Regulation of Organ Transplantation and Procurement: A Market Design Lab Experiment

Alex Chan and Alvin E. Roth^{*}

January 29, 2022

Abstract

We conduct a lab experiment that shows current rules regulating transplant centers (TCs) and organ procurement organizations (OPOs) create perverse incentives that inefficiently reduce both organ recovery and beneficial transplantations. We model the decision environment with a 2-player multi-period game between an OPO and a TC. In the condition that simulates current rules, OPOs recover only highest-quality kidneys and forgo valuable recovery opportunities, and TCs decline some beneficial transplants and perform some unnecessary transplants. Alternative regulations that reward TCs and OPOs together for health outcomes in the entire patient pool lead to behaviors that increase organ recovery and appropriate transplants. (*JEL* Codes: C92, D47, I18)

Keywords: Organ Transplantation, Lab Experiment, Market Design, Regulation, Healthcare

^{*}Email addresses: chanalex@stanford.edu (A.C.); alroth@stanford.edu (A.E.R.). Mailing address for A.E.R.: Department of Economics, 344 Landau Economics, Stanford University, Stanford, CA 94305. Fax for A.E.R.: 650-725-5702. Phone for A.E.R.: 650-725-9147. This study was approved by the Institutional Review Board at Stanford University (IRB-44868).

1 Introduction

Deceased donor organ transplantation is in turmoil. Over 110,000 Americans are waiting for an organ transplant and over 5,770 died waiting for an organ in 2019. In the same year, 5,957 recovered organs were discarded. There are multiple Congressional investigations, and active proposals to improve regulations of deceased organ procurement and transplantation, put forth by non-governmental agencies like the Federation of American Scientists and the National Academies of Sciences, Engineering, and Medicine (US Congress (2021), Federation of American Scientists (2021), NASEM et al. (2018), NASEM et al. (2021)). These investigations and proposals look for ways to increase organ transplantation, by increasing recoveries and reducing discard of already recovered organs. The bulk of this discussion had been focused on the regulation and performance measurement of organ procurement organizations (OPOs). Separate discussions have occurred regarding performance monitoring of transplant centers (TCs). Among steps taken to date are a major policy change to remove 1-year graft and patient survival standards as a TC re-certification requirement by the Centers for Medicare and Medicaid Services (CMS) (HHS (2019)). Although OPOs and TCs interact with each other, these regulatory discussions largely considered them separately.¹ Since the incentives and opportunities facing OPOs and TCs are intertwined, such fragmented approaches might be inefficient.

There are two prominent areas where mis-incentives might drive undesirable OPO and TC behavior. First, TCs are penalized (by the Organ Procurement and Transplantation Network [OPTN] Membership and Professional Standards Committee [MPSC]) for adverse health outcomes among patients who the TCs transplant, but not for those who remain untransplanted on the waitlist. So, TCs might be incentivized to restrict transplants to healthier patients using higher quality kidneys even when some sicker patients could have benefited from a lower quality kidney, that is instead discarded. Second, while the TCs are incentivized to be selective with kidneys offered to them by OPOs, the OPOs are incentivized to avoid recovering kidneys that will be declined by TCs. That could induce OPOs to forgo opportunities to recover organs from donors who are unlikely to yield readily accepted organs, especially deceased donors who fall along the margins of

¹The American Society of Transplantation is working with the American Society of Transplant Surgeons to review performance metrics but only focused on TCs (Phend (2020)), while groups like Federation of American Scientists and ORGANIZE are focused on OPOs (Rosenberg et al. (2020)).

what qualifies the deceased as eligible donors.

We use a lab experiment to investigate how a holistic approach that focuses on aligning OPO's and TC's interests to promote population health might offer a better design for organ transplant regulation. We focus on procurement and transplantation of kidneys. We show that moving from the current fragmented regulation to holistic regulation where TCs and OPOs are jointly rewarded for health improvements for all patients (transplanted or waitlisted but not transplanted) can shift the organ recovery and transplantation rates.² We also look at the nature of these shifts to see whether they benefit patients on average and whether the forgone recovery or transplantation opportunities could have benefited patients. While no actual transplantation takes place in our experiment, the decision architecture approximates the environment that OPOs and TCs face. Besides providing clean identification and counterfactuals, our experiment also lets us observe important outcomes that are typically unobservable in the wild. For instance, we are generally unable to observe unrecovered kidneys, or whether discarded organs could have benefited some transplant candidates, but we can observe the analogous outcomes in our experimental setting.

We adopt a simplified representation of the OPO-TC dynamic in a lab setting to represent the interactions between one OPO and one TC where the former decides whether to recover two kidneys from a deceased donor, and the latter (if the kidneys are recovered) gets the option to perform kidney transplants with two kidney transplant candidates (one sicker and riskier, and one healthier and safer).

To simulate the effects of different incentive systems, we compare player behavior across *status quo* and *holistic* regulation conditions. The conditions differ in how payoffs are determined for the players. The instructions to subjects are stated in abstract terms (involving red and blue balls in urns or jars), not in terms of patients and organs. The risks associated with transplanting a particular kidney into a particular patient are based on the risk profiles of both patient and kidney.

Comparing the average behavior of the players across the conditions, we can assess whether there is evidence for shifts in the behaviors representing organ recovery and discards under alternative incentive schemes. Second, we can review the true qualities

 $^{^{2}}$ For convenience, we speak of the waitlist as if it represents the whole pool of transplant-eligible patients. However, we will note in the conclusion that a non-manipulable measure of transplant-eligible patients remains to be constructed.

of the "kidneys" and "patients" to assess whether some forgone recoveries and/or discards could have improved the underlying probabilities of patients getting "good health outcomes." Our experiment also allow us to evaluate the policy impact on health outcomes. We deliberately use the term "good health outcome" to broadly represent desired health outcomes to include not just one-year graft survival but outcomes like more patient quality-adjusted life years (QALYs) or a broader set of outcomes that include improved QALYs across multiple stages of the patients' life-cycle of care as well as financial benefits.

We have several main findings. First, more kidneys are recovered and offered but discards do not rise under the holistic regulation condition, so more kidneys are transplanted. This suggests that the incentives under the current regulatory regime might be inducing OPOs to under-recover kidneys and TCs to over-discard recovered kidneys. Second, there are significant differences between conditions in recovery and discards that are "missed opportunities" that could have helped some patients. Third, there is evidence for gains in good health outcomes under the holistic regulation relative to the status quo. Health benefits are particularly pronounced for the sickest and healthiest transplant candidates - there are more helpful transplants for sick patients and fewer inappropriate transplants for healthier patients.

The problems and limitations of regulatory enforcement of performance standards in organ transplantation using simple metrics have been widely debated in the clinical community.³ While an exhaustive literature review is beyond the scope of this paper, we note that problems with this metric-based regulation approach have been well understood in the transplant literature (Chandraker et al. (2019)). In economics, Stith and Hirth (2016) examine the effects of the 2007 CMS implementation of Conditions of Participation quality standards based on 1-year post-transplant outcomes on TC strategic behavior. More kidney transplant candidates were removed from the waitlist for being "too sick" and fewer kidney transplants were undertaken following a TC's breaching of CMS's quality tolerance band relative to those that did not. However, little is known about whether and how TC post-transplant metric-based regulations and/or OPO regulations that impose a standard on donors per "eligible death referral" affect OPO recovery activities or strategic behaviors.

³Examples of this work include but are not limited to: Schold et al. (2010); Schold et al. (2013a); White et al. (2015); Schold et al. (2019); Schold et al. (2013b); Schold and Howard (2006); Buccini et al. (2014); Weinhandl et al. (2009); Howard et al. (2009); Abecassis et al. (2009); Massie and Segev (2013); Kasiske et al. (2019); Goldfarb (2020).

Our study is the first to evaluate the potential impact of regulations on OPO and TC behavior together, and provide evidence for the potential benefits of a holistic regulatory approach.

Kessler and Roth (2012) and Kessler and Roth (2014) study how changes in the rules governing organ waiting lists might impact potential donors' decision to register as an organ donor.⁴ Here, we propose a novel laboratory game design that can provide a foundation for future research on the market design and regulations surrounding the supply chain for transplantable organs.

The paper is organized as follows. Section 2 provides a brief overview of the policies and incentives facing the US deceased-donor organ transplantation industry. Section 3 describes the experiment and empirical strategy. Section 4 describes the data and presents the key results. Section 5 discusses the implications of this study for the transplantation policy. Section 6 concludes.

2 Background: Organ Transplantation Policy Today

A deceased organ transplantation starts with a death. A hospital must contact its local OPO about every patient's death. This has been required by law since 1998 (Administration (1998), Siminoff et al. (2001)). The OPO will then obtain information to assess whether the deceased is eligible for organ donation. Given limited resources, OPOs have to prioritize which cases to pursue and which ones to pursue first. There are 57 OPOs in the US, each exclusively responsible for recovering organs in their designated donor service area (DSA) (Mone (2002)).⁵

Should an organ be recovered by an OPO, deceased organ allocation would happen through the OPTN platform and offers of the recovered organ will be made to patients on waitlists maintained by TCs based on pre-determined priority rules. TCs do not control the prioritization of patients on their waitlists. What TCs have the freedom to do is to determine whether they want to reject an organ offered for a particular patient on their waitlist, or advise that patient to accept the organ.⁶

 $^{^{4}}$ Those studies were followed by field studies of how such changes played out in Israel (Stoler et al. (2017)).

⁵While Mone (2002) lists 58 OPOs, two OPOs in New England merged since.

⁶TCs can manipulate who is on their waitlist but for simplicity we speak of waitlist and the transplanteligible patient pool interchangeably here. However, we recognize that a patient pool measure that is not easily manipulable by TCs will be necessary for actual policy design: see the concluding discussion.

2.1 Transplant Center Regulations

TC regulation has gone through a few iterations. The Transplantation Amendment Act 1990 led to the first publicly reported transplant center-specific report from United Network for Organ Sharing (UNOS) and OPTN in 1992 (reporting results from October 1, 1987 to December 31, 1989). In 1993, the OPTN/UNOS Board of Directors approved using center-specific reports to "flag" programs for poor performance but it was not until 1997 that OPTN MPSC published a method for "flagging" underperforming transplant programs (Kasiske et al. (2019), Jay and Schold (2017), Chandraker et al. (2019), Luskin and Nathan (2015)).

A major shift in TC incentives came when the 2007 CMS regulations for transplant programs were published in the Final Rules for Approval and Reapproval. Under these rules, a minimum standard was set for 1-year post-transplant patient and graft survival as conditions for center certification and maintenance of funding with CMS. At the same time, the OPTN MPSC published similar but not identical 1-year post-transplant metrics to "flag" TCs. Based on this, the OPTN can designate a TC as a member-not-in-goodstanding and impose costly peer reviews while jeopardizing TCs' center of excellence (CoE) designations with private insurers. A loss of CoE status can spell a loss of privately insured patients and drop in reimbursements. In October 2019, CMS eliminated one-year post-transplant outcome requirements as a condition of Medicare recertification of TCs (HHS (2019)) but OPTN continues to impose outcomes standards. As of the end of 2021, OPTN continues to analyze and publish detailed data on one-year post-transplant performance and can recommend that a TC be shut down for not meeting performance benchmarks (Phend (2020)).

The use of a single metric, one-year post-transplant graft and patient survival, for identifying under-performing transplant programs by various bodies is not without its critics. Reliance on a metric that focuses only on two outcomes of patients selected into transplantation but ignores patients who remain on the waitlist, strongly disincentivizes the use of imperfect organs. Although the MPSC recognizes the need to evaluate multiple phases of transplant care and not just look at post-transplant, short-term outcomes, reforms to this approach to performance management for TCs remain at discussion stages as of late 2021 (OPTN and Network (2021)).

2.2 Organ Procurement Organization Regulations

CMS evaluates and recertifies OPOs every 4 years. Recertification is based on compliance with 3 outcome measures that proceed from the donor conversion metric (how often do "eligible" deceased patients become potential donors) and the yield metric (how often are donated organs actually transplanted). The donor conversion and yield metrics in their current formulation were adopted in 2000.⁷ While no OPO in the last 20 years has been decertified for poor performance, LiveOnNY was threatened with a shutdown in 2014 and then again in mid-2018 due to poor performance scores for nearly a decade and organ recovery rates that were among the lowest in the nation (Kindy and Bernstein (2019)). OPOs can also incur significant costs associated with review and quality improvement initiatives, which can both be triggered by bad performance metrics.

One aspect of these metrics that has been particularly controversial is the "eligible deaths" denominator of the main OPO metrics. "Eligible" referrals are determined to be medically suitable if the patient has been declared brain dead, on a ventilator, and does not have any of a defined list of medical diagnoses that preclude organ transplantation. The number is reported by the OPOs, and some commentators (Goldberg et al. (2017), Rosenberg et al. (2020)) argue that this gives OPOs the room to cherry-pick deceased donors to recover, passing over potential donors whose organs are likely not acceptable to TCs. A 2011 OPTN review of OPOs reports "large inconsistencies and variations in how OPOs reported [eligible deaths] data" (Luskin and Nathan (2015)).

While the degree of manipulability of reported "eligible deaths" numbers remain controversial, it is clear that regulation of OPOs based on the two metrics in their current form can inadvertently offer incentives to under-recover organs. First, "eligible" does not include Donation after Circulatory Determination of Death (DCD), which makes up 15-20% of actual donors today. It also does not include older donors over 75, who could potentially benefit *some* transplant candidates (Aubert et al. (2019)), especially older candidates or candidates who expect a longer wait time for a kidney. Second, the decisions to put a patient on a ventilator and to invest the neurology resource to declare brain death can be influenced by multiple parties, including the OPO. A risk averse decision maker might want to make decisions that lower the odds that a donor on the margin of

⁷The same 2000 legislation also established the DSA monopolies and extended OPO recertification from two to four years to recognize that small OPOs had significant swings in donation potential year to year and to give time for performance improvement efforts to mature.

being acceptable to TCs for transplantation is declared brain dead in time (before their heart stops).

More importantly, reliance on the donation rate and transplantation rate metrics to judge OPOs neglects a key driver of OPO behavior: TCs' propensity to accept and use an organ for transplantation. OPOs have little incentive to recover organs from a deceased donor if those organs will not be accepted by any TC. Regulating OPOs without aligning TC's interests with increased organ recoveries creates fragmented incentives that may lead to inefficiencies.

On November 20, 2020, President Trump's Executive Order on Advancing American Kidney Health led to a new CMS final rule that updates OPO Conditions of Participation (Trump (2019), Lentine and Mannon (2020), HHS (2020)). Beginning in 2022, the "eligible deaths" denominator will be replaced by the number of organ donors in the OPO's DSA as a percentage of inpatient deaths among patients 75 years old or younger with a primary cause of death that is consistent with organ donation. This final rule will also impose stringent re-certification requirements including one that designates OPOs with below-median donation rate and transplantation rate measures as a Tier 3 OPO. Tier 3 OPOs will be decertified and will not be able to compete for any other open DSA contracts.

Whether these new OPO regulations will be constructive is unclear. DCD donors and donors over 75 are still not included in the metrics' denominator and recovery efforts targeting them therefore not directly incentivized. The regulation of TCs remain not directly aligned with that of OPOs.

In summary, as of late 2021, TCs and OPOs face separate incentives. TCs face incentives to get good post-transplant health outcomes for candidates that they can select into receiving a kidney transplant, but no direct incentives to improve the health of those who remain untransplanted on the kidney waitlist. OPOs face incentives to not recover all possible organs, especially organs from older donors who will not enter the denominators of the OPO performance metrics. OPOs might also limit the resources devoted to recovering from donors likely to yield organs of a lower quality that might not meet the acceptance thresholds of risk averse TCs. We simulate the consequences of such incentives in our lab experiment.

3 Experimental Design

We model the decision environment facing OPOs and TCs with a 2-player multi-round game. In the game, urns containing either red or blue balls represent patients, jars containing red or blue balls represent kidneys, the mixing of the balls from a jar into an urn represents a kidney transplantation, blue (red) balls represent a good (bad) health outcome, and the drawing of a ball from an urn (mixed with a jar or not) represents the manifestation of the health outcome.

Each pair of subjects plays 10 rounds of the game. Upon arrival to the game with another subject, each subject is randomly assigned a role as either Player 1 or Player 2. At the beginning of each round, Player 1 will receive a pair of identical jars (which can be thought of as a pair of kidneys from a deceased donor but are presented to subjects only as jars of balls) and Player 2 will receive 2 non-identical urns (representing two different patients on the transplant waitlist). The jars and the urns each have 100 balls at the beginning of each round, each ball is either blue or red. The percentage of balls that are blue in an urn can be thought of as the baseline probability of survival or good health outcome by other measures for a patient. One of Player 2's urns has a lower expected number of blue balls than the other. This could be thought of as the situation where a TC has sicker and healthier patients on the waitlist.

In each round, Player 1 chooses whether to offer his pair of jars to Player 2.⁸ The round ends if Player 1 chooses not to offer Player 2 the pair of jars. Player 1 can only offer two jars or none. If Player 1 offered the pair of jars to Player 2, Player 2 makes a decision to either decline the jars, mix all the balls from one of the two jars into only one of her urns, or mix all the balls from each of the two jars into each of her urns (one jar into each urn). Think of Player 1's decision as a rough representation of an OPO's decision to recover kidneys after getting a referral for a potential eligible deceased donor: offering a pair of jars can be interpreted as declaring a deceased person eligible for donation and recovering their two kidneys. Think of Player 2's decision as a representation of a TC's decision to accept one or both kidneys for transplantation or not. See Figure I for a schematic of the stage game.

⁸Subjects in our experiment may include individuals who identify with different genders across the gender spectrum. For expository purposes, we will use he/him pronouns for player 1 (player representing the OPO), and use she/her pronouns for player 2 (player representing the TC).

Insert Figure I about here.

One ball is drawn from each one of Player 2's urns at the end of each round. Transplants can change the risk of a red ball, by mixing the jar (kidney) with the urn (patient). If Player 2 mixed the balls in a jar into one of her urns, the draw from that urn is made after the balls from the jar were mixed in (ie. the health outcome is assessed posttransplant). This is a highly simplified representation of the manifestation of patient outcomes: if a red ball is drawn from an urn, it is akin to a bad health outcome for the patient represented by that urn.

3.1 Information Available to Players

We are modeling the OPO's information at the point of deciding whether to recover kidneys from a deceased person, and the TC's information after examining the patients and the recovered kidneys.

It is common knowledge to both players that there are 2 types of urns ("High Blue" and "Low Blue") and 2 types of jars ("Low Quality" and "High Quality"). Player 2 will have one urn (patient) of each type in every round while Player 1 will get a pair of identical jars.

In each round, Player 1 will not be able to observe the exact number of blue balls in each of the jars that he received but will be told whether he received a pair of "High Quality" or "Low Quality" jars for that round. A "High Quality" jar has a number of blue balls (out of 100) that is drawn from a uniform distribution U[70, 100] whereas a "Low Quality" jar has a number of blue balls (out of 100) that is drawn from a uniform distribution U[0, 70]. Player 1 is also aware that Player 2 has an urn of each type. He knows that a "High Blue" urn has a number of blue balls (out of 100) that is drawn from a uniform distribution U[40, 100], and a "Low Blue" urn has a number of blue balls (out of 100) that is drawn from a uniform distribution U[0, 60]. While he knows that Player 2 will have an urn of each type, he will neither observe nor receive any signals about the exact number of blue balls of either of her urns.

Unlike Player 1, Player 2 can see the actual number of blue balls in the jars as well as the actual number of blue balls in each of her urns before making any decisions in each round. This is a simplified way to represent the knowledge that a TC has about the transplant candidates and kidneys already recovered from a deceased donor and offered to the TC.

It is common knowledge to both players that Player 1 can only observe jar and urn types while Player 2 can observe the exact number of blue and red balls in each jar and urn.

The actual instruction screens are in the online Appendix.

3.2 Control and Treatment Conditions

Pairs of subjects are randomly assigned at the beginning of the game to have either a payoff scheme that represents the status quo rules governing OPOs and TCs or a payoff scheme that represents our proposed holistic regulatory rules. The former rewards (penalizes) OPOs for TC-accepted (TC-declined) kidneys from deceased donors declared by OPOs to be eligible and rewards (penalizes) TCs for good (bad) patient health outcomes for transplants that TCs chose to carry out by accepting a kidney. In contrast, the proposed holistic approach emphasizes aligning the interests of OPOs and TCs by rewarding each of them for improvements in good health outcomes among both patients selected for a transplantation and patients who remain on the waitlist.

In the status quo condition, subjects are paid based on separate and different payoff schemes. For each round, Player 1 earns \$0.10 by offering two jars that were subsequently both accepted by Player 2, gets a \$0.30 penalty by offering two jars that were subsequently both declined by Player 2, and gets \$0.00 by either not offering the jars at all or offering two jars when only one jar was accepted by Player 2. For each round, Player 2 earns nothing if Player 1 did not offer the jars or if she declined both jars, earns \$0.25 by mixing a jar with an urn and drawing a blue ball from the mixed urn, and gets a \$1.00 penalty instead if a red ball was drawn from the mixed urn.⁹ Under these payoffs where Player 2 is paid only if she mixed jars and urns, it is sometimes income-maximizing for Player 2 to mix a jar with fewer blue balls (lower quality) than an urn as long as the mixed urn still yields a high enough chance to draw a blue ball (notice that this "makes an urn worse" and can be thought of as an inappropriate transplant: the patient's expected health outcome would be better in the absence of a transplant.).

 $^{^{9}}$ In other words, Player 2 gets a \$0.50 reward for mixing jars into both urns and drawing blue balls from both, a \$0.75 penalty for mixing jars into both urns and drawing a blue ball from one and a red ball from the other, and a \$2.00 penalty for mixing jars into both urns and drawing red balls from both.

In the *holistic* condition, both subjects are paid based on the number of blue and red balls drawn from both urns in each round. For each round, Player 1 earns \$0.16 for each blue ball drawn from the urns whether mixing happened or not. He gets a \$0.08 penalty for each red ball drawn from the urns. Player 2 earns \$0.20 for each blue ball drawn from the urns whether mixing happened or not. She gets a \$0.10 penalty for each red ball drawn from the urns.

Note that in the status quo condition only the health outcomes of transplanted patients are used for the determination of payoffs, while in the holistic condition health outcomes of both patients are measured every period.

Subjects are told exactly how many rounds they would play the game, as well as the payoff scheme for both themselves and the other player with whom they are playing the 10 rounds of the game. At the end of each round, subjects are informed of the payoffs they received for that round as well as what the other player received for that round. Both players are presented with a table to help them keep track of actions taken by both players and their respective earnings from each of the previous and current rounds.

If all players are maximizing expected income perfectly and it is common knowledge that they do, we would expect that very few transplants would be performed in the status quo condition and no organs would be recovered as a result.¹⁰ This would not be true in the holistic condition where expected income maximizers will transplant every jar that has more blue balls than an urn into that urn, and every pair of jars will be recovered and offered.¹¹ But based on results from our pilots described in footnote 14, we see much higher rates of transplantation (mixing) than expected under the status quo condition.¹² We are concentrating on the comparative statics results comparing between conditions in this experiment, so we considered the bias towards action (to transplant) when we set these payoff levels.

¹⁰An expected income maximizing Player 2 (TC) will mix a jar into a High Blue urn less than 42% of the time and into a Low Blue urn less than 2% of the time. Player 2 will mix both jars less than 1% of the time, making the expected income of Player 1 strictly negative whenever he recovers a pair of jars as he could earn at most \$0.00 but risks losing \$0.30 over half the time. However, as we will discuss in footnote 15, a Player 1 who is perfectly maximizing expected income and responding to Player 2's actual acceptance behavior in the experiment should recover and offer all High Quality Jars even though he should not recover Low Quality Jars.

¹¹Any opportunity to improve the odds of drawing a blue ball from an urn will (weakly) increase the expected income for both players.

¹²This bias towards action bears some resemblance to the propensities of actual TCs and OPOs to perform transplants and recover organs. This had translated to risky actions (e.g. risky transplants) that led to bad outcomes and occasional penalties from UNOS, payers and other regulatory bodies.

All experimental sessions were conducted online using US-based subjects recruited via the Prolific platform. Batches of subjects were given a 10-minute window to show up and participate in the game. Subjects' basic demographic information were collected as they arrived, and then admitted to a waiting room. They are then paired off based on arrival time to the waiting room to play 10 rounds of the game.

4 Data and Results

4.1 Sample Population and Check for Balance

The experimental results are from 324 subjects who participated in pairs in one of 162 sessions in the Summer of 2021.¹³ There were 83 sessions under the holistic condition and 79 under the status quo condition. The experiment lasted up to 59:57 minutes and average earnings were \$5.34 per subject, in addition to a \$5.00 show-up fee. The experiment was conducted using software written in JavaScript and hosted on AWS cloud environment. Subjects were recruited through Prolific and redirected to the game website after completing a brief Qualtrics survey on their demographic background. Payoffs and actions are recorded in a PostgreSQL relational database that can be downloaded as csv files.

Table I presents summary statistics, with the full sample in columns 1 and 2. The last column of Table I investigates the balance between the treatment groups. Overall, balance is achieved across subject demographics (with the exception of the age groups 35-44 and 45-54) as well as the number of blue balls in the jars and urns that players encounter.¹⁴

Insert Table I about here.

Figure II presents descriptive statistics that compares the key outcomes of the status

 $^{^{13}164}$ sessions were completed but one subject participated in 2 different sessions. The results in the paper uses the sample where this subject and the 2 players who played against them were dropped. Including the results from these 2 dropped sessions does not change the results in this paper.

¹⁴Before the actual run of our experiment, we piloted variants of this experiment where the OPO only gets 1 jar for each round between June 15, 2021 and July 13, 2021. The raw data and results from regressions analogous to the ones reported in this paper are available upon request. The results on recovery, discards, and transplant rate are similar to the final experiment reported in this paper. The key difference are (1) with only 1 jar to use, the players cannot improve the final "health outcomes" as prominently as when there are 2 jars per round; and (2) within the 1-jar variants in the pilot stage, we varied the levels of the incentives but the behavior remain largely similar in the signs of the effects.

quo and holistic conditions.

Insert Figure II about here.

4.2 Recovery, Discards, Transplants

The first pair of bars of Figure II displays the average rates of jar recovery and offering by Player 1. Overall, Player 1s recover 17.4 percentage points fewer jars under status quo compared to holistic regulations: 58.6% and 76.0% of the jars are recovered under status quo and holistic conditions respectively.¹⁵ The first column of Table II shows the results unconditional on jar quality type while the second and third column show results conditional on jar quality type.¹⁶ The difference is most stark for low quality (fewer blue balls on average) jars: 29.9 percentage points fewer jars under status quo. Both of these results are significant. However, there is no statistically significant¹⁷ difference in recovery rates when the jars are high quality jars. Player 1s recover similarly when the observed jar quality type is high (number of blue balls), but the recovery behavior diverges when the observed jar quality type is low.

Insert Table II about here.

The second pair of bars of Figure II and columns 4 to 6 of Table II show the results on instances where a pair of jars could have benefited at least one urn but was not recovered. Under the status quo condition, Player 1 "missed recovery opportunities" 26.2% of the time while the same is 15.7% only under the holistic condition. The differences are significantly higher overall and when jars are of a low quality type, by 10.5 percentage points and 16.6 percentage points respectively (columns 4 and 5 of Table II). This indicates that one out of four times, it would been better for the Player 1 (the OPO) to recover the jars ("kidneys") for the objective of increasing the expected number of blue balls drawn from urns (better population health outcomes) but he did not under status quo.

¹⁷At conventional levels.

¹⁵In the data, Player 2s accepted both jars 42.0% and rejected both jars 5.0% of the time when the jar type is High Quality, and accepted both jars 16.8% and rejected both jars 39.2% of the time when the jar type is Low Quality. If a Player 1 is maximizing expected income with expectation of jar acceptance that match these summary statistics, he will recover the jars 49.1% of the time under the status quo condition. As mentioned earlier, he will recover 100.0% of the time under the holistic condition.

¹⁶Player 1s, the OPO players, can only observe jar quality type when they make the recovery decision but not actual quality (number of blue balls) Therefore, we are only controlling for jar quality types.

The third pair of bars of Figure II and the first column of Table III show that the average quality (average number of blue balls) of recovered and offered jars is higher under the status quo condition. The average recovered jar has 65.1 blue balls (out of 100) under the holistic condition while the average recovered jar has 71.0 blue balls. This is because of the cherry picking done by Player 1s as evidenced by the lower recovery rates above: average quality of recovered is similar if we only look at jars that are ex ante high quality typed.

Insert Table III about here.

The fourth pair of bars of Figure II and columns 2 to 4 of Table III display the results on jar discards by Player 2 when Player 1 offered her a jar. Despite the lower average quality of offered jars under the holistic regulations condition, Player 2s are not any more likely to discard jars than in status quo (see columns 2 to 4 of Table III). The threshold for accepting a jar to carry out a transplant is significantly lower under the holistic regulations condition as indicated by 7.0 fewer blue balls in the average post-transplant urn (see column 3 of Supplementary Table A4). The player representing TCs accepts "Low Quality" jars 22.1 percentage points less often when offered jars under status quo (Supplementary Table A4 column 6). More generally, she is also accepting jars less aggressively for transplants that increase the percentage of blue balls in an urn after transplantation in the status quo condition - the average jar (kidney) accepted for such transplants has 4.8 more blue balls under status quo (Supplementary Table A4 column 5). Under the holistic condition, they are more aggressively transplanting patients who are sicker on the baseline (urns transplanted under status quo has 13 more blue balls before transplantation; see Supplementary Table A4). Furthermore, the fifth pair of bars of Figure II and columns 5 to 7 in Table III show that Player 2s (TC) discard jars that could have helped the blue-ball-odds of an urn in a level that is 23.8 percentage points more under status quo condition (49.8%) than under holistic regulations (26.0%).

The sixth pair of bars of Figure II and Table IV displays the results on mixing ("transplants") by Player 2. The first 3 columns in Table IV shows the results on mixing rate. On average, Player 2s mix 35.5% of the urns under status quo and 44.3% of the urns under holistic conditions.¹⁸ In other words, Player 2s mix about 8.8 percentage points

 $^{^{18}}$ Conditional on the actual recoveries and offers by Player 1 observed in the data for each round of

less under status quo compared to holistic regulations. These results are significant at conventional levels, controlling for jar/urn quality or not (see Table IV).

Insert Table IV about here.

Conditional on jars being recovered and offered by Player 1, Player 2 can also harm the odds of a good outcome for an urn by mixing a jar with fewer blue balls than that urn into the urn ("harmful transplants"). The seventh pair of bars of Figure II and columns 4 to 6 in Table IV show that Player 2s conduct many more harmful transplants under the status quo condition (22.9%) than under holistic regulations (8.3%). These results are significant at conventional levels, controlling for jar/urn quality or not (see Table IV).

4.3 Impact on "Health Outcomes"

Based on the number of blue and red balls in the urns at the end of each round,¹⁹ the expected number of red balls drawn is 46.2% and 42.1% for status quo and holistic conditions respectively (last pair of bars of Figure II).²⁰ In other words, we expect that a red ball or a bad outcome is 4.1 percentage points more likely to be drawn from an urn under the status quo condition compared to under holistic regulations (see column 1 of Table V). In the actual draw in our experiment, 46.2% and 42.8% of the balls drawn are red under status quo and holistic conditions (see Supplementary Table A1).²¹ The largest gains in health outcomes under holistic regulations are expected to come from the Low Blue urns (representing sicker patients) (see columns 2 and 3 in Table V).

each game, a perfectly expected income maximizing Player 2 would have mixed 9% of the urns under status quo and 48% of the urns under holistic conditions. The bias for action is statistically significant in the status quo condition but not in the holistic condition: the percentage of urns mixed is statistically similar at conventional levels between the observed and perfectly-expected-income-maximizing-Player-2-level conditional on actual recoveries for holistic but not for status quo.

¹⁹After mixing if mixing occurred. The exercise to assess the actual number of blue and red balls in the urn at the end of round can help us evaluate the limit bad outcome rate for if we repeated this experiment many times. This metric is arguably more relevant for the purposes to assess whether the incentive systems induced differences in behavior.

²⁰Conditional on the actual recoveries and offers by Player 1 observed in the data for each round of each game, a perfectly expected income maximizing Player 2 will mix in a way such that 41.6% and 38.8% of the balls drawn are expected to be red under status quo and holistic conditions respectively. While the levels of bad outcome rates are higher in the data than if the Player 2s are perfectly expected income maximizers, the comparative statics results are similar.

 $^{^{21}}$ The difference of 3.4 percentage points is significant at the 10% level. Since we observe the actual lotteries associated with the drawing of a ball from each urn, the observed outcome is a noisy signal of the actual expected health outcomes compared to the exact expected value conditional on the realized urns. Therefore, we do not have to rely on the actual draws to understand the impact on expected health outcome.

Insert Table V about here.

The difference in bad outcome rate is most notable among the urns representing the healthiest and sickest patients. Looking only at rounds where at least one urn has more than 90% blue balls or less than 10% blue balls, a red ball is expected to be 7.0 percentage points more likely to be drawn from an urn under the status quo condition compared to under holistic condition (see column 4 of Table V).²² Bad outcome differences are expected to be only 2.6 percentage points more likely under the status quo condition compared to under holistic regulations in rounds where neither urns have between 10% to 90% blue balls (see column 7 of Table V).²³

The results on missed recoveries, bad discards, and bad mixings (transplants) offer evidence for efficiency loss. The results on red ball rates tell us that holistic regulations can benefit the "well-being" of the overall population, and this benefit will mostly come from improvements from the "sickest" and "healthiest."

5 Discussion

Many factors and decision-makers that drive access and outcomes of transplantation act outside the walls of transplant centers. Better coordination between OPOs and TCs might help reduce instances where organs are discarded while patients perished on the waitlist. With better coordination, TCs can strategically work up patients who are far down the waitlist so that they can swiftly respond and receive kidneys that are declined by patients higher up on the waitlist without delays that could push the organ's cold time beyond the acceptable range. The 2019 Advancing American Kidney Health (AAKH) initiatives begin to address the fact that multiple players influence performance and offer novel opportunities for collaboration between historically insular stakeholders like TCs and dialysis centers (by financially incentivizing dialysis centers to refer more patients to transplantation). This should also involve both public and private payers who will need to introduce non-conflicting payment models that motivates cost-effective QALY-improving

 $^{^{22}}$ In the actual draw in our experiment, a red ball is 5 – 7 percentage points more likely to be drawn from an urn at rounds where at least one urn has more than 90% blue balls or less than 10% blue balls under the status quo condition compared to under holistic regulations (see Supplementary Table A2).

 $^{^{23}}$ In the actual draw, bad outcome differences are not significant (at conventional levels) in rounds where neither urns have between 10% to 90% blue balls. See Supplementary Table A3.

transplants. But integrated decision-making is hampered by the fact that organ policy remains splintered between different federal agencies and UNOS.

6 Conclusion

Current transplantation policy gives rise to adverse incentives for OPOs to under-recover kidneys while incentivizing TCs to cherry-pick kidneys and patients. Using a laboratory experiment to model the decision architecture facing OPOs and TCs, we show that status quo regulations present incentives for TCs and OPOs to behave in a way not favorable to maximize access to transplantation. In particular, TCs missed opportunities to transplant that could have improved the health of the patients and discard recovered organs. TCs might also conduct transplants that are not beneficial when they are paid to do more transplants rather than to maximize the health outcomes of patients. In the same time, OPOs recover fewer organs.

Holistic regulation that align OPO and TC interests by rewarding them based on the health outcomes of the entire patient pool led to more organ recoveries by OPOs, higher utilization of organs by TCs who transplant sicker patients, and more appropriate transplantation by TCs.

These findings suggest that we need to move beyond current discussions on the revision of isolated metrics and their enforcement, but start to reformulate the regulation conversation to consider the TCs and OPOs together. Our experiment illustrate the benefits of an alternative, holistic regulatory approach but does not outline the details of implementation and the precise way a patient pool should be identified and attributed to different TCs and OPOs. Future work should translate the holistic regulatory approach into actual legislations and rules. For instance, we will need to develop a non-manipulable measure of transplant-eligible patients (distinct from currently TC-managed waitlists) and a logic to attribute patients to specific TCs.²⁴ Nonetheless, insights from this study should point

²⁴Patient attribution to providers belongs to a broader, on-going policy discussion in healthcare. For instance, patients are attributed to an accountable care organization (ACOs) based on their patterns of primary care use, and each ACO is held accountable for the cost of all services received by the patients attributed to them, even those received outside the ACO. ACOs are also often measured and partially paid based on health outcomes for patients attributed to them (McClellan et al. (2010)). Even after over a decade, ACO patient attribution remains imperfect and prone to strategic manipulation. Interviews with industry experts reveal that there are private sector consultants who help ACOs strategically construct the most profitable patient panels under the terms of their respective payment contracts. Future work that identify a feasible attribution logic will be pivotal to the implementation of a holistic regulatory

regulatory bodies to take note of the importance of aligned, holistic incentives.

STANFORD UNIVERSITY STANFORD UNIVERSITY AND NATIONAL BUREAU OF ECONOMIC RESEARCH

References

- Abecassis, Michael Messod, R Burke, GB Klintmalm, Arthur J Matas, RM Merion, D Millman, K Olthoff, JP Roberts, and American Society of Transplant Surgeons, "American Society of Transplant Surgeons transplant center outcomes requirements—a threat to innovation," American Journal of Transplantation, 2009, 9 (6), 1279–1286.
- Administration, Health Care Financing, "Medicare and Medicaid programs: hospital conditions of participation; identification of potential organ, tissue and eye donors and transplant hospitals' provision of transplant related data," *Federal Register*, 1998, 63 (119), 33856.
- Aubert, Olivier, Peter P Reese, Benoit Audry, Yassine Bouatou, Marc Raynaud, Denis Viglietti, Christophe Legendre, Denis Glotz, Jean-Phillipe Empana, Xavier Jouven et al., "Disparities in acceptance of deceased donor kidneys between the United States and France and estimated effects of increased US acceptance," JAMA internal medicine, 2019, 179 (10), 1365–1374.
- Buccini, LD, DL Segev, J Fung, C Miller, D Kelly, C Quintini, and JD Schold, "Association between liver transplant center performance evaluations and transplant volume," *American Journal of Transplantation*, 2014, 14 (9), 2097–2105.
- Chandraker, Anil, Kenneth A Andreoni, Robert S Gaston, John Gill, Jayme E Locke, Amit K Mathur, Douglas J Norman, Rachel E Patzer, Abbas Rana, Lloyd E Ratner et al., "Time for reform in transplant program–specific reporting: AST/ASTS transplant metrics taskforce," American Journal of Transplantation, 2019, 19 (7), 1888–1895.

approach for organ transplantation.

- **Congress, USA US**, "The Urgent Need to Reform the Organ Transplantation System to Secure More Organs for Waiting, Ailing, and Dying Patients," 05 2021.
- for Medicare Medicaid Services (CMS) HHS, Centers, "Medicare and Medicaid Programs; Regulatory Provisions to Promote Program Efficiency, Transparency, and Burden Reduction; Fire Safety Requirements for Certain Dialysis," 09 2019.
- ____, "Medicare and Medicaid Programs; Organ Procurement Organizations Conditions for Coverage: Revisions to the Outcome Measure Requirements for Organ Procurement Organizations," 12 2020.
- Goldberg, D, MJ Kallan, L Fu, M Ciccarone, J Ramirez, P Rosenberg, J Arnold, G Segal, KP Moritsugu, H Nathan et al., "Changing metrics of organ procurement organization performance in order to increase organ donation rates in the United States," American Journal of Transplantation, 2017, 17 (12), 3183–3192.
- Goldfarb, David A, "Re: Time for Reform in Transplant Program-Specific Reporting: AST/ASTS Transplant Metrics Taskforce," *The Journal of urology*, 2020, *203* (1), 20–21.
- Howard, Richard J, Danielle L Cornell, and Jesse D Schold, "CMS oversight, OPOs and transplant centers and the law of unintended consequences," *Clinical transplantation*, 2009, 23 (6), 778–783.
- Jay, Colleen and Jesse D Schold, "Measuring transplant center performance: the goals are not controversial but the methods and consequences can be," *Current transplantation reports*, 2017, 4 (1), 52–58.
- Kasiske, Bertram L, Nicholas Salkowski, Andrew Wey, Ajay K Israni, and Jon J Snyder, "Scientific Registry of Transplant Recipients program-specific reports: where we have been and where we are going," *Current opinion in organ transplantation*, 2019, 24 (1), 58–63.
- Kessler, Judd B and Alvin E Roth, "Organ allocation policy and the decision to donate," American Economic Review, 2012, 102 (5), 2018–47.
- and _ , "Loopholes undermine donation: An experiment motivated by an organ donation priority loophole in Israel," *Journal of Public Economics*, 2014, 114, 19–28.

- Kindy, K and L Bernstein, "Despite low performance, organ collection group gets new federal contract," *Washington Post*, 2019.
- Lentine, Krista L and Roslyn B Mannon, "The Advancing American Kidney Health (AAKH) executive order: promise and caveats for expanding access to kidney transplantation," *Kidney360*, 2020, 1 (6), 557.
- Luskin, R and H Nathan, "Eligible death statistic: not a true measure of OPO performance nor the potential to increase transplantation," 2015.
- Massie, Allan B and DL Segev, "Rates of false flagging due to statistical artifact in CMS evaluations of transplant programs: Results of a stochastic simulation," American Journal of Transplantation, 2013, 13 (8), 2044–2051.
- McClellan, Mark, Aaron N McKethan, Julie L Lewis, Joachim Roski, and Elliott S Fisher, "A national strategy to put accountable care into practice," *Health Affairs*, 2010, 29 (5), 982–990.
- Mone, Thomas D, "The business of organ procurement," Current Opinion in Organ Transplantation, 2002, 7 (1), 60–64.
- of American Scientists, FAS Federation, "FAS Organ Procurement Organization Innovation Cohort Shares Data to Advance Organ Recovery Research," 10 2021.
- of Sciences, Engineering NASEM National Academies, Medicine et al., "Opportunities for organ donor intervention research: Saving lives by improving the quality and quantity of organs for transplantation," 2018.
- _ , _ , and _ , "A Fairer and More Equitable, Cost-Effective, and Transparent System of Donor Organ Procurement, Allocation, and Distribution - A Virtual Workshop," 2021.
- **OPTN, Organ Procurement and Transplantation Network**, "Enhance Transplant Program Performance Monitoring System OPTN Membership and Professional Standards Committee," 2021.
- Phend, Crystal, "Easing Transplant Outcome Scrutiny: A Small but Welcome Step," 01 2020.

- Rosenberg, Paul, Michael Ciccarone, Bradley Seeman, David Washer, Gayle Martin, Jackson Tse, and Erica Lezama, "Transforming Organ Donation in America," *Bridgespan. org*, 2020.
- Schold, JD and RJ Howard, "Prediction models assessing transplant center performance: can a little knowledge be a dangerous thing?," American Journal of Transplantation, 2006, 6 (2), 245–246.
- __, LD Buccini, ELG Heaphy, DA Goldfarb, AR Sehgal, J Fung, ED Poggio, and MW Kattan, "The prognostic value of kidney transplant center report cards," *American Journal of Transplantation*, 2013, 13 (7), 1703–1712.
- Schold, Jesse D, Charlotte J Arrington, and Greg Levine, "Significant alterations in reported clinical practice associated with increased oversight of organ transplant center performance," *Progress in Transplantation*, 2010, 20 (3), 279–287.
- _ , LD Buccini, TR Srinivas, RT Srinivas, ED Poggio, SM Flechner, C Soria, DL Segev, J Fung, and DA Goldfarb, "The association of center performance evaluations and kidney transplant volume in the United States," American Journal of Transplantation, 2013, 13 (1), 67–75.
- -, Rachel E Patzer, Timothy L Pruett, and Sumit Mohan, "Quality metrics in kidney transplantation: current landscape, trials and tribulations, lessons learned, and a call for reform," *American Journal of Kidney Diseases*, 2019, 74 (3), 382–389.
- Siminoff, Laura A, Robert M Arnold, and Joan Hewlett, "The process of organ donation and its effect on consent," *Clinical Transplantation*, 2001, 15 (1), 39–47.
- Stith, Sarah S and Richard A Hirth, "The effect of performance standards on health care provider behavior: evidence from kidney transplantation," *Journal of Economics* & Management Strategy, 2016, 25 (4), 789–825.
- Stoler, Avraham, Judd B Kessler, Tamar Ashkenazi, Alvin E Roth, and Jacob Lavee, "Incentivizing organ donor registrations with organ allocation priority," *Health Economics*, 2017, 26 (4), 500–510.

- Trump, DJ, "Executive order on advancing American kidney health," The White House https://www. whitehouse. gov/presidential-actions/executive-order-advancingamerican-kidney-health, 2019.
- Weinhandl, Eric D, JJ Snyder, AK Israni, and Bert L Kasiske, "Effect of comorbidity adjustment on CMS criteria for kidney transplant center performance," American Journal of Transplantation, 2009, 9 (3), 506–516.
- White, Sarah L, Dawn M Zinsser, Matthew Paul, Gregory N Levine, Tempie Shearon, Valarie B Ashby, John C Magee, Yi Li, and Alan B Leichtman, "Patient selection and volume in the era surrounding implementation of Medicare conditions of participation for transplant programs," *Health services research*, 2015, 50 (2), 330–350.



Figure I: Stage Game for the OPO Player (P1) and TC Player (P2)

Notes: This figure presents the flow of the stage game in this experiment. Each pair of players play this stage game 10 times. The jars (with the handle) can be thought of as kidneys that can be recovered and offered. The urns can be thought of as transplant candidates. The chance of drawing a blue ball from an urn can be thought of as the odds of the patient represented by that urn getting a good health outcome.



Figure II: Outcome Variable Comparison

Notes: This figure reports the average levels of the outcome variables of interest, by treatment condition, of the 324 subjects in the main sample. "Jar Recovery" is the percentage of the pairs of jars that are recovered and offered by Player 1. "Missed Recovery" is the percentage of the pairs of jars that could have improved the odds of a good outcome (drawing a blue ball) of at least one urn but are NOT recovered/offered by Player 1. "Offered Jar Quality" is the percentage of balls that are blue in each of the recovered and offered jars, "Discard" is the percentage of recovered jars that are discarded/rejected. "Bad Discard" is the percentage of discarded jars that has more blue balls than at least one urn (benefits at least one urn). 'Transplant" is the percentage of urns where a mixing happened. "Harmful Transplant" is the percentage of mixings that occurred which made the odds of blue balls worse for an urn. "Expected Bad Outcome" is the proportion of red balls in an urn at the end of a round. Table shows averages ("mean") by treatment condition.

	То	tal	Holi	istic	Statu	s Quo	
	mean	s.d.	mean	s.d.	mean	s.d.	Difference p -value
Female	0.45	0.50	0.49	0.50	0.41	0.49	0.17
White	0.78	0.41	0.77	0.43	0.80	0.40	0.48
Black	0.09	0.28	0.08	0.27	0.09	0.29	0.60
Asian	0.07	0.26	0.09	0.29	0.05	0.22	0.16
Other Race	0.06	0.24	0.07	0.25	0.06	0.23	0.73
Hispanic	0.06	0.24	0.05	0.21	0.07	0.26	0.41
Employed Full-time	0.55	0.50	0.53	0.50	0.57	0.50	0.48
Unemployed	0.07	0.25	0.07	0.26	0.06	0.23	0.58
College Graduate	0.60	0.49	0.62	0.49	0.58	0.50	0.42
High School Graduate	0.99	0.08	1.00	0.00	0.99	0.11	0.15
Age 18-24	0.04	0.19	0.03	0.17	0.04	0.21	0.50
Age 25-34	0.33	0.47	0.36	0.48	0.29	0.46	0.18
Age 35-44	0.29	0.45	0.25	0.43	0.34	0.47	0.08^{*}
Age 45-54	0.18	0.38	0.22	0.41	0.14	0.35	0.07^{*}
Age 55-64	0.11	0.32	0.11	0.31	0.12	0.33	0.74
Age 65-74	0.04	0.20	0.02	0.15	0.06	0.23	0.13
Age 75-84	0.01	0.11	0.01	0.11	0.01	0.11	0.96
# Blue Balls in Jars	59.00	30.24	59.88	31.00	57.32	29.56	0.45
# Blue Balls in Urns	51.00	11.81	51.28	11.00	51.09	12.61	0.88
# Subjects	324		166		158		
N	$3,\!240$		$1,\!660$		$1,\!580$		

Table I: Summary Statistics and Balance Checks

Notes: This table reports the background characteristics of the 324 subjects in the main sample, pooled and by treatment group. "Female" indicates the share of female sex; "White," "Black," "Asian," "Hispanic," and "Other Race" indicate the shares of subjects belonging to each of these categories. Age data was recorded in intervals, each one of the age categories indicate shares of subjects in these age buckets. "High School Graduate" indicate the share of subjects who did not graduate from high school, "College Graduate" indicate indicate the share of subjects who reported that they have a bachelor's or advanced degree. "Employed" indicate the share who are employed full-time, while "Unemployed" indicates the share of subjects who selected are not employed either full-time or part-time. "# Blue Balls in Jars," "# Blue Balls in High Blue Urn," and "# Blue Balls in Low Blue Urn" indicate the average number of blue balls (out of 100 balls that are either blue or red) in the jars and urns encountered by the subjects during the game. Table shows averages ("mean") and standard deviations ("s.d."). The Difference p-value column reports the p-value for the test of equality between the treatment and control groups. Stars indicate whether this difference is significant.

	(1)	(2)	(3)	(4)	(5)	(6)
		Jar Recovery	Rate	Mi	ssed Beneficial	Recovery
	All	Low Quality	High Quality	All	Low Quality	High Quality
Status Quo	-0.174^{**}	-0.299**	-0.042	0.105^{**}	0.166^{**}	0.042
	(0.035)	(0.055)	(0.038)	(0.026)	(0.036)	(0.038)
Constant	0.760**	0.610^{**}	0.913^{**}	0.157^{**}	0.225**	0.0874**
	(0.024)	(0.041)	(0.022)	(0.017)	(0.024)	(0.022)
N	1620	820	800	1620	820	800
R^2	0.035	0.090	0.004	0.017	0.032	0.004

Table II: Impact on Jar Recovery Rates and Missed Opportunities for Beneficial Recoveries

Standard errors (Robust, clustered by player-pairings) in parentheses $p^+ p < 0.1$, * p < 0.05, ** p < 0.01

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 Status Quo_i + \epsilon_i$. Player 1 (OPO) can either recover a pair of jars or not (1=recover/offer jar; 0=not recover/offer jar). The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. The 1st and 4th columns reports results unconditional on jar quality type, the 2nd and 5th columns reports results conditional on the jar quality, and the 3rd and 6th columns reports results conditional on the jar quality being high quality. "Jar Recovery Rate" is the percentage of the pairs of jars that are recovered and offered by Player 1. "Missed Beneficial Recovery" is the percentage of the pairs of jars that could have improved the odds of a good outcome (drawing a blue ball) of at least one urn but are NOT recovered/offered by Player 1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Jar # Blue Recovered	Jar Disc	ard Rate	Recovered	% c	of Bad Disc	eards
Status Quo	5.985**	-0.022	0.020	0.024	0.238^{**}	0.143^{**}	0.060^{*}
	(2.106)	(0.029)	(0.027)	(0.027)	(0.046)	(0.039)	(0.030)
Jar # Blue Balls			-0.007**	-0.007**		0.008^{**}	0.010^{**}
			(0.000)	(0.000)		(0.001)	(0.000)
High Urn # Blue Balls				0.003**			-0.013**
8 //				(0.001)			(0.001)
Low Urn $\#$ Blue Balls				0.000			-0.002*
				(0.001)			(0.001)
Constant	65 060**	0 /10**	0 070**	0 696**	0.960**	0 101**	0 777**
Constant	(1, 270)	(0.418)	(0.010)	(0.050)	(0.200)	-0.191	(0.072)
	(1.379)	(0.021)	(0.030)	(0.054)	(0.032)	(0.021)	(0.053)
N	1094	1094	1094	1094	702	702	702
R^2	0.012	0.001	0.300	0.317	0.064	0.323	0.564

Table III: Impact on Recovered Jar Quality, Discard Rates, and Discards that could have Benefited Urn(s)

Standard errors (Robust, clustered by player-pairings) in parentheses

+ p < 0.1, * p < 0.05, ** p < 0.01

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + X_i\gamma + \epsilon_i$. Player 2 (TC) can either discard 0, 1 or 2 jars *if a pair of jars were offered by Player 1*. The results here are conditional on Player 1 having recovered jars and made an offer to Player 2. The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. "Jar # Blue Balls" is the number of blue balls in each of the jars. "High Urn # Blue Balls" is the number of blue balls in the high blue urn. "Low Urn # Blue Balls" is the number of blue balls in the low blue urn. "Jar # Blue | Recovered" is the number of blue balls in each of the recovered jars that are discarded/rejected. "% of Bad Discards" is the percentage of discarded jars that has more blue balls than at least one urn (benefits at least one urn).

	(1)	(2)	(3)	(4)	(5)	(6)
	Mixing	(Transplan	nt) Rate	% Mixin	igs Made U	Irn Worse
Status Quo	-0.088**	-0.087**	-0.088**	0.146^{**}	0.166^{**}	0.163^{**}
	(0.028)	(0.025)	(0.025)	(0.030)	(0.023)	(0.023)
Jar # Blue Balls		0.008**	0.008**		-0.008**	-0.008**
		(0.000)	(0.000)		(0.001)	(0.001)
High Urn $\#$ Blue Balls			-0.001**			0.004^{**}
			(0.001)			(0.001)
Low Urn $\#$ Blue Balls			-0.000			0.002**
			(0.000)			(0.001)
Constant	0.443**	-0.018	0.086^{+}	0.083**	0.685^{**}	0.331^{**}
	(0.020)	(0.023)	(0.047)	(0.016)	(0.056)	(0.062)
N	1620	1620	1620	903	903	903
R^2	0.012	0.332	0.336	0.042	0.297	0.352

Table IV: Impact on Mixing ("Transplant") Rate, and Mixings that Gave an Urn Worse Odds for Blue

Standard errors (Robust, clustered by player-pairings) in parentheses

+ p < 0.1, * p < 0.05, ** p < 0.01

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 Status Quo_i + X_i\gamma + \epsilon_i$. Player 2 (TC) can either mix 0, 1 or 2 jars into the urns *if a pair of jars were offered by Player 1*. The results here are conditional on Player 1 having recovered jars and made an offer to Player 2. The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. "Jar # Blue Balls" is the number of blue balls in each of the jars. "High Urn # Blue Balls" is the number of blue balls in the high blue urn. "Low Urn # Blue Balls" is the number of blue balls in the low blue urn. "Mixing (Transplant) Rate" is the percentage of urns where a mixing happened, this is the number of transplant(s) divided by 2 per round. "% Mixings Made Urn Worse" is the percentage of mixings that occurred which made the odds of blue balls worse for an urn.

	(1)	(2)	(3)	(4)	(5)	(9)	(2)	(8)	(6)
		All Urns		At Least	One Urn Sick	test or Healthiest	No Urn	Sickest or]	Healthiest
	All	Low Blue	High Blue	All	Low Blue	High Blue	All	Low Blue	High Blue
Status Quo	0.041^{**}	0.055^{**}	0.027^{**}	0.070^{**}	0.101^{**}	0.039^{*}	0.026^{**}	0.034^{**}	0.018^{*}
	(0.007)	(0.012)	(0.008)	(0.016)	(0.022)	(0.015)	(0.007)	(0.011)	(0.009)
Constant	0.421^{**}	0.578^{**}	0.264^{**}	0.401^{**}	0.638^{**}	0.164^{**}	0.432^{**}	0.548^{**}	0.315^{**}
	(0.006)	(0.008)	(0.006)	(0.011)	(0.016)	(0.010)	(0.006)	(0.008)	(0.007)
N	3240	1620	1620	1078	539	539	2162	1081	1081
R^2	0.007	0.020	0.007	0.012	0.054	0.015	0.004	0.010	0.004

Table V: Impact on Expected Bad Outcomes (Red Balls Drawn) Based on Actual Mixing Behavior

+ p < 0.1, * p < 0.05, ** p < 0.01

the end of the round, representing the percentage of transplant candidates/patients expected to get a bad health outcome (e.g. expected mortality rate). The results here are conditional on actual Player 1 and Player 2 behavior in the game but not the actual draws. The independent variable "Status Quo" indicates Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + X_i\gamma + \epsilon_i$. We model bad health outcomes as the drawing of red balls. The outcome of interest reported in this table is "% Red Ball" This outcome is the percentage of balls that are red balls in urns at the status quo condition where the incentives resemble the current fragmented regulations. The dependent variable in this table is the percentage of balls drawn from an urn at the end of a round being red.

A Online Appendix

This is the online appendix for "Regulation of Organ Transplantation and Procurement: A Market Design Lab Experiment" by Alex Chan and Alvin E. Roth.

- Section A.1: Experiment Details
 - Section A.1.1: Recruitment On Prolific
 - Section A.1.2: Overview of Experiment
 - Section A.1.3: Game Introduction
 - Section A.1.4: Comprehension Quiz
 - Section A.1.5: Stage Game Screens
 - Section A.1.6: Demographic Information Collection
- Section A.2: Supplementary Tables

A.1 Experimental Details

In this Section, we describe the design of our experiment in detail, including the recruitment screen on Prolific (A.1.1), and the game screens (A.1.2).

A.1.1 Recruitment On Prolific

Subjects were recruited on the Prolific platform (see Section 4.1). The subjects in the game participated between August 2 2021 and August 12 2021. Prolific recruitment posting were posted during 7 days (August 2, 3, 4, 5, 7, 9, 12) in this period where subjects were recruited to play the game during a 10-minute window on each of these 7 days. The narrow window for participation is to increase the number of subjects who are online simultaneously, so that we can pair them off into our two-player game. See Figure A1 for the recruitment information on Prolific. Subjects who managed to get paired off become part of our sample.



Figure A1: Recruitment screen posted on Prolific

A.1.2 Overview of Experiment

After a participation consent screen, subjects demographic information were collected (see Section A.1.6). After these short steps, the subjects are re-directed to a game page where they will enter their Prolific ID (see Figure A2) and get assigned to a online waiting room until they are matched with another subject (see Figure A3). Pairing is done randomly based on arrival time to the game. Each pair is then randomized into a treatment condition (status quo or holistic) using Bernoulli draws as subjects arrive.

Welcome	
	Welcome
	If you are a new player, the please enter your Prolific ID and press the left button here
	Enter Prolific ID
	ENTER NEW GAME
	If you are a returning player, then please enter your room code and the Prolific ID you used and press the right button here
	Enter Prolific ID Enter Room Code
	RESUME GAME

Figure A2: Game Launch Page

Welcome
Welcome
Waiting for next player to join
If the game has not started in 5 minutes or the room code details are empty, please go back to the home screen and start a new room again.
Room code (Write this down): 2zUvq
Your alias (Write this down): Alex
You will join the game as player1

Figure A3: Game Online Waiting Room Page

This Appendix Section should serve to provide additional details about the actual experience of subjects of this experiment as they play this online, two-player game. We will share the screens broken down into the following three sections (Sections A.1.3, A.1.4 and A.1.5).

A.1.3 Game Introduction

Subjects randomized into either the status quo and holistic conditions will go through 6 introductory screens that are identical. These 6 screens serve to explain game details like the available actions to both players (Figures A4 and A9), number of rounds (Figure A4), high-level description of the games including the fact that balls will be drawn from urns at the end of each round (Figure A5), the nature of the jars and urns (how many, the distribution from which balls are drawn from) (Figures A6, A7), as well as information available to each player (Figure A8). Essentially, the first 6 screen outlines the game as described in Section 3 except information regarding the actual payoff schemes.



Figure A5: Game Screen 2

(Page 3 of 10)





Instructions for Player 2

In this study, you have been assigned the role of Player 2. You have been randomly matched with another participant who will be in the role of Player 1. Your earnings will depend on your decisions, as well as on the decisions of Player 1. Three will be 10 rounds of this study. For all 10 rounds you will be paired with the same Player 1, who will participate at the same time as you.

In each of the 10 rounds Player 1 will be given two identical jars of balls and you will be given two urns of balls. A different pair of jars and 2 different urns will given for each round (the jars or urns do not carryover to the next round(s)). Player 1's earnings for each round can depend on whether Player 1 offered the pair of jars to you, whether you accepted Player 1's jars, and how many blue balls were drawn from your urns at the end of the round.

Wait 5 seconds before the next page can appear...

Figure A6: Game Screen 3

(Page 4 of 10)

There are 2 types of ums: "High Blue" and "Low Blue". You will have one um of each type in every round. The ums each have 100 balls at the beginning of each round, each ball is either blue or red. The chance of drawing blue balls from an um is higher when there are more blue balls relative to red balls in that um. You can change the chance of drawing blue balls from an um at the end of the round by mixing the balls in one of the jars from Player 1 into that um.

There are 2 types of jars: "Low Quality" and "High Quality". The quality of each jar is indicated by the number of blue balls: it can range from 0 to 100. When mixed into an urn, a jar can increase or decrease the average chance of drawing blue balls from the urn at the end of the round. With mixing, the number of blue balls in an urn at the end of the round will be equal to the sum of the number of blue balls in the jar and the number of blue balls in that urn (instead of the number of blue balls in that urn at the beginning of the round). For example, if the number of blue balls in a jar is higher than the number of blue balls in an urn, mixing the balls from that urn will increase the chance of drawing blue balls from that urn at the end of the round.

Wait 5 seconds before the next page can appear...

Figure A7: Game Screen 4

(Page 5 of 10)

In each round, you can see the type and the actual number of blue balls in Player 1's jars as well as the type and actual number of blue balls in each of your urns before making any decisions in each round.

Unlike you, Player 1 will not be able to observe the exact number of blue balls in the jars that he/she received but he/she will be told whether he/she received a pair of identical "High Quality" or a "Low Quality" jars for that round. A "High Quality" jar has a number of blue balls (out of 100) that can be equal to any number between 70-100 with equal chance, whereas a "Low Quality" jars has a number of blue balls (out of 100) that can be equal to any number between 70-100 with equal chance. Blue" una swell as a "Low Blue" un. Player 1 also knows that a "High Blue" unh has a number of blue balls (out of 100) (before mixing) that can be equal to any number between 40-100 with equal chance while a "Low Blue" un has a number of blue balls (out of 100) (before mixing) that can be equal to any number between 40-100 with equal chance while a "Low Blue" un has a number of blue balls (out of 100) (before mixing) that can be equal to any number between 40-100 with equal chance while a "Low Blue" un has a number of blue balls (out of 100) (before mixing) that can be equal to any number between 40-100 with equal chance while a "Low Blue" un has a number of blue balls (out of 100) (before mixing) that can be equal to any number between 40-100 with equal chance while a "Low Blue" un has a number of blue balls (out of 100) (before mixing) that can be equal to any number between 40-100 with equal chance while a "Low Blue" and has a number of blue balls of either of your urns.

You, Player 2, can see the actual number of blue balls in the jars as well as the actual number of blue balls in each of your urns before making any decisions in each round.

Wait 5 seconds before the next page can appear...

1 2 3 4 5 >

Figure A8: Game Screen 5

(Page 6 of 10)

Instructions for Player 2

In each of the 10 rounds, if Player 1 offered you the jars, you can decide whether to:

- Decline both of Player 1's jars
 OR
- Mix all the balls from a jar offered by Player 1 into in your High Blue urn and decline the other jar OR
- Mix all the balls from a jar offered by Player 1 into in your Low Blue urn and decline the other jar
- OR

• Mix all the balls from a jar offered by Player 1 into in your High Blue urn and mix all the balls from the other jar offered by Player 1 into in your Low Blue urn

If Player 1 did not make an offer of jars in a round, you will not move for that round and will earn zero for that round. You, Player 2, cannot mix the balls from both jars into a single



Wait 5 seconds before the next page can appear.

< 1 2 3 4 5 6 >

Figure A9: Game Screen 6

After these introduction screens, four more screens will be shown to outline how subjects will be paid and to give subjects a chance to see how the stage-game interfaces for them and the player they are playing against look like.

First, subjects will see a verbal description of the incentive scheme associated with their assigned treatment condition. Figure A10 shows this screen for status quo and Figure A11 shows it for holistic.

(Page 7 of 10)
Your goal in the game is to minimize the percentage of red balls among balls drawn from urns where mixing occurred.
Player 1 earns by offering and getting more jars accepted, gets a penalty if both of his/her jars was declined, and gets zero if he/she made no offer or if only one of the jars were accepted.
Player 2 earns by mixing balls from jars to urns. He/she earns more by mixing more jars but he/she is penalized for for each red ball was drawn from every urn that was mixed with a jar. Player 2 gets zero if Player 1 made no offer.
We provide more details on the payment schemes for both Players in the next screen.
Wait 5 seconds before the next page can appear < 1

Figure A10: Game Screen 7 for Status Quo

(Page 7 of 10)

The aim of the game is to maximize number of blue balls drawn from all the urns.					
Both Player 1 and Player 2 get rewards based on the number of blue balls and penalties based on the number of red balls drawn at the end of each round.					
We provide more details on the payment schemes for both Players in the next screen.					
Wait 5 seconds before the next page can appear					
1 2 3 4 5 6 7 >					

Figure A11: Game Screen 7 for Holistic

Next, subjects will see a schematic and a more detailed verbal description of how payoffs are determined. It clearly depicts how payoff levels for the subject and their counterpart in the game are determined based on each other's actions. Figure A12 shows this screen for status quo and Figure A13 shows it for holistic.

(Page 8 of 10)

Wait 5 seconds before the next page can appear. < 1 ... 5 6 7 8 9 →



Figure A12: Game Screen 8 for Status Quo





Then, subjects will see what they would expect to see as the game interface in the stage game, as well as what the other play would see. Player 1's stage game screen is always shown first and then Player 2's. Figures A14 and A15 shows these screens for status quo and Figures A16 and A17 shows them for holistic.

(Page 9 of 10)	
For your reference, Player 1's decision screen in the game in e	each stage will look as follows:
	Round Details For this round, you are randomly assigned a pair of jars of the following type: high quality (70-100 blue balls)
	Available Actions Now, please select an action by clicking a box below for the current round of the study. Player2 will then decide whether to reject this offer, accept 1 of your jars, or accept both of your jars. After that, we will go to the next round, and so on until round 10 Your reward amount if you Player 2 accepts BOTH of your jars. 0 tokens Your reward/penalty amount if you Player 2 rejects your jars: 30 tokens Your penalty amount if you Player 2 rejects your jars: 30 tokens
	Offer your jars to Player 2 Do not offer your jars to Player 2 SUMMAT RESPONSE
Wait 5 seconds before the next page can appear	

Figure A14: Game Screen 9 for Status Quo

(Page 10 of 10)

For your reference, Player 2's decision screen in the game in each stage will look as follows:

	Round Details
	For this round, your High Blue um, your Low Blue um, and Player 1's jars are randomly assigned the following number of blue balls and red balls:
	Player1's Jar Quality (same for both jars): high quality (100 blue balls and 0 red balls)
	High Blue Um: 68 blue balls and 32 red balls
	Low Blue Urn: 21 blue balls and 79 red balls
	Available Actions
	Player 1 has decided to offer you their jars. Please choose an action:
	Your reward amount for EACH blue ball you draw if you mix an um: 25 tokens
	Your penalty amount for EAOH red ball you draw if you mix an um: -100 tokens
	Reject the jars offered by Player 1. You will earn 0 tokens by choosing this option for this round.
	O Mix One of Player 1's jar with your High Blue um. The High Blue um will now have 200 balls where 168 or 84.0% of the balls will be blue for this um (before a ball is drawn from it).
	O Mix One of Player 1's jar with your Low Blue um. The Low Blue um will now have 200 balls where 121 or 60.5% of the balls will be blue for this um (before a ball is drawn from R).
	Mix One of Player 1's jar with your High Blue um and, and mix the other jar with your Low Blue um. Both ums will now have 200 balls. You will have 84.0% blue balls in the High Blue um and 68.5% blue balls in the Low Blue um (before the balls are drawn).
	SUBMIT RESPONSE AND DRAW BALL
< 1 6 7 8 9 10 >	
	Close Instructions

Figure A15: Game Screen 10 for Status Quo

(Page 9 of 10)

For your reference, Player 1's decision screen in the game in each stage will look as follows:

	Round Details
	For this round, you are randomly assigned a pair of jars of the following type:
	low quality (0-70 blue balls)
	Available Actions
Now, please select an action b	ay clicking a box below for the current round of the study. Player2 will then decide whether to reject this offer, accept 1 of your jars, or accept both of your jars. After that, we will go to the next round, and so on until round 10
	Your reward amount if you Player 2 draws 2 blue balls: 32 tokens
	Your reward amount if you Player 2 draws 1 blue ball and 1 red ball: 8 tokens
	Your penalty amount if you Player 2 draw 2 red balls: -16 tokens
	Offer your jars to Player 2
	O Do not offer your jars to Player 2
	SUBMIT RESPONSE
Wait 5 seconds before the next page can appear	

< 1 ... 5 6 7 8 9 >

Figure A16: Game Screen 9 for Holistic

(Page 10 of 10)



For your reference, Player 2's decision screen in the game in each stage will look as follows:

Figure A17: Game Screen 10 for Holistic

A.1.4 Comprehension Quiz

To reinforce comprehension of the rules of the game, the introduction screens are followed by five comprehension questions before the actual game play commences. The correct answers, along with a pop-up text box to explain why such answers are correct, to these questions are provided to the subjects if they got a question wrong. The subject will not be able to advance until they correctly answered all the five questions. The order of these questions are randomized. The questions are shown in the figures below.

If the question shown in Figure A18 is answered incorrectly, the following answer is shown for both conditions:

"Answer is 40%. Since the jar and the urn has the same number of balls in total and the jar has 50% blue balls and the urn has 30% blue balls, the percentage of blue balls will be the average of the 2, or 40%. Another way to see this is that 50 blue balls from jar + 30 blue balls from urn = 80 blue balls in total. Dividing this by 200 balls (100 balls from jar and 100 balls from urn), we get 40%."

If there are 50 blue balls and 50 red balls in a jar and 30 blue balls and 70 red balls in a urn, what is the percentage of blue balls in the urn after we mix in the balls from the jar?



Figure A18: Comprehension Question 1

If the question shown in Figure A19 is answered incorrectly, the following answer is shown for both conditions:

• "Answer is Higher than before mixing. Before mixing, the urn has a lower percentage of blue balls. Since the jar has a higher percentage of blue balls, mixing will increase the percentage of blue balls in the urn. So if the one wants to have the highest chance of drawing a blue ball, one would mix." Assume that you know the exact numbers of blue balls in each urn (as Player 2 does). If Player 2 received an offer and mixed the balls from one of Player 1's jar to one of Player 2's urns with fewer blue balls than the jar, what is the chance of drawing a blue ball from the urn after mixing relative to before mixing?



Figure A19: Comprehension Question 2

If the question shown in Figure A20 is answered incorrectly, the following answer is shown depending on the treatment condition:

- For Status Quo: "Answer is Unclear with given information. It depends whether the mixed urn(s) will provide good enough odds for Player 2. For example, if Player 2 is given jars with 99 blue balls and 1 red ball, the mixed High Blue urn will still have a 99.5% chance of drawing a blue ball. While 99.5% is lower than 100%, Player 2 might still find it worth his/her while to make the bet. On the other hand, if the jar has fewer blue balls, say 1 blue ball and 99 red balls, mixing it with the urn with 100 blue balls will now lower the chance of drawing a blue ball while some more risk loving Player 2 might still want to mix for a chance to win some earnings, some others might rather not take the gamble. Also, note that Player 2 will never mix the jar with the urn with 100 red balls as mixing any given jar with the urn with 100 blue balls will give him/her better odds of drawing blue balls."
- For Holistic: "Answer is Reject any offer from Player 1 and Mix the balls from Player 1's jar with the urn with 100 blue balls. Player 2 will never reject any jar offer that can at least improve the odds of one of his/her urns. Given that the

urn with 100 red balls has 0% chance of drawing a blue ball, any chance to add a blue ball in the mix will be an improvement - therefore he/she will never reject a jar. Similarly, with whatever jar (not 100 red balls or 100 blue balls) that Player 2 received, he/she can only make the odds of drawing a blue ball from the urn with 100 blue balls worse, as the current chance is 100%. Therefore, he/she will never mix a jar with the urn with 100 blue balls instead of the urn with 100 red balls."





Figure A20: Comprehension Question 3

If the question shown in Figure A21 is answered incorrectly, the following answer is shown for both conditions:

• "Answer is No. The jar has 49% blue balls and the urn has 50% blue balls. Mixing a jar with lower percentage of blue balls that the urn will only lower the percentage of blue balls in the urn. So if the one wants to have the highest chance of drawing a blue ball, one would not mix and just draw from the urn." If there are 49 blue balls and 51 red balls in a jar and 50 blue balls and 50 red balls in Player 2's urn in a specific round of the study, would Player 2 want to mix the balls from the jar into his/her urn if he/she is trying to increase the chance that a blue ball is drawn randomly from the urn at the end of the round?



Figure A21: Comprehension Question 4

If the question shown in Figure A20 is answered incorrectly, the following answer is shown depending on the treatment condition:

- For Status Quo: "Answer is Definitely not make an offer to Player 2. This is because Player 1 will lose tokens if he/she made an offer to Player 2 and it was not accepted. Therefore, if Player 1 knows that the jars will be rejected, he/she will never offer it in the first place."
- For Holistic: "Answer is Unclear/Do not know. Player 1's payoff depends on what color balls are drawn from the 2 urns. While he/she can influence the odds of drawing blue balls by offering Player 2 a pair of jars as an option of changing the percentage of blue balls for one or both of his/her urns, if Player 2 were to not accept the jars for sure Player 1 loses nothing. However, if Player 1 is not completely sure if Player 2 would accept the jars. If Player 1 is not 100% sure that Player 2 would accept both jars but 99% sure that Player 2 would reject the jars, it is still in Player 1's interest to offer the jars in the off chance (1% chance) that Player 2 might be able to accept and improve the odds of drawing a blue ball from one or both of the urns."

If Player 1 believes that Player 2 will certainly reject both jars in a given round, what would Player 1 most likely do?



Figure A22: Comprehension Question 5

A.1.5 Stage Game Screens

Finally, the subjects play the stage game for 10 rounds.

Player 1 gets action options (in a multiple choice format) and a reminder that shows them the level of payoffs he should expect to receive based on Player 2's and their own actions (and the draw of the balls if applicable). This decision screen for Player 1s in the status quo condition is shown in A23 and shown in A24 for holistic.

OPEN/CLOSE INSTRUCTIONS Round: 1 Room: G2QhP Alias: Alex
Round Details
For this round, you are randomly assigned a pair of jars of the following type:
high quality (70-100 blue balls)
Available Actions
Now, please select an action by clicking a box below for the current round of the study. Player2 will then decide
whether to reject this offer, accept 1 of your jars, or accept both of your jars. After that, we will go to the next round,
and so on until round 10
Your reward amount if you Player 2 accepts BOTH of your jars: 10 tokens
Your reward/penalty amount if you Player 2 accepts ONE of your jars: 0 tokens
Your penalty amount if you Player 2 rejects your jars: -30 tokens
Offer your jars to Player 2
O Do not offer your jars to Player 2
SUBMIT RESPONSE

Figure A23: Player 1 Stage Game Decision Screen for Status Quo

OPEN/CLOSE INSTRUCTIONS Round: 2 Room: 2zUvq Alias: Alex	
Round Details	
For this round, you are randomly assigned a pair of jars of the following type:	
low quality (0-70 blue balls)	
Available Actions	
Now, please select an action by clicking a box below for the current round of the study. Player2 will then decide whether to reject this offer, accept 1 of your jars, or accept both of your jars. After that, we will go to the next round, and so on until round 10	
Your reward amount if you Player 2 draws 2 blue balls: 32 tokens	
Your reward amount if you Player 2 draws 1 blue ball and 1 red ball: 8 tokens	
Your penalty amount if you Player 2 draw 2 red balls: -16 tokens	
Offer your jars to Player 2	
O not offer your jars to Player 2	
SUBMIT RESPONSE	

Figure A24: Player 1 Stage Game Decision Screen for Holistic

After Player 1 moves, Player 2 will get to move in the round. If Player 1 offered Player 2 the jars, Player 2 sees a screen that show her action options and a reminder that shows her the level of payoffs she should expect to receive based on the draw of a ball from the urns at the end of the round. Note that along with each action option for Player 2, the expected number of blue/red balls associated with the urn post-mixing is also provided to ease the need for Player 2 to do the calculations. This decision screen for Player 2s

in the Status Quo condition is shown in A25 and shown in A26 for those in the Holistic condition.

V/CLOSE INSTRUCTIONS	Round: 2 Room: G2	2QhP Alias: Alvin				
			Rou	nd Details		
	For this	s round, your High Blue urn, you	r Low Blue urn, and Player 1's	jars are randomly assign	ed the following number of b	ue balls and red balls:
		Player1's Jar G	Quality (same for both jar	s): low quality (59 bl	ue balls and 41 red ball	s)
			High Blue Urn: 58 b	olue balls and 42 red	balls	
			Low Blue Urn: 57 b	lue balls and 43 red	balls	
			Availa	ble Actions		
		1	Player 1 has decided to offer y	ou their jars. Please choo	se an action:	
		Your	reward amount for EACH blue	e ball you draw if you mix	an urn: 25 tokens	
		Your	penalty amount for EACH red	ball you draw if you mix a	n urn: -100 tokens	
	O Reject the jars of	ffered by Player 1. You will earn	0 tokens by choosing this opt	on for this round.		
	Mix One of Playe drawn from it).	ər 1's jar with your High Blue um	. The High Blue urn will now h	ave 200 balls where 117	or 58.5% of the balls will be b	lue for this urn (before a ball is
	Mix One of Playe drawn from it).	er 1's jar with your Low Blue urn.	The Low Blue urn will now ha	ve 200 balls where 116 o	r 58.0% of the balls will be bl	ue for this urn (before a ball is
	Mix One of Playe in the High Blue	er 1's jar with your High Blue um um and 58.0% blue balls in the	and, and mix the other jar wit Low Blue urn (before the balls	h your Low Blue urn. Both are drawn).	urns will now have 200 balls	. You will have 58.5% blue balls
			SUBMIT RESP	ONSE AND DRAW BALL		
	Scorecard					
	Round	Player 1 Decision	Player 2 Decision	Ball Drawn	Your Bank	Earnings that round
	1	Offer	MixWithHighBlue	blue ball	525	25

Figure A25: Player 2 Stage Game Decision Screen for Status Quo



Figure A26: Player 2 Stage Game Decision Screen for Holistic

After both players have made their decisions for each round, a screen showing the actual ball(s) drawn from the urn(s) at the end of the round and the earnings for the player in the round is shown. A table summarizing the actions taken by each player,

actual ball drawn, and each player's earnings and bank in each previous round is shown at the bottom of this page. This table is also shown in the bottom of each decision screen. The post-decision screen summarizing actions, balls drawn and payoffs for each round is shown below (Figure A27 for Player 1 and Figure A28 for Player 2.).

OPEN/CLOSE INSTRUCTIONS	Round: 1	Room: IM1kA Alia	as: Alex			
			Round	Details		
		For this round, ye	ou are randomly assig	ned a pair of jars of	the following type:	:
			low quality (0-	-70 blue balls)		
			Res	sults		
			All players mad	de their moves	s!	
		The	ball drawn wa	s: 1 blue and	1 red	
			You earned	d: 8 tokens		
		Pres	s OK to continu	ue to the next	round	
			o	к		
	Scorecard					
	Round	Player 1 Decision	Player 2 Decision	Ball Drawn	Your Bank	Earnings that round
	1	Offer	MixWithHighBlue	1 blue and 1 red	508	8

Figure A27: Player 1 Stage Game Post-Decision Screen

OPEN/CLOSE INSTRUCTIONS ROL	und: 2 Room: 2zUvq	Alias: alvin				
			Round	Details		
	For this round	d, your High Blue urn, your Lo	ow Blue um, and Player 1's jan	s are randomly assigned the	following number of blue bal	s and red balls:
		Player1's Jar Qua	lity (same for both jars):	low quality (14 blue bal	lls and 86 rød balls)	
			High Blue Urn: 89 blue	balls and 11 red balls		
			Low Blue Urn: 49 blue	balls and 51 red balls		
			Res	sults		
			All players mad	de their moves!		
			The ball drawn	vas: 2 blue balls		
			You earned	I: 40 tokens		
			Press OK to continu	ue to the next round		
			c	к		
	Scorecard					
	Round	Player 1 Decision	Player 2 Decision	Ball Drawn	Your Bank	Earnings that round
	1	Offer	MixWithHighBlue	1 blue and 1 red	510	10
	2	Offer	MixWithHighBlue	2 blue balls	550	40

Figure A28: Player 2 Stage Game Post-Decision Screen

A.1.6 Demographic Information Collection

We constructed the survey tool to gather demographic information using Qualtrics software for customers who clicked the URL on Prolific to participate in the experiment. Upon opening the link, respondents must read and consent to continue (see Figure A29).

Protocol Director: Professor Alvin Roth

Protocol Number: IRB-44868 IRB Approval Date: March 31, 2021 Expiration Date: March 31, 2022

Description: You are invited to participate in a research study on decision-making. You will be asked to read several pages of instructions. Then you will be asked to make several choices that will determine the precise amount you will be paid, and then possibly answer several survey questions.

Risk and benefits: We cannot and do not guarantee or promise that you will receive any benefits from this study. There are no risks associated with this study.

Time involvement: Your participation in this experiment will take approximately as long as is indicated in the advertisement.

Payments: You will be compensated at the advertised rate.

Subject's rights: If you have read this form and have decided to participate in this project, please understand your participation is voluntary and you have the right to discontinue participation at any time without penalty or loss of benefits to which you are otherwise entitled. You have the right to refuse to answer particular questions. Your individual privacy will be maintained in all published and written data resulting from the study.

Contact information:

Questions, concerns, or complaints: If you have any questions, concerns or complaints about this research study, its procedures, risk and benefits, you should ask the Protocol Director, Alvin Roth, (650)-725-9147.

Independent Contact: If you are not satisfied with how this study is being conducted, or if you have any concerns, complaints, or general questions about the research or your rights as a participant, please contact the Stanford Institutional Review Board (IRB) to speak to someone independent of the research team at (650)-723-2480 or toll free at 1-866-680-2960 or via email at irbnonmed@stanford.edu. You can also write to the Stanford IRB, Stanford University, Stanford, CA 94305-5401.

By continuing with this study, you are consenting to participate. Please make a copy of this consent form for your own records. You can do so by right-clicking and selecting "print" in most browsers. If you cannot do so on your browser, please contact the protocol director for a copy of the consent form.

I consent, begin the study
I do not consent, I do not wish to participate

Figure A29: Consent screen

Then, the subjects are asked to answer a series of questions listed below before they are redirected to the game itself. These question gather information about the subject on their (in the order of appearance): race (Figure A30) and ethnicity (A31), sex (Figure A32), state of residence (Figure A33), age (Figure A34), employment status (Figure A35), and education (Figure A36).

Choose one or more races that you consider yourself to be:			
White	Asian		
Black or African American	Native Hawaiian or Pacific Islander		
American Indian or Alaska Native	Other		

Figure A30: Survey Question about Subject's Race

Are y	ou Spanish, Hispanic, or Latino or none of these?
Yes	3
No	ne of these

Figure A31: Survey Question about Subject's Ethnicity

What is your se	ex?		
Male			
Female			

Figure A32: Survey Question about Subject's Sex

In which state do you currently	reside?
	$\mathbf{\mathbf{v}}$

Figure A33: Survey Question about Subject's State of Residence

What is your age?
Under 18
18 - 24
25 - 34
35 - 44
45 - 54
55 - 64
65 - 74
75 - 84
85 or older

Figure A34: Survey Question about Subject's Age

What is your current employment status?
Employed full time (40 or more hours per week)
Employed part time (up to 39 hours per week)
Unemployed and currently looking for work
Unemployed not currently looking for work
Student
Retired
Homemaker
Self-employed
Unable to work

Figure A35: Survey Question about Subject's Employment Status

What is the highest degree or level of school you have completed?
Less than a high school diploma
High school degree or equivalent (e.g. GED)
Some college, no degree
Associate degree (e.g. AA, AS)
Bachelor's degree (e.g. BA, BS)
Master's degree (e.g. MA, MS, MEd)
Doctorate or professional degree (e.g. MD, DDS, PhD)

Figure A36: Survey Question about Subject's Education Status

After providing demographic information, the subjects will asked to submit their Prolific ID and will be automatically redirected to the game itself (see Figure A37).

(IMPORTANT) Read the following carefully:
Welcome to this study of decision-making. In this study, you will play a game with another participant. You will receive your <u>show-up payment if you and the player paired with you</u> <u>finish all 10 rounds of the game</u> . Besides your show-up payment, you can win additional earnings through the game.
Note that you need to show up during the time designated in the Prolific Request in order to be matched with another player. We cannot guarantee that you will be matched with another player if you are not showing up during the designated timeframes.
In the game, you must <mark>use your Prolific ID as your alias in order to get paid</mark> .
What is your Prolific ID?

Figure A37: Transition Screen to Game

A.2 Supplementary Tables

	(1)	(2)	(3)	(4)	(5)	(6)
	% Re	d Balls	$\% \mathrm{Re}$	d Balls	% Re	d Balls
	for al	l Urns	for Low	Blue Urns	for High	Blue Urns
Status Quo	0.034^{+}	0.031^{+}	0.026	0.031	0.041^{+}	0.032
	(0.018)	(0.0167)	(0.024)	(0.023)	(0.022)	(0.021)
Ion # Dluo Dolla		0.009**		0 009**		$0.001 \pm$
Jar $\#$ Drue Dans		-0.002		-0.005		-0.001
		(0.000)		(0.000)		(0.000)
High Urn # Blue Balls		-0.004**		-0.000		-0.007**
High OH $\#$ Dhe Dahs		-0.004		(0.001)		(0,001)
		(0.000)		(0.001)		(0.001)
Low Urn # Blue Balls		-0.004**		-0.008**		0.000
,,		(0.000)		(0.001)		(0.001)
Constant	0.428^{**}	0.941^{**}	0.589^{**}	1.055^{**}	0.267^{**}	0.826^{**}
	(0.013)	(0.037)	(0.018)	(0.056)	(0.015)	(0.053)
N	1620	1620	1620	1620	1620	1620
R^2	0.003	0.115	0.001	0.120	0.002	0.086

Table A1: Impact on Bad Outcomes (Red Balls Drawn)

Standard errors (Robust, clustered by player-pairings) in parentheses

+ p < 0.1, * p < 0.05, ** p < 0.01

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 Status Quo_i + X_i \gamma + \epsilon_i$. The ball drawn from each urn at the end of each round could be either Blue or Red: We model bad health outcomes as the drawing of red balls. The outcome of interest reported in this table is "% Red Ball." This outcome is the percentage urns from which a red ball was drawn at the end of the round of that urn, representing the percentage of transplant candidates/patients getting a bad health outcome (e.g. mortality rate). The results here are conditional on Player 1 having recovered jars and made an offer to Player 2. The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. "Jar # Blue Balls" is the number of blue balls in each of the jars. "High Urn # Blue Balls" is the number of blue balls urn. "Low Urn # Blue Balls" is the number of blue balls in the low blue urn.

	(1)	(2)	(3)	(4)	(5)	(6)
	% Re	d Balls for	% Red	Balls for	% Red 1	Balls for
	All Urns	when One Urn	Sickes	t Urns	Healthie	est Urns
	is Health	iest or Sickest				
Status Quo	0.072^{*}	0.046^{+}	0.130**	0.100^{*}	0.074^{*}	0.063^{+}
	(0.032)	(0.027)	(0.045)	(0.044)	(0.036)	(0.036)
Jar # Blue Balls		-0.003**		-0.004**		-0.001
		(0.000)		(0.001)		(0.001)
High Urn # Blue Balls		-0.006**		-0.003**		-0.008
		(0.001)		(0.001)		(0.006)
Low Urn $\#$ Blue Balls		-0.003**		-0.014^{+}		0.001
		(0.001)		(0.007)		(0.001)
Constant	0.411**	1.150^{**}	0.731**	1.273**	0.062**	0.832
	(0.022)	(0.065)	(0.035)	(0.103)	(0.020)	(0.572)
N	539	539	300	300	295	295
R^2	0.012	0.255	0.026	0.129	0.016	0.038

Table A2: Impact on Bad Outcomes (Red Balls Drawn) for the "Healthiest" (Urns with $\geq 90\%$ Blue Balls) and "Sickest" (Urns with $\leq 10\%$ Blue Balls)

Standard errors (Robust, clustered by player-pairings) in parentheses

+ p < 0.1, * p < 0.05, ** p < 0.01

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 Status Quo_i + X_i\gamma + \epsilon_i$. This table reports only results from urns that are among the "Healthiest" (Urns with $\geq 90\%$ Blue Balls) and "Sickest" (Urns with $\leq 10\%$ Blue Balls. The ball drawn from each urn at the end of each round could be either Blue or Red: We model bad health outcomes as the drawing of red balls. The outcome of interest reported in this table is "% Red Ball." This outcome is the percentage urns from which a red ball was drawn at the end of the round of that urn, representing the percentage of transplant candidates/patients getting a bad health outcome (e.g. mortality rate). The results here are conditional on Player 1 having recovered jars and made an offer to Player 2. The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. "Jar # Blue Balls" is the number of blue balls in each of the jars. "High Urn # Blue Balls" is the number of blue balls in the high blue urn. "Low Urn # Blue Balls" is the number of blue balls in the low blue urn.

	(1)	(2)	(3)	(4)	(5)	(6)
	$\% \mathrm{Re}$	d Balls	$\% \mathrm{Re}$	d Balls	% Re	d Balls
	for al	l Urns	for Low	Blue Urns	for High	Blue Urns
Status Quo	0.015	0.019	0.004	0.014	0.027	0.023
	(0.021)	(0.021)	(0.025)	(0.025)	(0.027)	(0.026)
Jar # Blue Balls		-0.001**		-0.003**		-0.001
		(0.000)		(0.000)		(0.000)
High Urn $\#$ Blue Balls		-0.003**		0.000		-0.007**
		(0.001)		(0.001)		(0.001)
Low Urn $\#$ Blue Balls		-0.003**		-0.008**		-0.000
		(0.000)		(0.001)		(0.001)
Constant	0.437**	0.856**	0.556^{**}	1.012**	0.317**	0.819**
	(0.015)	(0.055)	(0.019)	(0.067)	(0.018)	(0.067)
N	1081	1081	1320	1320	1325	1325
R^2	0.000	0.058	0.000	0.086	0.001	0.049

Table A3: Impact on Bad Outcomes (Red Balls Drawn) for those NOT "Healthiest" (Urns with $\geq 90\%$ Blue Balls) or "Sickest" (Urns with $\leq 10\%$ Blue Balls)

Standard errors (Robust, clustered by player-pairings) in parentheses ⁺ p < 0.1, * p < 0.05, ** p < 0.01

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_0$ $\beta_1 Status Quo_i + X_i \gamma + \epsilon_i$. This table reports only results from urns that are among those NOT "Healthiest" (Urns with $\geq 90\%$ Blue Balls) or "Sickest" (Urns with $\leq 10\%$ Blue Balls). The ball drawn from each urn at the end of each round could be either Blue or Red: We model bad health outcomes as the drawing of red balls. The outcome of interest reported in this table is "% Red Ball." This outcome is the percentage urns from which a red ball was drawn at the end of the round of that urn, representing the percentage of transplant candidates/patients getting a bad health outcome (e.g. mortality rate). The results here are conditional on Player 1 having recovered jars and made an offer to Player 2. The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. "Jar # Blue Balls" is the number of blue balls in each of the jars. "High Urn # Blue Balls" is the number of blue balls in the high blue urn. "Low Urn # Blue Balls" is the number of blue balls in the low blue urn.

ow Quality Jars Ac	cepted					
	(1)	(2)	(3)	(4)	(5)	(9)
	pre-TX Urn	# Blue Balls	TXed Urn # Blue Balls	TX	ed Jar $\#$ Blue Balls	% Low Quality Jars Accepted offered
Status Quo	13.47^{**}	13.40^{**}	6.990^{**}	0.514	4.797^{**}	-0.221**
·	(1.567)	(1.553)	(1.277)	(1.836)	(1.135)	(0.0423)
Jar # Blue Balls		0.120^{**} (0.0329)				
Constant	A1 01**	00 £1 **	*** 0 01	**76 34	10 01**	21 0×*
COLLEVALL	41.01	10.26	03.01	10.04	10.01	CT1.U
	(1.016)	(2.854)	(0.791)	(1.053)	(0.896)	(0.0264)
Which transplants?	All	All	All	All	Did not made urn worse	Transplanted or not
N	1295	1295	1295	1295	1089	1094
R^2	0.064	0.073	0.039	0.000	0.023	0.050
Standard errors in p + $p < 0.1$, * $p < 0.05$	arentheses $b_{\star}^{\star \star} p < 0.01$					

centage	
Perc	
and	
Transplanted	
Those	
) for	
(Kidneys)	
Jars	
and	
(Patients)	
Urns	
alls for ¹	
f Blue E	
iber o	
Num	
Average	ed
ss in	cept
Difference	y Jars Ac
A4:	J ualit
Table	Low G

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + X_i\gamma + \epsilon_i$. The dependent variables are the number of blue balls of an urn (columns 1 and 2 are for the urn pre-transplantation, while column 3 is for the urn post-transplantation) or a jar (columns 4-5 are for the jars used in transplantation), and the percentage of Low Quality Jars accepted (column 6). The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. "Jar # Blue Balls" is the number of blue balls in the jar.