Institute for Fiscal Studies

Herding in Quality Assessment: An Application to Organ Transplantation

IFS Working Paper W20/22

Stephanie De Mel
Kaivan Munshi
Soenje Reiche
Hamid Sabourian

Economic and Social Research Council

Herding in Quality Assessment: An Application to Organ Transplantation*

Stephanie De Mel† Kaivan Munshi‡ Soenje Reiche§ Hamid Sabourian¶

June 2, 2020

Abstract

There are many economic environments in which an object is offered sequentially to prospective buyers. It is often observed that once the object for sale is turned down by one or more agents, those that follow do the same. One explanation that has been proposed for this phenomenon, which goes back to Banerjee (1992) and Bikhchandani et al. (1992) is that agents making choices further down the line rationally ignore their own assessment of the object's quality and herd behind their predecessors. We develop novel tests to detect information-based herding, based on heterogeneity in agent ability, together with a methodology to quantify its welfare consequences, that are applied to organ transplantation in the U.K. We find that herding is common and is an important contributor to the high rate at which organs are rejected by transplant centers (and subsequently discarded). However, herding does not raise discard rates much above the level that would be obtained if private assessments were made publicly available. Instead, the (limited) information contained in predecessors' decisions substantially reduces the acceptance of bad organs. This is because in our application (i) high ability centers are often willing to deviate from the herd and follow their own positive signals, and (ii) sequences are exogenously terminated relatively quickly.

Keywords. Social learning. Herd behavior. Information Cascades. Organ transplant decisions.
JEL. J12. J16. D31. I3.

*We are grateful to Kourosh Saeb-Parsy and Patrick Trotter for their collaboration on this project and to numerous seminar participants for their constructive comments. Research support from the Keynes Fund, Cambridge-INET, and the Agence Nationale de la Recherche (ANR) under the EUR Project ANR-17-EURE-0010 is gratefully acknowledged. We are responsible for any errors that may remain.

†Institute for Fiscal Studies
‡Yale University
§Yale University
¶University of Cambridge

1 Introduction

There are many economic environments in which prospective buyers, acting sequentially, must choose whether or not to acquire an object. Examples of such environments include venture capital and property development, where a startup or a piece of land is offered for sale, and the labor market, notably the draft in professional sports leagues and the academic job market. It is often observed in these environments that once the object for sale is turned down by one or more agents, those that follow do the same. One explanation for this correlation in decisions is that the object is (correctly) assessed to be of poor quality by all agents. An alternative explanation, which goes back to seminal contributions by Banerjee (1992) and Bikhchandani et al. (1992), is that agents who must make choices further down the line (rationally) ignore their own assessment of the object's quality and herd behind their predecessors. The statistical identification of information–based herding is a challenging problem. In this paper, we develop novel strategies to detect herding in settings where decisions are characterized by a sequence of rejections followed by either an acceptance or termination of trade, together with a methodology to quantify its welfare consequences.

We apply our methodology to organ transplantation in the United Kingdom. The organ transplant program in the U.K. is organized around a nationwide network of centers (hospitals). When a deceased donor organ becomes available, all patients on the National Transplant Registry are assigned a priority rank based on a predetermined allocation algorithm. Transplant centers are offered the organ in order of their patients' priority, until the organ either is accepted or, having deteriorated with time, is no longer viable and is discarded by the National Health Service Blood and Transplantation (NHSBT). We will see that the organization of the U.K. transplant program makes it an ideal test-case for our methodology. This is also a setting in which information-based herding may have large practical consequences. Currently, demand outstrips the availability of both livers and kidneys, the two organs that dominate transplantation activity in the U.K. and that constitute the focus of our analysis. NHSBT statistics indicate that, five years after being listed, approximately 20% of patients on the National Registry have either died or been removed from the waiting list as their condition has deteriorated below the minimum eligibility criterion for transplantation. Simultaneously, however, 40% of livers and 22% of kidneys are discarded. Our methodology allows us to quantify the contribution of information-based herding to this relatively high discard rate.

To understand why transplant centers might rationally condition their decisions on those of the centers that preceded them, and why such herding behavior could generate inefficiencies, suppose that there are two types of organs: good (G) and bad (B). The optimal decision is to accept a good organ and reject a bad organ. Each transplant center makes an assessment of the quality of the organ that is made available to it. This assessment, which we characterize as an information signal, is not directly observed by other transplant centers. As is common in the literature on herd behaviour; e.g. Bikhchandani et al. (1992), Anderson and Holt (1997), and other references cited in Chamley (2004), we assume that signals are binary: good (g) and bad (b). Centers are not systematically misinformed;

they are thus more likely to receive a g (b) signal when an organ is good (bad). For expositional convenience, we assume that transplant centers have a common negative prior about the quality of offered organs; thus, in the absence of any other information, each center's decision is to reject the organ. Additionally, we assume that, if a center receives a g signal, this dominates its negative prior and it will accept the offered organ (in the absence of any other information).

It follows that the first center to be offered an organ will reject it on receipt of a b signal (which reinforces its prior), but will accept it on receipt of a g signal. The second center in line is only offered the organ if the first center rejects. It knows that the first center only rejects following a b signal. Thus, if the second center also receives a b signal, this reinforces the information contained in both the first center's b signal and the negative prior, and it will certainly reject the organ. If, however, the second center receives a g signal, it knows that its signal is not aligned with that of the first center. Assuming, for the purpose of this example, that all transplant centers receive signals of equal precision; i.e. that they are all equally competent in assessing the quality of an organ, the first and second centers' signals cancel each other out, and the second center also rejects (based on its negative prior). Next, consider the third center's decision: it knows that the second center rejects the organ regardless of it's signal, so the second center's decision gives the third center no additional information. Accordingly, the third center behaves as if it were second in line. Following the preceding argument, it also rejects the organ, regardless of the signal it receives. This process is repeated along the entire waiting list, regardless of the sequence of signals received by centers.[1]

While it is individually rational for centers to ignore their signals in the manner described above, herding can result in the under-utilization of viable organs. To see why this is the case, suppose that the first center in line for an organ receives a b signal, but all the centers that follow receive g signals. This organ is rejected by all centers despite the high likelihood that it is a G organ. Banerjee (1992) and Bikhchandani et al. (1992) construct similar examples of this particular type of herding, which is referred to in the literature as an "information cascade". For the purpose of this paper, herding arises when an agent follows his predecessor's decision and his successor cannot *fully* recover the signal he received. An information cascade is an extreme type of herding in which agents completely ignore their own signals when they follow their predecessors; agents further down the line learn nothing from their decisions. In our model, described below, there are two actions, accept or reject, two types of organs (states), G and B, and two signals, g and b. Herding and information cascades are synonymous in this setting and we will use these terms interchangeably in the discussion that follows.[2]

[1] If centers have a positive prior about organ quality, then it is straightforward to verify that the first two centers will follow their signals. If they both reject, then all centers that follow will do the same.

[2] When the distribution of signals is unbounded, agents always place some weight on their own signals; i.e. information cascades are ruled out (Smith and Sørensen, 2000). With a continuous distribution and bounded support, herding and information cascades can co-exist. Either way, there is always an information loss when decisions are binary because predecessors' signals cannot be recovered perfectly (Vives, 1996). Unlike Çelen and Kariv (2004) who design a lab experiment to distinguish between herding and information cascades, our objective in this analysis is to simply establish that there is a specific type of information inefficiency, highlighted in the canonical models of herding, and our tests suffice for this purpose.

Over the past three decades, the literature on information-based herding has advanced on multiple fronts. Early theoretical contributions examined the robustness of information cascades to alternative signal and choice structures (see Gale (1996) and Chamley (2004) for overviews). Theoretical work on financial markets; e.g. Avery and Zemsky (1998) and Park and Sabourian (2011) showed that herding can arise even when the cost of conformity (the asset price) is increasing in the number of agents who have made the same (investment) decision. In parallel, empirical contributions in finance and development sought to establish the key implication of social learning, a general phenomenon subsuming herding but not necessarily involving information loss, which is that agents condition their decisions on their predecessors' decisions (Lakonishok et al., 1992; Grinblatt et al., 1995; Wermers, 1999; Foster and Rosenzweig, 1995; Munshi, 2004; Conley and Udry, 2010). More recently, the theoretical focus has shifted to model misspecification (Bohren, 2016; Frick et al., 2019) and learning on networks (Golub and Sadler, 2016). The empirical literature has developed sharper tests of social learning by exploiting experimentally induced variation in predecessors' decisions (Dupas, 2014) and has moved to estimating sophisticated structural models of herding (Zhang, 2010; Cipriani and Guarino, 2014). Although much progress has been made, there remains one important gap: there has been no test, in a real world setting, of the information inefficiency (loss) that lies at the heart of the early herding literature. The structural analyses cited above quantify this inefficiency, but they do not validate the restrictions they impose on underlying agent and object heterogeneity (more on this below). Our analysis is the first we are aware of to develop robust reduced form tests of the information loss associated with herding. Only after these tests are complete do we proceed to the structural analysis, which quantifies the welfare consequences of this information loss in a setting where it is likely to be of practical importance. Apart from the credibility that comes with the reduced form tests, the additional value of our two-step analysis is that it allows us to independently cross-validate the model.

When choices are sequential, one implication of herding, and social learning more generally, is that the same patient (and her associated center) are more likely to reject an organ when they are further down the line. Zhang (2010) exploits this source of variation to estimate a structural model of herding in organ transplant decisions in the U.S. However, other explanations for this observation are also available. For example, lower quality organs, which are less likely to be accepted at any position on average, travel further down the line even without herding. Furthermore, patient priority is determined, in part, by the organ-patient match. Patients lower in line will have a worse match on average and, moreover, the organ is more likely to have deteriorated with time. Zhang controls for these factors in her analysis, but such conditioning is always imperfect. We break new ground by developing tests of herding and its associated information loss that leverage variation in decisions at the *same* position. This variation arises because (i) centers differ in their ability to distinguish between good and bad organs, and (ii) the ordering of centers varies from one organ to the next depending on the priority of their patients (and not on center ability). To the best of our knowledge, this is the first empirical analysis of herding to incorporate agent ability.

An obvious implication of herding with center heterogeneity is that a center in second position will be more likely to reject the organ when it follows a higher-ability center in first position (who must necessarily have rejected). This is because the b signal received by the center in first position is more informative about the state of the world; i.e. that the organ is a bad (B) organ. Letting p_2 be the probability that center 2 rejects (conditional on being offered the organ) and q_1 be center 1's ability, this implies that α_1 is positive in the following equation:

$$p_2 = \alpha_0 + \alpha_1 q_1. \tag{1}$$

While an estimated $\alpha_1 > 0$ is consistent with herding, a special feature of our application, which involves sequential decision making with a single object being passed along, is that this result could be obtained even if centers independently follow their own signals with no regard for the decisions of their predecessors. This is because a higher-ability center is more likely to accept a good organ and less likely to accept a bad organ. Compared to a center of lower ability, it thus passes on a worse pool of organs when in first position.

Our first test of herding is thus based on an augmented version of the preceding equation specification:

$$p_2 = \alpha_0 + \alpha_1 q_1 + \alpha_2 q_2 + \alpha_3 q_1 \cdot q_2, \tag{2}$$

where q_2 is center 2's ability and our focus is on the interaction term, $q_1 \cdot q_2$. In our model, both choices and signals are binary, allowing center heterogeneity to be incorporated relatively easily; higher ability centers are more likely to receive a $g(b)$ signal with a $G(B)$ organ. Given this structure, a center can either independently follow its own signal or it can herd behind its predecessors, in which case it will reject for sure (regardless of the signal it receives). When centers follow their own signals, center 2 is more responsive to the deterioration in the organ pool that results from center 1 being of relatively high ability when it too is of higher ability. That is, we expect α_3 to be positive. When centers ignore their own signals and herd behind their predecessors, however, this effect could be reversed, resulting in a negative value for α_3. This is because lower ability centers in second position are more likely to abandon their signals and reject with certainty, especially when following higher-ability centers. A notable feature of our test is that we are able to derive precise (observable) conditions under which α_3 is negative if herding is present.

Our second test of herding is based on the behavior of centers in third position. To implement this test, we estimate the following equation, with the probability of rejecting an organ in third position, p_3, as the dependent variable:

$$p_3 = \lambda_0 + \lambda_1 q_1 + \lambda_2 q_2. \tag{3}$$

Consider the case in which centers follow their own signals. As above, a higher-ability center passes on a relatively worse pool of organs to following centers. Thus, λ_1, λ_2 are both positive. In addition, λ_1 and λ_2 are equal to each other when we restrict attention to organs for which q_1 equals q_2. However,

once we introduce the possibility of herding (this can happen in second, but not first, position), λ_2 is strictly smaller than λ_1. This is because centers that herd (and reject with certainty, regardless of their signal) do not alter the quality of the organ pool and, thus, the rejection probability of center 3.

We implement the tests of herding with administrative data obtained from NHSBT. These data cover the universe of deceased-donor livers and kidneys offered between 2006 and 2015 in the United Kingdom. The data includes the sequence of centers that were offered each organ, as well as their decisions, which – with the possible exception of the final center in every sequence – must necessarily be rejections. The first step in estimating equations (2) and (3) is to construct a measure of center ability. Higher-ability centers are better than their low-ability counterparts at detecting both G and B organs. Thus, when the pool of organs is poor, higher-ability centers reject more often than do lower-ability centers in first position (where they are always following their own signals). Conversely, when the pool of organs has high average quality, higher-ability centers accept more often. We use this intuition to construct measures of ability for livers and kidneys, based on each center's rejection rate in first position. Using these measures, we independently detect herding with both tests.

Having established that herding is present, the next step is to estimate its prevalence and to quantify its welfare consequences. Accordingly, we estimate the structural parameters of the model and then conduct counter-factual simulations. Heterogeneity in center ability is a distinguishing feature of our model; while this helps us identify herding, it also makes the structural estimation more challenging. Our model has two ability parameters – the probability of receiving a g signal with a G organ and the probability of receiving a b signal with a B organ – that must be estimated separately for each center. Given the large number of centers, we estimate these ability parameters outside the model. Indices of organ quality, which are predictive of transplant success, have recently been proposed in the organ transplant literature for both livers and kidneys. These "risk indices" are constructed from retrospective NHSBT data that covers the universe of organ transplant decisions and subsequent patient outcomes over many years; the set of organ characteristics included in these indices, together with the (estimated) weights placed on these characteristics, will thus predict transplant success with a relatively high degree of precision. Individual centers, in contrast, base their decisions on their own (limited) experiences and the organ-specific information that was received at the time of decision-making. We nevertheless expect quality assessment and the associated decisions taken by higher-ability centers to track more closely with the risk index. We use this association to construct center-specific measures of ability, to characterize the organ-center specific distribution of signals, and to compute the fraction of G organs in the population of organs.

The risk indices allow us to obtain internally consistent estimates of all the model's parameters, with one exception: the threshold belief that an organ is good, above which centers accept it. An increase in this threshold has no bearing on the decisions of centers that follow their own signal; they will accept if they receive a g signal and reject if they receive a b signal. It does, however, increase the fraction of centers that herd and ignore their own signal, rejecting with certainty. We estimate

this parameter using the simulated method of moments (based on repeated draws of the information signals) by matching rejection rates predicted by the model to the data. Although the estimation is computationally straightforward, the model places strong restrictions on the parameter values that can be obtained. We verify that the estimated threshold satisfies a key assumption of the model, which is that centers follow their own signals in first position. We also verify that the ability parameters estimated from the risk indices correlate closely with the independently derived ability measure used in the reduced form tests. Finally, we verify that the model does a good job of matching the data; indeed, our model's goodness of fit is substantially better than that of the alternative "no-learning" model, in which centers ignore their predecessors' decisions and always follow their signals.

To measure the prevalence of herding, we compute the fraction of decisions in our data for which centers are predicted to have ignored their signals (given our estimate of the threshold belief). Based on this estimate, it is common for centers to ignore their signals: this occurs 42% of the time for livers and 32% of the time for kidneys, with an increase in these statistics at higher positions. However, a more important question is whether such herding has substantial welfare consequences. To answer this question, it is necessary to specify a benchmark, which we define as the counter-factual outcome in which information signals are pooled (common knowledge). In our model, centers know which of their predecessors followed their own signals and, consequently, must have received a b signal when they rejected. The missing information is associated with centers who herd, as the centers that follow them learn nothing from their decisions. With pooled information, by contrast, signals received by all preceding centers are utilized in the decision-making process. We construct the pooled information benchmark by drawing signals for those centers who are predicted by the model to herd (this is possible because the risk index provides us with more information than was historically available to individual centers). The signals we draw are also used to predict decisions in the alternative no-learning model, in which centers always follow their own signals.

Herding, as in the example above, is associated with false rejections, relative to the pooled information benchmark. This is because g signals are ignored by centers who herd, and this has spillover effects on the decisions of centers that follow. Conversely, no-learning is associated with false acceptances, relative to the same benchmark, because useful information contained in previous rejection decisions is ignored. Our estimates quantify these opposing effects. We find that the rate of false rejections with the herding model is modest. This is because centers are heterogeneous in their abilities and high ability centers are often willing to deviate from the herd and follow their g signals. In addition, while false rejections tend to be concentrated at higher positions, sequences are terminated relatively quickly by NHSBT. In contrast, the false acceptance rate with the no-learning model is substantial. These findings motivate the final step of the analysis, in which we compare discard rates under the herding and no-learning models against the pooled information benchmark. The decision to discard an organ is taken by NHSBT and is outside our model. Our interest is in whether an organ that is discarded in the data (with herding) would have been accepted at an earlier position with the

alternative models. As decisions are similar under herding and the pooled information benchmark, we do not expect discard rates to diverge substantially. As expected, discard rates with herding are roughly 10% higher than the pooled information benchmark. In contrast, discard rates under the assumption of no learning are 35% lower than the same benchmark, on account of the many false acceptances that arise when centers ignore the information that is contained in their predecessors' decisions. These results collectively indicate that herding contributes substantially to the high discard rate observed in the data, but that this increase in the discard rate does not result in significant information loss (inefficiency) relative to the pooled information model. Centers often ignore their own signals, but their reliance on their predecessors actually protects them from accepting bad organs. In other environments, however, where agents are homogeneous or where queue lengths are longer (as with organ transplantation in the U.S.), the standard inefficiencies associated with herding will likely be more substantial.

2 Institutional Setting

The shortage of suitable donor organs has always been the primary challenge faced by organ transplant programs. In response to this challenge, many countries, including the United Kingdom, have established national allocation schemes for the distribution of organs supplied by deceased donors. Organs obtained from deceased donors are classified according to the manner of death as either DBD (donation after brain death) or DCD (donation after cardiac death). Although DBD and DCD organs do not vary systematically with respect to *ex ante* quality and the same broad allocation protocols are utilized by the National Health Service Blood and Transplant (NHSBT) for both types of organs, DCD organs are useable for a shorter period of time before they must be discarded from the donor pool and set aside for research (Watson and Dark, 2012).

Our analysis focuses on livers and kidneys, for which donors have been matched to recipients in the United Kingdom through a national allocation scheme since the late 1990s. These two types of organs continue to dominate transplantation activity: NHSBT statistics indicate that over 80% of livers and kidneys obtained from DBD donors in 2014-2015 were transplanted, while the corresponding statistics for pancreases, hearts, and lungs were less than 35%. For DBD livers and kidneys, a Transplant Benefit Score (TBS), which puts weight on both the patient's need for a transplant and the patient's organ-specific quality of life after the transplant, is used to rank all patients listed on the National Registry when a given organ becomes available. The TBS is calculated using a fixed set of donor and recipient characteristics. Transplantation delays are substantially more costly for DCD organs and, hence, the proximity between donor and recipient is also a factor in drawing up the priority list for them.

When an organ becomes available, it is offered to patients in order of their priority. Each patient's hospital (transplant center) has 45 minutes to accept or decline the offer. The attending surgeon has an open-ended conversation with the NHSBT administrator about the characteristics of the organ

and the donor, as well as other factors that are relevant for that particular case, before arriving at a decision. This sequential process continues until the organ has been accepted or until too much time has elapsed for it to remain useable. For livers, DBD (DCD) organs should be transplanted within 12 (6) hours, while the corresponding cutoffs for kidneys are 18 (12) hours. This is a narrow time window, leaving room for just a few centers to make decisions before an organ is discarded by NHSBT. We will see below that queue lengths for organs rarely exceed eight centers.

The allocation of deceased-donor organs in the United Kingdom differs in important respects from the allocation procedure in the United States. Zhang (2010), using data on kidney donations in Texas, documents that on average an organ is accepted by the 34^{th} patient in line, who has already turned down 15 offers. Such long queue lengths are possible because organs are only discarded after 48 hours. Under these circumstances, the condition of the organ becomes a major consideration in decision-making, particularly at higher positions. The mismatch between organ and recipient also becomes relevant (in Zhang's data, kidneys are accepted as late as the 77^{th} position). Given this mismatch, patients consider (and reject) many organs before finally accepting and, hence, dynamic considerations enter the decision rule. Both Zhang (2010) and Agarwal et al. (2019), who also study kidney allocation in the United States, model the acceptance decision as an optimal stopping problem. The institutional environment in the United Kingdom, where organs are almost always accepted by patients towards the very top of the national priority list and where mismatch, deterioration, and the associated strategic inter-temporal considerations are thus less relevant, allows us to ignore these factors in our analysis and focus on a new aspect of decision-making, which is the ability of centers to correctly assess the quality of the organs they are offered.

Table 1 provides direct evidence that the patient-organ mismatch and organ deterioration, which are both necessarily increasing with a center's position in the queue for a given organ, are less relevant in the United Kingdom. This table reports the relationship between the most stringent (conventional) measure of transplant success – whether the organ survives at least three years – and the position in the queue of the transplanting center, for all livers and kidneys that were transplanted between 2006 and 2015.[3] If mismatch and deterioration are relevant, then organs transplanted at higher positions will have worse outcomes. Because lower quality organs will travel further down the line on average, we include a recently constructed risk index of organ quality (described in greater detail below) in the estimating equation. As observed in Table 1, the probability of transplant success is (not surprisingly) declining significantly in the risk index. Conditional on the risk index, the transplanting center's position in the queue has a negligible effect on transplant success.[4] This indicates that neither mismatch nor deterioration (at higher positions) are relevant in this setting. One alternative

[3]A new National Kidney Allocation Scheme was initiated in 2006 and a new National Liver Allocation Scheme was initiated in 2015. The analysis thus covers a period during which both livers and kidneys were allocated in a uniform manner.

[4]The estimates reported in Table 1 are based on the linear probability model because the marginal effects are easy to interpret. Average marginal effects with the probit model are almost identical to the estimates reported in the table. The estimated marginal effects imply that a two standard deviation increase in center position would reduce the survival probability for both livers and kidneys by 0.02.

Table 1: Determinants of Transplant Success

Dependent variable:	organ survives for at least three years	
Organ:	liver	kidney
	(1)	(2)
Organ risk index	-0.0463***	-0.130***
	(0.011)	(0.011)
Center ability	-0.00489	-0.0155
	(0.082)	(0.032)
Position in queue	-0.00644	-0.00471*
	(0.004)	(0.003)
Mean of dependent variable	0.729	0.771
N	6243	11755

Note: heteroscedasticity-robust standard errors in parentheses
Center ability is the measure used in the herding tests
Position in queue ranges from 1 to 8
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

explanation for this result is that centers at higher positions account for the mismatch and deterioration and are more conservative when accepting organs. Based on our structural estimates, however, there is no evidence that the cutoff belief above which centers accept an organ is increasing with their position in the sequence.

While centers may differ in their ability to distinguish between good and bad organs, we assume that this particular dimension of ability is independent of their competence in implementing transplant procedures. The estimating equation in Table 1 includes the measure of center ability that we will use in the herding tests, which is based on the center's rejection rate when in first position. Although this measure will have a strong effect on decisions, we see (conditional on the risk index) that it has no impact on transplant success. This finding indicates that higher ability centers, as we define them, do not have greater competence in implementing transplants *and* that their patients do not differ with respect to fixed characteristics such as age and health condition that independently determine transplant outcomes.[5] The results in Table 1, taken together, will allow us to simplify both the model and the empirical analysis that follow.

[5]Transplant centers service precisely defined regions. Given the length of time that patients must wait for a transplant and the fact that decisions must be made and the transplant itself must be undertaken within a matter of hours, selective sorting by patients into particular centers is unlikely to be a consideration in practice. However, variation in population characteristics across regions could, in principle, generate variation in patient characteristics across centers. The results in Table 1 indicate that this is not the case.

3 A Model of Organ Transplantation

3.1 Organs, Centers and Signals

Organs can be either of good (G) or bad (B) quality. The outcome of an organ transplant is denoted by H if the organ is good, and by L if it is bad, with $H > 0 > L$. Payoffs H and L are realized independently of the center (hospital) undertaking the transplant and the identity of the patient who receives the organ. This implies that centers are equally competent in implementing the transplant procedure and that fixed patient characteristics, the organ-recipient mismatch, and organ deterioration can be ignored when modeling transplant decisions (as verified above). We normalize the outcome of not transplanting an organ to 0. Although centers do not know the quality of a particular organ with certainty, the fraction of organs that are G organs is common knowledge. The prior on organ quality, denoted by π, is thus the probability that a random organ from the pool is a G organ. We define the *cut-off* belief $\tilde{\pi}$ as the belief at which every hospital is indifferent between accepting or rejecting an organ; i.e. $\tilde{\pi}H + (1-\tilde{\pi})L = 0$, or $\tilde{\pi} = \frac{-L}{H-L}$.

Centers individually assess organ quality before making a decision. This assessment is based on each center's past experience and the organ-specific information it receives from the NHSBT administrator. More competent centers are better able to acquire salient information from the administrator and to utilize that information. We characterize each center's independent assessment of an organ by a private information signal $s \in \{g, b\}$, where a g signal indicates that the organ is good, while b indicates that it is bad. We denote a center by j and its ability by $q_j \in \left[\underline{q}, \overline{q}\right] \subset \mathbb{R}$. A center's ability determines the probability γ_j with which it correctly identifies a G organ and the probability β_j with which it correctly identifies a B organ. In particular, for any center j, $prob(g \mid G) = \gamma_j = \gamma(q_j)$ and $prob(b \mid B) = \beta_j = \beta(q_j)$, where γ, β are strictly increasing functions.

Centers are not systematically misinformed; i.e. each center is more likely to receive a b signal with a bad organ than with a good organ. This requires:

Assumption 1: *For all* j, $\beta_j \geq 1 - \gamma_j$.

This assumption will certainly be satisfied if the lowest ability center, with ability $\underline{q}$, is completely uninformed; i.e. is equally likely to receive a b (g) signal with a G organ or with a B organ. This implies $\beta(\underline{q}) = 1 - \gamma(\underline{q})$.

Centers in first position update their prior belief that an organ is good, π, upon receiving their signals. Formally, center j's updated belief that an organ is good, upon receiving signals g and b in first position is denoted by

$$\begin{aligned} \pi_j(g) &= \frac{\pi\gamma_j}{\pi\gamma_j + (1-\pi)(1-\beta_j)} \\ \pi_j(b) &= \frac{\pi(1-\gamma_j)}{\pi(1-\gamma_j) + (1-\pi)\beta_j}, \end{aligned}$$

respectively. Given Assumption 1, the belief shifts up (down) upon receipt of a $g(b)$ signal:

$$\pi_j(g) \geq \pi \geq \pi_j(b). \tag{4}$$

In addition to Assumption 1, we further assume that centers always follow their own signals in first position (absent any other information) such that each center accepts the organ if it receives a g signal and declines the organ upon receipt of a b signal. This is equivalent to the following:

Assumption 2: *For all j, $\pi_j(g) \geq \tilde{\pi} > \pi_j(b)$.*

3.2 Transplant Decisions

Organs are offered sequentially to centers on the basis of a predetermined algorithm. The priority list for a given organ is based on recipients' characteristics and their match with the organ and is independent of the centers to which they are attached. As verified below, the average quality of organs received by centers when in first position, and the subsequent ordering of centers for a given organ, will thus be independent of their ability. We also assume that fixed patient characteristics, the organ-patient mismatch, and organ deterioration can be ignored when modeling decision-making (as they were seen to have no impact on transplant outcomes).

We next describe the evolution of beliefs, and associated decisions, for different centers in line for a given organ. To simplify notation for the rest of this section and for the tests of herding in Section 4, we identify a center by its position in line, such that the center at position j has ability q_j. Center 1 receives a signal and, given Assumption 2, accepts after a g signal and declines after a b signal. If the organ is accepted, it is transplanted by center 1 and results in payoff H or L, depending on its quality. If it is declined, an administrator from NHSBT decides either to offer the organ to the next center or to set it aside for research. The decision to discard an organ is based on its condition or useability, which is determined by the delay in retrieving the organ and the subsequent delay prior to transplantation. NHSBT administrators monitor the condition of the organ during the offering process, discarding it as soon as it is considered to be unsuitable for transplantation. Because the decision to discard an organ is orthogonal to its quality, this will not affect the next center's prior on organ quality.

Centers positioned further along in the sequence learn from the (rejection) decisions of their predecessors. Each center knows the identity of its predecessors and the order in which they made their decisions. If this were not the case, then all the tests reported below would fail to be supported by the data. We use an iterative process to describe centers' equilibrium beliefs and strategies moving down the line. The equilibrium concept that characterizes the learning process is Perfect Bayesian Nash Equilibrium.

If center 2 is offered an organ, it knows that center 1 must have received a b signal, given Assumption 2. Its prior belief before it receives its private signal, which is public information, is denoted by π_2;

thus

$$\pi_2 = \pi_1(b) = \frac{\pi(1-\gamma_1)}{\pi(1-\gamma_1)+(1-\pi)\beta_1}. \tag{5}$$

Its posterior belief, upon receiving signals g and b, respectively, is given by

$$\pi_2(g) = \frac{\pi_2\gamma_2}{\pi_2\gamma_2+(1-\pi_2)(1-\beta_2)},$$

and

$$\pi_2(b) = \frac{\pi_2(1-\gamma_2)}{\pi_2(1-\gamma_2)+\pi_2\beta_2}.$$

Center 2 always rejects the organ if it receives a b signal, because its prior belief, π_2 (which is lower than $\tilde{\pi}$ from Assumption 2), is downgraded even further following a b signal. Center 2 could reject the organ even if it receives a g signal – which implies that it is herding – if this updating does not raise its posterior above $\tilde{\pi}$. To summarize, center 2's optimal decision is to accept the organ if it received a g signal and $\pi_2(g) \geq \tilde{\pi}$, and to decline otherwise.

Next, center 3 knows center 2's decision-making process and its prior belief, π_2, but does not necessarily know center 2's signal. If center 2 herds, its decision provides no information about its signal to center 3, and the latter's public belief π_3 is therefore equal to π_2. If, on the other hand, center 2 uses its signal to make its decision ($\pi_2(g) \geq \tilde{\pi}$), center 3 infers from center 2's rejection that it must have received a b signal, and therefore has a public belief equal to $\pi_2(b)$. Thus,

$$\pi_3 = \begin{cases} \pi_2, & \text{if } \pi_2(g) < \tilde{\pi} \\ \pi_2(b) & \text{otherwise.} \end{cases}$$

The preceding discussion can be easily generalized. In the same way as center 2, center $n > 3$, given its public belief (π_n), forms its posterior belief (either $\pi_n(g)$ or $\pi_n(b)$), and then chooses optimally either to accept or to decline the organ. Then, as with center 3, center $n+1$'s public belief π_{n+1} equals π_n if center n herds and $\pi_n(b)$ otherwise.

3.3 Center Heterogeneity

As discussed, the novelty of our tests of herding is that they are based on variation in center decisions at the same position. This variation arises because centers differ in their ability to distinguish between good and bad organs and because the order of centers varies from one organ to the next. The first step in deriving our tests is thus to use the model to construct a measure of center ability.

Our measure of a center's ability is based on its observed rejection rate when in first position $\overline{p}_1$. By Assumption 2, all centers follow their signals in first position. Thus, for any specific organ, the probability that center 1 with ability q_1, rejects the organ, $p_1(q_1)$, is equal to the probability that it receives a b signal:

$$p_1(q_1) = \pi(1-\gamma_1)+(1-\pi)\beta_1. \tag{6}$$

Note that this probability depends only on the probability that the organ is good π and the ability of the center q_1. Thus, it is the same for all organs that the center receives in first position. This implies that $p_1(q_1) = \overline{p}_1$, and we will thus use these terms interchangeably.

By (6), we also have that

$$\frac{dp_1(q_1)}{dq_1} = -\pi\gamma'(q_1) + (1-\pi)\beta'(q_1). \tag{7}$$

It is evident from (7) that $p_1(q_1)$ could be increasing or decreasing in q_1 because more able centers are better at detecting both good and bad organs; i.e. $\gamma'(q_1)$ and $\beta'(q_1)$ are both positive. In an inferior organ pool, with many bad organs, the $\beta'(q_1)$ term dominates and $p_1(q_1)$ is increasing in q_1. In a superior organ pool, with many good organs, the $\gamma'(q_1)$ term dominates and $p_1(q_1)$ is decreasing in q_1. We allow for both possibilities, with the restriction that the probability of rejection in first position, for a given organ pool, is either monotonically increasing or decreasing in ability for all centers; i.e. either $\frac{dp_1(q_1)}{dq_1} > 0$ for any $q_1 \in [\underline{q}, \overline{q}]$ or $\frac{dp_1(q_1)}{dq_1} < 0$ for any $q_1 \in [\underline{q}, \overline{q}]$.

To determine whether the probability of rejection in first position is increasing or decreasing in center ability, we examine the decisions of centers in second position. To begin with, assume that center 2 follows its own signal. This would be the case if it ignores its predecessors' decisions or if it does not herd; that is, its posterior belief upon receiving a g signal exceeds $\tilde{\pi}$. In this case, the probability that center 2 rejects an organ is the probability that it receives a b signal, conditional on center 1 also having received a b signal. For a given organ type (B, G), the variation in signals received by a center reflects the mistakes that it makes in assessing organs. While higher ability centers make fewer mistakes on average, these mistakes are independent across organs for any center, and across centers for a given organ. The signals received by center 1 and center 2 will thus be independent, conditional on the type of organ. Assuming that center 2 follows its own signal, the probability that center 2 rejects an organ conditional on center 1 rejecting is given by

$$p_2(q_1, q_2) \equiv \frac{\pi(1-\gamma_1)(1-\gamma_2) + (1-\pi)\beta_1\beta_2}{\pi(1-\gamma_1) + (1-\pi)\beta_1}. \tag{8}$$

We can then compute, using Assumption 2, the manner in which center 2's rejection probability $p_2(q_1, q_2)$, which we also refer to as p_2 in the discussion that follows, varies with center 1's ability:

$$\frac{\partial p_2(q_1, q_2)}{\partial q_1} = \frac{\pi(1-\pi)(\gamma_1'\beta_1 + \beta_1'(1-\gamma_1))(\beta_2 - (1-\gamma_2))}{(\pi(1-\gamma_1) + (1-\pi)\beta_1)^2} \geq 0. \tag{9}$$

If center 2 follows its own signal, then it is more likely to reject when its predecessor has higher ability. Moreover, an increase in center 1's ability makes it more likely that center 2 herds, in which case it rejects for sure (regardless of its signal). This is because a higher-ability predecessor's rejection has a bigger impact on center 2's prior belief, thereby increasing the likelihood that its posterior belief will remain below $\tilde{\pi}$ even when it receives a g signal. In general, center 2 is more likely to reject an

organ when center 1 has high ability, regardless of whether centers learn from their predecessors or not.

Based on the preceding discussion, if we observe that centers in second position are more (less) likely to reject an organ when they follow centers with a higher rejection rate in first position, $\overline{p}_1$, then $\overline{p}_1$ or, equivalently, $p_1(q_1)$, must be positively (negatively) associated with center ability. This is true even if average organ quality in first position varies across centers; higher ability centers always pass on a worse pool of organs and, hence, the sign of the estimated $p_2-\overline{p}_1$ relationship would still allow us to infer whether $\overline{p}_1$ is increasing or decreasing in ability. To use $\overline{p}_1$ as a measure of ability, however, we require that average organ quality in first position be independent of center ability, as assumed above and verified below. We will see later that the sign of the $p_2-\overline{p}_1$ relationship is positive for livers and negative for kidneys. We will therefore use $\overline{p}_1$ to measure center ability for liver transplants and $1-\overline{p}_1$ to measure center ability for kidney transplants in our tests of herding.

3.4 A Test of Herding (based on decisions in second position)

The preceding discussion indicates that center 2 is more likely to decline an organ when it follows a higher-ability center, with and without herding. To test for herding in second position we thus look more closely at the p_2-q_1 relationship, by examining how this relationship varies with q_2. In particular, we estimate equation (2), $p_2 = \alpha_0 + \alpha_1 q_1 + \alpha_2 q_2 + \alpha_3 q_1 \cdot q_2$, and focus attention on the interaction term.

To begin with, assume that center 2 does not herd, such that its rejection probability for a given organ is described by equation (8). The cross-partial with respect to q_1 and q_2, which is essentially the coefficient on the interaction term, is then

$$\frac{\partial^2 p_2(q_1,q_2)}{\partial q_1 \partial q_2} = \frac{\pi(1-\pi)(\gamma_1'\beta_1 + \beta_1'(1-\gamma_1))(\beta_2' + \gamma_2'))}{(\pi(1-\gamma_1)+(1-\pi)\beta_1)^2} > 0, \tag{10}$$

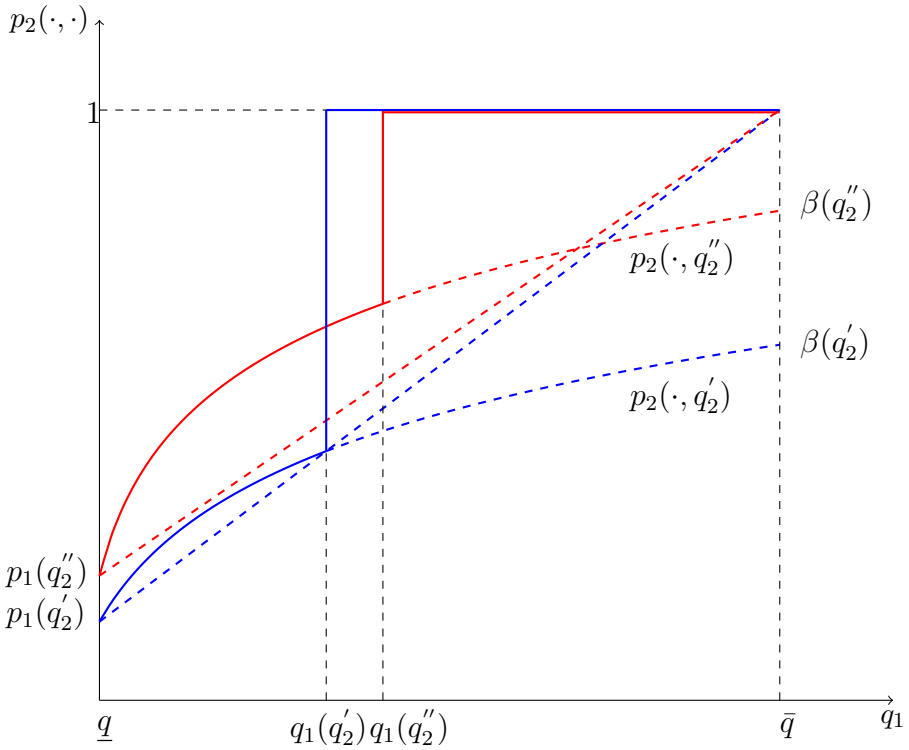
That is, the effect of an increase in center 1's ability on center 2's rejection probability is larger when center 2 has higher ability. This result is obtained because (i) an increase in q_1 reduces the quality of the organ pool passed on to center 2, i.e. the prior belief π_2 goes down and (ii) center 2's decision, p_2, is more sensitive to π_2 when it has higher ability. In the extreme case, if center 2 is completely uninformed; i.e. $\beta_2 = 1-\gamma_2$, then its decision will be unaffected by the change in the quality of the organ pool.

It follows from equation (10) that the cross partial is positive in the absence of herding. This result, however, does not necessarily hold when center 2 herds. In particular, there are now two effects: the *quality selection effect* and the *herding effect*. The former, which we described above, implies that a low-ability center reacts less to an increase in its predecessor's ability than does a high-ability center, because the former is less sensitive to the quality of its organ pool. The latter effect works in the opposite direction: as the rejection by a high-ability center 1 represents worse news about underlying

organ quality than does the rejection of a low-ability center, it is more likely that a weak center 2 herds and also rejects. Which effect dominates depends on the underlying organ pool and on the ability of the relevant centers. This is best seen in the following two figures. In drawing these figures, we have assumed that the lowest-ability center is completely uninformed and that the highest-ability center is perfectly informed, although the results do not rely on those assumptions; i.e. we assume $\beta(\underline{q}) = 1 - \gamma(\underline{q})$ and $\beta(\overline{q}) = \gamma(\overline{q}) = 1$. The figures show the rejection probabilities of two centers at position 2 as a function of center 1's ability q_1. The abilities of the two centers are q_2' and q_2''; and we assume that $q_2' < q_2''$.

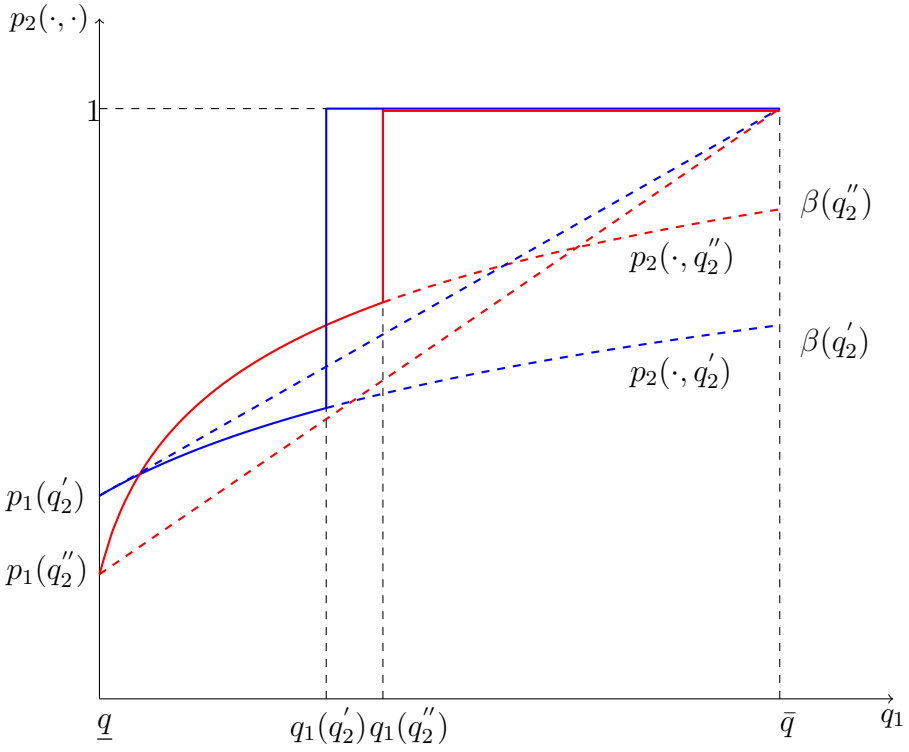
The curves labelled $p_2(\cdot, q_2')$ and $p_2(\cdot, q_2'')$ in the figures are the centers' respective rejection probabilities *without herding*, given by the expression in equation (8). If center 2 follows center 1 with $q_1 = \underline{q}$ it faces an organ pool with quality π and, consequently, draws signals as if it were in first position. This implies that $p_2(\underline{q}, q_2) = p_1(q_2)$, which pins down the intercept of each curve. Note that Figure 1 assumes that $p_1(q_2') < p_1(q_2'')$, which we later see applies to livers, while Figure 2 assumes that the inequality is reversed, which is relevant for kidneys. This represents the only difference between the two figures. If center 2 follows center 1 with $q_1 = \bar{q}$, the organ is a B organ for certain, so center 2 draws signals from a B organ. This explains why the curves reach heights of $p_2(\overline{q}, q_2') = \beta(q_2')$ and $p_2(\overline{q}, q_2'') = \beta(q_2'')$, where $\beta(q_2') < \beta(q_2'')$.

Figure 1: $\frac{dp_1(q)}{dq} > 0$



With herding the rejection probability of any center 2 with ability q_2 jumps to 1 at some threshold $q_1(q_2)$; thus, for $q_1 < q_1(q_2)$, center 2 uses its signal and rejects according to the expression in equation

Figure 2: $\frac{dp_1(q)}{dq} < 0$



(8) while, for $q_1 \geq q_1(q_2)$, center 2 herds and always rejects. Note that $q_1(q_2') < q_1(q_2'')$ as a lower-ability center positioned at 2 starts herding sooner than does a high-ability one.

We now use the figures to derive the effect of the interaction term $q_1 \cdot q_2$, on center 2's rejection probability with herding. Expression (10) allows us to obtain a measure of this effect for each pair (q_1, q_2) in the absence of herding. With herding, we cannot use the same method, because center 2's rejection probability contains a jump when center 2 starts to herd and, hence, the derivative is not well-defined at each point. Instead, for each q_2, we compute the"average effect" of an increase in q_1; this is the slope of the line starting at $(\underline{q}, p_1(q_2))$ and going to $(\bar{q}, \beta(q_2))$ if there is no herding and to $(\bar{q}, 1)$, otherwise.[6] We then examine how this slope varies with q_2 (q_2' versus q_2'').

Consistent with the cross-partial expression in (10), the average slope with respect to q_1 is increasing in q_2 in the absence of herding in both figures:

$$\frac{\beta(q_2') - p_1(q_2')}{\bar{q} - \underline{q}} < \frac{\beta(q_2'') - p_1(q_2'')}{\bar{q} - \underline{q}}$$

When we incorporate the effect of herding, however, we see that the slope with respect to q_1 in Figure 1 is decreasing in q_2:

$$\frac{1 - p_1(q_2')}{\bar{q} - \underline{q}} > \frac{1 - p_1(q_2'')}{\bar{q} - \underline{q}}.$$

[6]The implicit assumption when computing the average effect is that the distribution of q_1 is uniform on $[\underline{q}, \bar{q}]$.

In contrast, the slope with respect to q_1 in Figure 2 is increasing in q_2. Thus, when center ability is *increasing* in the probability of rejection in first position, $\left(\frac{dp_1(q)}{dq} > 0\right)$, as observed in Figure 1 and later for livers, we predict that the interaction effect is reversed with herding: lower-ability centers at position 2 react more to an increase in center 1's ability because the herding effect dominates. When center ability is *decreasing* in the probability of rejection in first position, $\left(\frac{dp_1(q)}{dq} < 0\right)$, as observed in Figure 2 and later for kidneys, we predict that the interaction effect goes in the same direction as it does in the absence of herding; the quality selection effect then dominates, such that higher-ability centers at position 2 react more to an increase in center 1's ability. We summarise the above discussion as follows:

Test 1 *Without herding, the cross-partial effect of an increase in both center 1's and center 2's ability on center 2's rejection probability, for a given organ, is strictly positive. With herding, the average cross-partial effect is strictly positive if center ability is decreasing in the probability of rejection in first position and negative if center ability is increasing in the probability of rejection in first position.*

3.5 A Test of Herding (based on decisions in third position)

Our second test of herding is based on decisions at position 3, in particular, on the relationship between these decisions and center abilities at position 1 (q_1) and position 2 (q_2), as expressed in equation (3): $p_3 = \lambda_0 + \lambda_1 q_1 + \lambda_2 q_2$. In deriving this test we assume that center 3 does not herd - centers that herd at third position always reject, and their decision is thus unaffected by marginal changes in q_1 and q_2.

To develop our second test of herding we investigate the effect of a marginal increase in q_1 and q_2 on center 3's rejection probability p_3. We first consider the case in which center 2 does not herd; then p_3 is the probability that center 3 receives a b signal, conditional on both center 1 and center 2 also having received b signals. In this case p_3 is given by

$$p_3(q_1, q_2, q_3) = \frac{\pi(1-\gamma_1)(1-\gamma_2)(1-\gamma_3) + (1-\pi)\beta_1\beta_2\beta_3}{\pi(1-\gamma_1)(1-\gamma_2) + (1-\pi)\beta_1\beta_2}. \tag{11}$$

It is easy to see that an increase in either center 1's or center 2's ability decreases the quality of the organ pool passed on to center 3, which increases the latter's rejection probability. Formally,

$$\frac{\partial p_3(q_1, q_2, q_3)}{\partial q_1} = \Theta\beta_2(1-\gamma_2)(\gamma_1'\beta_1 + \beta_1'(1-\gamma_1)), \tag{12}$$

$$\frac{\partial p_3(q_1, q_2, q_3)}{\partial q_2} = \Theta\beta_1(1-\gamma_1)(\gamma_2'\beta_2 + \beta_2'(1-\gamma_2)), \tag{13}$$

where, $\Theta = \frac{\pi(1-\pi)(\beta_3-(1-\gamma_3))}{[\pi(1-\gamma_1)(1-\gamma_2)+(1-\pi)\beta_1\beta_2]^2}$. Both expressions clearly are strictly positive, but their exact magnitudes depend on the abilities of centers 1 and 2. When $q_1 \neq q_2$, $\frac{\partial p_3(q_1,q_2,q_3)}{\partial q_1}$ could be larger or smaller than $\frac{\partial p_3(q_1,q_2,q_3)}{\partial q_2}$; for $q_1 = q_2$, however, $\frac{\partial p_3(q_1,q_2,q_3)}{\partial q_1}$ must equal $\frac{\partial p_3(q_1,q_2,q_3)}{\partial q_2}$.

Now consider a situation in which center 2 herds. Because it rejects, regardless of the signal it receives, its decision has no effect on the quality of the organ pool passed on to center 3. By contrast,

center 1 always uses its signal. Center 3's rejection probability, p_3, is thus the probability that center 3 receives a b signal, conditional only on center 1 having received a b signal:

$$p_3^h(q_1, q_2, q_3) = \frac{\pi(1-\gamma_1)(1-\gamma_3) + (1-\pi)\beta_1\beta_3}{\pi(1-\gamma_1) + (1-\pi)\beta_1}. \tag{14}$$

In this case, the effect of an increase in the ability of center 1, though different to the case without herding (see equation (12)), is still positive, while the effect of an increase in the ability of center 2 is zero:

$$\frac{\partial p_3^h(q_1, q_2, q_3)}{\partial q_1} = \Pi(\gamma_1'\beta_1 + \beta_1'(1-\gamma_1)) \tag{15}$$

$$\frac{\partial p_3^h(q_1, q_2, q_3)}{\partial q_2} = 0, \tag{16}$$

where $\Pi = \frac{\pi(1-\pi)(\beta_3-(1-\gamma_3))}{[\pi(1-\gamma_1)+(1-\pi)\beta_1]^2}$. Summarizing the preceding discussion:

Test 2 *If centers* 1 *and* 2 *have identical abilities then, without herding, the effect of an increase in center* 1*'s ability on center 3's rejection probability equals the effect of an increase in center 2's ability. With herding, the effect of an increase in center* 1*'s ability is larger than the effect of an increase in center* 2*'s ability.*

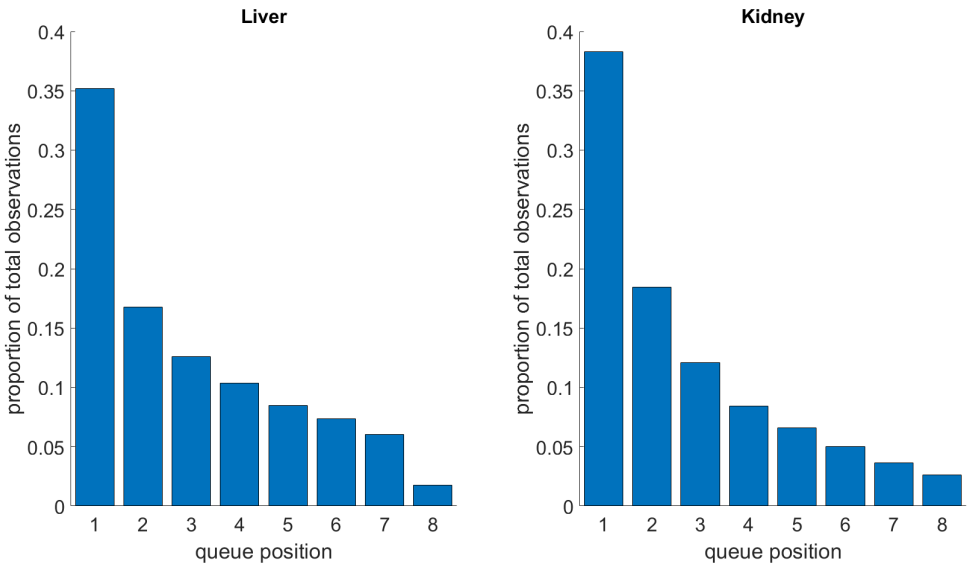
4 Testing the Model

4.1 The Data

The data that we use to test the model consists of the sequence of decisions taken by centers for each deceased-donor organ (liver or kidney) that was offered for transplantation in the 2006-2015 period. Each center that is offered an organ can either accept or reject it. If an organ is rejected, it is offered to the next center in line, unless NHSBT assesses that the condition of the organ has deteriorated to the point that it is no longer useable, in which case it is discarded, i.e. set aside for research. There are thus two possible end-points for an organ: it is accepted or it is discarded. Prior to either end-point, all decisions must necessarily be rejections.

The deterioration that results in an organ being discarded can be caused by delays in retrieving the organ (warm ischaemia) or by subsequent delays in transplantation (cold ischaemia). The U.K. is much more conservative with regard to organ deterioration than the U.S. and, hence, sequence lengths tend to be short. When sequence lengths are short, most decisions are concentrated in early positions, and this is what we observe in Figure 3. For livers, 35% of observations are in first position, with a steep decline in the fraction of decisions at higher positions. For kidneys, nearly 40% of observed decisions are in first position, followed by an even steeper decline in the fraction of decisions at higher positions. There are relatively few decisions past the eighth position for either type of organ.

Figure 3: Proportion of Observations by Queue Position



Another useful way to describe the data would be to plot the fraction of rejections and the fraction of discarded organs (conditional on rejection) at each position. We see in Figure 4 that organs are discarded as early as the first position, presumably because such organs enter the offering sequence in relatively poor condition. Discard rates remain fairly stable at higher positions, except for a spike at positions 7 and 8 for livers. By contrast, rejection rates, which start at around 60% in first position for both livers and kidneys, increase steadily with position. Notice that rejection rates are systematically lower for kidneys; this results in shorter sequence lengths, explaining the steeper decline in the fraction of organs by position that we documented for kidneys in Figure 3.

4.2 Center Ability

The measure of center ability that we use for the herding tests is based on the center's rejection rate in first position. Figure 5 plots this statistic for all transplantation centers in the United Kingdom, separately for livers and kidneys. There are eight liver transplant centers in total, and we see that their probability of rejection in first position ranges from 0.3 to nearly 0.8. This range is quite wide, even if we ignore one outlying center with a relatively low rejection rate. There are many more kidney transplantation centers (twenty-four), and here again we see wide variation in the probability of rejection. Ignoring two centers with relatively low or high probabilities, the rejection rates range from around 0.4 to 0.8. Thus, there appears to be substantial variation in our measure of ability across transplantation centers.

As noted, the rejection probability in first position can be positively or negatively associated with center ability, depending on the sign of expression (7). We determine the sign of the expression by

Figure 4: Probability of Rejection and Discarding by Queue Position

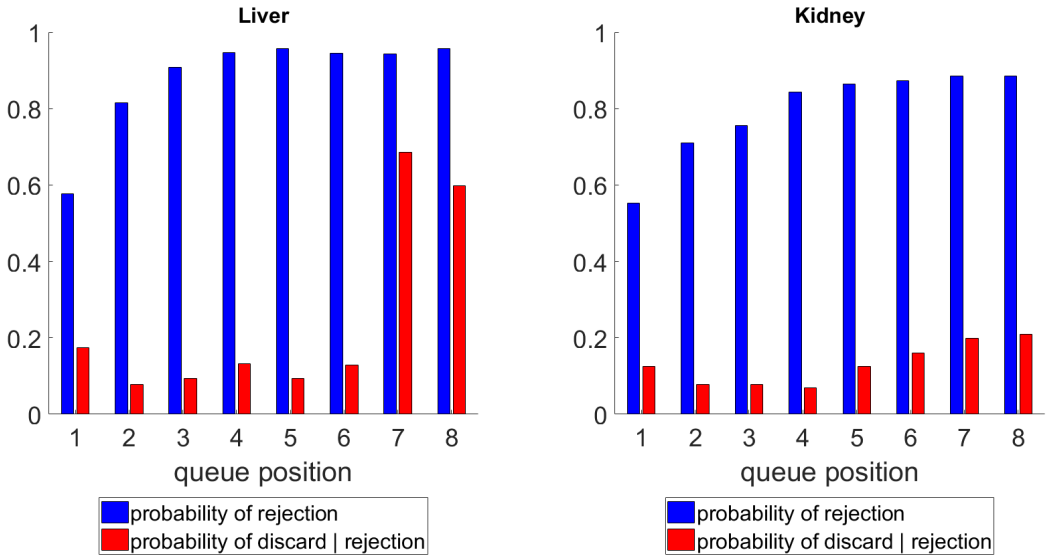
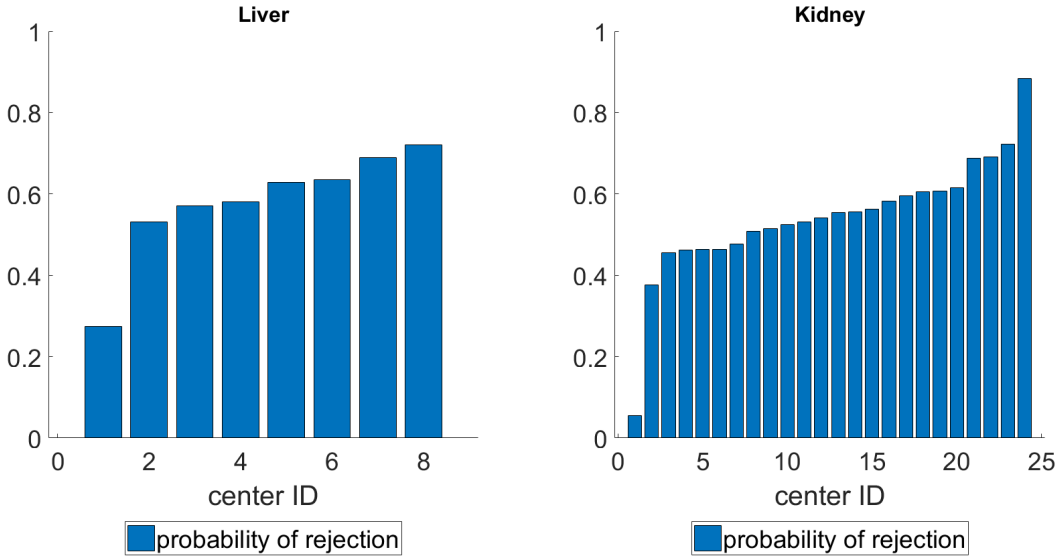


Figure 5: Rejections by Centers in First Position



estimating the relationship between the probability that an organ is rejected in second position and center 1's rejection rate in first position: $p_2 = \alpha_0 + \alpha_1 \overline{p}_1$. If the estimated α_1 coefficient is positive (negative), then center ability is increasing (decreasing) in $\overline{p}_1$.

Table 2: Measuring Center Ability

Dependent variable:	decision in second position			
Organ:	liver		kidney	
	(1)	(2)	(3)	(4)
$\bar{p_1}$	0.300***	0.440***	-0.281***	-0.374**
	(0.068)	(0.089)	(0.036)	(0.038)
Center 2 fixed effects	No	Yes	No	Yes
Organ risk index	No	Yes	No	Yes
Mean of dependent variable	0.816	0.816	0.710	0.710
N	6383	5684	9257	8764

Note: heteroscedasticity-robust standard errors in parentheses
Decision in second position: reject = 1, accept = 0
$\bar{p_1}$ measures center 1's rejection rate in first position
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The relationship between the decision in second position for each organ that reaches that position and center 1's probability of rejection when in first position, $\overline{p}_1$, is reported in Table 2, Columns 1 and 3, for livers and kidneys respectively.[7] We also report results with an augmented specification that includes center 2 fixed effects and a risk index of organ quality (described in detail below) in Columns 2 and 4. The coefficient on $\overline{p}_1$ is positive and significant for livers, and negative and significant for kidneys. By expression (9), centers in second position are unambiguously more likely to reject when they follow a higher ability center. The results in Table 2 thus imply that center ability is increasing (decreasing) in the first-position rejection rate for livers (kidneys).

Notice that the $\overline{p}_1$ coefficients estimated with the benchmark and augmented specifications are statistically indistinguishable, despite the fact that these coefficients are very precisely estimated. This verifies two assumptions of the model: (i) the ordering of centers is independent of their ability, and (ii) organ quality does not vary systematically across centers in first position. If center abilities, q_1 and q_2 were correlated, then adding center 2 fixed effects, which subsume q_2, in the estimating equation would shift the $\overline{p}_1$ coefficient. If center ability, which is associated with $\overline{p}_1$, was correlated

[7] We use the linear probability model to estimate our measures of ability and for the tests of herding because it follows directly from the specified estimating equations and because the marginal effects are easy to interpret. However, this implies that the error term will be heteroscedastic and, hence, robust standard errors are reported in this table and the tables that follow.

with organ quality, then adding the risk index to the estimating equation would similarly change the estimated $\overline{p}_1$ coefficient. Having established that organ quality does not vary systematically across centers in first position, variation in $\overline{p}_1$ can be attributed entirely to differences in center ability. We thus measure center ability by $\overline{p}_1$ for livers and by $1-\overline{p}_1$ for kidneys in the tests of herding that follow.

Notice also that we do not cluster standard errors in Table 2. The dependent variable in Table 2 is the decision by center 2 to reject or accept a given organ and the residual in the estimating equation incorporates the organ-specific signal that center 2 receives. Center 1 must have rejected the organ and, given the assumption (verified below) that all centers follow their signals in first position, must have received a b signal. However, there will be variation in the type of organs that it passes down on account of the mistakes that it makes. While it will often correctly receive a b signal for a B organ, it will also on occasion receive a b signal for a G organ. Given the assumption in the model that signals are independent, conditional on organ type, the mistakes will be independent across organs. The signals received by center 2 will also be independent across organs, conditional on the type of organ that it receives. It follows that the error (residual) term in the estimating equation is independent across organs for a given center 1-center 2 pair.[8] By the same argument, the error term will be independent across organs with different center-pairs in first and second position. The estimated standard errors in Table 2 should not be clustered and this is also true, for the same reasons, for the tests of herding that follow.[9]

4.3 Testing for Herding

The comparative statics for the herding tests are derived by exogenously varying the ability of centers at different positions in the sequence for a given organ. To implement the herding tests, we would thus ideally want to randomly vary the ordering of centers for the same organ. Given that this experiment is evidently infeasible, we exploit the fact that centers are assigned quasi-randomly to organs; in particular, the ordering of centers, which is based on patient priority, is independent of their ability (as verified empirically above). This allows us to use data from multiple organs, with different sequences of centers, for the tests of herding.

Our first test of herding is based on equation (2): $p_2 = \alpha_0 + \alpha_1 q_1 + \alpha_2 q_2 + \alpha_3 q_1 \cdot q_2$, where p_2 is the probability that a given organ is rejected in second position and q_1, q_2 measure the ability of center 1 and center 2, respectively. As described in Test 1, we expect the cross-partial effect (i.e. the effect of the interaction term $q_1 \cdot q_2$) to be negative when each center's ability is increasing in its first-position rejection rate. Based on the preceding results, this will be the case with livers. In contrast, we expect the cross-partial effect to be positive for kidneys, as center ability is negatively associated with the

[8] When center 2 herds after center 1, it always rejects and, hence, there is no variation in the error term across organs for that center-pair. However, it continues to be the case that the error term for a given organ provides no additional information about the error term for other organs and, hence, satisfies the independence assumption.

[9] Abadie et al. (2017) note that if the researcher assesses that the assignment mechanism is not clustered, as we do based on the model, then the standard errors should not be clustered. This is true irrespective of whether such an adjustment would change the standard errors.

first-position rejection rate.

Table 3 reports the estimated relationship between the decision in second position for each organ that reaches that position and center ability in first and second position, together with the interaction term. To interpret the estimated coefficients, it is convenient to normalize so that $\underline{q} = 0$. The coefficient on q_1 then applies to the case where q_2 equals zero. Assume that a center with ability 0 is completely uninformed; i.e. $\beta(0) = 1 - \gamma(0)$. This implies that without herding, a center with $q_2 = 0$ makes decisions that are independent of the quality of the organs that it receives and, hence, are independent of q_1. However, with herding, the probability that such a center rejects for sure is increasing in q_1. The predicted effect is ambiguous, and we find that α_1, the coefficient on q_1, is small and imprecisely estimated for kidneys and much larger and significant at the one percent level for livers. α_2, the coefficient on q_2, applies to the case in which q_1 equals zero. In this case, the first center's decision has no effect on the quality of the organ pool that is passed on and, moreover, does not increase the likelihood that the center that follows will herd. When $q_1 = 0$, center 2 effectively behaves as if it is in first position and α_2 corresponds to the intercept in Figures 1 and 2, which is increasing (decreasing) in q_2 for livers (kidneys). As predicted, the coefficient on q_2 is positive and significant for livers and negative and significant for kidneys. Our test of herding, however, is based on the interaction coefficient. As predicted by the model, the interaction coefficient is negative and significant for livers, and positive and significant for kidneys. In contrast, the interaction coefficient would be positive and significant for both livers and kidneys in the absence of herding.

Table 3: First Test of Herding (based on decisions in second position)

Dependent variable:	decision in second position	
Organ:	liver	kidney
	(1)	(2)
Center 1 ability (q_1)	2.135**	-0.0245
	(0.713)	(0.097)
Center 2 ability (q_2)	2.588***	-0.999***
	(0.688)	(0.102)
($q_1 \times q_2$)	-2.668**	1.003***
	(1.103)	(0.246)
N	6383	9257

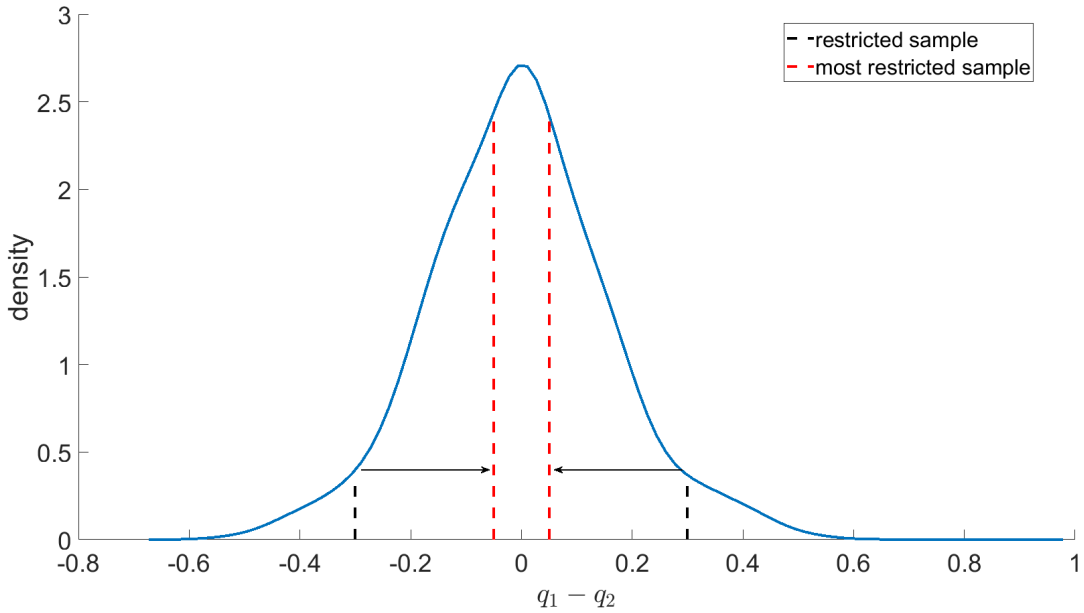
Note: heteroscedasticity-robust standard errors in parentheses.
Decision in second position: reject = 1, accept = 0
The constant term cannot be interpreted and is thus not reported.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.05$

Our second test of herding is based on the decisions made by centers in third position, as specified

in equation (3): $p_3 = \lambda_0 + \lambda_1 q_1 + \lambda_2 q_2$. As described in Test 2, we expect these decisions to vary more with first-position ability, q_1, than with second-position ability, q_2, when there is some amount of herding at position 2. The implicit assumption, which we verify below, is that centers always follow their signals in position 1. By contrast, decisions in third position will be equally responsive to q_1 and q_2 in the absence of herding (subject to the additional condition that $q_1 = q_2$).

Table 4 reports the estimated relationship between the decision in third position for each organ that reached that position, and center abilities q_1 and q_2. Center ability, for livers and kidneys, is measured as in Table 3. As predicted by the model when herding is present, the coefficient on q_1 is substantially larger than the coefficient on q_2; it is twice as large for livers and 50% larger for kidneys. The coefficients on q_1 and q_2 (λ_1 and λ_2 respectively) are imprecisely estimated for livers, and we cannot reject the hypothesis that $\lambda_1 \leq \lambda_2$. The corresponding coefficients for kidneys are, however, statistically significant; we can reject the hypothesis that $\lambda_1 \leq \lambda_2$ at the 5 per cent level.[10]

Figure 6: Ability Differential ($q_1 - q_2$) Distribution



Note: sample includes all kidneys that reached third position.

The data requirements to implement the second test of herding are quite stringent: (i) A substantial fraction of centers should herd in second position. (ii) A substantial fraction of centers should *not* herd in third position (if they did, then variation in q_1, q_2 would have no consequence for their decisions). (iii) There should be substantial variation in decisions – accept versus reject – in third position for

[10]In the model, information signals are independent across centers for a given organ. The implicit assumption is that no center decides more than once for a given organ. Given that centers have multiple patients on the waiting list, it is possible that this requirement will not be satisfied in practice. It turns out that sequences in which the same center makes repeated decisions are rare in the data; 7% of all decisions for kidneys and 3% of all decisions for livers are repeat decisions made by a center with the same organ. Restricting attention to the first three positions, which we use for the herding tests, these statistics decline even further to 5% and 2%, respectively.

Table 4: Second Test of Herding (based on decisions in third position)

Dependent variable:	decision in third position	
Organ:	liver	kidney
	(1)	(2)
Center 1 ability (q_1)	0.104	0.352***
	(0.066)	(0.045)
Center 2 ability (q_2)	0.0529	0.220***
	(0.064)	(0.046)
Constant	0.820***	0.541***
	(0.067)	(0.024)
F-statistic ($\lambda_1 \leq \lambda_2$)	0.47	3.43
p-value	[0.247]	[0.032]
$\bar{q_1}$	0.60	0.40
$\bar{q_2}$	0.65	0.40
N	4819	6084

Note: λ_1, λ_2 are the coefficients on q_1, q_2, respectively
$\bar{q_1}$ and $\bar{q_2}$ denote the sample means of q_1 and q_2, respectively.
Heteroscedasticity-robust standard errors in parentheses
Decision in third position: reject = 1, accept = 0
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Second Test of Herding (restricted samples)

Dependent variable:	decision in third position			
$(q_1 - q_2)$ range:	[-0.30,0.30]	[-0.20,0.20]	[-0.10,0.10]	[-0.05,0.05]
	(1)	(2)	(3)	(4)
Center 1 ability (q_1)	0.396***	0.427***	0.638***	1.020**
	(0.053)	(0.064)	(0.136)	(0.468)
Center 2 ability (q_2)	0.153**	0.132**	0.00181	-0.437
	(0.055)	(0.065)	(0.140)	(0.470)
Constant	0.554***	0.550***	0.513***	0.515***
	(0.025)	(0.027)	(0.039)	(0.045)
F-statistic ($\lambda_1 \leq \lambda_2$)	7.27	6.81	5.94	2.44
p-value	[0.004]	[0.004]	[0.007]	[0.059]
N	5603	5071	3069	1665

Note: heteroscedasticity-robust standard errors in parentheses

Alternative samples restricted to kidneys within a pre-specified ability differential $(q_1 - q_2)$ range

Decision in third position: reject = 1, accept = 0

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

the test to have statistical power. Based on the structural parameter estimates, we will see below that conditions (i) and (ii) are satisfied for both livers and kidneys. The important difference between the two organ types is that by the third position, over 90% of decisions for livers are rejections. This lack of variation might explain why the coefficients on q_1 and q_2 are imprecisely estimated in Column 1 of Table 4. While livers are most useful for identifying herding with the first test, we thus focus on kidneys for the second test.

Test 2 is derived for the case where centers at position 1 and 2 have equal ability; i.e. $q_1 = q_2$. Although, average ability in first and second position, $\overline{q}_1$ and $\overline{q}_2$, respectively, are equal for kidneys in Column 2, the more stringent requirement to test the model is that these abilities should be equal for each organ. Figure 6 describes the distribution of the ability differential, $q_1 - q_2$, for all kidneys that reached at least third position (and are thus used for the second test of herding). Although the distribution is centered at zero, there is substantial variation in the ability differential statistic. Table 5 takes account of this variation in $q_1 - q_2$ by implementing the second test of herding with an increasingly restricted sample of kidneys; i.e. by gradually narrowing the ability differential range. We see that the key result from Table 4, which is that the coefficient on q_1 is significantly larger than the coefficient on q_2 for kidneys, is retained as we reduce the sample. Indeed, this result is even obtained with the most stringent ability-differential restriction in Column 4, by which point the sample is just one-quarter of the full sample of kidneys.

5 Structural Estimation and Quantification

5.1 Estimation

Our model has the following features. There are two types of organs G and B. The fraction of good organs in the population of organs is π and the cutoff belief (that the organ is good) above which centers accept an organ is $\tilde{\pi}$. Centers are heterogeneous in their ability to identify organ types with β_j describing the probability that center j receives a b signal when the organ is bad and γ_j describing the probability that it receives a g signal when the organ is good.[11] Therefore, there are four sets of parameters: π, $\tilde{\pi}$, β_j and γ_j, for each j.

In the model, a given organ is either good (G) or bad (B). The type of organ and the center's ability determine the probability that it will receive a g or, conversely, a b signal, and hence the decision that it makes. To predict decisions in the model it is thus necessary to generate signals that are organ and center specific. To do this, we take advantage of two indices of organ quality that have previously been developed specifically for the United Kingdom: the UK KDRI (Kidney Donor Risk Index) and the UK DLI (Donor Liver Index). We use these indices to simultaneously generate organ-center specific signals and to estimate β_j, γ_j, and π, as shown below.

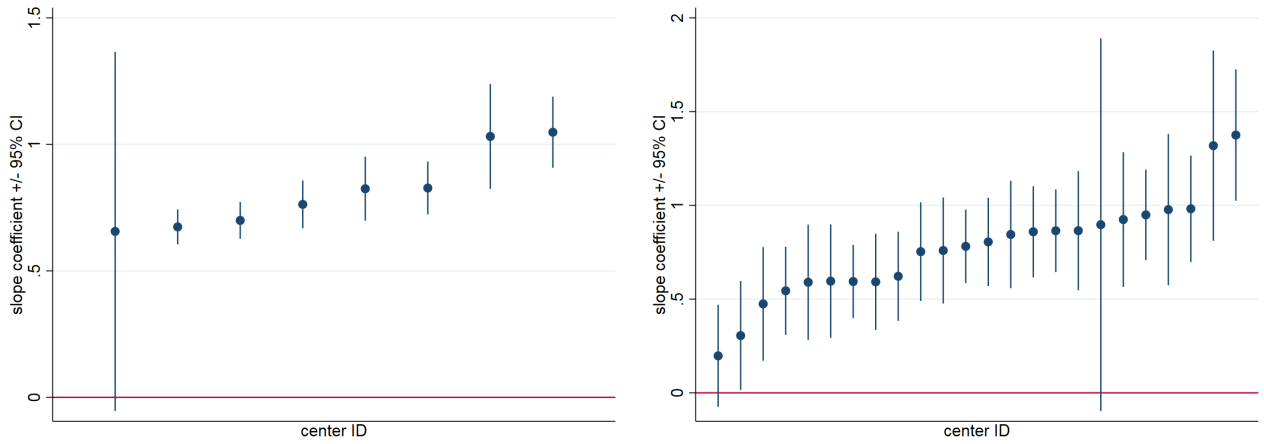
Indices of liver and kidney quality were first constructed in the United States, but have recently been adapted to the U.K. population. The UK KDRI is based on U.K. National Transplant Registry data covering over 7000 recipients who received deceased-donor kidneys between January 1, 2000 and December 31, 2007 (Watson et al., 2012). Various recipient and transplant factors were included in a model of transplant success, measured by patient survival, and the UK KDRI consists of those donor and organ characteristics that were found to be significant determinants of success (with optimal, estimated weights on each of those characteristics). More recently, data from all liver transplants from deceased donors between January 1, 2000 and December 31, 2014 have been used to construct the UK DLI (Collett et al., 2017). As with the UK KDRI, donor, recipient, and transplant data were used to identify factors associated with graft survival. Those donor and organ characteristics that were found to be significant determinants of transplant success, appropriately weighted, are included in the UK DLI.

The risk indices, which are (negatively) associated with the probability that an organ is a G organ, are based on outcomes generated by thousands of transplants over many years. The set of organ and donor characteristics included in the indices, and the weights placed on these characteristics, taken together, will accurately predict transplant outcomes or, equivalently, the quality of the organ. In contrast, transplant centers must base their assessment of an organ's quality, g or b, on their past experiences with a limited set of outcomes and the organ-specific information that they receive from NHSBT. The risk indices were originally developed to aid centers in their decision-making, and a proposal to incorporate the UK KDRI into the National Kidney Offering Scheme was presented at the

[11] Note that j now defines a center's identity alone and is no longer associated with its position in the sequence for a given organ.

2018 Blood and Transplantation Congress. At the time of writing, however, neither the UK KDRI nor the UK DLI, the latter of which was developed in 2017, are made available to transplant surgeons when they make their decisions. While centers may not have had explicit knowledge of the risk indices during the period of our analysis (2006-2015), we nevertheless expect that their assessments and, hence, their (rejection) decisions will track with the risk index.

Figure 7: Center-Specific Probit Slope Coefficient Estimates



(a) liver (b) kidney

Note: estimates based on the relationship between the probability of rejection in first position and the risk index.

To test the preceding hypothesis, we estimate the relationship between the probability that an organ i is rejected and its risk index R^i, separately by center, restricting the sample to decisions that were made when centers were in first position (and therefore, by assumption, following their signals). We use the probit model for the estimation because this ensures that predicted rejection probabilities lie in the unit interval; these predicted values will be needed to estimate β_j, γ_j, and π, as discussed below. Figure 7 reports probit estimates of the Risk Index (slope) coefficient, with the corresponding 95% confidence interval, by center. The estimated coefficient is positive and significant, almost without exception, both for livers and kidneys. Notice, however, that there is substantial variation in this coefficient across centers. We would expect rejection decisions by higher ability centers to track more closely with the risk index, and we will build on this intuition to estimate β_j, γ_j below.

Based on our interpretation of the risk index, the probability that organ i is a good organ, π^i, is decreasing in its risk index, R^i. We model this by assuming that $\pi^i = \pi(R^i)$ for some decreasing function $\pi(.)$. We place two additional restrictions on the $\pi(.)$ function: $\pi(\underline{R}) = 1$, $\pi(\overline{R}) = 0$. This is simply saying that an organ is a G organ with probability one (zero) at the bottom (top) of the risk index distribution. Given that the probability that center j will reject organ i in first position, p^i_{j1}, is equal to the probability that it receives a b signal, the following must hold

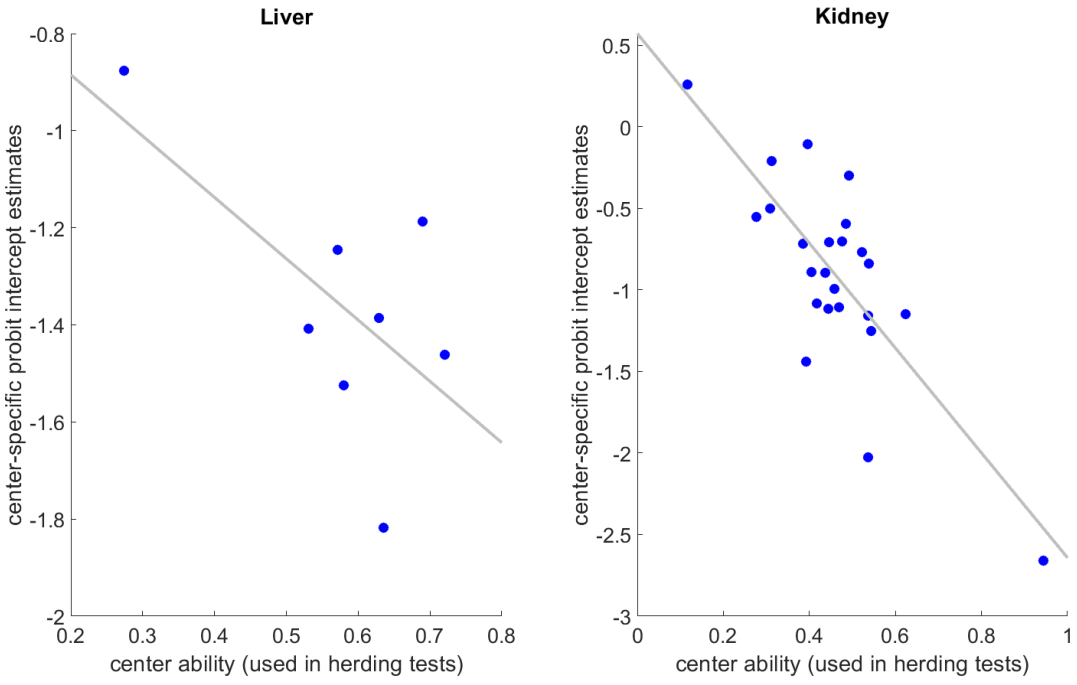
$$p^i_{j1} = \pi^i(1-\gamma_j) + (1-\pi^i)\beta_j. \tag{17}$$

Rearranging terms,

$$p^i_{j1} = (1-\gamma_j) + (1-\pi^i)(\beta_j - (1-\gamma_j)). \tag{18}$$

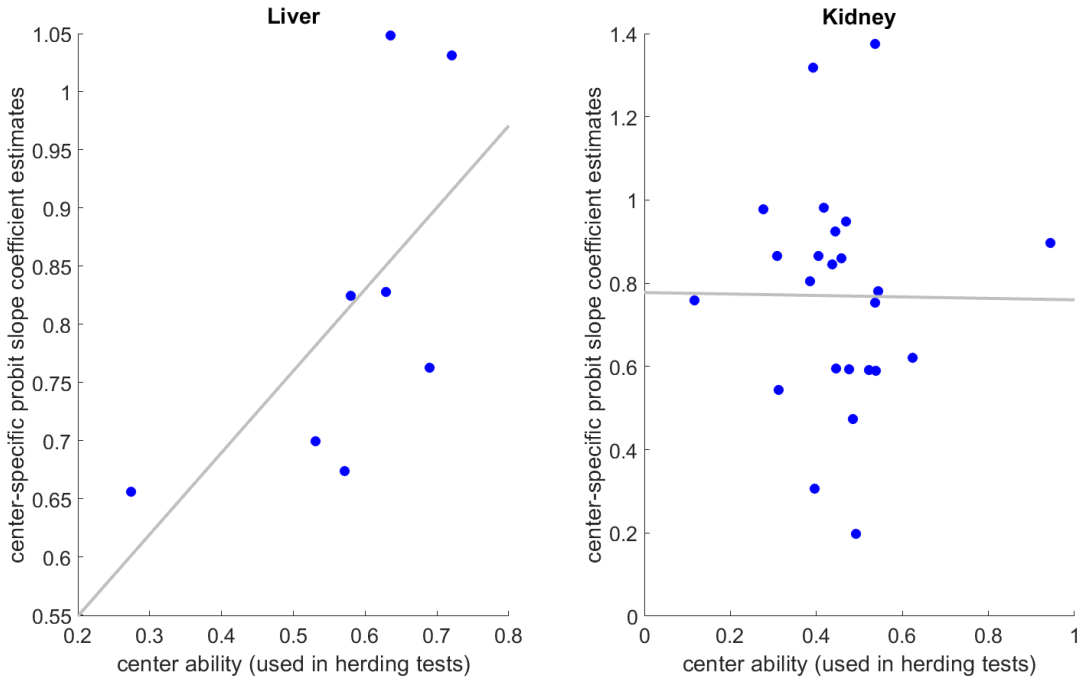
Given the restrictions we have imposed on the $\pi(.)$ function, $(1-\pi^i)$ is increasing in R^i and is equal to zero at $R^i = \underline{R}$. When estimating the $p^i_{j1}-R^i$ relationship, as above, the intercept $(1-\gamma_j)$ is decreasing in center ability, while the slope, associated with $\beta_j - (1-\gamma_j)$, is increasing with ability. Intuitively, higher ability centers will reject fewer organs with a low risk index, hence the smaller intercept, while simultaneously rejecting more organs with a high risk index, which results in the steeper slope. We will use the estimated intercept and slope to construct β_j, γ_j momentarily, but first we use this insight to assess the consistency of our independently constructed measures of center ability, used in the herding tests and based on the risk indices, respectively. Figures 8 and 9 report the relationship between the intercept and slope, estimated using the risk index, and the measure of ability used in the herding tests, based on each center's rejection rate in first position. Cross-validating our ability measures, the observed associations match the model's predictions with one exception (the slope coefficient for kidneys, where no relationship is discernable).

Figure 8: Cross-Validating Ability Measures (intercept)



At $R^i = \underline{R}$, $\pi^i = 1$ and, hence, $p^i_{j1} = (1-\gamma_j)$ from equation (18). At $R^i = \overline{R}$, $\pi^i = 0$ and, hence,

Figure 9: Cross-Validating Ability Measures (slope)



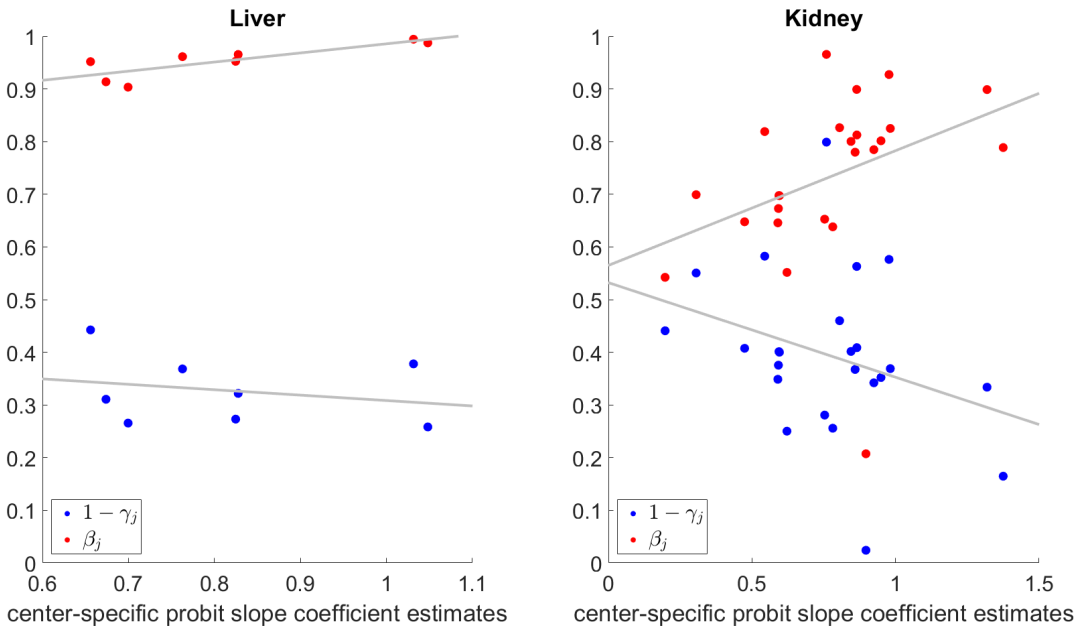
$p^i_{j1} = \beta_j$. Using the $p^i_{j1}-R^i$ relationship that we have estimated for each center, the predicted p^i_{j1} at $R^i = \underline{R}$ provides an estimate of $1-\gamma_j$ and the predicted p^i_{j1} at $R^i = \overline{R}$ provides an estimate of β_j. Figure 10 reports the estimated β_j and $1-\gamma_j$, for all centers, separately for livers and kidneys. These predictions are computed at the 95^{th} percentile and the 5^{th} percentile of the risk index distribution, respectively. Assumption 1 in the model states that $\beta_j \geq 1-\gamma_j$; i.e. that centers are not systematically misinformed. We see in Figure 10 that this assumption is satisfied for each center, both for livers and for kidneys.[12]

The risk indices can also be used to compute π (the fraction of G organs in the population of organs). Given the estimated $p^i_{j1}-R^i$ relationship for each center, p^i_{j1} can be predicted for any organ i with risk index R^i. Given the predicted p^i_{j1} and the estimated β_j, γ_j, we can recover $\pi^i = \pi(R^i)$ from equation (17). Although, in principle, π^i corresponding to a given R^i should be the same for all centers, noise in the estimated β_j and γ_j could generate some variation in practice. Our best estimate of π^i is thus the average across all centers that were offered organ i. Averaging the estimated π^i across all organs, we arrive at an estimate of π, the fraction of G organs in the population of organs: 0.57 for livers and 0.60 for kidneys.

Having constructed measures of center ability (β_j and γ_j) and estimated π, all that now remains

[12]Notice also that, while β_j is increasing and $(1-\gamma_j)$ is decreasing in the estimated risk index coefficient, the relationships are not monotone. This is because β_j, γ_j are based on both the estimated slope and the estimated intercept of the $p^i_{j1} - R^i$ relationship.

Figure 10: Estimates of Center Ability



is to estimate $\tilde{\pi}$, the cutoff belief that an organ is a G organ above which centers accept the offer. All centers follow their signal in first position in the model, which implies that their belief following a g (b) signal lies above (below) $\tilde{\pi}$. While some centers continue to follow their signals in later positions, others will start to herd (i.e. to reject offers regardless of whether they receive a g or a b signal). This is because their beliefs always lie below $\tilde{\pi}$. As $\tilde{\pi}$ increases, the fraction of centers that herd thus increases, with an accompanying increase in the rejection rate (the decisions of centers that follow their signals remain unchanged). To estimate $\tilde{\pi}$ we thus match the overall rejection rate in the data to the rejection rate predicted by the model; there is a unique value of $\tilde{\pi}$ at which the actual and predicted rejection rates match and this will be our best estimate of the $\tilde{\pi}$ parameter.

The simulated method of moments is used to estimate $\tilde{\pi}$. To draw signals for the estimation, we take advantage of the fact that our center-specific probit estimates of the $p^i_{j1}-R_i$ relationship allow us to predict the probability of rejection in first position for any organ-center pair. Since centers always follow their signals in first position, as verified below, this provides us with the probability that the center would receive a b signal in first position and, for that matter, in any position. We draw signals in this way, and then predict decisions at each position (given the previously-estimated values of β_j, γ_j, and π). The average over multiple draws of the signals predicts the overall rejection rate for a given $\tilde{\pi}$, and we then search over all $\tilde{\pi}$ to find the value at which the actual and predicted rejection rates match.

The data are effectively generated by a single draw from the signal distribution. Even if the model was correctly specified and the correct value of $\tilde{\pi}$ was selected by the econometrician, actual decisions

and predicted decisions would evidently not match at each organ-position. However, as long as a large number of centers follow their signals, this sampling error will wash out and actual and predicted rejection rates, overall, will match when the correct $\tilde{\pi}$ is selected.[13] Table 6 reports $\tilde{\pi}$ estimates, with bootstrapped standard errors, separately for livers and kidneys. As we estimate a single parameter, we need only match on a single moment; we can utilize the rejection rate at any position (except for the first position where we do not draw signals) for this purpose and the benchmark specification matches on the rejection rate in second position.

Recall from Table 1 that transplant success does not vary by position in the U.K. Our interpretation of this finding was that the organ-recipient mismatch and organ deterioration, which are increasing in higher positions, are not salient for decision-making.[14] An alternative explanation is that centers are more conservative; i.e. set a higher $\tilde{\pi}$ at higher positions in response to the increased mismatch and deterioration. Rejecting this explanation, we see in Table 6 that the $\tilde{\pi}$ estimate remains very stable when we match on additional positions all the way up to position 5; the only exception is when we add position 3 for estimation with kidneys, but the $\tilde{\pi}$ estimate actually declines in this case.[15]

Table 6: Structural Parameter Estimates

Organ:	liver				kidney			
Number of matched moments:	one (1)	two (2)	three (3)	four (4)	one (1)	two (2)	three (3)	four (4)
$\tilde{\pi}$	0.87	0.86	0.86	0.86	0.58	0.52	0.50	0.50
	(0.0038)	(0.0058)	(0.0055)	(0.005)	(0.0044)	(0.0018)	(0.0066)	(0.0015)
N	5029	3780	3109	2548	7691	5031	3508	2747

Note: when matching moments, we begin with the probability of rejection in second position, and sequentially add the corresponding probabilities in third, fourth and fifth positions.

Bootstrapped standard errors in parentheses.

As multiple positions (moments) are available for estimation, we might think it possible to simultaneously estimate both $\tilde{\pi}$ and π. To see why this is infeasible, however, suppose that we pre-select a value of π, but leave the estimated β_j and γ_j unchanged. The generated signals will no longer

[13] A special feature of our data is that the order of centers who would have been approached, past the position where an organ is accepted or discarded, is unavailable. Thus, if the model predicts a rejection at the final position in a sequence where an organ was accepted in the data, then we can go no further. To be consistent, if the model predicts an acceptance at a position where an organ was rejected in the data, we proceed no further (and subsequent positions are not utilized for estimation).

[14] The implicit assumption is that our outcome measure matches the measure that centers use in their decision-making. Three-year survival is the most stringent measure of transplant success that is used in the transplantation literature. Given that center position has no effect on this measure, it is unlikely that it will affect a less stringent measure, and hence our result effectively covers the range of success measures that centers might use in their decision-making.

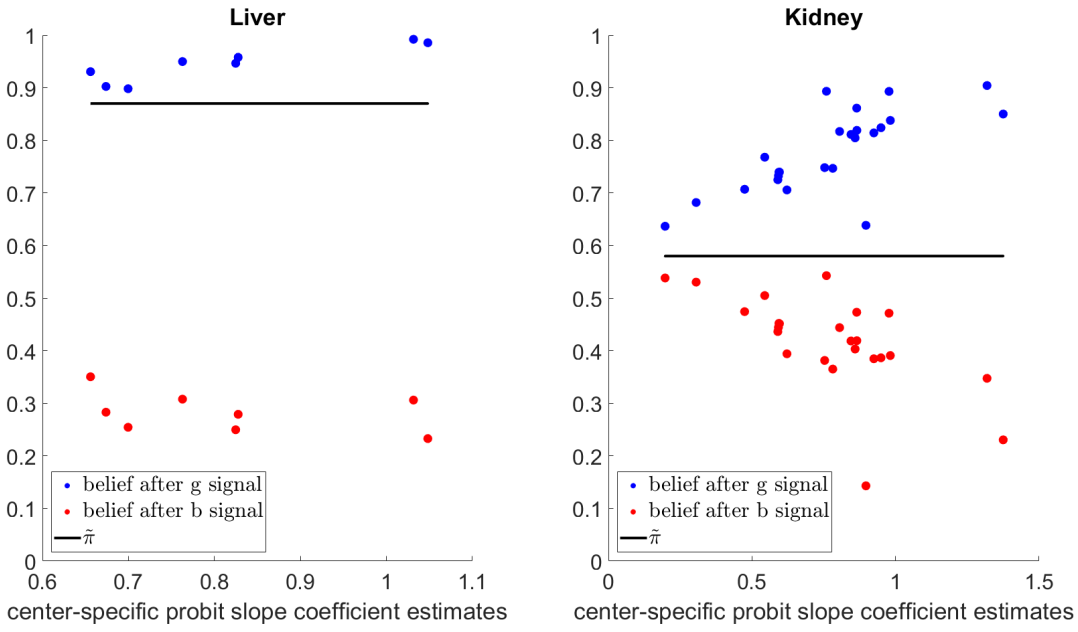
[15] When matching on multiple positions, we compute the error in the rejection rate at each position and then take the unweighted average across all positions to compute the overall error. Our best estimate of $\tilde{\pi}$ is the value that minimizes the overall error.

be organ-specific but, nevertheless, as π increases, the rejection rate among centers that follow their signals will decrease. To bring the overall rejection rate back in line with the data, $\tilde{\pi}$ must increase (and, with it, the fraction of centers who herd). It follows that, for every value of π, there exists a $\tilde{\pi}$ such that the actual and predicted rejection rates match. Therefore, $\tilde{\pi}$ and π cannot be estimated simultaneously, highlighting the important role played by the risk index in our analysis.

5.2 Validation and Goodness of Fit

As we need only estimate a single parameter, $\tilde{\pi}$, it is straightforward to establish that this parameter is identified; i.e. that a unique value exists at which the model best fits the data. This value must, however, also satisfy additional restrictions implied by the model. In particular, Assumption 2 requires that each center's belief that the organ is good must lie above (below) $\tilde{\pi}$ when it receives a g (b) signal in first position. This ensures that it follows its own signal. Figure 11 verifies that this important assumption is satisfied for each center, both for livers and for kidneys.[16]

Figure 11: Updated Beliefs in First Position, by Center

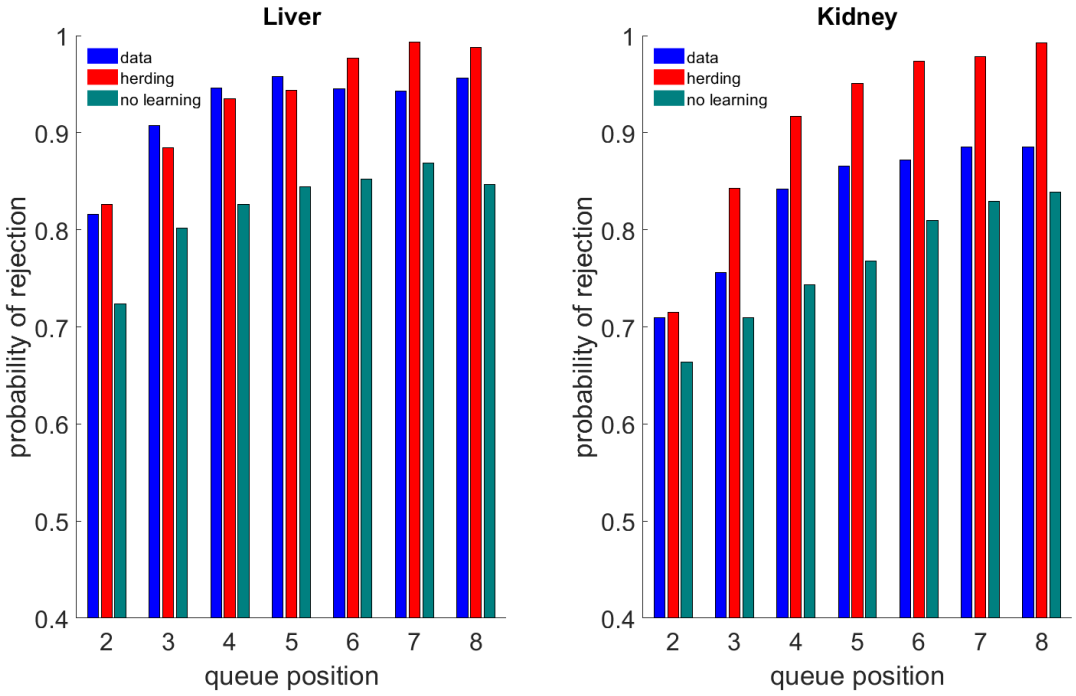


Having estimated the model and validated its key assumptions, the next step is to assess the model's goodness of fit with the data. As we use the overall rejection rate to estimate $\tilde{\pi}$, we begin by comparing the actual rejection rate and the predicted rejection rate, by position, separately for livers and kidneys. The benchmark $\tilde{\pi}$ estimate is obtained by matching rejection rates in second position,

[16] Centers must also necessarily follow their own (g) signal when they accept. Based on the parameter estimates and the sequence of preceding centers, the model predicts that centers were following their own signals 80% of the time for kidneys and 81% of the time for livers at those positions where acceptances were observed in the data.

and all the results that follow are based on this estimate. Thus, we would expect a close match in second position, but not necessarily at higher positions. As a basis for comparison, we also report rejection rates from an alternative model with no (social) learning. Center ability and the signal-generating process in the alternative no learning model are determined in the same way as in our model, but decisions are now based exclusively on the signals received by each center (without regard to the decisions of preceding centers). We see in Figure 12 that rejection rates predicted by our model, which allows for social learning (herding), exceed the corresponding rates in the data, particularly at higher positions, while the alternative no-learning model systematically under-predicts rejection. The error associated with our model is not, however, substantial: overall, predicted rejection rates exceed actual rejection rates by just 3% for kidneys and 7% for livers

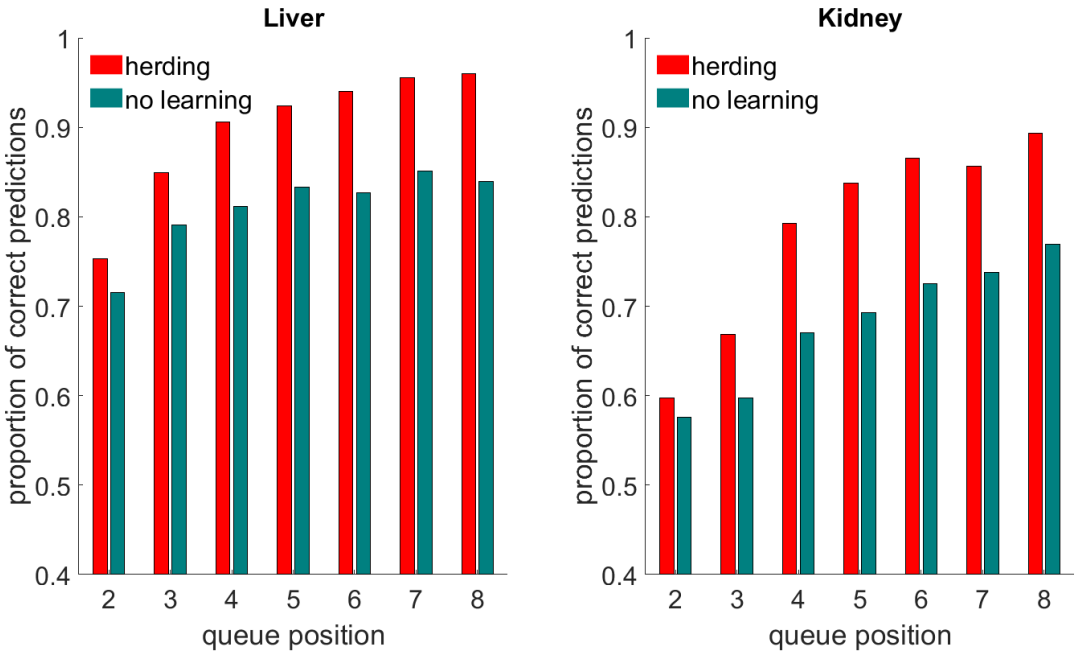
Figure 12: Goodness of Fit (Probability of Rejection)



An alternative metric to compare the performance of the two models is the fraction of correctly-predicted decisions – that is, acceptances (rejections) in the data that are predicted to be acceptances (rejections). Based on this metric, the herding model clearly out-performs the no-learning model, both for livers and for kidneys, as observed in Figure 13.[17] This result complements the reduced-form tests of the model, providing independent evidence that centers are herding.

[17]When we compare the actual and predicted rejection rate in Figure 12; i.e. the data versus the alternative models, sampling error at each organ-position washes out when it is averaged across all organs at a given position. In contrast, the mistakes in Figure 13 are all positive and, hence, the sampling error does not wash out in the same way. The comparison of the herding and the no learning models, nevertheless, remains valid.

Figure 13: Goodness of Fit (Proportion of Correct Predictions)



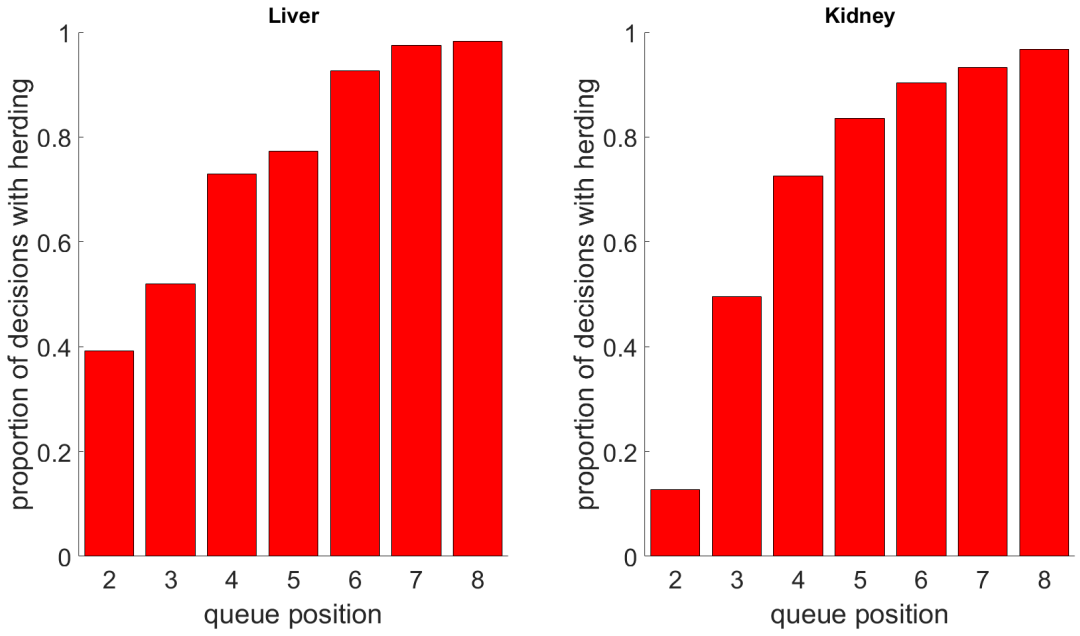
5.3 Quantification and Counter-Factual Simulations

While both the reduced-form tests and tests of the structural model against the no-learning alternative indicate that centers are herding, we would like to quantify the prevalence of this behavior. Given the estimates of β_j, γ_j, π, $\tilde{\pi}$ and the sequence of centers associated with each organ, we can determine whether a given center at a given position is herding – i.e. that its prior belief based on preceding decisions is so far below $\tilde{\pi}$ that it will reject regardless of the signal it receives. Figure 14 reports the prevalence of such herding, by position, for livers and kidneys. Herding is very common. For livers, about 40% of centers in second position herd. There is a steep increase in herding at higher positions and, by the sixth position, almost all centers herd. Herding is less prevalent, on average, for kidneys. Nevertheless, over 10% of centers in second position herd, there is a sharp increase to 50% at third position, and over 90% of centers herd by the sixth position.[18]

Having measured the prevalence of herding in organ transplant decisions, we next consider the efficiency consequences of this behavior. To do this, we compare discard rates (and associated decisions) under three models: (i) our model in which signals received by centers that herd are not observed by those that follow, (ii) a counter-factual no-learning model in which centers rely exclusively on their signals, and (iii) a counter-factual pooled information model in which centers observe and use the signals of all their predecessors.

[18]Based on the model, centers should always reject when they are herding. Centers reject 91% of the time for livers and 84% of the time for kidneys in positions where the estimated model predicts they will herd. In contrast, rejection rates are 70% for livers and 63% for kidneys in positions where centers are not predicted to herd.

Figure 14: Prevalence of Herding



The standard practice with counter-factual analyses is to compare outcomes predicted by the estimated model with simulated outcomes under alternative scenarios. The complication that arises when predicting discards with our herding model is that the order of centers who would have followed, past the point where an organ is accepted or discarded in the data, is unavailable. Given that predictions are thus based on observed data and given that observed sequences consist exclusively of rejections, with the possible exception of an acceptance at the end of a sequence, the model will under-predict discards, and sequence lengths, even if it is correctly specified (by predicting acceptances in place of rejections just by chance).[19] Given the special structure of our data, we use the actual discard rate and actual decisions to characterize outcomes with the herding model.

We begin by examining the decisions that underlie discard rates under the alternative models. A false rejection (acceptance) arises in our model, with herding, when a center that rejects (accepts) an organ would have reversed that decision if all preceding signals were made available; i.e. with pooled information. When centers are homogeneous in their abilities, as in the example that we used to motivate our analysis, false rejections arise when the first center receives a b signal and rejects a good organ. All centers that follow will ignore their g signals and do the same. However, this also rules out false acceptances; for such acceptances to arise, a center at the end of a sequence must be willing to deviate from the herd and accept the organ. Once we introduce heterogeneity in center ability, false rejections will decline because centers with sufficiently high ability (and a g signal) are willing to

[19]Some acceptances in the data will be predicted by the model to be rejections, but this will be less frequent than reversals in the opposite direction since rejections are more common. Moreover, we cannot infer that an organ for which an acceptance in the data is reversed by the model will necessarily be subsequently discarded.

deviate from the herd, but by the same argument, false acceptances can also emerge.

Under the no-learning model, centers follow their own signals, ignoring the negative information contained in their predecessors' rejections. The resulting false acceptances, relative to the pooled information benchmark, are more frequent than with the herding model, where they only arise because b signals received by predecessors who herd (and reject) are ignored. False rejections can also arise with the no-learning model if a center that received a b signal and rejected would have reversed its decision if preceding signals were made available and utilized, and a sufficiently large number of those signals were g signals (these would have to be centers that were herding in the data). These are stringent conditions and, hence, false rejections are likely to be rare with no-learning.

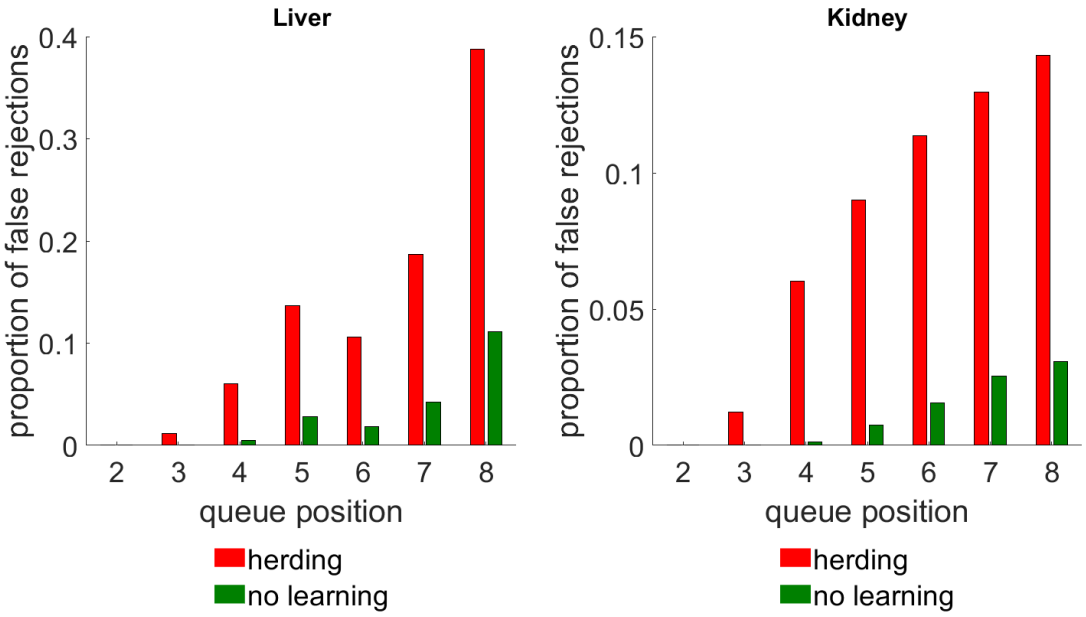
To predict decisions with the counter-factual pooled information and no-learning models, we only draw signals where necessary. The presumption is that the estimated herding model correctly characterizes the data. Thus, when a center accepts an organ, we assume that it must have followed its (g) signal. Similarly, when our model predicts that a center was following its signal, and it was observed to reject the organ, we assume that it received a b signal. It is only when our model predicts that a center was herding that we draw signals, using the organ's risk index and the center's ability to determine the probability that it received a b signal, as described above. Decisions generated by the pooled information model for each draw of the signals are compared with actual decisions (corresponding to the herding model) at each organ-position, and the resulting discrepancy is then averaged over multiple draws to compute the fraction of false rejections and false acceptances for the herding model. The same procedure is used to compare no-learning with pooled information, except that, with no-learning, centers always follow their signals.

Figure 15 reports the fraction of false rejections among all rejections, by position, in the herding and no-learning models, relative to the pooled information model. As predicted, the proportion of false rejections is greater for the herding model than under no-learning. Notice, however, that the rate of false rejections is relatively low even with herding, which can be explained (in part) by the heterogeneity in center ability. Recall that without this heterogeneity, once a center herds, all the centers that follow will do the same. Based on our structural estimates, at least one center in the sequence herds for 41% of livers and 29% of kidneys. Among these organs, at least one center follows its own signal after herding had commenced, for 59% of livers and 44% of kidneys. Moreover, false rejections can only start at the third position (if there is herding in second position) and will subsequently increase mechanically by position, together with the incidence of herding. In our data we see that these rejections increase from the fourth position, by which point most organs are discarded. False rejections, which are low to begin with, will thus have little impact on discard rates.

Figure 16 reports the fraction of false acceptances among all acceptances, comparing the herding model and the no-learning model to the pooled information benchmark. Once again, as predicted, we see in Figure 16 that false acceptances are more common with the no-learning model than with the herding model, beginning as early as position 2. Even one false acceptance (of a bad organ) is

extremely socially costly, and we see that such acceptances with the no-learning model are actually quite frequent, relative to the pooled information benchmark. In contrast, the herding model generates almost no false acceptances. This implies that herding may actually protect centers from accepting bad organs, and we will return to this point below.

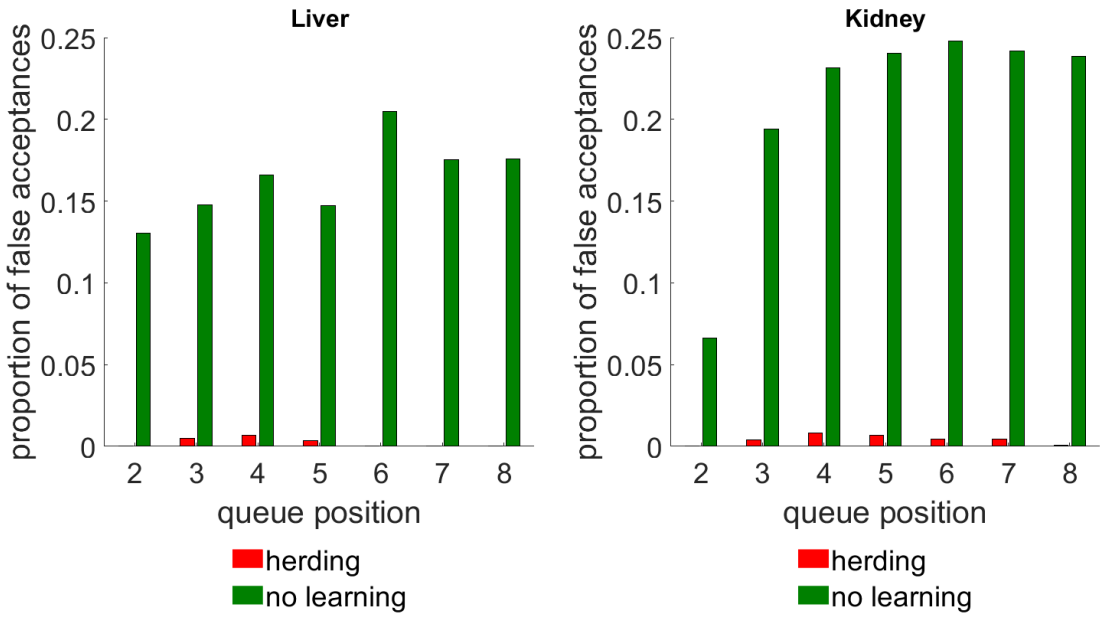
Figure 15: False Rejections as a Proportion of All Rejections



While the position-specific analysis described above paints a comprehensive picture of decision-making under different models relative to the pooled information benchmark, the important consideration from a social welfare perspective is that good organs are accepted and bad organs discarded. For example, the case in which a center herds behind its predecessors and rejects a good organ is costly for its patient. There is, however, no welfare loss as long as patients are treated interchangeably and the organ is accepted further down the line. We thus complete the analysis by comparing discard rates with pooled information and no-learning against the data; i.e. the herding model.

In our data, an organ is either accepted for transplantation or discarded (set aside for research). The decision to discard an organ is taken by NHSBT and is based on its usability, which depends, in turn, on its condition (which is distinct from its quality). Thus, the discard decision is treated as exogenously determined in our analysis. We do not know what would have happened if an organ that was accepted at a particular position in the data is rejected at all positions up to and including that point with one of the counter-factual models; that organ could have been accepted or discarded further down the line. The analysis that follows thus focuses on discarded organs, which is an important group

Figure 16: False Acceptances as a Proportion of All Acceptances



to study, and our objective is to determine how many of those organs would have been utilized under the counter-factual models.

Table 7 reports the discard rate for the herding model (in the data), the pooled information model, and the no-learning model. We see that discard rates in the data are substantially higher for livers than for kidneys: 0.40 versus 0.22. Based on the estimated model, this is because centers are much more conservative when accepting livers; recall that the fraction of good organs, π, is roughly the same for livers and kidneys (0.57 versus 0.60) whereas the cutoff belief, $\tilde{\pi}$, is substantially higher for livers (0.87 versus 0.58). $\tilde{\pi} = \frac{-L}{H-L}$ is decreasing in L and H. While it seems reasonable to assume that the payoff from transplanting a good organ, H, is the same for livers and kidneys (the value of an extended life), the payoff from transplanting a bad organ, L, is likely to be substantially lower for livers (because a failed transplant inevitably results in the recipient's death).

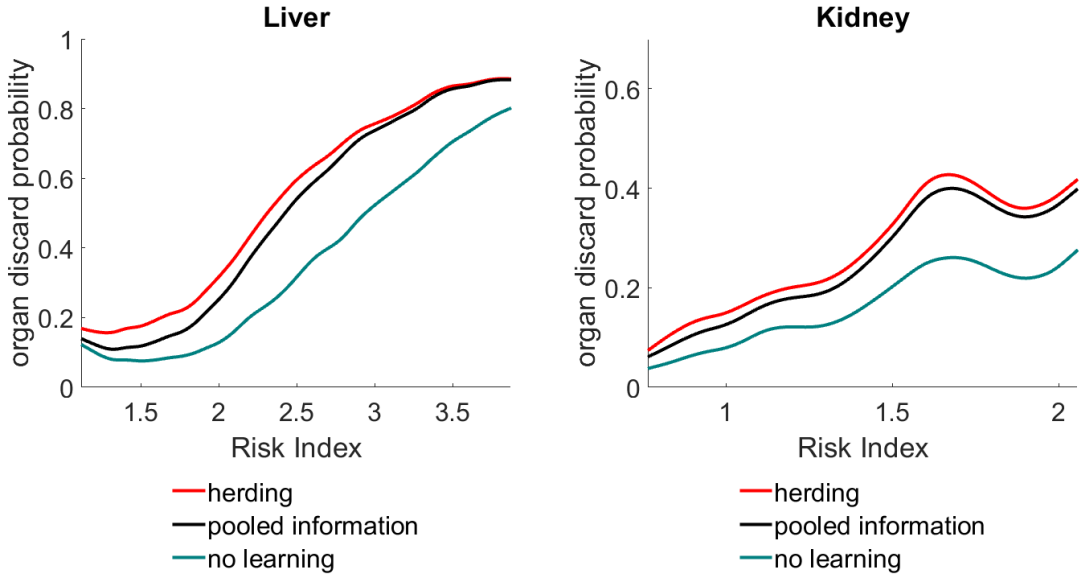
Table 7: Organ Discard Rates

Organ:	liver (1)	kidney (2)
Herding	0.40	0.22
Pooled information	0.36	0.20
No learning	0.25	0.13

Comparing the alternative models in Table 7, the organs that are discarded in the data (with the

herding model) would only have been accepted in the pooled information model if they were falsely rejected. We have seen that false rejection rates are low with the herding model. Moreover, sequence lengths are short in the data, which implies that false rejections (which tend to be concentrated at higher positions) will have a relatively small effect on discard rates. As expected, the discard rate would decline by just 4 percentage points for livers and 2 percentage points for kidneys with pooled information. In contrast, discard rates would decline substantially with no learning, by 15 and 8 percentage points, respectively. To place these statistics in perspective, 4253 livers and 3582 kidneys were discarded in our sample period. Of these discarded organs, 467 livers and 360 kidneys would have been retrieved with pooled information. As many as 1576 livers and 1471 kidneys would have been retrieved with no learning; i.e. by suppressing previous decisions.

Figure 17: Organ Discard Rates by Risk Index



What matters for welfare is not just the number of organs that are discarded (or retrieved) but also their quality. Figure 17 reports the discard rate in the data (with the herding model), for the pooled information model, and for the no-learning model across the range of risk indices. We have seen in Table 7 that discard rates with the pooled information model are only slightly lower than the corresponding rates in the data (with the herding model). As observed in Figure 17, this is true across the range of risk indices. This contrasts with discard rates for the no-learning model, which are substantially lower than those in the herding model and the pooled information model at higher risk indices. The false acceptances in the no-learning model that were documented in Figure 16 appear to be concentrated at higher risk indices, precisely where they are most dangerous. With organ transplantation, herding appears to protect centers from costly false acceptance decisions, without

substantially raising the discard rate through false rejections.

6 Conclusion

There are many economic environments in which prospective buyers, acting sequentially, must choose whether or not to acquire an object. It is often observed that (rejection) decisions across buyers are correlated. One explanation for this correlation is that buyers correctly, and independently, assess that the object is of poor quality. An alternative explanation, which goes back to Banerjee (1992) and Bikhchandani et al. (1992) is that agents further in line herd behind their predecessors and ignore their own assessment of the object's quality. Although the herding literature has advanced on many fronts since those canonical papers were published, there has been no test of the information inefficiency that lies at the heart of their analyses. We develop tests to detect information-based herding, which are based on heterogeneity in agent ability, together with a methodology to quantify its welfare consequences, which we apply to organ transplant decisions in the United Kingdom.

Organ transplantation in the U.K. is an ideal test-case for our methodology for a number of reasons. First, the ordering of centers for a given organ is independent of their ability to distinguish between good and bad organs. Second, the payoff from transplanting an organ is independent of center ability. In other applications, such as investment decisions or the labor market, higher ability agents could get to choose earlier on average and might have superior payoffs. As a result, lower ability agents might have a different belief threshold for acceptance. These additional considerations will complicate the analysis, but the basic principles of our methodology would still apply. Agent-specific measures of ability can be constructed as long as there is sufficient historical data on decisions and outcomes, which would also allow for something like the risk index to be constructed to measure object quality. Our tests of herding, the structural estimation, and the counter-factual simulations, suitably modified, may then be implemented. Herding has the potential to generate serious inefficiencies in many economic environments and, hence, this would seem to be an important area for future research.

Apart from its methodological contribution, our analysis applies to a setting where herding could have large practical consequences. Each year in the U.K., many patients on the wait list die or are removed from the list because their condition has deteriorated past the point where they are eligible for a kidney or liver transplant. At the same time, a large fraction of organs are discarded. Given the sequential nature of the decision-making process and the uncertainty in assessing an organ's quality, it is entirely possible that centers are herding and that viable organs are being discarded. While our analysis indicates that herding is indeed common, we find that the associated information inefficiency may not be very significant. Counter-factual analyses, based on the estimated structural parameters of the model, indicate that making private assessments of organ quality publicly available would have little impact on both the discard rate and the quality of accepted organs. This is because of the special nature of our application: centers are heterogeneous in their abilities and sequence lengths are short. In other settings, the inefficiencies typically associated with herding will likely be more substantial.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge, "When should you adjust standard errors for clustering?," Technical Report, National Bureau of Economic Research 2017.

Agarwal, Nikhil, Itai Ashlagi, Michael A Rees, Paulo J Somaini, and Daniel C Waldinger, "An empirical framework for sequential assignment: The allocation of deceased donor kidneys," Technical Report, National Bureau of Economic Research 2019.

Anderson, Lisa R and Charles A Holt, "Information cascades in the laboratory," *The American Economic Review*, 1997, pp. 847–862.

Avery, Christopher and Peter Zemsky, "Multidimensional uncertainty and herd behavior in financial markets," *American Economic Review*, 1998, *88* (4), 724–748.

Banerjee, Abhijit V, "A simple model of herd behavior," *The Quarterly Journal of Economics*, 1992, *107* (3), 797–817.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of Political Economy*, 1992, *100* (5), 992–1026.

Bohren, J Aislinn, "Informational herding with model misspecification," *Journal of Economic Theory*, 2016, *163*, 222–247.

Çelen, Boğaçhan and Shachar Kariv, "Distinguishing informational cascades from herd behavior in the laboratory," *American Economic Review*, 2004, *94* (3), 484–498.

Chamley, Christophe P, *Rational herds: Economic models of social learning*, Cambridge University Press, 2004.

Cipriani, Marco and Antonio Guarino, "Estimating a structural model of herd behavior in financial markets," *American Economic Review*, 2014, *104* (1), 224–51.

Collett, David, Peter J Friend, and Christopher JE Watson, "Factors associated with short- and long-term liver graft survival in the United Kingdom: development of a UK donor liver index," *Transplantation*, 2017, *101* (4), 786–792.

Conley, Timothy G and Christopher R Udry, "Learning about a new technology: Pineapple in Ghana," *American Economic Review*, 2010, *100* (1), 35–69.

Dupas, Pascaline, "Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment," *Econometrica*, 2014, *82* (1), 197–228.

Foster, Andrew D and Mark R Rosenzweig, "Learning by doing and learning from others: Human capital and technical change in agriculture," *Journal of Political Economy*, 1995, *103* (6), 1176–1209.

Frick, Mira, Ryota Iijima, and Yuhta Ishii, "Misinterpreting Others and the Fragility of Social Learning," 2019.

Gale, Douglas, "What have we learned from social learning?," *European Economic Review*, 1996, *40* (3-5), 617–628.

Golub, Benjamin and Evan Sadler, "Learning in social networks," *The Oxford Handbook of the Economics of Networks*, 2016.

Grinblatt, Mark, Sheridan Titman, and Russ Wermers, "Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior," *The American Economic Review*, 1995, pp. 1088–1105.

Lakonishok, Josef, Andrei Shleifer, and Robert W Vishny, "The impact of institutional trading on stock prices," *Journal of Financial Economics*, 1992, *32* (1), 23–43.

Munshi, Kaivan, "Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution," *Journal of Development Economics*, 2004, *73* (1), 185–213.

Park, Andreas and Hamid Sabourian, "Herding and contrarian behavior in financial markets," *Econometrica*, 2011, *79* (4), 973–1026.

Smith, Lones and Peter Sørensen, "Pathological outcomes of observational learning," *Econometrica*, 2000, *68* (2), 371–398.

Vives, Xavier, "Social learning and rational expectations," *European Economic Review*, 1996, *40* (3-5), 589–601.

Watson, Christopher JE, Rachel J Johnson, Rhiannon Birch, Dave Collett, and J Andrew Bradley, "A simplified donor risk index for predicting outcome after deceased donor kidney transplantation," *Transplantation*, 2012, *93* (3), 314–318.

Watson, CJE and JH Dark, "Organ transplantation: historical perspective and current practice," *British Journal of Anaesthesia*, 2012, *108* (suppl_1), i29–i42.

Wermers, Russ, "Mutual fund herding and the impact on stock prices," *the Journal of Finance*, 1999, *54* (2), 581–622.

Zhang, Juanjuan, "The sound of silence: Observational learning in the US kidney market," *Marketing Science*, 2010, *29* (2), 315–335.