for the patient, the donor, and the surgical team. In other words, the patient receives a new kidney with certainty, and her health is improved immediately (i.e. there is virtually no waiting). Because each member of the social network derives utility from the improved health of the patient, each member of the social network realizes a higher level of utility immediately and with certainty. In contrast, there is no guarantee that the patient will ever be offered a publicly available kidney while on the waitlist. In addition, even if the patient receives a kidney from the waitlist, the median waiting time for a kidney in the United States in over 3 years. While the patient waits on the waitlist, her condition could worsen, and this deterioration in the patient's health would lower the utility level of the patient and the utility of the members of her social

[4] In this sense, the demand for a transplantable kidney is a derived demand, as it is a key input into the production of health for the patient. That is, the patient and her social network derive utility from the health of the patient, and not directly from the kidney itself.

network. The impact of this waiting period on the decision to opt for living donation depends on the time preferences of the patient and the members of her social network.

Despite the benefits of living kidney transplantation, there are important costs associated with relying on the private supply of kidneys that must be considered. These costs are borne privately by the patient and her social network, and can take on different forms. First, the risk of complication or death that the living donor will be exposed to as a result of making the donation can be considered a transaction cost; the exchange cannot be made without the surgery, and thus without exposure to these risks. While the mortality and complication rates for living donors are low, patients and potential donors may over-estimate these risks, and may also have differing levels of willingness to accept risk (Young et al 2008). In addition, the donor may incur out of pocket financial costs, such as lost wages and travel expenses. There is also some evidence that living donors have difficulty obtaining life insurance after donation, or may pay higher premiums for coverage (Yang et al. 2007).

Going inside the "black box" of the decision process of the patient and her social network, there are additional costs that are associated with identifying and selecting a potential living donor. First, the patient may incur search costs – which are likely to be psychological in nature - associated with recruiting potential living donors. Patients can be reluctant to ask a member of their social network for a kidney out of concern for the donor's health and well-being (Pradel et al 2003). Second, there may be bargaining costs within the social network in terms of identifying who becomes the living donor. The health of the patient can be viewed as a public good, so members of the patient's social network may attempt to "free-ride". There is evidence that families do strategize when considering living donation. For example, one Canadian study found that wives are more likely to donate to their husbands, but that husbands are less likely to

donate to their wives. This disparity can be explained in the context of a family bargaining framework: if the husband is the primary earner, then the opportunity cost of his time associated with making a living donation may be relatively large. Therefore, for a female patient, the family may collectively decide that the patient's husband should not be the donor, since the donation process could result in a disruption of income or consumption for the family (Zimmerman et al. 2000).

A publicly supplied kidney represents a less costly treatment option for a patient and her social network because it does not carry with it the costs discussed above, as these kidneys are from anonymous donors. The hypothesis of this paper is that patients with a relatively high probability of receiving a publicly provided kidney will be less likely to receive a privately supplied kidney from a living donor, all else equal. There is one study in the economics literature that examines this issue. Howard (forthcoming) uses geographic variation in the supply and demand of deceased donor kidneys to identify the substitutability of living and deceased kidney donation. He finds that patients with a higher predicted wait time for a deceased donor kidney are more likely to opt for living donor kidney transplantation. His results are consistent with the hypothesis put forth here, as well as with the empirical results of this paper, although motivated and identified in different ways.

The framework discussed here highlights the potential difficulty in empirically identifying the effect of having differentially better access to the pool of publicly available kidneys on the probability that a patient receives a kidney from a living donor. There are many factors that go into the decision-making process that are unobserved to the researcher, like the size of a patient's social network, the value that the members of the social network place on the improved health of the patient, the willingness of the patient to accept a kidney from a member

of her social network, etc. In some respect, it is unlikely that any of these unobserved factors are correlated with the ability of the patient to receive a publicly available kidney. But there are other factors, like the income level of the patient, that are unobserved and could be correlated with the expected probability of receiving a publicly provided kidney. Higher income patients can more easily afford to pay multiple registration fees and they may register at multiple transplant centers to increase the likelihood of receiving a kidney from a deceased donor. If higher income patients (and their social networks) have systematically different preferences for living donation, then omitting income from an estimating equation could lead to bias in the estimated regression coefficients.

1.4 Identification Strategy and Estimation

In order to identify the causal crowd-out effect, I exploit a discontinuous change in patients' ability to access the pool of publicly available kidneys created by the kidney allocation policy. As described earlier, patients with a PRA score of 80 or higher receive 4 extra points on their waitlist score if they are an acceptable match to a deceased donor kidney. Therefore, relative to patients with a PRA score just below 80, patients just above are more likely to receive a publicly available kidney from the waitlist. Therefore, the treatment of interest – the ability to access the pool of publicly available kidneys – changes discontinuously at the PRA cutoff of 80. If having improved access to the pool of publicly available kidneys creates a disincentive for patients and their social networks to opt for living donation, then there will exist a discontinuous decrease in the probability that a patient receives a kidney from a living donor at the PRA cutoff.

To test this, I first specify a first-stage equation that estimates the discontinuous change (increase) in the probability that a patient receives a publicly available deceased donor kidney that occurs at the PRA cutoff. In particular, I fit a model given by:

$$(1.1)\ DDT_i = \pi_0 + \pi_1 PRA80_i + PRA80_i \sum_{q=1}^{2} (\pi_{Rq})(PRA_i^q) + (1 - PRA80_i) \sum_{q=1}^{2} (\pi_{Lq})(PRA_i^q) + \nu_i$$

In equation (1.1), DDT_i denotes the probability that patient *i* will receive a publicly available kidney from the waitlist, $PRA80_i$ equals one if the patient's PRA value is 80 or higher, and PRA_i is the patient's PRA value. Following the literature (Lee and Lemieux 2009), I estimate equation (1.1) using a global polynomial approach, and I use a quadratic in PRA score to flexibly estimate the underlying trend in the conditional expectation that a patient will receive a kidney from the waitlist.[5] The estimated coefficient $\widehat{\pi_1}$ captures the discontinuous increase in the probability of receiving a publicly available kidney at the PRA cutoff. Estimation of the first-stage equation is meant to demonstrate the expected probability that a patient receives a publicly available kidney *if* she opts against living donation.

Next, I estimate a reduced-form equation that estimates the discontinuity in the probability that a patient receives a kidney from a living donor:

$$(1.2)\ LDKT_i = \gamma_0 + \gamma_1 PRA80_i + PRA80_i \sum_{q=1}^{2} (\gamma_{Rq})(PRA_i^q) + (1 - PRA80_i) \sum_{q=1}^{2} (\gamma_{Lq})(PRA_i^q) + \varepsilon_i$$

[5] Based on the Schwarz criterion, the quadratic specification was selected over other models that included between one and nine polynomial terms.

Strictly speaking, the estimated coefficient $\widehat{\gamma_1}$ captures the treatment effect of having a PRA value at or above the cutoff of 80 on the probability that a patient receives a kidney from a living donor. Taking the ratio $\frac{\widehat{\gamma_1}}{\widehat{\pi_1}}$ yields an estimate of the local average treatment effect of increasing the probability of receiving a publicly available kidney on the probability of receiving a kidney that was supplied privately from a living donor. If the crowd-out hypothesis is true, then $\widehat{\gamma_1}$ should be negative, as should $\frac{\widehat{\gamma_1}}{\widehat{\pi_1}}$. I estimate equations (1.1) and (1.2) using a linear probability model, and cluster standard errors at each PRA value to account for heteroskedasticity introduced by specification error. I estimate $\frac{\widehat{\gamma_1}}{\widehat{\pi_1}}$ using indirect least squares (ILS) and bootstrap the standard error of the ratio.

The identifying assumption is that patients in the immediate neighborhood of the PRA cutoff value are alike, on average, in all ways that affect the probability that they will receive a kidney from a living donor, except for the fact that patients just above the cutoff have a greater ability to access the pool of publicly available kidneys, relative to patients just below. I partially test this assumption by showing that observable characteristics of patients do not appear to change in a discontinuous manner at the cutoff. Note that this does not rule the possibility that patients on either side of the cutoff are different, on average, along dimensions that are unobservable to the researcher. These unobservable characteristics might include size of the patient's social network, the levels of altruism of potential donors and the patient, and the income level of patients and members of their social networks. In the current setting, however, it seems unlikely that these types of unobservable characteristics that also affect the probability of

receiving a kidney from a living donor would also be changing discontinuously at the cutoff. A patient's PRA score is the result of a laboratory blood test, which is presumably difficult to manipulate. Also, the PRA cutoff value of 80 has been described by some in the transplant community as being "arbitrary" and "artificial", which implies that, from a medical point of view, there is nothing significant about a PRA value of exactly eighty.[6] Therefore, it is likely that patients on either side of the cutoff are alike, on average, in all ways that affect the probability that they will receive a kidney from a living donor, except that patients just above the cutoff have relatively better access to the pool of publicly available kidneys. Nonetheless, I estimate equations (1.1) and (1.2) with and without individual level covariates to indirectly test the validity of the identifying assumptions. I also provide additional tests of the identifying assumptions later in the paper.

One challenge from using the global polynomial approach to estimate equations (1.1) and (1.2) is that the distribution of observations is heavily skewed at the lower end of the PRA distribution: over 70 percent of observations in the data have a PRA between 0 and 5. Any regression will try to best fit the data at this cluster in order to minimize squared errors. Therefore, the fitted regression line in other ranges of the PRA score, particularly near the PRA cutoff of 80, may actually reflect variation in the conditional expectation of $LDKT$ at the cluster at the lower end of the PRA distribution, rather than the variation in $LDKT$ in the neighborhood of the PRA cutoff. In order to avoid this problem, I estimate the regressions using a sub-sample of observations with a PRA of 6 or higher.[7] While this strategy eliminates a large number of

[6] These comments were made during the proposition and subsequent period of public comment of the new kidney allocation policy. Available here (2 links):
http://www.unos.org/SharedContentDocuments/KidneyAllocationSlides_Reduced.pdf
http://www.unos.org/SharedContentDocuments/KidneyAllocationSystem--RequestForInformation.pdf

[7] The empirical results are robust against using different ranges of PRA values to estimate the discontinuities; these results are presented later in the paper.

observations, it reduces the bias that may be created by including them in the estimation sample. In addition, the remaining estimation sample has a sufficient number of observations in the local area of the PRA cutoff to yield precisely estimated coefficients.

1.5 Data and Sample Construction

To fit equations (1.1) and (1.2), I use data from the United Network of Organ Sharing (UNOS) Standard Transplant Analysis and Research Files (STAR) on all kidney waitlist activity as of August 2008. These data include additions to the waitlist (new waitlist registrations) or removals from the waitlist, which can result from transplant, patient death, or if the patient is too ill to be considered a viable transplant recipient. The dataset includes detailed information on transplant candidates collected at the time of registration on the kidney waitlist, information about the transplant procedure (if a transplant is performed), and follow up information on the patient's outcomes after the transplant (again, if a transplant is performed).

For each patient that registers on the waitlist, two measures of PRA are collected and recorded: current PRA, and peak PRA. The current PRA refers to the patient's most recent PRA score while on the waitlist. The peak PRA score refers to the patient's highest ever value of her PRA score. Either score can be used for the purposes of calculating the patient's waitlist score, and the choice of which PRA value to use is made by each transplant center. Also included in the dataset is a variable that indicates which PRA score – current or peak – is used in the calculation of each patient's waitlist score. Using these three variables, I formulate the relevant PRA score that is used for each transplant candidate. This constructed PRA score serves as the running variable in the regression discontinuity estimation.

The data captures a patient's position in the PRA distribution at a particular point in time, but does not report the entire history of laboratory results. One concern is that once a patient is listed on the waitlist, she may experience a "sensitizing" event that would increase her PRA score (and in particular, place her above the PRA cutoff of 80). For example, the patient may start off below the cutoff (and be ineligible for the extra waitlist points), but is then captured in the data above the cutoff because she has experienced a sensitizing event that has increased her level of sensitization. In this case, the running variable in the regression discontinuity design is itself changing, which would mean that patients could potentially be eligible for the extra waitlist points at some points in time, but not at others. It also raises the concern that patients may intentionally expose themselves to sensitizing events to move themselves up the PRA distribution, and possibly qualify for the extra waitlist points.

The three most common sensitizing events are pregnancy, previous transplant, and blood transfusions. In the current context, it is unlikely that patients in the data sample are experiencing these events. Women with kidney failure are less likely to become pregnant because they often have irregular menstruation. In addition, because a pregnancy would greatly increase the risk of high blood pressure for both the mother and the baby, women are strongly discouraged from trying to become pregnant until after they receive a transplant. Therefore, it is unlikely that any women in the estimation sample are becoming pregnant (and potentially more sensitized) after registering on the waitlist. I exclude any patients that have had a previous transplant, and also exclude patients who are simultaneously listed for another organ transplant. Therefore, no patients in the sample should experience sensitization because of receipt of a non-kidney transplant. Finally, patients with kidney failure also often suffer from anemia, a condition that was once treated with blood transfusions. However, the advent of the use of epoetin (a synthetic

hormone used to stimulate red blood cell creation by the bone marrow) in the mid-1990's has greatly reduced the reliance on blood transfusions as a treatment option for dialysis patients. Although I do not directly observe if a patient has received a blood transfusion, I start the sample in 1997 to help reduce the chance that patients receive blood transfusions to treat anemia, and thus become sensitized after being placed on the waitlist.

In addition to the above sample restrictions, I also limit the sample to adult patients (at least 18 years old) and to patients with non-missing values for the three PRA variables discussed above. I exclude patients from Puerto Rico, the US Virgin Islands, and Guam. I also drop observations in which at least one of the following covariates has a missing value: age at registration, gender, race, primary source of insurance, primary diagnosis at registration, blood type, date of waitlist registration, and permanent state of residence. In the models, I include measures of a patient's education level, and whether a patient needs assistance with activities of daily living (ADLs). Data for these two measures are frequently missing. Rather than drop observations with missing information on education or assistance with ADLs, I create dummy variables that indicate whether an observation has a missing value for these two characteristics. The dependent variable in equation (1.1) is an indicator for whether a patient receives a publicly supplied kidney from the waitlist. The dependent variable for equation (1.2) is an indicator for whether a patient receives a privately supplied kidney from a living donor.

The sample restrictions described above yield a sample of 145,159 patients who registered on the waitlist between 1997 and 2006. The final estimation sample includes 39,111 patients who have a PRA score of 6 or higher. It is important to note that the sample does not include any individual that received a living donor transplant without also registering on the waitlist. This is because PRA (the running variable in the regression discontinuity design) is only

observed for patients that are on the waitlist. Over the time period studied here, roughly two-thirds of recipients of a living donor kidney were also registered on the waitlist.[8] For the remaining one-third of living donor kidney recipients, I do observe some patient characteristics like age, race, gender, and source of insurance, which are measured at the time of the transplant. However, PRA is largely uncorrelated with many patient characteristics, so it is difficult to use the characteristics of these patients to project where they would lie in the PRA distribution *if* they were in the estimation sample. The only observable characteristic that is strongly correlated with PRA is gender: because pregnancy is one of the main sources of sensitization, patients in the neighborhood of the PRA cutoff are predominantly female. In particular, females make up 77 percent of the estimation sample with PRA values between 75 and 85. Based on the data that is collected at the transplant, living donor recipients who do not register on the waitlist are more likely to be male; only 41 percent of living donor kidney recipients who did not register on the waitlist are women. Therefore, at least along one dimension, it appears that *if* these missing living donor kidney recipients were in the estimation sample, they would be more likely to be from the lower end of the PRA distribution. In other words, this limited evidence suggests that the living donor recipients that are missing from the estimation sample are not differentially more likely to have PRA values that are in the neighborhood of 80.

[8] Based on author's calculations using Organ Procurement and Transplantation (OPTN) data found here: http://optn.transplant.hrsa.gov/data/. Many transplant centers will place patients on the waitlist, even if the patient has indicated that she will opt for living donor kidney transplantation. The rationale for listing these patients is that it creates an "insurance policy" in the event that the living donor backs out of the surgery.

1.6 Results

Table 1.1 reports the sample means of patient characteristics, and compares average characteristics of patients that receive a kidney from a living donor ("LDK Recipients" in the table) against those that do not.[9] Clearly, there are many factors that are correlated with the probability that a patient receives a kidney from a living donor, namely age, race, education, and insurance status. There are also statistically significant differences in the primary diagnosis and blood types between living donor kidney recipients and non-recipients. In addition, there are secular trends in the probability that patients receive kidneys from living donors. These differences in observable characteristics underscore the challenge of empirically identifying the causal effect of having improved access to the pool of publicly available kidneys on the probability that a patient receives a kidney from a living donor. Simply controlling for these characteristics in a regression will not address the likelihood that living donor kidney recipients (and their social networks) are different, on average, than non-recipients along unobserved dimensions.

First I present the graphical evidence that underlies the identification strategy discussed above. Figure 1.1 plots the unadjusted mean probability of receiving a publicly available kidney against the range of PRA values. The probability of receiving a publicly provided kidney from the waitlist increases discontinuously at the PRA cutoff of 80; patients just below the PRA cutoff have approximately a 25 percent chance of receiving a kidney from the waitlist, while patients just above the cutoff have about a 40 percent chance. Therefore, the ability to access the pool of publicly available kidneys increases by 60 percent for patients who are just above the PRA cutoff, relative to patients just below. Figure 1.2 plots the outcome of interest – the probability

[9] The group does not receive a kidney from a living donor include patients that receive a publicly available kidney, as well as patients that do not receive any transplant at all.

Table 1.1: Descriptive Statistics

	LDK Recipients		Non-LDK Recipients		
	Mean	S.E	Mean	S.E	p-value
Age at Registration (in years)	46.2	0.22	50.1	0.07	0.0
Female	0.586	0.008	0.596	0.003	0.235
Race					
White	0.567	0.008	0.399	0.003	0.000
Black	0.241	0.007	0.387	0.003	0.000
Hispanic	0.132	0.006	0.142	0.002	0.126
Other	0.060	0.004	0.072	0.001	0.006
Education					
High school only or less	0.399	0.008	0.498	0.003	0.000
At least some college	0.426	0.008	0.317	0.002	0.000
Missing	0.175	0.006	0.185	0.002	0.121
Primary Insurance					
Private	0.593	0.008	0.375	0.003	0.000
Medicare	0.315	0.008	0.506	0.003	0.000
Medicaid	0.066	0.004	0.092	0.002	0.000
Other	0.026	0.003	0.027	0.001	0.787
Functional Status					
No Assistance with ADL	0.642	0.008	0.642	0.003	0.991
Some Assistance with ADL	0.028	0.003	0.051	0.001	0.000
Total Assistance with ADL	0.001	0.001	0.001	0.000	0.087
Missing	0.329	0.008	0.305	0.002	0.003
Blood Type					
A	0.328	0.008	0.278	0.002	0.000
B	0.141	0.006	0.155	0.002	0.030
AB	0.032	0.003	0.032	0.001	0.852
O	0.486	0.008	0.523	0.003	0.000
Other	0.013	0.002	0.013	0.001	0.988
Primary Diagnosis					
Diabetes Type 1, Insulin Dependent	0.030	0.003	0.040	0.001	0.006
Diabetes Type 2, Non - Insulin Dependent	0.039	0.003	0.070	0.001	0.000
Diabetes Type 2, Insulin Dependent	0.048	0.004	0.092	0.002	0.000
Hypertensive Nephrosclerosis	0.155	0.006	0.201	0.002	0.000
Polycystic Kidneys	0.102	0.005	0.068	0.001	0.000
Malignant Hypertension	0.029	0.003	0.044	0.001	0.000
Other	0.597	0.008	0.486	0.003	0.000
Year of Waitlist Registration					
1997	0.041	0.003	0.070	0.001	0.000
1998	0.069	0.004	0.085	0.001	0.000
1999	0.071	0.004	0.087	0.002	0.000
2000	0.087	0.005	0.093	0.002	0.260
2001	0.087	0.002	0.098	0.002	0.047
2002	0.111	0.005	0.103	0.002	0.151
2003	0.127	0.005	0.109	0.002	0.000
2004	0.142	0.006	0.113	0.002	0.000
2005	0.137	0.006	0.121	0.002	0.004
2006	0.128	0.005	0.121	0.002	0.232

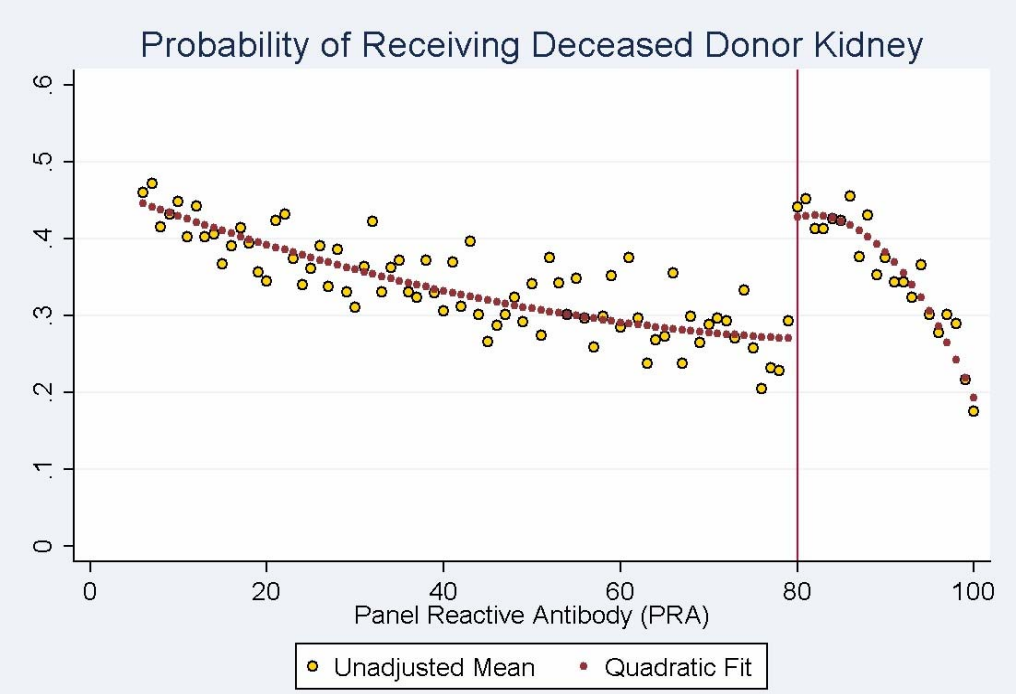

Figure 1.1

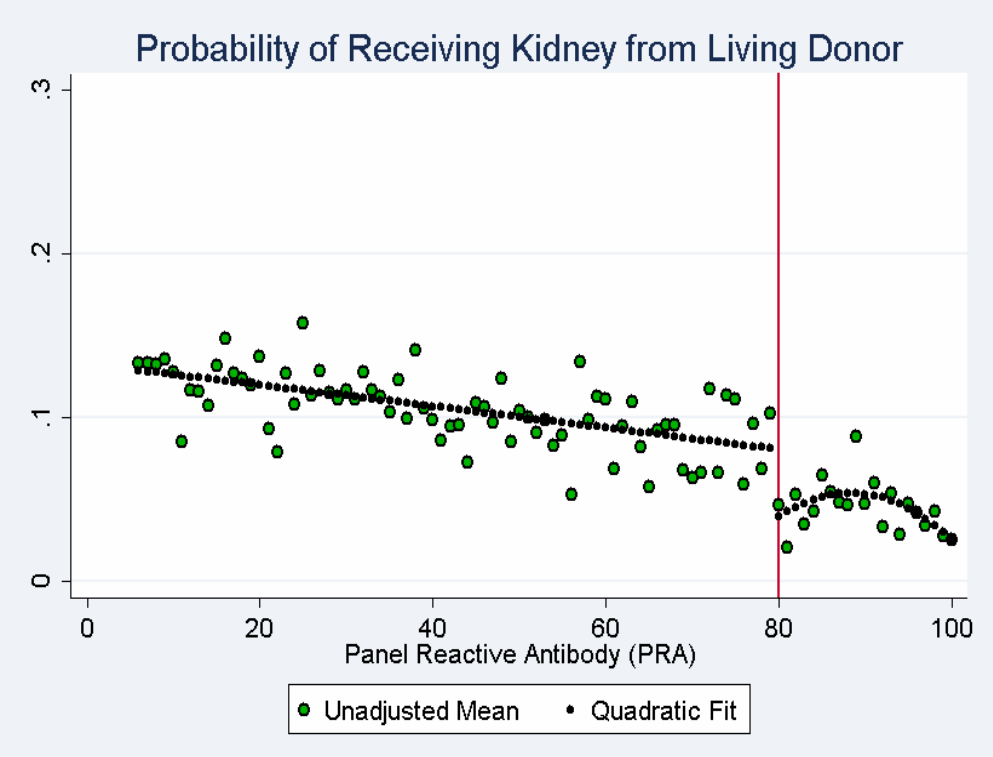

Figure 1.2

that a patient receives a kidney from a living donor – against the PRA score. Consistent with the conceptual framework discussed above, there appears to be a break at the PRA cutoff in the trend of the conditional expectation that a patient receives a kidney from a living donor. In particular, it appears that patients who have better access to the pool of publicly available kidneys are less likely to receive a privately supplied kidney from a living donor. Visually, this discontinuous decrease appears to be small in magnitude, but note that the overall rate of living donor transplantation is also relatively low in the neighborhood of the cutoff.

Table 1.2 reports the estimated values of π_1 and γ_1, which measure the discontinuity in the probability of receiving a kidney from the public supply, and the discontinuity in the probability of receiving a kidney from a living donor, respectively. Panel A shows that the expected probability of receiving a publicly provided kidney increases by 13 to 16 percentage points for patients just above the PRA cutoff. For patients just below the PRA cutoff (with PRA values between 75 and 79), the unadjusted expected probability of receiving a kidney from the waitlist is 24.1 percent. Therefore, having a PRA value just above the cutoff increases the probability that a patient can access the pool of publicly available kidneys by 54 to 66 percent. Panel B of Table 1.2 shows that the probability that patients just above the PRA cutoff receive a kidney from a living donor decreases by 4.1 to 4.4 percentage points, on average. On a base of 8.7 percentage points, this decrease represents a 47 to 50 percent reduction in the probability of living donor transplantation, as compared to patients just below the cutoff. Therefore, having improved access to the pool of publicly available kidneys appears to create a disincentive for patients to opt for living donation. Panel B also shows that there are some individual characteristics that are strong predictors of whether a patient will receive a kidney from a living donor. In particular, older patients, non-white patients, and less educated patients are less likely

Table 1.2: Discontinuity Estimates

Panel A: Probability of Receiving Deceased Donor Kidney (first-stage)			
Discontinuity Estimate	0.159***	0.143***	0.129***
	(0.017)	(0.016)	(0.014)
Polynomial Order	2nd	2nd	2nd
Individual Characteristics Included?	N	Y	Y
Year Dummies Included?	N	N	Y
Transplant Center Dummies Included?	N	N	Y
Number of Observations	35,371	35,371	35,371
Panel B: Probability of Receiving a Kidney from a Living Donor (reduced-form)			
Discontinuity Estimate	-0.041***	-0.044***	-0.043***
	(0.010)	(0.009)	(0.009)
Age at Registration (in years)		-.002***	-.002***
		(.000)	(.000)
Female		.019***	.022***
		(.003)	(.004)
Race (*White* omitted):			
Black		-.060***	-.061***
		(.005)	(.005)
Hispanic		-.024***	-.013**
		(.004)	(.005)
Other		-.045***	-.041***
		(.008)	(.008)
Education (*HS or less* omitted):			
At least some college		.027***	.026***
		(.003)	(.003)
Missing		-.005	.005
		(.007)	(.004)
Primary Insurance (*Private* omitted):			
Medicare		-.059***	-.052***
		(.004)	(.004)
Medicaid		-.061***	-.058***
		(.006)	(.006)
Other		-.037***	-.043***
		(.011)	(.012)

Table 1.2: Discontinuity Estimates (continued)

Functional Status (*No Assistance* omitted)			
Some Assistance with ADL		-.021***	-.015**
		(.006)	(.006)
Total Assistance with ADL		.005	.008
		(.037)	(.037)
Missing		-.001	.005
		(.003)	(.006)
Blood Type (*Type O* omitted):			
A		.013***	.011**
		(.004)	(.004)
B		.007**	.007**
		(.003)	(.003)
AB		.007	.003
		(.008)	(.008)
Primary Diagnosis (*Other Diagnosis* omitted*):*			
Diabetes Type 1, Insulin Dependent		-.050***	-.047***
		(.007)	(.007)
Diabetes Type 2, Non - Insulin Dependent		-.030***	-.032***
		(.006)	(.006)
Diabetes Type 2, Insulin Dependent		-.034***	-.033***
		(.005)	(.005)
Hypertensive Nephrosclerosis		-.010***	-.008**
		(.004)	(.004)
Polycystic Kidneys		.001	.003
		(.007)	(.007)
Malignant Hypertension		-.023**	-.020**
		(.007)	(.008)
Polynomial Order	2nd	2nd	2nd
Individual Characteristics Included?	N	Y	Y
Year Dummies Included?	N	N	Y
Transplant Center Dummies Included?	N	N	Y
Number of Observations	39,111	39,111	39,111

Notes:

1. The number of observations used in the first-stage regression (Panel A) is lower than the number of observations used in the reduced-form equation because it is limited to patients that do not receive a kidney from a living donor. That is, the estimates in Panel A reflect the discontinuous increase in the probability of receiving a kidney from the waitlist if they forgo LDKT
2. Clustered Standard Errors in parentheses (clustered at PRA level)
3. *** p < .01, ** p < .05, * p < .1

to receive a kidney from a living donor. Privately insured patients are more likely to receive a kidney from a living donor, as compared to patients with other sources of insurance (Medicare, Medicaid, and other).

Panel B of Table 1.2 also provides an indirect test of the identifying assumptions discussed earlier. In columns 2 and 3 of the table, individual characteristics are included in the estimating equations. Yet when these characteristics are omitted from the analysis (as in the specification presented in column 1), the point estimate of the discontinuity remains virtually unchanged. In other words, even though many individual characteristics are strongly correlated with the probability that a patient receives a kidney from a living donor, because these factors are not changing discontinuously at the PRA cutoff, their exclusion from the specification in column 1 does not seem to result in omitted variables bias.

Table 1.3 presents the estimate of $\frac{\gamma_1}{\pi_1}$, which is the indirect least squares (ILS) estimate of the effect of increasing the probability of receiving a publicly provided kidney on the probability that a patient will receive a privately supplied kidney from a living donor. The estimates suggest that if the probability of receiving a publicly provided kidney increases from 0 to 100 percent, the probability that a patient receives a kidney from a living donor decreases by 26 to 34 percentage points. A more intuitive interpretation is that a 10 percentage point increase in the probability of accessing the pool of publicly available kidneys would decrease the likelihood that a patient receives a kidney from a living donor by 2.6 to 3.4 percentage points. In regards to the crowd-out hypothesis, these estimates imply that there exists an element of crowd-out of living kidney donation, but that the crowd-out is less than full. Therefore, policies aimed at increasing the pool of publicly available kidneys may have the unintended consequence

Table 1.3: Indirect Least Squares Estimates

Probability of Receiving Deceased Donor Kidney (first-stage)	0.159	0.143	0.129
Probability of Receiving Kidney from Living Donor (reduced-form)	-0.041	-0.044	-0.043
ILS Estimate (reduced-form / first-stage)	-0.259***	-0.306***	-0.335***
	(0.069)	(0.075)	(0.084)
Individual Characteristics Included?	N	Y	Y
Year Dummies Included?	N	N	Y
Transplant Center Dummies Included?	N	N	Y
Notes:			
1. Bootstrapped Standard Errors appear in parentheses			
2. *** p < .01, ** p < .05, * p < .1			

of crowding out some living kidney donation, but such a policy would yield a net increase in the number of total kidney transplants performed.

1.7 Validity of the Regression Discontinuity Design and Robustness Checks

The results in Table 1.2 provide indirect support of the identifying assumptions of the regression discontinuity design. In order to provide more direct evidence, I estimate the following equation:

(1.3) $$LDKT_i = \beta_0 + \beta_1 PRA_i + \phi' X_i + \xi_i$$

This equation is analogous to equation (1.2), except that PRA only enters linearly in equation (1.3), and includes patient characteristics (captured in the vector X_i).[10] I estimate equation (1.3), and then calculate the predicted value of $LDKT_i$ (denoted by $\widehat{LDKT_i}$) for each observation. I then plot the conditional expectation of $\widehat{LDKT_i}$ against PRA values. If any patient characteristics are changing discontinuously at the PRA cutoff, then the there will be a break in the conditional

[10] In other words, equation (1.3) is the version of equation (1.2) without the quadratic term in PRA, the indicator $PRA80_i$, and without the interaction terms of the indicator variable and linear and quadratic terms of PRA.

expectation of $\widehat{LDKT_i}$ at the PRA cutoff.[11] If this occurs, then it implies that the identifying assumptions do not hold; namely, that patients in the immediate neighborhood are different, on average, along observable dimensions that affect the probability that they receive a kidney from a living donor. If this is true, then the estimates presented in Tables 1.2 and 1.3 cannot be interpreted as the causal effect of having improved access to the pool of publicly available kidneys on the probability that a patient receives a kidney from a living donor.

Figure 1.3 plots the conditional expectation of predicted values of $LDKT_i$ from estimating equation (1.3) against the range of PRA values. Visually, it appears that the predicted values of $LDKT_i$ are trending smoothly through the PRA cutoff; no breaks are visible. This implies that observable characteristics are not changing discontinuously at the cutoff. Note that this does not rule out the possibility that patients on either side of the cutoff are different, on average, along unobserved dimensions. But considering the specific context (i.e. the running variable is the result from a laboratory blood test), and the fact that the discontinuity estimates are robust to the inclusion of a wide range of covariates, it seems that the identifying assumptions are satisfied.

In order to test the robustness of the empirical results to the choice of PRA range, I re-estimate equation (1.2) using varying ranges of PRA. Table 1.4 reports the results of this analysis. The top left cell reports the discontinuity estimate presented in Panel B of Table 1.2, which is estimated using observations with PRA values between 6 and 100. Moving down the rows, the lower bound of the PRA range moves closer to the PRA cutoff of 80 (from the left). Likewise, moving across the columns (from left to right), the upper bound of the PRA range

[11] In this sense, this is like an "omnibus" check of the observable characteristics. One-by-one visual inspection of the conditional expectation of the covariates implies that observable characteristics are trending smoothly the PRA cutoff. For brevity, these graphs are omitted from the paper, but they are available from the author upon request.

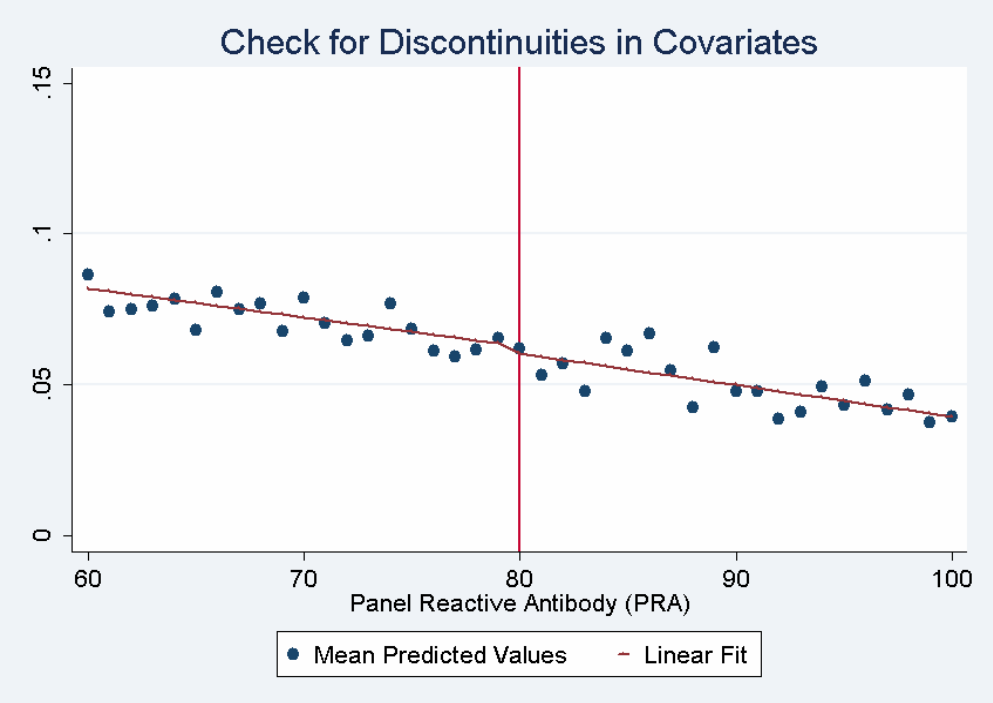

Figure 1.3

moves closer to the PRA cutoff (from the right). The point estimates remains strongly statistically significant over a wide spectrum of possible PRA ranges used in estimation, and are largely similar in magnitude. The estimate reported in the southeast corner of the table results from the most restrictive sub-sample; only observations within 5 points of either side of the PRA cutoff are used. This point estimate is fairly close to the others reported in the table, although it is less precisely estimated.

While the sample was constructed to exclude patients that might experience a sensitizing event while on the waitlist (and thus a changing PRA score), one concern is that patient PRA values change over time, even in the absence of a sensitizing event. This could occur if the laboratory test used to calculate PRA generates different PRA scores from month to month (perhaps due to random chance). Such random variation would be troubling in the current

Table 1.4: Reduced-Form Estimate Using Different PRA Ranges

		Upper Bound of PRA Range (inclusive)				
		100	99	94	89	84
Lower Bound of PRA Range (inclusive)	6	**-.043*** **	-.043***	-.042***	-.039***	-.042***
		(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
	9	-.038***	-.038***	-.037***	-.034***	-.037***
		(0.009)	(0.009)	(0.009)	(0.008)	(0.008)
	14	-.044***	-.044***	-.043***	-.040***	-.043***
		(0.009)	(0.009)	(0.009)	(0.008)	(0.009)
	19	-.041***	-.040***	-.039***	-.036***	-.039***
		(0.010)	(0.010)	(0.010)	(0.009)	(0.010)
	24	-.046***	-.046***	-.045***	-.042***	-.044***
		(0.010)	(0.010)	(0.011)	(0.010)	(0.010)
	29	-.045***	-.044***	-.044***	-.040***	-.042***
		(0.010)	(0.010)	(0.011)	(0.010)	(0.010)
	34	-.044***	-.044***	-.042***	-.039***	-.040***
		(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
	39	-.039***	-.039***	-.038***	-.035***	-.037***
		(0.011)	(0.011)	(0.012)	(0.011)	(0.011)
	44	-.044***	-.044***	-.043***	-.040***	-.042***
		(0.012)	(0.012)	(0.012)	(0.011)	(0.012)
	49	-.044***	-.044***	-.043***	-.040***	-.041***
		(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
	54	-.045***	-.045***	-.044***	-.041***	-.043***
		(0.013)	(0.013)	(0.013)	(0.012)	(0.013)
	59	-.052***	-.053***	-.051***	-.049***	-.050***
		(0.014)	(0.014)	(0.013)	(0.013)	(0.013)
	64	-.045***	-.045***	-.044***	-.041***	-.044***
		(0.017)	(0.017)	(0.017)	(0.016)	(0.015)
	69	-.036**	-.039***	-.047***	-.050***	-.050***
		(0.015)	(0.015)	(0.014)	(0.014)	(0.014)
	74	-.02	-.025	-.034*	-.038**	-.036*
		(0.020)	(0.020)	(0.018)	(0.018)	(0.019)

Notes:

1. Table reports the results of 75 separate regressions, where a different range of PRA scores was used in each regression to estimate the discontinuity in the probability that a patient receives a kidney from a living donor.
2. The estimate in the northwest corner is the previously reported discontinuity estimate reported in Table 2.
3. All equations include controls for patient characteristics, registration year fixed effects, and transplant center fixed effects
4. All equations use a quadratic of PRA except the ones used to estimate the discontinuities reported in the last two rows. In these cases, a linear specification was used because the conditional expectation of the the dependent variable appears "locally linear" near the cutoff.
5. Clustered standard errors appear in parentheses.
6. *** p < .01, ** p < .05, * p < .1

context because it would move patients above and below the PRA cutoff, thus giving and then removing priority access to the pool of publicly provided kidneys in a random fashion over time. I address this concern in two ways. First, I contacted histocompatibility labs to inquire how often it happens that a patient can "bounce" around the PRA cutoff. Based on the responses of the labs I contacted, it appears that this random variation happens with relatively low frequency. In addition, even though patients submit monthly blood samples, their PRA is not necessarily calculated every month. The testing procedure is costly, and is not always fully reimbursed, so some labs calculate the PRA score much less frequently than once a month. This would limit the number of chances for a patient's PRA score to change over time.

Empirically, I partially address this concern by examining two sub-samples of patients. Recall that the data reports two PRA values: current PRA and peak PRA. First, I eliminate patients who show evidence of "straddling" the PRA cutoff; these are patients whose current PRA is less than 80, but their peak PRA is 80 or higher. These patients have, at one point in time, had a PRA value above the PRA cutoff, but currently do not. That is, they have exhibited exactly the variation in PRA that would cause concern in the context of utilizing the regression discontinuity design. Among the 39,111 patients in the estimation sample, 4,418 patients (11%) have PRA values that "straddle" the cutoff. I exclude these observations and re-estimate equation (1.2) on the 34,693 observations that do not "straddle" the PRA cutoff. For the second sub-sample, I eliminate any patient whose current PRA is not equal to their peak PRA. These are patients who have experienced some variation in their PRA scores over time. The remaining patients exhibit some "stability" in their PRA values. Among the estimation sample, 13,885 patients (36%) have "stable" PRA values in that their current and peak PRA take on the same

values in the data. This approach is more conservative in that it restricts the analysis to patients for whom I do not observe variation in their PRA values in the data.

Table 1.5 reports the estimates from re-estimating equation (1.2) using the two sub-samples. The first row reports the estimates using the full estimation sample that appear in Panel B of Table 1.2. Eliminating patients that have PRA values that "straddle" the cutoff does not have any significant impact on the point estimates of the discontinuity in the probability of receiving a kidney from a living donor. Interestingly, though, the point estimates increase in magnitude when the analysis is conducted on the sub-sample of patients with "stable" PRA values. While it is not immediately clear why there would be heterogeneous treatment effects for patients with relatively stable PRA values and those that do not, one possible explanation is that

Table 1.5: Analysis on Sub-samples

Probability of Receiving Kidney from Living Donor			
Discontinuity Estimate	-0.041***	-0.044***	-0.043***
Sample: Full estimation sample (n = 39,111)	(0.010)	(0.009)	(0.009)
Discontinuity Estimate	-0.045***	-0.049***	-0.052***
Sample: Eliminate observations that "straddle" cutoff (n = 34,693)	(0.011)	(0.011)	(0.011)
Discontinuity Estimate	-0.070***	-0.071***	-0.071***
Sample: Patients with "stable" PRA (n = 13,885)	(0.017)	(0.018)	(0.017)
Polynomial Order	2nd	2nd	2nd
Individual Characteristics Included?	N	Y	Y
Year Dummies Included?	N	N	Y
Transplant Center Dummies Included?	N	N	Y
Notes:			
1. Clustered Standard Errors in parentheses (clustered at PRA level)			
2. *** p < .01, ** p < .05, * p < .1			

patients with stable PRA values know with more certainty their position on the PRA scale, and in particular, on which side of the cutoff they lie. A patient who has experienced variation in her PRA value will know with relatively less certainty, at any given point in time, where their PRA value lies on the PRA distribution. Therefore, it could be that patients who have PRA values near the cutoff, but who have also exhibited variation in their PRA values in the past, may hold off on opting for living donor transplantation with the expectation (or hope) that they will end up with a high enough PRA value to place them above the cutoff (and thus qualify for improved access to the pool of publicly available kidneys).[12]

Another concern is that the point estimates reflect a spurious change in the conditional expectation of $LDKT_i$ at the PRA cutoff, rather than any kind of meaningful economic behavior. In order to test for this, I re-estimate equation (1.2) using a range of placebo cutoffs. In particular, I estimate the "discontinuities" at each PRA value between 10 and 75, and between 85 and 95. This yields 77 additional discontinuity estimates, in addition to the discontinuity calculated at the PRA cutoff of 80. Figure 1.4 plots the histogram of these 78 point estimates. As expected, the mean and median point estimate estimated at the placebo cutoffs is close to zero. More importantly, the discontinuity estimated at the PRA cutoff of 80 is in the tail of the distribution of these estimates. This implies that the discontinuity estimates reported in Table 2 are not a "fluke" of the data.

1.8 Conclusion

This paper examines whether patients with improved access to the pool of publicly available kidneys are less likely to receive a privately supplied kidney from a living donor. Using

[12] If this hypothesis is true, then the presence of the PRA "notch" at 80 is still distorting private behavior.

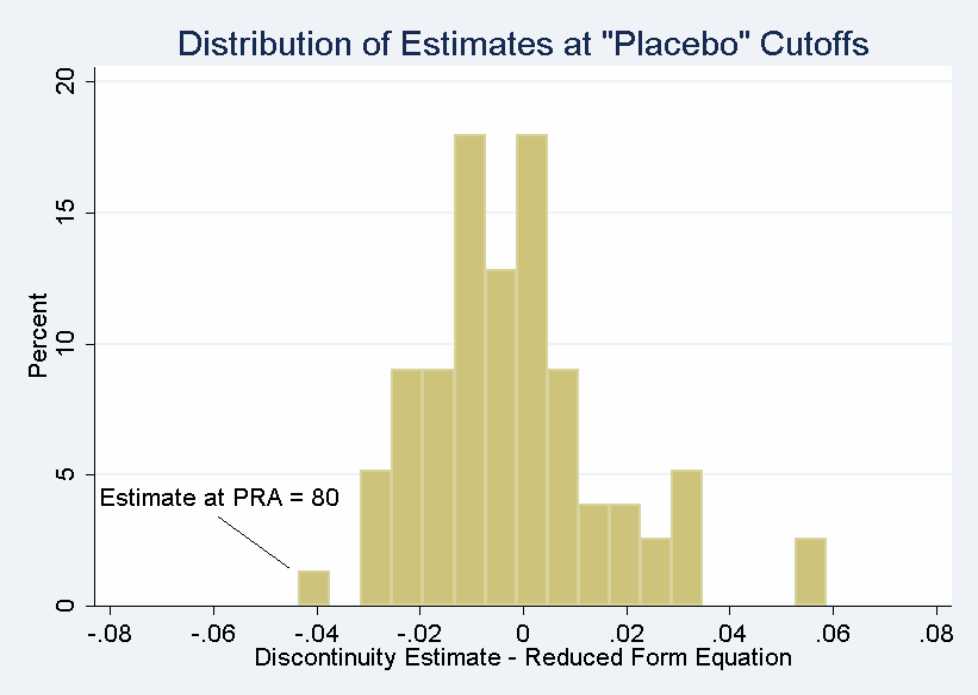

Figure 1.4

data on the universe of kidney waitlist and transplant activity from 1997 through 2006, I confirm this hypothesis empirically. Identification comes from a discontinuity in the probability that the patient can receive a publicly available kidney that is generated by the kidney allocation policy. The estimates imply that a 10 percentage point increase in the probability that a patient will receive a publicly provided kidney from the waitlist decreases the probability that a patient will receive a kidney privately from a living donor by 2.6 to 3.4 percentage points. These estimates appear to be internally valid, and are robust to a variety of specification checks.

The results imply that the public provision of kidneys may not fully crowd-out the private supply of kidneys from living donors. In particular, the crowd-out effect estimated here is smaller than the crowd-out implied by public programs like public health insurance (Cutler and Gruber 1996 ; Gruber and Simon 2008) and unemployment insurance (Cullen and Gruber 2000). Given the potential costs and disutility that can be imposed on the patient and her social network

by opting for living kidney donation, a natural question is why the effect estimated here is not larger. There are a few factors to consider. First, patients and members of their social network may be exhibiting some level of risk aversion. Even though the allocation policy increases the probability of receiving a publicly available kidney for patients just above the PRA cutoff, it does not ensure with 100 percent certainty that the patient will receive a kidney from the waitlist. In the face of this uncertainty, some patients and members of their social network may always opt for living donation since that option yields a kidney (and improved health) for the patient with certainty.

Another possibility is that there are quality differences between kidneys from living donors and kidneys procured from deceased donors. Indeed, it is a well-established fact in the transplant literature that kidneys from living donors survive longer than kidneys from deceased donors. If some patients and their families demand higher quality kidneys, then they may always opt for living donation, regardless of the patient's ability to access the pool of publicly available kidneys. Put differently, as long as the quality level of publicly available kidneys remains sufficiently low, then there may exist some portion of patients that will always opt for living kidney donation. This behavior would be consistent with the predictions of the theoretical work on the public provision of private goods, as in Besley and Coate (1991).

Finally, the results of this paper support the findings of Howard (forthcoming), and have salient policy implications. As the transplant community considers policies that are aimed at increasing the number of deceased donors, such as presumed consent for organ donation, it is important that it understands the potential for crowd-out of living kidney donation. The results of this analysis imply that the crowd-out would be less than full, which suggests that the effect of expanding the pool of publicly available kidneys would be a net increase in the total number of

kidney transplants performed. In addition, while “crowd-out” is often considered in pejorative terms, in this case it may not represent a social inefficiency, from an economic point of view. In particular, given the potential disutility and anxiety associated with having a loved one go through the donation process, the ability to substitute away from living donation toward utilizing publicly provided kidneys for transplantation may be welfare enhancing for patients and their families.

CHAPTER 2

Exploring the Volume-Outcome Effect: Evidence from Kidney Transplantation

2.1 Introduction

The association between provider volume and patient outcomes is well documented in the medical literature. In comprehensive literature reviews, both Halm, Lee, and Chassin (2002) and Choudhury, Dagash, and Pierro (2007) conclude that over 70 percent of articles on the topic find that patients that are treated by higher volume providers experience relatively good outcomes. Recently, economists and health services researchers have questioned whether this association represents a causal relationship, and if so, which direction the causation runs. There are three scenarios to consider. First, by performing a procedure more often, a provider (either an individual physician, or the entire hospital) becomes more skilled at the procedure, which reduces the probability of complications for the patient. This "practice makes perfect" hypothesis implies that there exist some kind of economies of scale with respect to patient outcomes (Gaynor, Seider, and Vogt 2005). An alternate explanation is that the causal relationship runs from outcomes to volume. In this "selective referral" hypothesis (Luft, Hunt, and Maerki 1987), providers vary in unobserved (to the researcher) quality, and physicians refer their patients to the higher quality providers. Finally, patients may self-select to higher or lower volume providers based on unobserved dimensions of health. For example, the observed pattern in the medical literature could be explained by "sicker" patients sorting themselves to low volume providers. On the other hand, sicker patients could differentially sort themselves to higher volume providers, in which case the magnitude of the volume-outcome relationship may be understated if patient selection is not accounted for.

In this paper, I examine the volume-outcome relationship in the context of kidney transplantation. In 2010, 13,472 adults received a kidney transplant; 63 percent of transplanted kidneys were from deceased donors, and the remaining 37 percent from living donors.[13] While kidney transplantation remains the best treatment option for patients with kidney failure, the success of the procedure is not always certain. Based on national statistics, approximately 8 percent of recipients will experience failure of the transplanted kidney within one year of their transplant (OPTN/SRTR 2009). Many factors contribute to graft survival – the quality of the transplanted organ, the immunological match between the donor and the recipient, and even the amount of time between when the organ is recovered and when it is transplanted.[14] The question I address in this paper is whether there exists a causal effect of performing more transplants on outcomes for transplant recipients. Axelrod et al. (2004) document that higher volume kidney transplant centers have lower rates of one-year graft failure rates than lower volume transplant centers. However, the authors do not address the likelihood that there are other differences between transplant centers other than volume that are correlated with patient outcomes.

Identifying the true causal effect of transplant center volume on recipient outcomes is important because Medicare is the primary payer for kidney transplants in the United States; in 2008, the program spent $23 billion on the care of patients with end stage renal disease, $242 million of which was spent on kidney transplants (USRDS 2010). When a recipient experiences graft failure, she will return to dialysis, which is the least cost-effective treatment option for patients with kidney failure (Mullins et al 2003). Medicare pays approximately $91,000 per year for a patient that has experienced graft failure (USRDS 2010). If there does exist a causal

[13] Data taken from Organ Procurement and Transplantation Network (OPTN) data website, available here: http://optn.transplant.hrsa.gov/data/

[14] The “graft” is the transplanted kidney.

volume-outcome relationship, then Medicare (and other insurers) may realize cost savings if transplants were concentrated at a fewer number of transplant centers.

A unique feature of addressing the volume-outcome relationship in kidney transplantation is that the number of transplants that a transplant center performs is limited by the available supply of transplantable kidneys. This fact allows me to employ several strategies that plausibly identify the causal effect of volume on patient outcomes. The first is to estimate transplant center-by-year fixed effects models which would eliminate any biases generated by differences in unobserved center quality that is likely correlated with volume. In this framework, within-transplant center variation in volume, driven by plausibly exogenous shocks in the supply of kidneys, is used to explain within-center variation in outcomes for otherwise identical patients. As discussed later, the use of transplant center fixed effects may not fully address systematic unobserved differences in patients that are correlated with volume. Thus, I also use an instrumental variables approach to exploit plausibly exogenous shifts in transplant center volume. Usually, kidneys are allocated regionally, with recovered organs being offered first to patients registered at local transplant centers. One exception to this rule is that if a patient on the waitlist is a perfect immunological match to a recovered kidney, that patient is offered the kidney regardless of her geographical distance from the kidney. That is, when a kidney is a "perfect match" to someone on the national waitlist, the usual regional allocation policy is superseded and the kidney is offered nationally. From the perspective of the recipient's transplant center, the offer of a "perfect match" kidney represents one additional transplant that it would not have otherwise performed. Therefore, in this second estimation approach, within-transplant center variation in the number of "perfect matches" will be used in the first stage to predict within-

center variation in volume. Because these "perfect matches" occur in a random fashion, I argue that they meet the usual requirements for instrument validity.

The results suggest that "naïve" models that do not account for between hospital differences in underlying quality overstate the volume-outcome relationship. For example, results from pooled OLS models suggest that transplant volume significantly reduces the probability of adverse patient outcomes, both in the short term (i.e. one week post-transplant) and longer term (one year post-transplant). Estimates from fixed-effects and instrumental variables models, however, suggest a smaller and often statistically insignificant relationship between transplant center volume and patient outcomes. Taken together, the results imply that much, if not all, of the observed volume-outcome relationship observed in kidney transplantation is due to underlying differences across transplant centers. From a policy perspective, these results require careful consideration. On one hand, transplant candidates should consider registering at higher volume transplant centers because those centers have, on average, better patient outcomes than lower volume transplant centers. On the other hand, it is unknown if there might exist diseconomies of scale if transplants became more concentrated at relatively few transplant centers.

The rest of the paper is organized as follows. The next section provides background on the volume-outcome literature in general, and specifically in the case of kidney transplantation. The third section describes the institutional setting. The fourth section provides a discussion of which measures of provider volume are typically used in the literature, and which one I use in this analysis. The fifth section discusses the identification strategy and estimation, which is followed by a description of the dataset and construction of the estimation sample. The seventh

section presents the empirical results, which is followed by a discussion of the policy implications of these empirical findings. The last section concludes the paper.

2.2 The Volume-Outcome Effect

There exists a large number of studies on the volume-outcome relationship in medical care. The consensus of these studies is that patients who receive their care at providers that perform a relatively large number of a procedure experience better outcomes than patients who have their procedure performed at lower volume providers. This relationship is usually attributed to "learning-by-doing" or "practice makes perfect:" higher volume providers have a greater stock of experience that they can draw upon, thus yielding improved patient outcomes. If "practice makes perfect" explains the observed volume-outcome relationship, then the direction of causality runs from volume to outcomes. While this explanation has intuitive appeal, the empirical challenge lies in disentangling the volume effect from other mechanisms that also explain the observed pattern in the data. For example, provider volume may simply be a proxy for underlying quality of providers. In other words, patients are attracted to, or are referred to, providers who have better patient outcomes (this is the "selective referral" hypothesis (Luft, Hunt, and Maerki 1987)). If this is true, the direction of causality is reversed: outcomes drive volume. In addition, patients may differentially sort themselves to high or low volume providers along unobserved (to the researcher) dimensions of underlying health. Therefore, a regression of patient outcomes on provider volume does not necessarily yield an unbiased estimate of the effect of performing an additional procedure on patient outcomes.

There are two main strategies used in the health economics and health services research that attempt to "control" for any kind of selective sorting by patients. The first is to include

provider fixed-effects in the estimations to control for time-invariant differences across providers that are correlated with both volume and outcomes. Therefore, the volume effect is identified from within-provider variation in the number of procedures performed over time. Hamilton and Hamilton (1997) show that differences in outcomes for hip surgery patients in Quebec are explained largely by between-hospital differences in unobserved determinants of outcomes; the fixed-effect estimate of the volume effect is statistically insignificant. Ho (2002) shows that within-hospital increases in PTCA volume are associated with relatively small, yet statistically significant, improvements in patient outcomes. In either case, the fixed effects specification assumes that within-provider variation in volume is uncorrelated with other determinants of outcomes that are unobserved to the researcher. This may be an innocuous assumption over shorter panels, but may be less tenable over longer time spans, because unobserved factors that affect outcomes may change within a provider over time. For example, a hospital may contract with higher quality surgeons or nurses over time. In this scenario, the hospital attracts more patients because of the improved quality of the surgical team, and thus within-hospital variation in volume is correlated with unobserved (to the researcher) changes in underlying quality at the hospital.

More recent studies have used an instrumental variables (IV) approach (Gaynor, Seider, and Vogt 2005; Gowrisankaran, Ho, and Town 2006; Tsai et al 2006; Huesch 2009). In these studies, patient distance from a provider serves as a plausibly exogenous predictor of provider volume. In the first stage of the analysis, distance from the patient's home to each provider in the patient's choice set is assumed to be an exogenous predictor for where the patient receives her care. The identifying assumption is that patients do not choose their residence based on unobserved preferences for having (future) care performed at a particular provider. Likewise, it

is assumed that providers do not locate where they expect demand for their services to be the greatest. For each provider a measure of expected volume serves as an instrument for actual volume. Across all of these studies, the volume-outcome effect is much smaller in the IV specification than in the ordinary least squares (OLS) specification, and is usually statistically insignificant. However, Gaynor, Seider, and Vogt (2005) and Gowrisankaran, Ho, and Town (2006) conclude that provider volume can be treated as exogenous.[15]

There is one paper in the medical literature that has examined the volume-outcome relationship in the context of adult kidney transplantation. Axelrod et al (2004) compare the outcomes of transplant recipients, defined as graft failure at one year post-transplant, across hospitals that lie in different quartiles of average annual volume across the study period (1996 through 2000). Their results suggest that transplant recipients who have their transplant performed at either "low" or "very low" volume transplant centers are more likely to experience graft failure within one year of their transplant. In addition, they find that most of this effect is driven by short-term graft failure (graft failure within one month of transplant). However, the analysis does not account for differences between transplant centers that may be correlated with both transplant volume and patient outcomes. One advantage of their study, however, is that they include clinical measures of patient illness, which is an improvement over studies that simply control for patient demographics (Tsai et al. 2006).

Should transplant center volume be treated as exogenous? The available evidence suggests not, and points to both selective referral and patient selection as plausible explanations for Axelrod et al's (2004) findings. First, unlike other procedures studied in the volume-outcome

[15] In both studies, the estimates from the IV specification are statistically insignificant. Gaynor et al (2005) reject the null that the IV estimate is statistically different from the estimate generated by the specification that treats volume as exogenous. However, the p-value of the test is .06, which is very close to the usual p-value of .05 that is used to reject the null hypothesis. Gowrisankaran et al (2006) state that "any evidence suggesting endogeneity was not overwhelming", but do not provide any statistical evidence for their claim.

literature, transplants do not occur immediately after the patient chooses her provider. Kidney failure is a gradual disease that is most often caused by hypertension and diabetes. When patients learn that they need a kidney transplant, they choose the transplant center at which they will receive their transplant, as well as their pre- and post-surgical care. Because of excess demand for kidneys in the United States, the median waiting time for a transplant is over 3 years, although there is significant regional variation in waiting times (OPTN/SRTR 2009). In the meantime, patients with kidney failure can rely on dialysis as a maintenance treatment. Therefore, relative to other surgical procedures, potential transplant candidates have significant time to choose which transplant center at which they will receive their care. They may do this passively (i.e. they go with the recommendation of their nephrologists), or they may actively compare transplant centers using transplant center performance reports that are publicly available on the Internet. These reports are published twice a year by the Scientific Registry of Transplant Recipients (SRTR) and include information (at the center level) on the number of transplants performed, the median time to transplant for patients, and actual graft and patient survival rates. In addition, the reports include expected rates of graft and patient survival that are constructed for each transplant center using the characteristics of transplant recipients at that particular center. The reports also provide a test of whether the difference between the actual and expected graft and patient survival rates are statistically significant. The difference between actual and expected performance can be considered a proxy for the quality of care provided at a particular center.

Howard (2008) tests whether transplant candidates sort themselves based on this measure of quality across transplant centers within their choice set. In particular, he shows that private insurers contract with the transplant centers that perform better (in regards to recipient

outcomes), and therefore privately insured individuals are directed to better performing centers, which is consistent with the selective-referral hypothesis. A surprising result is that more educated patients are differentially likely to choose a better performing transplant center than less educated patients, as would be expected. One caveat is that the data span registrations from 2000 to 2002, when Internet use was less widespread than it is currently. Therefore, the SRTR reports may not have been as visible to transplant candidates as they are today.

Furthermore, the descriptive statistics provided in Axelrod et al (2004) show that there are significant differences in the observable characteristics of both transplant recipients and the donors across the volume quartiles. In particular, when comparing "very low" to "high" volume transplant centers, the authors show statistically significant differences in transplant recipient age, race, primary diagnosis, time on dialysis, as well as significant differences in donor age, status (living versus deceased), ischemic time, and expanded criteria designation.[16] Because these characteristics are also independent predictors of outcomes, simply controlling for these factors may not account for unobserved heterogeneity in patient characteristics that varies systematically by transplant center volume that also affects patient outcomes.

A natural question is why one should expect patient outcomes to be related to transplant volume, especially since the surgical procedure itself has changed very little since 1951 and rates of surgical complications are low (Humar and Matas 2005). Some of the causes of graft failure are directly related to surgical technique: renal artery thrombosis can be caused by the kinking of blood vessels and renal artery stenosis can be caused by improper suturing technique (Humar and Matas 2005). Apart from more practice at the surgical technique, increased volume and

[16] Expanded criteria donors (ECDs) are donors that meet the following criteria: a donor age 60 or older, or a donor age 50 or older with two of the following three conditions: history of high blood pressure, high serum creatinine (an indication of impaired kidney function), or death resulting from stroke. Kidneys from ECDs are considered to be of marginal quality.

experience dealing with recipients may improve the ability of transplant centers to recognize, and treat, the early signs of complications. Larger volume transplant centers will have more experience evaluating the quality of donor kidneys, a key determinant of patient outcomes, and whether a kidney from donor *X* would be an appropriate match for patient *Y*. Post-transplant, the probability of graft failure or patient mortality will depend in part on the patient's adherence to immunosuppressive medications. Larger volume transplant centers may be better able to monitor their recipients' post-transplant medication regimen than lower volume centers. On a similar note, the availability of new immunosuppressive medications have increased over time, which has provided transplant centers more flexibility in tailoring patient-specific post-transplant drug regimens. Danovitch (2001) points out that "transplant centers tend to be loyal to their own [immunosuppressive] protocols, which often have been developed in response to local experience." Therefore, larger transplant programs have more experience to draw upon when it comes to designing effective immunosuppressive drug regimens for their patients. This experience can translate to lower rates of graft failure or patient mortality.

2.3 Institutional Setting

The goal of this paper is to identify the volume effect in kidney transplantation, if one exists. A description of the institutional setting is required to motivate the empirical strategy. As noted earlier, one unique aspect of examining the volume-outcome relationship in kidney transplantation is the fact that the number of transplants that a center performs is limited by the number of available transplantable kidneys. The supply of transplantable kidneys comes from two sources: deceased donors and living donors. Kidneys from deceased donors are allocated according to strict allocation rules developed and maintained by the United Network of Organ

Sharing (UNOS). These kidneys are allocated regionally, with kidneys first offered to patients registered to transplant centers within the same local area of the donor (i.e. within the same organ procurement organization (OPO) service area). Among blood-compatible patients within the OPO, patients with the longest waiting time are typically offered the kidney first, although there are some exceptions to that rule. Therefore, transplant center volume is a function of the local supply of kidneys, as well as the number of candidates waitlisted at a transplant center. Within a particular OPO, the transplant center with more candidates will, on average, perform more transplants than centers with fewer candidates on the waitlist.

Therefore, the key question is what drives variation in the size of waitlists across transplant centers, even within a local area. That is, why are transplant candidates more likely to register at some transplant centers rather than others? There are three factors to consider. The first is distance from the patient's home to the transplant center. Once a kidney becomes available for a patient, the patient has a window of time to get to her transplant center. If the patient cannot get to the transplant hospital in sufficient time, then she will be passed over and the kidney will be offered to another candidate on the waitlist.[17] In addition, while the candidate waits for her transplant, she must go to her transplant center for periodic examinations to identify changes in the patient's condition, or her suitability as a transplant candidate. Therefore, the number of transplant centers within a patient's choice set will depend on her proximity to those transplant centers. Second, the patient's insurer will play a large role in a patient's choice of transplant center. Patients covered by Medicare can register at any of the certified transplant centers in the country. Privately insured patients will be limited to the transplant centers with

[17]There is no published rule that strictly defines the window of time that a patient has to get to her transplant center. The decision of whether the patient can get to her transplant center in sufficient time is likely made on a case by case basis.

whom their insurer has contracted. Medicaid patients will be limited to transplant centers within their state of residence.

Finally, as discussed earlier, patients may be sort themselves to better performing transplant centers within their choice set. They may do this actively by using the biannual center-specific performance reports published by SRTR. Or they may be directed to better performing centers by their physicians or their insurer (Howard 2008). This is consistent with the selective referral hypothesis whereby volume is caused by unobserved (to the researcher) underlying differences in the quality of care provided by transplant centers. The challenge then is to disentangle any kind of selective sorting from the true volume effect.

2.4 Which Volume Measure to Use?

Before discussing the estimation strategy, I turn to the issue of which volume measure to use in the analysis. In the literature, the usual measures of volume are either the number of procedures performed at a hospital in the year of a patient's procedure (as in Gaynor, Seider, and Vogt (2005)) or the number of procedures performed in the quarter of a patient's procedure (as in Gowrisankaran, Ho, and Town (2006) and Huesch (2009)). These measures of volume are likely an artifact of the datasets that are used for analysis.[18] In addition, because these are contemporaneous measures, they implicitly include procedures that are performed by a provider after a given patient's procedure.[19] Apart from these issues, the literature does not provide a discussion of which volume measure is the optimal measure of recent provider experience. That is, what is the appropriate "lookback" window to use when calculating a provider's volume? In

[18] These studies use state-level hospital discharge datasets that mask the exact date of a procedure, and instead include only the quarter or the year in which a procedure was performed.
[19] For example, in these studies, a patient who is treated in January 2006 would be assigned that hospital's 2006 level of volume.

the current context, I assert that 30 day volume is the best available measure to use to capture the effect of volume changes on transplant center performance. As I empirically show later, I find strong evidence of depreciation of experience from one 30 day "block" to another. In particular, holding constant 30 day volume, volume in the 31 to 60 days or the 61 to 90 days prior to a patient's transplant has no economically or statistically significant effect on transplant center performance, which implies that any benefit of transplant center experience in the 31 to 90 days prior to a patient's transplant depreciates rather quickly. This finding is consistent with Gowrisankaran, Ho, and Town (2006) and Huesch (2009) which both estimate large rates of depreciation of hospital and surgeon experience. Therefore, if experience depreciates rapidly from one 30 day "block" to another, then it stands to reason that the human capital of the transplant team may appreciate during above-average periods of transplant volume.

As discussed earlier, the surgical procedure to transplant a kidney has changed very little since 1951. Therefore, one might not expect changes in recent volume to affect the productivity of transplant surgeons since the procedure is well-established. However, kidney transplantation is a team-oriented process, which involves surgeons, transplant coordinators, nurses, etc. While an increase in volume may not affect surgeon performance, it may affect the performance of other team members, which is reflected in patient outcomes.

2.5 Empirical Strategy

As a first step in thinking about how volume affects the outcomes of kidney transplant recipients, I start with a baseline pooled OLS model of transplant recipient outcomes:

$$(2.1)\quad y_{iht} = \beta_0 + \beta_1 Vol_{iht,-30} + \beta_2 X_{iht}^{Dem} + \beta_3 X_{iht}^{Clin} + \beta_4 X_{iht}^{Don} + \varepsilon_{iht}$$

where y_{iht} represents the outcome of transplant recipient *i* who received her transplant in year *t* at transplant center *h*.[20] X_{iht}^{Dem}, X_{iht}^{Clin}, and X_{iht}^{Don} are vectors of patient demographic characteristics, patient clinical information, and donor characteristics, respectively. I include these covariates to adjust for differences in patient characteristics and quality of donor organs, which may be correlated with transplant center volume. Of interest is the parameter β_1, which captures the conditional correlation of the level of transplant volume at transplant center *h* in the 30 days preceding recipient *i*'s transplant with the recipient's post-transplant outcomes. Equation (2.1) uses both between and within transplant center variation in 30 day volume to estimate the effect of 30 day volume on patient outcomes. Given the discussion above, it is likely that the estimated effect will be biased, either due to patient selection along unobserved dimensions of health, or due to selective referral. Nonetheless, estimation of (2.1) provides a baseline set of results which serve as a starting point for the remaining analyses.

Decomposing the error term in (2.1) illustrates why using pooled OLS to estimate (2.1) will likely lead to a biased estimate of the volume effect:

(2.2) $$\varepsilon_{iht} = \alpha_{ht} + \lambda_i + \omega_{iht}$$

In (2.2), α_{ht} is an unobserved attribute (or a set of attributes) of transplant center *h* in year *t*, λ_i represents unobserved characteristics of transplant recipient *i*, and ω_{iht} is an idiosyncratic error

[20] It is important to note that the data do not include surgeon identifiers, so this analysis is done at the transplant center level. Since transplantation - from pre-transplant coordination, to donor evaluation, to the surgical process itself, and to post-transplant care - is a team-orientated process, identifying the volume effect at the organizational level is arguably more intuitive than focusing narrowly on the volume effect at the surgeon level.

term that is mean zero and assumed to be uncorrelated with any of the regressors in (2.1). Under the selective referral hypothesis, α_{ht} represents unobserved transplant center quality that varies over time, with the assumption that $\text{cov}(\alpha_{ht}, Vol_{ht,-30}) > 0$. Under this scenario, estimating (2.1) using OLS will generate a biased estimate of the volume effect. For example, if the outcome of interest is patient mortality (i.e. $y_{iht} = 1$ if the recipient dies, and 0 otherwise), the fact that $\text{cov}(\alpha_{ht}, Vol_{ht,-30}) > 0$ and $\text{cov}(\alpha_{ht}, y_{iht}) < 0$ indicates that the OLS estimate of the volume effect will be negatively biased. That is, failing to account for unobserved differences in underlying quality that are also correlated with volume will overstate (in magnitude) the size of the volume effect.

To address these sources of bias, I add transplant center-by-year fixed effects to the specification outlined in (2.1):

$$(2.3)\quad y_{iht} = \beta_0 + \beta_1 Vol_{iht,-30} + \beta_2 X_{iht}^{Dem} + \beta_3 X_{iht}^{Clin} + \beta_4 X_{iht}^{Don} + D_{ht} + \xi_{iht}$$

The volume effect estimated by (2.3) is identified from variation in 30 day volume within a given year at a given transplant center. The identifying assumption is that conditional on the fixed effects, the residual variation in volume is uncorrelated with ξ_{iht}, where $\xi_{iht} = \lambda_i + \omega_{iht}$. In the current setting, this may be a plausible assumption: changes in transplant center volume will be driven by exogenous supply shocks in the number of available transplantable kidneys. That is, the volume effect is identified by comparing otherwise observably identical recipients that go to the same transplant center in the same year, but that receive their transplants at times when the

transplant center has experienced changes in its 30 day volume, generated by the randomness of offers from the supply of transplantable kidneys.

A natural question is whether there is sufficient variation in 30 day volume within a transplant center in a given year to identify any effects on transplant center performance. Clearly, longer "lookback" windows (e.g. 60 or 90 day volume) will have less variation within a transplant center within a year. However, in the empirical evidence I present later, it is clear that there is enough variation in 30 day volume among transplant recipients within a transplant center in a year to identify the effects of changes in transplant center volume on performance, as measured by patient outcomes.

The motivation for using the fixed-effects specification to estimate the volume effect is that the volume and timing of transplants are determined exogenously by supply shocks in the number of kidneys that are made available to particular transplant centers. These shocks would apply to the supply of deceased donor kidneys, which are subject to the strict allocation rules described earlier. But approximately one-third of kidneys come from living donors, which are not subject to the same kinds of rules that govern the allocation of deceased donor kidneys. One concern is that the timing and number of living donor transplants may be correlated with the volume of deceased donor transplants performed. Because transplant recipients that opt for living donation may be different, on average, than those that opt to wait for a deceased donor along unobserved dimensions, this kind of correlation would lead to biased fixed-effects estimates.[21] Consider the following example. A transplant center schedules its living donor transplants such

[21] The donors are likely to be systemically different as well. Living donors go through intensive medical screening to ensure that they are healthy enough for the surgery. Part of the rationale for the intensive screening is that organ donation is a surgical procedure from which the living donor derives no direct medical benefit, but that does introduce the possibility of complication or death. This is an apparent deviation from the "first, do no harm" ethos in medicine. Thus, transplant programs aim to only accept the healthiest living donors. Therefore, kidneys from living donors are likely to be healthier, on average, than kidneys from deceased donors, which would be correlated with improved recipient outcomes.

that the number of living donor transplants is uniformly distributed over time within a year.[22] In periods with zero deceased donors, the total transplant volume of the center will simply equal the number of living donor transplants performed in that period. When the transplant center receives deceased donor kidneys from the local supply, then the total volume of transplants performed in a particular period is the sum of the living donor transplants and the deceased donor transplants performed in that period. In the fixed-effects framework, periods with few or no deceased donors (i.e. periods when just living donor transplants are performed) are more likely to be below the center's average level of total volume. In other words, the probability that a patient receives a living donor transplant will be systematically correlated with the within-transplant center variation in total transplant volume (in particular, living donor recipients are more likely to receive their transplants during "below average" periods of total transplant volume). While donor status (living versus deceased) can be controlled for, there may be unobserved differences in patient and donor health that are also correlated with deviations from the transplant center-specific average level of volume over time that are not adequately accounted for by the inclusion of donor status information. In the context of equation (2), the transplant center-by-year fixed effects, will not account for systematic differences in the patient specific component of the error term, λ_i.

One strategy to circumvent this challenge would be to run the analysis only on the recipients of deceased donor transplants. However, this would eliminate roughly one third of all transplant recipients from the analysis, and there is no a priori reason to believe that surgical process (or post-surgical treatment) of transplanting a kidney from a living donor is any different

[22] The idea that living donor transplants are scheduled bears out in the data. Less than one percent of all living donor transplants are performed on the weekend, and among weekdays, Tuesdays, Wednesdays, and Thursdays appear to be more popular days than either Monday or Friday. Deceased donor transplants, on the other hand, are uniformly distributed across all seven days of the week. See Chapter 3 for a more detailed discussion.

than transplanting a kidney from a deceased donor. Another strategy is to find an instrument that is uncorrelated with λ_i, and is also predictive of transplant center volume. The rules of the kidney allocation policy offer a plausibly valid instrument: perfect match kidneys. As discussed earlier, kidneys from deceased donors are typically transplanted locally (within an OPO service area). One exception to this rule is when a patient on the national waitlist is a perfect immunological match to a deceased donor kidney. Each time a kidney is recovered, the antigen profile of the donor is compared to the entire national waitlist of transplant candidates.[23] If a patient on the national waitlist is a perfect antigen match to the donor, then that patient is offered the kidney regardless of her geographical location.[24] For example, a kidney recovered in Florida that would otherwise be transplanted in a patient in Florida will instead be offered to a patient elsewhere (e.g. in California), if that patient is a perfect antigen match to the kidney. From the viewpoint of the recipient's transplant center, the offer of a perfect match kidney represents one additional transplant that it would not have otherwise performed in a given time period.

This aspect of the allocation policy is not unlike a lottery for a kidney. Each time a kidney is recovered anywhere in the United States, any person on the national waitlist has a chance to receive that kidney. For the purposes of my identification strategy, it is important to note that the only factor that determines whether a patient is offered a perfect match kidney is the level of the antigen matching. Neither patient characteristics nor characteristics of the patient's transplant center are taken into account when the offer is made. Given the data, I can empirically test the claim that the number of perfect match kidneys offered to a particular transplant center in

[23] Antigens are proteins that can trigger a person's immune system. Antibodies are "programmed" to attack antigens that are foreign to a person's body. In the context of kidney transplantation, antigen matching is important because an increased number of antigen mismatches increases the likelihood that the recipient's body will reject the transplanted kidney.

[24] There are six chances for an antigen mismatch. Therefore, a "perfect match" is one is which the donor and recipient express the same antigens, which greatly reduces the risk of organ rejection. This forms the rationale of perfect match/zero mismatch kidney sharing policy.

the last 30 days of recipient *i*'s transplant is uncorrelated with the recipient's (and her donor's) characteristics. As stated above, a perfect match kidney represents one additional transplant that the recipient's transplant center would not have otherwise performed in a given period of time. Therefore, I also expect that the instrument will be strongly correlated with 30 day transplant volume.

To implement the instrumental variables approach, I estimate the first-stage equation:

$$(2.4)\quad Vol_{iht,-30} = \pi_0 + \pi_1 PM_{iht,-30} + \pi_2 X_{iht}^{Dem} + \pi_3 X_{iht}^{Clin} + \pi_4 X_{iht}^{Don} + D_{ht} + \psi_{iht}$$

where $PM_{iht,-30}$ is the number of perfect match kidney transplants performed at the recipient *i*'s transplant center in the previous 30 days. The assumptions are 1) $\pi_1 > 0$ and 2) $\text{cov}(\xi_{iht}, \psi_{iht}) = 0$. The first assumption is empirically testable. The second assumption cannot be directly tested, but evidence that the instrument is uncorrelated with observable characteristics will provide support to the claim. Note that in equation (2.4), I exploit within-transplant center-year variation in the number of perfect match kidneys to predict within-transplant center-year variation in total transplant volume. The transplant center specific year fixed effects are required for the instrument to be valid. The number of perfect match kidneys offered to a particular transplant center will be a function of the number of transplant candidates registered at that transplant center. If variation in the number of registrations is driven by the selective referral hypothesis, then the instrument (in the cross-section) would be correlated with unobserved levels of quality that vary systematically between transplant centers. Therefore, the fixed effects removes bias associated with selective referral, and the instrumental variables approach should eliminate any remaining bias generated by within-transplant center variation in unobservable

patient characteristics that are correlated with both total transplant volume and patient outcomes. As usual, the fitted values of 30 day volume will replace the actual values of 30 day volume in the second stage equation.

To recap, the empirical strategy is three-pronged. First, I estimate pooled OLS equations to estimate the conditional correlation between 30 day volume and patient outcomes. Second, I add in transplant center specific year fixed effects, thereby addressing biases created by selective referral. As discussed above, the patient-specific component of the error term in the fixed effects equation may be correlated with within-transplant center volume. To address this, I use an instrumental variables approach as the last estimation strategy. For specifications where the dependent variable is binary, I estimate a linear probability model, which yields similar estimates to a probit specification, but allows for more intuitive interpretation of the estimated coefficients, as well as implementation of the IV estimator via two stage least squares. Standard errors are clustered at the transplant center-year level.

2.6 Data and Sample Construction

The data used in the analysis come from the United Network of Organ Sharing (UNOS) Standard Transplant Analysis and Research Files (STAR) on all kidney waitlist, transplant, and follow-up activity as of August 2008. These data include additions to the waitlist (new waitlist registrations) or removals from the waitlist, which can result from transplant, patient death, or if the patient is too ill to be considered a viable transplant recipient. The dataset includes detailed information on transplant candidates collected at the time of registration on the kidney waitlist, information about the transplant procedure (if a transplant is performed), and follow up information on the patient's outcomes after the transplant (again, if a transplant is performed).

The advantage of using this dataset is that it covers the universe of kidney transplants in the United States, and each transplant recipient is followed after the transplant is performed. Therefore, by using this dataset, I am able to overcome any shortcomings of using state hospital discharge records, which is common in the volume-outcome literature. For example, rather than observing just inpatient mortality, I observe outcomes in the longer run. In addition, the data include a number of clinical measures that allows me to more precisely adjust for patient risk, as compared to relying on the assumption that patient demographics adequately capture heterogeneity in underlying illness at the time of surgery.

I exclude pediatric transplant recipients from the analysis because they make up a small percentage of all transplants (4.4 percent in 2010), and they are a subsample that is quite different than the sample of adult transplant recipients.[25] This exclusion also removes all of the pediatric hospitals from the sample, where the vast majority of pediatric transplants are performed. Visual inspection of the names of the excluded transplant centers shows that all have the word "Children's" in their name, and that the names of the included hospitals do not have any indication in their title that they are a children's hospital.

The main volume measure is the number of transplants performed at a patient's transplant center in the 30 days preceding the recipient's transplant. This is a "finer" measure of recent volume than is typically used in the literature.[26]In addition, it avoids the issue of assigning contemporaneous volume, which would implicitly assume that future volume affects a patient's outcome. The rationale for using 30 day volume versus a broader look-back window was discussed earlier. Because the marginal benefit of performing one additional transplant is likely

[25] Data taken from Organ Procurement and Transplantation Network (OPTN) data website, available here: http://optn.transplant.hrsa.gov/data/

[26] For example, Gaynor, Seider, and Vogt (2005) use annual volume and Gowrisankaran, Ho, and Town (2006) and Huesch (2009) use quarterly volume.

decreasing in transplant center volume, I take the log transformation of 30 day volume to capture the non-linear effects.[27]

Apart from the 30 day volume measure, there are five additional groups of variables that require definition: patient outcome measures, patient demographic characteristics, patient clinical measures, characteristics of the kidney donor, and the instrument. The main outcomes of interest are graft failure and patient mortality at one week, one month, and one year post-transplant. Axelrod et al (2004) focus on one year graft failure rates, but show that much of the association that they find is manifested in the first month post-transplant. To build upon Axelrod et al (2004), I also include patient mortality as an outcome, which is the usual outcome measure used in the literature. In addition to these six measures (three time periods each for graft failure and patient mortality), I also examine two additional outcomes. The first is the probability that the patient returns to dialysis within one week of the transplant. Return to dialysis does not necessarily indicate graft failure, but is an indication of delayed graft function, which can result if the transplanted kidney is from an older donor or from a donor with history of hypertension or diabetes (Perico et al 2004). The last outcome is the recipient's glomerular filtration rate (GFR) at discharge. This is a measure of how well the patient's renal system is filtering serum creatinine (a waste by-product of muscle activity) from the bloodstream. Therefore, higher GFR levels indicate better kidney function immediately after transplant. I use a slightly different specification when GFR is the outcome of interest. In particular, I include the recipient's pre-transplant GFR as an explanatory variable. Pre-transplant GFR is missing for a number of observations (discussed below), so the GFR regressions are estimated using the subsample of observations that have non-missing values of this regressor.

[27] For observations where 30 day volume is equal to zero, I recode 30 day volume as .0001, and then take the log transformation. Results are robust to using other (small) values in place of zero.

Patient demographics include age at transplant, gender, race, education, and primary insurer. Because of the large number of observations with missing data on education status, I also include an indicator of whether education is missing or not. Patient clinical information includes indicators for the following: obesity, whether the patient was on dialysis at the time of transplant, whether the patient was hospitalized leading up to the transplant, poor functional status at time of transplant, patient primary diagnosis, previous kidney transplant recipient, patient "sensitization", and the stage of the patient's kidney disease. I calculated each recipient's body mass index (BMI) using height and weight data, and then I coded a variable indicating obesity (BMI of 30 or higher). I coded a patient as being sensitized if her panel reactive antibody (PRA) value was 20 or higher.[28] To develop the kidney disease stage variable, I converted each patient's serum creatinine level to GFR using the method described in Levey et al (2006).[29] Then, I coded a patient as having stage 5 kidney disease if her GFR value was below 15 mL/min/1.73m^2.[30] In addition to these indicator variables, I also control for the number of days that a patient waited for her transplant. As with educational status, I created indicator variables for missing BMI, PRA, GFR, and functional status information and included them in the analysis, rather than drop these observations.

The demographic and clinical variables are meant to address patient selection and to adequately adjust for a recipient's underlying health at the time of her transplant. In addition, information on the kidney donor is required to appropriately account for systematic differences

[28] PRA is a measure of the percent of the population against whom the patient is "sensitized against." Sensitized patients have immunological profiles that increase the odds of acute organ rejection. Most of the population has a PRA of zero (no level of sensitization), and the cutoff of 20 is used often in the transplant community to designate a patient as sensitized.

[29] Serum creatinine is a waste by-product of muscle activity that is typically filtered by the kidneys. Because muscle mass (and thus levels of serum creatinine) vary by age, race, and gender, the GFR calculation allows for a standardized measure of kidney function. Lower levels of GFR indicate impaired kidney function.

[30] Cutoff value taken from the guidelines provided by the National Kidney Foundation, accessed here on January 15, 2011: http://www.kidney.org/kidneydisease/ckd/knowgfr.cfm

in the quality of organs used by transplant centers. This set of variables includes the age, race, and gender of the donor, the donor status (living versus deceased), the number of HLA mismatches between the donor and the recipient, and whether the donor is an "expanded criteria donor" (ECD). As described in an earlier note, ECDs are deceased donors that meet the following criteria: age 60 or older, or age 50 or older with two of the following three conditions: history of high blood pressure, high serum creatinine (an indication of impaired kidney function), or death resulting from stroke. As their name implies, these are donors that are of marginal quality. But given the persistent shortage of transplantable kidneys in the United States, utilization of these kidneys has increased, because a transplant with a lower-quality kidney may be a superior treatment option over maintenance dialysis for some patients. HLA mismatches refer to the degree of immunological mismatches between the donor and recipient; these can range from zero (perfect match) to six (total mismatch). Given the set of variables above, the following would indicate increased risk of graft failure: increased donor age, deceased status, increased number of HLA mismatches, and ECD status. Donor race and gender may not affect outcomes per se, but may proxy for additional underlying differences between donors that may affect outcomes (like risky behavior or history of disease).

Finally, the instrument used in the estimation of equation (4) is the number of perfect match kidneys performed in the last 30 days at a recipient's transplant center. These are defined as transplants from kidneys that are 1) either perfect matches or zero mismatches and 2) shared nationally. While the difference is subtle, a perfect match kidney is one that is a perfect six-out-of-six antigen match to the recipient. A zero-mismatch kidney is one in which there is clinically no evidence of a mismatch between the recipient's antigen profile, and that of the donor. Therefore, perfect match kidneys are a subset of all zero mismatch kidneys. For simplicity, I will

refer to all of these kidneys as "perfect matches." There is no difference with respect to the allocation rules; all zero mismatch kidneys (of which perfect six-out-of-six matches are a subset) are subject to national sharing. The rationale for the second criteria (that they are shared nationally) is to reinforce the notion that the receipt of a perfect match kidney is an unforeseen and random event. In practice, this part of the definition of the instrument is not overly restrictive as most (approximately 85 percent) perfect match kidneys are shared nationally.

The data set is restricted to adult kidney transplants performed in the United States between 1996 and 2005. These transplants were performed at 243 transplant centers over the ten year sample period. Some transplant centers opened and others closed during this time period, so the estimation sample is an unbalanced panel. In all analyses, I exclude the recipients of perfect match kidneys. These recipients receive kidneys that are of higher quality (they are perfect matches, by definition) and spend less time on the kidney waitlist than recipients of non-perfect match kidneys. Therefore, while recipients of perfect match kidneys may be similar to other recipients, the characteristics of their transplants are quite different. In other words, perfect match recipients are used for the purposes of constructing the instrument, but are themselves excluded from any outcomes regressions.[31] The final estimation sample includes 117,137 adult transplants performed at all non-pediatric transplant centers in the United States (including Puerto Rico).

[31] In reality, their inclusion in the estimation sample has no effect on the results. But since perfect match recipients are likely to have better outcomes due to the lack of antigen mismatches, and reduced waiting time, it seems natural to exclude them to ensure that their experience do not drive any of the results or conclusions.

2.7 Results

2.7.1 Descriptive Analysis

To motivate the idea that volume is related to transplant recipient outcomes, I first plot average patient outcomes against average 30 day volume. In Figures 2.1 – 2.8, each point in the graph represents a transplant center in the sample. The horizontal axis is the transplant center sample average of 30 day transplant volume. At first blush, it appears that there exists a clear negative relationship between volume and adverse patient outcomes in Figures 2.1 – 2.7; transplant centers that perform more transplants have, on average, better patient outcomes that those transplant centers that perform fewer transplants. However, it also appears that the association between average transplant volume and average patient outcomes is "flat" over a large range of average volume. In addition, while it appears that some lower volume transplant centers have worse patient outcomes, on average, there are a number of lower volume transplant centers that are clustered at zero (i.e. no adverse patient events). This may reflect the fact that the mean will be a noisier statistic at lower volume centers than at higher volume centers due to smaller sample sizes. Figure 2.8 plots average post-transplant glomerular filtration rate (GFR), a clinical measure of kidney function, against average 30 day volume. Higher values of GFR indicate better kidney function. However, there does not appear to be a clear positive relationship between post-transplant GFR and volume, as would be predicted by the volume-outcome relationship.

Table 2.1 provides descriptive statistics for selected variables, split by tercile of average transplant center 30 day volume. The last column provides a statistical test of the difference in average outcomes and characteristics between patients and donors at the transplant centers in the lowest and highest quartiles of average 30 day volume. There are clear differences in patient

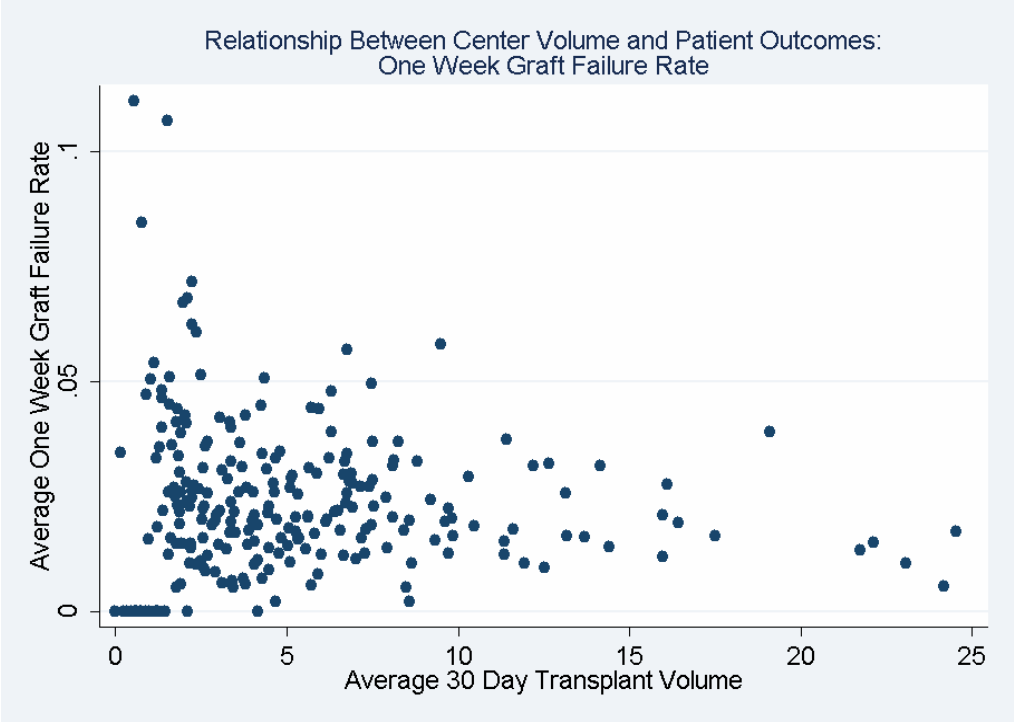

Figure 2.1

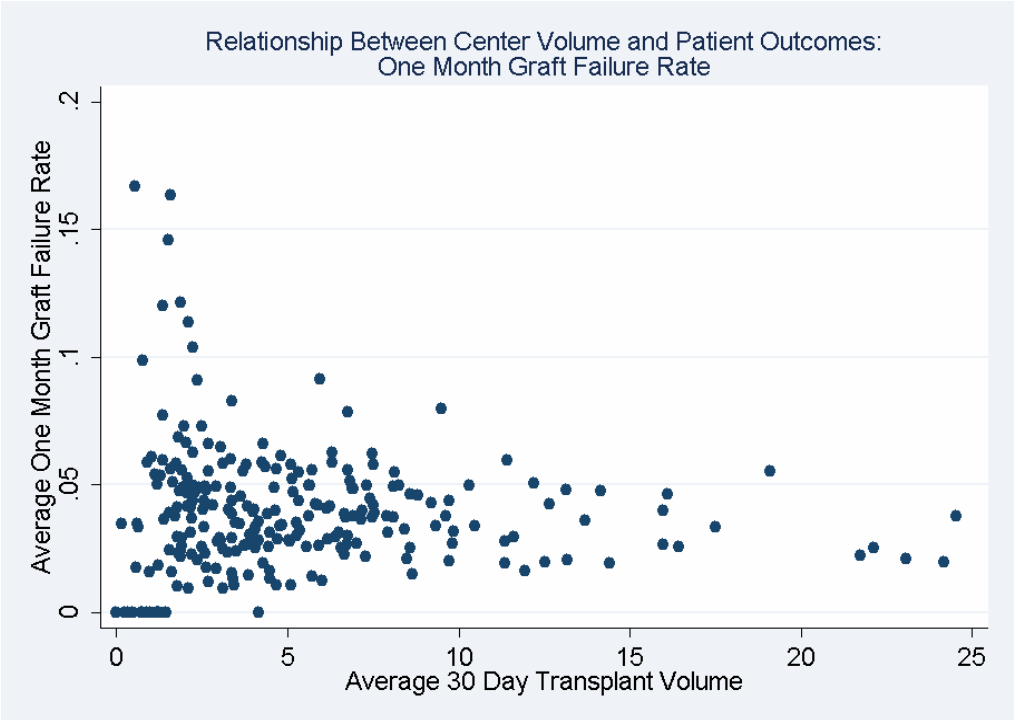

Figure 2.2

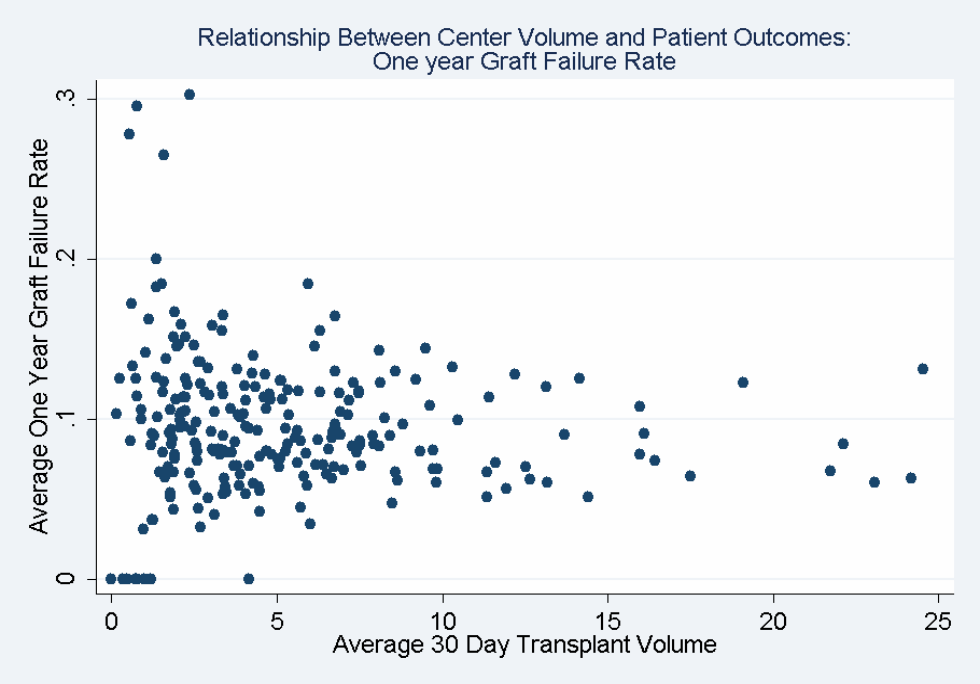

Figure 2.3

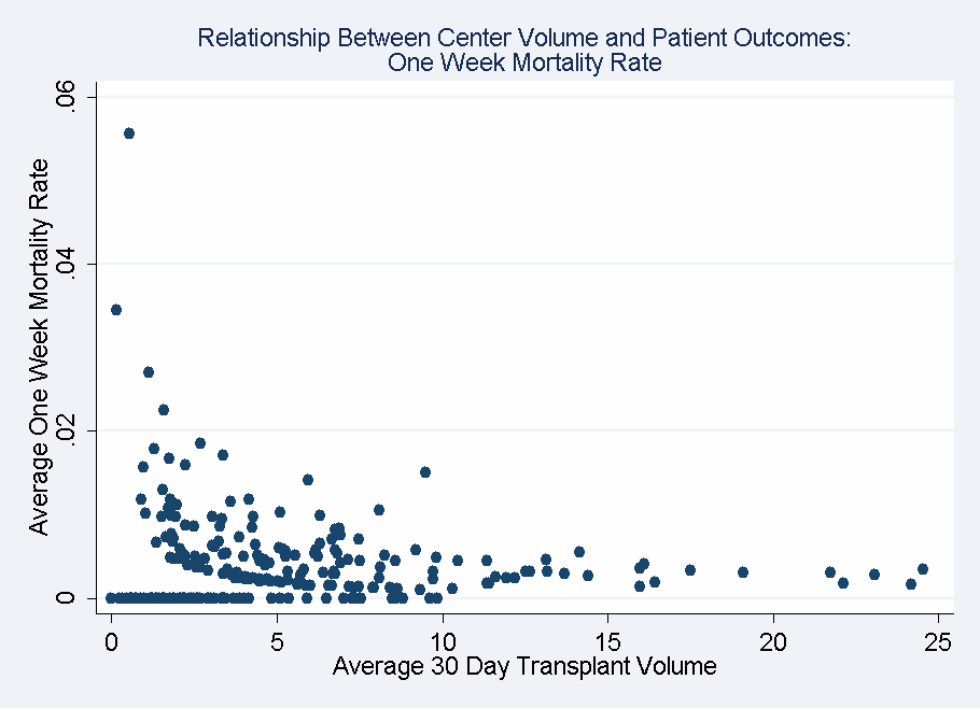

Figure 2.4

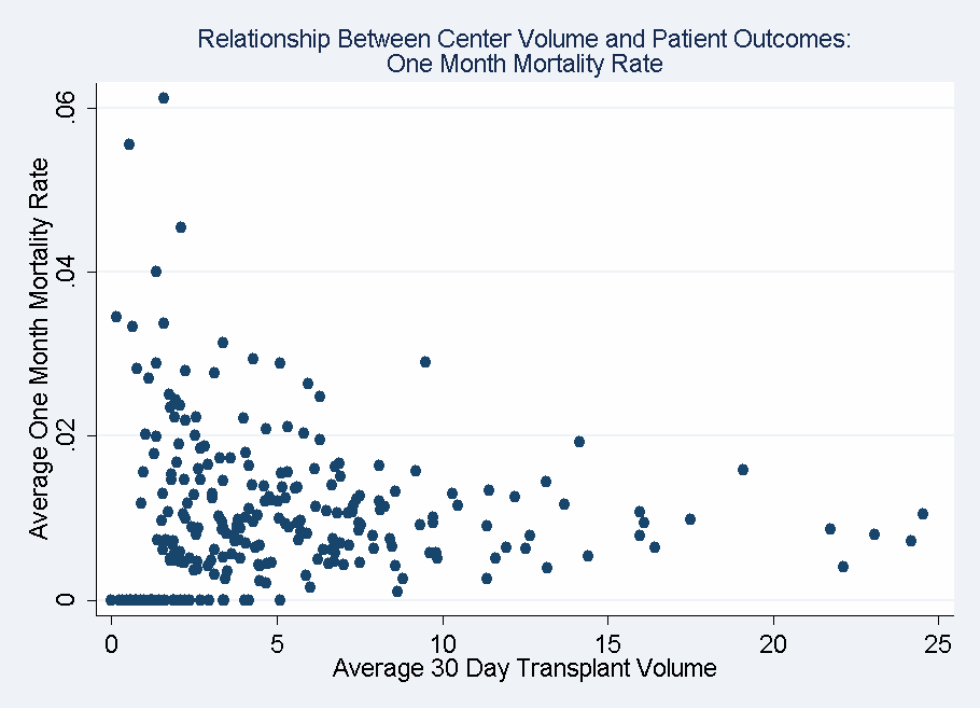

Figure 2.5

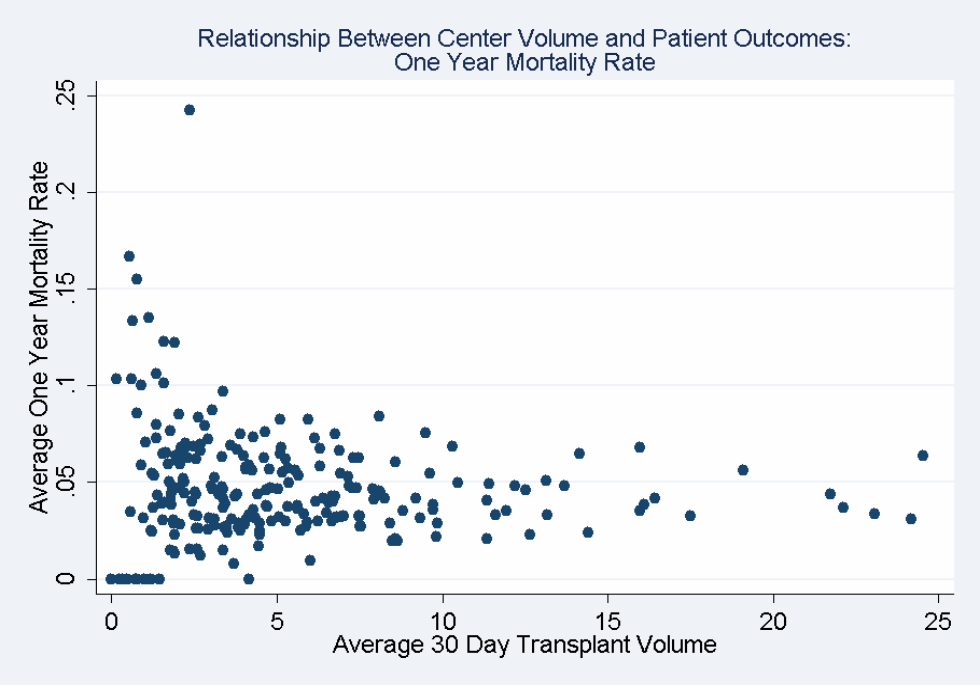

Figure 2.6

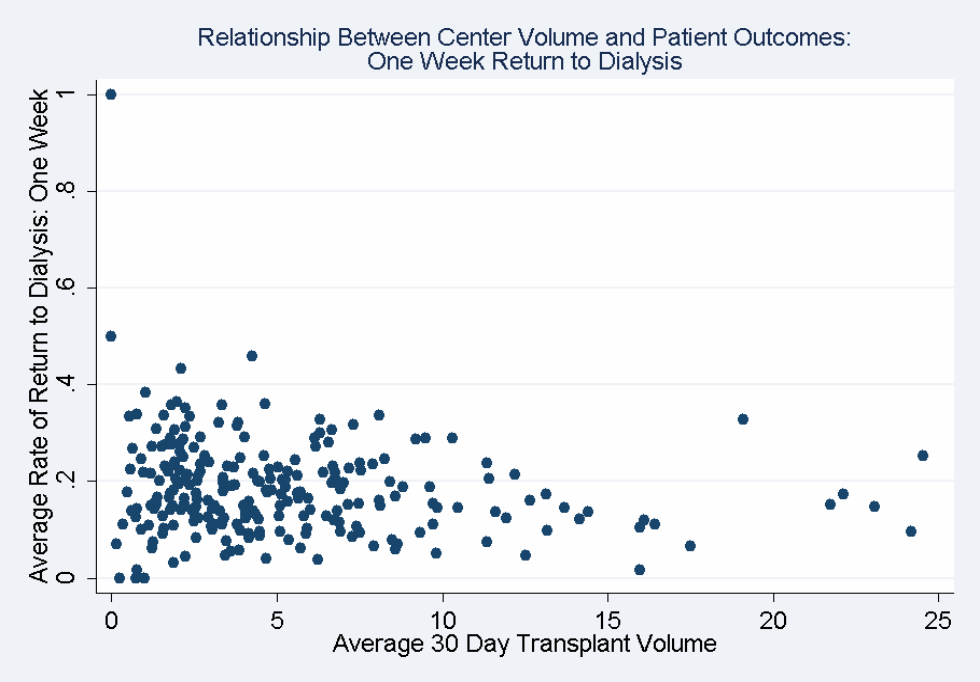

Figure 2.7

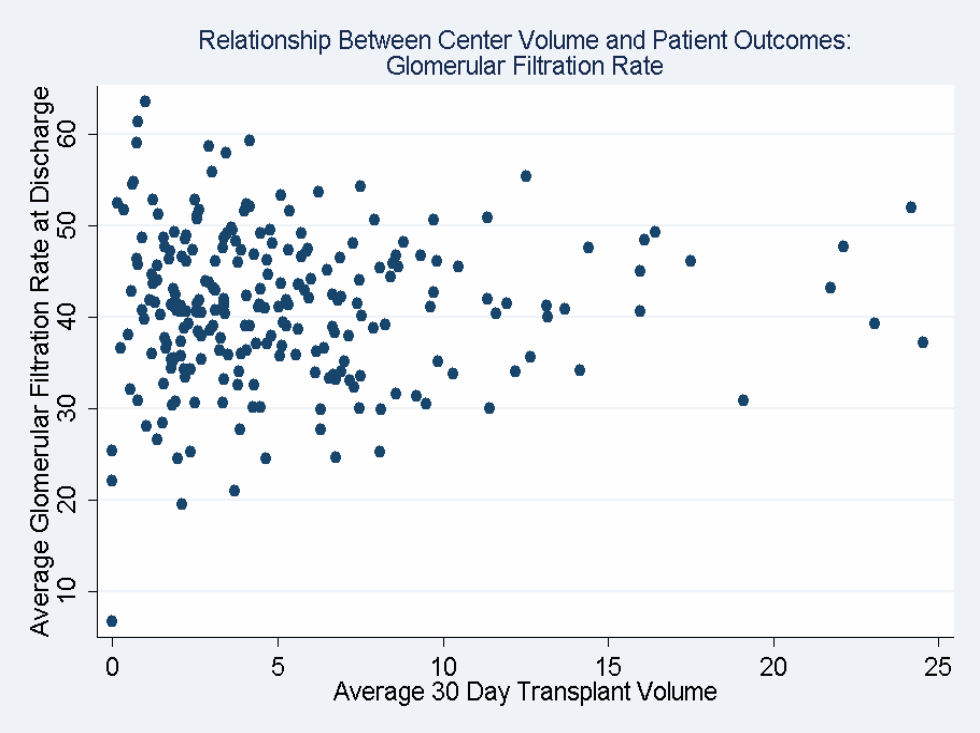

Figure 2.8

demographics, patient clinical values, and donor characteristics between the lower and higher volume transplant centers, as well as distinct differences in patient outcomes. On average, outcomes are better at higher volume transplant centers, although the magnitude of the differences in outcomes (highest versus lowest tercile) varies depending on the outcome. Table 2.1 also shows clear differences in the racial and ethnic makeup of transplant recipients across terciles of volume, as well as differences in their educational attainment and source of insurance. There is mixed evidence of positive selection of healthier patients (clinically) and donors at higher volume transplant centers. For example, transplant centers in the highest tercile of average 30 day volume have fewer obese patients, fewer patients on dialysis or in the hospital at the time of transplant, and also are more likely to use kidneys from living donors than transplant centers in the lowest quartile. On the other hand, patients at higher volume transplant centers are more like to have poor functional status, to be immunologically sensitized, and to have had a previous transplant. In addition, the donors at higher volume centers are slightly older and are more like to be expanded criteria donors. Therefore, it is difficult to conclude from the evidence in Table 2.1 what the net effect of patient or donor selection will have on the magnitude of the estimated volume effect.

2.7.2 Pooled OLS Estimates

In order to test for systematic sorting of patients and donors to higher or lower volume transplant centers, I first estimate equation (2.1) without any patient or donor characteristics, and then re-estimate (2.1) with patient and donor characteristics added as additional regressors. This

Table 2.1: Descriptive Statistics – Selected Variables

	Full Sample	First Tercile	Second Tercile	Third Tercile	p-value
Number of Observations	117,137	39,257	38,855	39,025	
Number of Transplant Centers	243	161	56	26	
Transplants in Last 30 Days	9.29	3.67	7.68	16.54	0.000
	(0.02)	(0.01)	(0.02)	(0.04)	
Outcomes					
Graft Failure: 1 Week	0.022	0.023	0.024	0.019	0.000
	(0.000)	(0.001)	(0.001)	(0.001)	
Graft Failure: 1 Month	0.037	0.039	0.040	0.033	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	
Graft Failure: 1 Year	0.092	0.096	0.094	0.086	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	
Patient Died: 1 Week	0.004	0.004	0.004	0.003	0.004
	(0.000)	(0.000)	(0.000)	(0.000)	
Patient Died: 1 Month	0.010	0.011	0.010	0.009	0.005
	(0.000)	(0.001)	(0.000)	(0.000)	
Patient Died: 1 Year	0.044	0.048	0.042	0.043	0.002
	(0.001)	(0.001)	(0.001)	(0.001)	
Patient Return to Dialysis: 1 Week	0.169	0.186	0.173	0.150	0.000
	(0.001)	(0.002)	(0.002)	(0.002)	
Glomerular Filtration Rate at Discharge	41.23	41.70	39.74	42.24	0.016
	(0.10)	(0.18)	(0.18)	(0.14)	
Patient Demographics					
Recipient Black	0.244	0.240	0.265	0.225	0.000
	(0.001)	(0.002)	(0.002)	(0.002)	
Recipient White	0.575	0.595	0.544	0.587	0.036
	(0.001)	(0.002)	(0.003)	(0.002)	
Recipient Hispanic	0.116	0.105	0.122	0.121	0.000
	(0.001)	(0.002)	(0.002)	(0.002)	
Recipient Other Race	0.065	0.060	0.069	0.066	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	
Primary Insurer: Medicare	0.520	0.558	0.510	0.491	0.000
	(0.001)	(0.003)	(0.003)	(0.003)	
Primary Insurer: Medicaid	0.046	0.045	0.054	0.039	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	
Primary Insurer: Private	0.414	0.367	0.420	0.454	0.000
	(0.001)	(0.002)	(0.003)	(0.003)	
Primary Insurer: Other	0.020	0.029	0.015	0.016	0.000
	(0.000)	(0.001)	(0.001)	(0.001)	
High School or Less	0.404	0.456	0.377	0.379	0.000
	(0.001)	(0.003)	(0.002)	(0.002)	
Education Status Missing	0.240	0.160	0.291	0.270	0.000
	(0.001)	(0.002)	(0.002)	(0.002)	
College or Higher	0.356	0.384	0.332	0.351	0.000
	(0.001)	(0.002)	(0.002)	(0.002)	

Table 2.1: Descriptive Statistics – Selected Variables (continued)

Patient Clinical Information					
Recipient Obese	0.219	0.232	0.213	0.213	0.000
	(0.001)	(0.002)	(0.002)	(0.002)	
On Dialysis	0.829	0.865	0.817	0.805	0.000
	(0.001)	(0.002)	(0.002)	(0.002)	
In Hospital	0.026	0.029	0.025	0.023	0.000
	(0.000)	(0.001)	(0.001)	(0.001)	
Poor Functional Status	0.249	0.201	0.232	0.315	0.000
	(0.001)	(0.002)	(0.002)	(0.002)	
Recipient Sensitized (PRA > 19%)	0.086	0.078	0.095	0.085	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	
Previous Kidney Transplant	0.083	0.077	0.086	0.086	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	
Donor / Transplant Characteritsics					
Donor Age	38.42	37.78	38.56	38.94	0.000
	(0.04)	(0.08)	(0.08)	(0.08)	
Donor: Living	0.419	0.393	0.432	0.432	0.000
	(0.001)	(0.002)	(0.003)	(0.003)	
Donor: Expanded Criteria	0.098	0.092	0.104	0.100	0.000
	(0.001)	(0.001)	(0.002)	(0.002)	

Notes

1. Table displays the descriptive statistics of the full sample (first column) and then by tercile of transplant center average 30-day volume (columns 2, 3 and 4). The last column displays the p-value from a t-test of the difference in means between the first and third tercile (column 2 versus column 4)
2. Terciles are calculated at transplant center level. For each transplant center in the estimation sample, I calculated average 30-day volume, and then took terciles based on that average. Therefore, each tercile has approximately the same number of transplants, but varying numbers of transplant centers. The transplant centers in the first tercile are, on average, lower volume transplant centers, for example.
3. The mean of each variable is displayed, and the standard error is reported in parentheses

provides an indirect test of systematic patient and donor selection, as well as a sense of how much of the volume-outcome association may be directly related to sorting, on net. Table 2.2 presents the pooled OLS estimates of the volume effect for each of the eight outcomes. The specifications in the odd numbered columns include only the log of 30 day volume and year dummies as explanatory variables. The specifications in the even numbered columns include controls for patient demographics, patient clinical information, and donor characteristics. With the exception of GFR at discharge (panel D of Table 2.2), 30 day volume is strongly associated with patient outcomes. In addition, excluding controls for patient demographics, patient clinical information, and donor characteristics leads to an under-estimate of the volume-outcome relationship, which suggests that on net, patients, donors, or both are on average less healthy at

higher volume transplant centers. The estimates in panel A of Table 2.2 suggest that a 10 percent increase in 30 day volume is associated with a 1.6 percentage point decrease in the probability of graft failure one week post-transplant, a 2.1 percentage point decrease in the probability of graft failure one month post-transplant, and a 2.4 percentage point decrease in the probability of graft failure one year post-transplant. Based on the descriptive statistics, a ten percent increase in 30 day volume is roughly equivalent to an increase of one transplant per 30 days.

Panel B of Table 2.2 shows that volume is negatively associated with the probability of patient mortality. For example, the estimates in panel B suggest that a 10 percent increase in 30 day transplant volume is associated with a .4 percentage point decrease in the probability of patient mortality one week post-transplant. The association becomes stronger as the post-transplant window is widened: the pooled OLS estimate of the association between volume and one-month mortality is approximately 75 percent larger in magnitude that the association measured at one week. Likewise, the volume-outcome association is over twice as large one year post-transplant as it is at one month post-transplant.

Panels C and D report the pooled OLS estimates for the remaining two outcomes. The results in panel C suggest that a 10 percent increase in 30 day volume is associated with a 5.6 percentage point decrease in the probability that the recipient returns to dialysis within one week of her transplant. Unlike the results in panels A or B, though, the volume-outcome association becomes smaller in magnitude when the covariates are added into the specification. Finally, panel D suggests that there is no statistically signification association between 30 day volume, and the recipient's glomerular filtration rate at discharge from the hospital.

Table 2.2: Pooled OLS Estimates

A. Post-Transplant Graft Failure	One Week Post TX		One Month Post TX		One Year Post TX	
	(1)	(2)	(3)	(4)	(5)	(6)
Center Volume in Previous 30 days (Log)	-0.00140***	-0.00159***	-0.00182***	-0.00212***	-0.00212***	-0.00244***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Observations	117,137	117,137	117,137	117,137	117,137	117,137
R-squared	0.000	0.008	0.001	0.013	0.001	0.036

B. Post-Transplant Patient Mortality	One Week Post TX		One Month Post TX		One Year Post TX	
	(7)	(8)	(9)	(10)	(11)	(12)
Center Volume in Previous 30 days (Log)	-0.00033***	-0.00038***	-0.00054**	-0.00068***	-0.00136***	-0.00152***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	117,137	117,137	117,137	117,137	117,137	117,137
R-squared	0.000	0.004	0.000	0.009	0.000	0.032

C. Return to Dialysis Within One Week	(13)	(14)
Center Volume in Previous 30 days (Log)	-0.00561***	-0.00427***
	(0.001)	(0.001)
Observations	117,137	117,137
R-squared	0.002	0.116

D. GFR at Discharge	(15)	(16)
Center Volume in Previous 30 days (Log)	0.02562	-0.00453
	(0.109)	(0.090)
Observations	83,872	83,872
R-squared	0.003	0.197

Notes:

1. The table displays the pooled OLS estimates of the effect of a one percent increase in 30 day transplant volume one each of the eight outcome variables. The results in the odd numbered columns are from specifications where the regressors are log of 30 day volume and year dummy variables. The results in the even numbered columns are from the specification where patient demographics, patient clinical information, and donor/transplant characteristics are added as covariates. The transplant center fixed effects are omitted from all regressions.
2. The sample size in panel D is smaller because a number of observations have missing information on pre-transplant GFR, which is included as a regressor in the specification where GFR at discharge is the outcome of interest.
3. Clustered standard errors are reported in parentheses (clustered at transplant center-year level)
4. *** p<0.01, ** p<0.05, * p<0.1

There is no a priori reason to expect that the pooled OLS estimates reported in Table 2.2 reflect an unbiased estimate of the volume effect in kidney transplantation. However, the results do provide strong evidence that the observed volume-outcome association is not due to favorable selection to higher volume transplant centers. When patient clinical characteristics and donor characteristics are added to the specifications, the magnitude of the volume effect increases, which suggests, if anything, that net effect of patient and donor selection to higher volume centers biases the estimated volume association towards zero.

2.7.3 Within Transplant Center-Year Variation in 30 Day Volume

Before turning to the fixed effects estimates, I present evidence that there is sufficient variation in 30 day volume within a transplant center within a particular year for the fixed effects estimation strategy to plausibly identify a volume-effect. First, I regressed the natural log of 30 day volume on the set of transplant center-year fixed dummy variables. The R-squared from this regression is .482, which implies that the fixed effects account for 48 percent of the variation in the natural log of 30 day volume. The fixed effects estimates are identified from the remaining 52 percent of 30 day volume variability.

Figures 2.9, 2.10, and 2.11 provide graphical evidence of the within transplant center-year variation in 30 day volume. Figure 2.9 displays the distribution of 30 day volume within a high volume transplant center in a given year; figures 2.10 and 2.11 do the same, for medium and low volume centers, respectively.[32]The figures show that there exists significant variation in 30 day volume within a transplant center in a given year. For example, 30 day volume ranges from

[32] These are randomly selected transplant centers/years from the estimation sample. For brevity, I provide graphical evidence for three transplant centers only, but the graphical evidence across all transplant center/years support the notion that there is significant variation within a particular transplant center in a given year in 30 day transplant volume.

10 to 34 for the high volume transplant center displayed in figure 2.9. There is somewhat less variability in 30 day volume at the low volume transplant center presented in figure 3.11. These results suggest that the volume effect in kidney transplantation can be plausibly identified using the transplant center-year fixed effects.

2.7.4 Fixed Effects Estimates

Table 2.3 reports the fixed effects estimates of the volume effect. For each of the eight outcomes, I estimated two fixed effects regressions. The first (in the odd numbered columns) exclude the covariates, and the second (in the even numbered columns) include the full set of covariates as regressors. With the exception of one week and one month graft failure, the volume-outcome association implied by the pooled OLS estimates largely goes away. The point

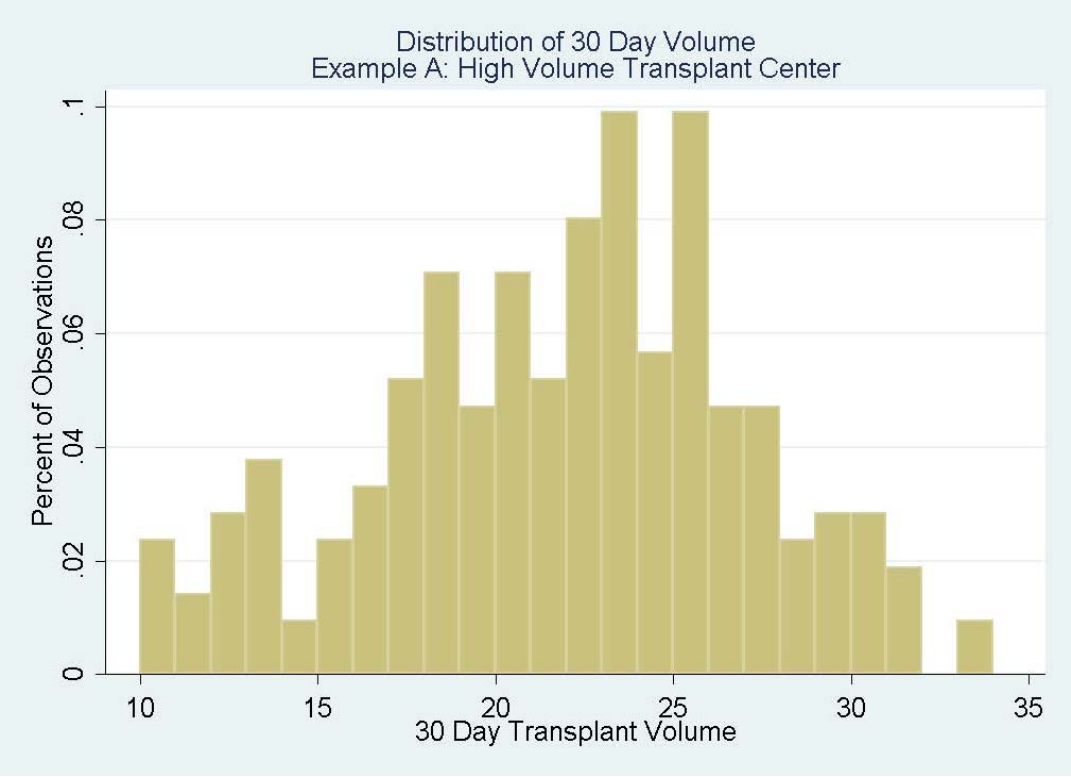

Figure 2.9

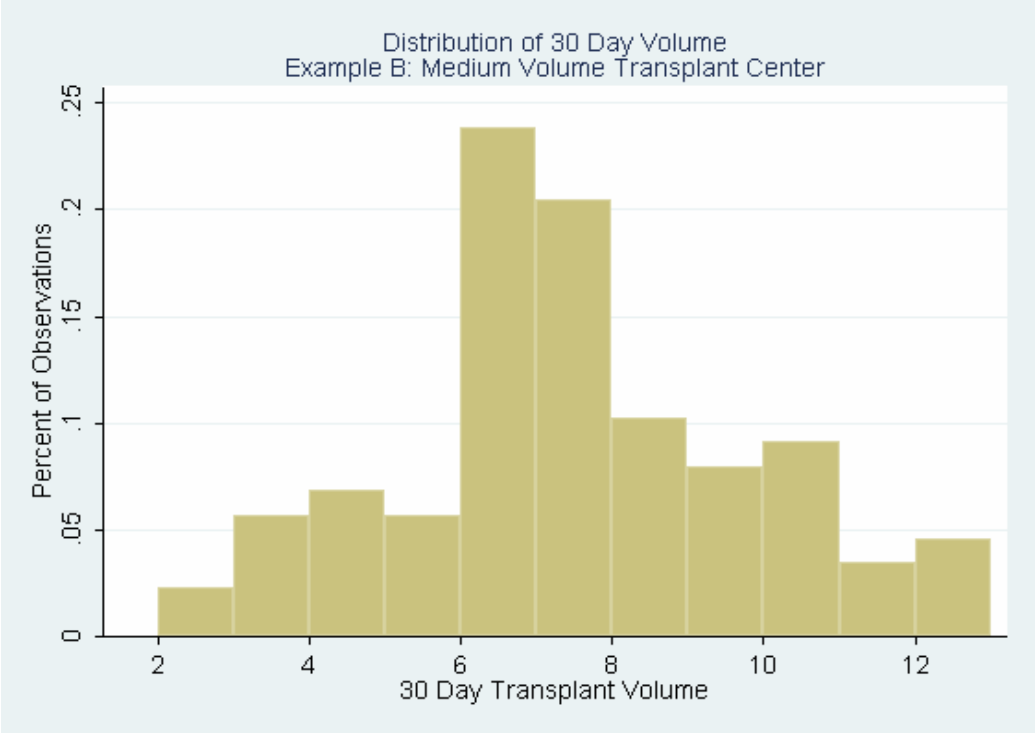

Figure 2.10

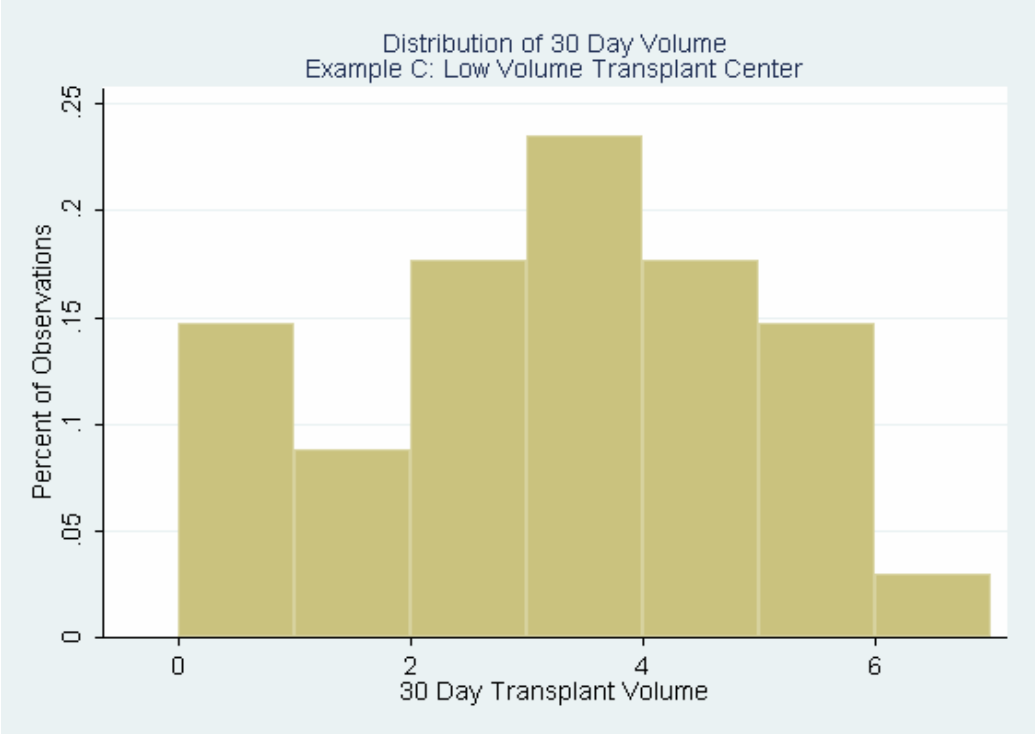

Figure 2.11

estimates of the volume effect are smaller in magnitude and imprecisely estimated. However, a volume effect appears to persist for graft failure in the short run. The fixed effects estimates imply that a 10 percent increase in 30 day volume (approximately one additional transplant per 30 days) is associated with a 1 percentage point decrease in the probability of one week graft failure, and a 1.4 percentage point decrease in one month graft failure. Based on the descriptive statistics, the estimates suggest that a 10 percent increase in 30 day volume decreases the rate of one week graft failure by 45 percent (1 percentage point divided by sample mean of 2.2 percent) and the rate of one month graft failure by 38 percent (1.4 percentage points divided by sample mean of 3.7 percent). These represent large decreases in the probability of graft failure in the short run, but the effect does not appear to persist into the longer run (one year post transplant).

The fixed effect estimates suggest that all of the relationship between patient mortality and volume is driven by between hospital differences in unobserved factors that are correlated with patient outcomes (panel B of Table 2.3). This finding is consistent with the conclusions of other fixed effects estimates of the volume effect, particularly Hamilton and Hamilton (1997). In addition, there is no evidence of a volume effect for return to dialysis and glomerular filtratation rate (panels C and D of Table 2.3). In fact, the sign of the volume effect for return to dialysis is positive, which suggests that an increase in 30 day volume increases the chances that a transplant recipient returns to dialysis within one week of her transplant, although the estimate is imprecisely estimated.

The results in Table 2.3 also provide indirect evidence of the internal validity of the fixed effects estimates. With the exception of panels C and D, the exclusion of the set of covariates from the specification leaves the estimated volume effect largely unchanged. This suggests that conditional on the transplant center-by-year fixed effects, 30 day volume is uncorrelated with the

Table 2.3: Fixed Effect Estimates

A. Post-Transplant Graft Failure	One Week Post TX		One Month Post TX		One Year Post TX	
	(1)	(2)	(3)	(4)	(5)	(6)
Center Volume in Previous 30 days (Log)	-0.00102**	-0.00103**	-0.00136**	-0.00137**	-0.00119	-0.00120
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	117,137	117,137	117,137	117,137	117,137	117,137
R-squared	0.027	0.034	0.026	0.038	0.030	0.062

B. Post-Transplant Patient Mortality	One Week Post TX		One Month Post TX		One Year Post TX	
	(7)	(8)	(9)	(10)	(11)	(12)
Center Volume in Previous 30 days (Log)	-0.00016	-0.00016	-0.00044	-0.00042	-0.00077	-0.00072
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Observations	117,137	117,137	117,137	117,137	117,137	117,137
R-squared	0.021	0.026	0.022	0.032	0.025	0.056

C. Return to Dialysis Within One Week	(13)	(14)
Center Volume in Previous 30 days (Log)	0.00062	0.00032
	(0.001)	(0.001)
Observations	117,137	117,137
R-squared	0.077	0.177

D. GFR at Discharge	(15)	(16)
Center Volume in Previous 30 days (Log)	-0.03159	-0.04068
	(0.084)	(0.074)
Observations	83,872	83,872
R-squared	0.110	0.276

Notes:

1. The table displays the fixed-effects estimates of the effect of a one percent increase in 30 day transplant volume on each of the eight outcome variables. The results in the odd numbered columns are from specifications where the regressors are log of 30 day volume and transplant center specific year dummy variables. The results in the even numbered columns are from the columns are from the specification where patient demographics, patient clinical information, and donor/transplant characteristics are added as covariates.
2. The sample size in panel D is smaller because a number of observations have missing information on pre-transplant GFR, which is included as a regressor in the specification where GFR at discharge is the outcome of interest.
3. Clustered standard errors are reported in parentheses (clustered at transplant center-year level)
4. *** p<0.01, ** p<0.05, * p<0.1

set of covariates, which confirms the claim that variation in 30 day volume is driven largely by exogenous supply shocks of transplantable kidneys.

To fully address this claim, I first estimate an additional fixed effects regression in which 30 day volume is regressed on the full set of covariates. The F-statistic from this regression is .83 (p-value = .755), so I cannot reject the null hypothesis that the set of covariates (jointly) is uncorrelated with 30 day volume. In addition, I also run 43 separate fixed effects regressions in which I regress each covariate on 30 day volume. The results of these regressions are reported in Table 2.4. Across almost all of the regressions, within-transplant center-by-year variation in 30 day volume is uncorrelated with deviations from the transplant center-by-year mean of each of the covariates. These results provide additional direct evidence that the fixed effects estimates reported in Table 2.3 are unbiased.

Closer inspection of the results in Table 2.4 confirms one concern discussed earlier: conditional on the fixed effects, 30 day volume is negatively correlated with the probability that the recipient received her kidney from a living donor. Even though donor status (deceased versus living) can be directly controlled for, there may exist systematic differences in the quality of the kidneys from deceased donors versus those from living donors, as well as differences in the underlying health of the recipients of these kidneys. In particular, if kidneys from living donors are of higher quality than those from deceased donors (which would be correlated with improved patient outcomes), then the fixed effects estimate may be biased toward zero.

2.7.5 Instrumental Variables Estimates

In order to address the concern that the fixed effects estimate may be biased, I also

Table 2.4: Correlation between 30 Day Volume and Covariates

Patient Demographics		Patient Clinical Information		Donor / Transplant Characteritsics	
Recipient Age	-0.04042	Recipient Obese	0.00085	Donor Age	-0.07674*
	(0.036)		(0.001)		(0.043)
Recipient Female	-0.00136	BMI Missing	0.00000	Donor: Female	0.00005
	(0.001)		(0.000)		(0.001)
Recipient Black	0.00069	On Dialysis	0.00047	Donor: Black	0.00139*
	(0.001)		(0.001)		(0.001)
Recipient White	-0.00186	In Hospital	-0.00027	Donor: White	-0.00154
	(0.001)		(0.000)		(0.001)
Recipient Hispanic	0.00032	Poor Functional Status	-0.00037	Donor: Hispanic	0.00019
	(0.001)		(0.001)		(0.001)
Recipient Other Race	0.00085	Functional Status Missing	-0.00026	Donor: Other Race	-0.00004
	(0.001)		(0.000)		(0.001)
		Glomerular Filtration Rate (GFR) at			
Primary Insurer: Medicare	0.00024	Transplant	-0.01402	Donor: Living	-0.00215*
	(0.001)		(0.034)		(0.001)
Primary Insurer: Medicaid	0.00024	Stage 5 Kidney Disease	0.00058	Donor: Living, No Waitlist	-0.00069
	(0.001)		(0.001)		(0.001)
Primary Insurer: Private	-0.00031	GFR / Kidney Disease Stage Missing	-0.00046	HLA Mismatches	0.00265
	(0.001)		(0.001)		(0.004)
Primary Insurer: Other	-0.00017	Recipient Sensitized (PRA > 19%)	-0.00012	Days Waiting	0.14143
	(0.000)		(0.001)		(1.426)
High School or Less	0.00038	PRA Value Missing	-0.00075	Donor: Expanded Criteria	0.00073
	(0.001)		(0.000)		(0.001)
Education Status Missing	0.00060	Primary Diagnosis: Diabetes	0.00042	Donor: Local to Recipient	-0.00007
	(0.001)		(0.001)		(0.001)
		Primary Diagnosis: Hypertensive			
College or Higher	-0.00098	Nephroscelerosis	-0.00064		
	(0.001)		(0.001)		
		Primary Diagnosis: Glomerular Disease	0.00031		
			(0.001)		
		Primary Diagnosis: Polycystic Kidneys	-0.00004		
			(0.001)		
		Primary Diagnosis: Other	-0.00006		
			(0.001)		
		Primary Diagnosis: Missing	0.00044		
			(0.000)		
		Previous Kidney Transplant	-0.00031		
			(0.001)		

Notes:

1. Table displays the results from 43 separate regressions. Each covariate was regressed on (log) 30 day transplant volume, and the set of transplant center-year dummies. The table displays the 43 estimates of the correlation between the covariates and 30 day transplant volume.
2. Clustered standard errors are reported in parentheses (clustered at transplant center-year level)
3. *** p<0.01, ** p<0.05, * p<0.1

implement an instrumental variables strategy. As discussed earlier, the IV is the number of perfectly matched kidneys that were transplanted a recipient's transplant center in the 30 days prior to her transplant. The assumptions are that the number of perfectly matched kidneys will be a strong predictor of total 30 day volume, and only affects patient outcomes through its effect on 30 day volume. As a partial check of the instrument's validity, I regress the instrument on each of the covariates (separately) and the set of transplant center-by-year dummy variables. Showing that the instrument is uncorrelated with observable patient demographics, clinical information, and donor characteristics provides evidence that the instrument is also uncorrelated with unobserved factors that are correlated with patient outcomes, like underlying patient or donor health.

Table 2.5 displays the results of these regressions. Conditional on the fixed effects, the instrument is uncorrelated with 40 of the 43 covariates. The instrument is correlated with patient and donor gender, as well as the probability that the patient's primary insurer is Medicaid, but these estimates are weakly statistically significant. More importantly, the instrument is uncorrelated with any of the measures of the patient's clinical health, and is also uncorrelated with donor status (living versus deceased). These results suggest that the instrument meets the usual standards for validity, and confirms the idea that the receipt of perfectly matched kidneys is a random event.

Table 2.6 reports the IV estimates. The instrument is strongly correlated with 30 day volume in the first stage, and the first stage F-stat is well above the usual threshold of 10 for instrument strength. For reference, I also report the fixed effects estimates from Table 2.3. In order to improve the precision of the estimates, I include all covariates in these regressions. The IV estimates are not statistically different from zero. In addition, the IV estimates for one week

Table 2.5: Correlation between Instrument and Covariates

Patient Demographics		Patient Clinical Information		Donor / Transplant Characteritsics	
Recipient Age	-0.01440	Recipient Obese	-0.00085	Donor Age	-0.01881
	(0.058)		(0.002)		(0.068)
Recipient Female	-0.00351*	BMI Missing	0.00025	Donor: Female	-0.00458**
	(0.002)		(0.001)		(0.002)
Recipient Black	-0.00149	On Dialysis	-0.00012	Donor: Black	-0.00057
	(0.002)		(0.002)		(0.001)
Recipient White	0.00079	In Hospital	-0.00042	Donor: White	-0.00118
	(0.002)		(0.001)		(0.002)
Recipient Hispanic	-0.00030	Poor Functional Status	-0.00051	Donor: Hispanic	0.00121
	(0.001)		(0.002)		(0.001)
Recipient Other Race	0.00100	Functional Status Missing	-0.00044	Donor: Other Race	0.00054
	(0.001)		(0.001)		(0.001)
		Glomerular Filtration Rate (GFR) at			
Primary Insurer: Medicare	0.00004	Transplant	-0.02553	Donor: Living	0.00233
	(0.002)		(0.096)		(0.002)
Primary Insurer: Medicaid	-0.00157*	Stage 5 Kidney Disease	0.00074	Donor: Living, No Waitlist	0.00058
	(0.001)		(0.002)		(0.001)
Primary Insurer: Private	0.00201	GFR / Kidney Disease Stage Missing	-0.00075	HLA Mismatches	-0.00351
	(0.002)		(0.002)		(0.006)
Primary Insurer: Other	-0.00048	Recipient Sensitized (PRA > 19%)	-0.00036	Days Waiting	-1.78321
	(0.000)		(0.001)		(2.482)
High School or Less	0.00003	PRA Value Missing	0.00069	Donor: Expanded Criteria	0.00011
	(0.002)		(0.001)		(0.001)
Education Status Missing	-0.00167	Primary Diagnosis: Diabetes	-0.00267	Donor: Local to Recipient	0.00144
	(0.001)		(0.002)		(0.001)
		Primary Diagnosis: Hypertensive			
College or Higher	0.00164	Nephroscelerosis	0.00208		
	(0.002)		(0.001)		
		Primary Diagnosis: Glomerular Disease	0.00111		
			(0.001)		
		Primary Diagnosis: Polycystic Kidneys	0.00048		
			(0.001)		
		Primary Diagnosis: Other	-0.00100		
			(0.002)		
		Primary Diagnosis: Missing	-0.00060		
			(0.001)		
		Previous Kidney Transplant	-0.00063		
			(0.001)		

Notes:

1. Table displays the results from 43 separate regressions. Each covariate was regressed on 30 day volume of perfect match kidneys (the IV), and the set of transplant center-year dummies. The table displays the 43 estimates of the correlation between the covariates and the IV.
2. Clustered standard errors are reported in parentheses (clustered at transplant center-year level)
3. *** p<0.01, ** p<0.05, * p<0.1

Table 2.6: Instrumental Variables Estimates

A. Post-Transplant Graft Failure	One Week Post TX		One Month Post TX		One Year Post TX	
	FE	IV-FE	FE	IV-FE	FE	IV-FE
Center Volume in Previous 30 days (Log)	-0.00103**	-0.00046	-0.00137**	-0.00108	-0.00120	-0.00061
	(0.000)	(0.004)	(0.001)	(0.005)	(0.001)	(0.007)
First Stage F-Stat		426.8		426.8		426.8
Observations	117,137	117,137	117,137	117,137	117,137	117,137
R-squared	0.034	0.034	0.038	0.038	0.062	0.062

B. Post-Transplant Patient Mortality	One Week Post TX		One Month Post TX		One Year Post TX	
	FE	IV-FE	FE	IV-FE	FE	IV-FE
Center Volume in Previous 30 days (Log)	-0.00016	0.00023	-0.00042	-0.00192	-0.00072	-0.00230
	(0.000)	(0.001)	(0.000)	(0.002)	(0.001)	(0.005)
First Stage F-Stat		426.8		426.8		426.8
Observations	117,137	117,137	117,137	117,137	117,137	117,137
R-squared	0.026	0.026	0.032	0.032	0.056	0.056

C. Return to Dialysis Within One Week	FE	IV-FE
Center Volume in Previous 30 days (Log)	0.00032	-0.01593*
	(0.001)	(0.010)
First Stage F-Stat		426.8
Observations	117,137	117,137
R-squared	0.177	0.174

D. GFR at Discharge	FE	IV-FE
Center Volume in Previous 30 days (Log)	-0.04068	-0.11882
	(0.074)	(0.755)
First Stage F-Stat		292.4
Observations	83,872	83,872
R-squared	0.276	0.276

Notes:

1. The table displays both the fixed effects estimates (re-reported from Table 2.3) and the IV-FE estimates of the effect of a one percent increase in 30 day volume on each of the eight outcome variables. The full set of covariates are included in each of the regressions.
2. In the first stage, log of 30 day volume is regressed on the IV (number of perfect match kidneys), the full set of covariates, and the set of transplant center-by-year dummy variables. The first stage F-statistic is reported above.
3. The sample size in panel D is smaller because a number of observations have missing information on pre-transplant GFR, which is included as a regressor in the specification where GFR at discharge is the outcome of interest.
4. Clustered standard errors are reported in parentheses (clustered at transplant center-year level)
5. *** p<0.01, ** p<0.05, * p<0.1

and one month graft failure are smaller in magnitude than the fixed effects estimates, which contradicts the expectation that the fixed effects estimates understate the volume effect. However, this may be due a local treatment effect; the compliers of the IV may have, on average, different characteristics than the sample as a whole. This may be the case if some transplant centers are more likely than others to receive perfectly matched kidneys, and the patients at these centers have different characteristics than the full population of transplant recipients. Whatever the case, the IV estimates confirm the notion that the observed volume-outcome relationship in kidney transplantation is largely due to between hospital differences, rather than to a volume effect, as would be predicted by the "practice makes perfect" hypothesis.

2.7.6 Effects by Transplant Center Size and Age

The fixed effects estimates suggest that the "practice makes perfect" hypothesis does not fully explain the volume-outcome relationship in kidney transplantation. However, a volume effect may persist for transplant centers with certain characteristics, and these effects may be "swamped" using the full sample of observations. In particular, one might expect a stronger volume effect among lower volume transplant centers. These centers have less accumulated experience than higher volume transplant centers, so an increase in 30 day volume may have a beneficial effect among these centers. That is, an increase in 30 day volume may have zero effect at a larger transplant center with much more experience transplanting kidneys. Therefore, I re-estimate the fixed effects estimates based on tercile of transplant center average of 30 day volume (these are the same terciles used in Table 2.1).

Table 2.7 reports the fixed effects estimates, broken out by tercile of center average 30 day volume. In order to improve the precision of the estimates, I include all covariates in each of

the regressions. For graft failure, it appears that the volume effect is stronger at lower volume transplants centers (in the first tercile) than at higher volume transplant centers (in the second and third terciles). The volume effect is statistically significant for lower volume transplant centers at one week and one month post transplant, but is statistically insignificant one year post transplant. These results suggest that most of the volume effect for one week and one month graft failure in the full sample is concentrated at lower volume transplant centers. This is an interesting result. Under the selective referral hypothesis, the reason that lower volume transplant centers may perform fewer transplants is that they may be lower quality providers, and therefore patients are discouraged from receiving their care at these centers. If lower volume transplant centers are in fact lower quality providers, then it appears that they are able to offset this disadvantage (with respect to patient outcomes) by performing additional transplants.

Unexpectedly, the volume effect appears to be stronger at higher volume transplant centers with respect to patient mortality, although the estimates are imprecisely estimated. The results suggest that a 10 percent increase in 30 day volume decreases the probability of one week patient mortality by 1.6 percentage points for patients that are transplanted at higher volume transplant centers. It is not immediately clear why such a volume effect would be present at higher volume centers and not at lower volume centers, or why there is no apparent volume effect at higher volume centers with respect to post transplant graft failure. One potential explanation is that patients who are transplanted at higher volume centers are, on average, sicker than the full sample of transplant recipients. The pooled OLS results indirectly confirm this notion. In that case, it may by that the effect of an increase in transplant volume manifests itself differently at different transplant centers because of patient and donor heterogeneity across

Table 2.7: Fixed Effects Estimates by Tercile of Volume

A. Post-Transplant Graft Failure	One Week Post TX			One Month Post TX			One Year Post TX		
	First Tercile	Second Tercile	Third Tercile	First Tercile	Second Tercile	Third Tercile	First Tercile	Second Tercile	Third Tercile
Center Volume in Previous 30 days (Log)	-0.00110**	-0.00106	0.00120	-0.00147**	-0.00160	0.00294	-0.00136	-0.00092	0.00281
	(0.000)	(0.001)	(0.002)	(0.001)	(0.002)	(0.003)	(0.001)	(0.002)	(0.004)
Observations	39,257	38,855	39,025	39,257	38,855	39,025	39,257	38,855	39,025
R-squared	0.051	0.031	0.019	0.054	0.037	0.024	0.076	0.060	0.051

B. Post-Transplant Patient Mortality	One Week Post TX			One Month Post TX			One Year Post TX		
	First Tercile	Second Tercile	Third Tercile	First Tercile	Second Tercile	Third Tercile	First Tercile	Second Tercile	Third Tercile
Center Volume in Previous 30 days (Log)	-0.00012	-0.00016	-0.00156**	-0.00040	-0.00014	-0.00197	-0.00085	0.00097	-0.00304
	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)
Observations	39,257	38,855	39,025	39,257	38,855	39,025	39,257	38,855	39,025
R-squared	0.042	0.022	0.010	0.048	0.032	0.017	0.074	0.053	0.042

C. Return to Dialysis Within One Week	First Tercile	Second Tercile	Third Tercile
Center Volume in Previous 30 days (Log)	0.00030	0.00022	0.00119
	(0.001)	(0.003)	(0.006)
Observations	39,257	38,855	39,025
R-squared	0.189	0.174	0.165

D. GFR at Discharge	First Tercile	Second Tercile	Third Tercile
Center Volume in Previous 30 days (Log)	-0.01887	-0.05251	-0.85904*
	(0.079)	(0.253)	(0.506)
Observations	27,997	28,390	27,485
R-squared	0.278	0.272	0.280

Notes:

1. The table displays the fixed-effects estimates of the effect of a one percent increase in 30 day transplant volume on each of the eight outcome variables. Each regression includes the full set of covariates as regressors.
2. Terciles are calculated at transplant center level. For each transplant center in the estimation sample, I calculated average 30-day volume, and then took terciles based on that average. Therefore, each tercile has approximately the same number of transplants, but varying numbers of transplant centers. The transplant centers in the first tercile are, on average, lower volume transplant centers, for example. These terciles correspond to the terciles used in the reporting of descriptive statistics in Table 2.1.
3. The sample size in panel D is smaller because a number of observations have missing information on pre-transplant GFR, which is included as a regressor in the specification where GFR at discharge is the outcome of interest
4. Clustered standard errors are reported in parentheses (clustered at transplant center-year level)
5. *** p<0.01, ** p<0.05, * p<0.1

transplant centers.

In addition to the size of the transplant center, the age of the transplant center may also determine whether an increase in transplant volume would cause the center to improve patient outcomes. Table 2.8 breaks down the transplants in the estimation sample by the year in which each transplant center in the sample opened. A very large percentage (84.1 percent) of transplants was performed at transplant centers that have been in operation since at least 1987.[33] Because these transplant centers have a large stock of accumulated experience to draw upon, the effect of performing one more transplant may be zero. That is, these centers may already be on the flat of their learning curves, so variation in 30 day volume has no effect on patient outcomes. "Younger" transplant centers, however, may truly be learning by doing, and therefore the volume effect may be stronger at relatively newer transplant programs.

In order to test this hypothesis, I re-estimate the fixed effects estimations on two subgroups. The "younger" subgroup includes transplants performed at centers that were opened after 1987 (81 transplant centers), and the "established" group includes transplants performed at transplant centers that have been open since at least 1987 (161 transplant centers). Table 2.9 presents the results. Contrary to expectations, it appears that with respect to one week and one month graft failure, the volume effect is larger (and statistically significant) at more established transplant programs. There is no apparent volume effect with respect to patient mortality among either of the two subgroups of transplant centers. While these results are difficult to reconcile against the notion that younger transplant programs would benefit more from an increase in volume, there may be unobserved characteristics between these two subgroups of transplant centers that explain why the volume effect (with respect to graft failure) is stronger at more

[33] As the note in the Table describes, the UNOS STAR data begins in 1987. Therefore, the earliest that I observe a transplant center performing a transplant is 1987.

Table 2.8: Estimation Sample by Year of Transplant Center Opening

Year of Opening	Number of Centers	Number of Transplants	Pct. Of Transplants	Cumm. Percentage
1987 or before	162	98,488	84.1%	84.1%
1988	14	4,019	3.4%	87.5%
1989	12	3,682	3.1%	90.7%
1990	10	3,152	2.7%	93.3%
1991	5	694	0.6%	93.9%
1992	2	891	0.8%	94.7%
1993	3	424	0.4%	95.1%
1994	4	335	0.3%	95.3%
1995	1	25	0.0%	95.4%
1996	4	377	0.3%	95.7%
1997	6	1,193	1.0%	96.7%
1998	0	0	0.0%	96.7%
1999	10	2,170	1.9%	98.6%
2000	3	755	0.6%	99.2%
2001	0	0	0.0%	99.2%
2002	2	146	0.1%	99.3%
2003	0	0	0.0%	99.3%
2004	3	766	0.7%	100.0%
2005	2	20	0.0%	100.0%
	243	117,137		

Notes:

1. The table displays the year in which each transplant center in the estimation sample opened. For example, of the 117,137 transplants performed between 1996 and 2005, 98,488 were performed at the 162 transplant centers that opened in 1987 or before. The UNOS STAR data begins in 1987, so the exact year of opening is unknown for centers whose first tranplant is observed in the data duirng 1987.
2. The estimation sample begins in 1996. The table shows that 10 transplant centers opened between 1996 and 2005, performing 5,427 transplants or 4.6% of the number of transplants in the estimation sample

established transplant programs. If these programs also experience higher rates of staff turnover (as compared to younger programs), then at any given point in time, the staff may not be as experienced as might be implied by the date the program opened. In this scenario, there may exist positive benefits of performing more transplants on patient outcomes even at more established programs. More work is needed to document the exact differences in these programs that may explain the results in Table 2.9.

2.7.7 Depreciation of Experience in Kidney Transplantation

As discussed earlier, one of the motivations for using 30 day volume as a measure of recent experience, as opposed to other "lookback" windows, is that there appears to be rapid depreciation of experience in kidney transplantation. Here I present the fixed effects estimates of the volume effect from specifications where I include lagged terms of 30 day volume. If volume in the 31 to 60 days (or 61 to 90 days) prior to a patient's transplant affects current outcomes, then the coefficients on the lagged terms should be negative and statistically significant. That is, holding constant 30 day volume, does the experience in the 31 to 60 (and 61 to 90) days prior to a patient's transplant have any marginal benefit on transplant center performance?[34] Table 2.10 presents the results of these regressions. For brevity, I only include the results for graft failure rates, since that is where any evidence of a volume effect is isolated. For reference, I report the fixed effects estimates from table 2.3 in the first column of table 2.10. The coefficients on the lagged terms in the second and third columns are both economically and statistically insignificant. This suggests that only very recent volume, measured by 30 day volume, affects short term graft failure rates, and that this experience does not carry over into the future. In additional regressions (not reported here), I also re-estimate the fixed effects equations using a 60 day measure of volume, and again using a 90 day measure of volume. In these specifications, the volume effect is not statistically different from zero. Taken together, these results suggest that experience depreciates rather quickly in kidney transplantation, and that the 30 day measure of recent volume is more appropriate than longer "lookback" windows.

[34] This approach is similar to the one used in Gaynor, Seider, and Vogt (2005), although the motivation for sing it is not exactly the same.

Table 2.9: Fixed Effects Estimates by Age of Transplant Center

A. Post-Transplant Graft Failure	One Week Post TX		One Month Post TX		One Year Post TX	
	"Younger"	"Established"	"Younger"	"Established"	"Younger"	"Established"
Center Volume in Previous 30 days (Log)	-0.00049	-0.00135**	-0.00025	-0.00204**	-0.00042	-0.00169
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	18,649	98,488	18,649	98,488	18,649	98,488
R-squared	0.047	0.031	0.055	0.035	0.074	0.060

B. Post-Transplant Patient Mortality	One Week Post TX		One Month Post TX		One Year Post TX	
	"Younger"	"Established"	"Younger"	"Established"	"Younger"	"Established"
Center Volume in Previous 30 days (Log)	0.00022	-0.00039	-0.00030	-0.00048	-0.00049	-0.00087
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Observations	18,649	98,488	18,649	98,488	18,649	98,488
R-squared	0.042	0.023	0.053	0.027	0.071	0.054

C. Return to Dialysis Within One Week	"Younger"	"Established"
Center Volume in Previous 30 days (Log)	-0.00113	0.00123
	(0.001)	(0.001)
Observations	18,649	98,488
R-squared	0.190	0.174

D. GFR at Discharge	"Younger"	"Established"
Center Volume in Previous 30 days (Log)	-0.01473	-0.06007
	(0.121)	(0.095)
Observations	13,651	70,221
R-squared	0.327	0.267

Notes:

1. The table displays the fixed-effects estimates of the effect of a one percent increase in 30 day transplant volume on each of the eight outcome variables. Each regression includes controls for patient demographics, patient clinical information, and donor/transplant characteristics.
2. "Younger" transplant centers are those that opened after 1987. "Established" transplant centers are those that opened in 1987 at the latest.
3. Clustered standard errors are reported in parentheses (clustered at transplant center-year level)
4. *** p<0.01, ** p<0.05, * p<0.1

Table 2.10: Depreciation of Experience

A: One Week Graft Failure			
Center Volume in Previous 30 days (Log)	-0.00103**	-0.00103**	-0.00104**
	(0.000)	(0.000)	(0.000)
Center Volume in Previous 31 - 60 days (Log)		-0.00012	-0.00013
		(0.000)	(0.000)
Center Volume in Previous 61 - 90 days (Log)			-0.00026
			(0.000)
Observations	117,137	117,137	117,137
R-squared	0.034	0.034	0.034
B. One Month Graft Failure			
Center Volume in Previous 30 days (Log)	-0.00137**	-0.00136**	-0.00136**
	(0.001)	(0.001)	(0.001)
Center Volume in Previous 31 - 60 days (Log)		0.00042	0.00042
		(0.000)	(0.000)
Center Volume in Previous 61 - 90 days (Log)			-0.00019
			(0.001)
Observations	117,137	117,137	117,137
R-squared	0.038	0.038	0.038
C. One Year Graft Failure			
Center Volume in Previous 30 days (Log)	-0.00120	-0.00120	-0.00118
	(0.001)	(0.001)	(0.001)
Center Volume in Previous 31 - 60 days (Log)		-0.00006	-0.00003
		(0.001)	(0.001)
Center Volume in Previous 61 - 90 days (Log)			0.00066
			(0.001)
Observations	117,137	117,137	117,137
R-squared	0.062	0.062	0.062
Notes:			
1. Each regression includes controls for patient demographics, patient clinical information, and donor/transplant characteristics.			
2. For reference, the fixed effects estimates from Table 2.3 are reported in the first column.			
3. Clustered standard errors are reported in parentheses (clustered at transplant center-year level).			
4. *** p<0.01, ** p<0.05, * p<0.1			

2.8 Policy Implications

How should the kidney transplant community utilize this analysis with regard to policy? The strong implication of the "practice makes perfect" hypothesis is that care could be concentrated at *any* hospital or provider and outcomes would improve through the volume effect. If this is true, then redundancies in care can be eliminated, and patient outcomes would improve. The results of this analysis are not clear-cut with respect to the regionalization of kidney transplants. First, it appears that much of the observed volume-outcome relationship documented by Axelrod et al (2004) is due to between hospital differences in the quality and processes of care that explain both differences in transplant volume and patient outcomes. However, with regard to graft failure, there does appear to be a causal volume effect, although it is limited to the short run. Second, the results suggest that most of the volume effect on graft failure is concentrated at lower volume transplant programs. There is also evidence of a volume effect on short term (one week) patient mortality that is concentrated at higher volume transplant programs.

If lower volume transplant programs were closed, and patients were instead directed to higher volume transplant programs, then the results suggest that rates of graft failure would not improve, but that rates of short term patient mortality would decrease. Of course, these improvements in short term patient mortality would have to be weighed against any reductions in patient access to care that may arise from regionalization of care. In addition, any price effects generated by a reduction in competition would need to be accounted for as well. One possibility would be to eliminate smaller transplant programs within urban areas in which there are already other transplant programs. Therefore, patients within a particular urban area (or who are already traveling to a particular urban area for care) would experience minimal disutility that would arise

from restricted access to care, but may realize improved outcomes by being transplanted at higher volume transplant centers. A challenge to this approach is defining the exact volume threshold that would be used to make regionalization decisions. While this paper documents heterogeneous volume effects by average transplant center size, it is outside the scope of the current analysis to derive optimal thresholds for regionalization of care.

In addition, the results of this analysis have implications for the public reporting of transplant center volume and outcomes. As discussed, patients or their doctors may utilize the published SRTR reports to infer the quality of transplant centers in the patient's choice set. One drawback of the SRTR reports is that it statistically compares the actual patient outcomes at a particular transplant center to the expected outcomes, based on a statistical model. However, for lower volume transplant centers, average actual outcomes (like the one year mortality rate of transplant recipients) will be "noisier" because they are based on smaller sample sizes. Therefore, in a statistical sense, it is harder to reject the null hypothesis that a lower volume transplant center's outcomes are worse than the national average, even if they actually are in reality. If this is true, then patients may overestimate the quality of care performed at lower volume transplant centers. If volume is directly related to the underlying quality of a transplant program, then perhaps simply reporting transplant center volume would provide sufficient information about the quality of a transplant program. This may lead patients and their doctors to "selectively avoid" lower volume, and thus lower quality, transplant programs.

2.9 Conclusion

This paper tests whether there is a volume effect in kidney transplantation. Identification of the volume effect comes from supply shocks of transplantable kidneys from the local donor

supply that are plausibly exogenous with respect to unobserved determinants of patient outcomes. Indeed, conditional on transplant center-by-year fixed effects, transplant center volume in the 30 days leading up to a patient's transplant appears to be uncorrelated with many of observable factors of patient outcomes. However, because there is correlation between 30 day volume and donor status (deceased versus living), I also estimate IV regressions that use perfectly matched kidneys, considered to be a random event, to predict 30 day volume in the first stage. The results suggest that much of the observed volume-outcome association in kidney transplantation is due to between hospital differences that are correlated with both volume and outcomes, such as the quality of the transplant team. While this analysis is not an explicit "horse race" between the "practice makes perfect" and "selective referral" hypotheses, the results are not generally supportive of the idea that practice makes perfect. While there appears to be a volume effect for graft failure in the short run, it appears that much of that is concentrated among lower volume transplant centers.

The results presented here are qualitatively similar to other volume-outcome studies in the health economics and health services research literature. In particular, it appears that across a variety of procedures, much of the volume-outcome relationship demonstrated in the medical literature is not explained by the practice makes perfect hypothesis. This is not a universal result; learning by doing may be present in other procedures and contexts. But the findings of this paper and others like it suggest that more attention should be paid to between provider differences to explain variation in patient outcomes. This paper, like others, fails to document specific differences between hospitals, such as processes of care, availability of advanced technology, and the training of staff, that may explain the differences in outcomes of patients that are treated at higher or lower volume hospitals. Additional research is needed to identify these attributes,

and the medical community should consider the sharing of best practices so that hospitals with worse patient outcomes can learn how to improve their performance.

CHAPTER 3

The Impact of Temporal Breaks on the Delivery of Healthcare: Evidence from Kidney Transplantation

3.1 Introduction

A number of articles in the health economics and health services research literature have tested whether healthcare providers "learn by doing," in which experience – usually measured by procedural volume – affects patient outcomes. The previous chapter provides a review of this literature, and adds to the literature by testing for a volume effect in the context of kidney transplantation. The available evidence suggests that little of the observed volume-outcome relationship in healthcare can be explained by the learning by doing hypothesis. That is, there does not appear to be a causal effect of performing one more procedure on improved patient outcomes. However, recent work in the area has focused on the role that human capital depreciation and organizational forgetting play in healthcare. The former, often simply called "forgetting", suggests that individual workers maintain a level of skill through repetition of tasks, and that human capital depreciates during breaks or interruptions in the production schedule. Organizational forgetting can arise from the human capital depreciation among all workers within a firm, through worker turnover, or from changes in the production processes that affect worker productivity.

This chapter relates the concepts of human capital depreciation and organizational forgetting in the context of kidney transplantation. In particular, I test whether breaks in the production schedule for transplant centers, measured by the number of elapsed days between transplants, negatively affect patient outcomes. If repetition prevents the depreciation of the skills of the transplant team, then longer gaps of time between transplants may have negative

affects on the outcomes of the patients that are treated after these breaks. If this is the case, then repetition of tasks may help healthcare providers "learn by not forgetting." That is, if transplant teams are already at or near the flat of their learning curve, then repetition keeps them near the flat, and temporal breaks in production temporarily moves them away from this minimum, generating worse patient outcomes.

The identification strategy in this chapter is similar to the one used in the previous chapter. Because the majority of kidney transplants are performed using kidneys from deceased donors, and the timing of the arrival of these kidneys at a transplant center is unpredictable, then the gap of time between any two transplants is plausibly exogenously determined. That is, the number of days since a transplant center has performed its last transplant should be uncorrelated with unobserved patient or donor characteristics that affect patient outcomes. The challenge lies in the fact that living donor transplants are scheduled procedures, and they make up rough one-third of all transplants performed. The scheduling of living donor transplants may "smooth" the transplant center's production schedule, which could lead to biased regression estimates.[35] I find distinct evidence that living donor transplants often occur after longer breaks in the production schedule, and in particular, are scheduled for particular days of the week at some transplant centers. While I am able to directly control for donor status (living versus deceased), living donor transplant recipients, and/or their donors, may be healthier in unobserved ways that affect patient outcomes. This correlation would lead to negatively biased estimates of the effect of breaks in the production schedule on transplant recipient outcomes.

[35] In addition, the scheduling of living donor transplants may reduce the amount of variation in temporal breaks within a given transplant center in a given year. However, there appears to be sufficient variation in the size of temporal breaks within a transplant center in given year to plausibly identify the effects of temporal breaks on transplant recipients' outcomes.

Additionally, the length of the temporal breaks may have a nonlinear effect on productivity, as measured by patient outcomes. For example, it may be possible to perform "too many, too soon." This may be especially true at smaller transplant centers with a smaller transplant staff; performing transplants too close to one another may lead to fatigue among the transplant team members, for example. On the other hand, the marginal effect of an increase in the length of the temporal break between transplants may diminish after a particular point in the temporal break distribution. Following Hockenberry, Lien, and Chou (2008), I use a dummy variable scheme to flexibly test for nonlinearities in the gradient of the effect of temporal breaks in the production schedule on patient outcomes.

Finally, I test whether the level of transplant center experience, measured by recent transplant volume, before the temporal break mitigates any negative productivity effects generated by the elapsed number of days between transplants. In this scenario, high volume transplant centers with larger stocks of overall experience may be less likely to "forget" during temporal breaks, as compared to lower volume transplant centers with lower levels of accumulated experience. If lower volume centers are more prone to depreciation of skills that arise from temporal breaks, then a case can be made for centralizing kidney transplants at higher volume transplant centers.

The empirical results do not generally support the hypothesis that temporal breaks between transplants affect the productivity of transplant centers, as measured by the conditional graft failure rates of patients transplanted immediately following the break. In models where the size of the temporal break is interacted with the measure of recent transplant center experience, the coefficients on the interaction terms are negative, which indicates that higher levels of transplant experience mitigate the deleterious effects that arise from temporal breaks between

transplants. However, much of this association appears to be concentrated at the lower end of the temporal break distribution. An unexpected result is that relatively long temporal breaks do not seem to have any impact on productivity (patient outcomes) upon return to production, which is where one would expect an effect. I confirm these results using the subsample of transplant recipients that receive their kidneys from deceased donors. As expected, the randomness of arrivals of deceased donor kidneys from the local donor supply appears to randomly assign temporal breaks among this population; observable patient, donor, and transplant characteristics are largely uncorrelated with the size of temporal breaks. Despite the apparent randomness of temporal breaks at transplant centers (especially for deceased donor kidney recipients), there still may exist unobserved heterogeneity that may explain why I cannot identify an effect of temporal breaks on transplant center productivity. First, as I discuss later in the chapter, there may be non-random sorting of surgeons or staff along unobserved dimensions of ability that is correlated with the size of temporal breaks. Second, I cannot observe the activities of the members of the transplant team during breaks between transplants. If members of the transplant staff perform productive activities during the days between transplants, these activities may offset any kind of depreciation of skills that would have otherwise occurred.

The rest of the chapter is organized as follows. The next section describes some of the literature pertaining to organizational forgetting, the depreciation of experience, and the impacts of temporal breaks on productivity. The third section describes the data and construction of the estimation sample. Before turning to the empirical strategy, I focus on the distribution of temporal breaks in the sample, and document evidence of scheduling of living donor transplants at transplant centers in the sample. With that background, I next present the empirical strategy and describe the estimating equations. The estimation results follow, first for the full sample, and

then for the subsample of deceased donor recipients. I then describe alternative measures and methods that I employed, all of which yielded similar results. After a discussion of the results, the last section concludes the chapter.

3.2. Human Capital Depreciation and Organizational Forgetting

Previous studies in the labor economics literature have addressed the idea that human capital depreciates when a worker is out of the productive workforce for some period of time. Using panel data, Mincer and Ofek (1982) document that "reentry wages" are statistically lower than wages at the time of labor force exit for women in their sample, and that the decrease in wages is proportional to the amount of time that the woman spent out of the workforce. They interpret these results as evidence of human capital depreciation; if wages proxy for actual productivity, then longer work interruptions lead to higher levels of depreciation, and thus lower wages upon reentry into the workforce. However, the authors do not directly observe worker productivity.

In experimental settings, forgetting has been demonstrated even in the simplest of tasks. Globerson, Levin, and Shtub (1989) recruited subjects to perform computer data entry. At baseline, each subject entered the data from 16 personnel records into a computer database. Subjects were then invited back after randomly assigned breaks and asked to repeat the task. Breaks between data entry episodes ranged from as little as one day for some subjects, and up to 82 days for others. The results suggest that the subjects became more efficient between the first and sixteen repetitions. After the break, however, subjects had decreased productivity (measured by time to completion), and the level of forgetting was proportional to the length of the break. Bailey (1989) documented similar results when subjects were tasked with the assembly and

disassembly of Erector Sets. As Hockenberry, Lien, and Chou (2008) point out, if the subjects in these studies exhibit significant forgetting when performing relatively routine and low skill tasks, then the impact of production breaks may have a stronger effect for high skill tasks like surgery.

Probably the best known paper on organizational forgetting is Argote, Beckman, and Epple (1990) which tests for the level of depreciation of knowledge in the building of World War II Liberty Ships. They estimate that only 3.2 percent of the stock of production knowledge at the beginning of a production year persisted until the end of the year. Subsequent research by Thompson (2001) on the Liberty Ship experience has shown that the results of the original paper are sensitive to the inclusion of measures of capital investment. Changes in stock of capital available to workers will alter their productivity, and create biased estimates of the depreciation of human capital (at the firm level) if ignored. Benkard (2000) estimates similar regressions as Argote, Beckman, and Epple (1990) in the case of aircraft manufacturing and finds that 61 percent of the firm's stock of experience survives through the calendar year.

The drawback of these studies is that they fail to identify exactly how an organization "forgets." One possibility is that the depreciation of experience at the firm level is simply a reflection of the depreciation of human capital of the individual workers within the firm. Another possibility is that worker turnover generates depreciation in experience at the level of the firm. In this case, human capital may not necessarily depreciate, but firm productivity suffers when experienced workers leave the firm, and are replaced by new workers that must learn the skills of the job. Finally, as pointed out by Thompson (2001) and Benkard (2000), the firm may either change its capital level over time, or the production process itself may change. In the former, workers may become more productive, although there may be a period of decreased productivity as workers learn how to operate the new capital. The latter explanation is similar in nature. A

change in the production process may cause temporary decreases in worker productivity as workers adapt to the new process. Benkard (2000) even documents anecdotal evidence of the reluctance of aircraft manufacturing officials to alter production processes out of concern that such changes would disrupt workers' routines and lead to lower levels of productivity.[36]

Studies in the health economics and health services research literature have offered extensions of the lines of research described above. One way in which the health literature is different, though, is in its definition of "productivity." Typically, worker productivity is measured by the amount of time he or she takes to complete a task, or by the number of workers needed to produce a given level of output. In the health literature, the productivity of a provider is usually defined as her ability to "produce" improved patient outcomes, or as Hockenberry, Lien, and Chou (2008) put it, her ability to "extend life." That is, the surgeon with the quickest average time in the operating room may not necessarily be the most productive surgeon if her patients experience adverse events like surgical complications or mortality. However, using this altered definition of productivity still allows for a parallel to be drawn between learning and forgetting in healthcare, compared to other industries and settings.

For example, Gowrisankaran, Ho, and Town (2006) estimate the extent to which experience depreciates from one quarter to the next for hospitals performing three surgical procedures, with respect to inpatient mortality. They find significant heterogeneity in their estimates of knowledge depreciation across the three procedures, although it should be noted that their depreciation parameters were estimated assuming that hospital volume is exogenously determined. Given the discussion of selective referral and patient selection in the last chapter, volume is likely endogenous, which would lead to bias in their depreciation estimates.

[36] As pointed out by Brachet and David (2009), the practical issue at hand is the availability of data at the level of the worker, as well as availability of data at the firm level on job turnover. Without these data, the underlying mechanism that explains organizational forgetting cannot be identified.

More recently, Huesch (2009) examines patterns of learning and forgetting among "new" cardiac surgeons in Florida. If there is learning and depreciation of human capital, it might be most profound for workers (surgeons) who have recently completed their training (medical school residency). He finds no evidence of learning, and that experience almost fully depreciates from one quarter to the next. The advantage of his paper is that he tests for forgetting at the individual worker lever, rather than that at the level of the firm (hospital).

Using individual and firm level data, Brachet and David (2009) are able to test which mechanisms lead to organizational forgetting in the provision of emergency medical services. In particular, they estimate organizational experience at the beginning of the year is depreciated by approximately 75 percent by the end of the year. In addition, they find that labor turnover is responsible for 62 percent of the organizational forgetting, and that human capital depreciation accounts for the remaining 38 percent. Lastly, they test whether breaks in production (at the individual level) is a pathway through which human capital depreciates. They find that an increase in the number of days between emergency calls for an emergency technician leads to a small, yet statistically significant, increase in the response and delivery time needed to get a patient to the hospital. The institutional setting lends credibility to their study, as the number and timing of emergency calls are difficult to predict, meaning that temporal breaks between emergency calls are plausibly exogenously determined.

Finally, Hockenberry, Lien, and Chou (2008) find that breaks in the production schedules of cardiac surgeons in Taiwan lead to negative outcomes for patients, measured by one month mortality. A drawback of the study, however, is that the nature of the breaks is poorly understood. Unlike in Brachet and David (2009), where breaks in the production schedule can be explained by demand shocks for emergency care, temporal distance between surgeries for

cardiac surgeons may not be exogenous. For example, a surgeon may schedule a break in his or her operating schedule to attend a professional conference. If this is the case, the break in production is endogenous, and is directly related to investment in human capital. On the other hand, if a surgeon has a negative patient outcome, he or she may be placed on leave while the incident is investigated. However, the results from both Hockenberry Lien, and Chou (2008) and Brachet and David (2009) imply that temporal breaks may plausibly have an effect on the productivity of workers in the healthcare industry, which serves as motivation for investigating the issue in the context of kidney transplantation in this chapter.

This review of the literature highlights that there are many different, yet related studies on the topic of learning and forgetting, both at the firm and the worker level, across a variety of industries. This paper falls in line with other papers that examine these issues in the healthcare industry. Therefore, the definition of "productivity" used here will be improved patient outcomes, rather than lower costs or improved speed at a production task.[37] Second, because the data available are at the transplant center level, this analysis is at the organizational level. I do not observe surgeon identifiers. However, because transplantation is very much a team-oriented process, and patient outcomes dependent on the collective skills of the team, it seems appropriate to keep the analysis at the transplant center level. Finally, the goal of this paper is test whether breaks in the schedule of a transplant center lead to worse performance following a break. Therefore, I do not estimate depreciation factors, as in Gowrisankaran, Ho, and Town (2006), Huesch (2009), or Brachet and David (2009). The empirical approach is similar to that of Hockenberry, Lien, and Chou (2008), although I argue that the nature of kidney transplantation and the random arrival of transplantable kidneys lends itself to a more credible research design.

[37] More accurately, productivity in this chapter is defined as the absence of a negative patient outcome.

3.3 Data and Estimation Sample

The data used in this chapter come from the United Network of Organ Sharing (UNOS) Standard Transplant Analysis and Research Files (STAR) on all kidney waitlist, transplant, and follow-up activity as of August 2008. The estimation sample is similar to the one used in the previous chapter, but there are some differences, which I outline here.

The key explanatory variable is the number of days between transplants at a given transplant center. More explicitly, this variable is defined at the patient level, and is equal to the number of days since the patient's transplant center performed its last transplant. I do not observe the exact time of day that a transplant was performed, so I cannot define "time since last transplant" any more precisely that at the level of the calendar day. There are days in which in a transplant center performs multiple transplants. In these instances, each patient is assigned the number of days since the transplant center's last transplant that did not occur on the same day of the patient.[38] Therefore, the minimum temporal breaks is equal to one day.[39] As discussed earlier, I do not observe which members of the transplant team are involved with a particular transplant, so I cannot estimate the impact of temporal breaks at the level of individual (surgeon, nurse, anesthesiologist, etc). In the sample, there were some extreme values of the number days since the transplant center performed its last transplant. To ensure that the results are not generated by these extreme values, I drop observations where it has been greater than 30 days since the transplant center performed its last transplant. This drops 3,905 transplants from the sample, which represent less than 3 percent of the entire sample.

[38] For example, assume that a transplant center performs a transplant on May 1st, and then performs two transplants on May 4th. Then the two observations on May 4th would be coded as "3" for the number of days since the transplant center performed its last transplant.

[39] I also created an indicator variable for whether the transplant center performs another transplant on the same day that a patient receives her transplant. The inclusion of this variable into the estimating equations did not yield a statistically significant coefficient, and did not change the general findings. Therefore, I omit it from the analysis presented here and focus on the temporal breaks in the production schedule of the transplant center.

Because recent experience may help mitigate the depreciation of skills generated from temporal breaks in production, I also include a measure of recent transplant center volume. This is the same measure of volume used in the previous chapter: the natural log of 30 day transplant volume. For each transplant in the sample, I calculate the number of transplants performed at the recipient's transplant center in the previous 30 days. In addition, I control for recent volume because the previous chapter showed that recent volume affects some patient outcomes. And since recent volume and the size of temporal breaks between transplants will be negatively correlated, excluding 30 day volume would generate biased estimates of the impact of temporal breaks on transplant center productivity.

The patient outcome of interest in this chapter is graft failure, measured at various time periods post-transplant (one week, one month, and one year). The results of the previous chapter suggest that a volume effect exists for short-term (one week and one month) graft failure, although this effect appears to be concentrated at lower volume transplant centers. Therefore, if there is evidence of learning for short-term graft failure, then this seem like a natural outcome to examine in order to identify any kind of organizational forgetting that may arise from temporal breaks in the production schedule of the transplant center.

In order to adequately adjust for patient selection, I include controls for patient demographics, patient clinical information, and donor and transplant characteristics, which are defined in the previous chapter. The sample is restricted to adult transplant recipients (age 18 or older), because pediatric transplants are a specialized kind of transplant that are predominantly performed at pediatric hospitals. The sample is restricted to transplants that were performed between 1996 and 2005, and the final estimation sample includes 125,125 adult transplants performed at 243 transplant centers over the sample period.

3.4 Distribution of Temporal Breaks, and Scheduling of Living Donor Transplants

Before discussing the empirical strategy and estimating equations, I first turn to the distribution of temporal breaks in the estimation sample, which is displayed in Figure 3.1. There are two features to note. First, the distribution is heavily skewed. Because high volume transplant centers perform the majority of transplants in the sample, and because temporal breaks will be, on average, shorter at high volume transplant centers, there is a large number of observations at the lower end of the distribution. The mean temporal break is 5.54 days, and the median temporal break between transplants lasts 4 days.

The other feature of the distribution is the number of "spikes" that occur at multiples of seven. That is, there appears to be heaping of the data at 7 days, 14 days, 21 days, and so on. This suggests that there is some kind of weekly scheduling of transplants at transplant centers in

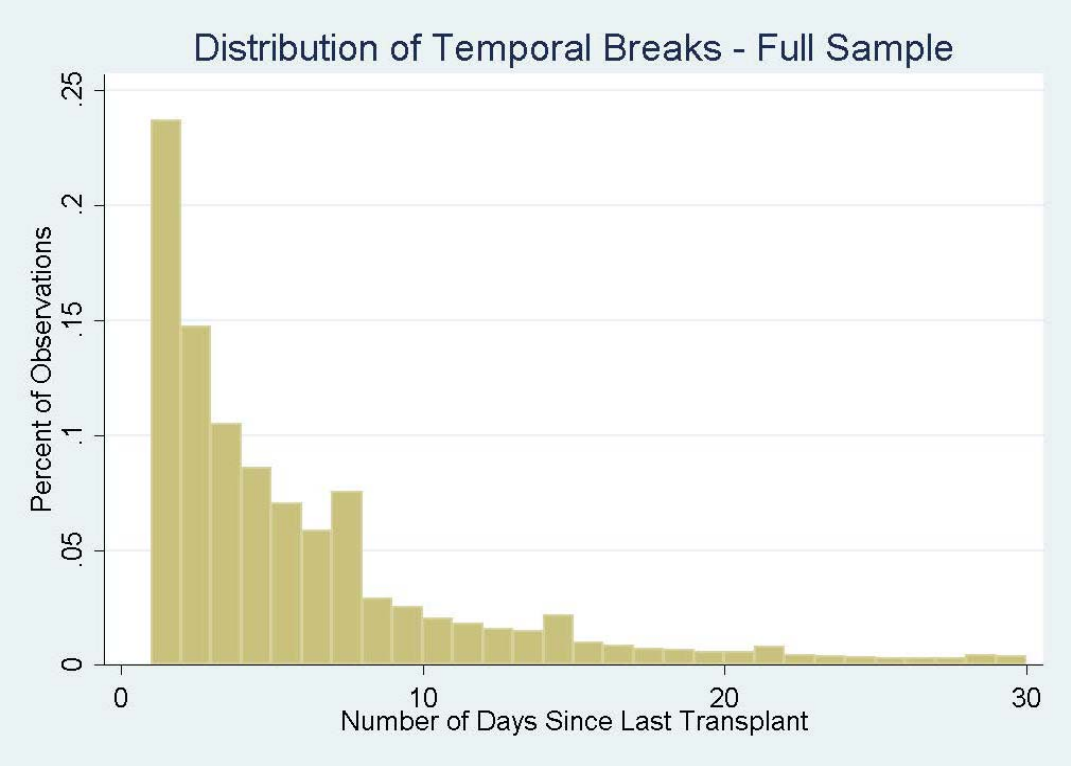

Figure 3.1

the sample. One would expect that given the randomness in the arrivals of transplantable kidneys that the scheduling of a transplant is somewhat impossible. However, living donor transplants are scheduled procedures, and make up approximately one third of all kidney transplants. Therefore, the clustering of the data at multiples of seven days may reflect the scheduling of living donor transplants at particular days of the week.

Figure 3.2 presents the distribution of temporal breaks for living donor transplant recipients only. The spikes in the distribution are more pronounced in this subsample and are suggestive of scheduling. However, one would still expect that the randomness of the arrival of deceased donor kidneys at a transplant center to eliminate any kind of heaping of the data at 7, 14, 21, etc. days for living donor recipients. This may be true for high volume transplant centers, but less so at lower volume transplant centers. That is, if living donor transplants are scheduled

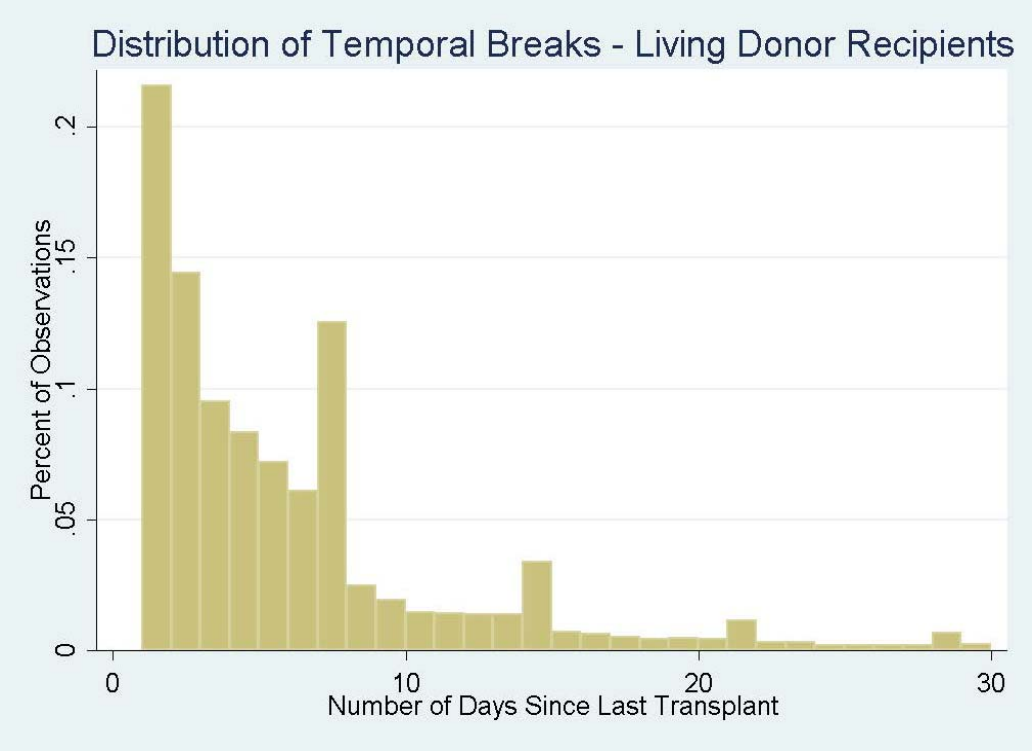

Figure 3.2

for particular days of the week, and low volume transplant centers experience large gaps in time between the arrivals of deceased donor kidneys from the local donor pool, then these spikes can be explained.

Figure 3.3 shows the distribution of temporal breaks for living donor recipients that have their transplant performed at a low volume transplant center, defined as being a transplant center in the lowest tercile of average annual transplant volume.[40] Again, the distribution has pronounced spikes at multiples of seven. Figure 3.4 displays the distribution of temporal breaks for living donor recipients at high volume transplant centers. Because high volume centers perform more deceased donor transplants than lower volume centers, and do so more frequently, then one would expect less heaping of the data at particular values of the temporal breaks (at

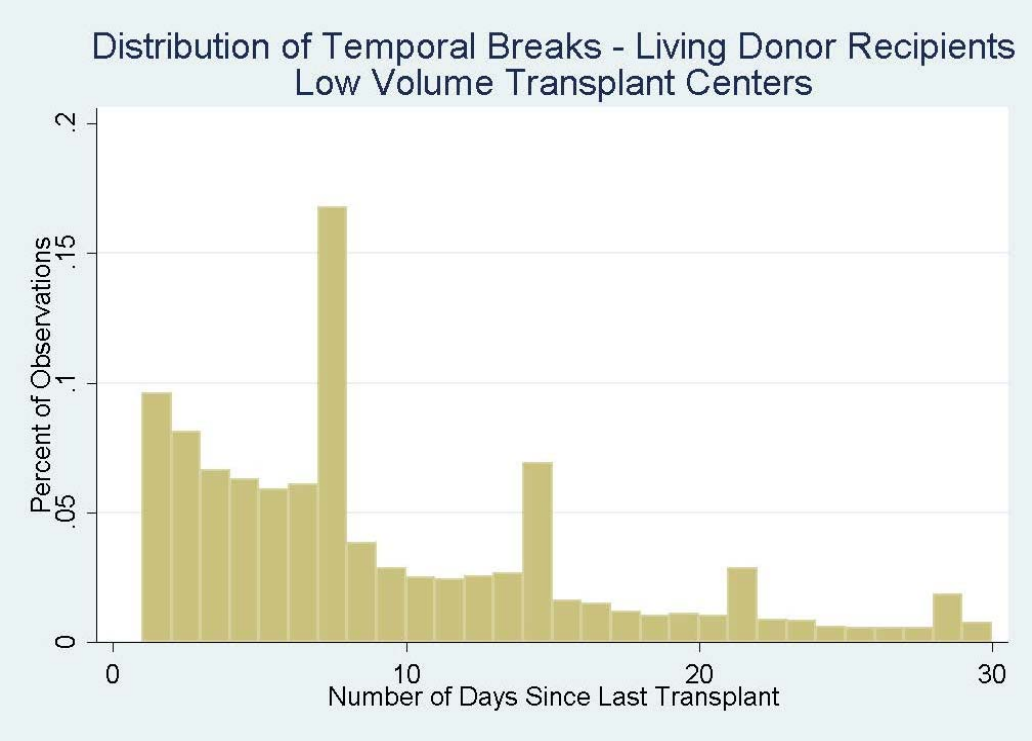

Figure 3.3

[40] In order to classify transplant centers as low or high volume, I first calculated the average annual transplant volume of each transplant center in the sample. I then broke the sample into terciles, whereby the lowest tercile is made up of transplants performed at the lower volume transplant centers in the sample. Note that each tercile has (roughly) the equivalent number of observations, but the number of transplant centers in each tercile varies.

least in relation to low volume transplant centers). Figure 3.4 confirms this; there is less evidence of heaping for the living donor transplant recipients at high volume transplant centers. These results suggest that the temporal breaks experienced by living donor recipients are more predictable at low volume centers.

In contrast, Figure 3.5 displays the distribution of temporal breaks experienced by recipients of deceased donor kidneys. Because the arrivals of these kidneys are plausibly unpredictable, then there should not be any evidence of heaping of the data at particular values in the distribution. The figure confirms this; while the distribution is heavily skewed, it appears to be "smooth", as expected.

Finally, the preceding analysis is predicated on the idea that transplant centers schedule their living donor transplants for particular days of the week. In order to detect this phenomenon, I calculated the exact day of the week that a transplant occurred. I tested whether the majority (at least 50 percent) of living donor transplants performed at a given transplant center occurred on one particular day of the week. I then repeated this exercise, using 75 percent as a cutoff. The results are presented in Table 3.1. Of the 243 transplant centers in the sample, 153 (63 percent) performed at least 50 percent of their living donor transplants on a particular day of the week, with Wednesday being the most popular day. Using the more restrictive cutoff (75 percent threshold), 73 transplant centers (30 percent) exhibited scheduling behavior, through which at least 75 percent of their living donor transplants were performed on a particular day of the week. As a contrast, this scheduling behavior is non-existent for deceased donor transplants, as expected.

The motivation behind these analyses is to get a better understanding of the variability and the distribution of temporal breaks in the sample. One might expect, ex ante, that breaks in

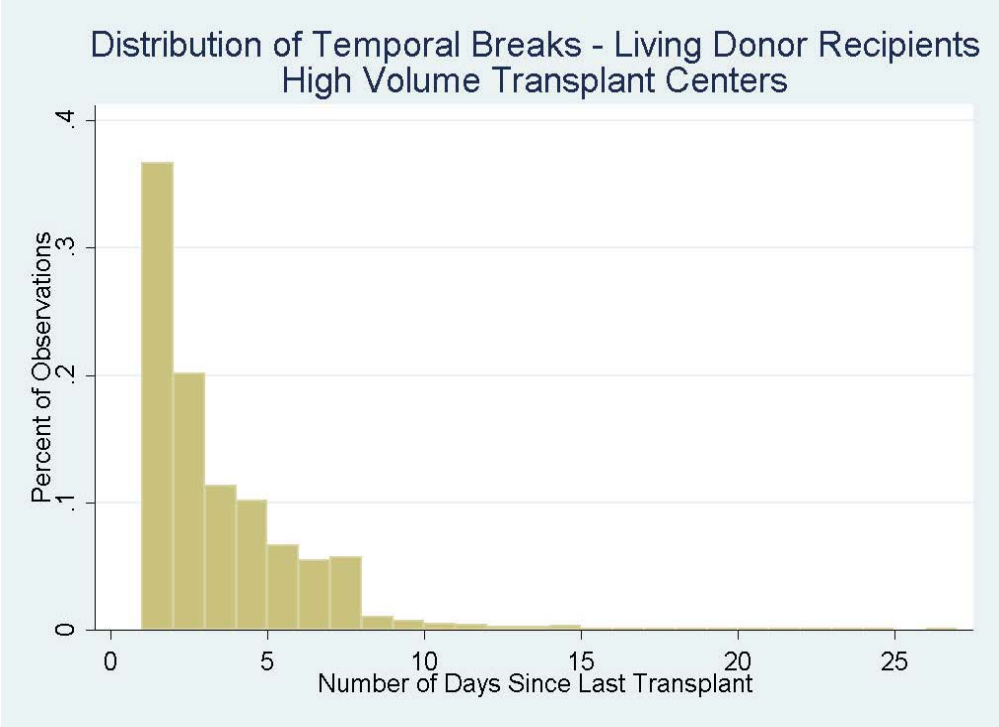

Figure 3.4

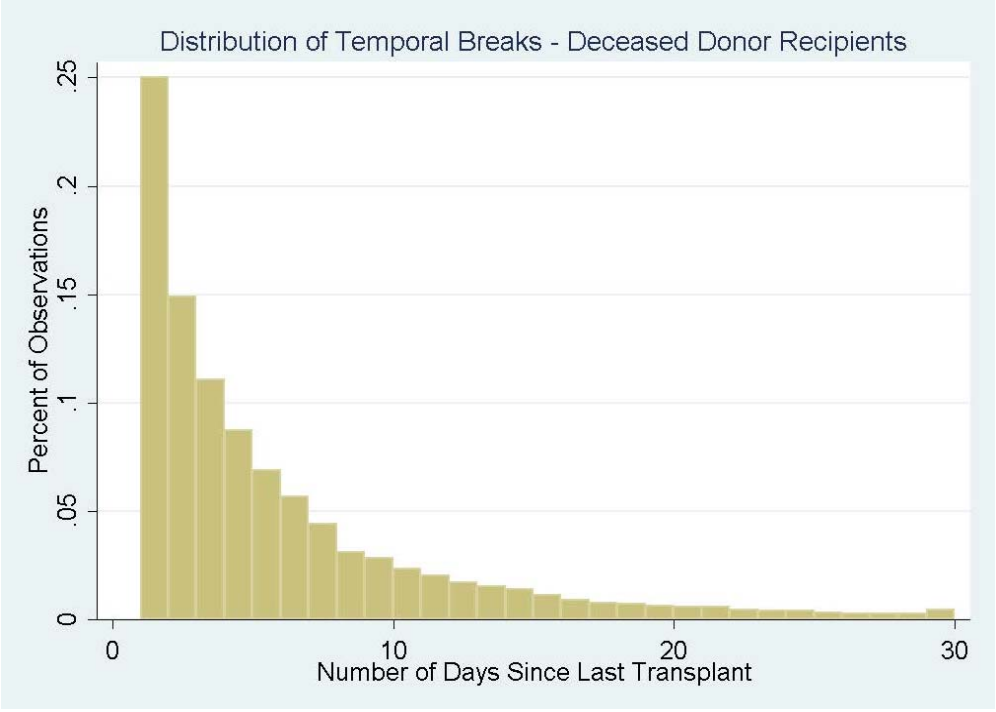

Figure 3.5

Table 3.1: Scheduling of Transplants on Particular Days of the Week

Living Donor Transplants: > 50 Percent Performed On...			Deceased Donor Transplants: > 50 Percent Performed On...		
	Number of Centers	Pct. Of Centers		Number of Centers	Pct. Of Centers
Sunday	0	0.0%	Sunday	0	0.0%
Monday	13	5.3%	Monday	0	0.0%
Tuesday	49	20.2%	Tuesday	0	0.0%
Wednesday	54	22.2%	Wednesday	0	0.0%
Thursday	33	13.6%	Thursday	0	0.0%
Friday	4	1.6%	Friday	0	0.0%
Saturday	0	0.0%	Saturday	0	0.0%
	153	**63.0%**		**0**	**0.0%**
Living Donor Transplants: > 75 Percent Performed On...			**Deceased Donor Transplants: > 75 Percent Performed On...**		
	Number of Centers	Pct. Of Centers		Number of Centers	Pct. Of Centers
Sunday	0	0.0%	Sunday	0	0.0%
Monday	8	3.3%	Monday	0	0.0%
Tuesday	20	8.2%	Tuesday	0	0.0%
Wednesday	29	11.9%	Wednesday	0	0.0%
Thursday	16	6.6%	Thursday	0	0.0%
Friday	0	0.0%	Friday	0	0.0%
Saturday	0	0.0%	Saturday	0	0.0%
	73	**30.0%**		**0**	**0.0%**

the production schedules of transplant centers would be as good as randomly assigned, due to the randomness of arrivals of transplantable kidneys from the local donor supply. However, I have found distinct evidence of heaping of the data at particular values of temporal breaks, specifically in multiples of seven. This heaping appears to be generated by the scheduling of living donor transplants on particular days of the week, which presents an empirical challenge in identifying the true causal effect of temporal breaks on patient outcomes.

3.5 Empirical Strategy

This paper tests whether temporal breaks in the production schedule of kidney transplant centers lead to lower levels of productivity after the break, as measured by patient outcomes. The preceding section dissects the distribution of temporal breaks in the sample, and indicates that temporal breaks may be correlated with both observable and unobservable patient or donor characteristics. With this in mind, I turn to the estimation strategy.

The general form of the estimating equation is:

$$(3.1)\quad y_{iht} = \beta_0 + \beta_1 Days_{iht} + \beta_2 Vol_{iht} + \beta_3 X_{iht}^{Dem} + \beta_4 X_{iht}^{Clin} + \beta_5 X_{iht}^{Don} + D_{ht} + \xi_{iht}$$

where y_{iht} represents the outcome of transplant recipient *i* who received her transplant in year *t* at transplant center *h*. As is the previous chapter, X_{iht}^{Dem}, X_{iht}^{Clin}, and X_{iht}^{Don} are vectors of patient demographic characteristics, patient clinical information, and donor characteristics, respectively. Recent experience is captured by Vol_{iht}, which the natural log of the number of transplants performed at transplant center *h* in the 30 days preceding recipient *i*'s transplant. The size of the temporal break in production experienced at transplant center *h* before recipient *i*'s transplant is captured by $Days_{iht}$, which is the number of elapsed days since the transplant center's last transplant. D_{ht} is a set of transplant center-by-year fixed effects. Therefore, the identification of the effect of temporal breaks in production comes from comparing two otherwise identical transplant recipients that are transplanted at the same transplant center in the same year, but are transplanted following different temporal breaks in the transplant center's production schedule.

Entering the size of the temporal break as a continuous variable into equation 3.1 might not yield regression estimates that fully capture the effect of temporal breaks on productivity. First, ordinary least squares will yield an estimate that is based on deviations from the conditional mean of temporal breaks, which does not provide any information about the effect of temporal breaks at all points in the distribution. In particular, if the effect of a break in production is nonlinear in the size of the break, then simply entering the number days between transplants as a continuous variable will likely mask these effects. Second, the size of the temporal break is correlated with the probability of receiving a kidney from a living donor, as discussed in the previous section. The average temporal break lasts 5.54 days, and the median

break is 4 days. But living donor transplants appear to be clustered where the temporal break is 7, 14, 21, etc. days, especially at low volume transplant centers. Therefore, living donor transplants are more likely to be performed after temporal breaks of above-average length. While I can directly control for donor status (living versus deceased), as well as for observable patient and donor characteristics, living donor transplant recipients and their donors may be healthier in unobserved ways, thus producing better outcomes. This would lead to negatively biased estimates (biased toward zero) of the effect of temporal breaks on patient outcomes.

In order to address these issues, I include dummy variables for various values of temporal breaks in the sample. The motivation for using dummy variables, rather than a continuous measure of temporal breaks, is that it allows me to detect effects along all points in the distribution. Also, I am able to directly control for transplants that are performed after temporal breaks of 7, 14, 21, etc. days, since these transplants are, on average, quite different than the transplants performed at other points in the distribution. Therefore, the primary estimating equation is given by:

(3.2) $$y_{iht} = \pi_0 + \gamma' iDays_{iht} + \pi_1 Vol_{iht} + \pi_2 X_{iht}^{Dem} + \pi_3 X_{iht}^{Clin} + \pi_4 X_{iht}^{Don} + D_{ht} + \xi_{iht}$$

which is similar to equation 3.1, except that a set of dummy variables ($iDays_{iht}$), replace the absolute number of days between transplants. I include indicator variables for the following values of the number of days since a transplant center's last transplant: < 4 days, 4 days, 5-6 days, 7 days, 8-13 days, 14 days, 15-20 days, 21 days, and > 21 days. The median temporal break is 4 days, and will serve as the reference group when I estimate equation 3.2. By explicitly including dummies for 7, 14, and 21 days, I partially address the fact that the transplants performed after these particular temporal breaks are more likely to be living donor transplants. Therefore, the estimated coefficients on these three dummy variables are likely to be negatively

biased, but the coefficients on the indicators for the other points in the distribution are plausibly unbiased.[41]

The usual assumption is that the more frequently a provider performs a procedure (i.e. the smaller the temporal break), the more productive it will be in terms of patient outcomes. In theory, that assumption may not be completely valid. While most economists would agree that longer temporal breaks will lead to the depreciation of skills, it may also be the case that doing "too much, too soon" may lead to a higher rate of mistakes. This may be true in kidney transplantation, especially for low volume transplant centers with smaller transplant staffs. For example, if a transplant team performs transplants on consecutive days, it may become fatigued, which may lead to errors and worse patient outcomes. If this is true, then there may be an optimal temporal break (that is greater than zero) that balances the benefit of maintaining skills against the detriment caused by provider fatigue. Therefore, while one would assume that the coefficient on the < 4 days dummy variable to be negative (indicating improved patient outcomes), I argue that the expected sign is ambiguous.

I also estimate a version of equation 3.2 that includes interaction terms of recent experience (measured by 30 day volume) with the set of temporal break dummy variables. In the previous chapter, recent volume was shown to reduce the probability of short term graft failure, and short term patient mortality in some specifications. By interacting recent volume with the size of the temporal break, I am able to test whether the benefit of recent volume mitigates any

[41] That is, relative to the reference group, who are transplanted after the medial temporal break of 4 days, patients that are transplanted after temporal breaks of <4 days, 5 or 6 days, 8 to 13 days, 15 to 20 days, and 22+ are alike in all ways that affect the probability of a negative outcomes expect for the fact that they are exposed to different temporal breaks. Those transplanted after temporal breaks of 7, 14, and 21 days are on average different than the observations in the reference group, primarily because they are more likely to receive their kidneys from living donors.

deleterious effect that is created by temporal breaks between transplants. The estimating equation in this case is:

(3.3) $$y_{iht} = \alpha_0 + \gamma' iDays_{iht} + \delta' iDays_{iht} \text{ x } Vol_{iht} + \alpha_1 Vol_{iht} + \alpha_2 X_{iht}^{Dem} + \alpha_3 X_{iht}^{Clin} + \alpha_4 X_{iht}^{Don} + D_{ht} + \xi_{iht}$$

If higher levels of recent experience help protect a transplant center against any negative effects that arise from temporal breaks in its production schedule, then I expect that the coefficients on the interaction terms (the vector δ) to be negative, which would indicate a reduction in the probability of a negative outcome like graft failure or patient death.

In equations 3.2 and 3.3, I utilize variation in temporal breaks within a transplant center in a given year to explain variation in transplant recipient outcomes. The fixed effects are needed to account for unobserved differences of the quality of care that is provided at transplant centers in the sample. For example, if unobserved quality drives volume (as in the selective referral hypothesis), and volume is negatively correlated with the size of temporal breaks, then simply using cross-sectional variation in temporal breaks will produce biased estimates of the impact of temporal breaks on patient outcomes. Even after conditioning on the set of fixed effects, temporal breaks may not be randomly assigned, as suggested by the apparent scheduling of living donor transplants on particular days of the week. The use of dummy variables for specific temporal breaks (especially at 7, 14, 21, etc.) help mitigate the concern that the scheduling of living donor transplants generates biased estimates of the effects of breaks in production on patient outcomes.

As a final estimation strategy, I estimate equations 3.2 and 3.3 using the subsample of transplants that come from deceased donor kidneys. As figure 3.5 shows, and table 3.1 implies, there is little evidence that deceased donor transplants can be, or are, scheduled in any particular way. That is, the random arrival of kidneys from the pool of deceased donors generates

randomness in the size of the temporal breaks for recipients of these kidneys. Note that in this analysis, the size of the temporal break is defined as the number of days between a deceased donor transplant, and the last transplant a center performed, regardless of donor status.[42] In this analysis, I also use a set of dummy variables to measure the size of the temporal breaks, but I redefine the set of dummy variables to indicate the following temporal breaks: < 4 days, 4 days, 5-7 days, 8-14 days, 15-21 days, and > 21 days. Because there are not any spikes in the distribution of temporal breaks for recipients of deceased donor transplants, I remove the dummy variables that indicate 7, 14, and 21 days, which are used when I estimate equations 3.2 and 3.3 on the full sample of transplants.

The question, then, is whether temporal breaks for recipients of deceased donor transplants can be considered as good as randomly assigned. The nature of kidney transplantation suggests that they might; the arrival of a transplantable kidney from the pool of local deceased donors is an unanticipated event with irregular timing.[43] But consider the case where a transplant center has multiple transplant surgeons, and these surgeons have varying levels of skills and expertise. Assume that a high ability surgeon typically performs a transplant, but during periods when the transplant center performs a relatively high number of transplants (such that the temporal break between transplants is short), the lower ability surgeons are called upon because either the high ability surgeon is unavailable, or is fatigued. In this scenario, there is non-random sorting of surgeons to patients that is correlated with the number of elapsed days between transplants, and unobserved heterogeneity in surgeon ability affects patient outcomes. In this

[42] In other words, the temporal break measured here is *not* the number of days since the last deceased donor transplant, but rather the number of days since the last transplant, regardless of the whether the last transplant used a kidney from a living or a deceased donor.

[43] Later in the chapter, I explicitly test whether the observable patient, donor, and transplant characteristics of deceased donor transplants are correlated with temporal breaks. Evidence that observable characteristics are uncorrelated with the size of temporal breaks lends credibility to the assumption that unobservable determinants of patient outcomes are also uncorrelated with the magnitude of temporal breaks.

chapter, I cannot address this kind of sorting of surgeons to patients within a transplant center because I do not observe any surgeon identifiers. Sorting of the nature that I just described would make it more difficult to detect the deleterious effect of longer temporal breaks because patients that are transplanted after relatively short temporal breaks may be more likely to be assigned to a lower ability surgeon, and thus have worse outcomes. In other words, the coefficients on the dummy variables for longer temporal breaks will be negatively biased, since these patients are more likely to be operated on by a higher ability surgeon. Ultimately, the fixed effects are not adequate to address this potential source of bias, and I cannot practically test for non-random sorting of surgeons to patients without data at the surgeon level. But the bias implied by this scenario suggests that any effect that I do estimate is likely a lower bound of the effect of temporal breaks on patient outcomes.

There are three outcomes of interest: one week, one month, and one year graft failure. These are measured with indicator variables, so I estimate equations 3.2 and 3.3 with linear probability models.[44] I also cluster standard errors at the transplant center-by-year level. All models include the full set of patient demographic, patient clinical, and donor/transplant characteristics to adjust for underlying differences in the health of the patient or donor that may be correlated with the size of temporal breaks. To be clear, for each outcome, I estimate four regressions: I estimate equations 3.2 and 3.3 using the full sample, and again using the subsample of deceased donor kidney recipients. Using the full sample, I include dummy variables for temporal breaks of 7, 14, and 21 days to account for the apparent scheduling of living donor transplants. When I use the subsample, I use an alternative dummy variable scheme

[44] The results are robust to using a probit model. I use the linear probability model because it generates coefficients that represent intuitive marginal effects.

since there does not appear to be heaping of observations at these specific temporal breaks for recipients of deceased donor kidneys.

3.6 Results

Table 3.2 presents average patient and donor characteristics for the full sample, and also split by the temporal break dummy variables.[45] At first blush, it appears that, on average, patient outcomes are better after shorter temporal breaks than after longer breaks, although an obvious gradient is not apparent. Part of the difficulty in detecting a gradient in patient outcomes comes from the relatively low rates of graft failure that are exhibited at temporal breaks of 7, 14, and 21 days. These lower graft failure rates correspond to the fact that the majority of transplants performed at these particular points in the temporal break distribution come from living donors, and recipients of living donor kidneys typically enjoy better outcomes than recipients of deceased donor kidneys. Patient and donor characteristics appear to be fairly balanced across the temporal break distribution, except at 7, 14, and 21 days. At these particular temporal breaks, recipients are slightly younger, are more likely to be white, are more likely to be privately insured, and are more likely to be college educated. These differences in patient characteristics are likely reflecting the heaping of living donor transplants at these points in the distribution.[46] Similarly, patients that receive their transplants after temporal breaks of 7, 14, or 21 days are less likely to be on dialysis at the time of transplant, and are less likely to be immunologically sensitized, suggesting that these recipients are, on average, healthier than the recipients that

[45] For brevity, I only report the descriptive statistics for a subset of the covariates. A "full" version of Table 3.2 is available from the author.

[46] The descriptive statistics presented in the first chapter confirm this relationship: recipients of living donor kidneys are, on average, younger, more educated, and more likely to be White and privately insured, as compared to those that do not opt for living donation.

Table 3.2: Descriptive Statistics

			Number of Days Since Last Transplant								
		Full Sample	< 4	4 Days	5 or 6	7	8 to 13	14	15 to 20	21	22+
Number of Observations		125,125	61,176	10,743	16,076	9,399	15,332	2,683	5,310	996	3,410
Number of Days Since Last Transplant	Mean	5.54	1.73	4.00	5.45	7.00	10.09	14.00	17.15	21.00	25.52
	St. Dev.	5.57	0.79		0.50		1.70		1.72		2.54
Number of Transplants Performed in Last 30 Days	Mean	9.54	12.60	9.87	8.44	6.66	5.36	3.85	3.23	2.43	1.85
	St. Dev.	7.20	7.81	6.27	5.39	4.16	3.39	2.28	1.99	1.41	1.14
Outcomes											
Graft Failure: One Week	Mean	0.022	0.021	0.022	0.024	0.017	0.023	0.023	0.025	0.017	0.024
	St. Dev.	0.145	0.143	0.145	0.153	0.131	0.151	0.150	0.155	0.130	0.152
Graft Failure: One Month	Mean	0.036	0.035	0.038	0.039	0.032	0.040	0.037	0.036	0.029	0.039
	St. Dev.	0.187	0.185	0.190	0.193	0.176	0.196	0.189	0.186	0.168	0.193
Graft Failure: One Year	Mean	0.091	0.091	0.093	0.093	0.077	0.093	0.083	0.093	0.092	0.094
	St. Dev.	0.287	0.288	0.290	0.290	0.267	0.290	0.276	0.290	0.290	0.292
Covariates											
Donor: Living	Mean	0.380	0.355	0.370	0.395	0.638	0.314	0.601	0.296	0.551	0.326
	St. Dev.	0.485	0.479	0.483	0.489	0.481	0.464	0.490	0.457	0.498	0.469
Recipient Age	Mean	47.33	47.62	47.35	47.21	46.09	47.38	45.94	47.27	46.19	47.27
	St. Dev.	13.22	13.21	13.22	13.24	13.41	13.20	12.99	13.02	12.97	13.18
Recipient White	Mean	0.590	0.582	0.596	0.592	0.638	0.578	0.643	0.574	0.616	0.590
	St. Dev.	0.492	0.493	0.491	0.491	0.481	0.494	0.479	0.495	0.486	0.492
Primary Insurer: Private	Mean	0.410	0.419	0.413	0.415	0.482	0.371	0.444	0.351	0.405	0.330
	St. Dev.	0.492	0.493	0.492	0.493	0.500	0.483	0.497	0.477	0.491	0.470
College or Higher	Mean	0.353	0.342	0.356	0.355	0.393	0.358	0.384	0.355	0.376	0.363
	St. Dev.	0.478	0.474	0.479	0.478	0.488	0.480	0.486	0.479	0.484	0.481
On Dialysis	Mean	0.832	0.826	0.826	0.827	0.794	0.863	0.806	0.866	0.830	0.878
	St. Dev.	0.374	0.379	0.379	0.378	0.405	0.344	0.395	0.340	0.376	0.327
Recipient Sensitized (PRA > 19%)	Mean	0.097	0.100	0.105	0.097	0.086	0.099	0.084	0.089	0.078	0.091
	St. Dev.	0.296	0.300	0.307	0.295	0.280	0.299	0.278	0.285	0.269	0.287
HLA Mismatches	Mean	3.242	3.282	3.223	3.224	3.115	3.245	3.091	3.232	3.187	3.194
	St. Dev.	1.810	1.810	1.823	1.816	1.768	1.815	1.783	1.821	1.755	1.828
Days Waiting	Mean	506.6	531.6	512.0	501.6	367.2	522.3	381.9	526.1	399.8	490.7
	St. Dev.	596.8	609.1	604.0	600.1	526.8	594.9	521.2	588.9	556.4	562.4

receive their transplants at different points in the temporal break distribution. In addition, they are better matched to their donors immunologically (fewer HLA (antigen) mismatches) and spend less time waiting for their transplants. Again, these differences are likely explained by the heaping of living donor transplants at these particular points in the temporal break distribution. Table 3.2 also shows that the size of temporal breaks is negatively correlated with recent (30 day) transplant volume, as expected.

Table 3.3 presents the results from the fixed effects estimations of equation 3.2. In each regression, I control for patient demographics, patient clinical information, and donor/transplant characteristics, as well as the set of transplant center-by-year dummy variables. The reference group in each regression is made up of patients who were transplanted after the median temporal break in the data (4 days). The results suggest that, all else equal, patients who are transplanted after below-median temporal breaks (less than 4 days) experience a lower probability of graft failure, as compared to the reference group, although this relationship is statistically insignificant. Generally speaking, it appears that patients transplanted after above-median temporal breaks (> 4 days) experience higher rates of graft failure than the reference group, but the implied gradient does not support the notion that graft failure rates increase linearly with the size of the temporal break. For example, patients transplanted after a temporal break of 5 or 6 days have an increased probability of one-week graft failure of about a quarter of a percentage point, as compared to the reference group, although the coefficient is imprecisely estimated. Relative to patients transplanted after a temporal break of 5 or 6 days, the results suggest that patients transplanted after a temporal break of 8 to 13 days have *lower* rates of one-week graft failure, which implies that the increased temporal break actually improved outcomes (although the graft failure rate is still higher than in the reference group). Again, this estimate is statistically

Table 3.3: Impact of Temporal Breaks on Graft Failure Rates – Full Sample

	Probability of Graft Failure Within…		
	One Week	One Month	One Year
Temporal Break			
< 4 days	-0.00044	-0.00204	-0.00209
	(0.002)	(0.002)	(0.003)
5 or 6 Days	0.00232	0.00069	0.00177
	(0.002)	(0.002)	(0.004)
7 Days	-0.00141	-0.00104	0.00117
	(0.002)	(0.003)	(0.004)
8 - 13 Days	0.00071	0.00016	-0.00448
	(0.002)	(0.002)	(0.004)
14 Days	0.00326	0.00288	0.00540
	(0.003)	(0.004)	(0.006)
15 - 20 Days	0.00171	-0.00552*	-0.00711
	(0.003)	(0.003)	(0.005)
21 Days	-0.00325	-0.00724	0.00896
	(0.005)	(0.006)	(0.010)
22 + Days	-0.00008	-0.00345	-0.00492
	(0.003)	(0.004)	(0.006)
(log) 30 Day Transplant Volume	0.00016	0.00009	0.00120
	(0.001)	(0.001)	(0.002)
Mean of outcome for reference group	0.022	0.038	0.093
Observations	125,125	125,125	125,125
R-squared	0.032	0.036	0.059
Notes:			
1. The temporal break refers to the number of days that have elaspsed since the transplant center's last transplant. The reference temporal break is 4 days, which is the median temporal break in the estimation sample.			
2. Each regression controls for patient demographics, patient clinical information, and donor/transplant characteristics, as well as the set of transplant center-by-year fixed effects.			
3. Recent transplant center experience is captured by the natural log of transplant center volume in the 30 days preceding each patient's transplant.			
4. Clustered standard errors are reported in parentheses (clustered at transplant center-year level).			
5. *** p<0.01, ** p<0.05, * p<0.1			

insignificant. If temporal breaks do have any negative impact on patient outcomes, then one would expect that graft failure rates would be higher after very long breaks, as compared to patients transplanted after the median temporal break. However, the coefficient on the dummy variable for a temporal break of at least 22 days is negative, implying these patients experience lower rates of one-week graft failure, as compared to the reference group. Once again, this estimate is statistically imprecise, but its sign is unexpected.

The estimates reported in table 3.3 do not indicate a statistically significant relationship between temporal breaks and patient outcomes. However, it may be that any effects of temporal breaks may be more pronounced at lower volume transplant centers, which have less accumulated experience to draw upon, and thus may be more likely to "forget" after longer than average temporal breaks between transplants. In other words, higher levels of transplant volume, a measure of higher general experience, may serve to protect a transplant staff from forgetting during temporal breaks between transplants. By estimating equation 3.3, I test whether increased experience, as measured by recent (30 day) volume, helps negate any deleterious effects of increased temporal breaks.

Table 3.4 presents the estimated coefficients from the model where the set of temporal break dummy variables are interacted with the (log of) recent transplant volume. As expected, the coefficients on the interaction terms are generally negative, indicating that any deleterious effects of temporal breaks are mitigated by higher levels of recent volume. However, the only statistically significant effects appear to be located at the lower end of the temporal break distribution. These results suggest that the effects of temporal breaks are more likely to manifest themselves at lower volume transplant centers, which have, on average, lower levels of 30 day volume. Table 3.5 presents the implied effects of temporal breaks at different points in the

Table 3.4: Interaction Model – Full Sample

	Probability of Graft Failure Within…		
	One Week	One Month	One Year
Temporal Break			
< 4 days	0.00882*	0.01076*	0.02325**
	(0.005)	(0.006)	(0.010)
5 or 6 Days	0.01359**	0.01364*	0.02631**
	(0.006)	(0.007)	(0.012)
7 Days	0.00045	0.00678	0.02514**
	(0.006)	(0.008)	(0.012)
8 - 13 Days	0.00080	0.00795	0.00644
	(0.005)	(0.007)	(0.011)
14 Days	0.00457	0.00558	0.01721
	(0.008)	(0.009)	(0.015)
15 - 20 Days	0.00568	-0.00099	0.00811
	(0.006)	(0.007)	(0.012)
21 Days	0.00533	0.00094	0.03479**
	(0.009)	(0.011)	(0.017)
22 + Days	0.00462	0.00582	0.01208
	(0.006)	(0.008)	(0.011)
(log) 30 Day Transplant Volume	0.00259	0.00430	0.00992**
	(0.002)	(0.003)	(0.004)
Interaction Terms			
< 4 Days x Volume	-0.00432**	-0.00601**	-0.01191***
	(0.002)	(0.003)	(0.004)
5 or 6 Days x Volume	-0.00559**	-0.00633*	-0.01195**
	(0.003)	(0.003)	(0.005)
7 Days x Volume	-0.00042	-0.00354	-0.01192**
	(0.003)	(0.004)	(0.006)
8 - 13 Days x Volume	0.00110	-0.00338	-0.00359
	(0.003)	(0.003)	(0.005)
14 Days x Volume	0.00099	0.00114	-0.00296
	(0.005)	(0.006)	(0.010)
15 - 20 Days x Volume	-0.00104	0.00030	-0.00535
	(0.004)	(0.005)	(0.007)
21 Days x Volume	-0.00687	-0.00311	-0.01856
	(0.008)	(0.010)	(0.015)
22 + Days x Volume	-0.00118	-0.00477	-0.00559
	(0.006)	(0.007)	(0.011)
Observations	125,125	125,125	125,125
R-squared	0.032	0.036	0.059

Notes:

1. The temporal break refers to the number of days that have elaspsed since the transplant center's last transplant. The reference temporal break is 4 days, which is the median temporal break in the estimation sample.
2. Each regression controls for patient demographics, patient clinical information, and donor/transplant characteristics, as well as the set of transplant center-by-year fixed effects.
3. Recent transplant center experience is captured by the natural log of transplant center volume in the 30 days preceding each patient's transplant.
4. Clustered standard errors are reported in parentheses (clustered at transplant center-year level).
5. *** p<0.01, ** p<0.05, * p<0.1

Table 3.5: Implied Marginal Effects by Percentile of Recent Volume

		Percentile of 30 Day Volume				
		10th (2)	25th (4)	50th (7)	75th (13)	90th (20)
One Week Graft Failure	Temporal Break					
	< 4 days	0.006	0.003	0.000	-0.002	-0.004
	5 or 6 Days	0.010	0.006	0.003	-0.001	-0.003
	7 Days	0.000	0.000	0.000	-0.001	-0.001
	8 - 13 Days	0.002	0.002	0.003	0.004	0.004
	14 Days	0.005	0.006	0.006	0.007	0.008
	15 - 20 Days	0.005	0.004	0.004	0.003	0.003
	21 Days	0.001	-0.004	-0.008	-0.012	-0.015
	22 + Days	0.004	0.003	0.002	0.002	0.001
		Percentile of 30 Day Volume				
		10th (2)	25th (4)	50th (7)	75th (13)	90th (20)
One Month Graft Failure	Temporal Break					
	< 4 days	0.007	0.002	-0.001	-0.005	-0.007
	5 or 6 Days	0.009	0.005	0.001	-0.003	-0.005
	7 Days	0.004	0.002	0.000	-0.002	-0.004
	8 - 13 Days	0.006	0.003	0.001	-0.001	-0.002
	14 Days	0.006	0.007	0.008	0.009	0.009
	15 - 20 Days	-0.001	-0.001	0.000	0.000	0.000
	21 Days	-0.001	-0.003	-0.005	-0.007	-0.008
	22 + Days	0.003	-0.001	-0.003	-0.006	-0.008
		Percentile of 30 Day Volume				
		10th (2)	25th (4)	50th (7)	75th (13)	90th (20)
One Year Graft Failure	Temporal Break					
	< 4 days	0.015	0.007	0.000	-0.007	-0.012
	5 or 6 Days	0.018	0.010	0.003	-0.004	-0.009
	7 Days	0.017	0.009	0.002	-0.005	-0.011
	8 - 13 Days	0.004	0.001	-0.001	-0.003	-0.004
	14 Days	0.015	0.013	0.011	0.010	0.008
	15 - 20 Days	0.004	0.001	-0.002	-0.006	-0.008
	21 Days	0.022	0.009	-0.001	-0.013	-0.021
	22 + Days	0.008	0.004	0.001	-0.002	-0.005

Notes:

1. The numbers in parentheses in the column headers are the number of transplants performed in the last 30 days at each point in the distribution. For example, 2 transplants in the previous 30 days is at the 10th percentile in the distribution of 30 day volume.

distribution of 30 day volume. After below-median temporal breaks (< 4 days), patients that are transplanted after below-median levels of 30 day volume experience an increase in the probability of graft failure, relative to the reference group. This is consistent with the idea that at lower-volume transplant centers, doing "too many, too soon" may lead to worse patient outcomes, although the mechanism through which this occurs (like transplant staff fatigue) cannot be identified here. As a contrast, patients that are transplanted at transplant centers coming off of relatively high 30 day volume experience a decrease in the probability of graft failure if they are transplanted after a below-median temporal break.

An unexpected finding is that the effect of temporal breaks does not appear to be linear in the size of the temporal break. In fact, the largest effects appear to be experienced by patients transplanted after very low 30 day volume (in the 10^{th} percentile of 30 day volume) and after a temporal break of 5 or 6 days. One would expect that longer temporal breaks, such as 22 or more days, would have the largest effect on patient outcomes, but this notion is not supported by the results in table 3.5.

3.7 Analysis on the Subsample of Deceased Donor Recipients

As discussed earlier, one of the empirical challenges in identifying the effects of temporal breaks on the outcomes of kidney transplant recipients is that there is clustering of living donor transplants at particular points in the temporal break distribution. The dummy variable scheme employed above partially addresses this challenge, but here I present the empirical results using the subsample for which temporal breaks are plausibly exogenous: deceased donor recipients. Table 3.6 presents evidence that temporal breaks are as good as randomly assigned for this subsample. I regressed each covariate on the set of temporal break dummy variables and the set

Table 3.6: Correlation of Temporal Breaks with Covariates – Deceased Donor Recipients

Patient Demographics		Patient Clinical Information		Donor / Transplant Characteritsics	
Recipient Age	0.206	Recipient Obese	0.873	Donor Age	0.358
Recipient Female	0.521	BMI Missing	0.979	Donor: Female	0.934
Recipient Black	0.297	On Dialysis	0.518	Donor: Black	0.407
Recipient White	0.321	In Hospital	0.156	Donor: White	0.102
Recipient Hispanic	0.671	Poor Functional Status	0.813	Donor: Hispanic	0.122
Recipient Other Race	0.084	Functional Status Missing	0.601	Donor: Other Race	0.691
Primary Insurer: Medicare	0.448	Glomerular Filtration Rate (GFR) at Transplant	0.387	HLA Mismatches	0.767
Primary Insurer: Medicaid	0.078	Stage 5 Kidney Disease	0.722	Days Waiting	0.989
Primary Insurer: Private	0.937	GFR / Kidney Disease Stage Missing	0.575	Donor: Expanded Criteria	0.129
Primary Insurer: Other	0.300	Recipient Sensitized (PRA > 19%)	0.151	Donor: Local to Recipient	0.096
High School or Less	0.759	PRA Value Missing	0.087		
Education Status Missing	0.237	Primary Diagnosis: Diabetes	0.641		
College or Higher	0.823	Primary Diagnosis: Hypertensive Nephroscelerosis	0.799		
		Primary Diagnosis: Glomerular Disease	0.440		
		Primary Diagnosis: Polycystic Kidneys	0.360		
		Primary Diagnosis: Other	0.121		
		Primary Diagnosis: Missing	0.168		
		Previous Kidney Transplant	0.124		

Notes:

1. Table displays the p-values to test the hypothesis that each covariate is correlated with the set of temporal break dummy variables. In particular, I estimated 41 separate regressions. For each, I regressed a covariate on the set of temporal break dummy variables and the set of transplant center-by-year fixed effects. The reported p-value corresponds to the F-statistic from each regression to test the hypothesis that the coefficients on the temporal break dummy variables are jointly equal to zero.
2. The regressions were estimated using the subsample of deceased donor transplant recipients (n = 77,374)

of transplant center-by-year fixed effects. The reported p-values in table 3.6 correspond to the F-statistic of each of these regressions. It appears that the temporal break dummy variables jointly have no statistical correlation with the majority of the covariates. Therefore, among the subsample of deceased donor recipients, temporal breaks appear to be as good as randomly assigned, at least with respect to observable patient, donor, and transplant characteristics. This finding is not surprising given the institutional setting: the arrivals of deceased donor kidneys from the local donor pool cannot be predicted and this randomness generates plausibly exogenous variation in the size of the temporal breaks experienced by the recipients of deceased donor kidneys.

Table 3.7 presents both the results of the baseline model (as in equation 3.2) and the interaction model (equation 3.3). Note that because there does not appear to be clustering of the

Table 3.7: Analysis Using Subsample of Deceased Donor Recipients

	Baseline Model			Interaction Model		
	One Week	One Month	One Year	One Week	One Month	One Year
Temporal Break						
< 4 days	-0.00103	-0.00372	-0.00610	0.01212*	0.00983	0.02277*
	(0.002)	(0.003)	(0.004)	(0.007)	(0.009)	(0.013)
5 - 7 Days	0.00188	0.00081	0.00173	0.01243*	0.01451	0.03715**
	(0.002)	(0.003)	(0.005)	(0.007)	(0.009)	(0.015)
8 - 14 Days	0.00187	0.00083	-0.00653	0.00392	0.00911	0.01047
	(0.003)	(0.003)	(0.005)	(0.007)	(0.009)	(0.014)
15 - 21 Days	-0.00019	-0.00798*	-0.00893	0.00375	-0.00400	0.00861
	(0.003)	(0.004)	(0.007)	(0.007)	(0.010)	(0.015)
22 + Days	0.00087	-0.00144	-0.00345	0.00789	0.01175	0.01830
	(0.004)	(0.006)	(0.009)	(0.008)	(0.010)	(0.015)
(log) 30 Day Transplant Volume	0.00001	0.00059	0.00135	0.00358	0.00555	0.01255**
	(0.001)	(0.002)	(0.003)	(0.003)	(0.004)	(0.006)
Interaction Terms						
< 4 Days x Volume				-0.00626**	-0.00653*	-0.01398**
				(0.003)	(0.004)	(0.006)
5 - 7 Days x Volume				-0.00528	-0.00689	-0.01803**
				(0.004)	(0.004)	(0.007)
8 - 14 Days x Volume				0.00026	-0.00359	-0.00696
				(0.004)	(0.005)	(0.007)
15 - 21 Days x Volume				0.00027	0.00155	-0.00531
				(0.005)	(0.006)	(0.010)
22 + Days x Volume				-0.00244	-0.01095	-0.00820
				(0.008)	(0.010)	(0.015)
Observations	77,374	77,374	77,374	77,374	77,374	77,374
R-squared	0.043	0.047	0.067	0.044	0.047	0.067

Notes:

1. The temporal break refers to the number of days that have elaspsed since the transplant center's last transplant. The reference temporal break is 4 days, which is the median temporal break in the sample of deceased donor recipients

2. Each regression controls for patient demographics, patient clinical information, and donor/transplant characteristics, as well as the set of transplant center-by-year fixed effects.

3. Recent transplant center experience is captured by the natural log of transplant center volume in the 30 days preceding each patient's transplant.

4. Clustered standard errors are reported in parentheses (clustered at transplant center-year level).

5. *** p<0.01, ** p<0.05, * p<0.1

data at particular points in the distribution of temporal breaks, I use a slightly different dummy variable scheme than I employ with the full sample. As before, the median temporal break for this subsample is 4 days, and the group of deceased donor transplant recipients that are transplanted after a 4 day temporal break serve as the reference group. The results of both models are generally consistent with the findings from the full sample. For example, the results from the baseline model are statistically insignificant, and the magnitudes of the coefficients do not support the expectation that longer temporal breaks generate worse patient outcomes, as compared to the reference group. The results from the interaction model are qualitatively similar to the results generated from the full sample. The coefficients on the interactions terms are generally negative, indicating that higher levels of recent experience mitigate any loss of productivity (indicated by a higher probability of graft failure) generated by temporal breaks between transplants at the patient's transplant center. However, the only statistically significant effects manifest themselves at the lower end of the temporal break distribution.

3.8 Specification Checks

The estimation results presented here suggest that there is a statistically insignificant relationship between temporal breaks between transplants and transplant center productivity, measured by the graft failure rates of the patients transplanted after the temporal break. I confirmed these findings by using different outcomes and different functional forms of temporal breaks. In particular, I also ran the regressions with patient mortality (also measured at one week, one month, and one year post-transplant) as the outcome variable. These regressions yielded statistically insignificant effects.

In addition, I tried using various functional forms of the temporal breaks as explanatory variables. First, I used quadratic, cubic, and quartic terms of the number of days between transplants. The estimated coefficients implied the non-linear effect suggested by the dummy variables, but the estimated slope coefficients were statistically insignificant. These specifications also yielded a lower R-squared than the specifications presented above. Second, I used different dummy variable schemes, including using different reference groups. Each set of alternate temporal break dummy variables yielded similar results as the ones reported here.

Finally, one concern is that recent volume (measured by the log of 30 day volume) and the set of transplant center-by-year fixed effects account for a large percentage of the variation in the size of temporal breaks. If this is the case, then the temporal break dummy variables may be closely collinear with both recent volume and the set of fixed effects, which would generate inflated standard errors. As a diagnostic check, I regressed the continuous measure of temporal breaks (the calculated number of days since the last transplant) on the measure of recent volume and the set of fixed effects. The R-squared from this regression is .364, which suggests that collinearity is not likely the explanation for the statistically insignificant relationship between temporal breaks and patient outcomes.

3.9 Discussion

The estimation results suggest that statistically there is little impact of temporal breaks on transplant center productivity, as measured by graft failure rates. In addition, the implied gradient of the effects are counterintuitive; conditional graft failure rates after relatively long temporal breaks (like 22 or more days) are often lower than the graft failure rates after shorter temporal breaks (like 5 or 6 days). These results (or “non-results”) deserve some consideration. First, there

may truly be no effect of temporal breaks on kidney transplant centers and their staff, and the estimated coefficients reflect this non-relationship. This is certainly a possibility since kidney transplantation is not a new procedure, and the surgical process itself has changed very little since 1951 (Humar and Matas 2005). In this case, the procedure may be so "familiar" that the temporal breaks measured here are not sufficiently long enough to allow for human capital depreciation or organizational forgetting. Second, there may actually be effects of temporal breaks on some members of the transplantation staff (or for particular aspects of the transplantation process) that are masked by using the transplant center as primary observational unit. For example, nurses that provide post-operative care for transplant recipients may experience depreciation of skills, but these effects are "swamped" in this analysis. Third, there may be non-random sorting of surgeons or staff along quality dimensions that are correlated with temporal breaks. As discussed earlier, lower ability staff may be utilized during periods of high transplant volume. This type of sorting would make it more difficult to detect deleterious effects of increased temporal breaks, since patients at different points in the temporal break distribution are exposed to different processes of care, even within the same transplant center in the same year. Finally, I cannot observe the activities of transplant staff members during longer temporal breaks. The implicit assumption of this analysis is that longer temporal breaks between transplants lead to organizational forgetting. But longer temporal breaks also allow the transplant staff to undertake training or otherwise productive activities that slow or negate skill depreciation. For example, staff members may attend professional conferences or simply interact with other transplant personnel which would lead to human capital appreciation. This kind of behavior may explain why conditional rates of graft failure were not as high as expected after relatively long temporal breaks (e.g. breaks of at least 22 days).

One of practical challenges to identifying the impact of temporal breaks at an organizational level (i.e. at the level of the transplant center) is that many surgical procedures, like transplantation, are team-oriented. Temporal breaks may have heterogeneous effects on different members of the transplant team, but these effects cannot be identified without knowing the composition of each team involved with each transplant. To my knowledge, data at this level of detail is rare, let alone for the specific case of kidney transplantation. Even if this data were available, it would not inform the researcher about the activities of the transplantation team during temporal breaks. This issue remains as a challenge for future research in this area.

3.10 Conclusion

This chapter addresses the question of whether temporal breaks in the production schedule of kidney transplant centers lead to lower productivity of the transplant center after a given break. The empirical evidence suggests that there is not a strong relationship between temporal breaks and patient outcomes. There is some evidence that higher levels of experience prior to a temporal break – measured by recent transplant volume – help mitigate any depreciation of skill or organizational forgetting that may arise from the breaks, but this relationship is also imprecisely estimated. The institutional setting lends credibility to the research design: the random arrival of transplantable kidneys at a transplant center generates plausibly exogenous variation in the number of elapsed days between transplants. One challenge, as describes above, is that living donor transplants appear to be explicitly scheduled for particular days of the week, which reduces some of the randomness (and variation) of temporal breaks. However, even using the subsample of deceased donor transplant recipients, I was unable to detect a statistically significant relationship between temporal breaks and transplant center

productivity, as measured by patient outcomes. In addition, the gradient implied by the estimated coefficients was unexpected and counterintuitive.

A key difference between this study and previous work by Hockenberry, Lien, and Chou (2008) and Brachet and David (2009) is that due to data limitations, I cannot address the impact of temporal breaks on individual worker (surgeon) productivity. Therefore, the findings of this chapter do not necessarily mean that there does not exist skill depreciation during temporal breaks for members of transplant teams, but rather that I cannot detect those effects at the organizational (transplant center) level. It is important to keep in mind, however, that even if I had surgeon level data, the fixed effects estimation cannot address the issue that I cannot observe surgeon activities between transplants. If surgeons, or other members of the transplant staff, are involved in other activities that increase human capital between transplants, then temporal breaks between kidney transplants may not have any detectable effect on surgeon or staff performance. That is, a temporal break between transplants does not necessarily indicate a temporal break in professional activities. This issue warrants consideration for future research in this area.

REFERENCES

2009 Annual Report of the U.S. Organ Procurement and Transplantation Network and the Scientific Registry of Transplant Recipients: Transplant Data 1999-2008. U.S. Department of Health and Human Services, Health Resources and Services Administration, Healthcare Systems Bureau, Division of Transplantation, Rockville, MD.

Al-Akash, S.I., R.B. Ettenger. 2005. Kidney transplantation in children. In *Handbook of Kidney Transplantation, 4th Edition*, edited by G. M. Danovitch. Lippincott Williams & Wilkins.

Argote, L., S.L. Beckman, and D. Epple. 1990. The persistence and transfer of learning in industrial settings. *Management Science* 36(2) 140-154.

Axelrod, D.A., M.K. Guidinger, K.P. McCullough, A.B. Leichtman, J.D. Puncha, and R.M. Meriona. 2004. Association of center volume with outcome after liver and kidney transplantation *American Journal of Transplantation* 4(6) 920-927.

Bailey, C.D. 1989. Forgetting and the learning curve: a laboratory study. *Management Science* 35(3) 340-352.

Benkard, C.L. 2000. Learning and forgetting: the dynamics of aircraft production. *American Economic Review* 90(4) 1034-1054.

Besley, T. and S. Coate. 1991. Public provision of private goods and the redistribution of income. *American Economic Review* 81(4) 979-984.

Besley, T., J. Hall, I. Preston. 1999. The demand for private health insurance: do waiting lists matter? *Journal of Public Economics* 72(2) 155-181.

Brachet, T. and G. David. 2009. On the determinants of organizational forgetting. A*merican Economic Journal: Microeconomics.* Revise and resubmit. Draft available here: http://works.bepress.com/tbrachet/5/.

Chowdhury, M.M., H. Dagash, and A. Pierro. 2007. A systematic review of the impact of volume of surgery and specialization on patient outcome. *British Journal of Surgery* 94(2) 145-161.

Cullen, J.B. and J. Gruber. 2000. Does unemployment insurance crowd out spousal labor supply? *Journal of Labor Economics* 18(3) 546-572.

Cutler, D.M. and J. Gruber. 1996. Does public insurance crowd out private insurance? *The Quarterly Journal of Economics* 111(2) 391-430.

Danovitch, G.M. 2001. Immunosuppressive medications for renal transplantation: a multiple choice question. *Kidney International* 59(1) 388-402.

Davis, C.L. 2009. Living kidney donors: current state of affairs. *Advances in Chronic Kidney Disease* 16(4) 242-249.

Gaynor, M., H. Seider, and W.B. Vogt. 2005. The volume-outcome effect, scale economies, and learning-by-doing. *American Economic Review, Papers and Proceedings* 95(2) 243-247.

Globerson, S., N. Levin, and A. Shtub. 1989. The impact of breaks on forgetting when performing a repetitive task. *IIE Transactions* 21(4) 376-381.

Goodman, W.G., G.M. Danovitch. 2005. Options for patients with kidney failure. In *Handbook of Kidney Transplantation, 4th Edition*, edited by G. M. Danovitch. Lippincott Williams & Wilkins.

Gouveia, M. 1997. Majority rule and the public provision of a private good. *Public Choice* 93(3-4) 221-244.

Gowrisankaran, G., V. Ho, and R.J.Town. 2006.Causality, learning and forgetting in surgery. Federal Trade Commission, draft(January 2006 version).

Gruber, J. and K. Simon. 2008. Crowd-out ten years later: Have recent public insurance expansions crowded out private health insurance? *Journal of Health Economics* 27(2) 201-217.

Halm, E.A., C. Lee, and M.R. Chassin. 2002. Is volume associate with outcome in health care? A systematic review and methodologic critique of the literature. *Annals of Internal Medicine* 137(6) 511-520.

Hamilton, B.H. and V.H. Hamilton. 1997. Estimating surgical volume-outcome relationships applying survival models: accounting for frailty and hospital fixed effects. *Health Economics* 6(4) 383-395.

Hockenberry, J., H. Lien, and S. Chou. 2008. The impacts of task repetition and temporal breaks in production on human capital and productivity. *Journal of Human Capital* 2(3) 303-335.

Ho, V. 2002. Learning and the evolution of medical technologies: the diffusion of coronary angioplasty. *Journal of Health Economics* 21(5) 873-885.

Howard, D.H. Forthcoming. Waiting time as a price for deceased donor kidneys. *Contemporary Economic Policy*.

Howard, D.H. 2008. Hospital quality and selective contracting: evidence from kidney transplantation. *Forum for Health Economics and Policy* 11(2).

Huesch, M.D. 2009. Learning by doing, scale effects, or neither? Cardiac surgeons after residency. *Health Services Research* 44(6) 1960-1982.

Humar, A. and A.J. Matas. 2005. Surgical complications after kidney transplantation. *Seminars in Dialysis* 18(6) 505-510.

Knoll, G.A. and G. Nichol. 2003. Dialysis, kidney transplantation, or pancreas transplantation for patients with diabetes mellitus and renal failure: a decision analysis of treatment options. *Journal of the American Society of Nephrology*14(2) 500-515.

Lee, D.S. and T. Lemieux. Regression discontinuity designs in economics. Working Paper Number 14723, NBER, Cambridge, MA.

Levey, A.S., J.Coresh, T. Greene, L.A. Stevens, Y. Zhang, S. Hendriksen, J.W. Kusek, and F.VanLente. 2006. Using standardized serum creatinine values in the Modification of Diet in Renal Disease Study equation for estimating glomerular filtration rate. *Annals of Internal Medicine* 145(4) 247-254.

Luft, H.S., S.S. Hunt, and S.C. Maerki. 1987. The volume-outcome relationship: practice makes perfect or selective referral patterns?*Health Services Research* 22(2) 157-182.

Mincer, J. and H. Ofek. 1982. Interrupted work careers: depreciation and restoration of human capital. *Journal of Human Resources* 17(1) 3-24.

Mullins, C.D., S.K. Thomas, F.G. Pradel, S.T. Bartlett. 2003. The economic impact of laparoscopic living-donor nephrectomy on kidney transplantation. *Transplantation* 75(9) 1505-1512.

Perico, N., D. Cattaneo, M.H. Sayegh, and G. Remuzzi. 2004. Delayed graft function in kidney transplantation. *Lancet* 364(9447) 1814-1827.

Pradel, F.G., M.R. Limcangco, C.D. Mullins, S.T Bartlett. 2003. Patients' attitudes about living donor transplantation and living donor nephrectomy. *American Journal of Kidney Diseases* 41(4) 849-858.

Thompson, P. 2001. How much did the Liberty Shipbuilders learn? New evidence for an old case study. *The Journal of Political Economy* 109(1) 103-137.

Tsai, A.C., M. Votruba, J.F.P.Bridges, and R.D. Cebul. 2006. Overcoming bias in estimating the volume-outcome relationship. *Health Services Research* 41(1) 252-264.

U.S. Renal Data System, USRDS 2010 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2010

Yang, R.C., H. Thiessen-Philbrook, S. Klarenbach, S. Vlaicu, A.X. Garg. 2007. Insurability of living organ donors: a systematic review. *American Journal of Transplantation* 7(6) 1542-1551.

Young, A., M. Karpinski, D. Treleaven, A. Waterman, C.R. Parikh, H. Thiessen-Philbrook, R.C. Yang, A.X. Garg. 2008. Differences in tolerance for health risk to the living donor among potential donors, recipients, and transplant professionals. *Kidney International* 73(10) 1159-1166.

Zimmerman, D., S. Donnelly, J. Miller, D. Stewart, S.E. Albert. 2000. Gender Disparity in Living Renal Transplant Donation. *American Journal of Kidney Diseases* 36(3) 534-540