

Feasibility Constraints and Protective Behavior in Efficient Kidney Exchange*

Antonio Nicoló[†]

Università degli Studi di Padova

Carmelo Rodríguez-Álvarez[‡]

Universidad Complutense de Madrid

November 12, 2008

1 Introduction

In the past decade economists became increasingly more involved in the design of markets/practical mechanisms (labor market clearinghouses Roth 2002, power markets Wilson 2002, school choice Abdulkadiroğlu and Sönmez 1999, 2003). With the recent seminal paper by Roth et al. (2004), the theory of mechanism design has found an important application in the design and implementation of matching mechanisms to allocate organs for transplantation. The complexity of institutional and feasibility constraints, the normative implications of these rules and their effects on patients' life, must be well taken into

*Nicoló thanks the Italian Ministry of University and Research for the financial support through grant 2005137858. Rodríguez-Álvarez gratefully acknowledges the financial support from the *Ministerio de Educación y Ciencia* through *Programa Ramón y Cajal* 2006, the *Consejería de Innovación Ciencia y Empresa (Junta de Andalucía)*, and the *Fundación Ramón Areces*.

[†]Dipartimento di Scienze Economiche “Marco Fanno”. Università degli Studi di Padova. Via del Santo 33, 37123 PADOVA. Italy. antonio.nicolo@unipd.it.

[‡]Departamento de Fundamentos del Análisis Económico II. Facultad CC. Económicas y Empresariales. Campus de Somosaguas. Universidad Complutense de Madrid. 28223 MADRID. Spain. carmelor@ccee.ucm.es.

account when designing them, and make the task of designing optimal rules a fascinating challenge.

The best treatment for End-Stage-Renal-Disease is kidney transplantation. Patients normally can undergo an alternative treatment –dialysis– but it implies a clear decline in the quality of life standards of the patients. Kidneys available for donation may come from a deceased donor or from willing living donors (normally a close relative). Unfortunately, a donated kidney may be unsuitable for transplantation (*incompatible*) to a given patient because the donor and the patient blood types and tissues may be incompatible, which would lead to the immediate rejection and loss of the graft (*positive crossmatch*). On the other hand, the probability of non immediate rejection of the graft appears to be related to the number of mismatches between donor and patient of the proteins that are present in the tissues of the donors' organ (what is called the Human Leukocyte Antigen loci, HLA). In October 2008, more than 400,000 people in the US are being treated for end-stage kidney failure, and of those more than 76,000 are listed for a deceased donor kidney transplant. In 2007, there were about 16628 kidney transplants in the US, and 6041 of those transplants were from living donors.¹ The figures vary significantly in different developed countries. For instance, in the same period, there were 2211 transplants from deceased donor kidneys and only 137 from living donor for about 6000 patients in the waiting list in Spain.²

Usually, if a willing living donor is not compatible with her intended donor, the donor is sent home and the patient to the cadaveric waiting list. However, the combination of dialysis and living donors makes possible new protocols in kidney transplantation that may generate evident efficiency gains. For instance, two couples of incompatible donor–patient may be mutually compatible and a swap of donors between the two couples would result in two successful transplantations. (Kidney Paired Exchange: KPE).³ Analogously, the donor's kidney may be transplanted to some patient in the deceased donor kidney waiting list and her initially intended patient receives an absolute priority over kidneys in the cadaveric waiting list. (Indirect Exchange or List Paired Exchange: LPE).⁴ In

¹See Organ Procurement and Transplantation Network webpage. www.optn.org.

²See Organización Nacional de Trasplantes webpage, www.ont.es.

³See Delmonico (2004); Delmonico et al. (2004); Segev et al. (2005b); Spital (2004); Segev et al. (2005a).

⁴Patients remaining in the cadaveric waiting list would benefit as well because the patientes who receive a kidney would drop from the standard waiting list. See Delmonico (2004); Kaplan et al. (2005);

fact, the result could be even more profitable because it is possible to design kidney exchanges involving more than two couples of incompatible donor–patients. Of course, when designing these kidney exchange procedures it is important that such mechanisms do not provide incentives for the patients to lie about their medical details in order to improve their chances of getting a match as good as possible.

The application of results in the theory of assignment of heterogeneous indivisible goods by Roth et al. (2004) has shown the huge potential gains of KPE and LPE programs. In subsequent work, Roth et al. (2005a) have proposed a mechanism design approach to KPE that encompasses the features of an acceptable kidney exchange program in New England.⁵ Roth et al. (2005a) assume that patients consider all compatible kidneys as homogeneous and patients’ sets of compatible kidneys are not known to Central Planner. Moreover, because incentives constraints imply that all the operations involved in an exchange must be carried out simultaneously and in the same facility, an exchange between k couples imply that $2k$ simultaneous operations.⁶ Hence, Roth et al. (2005a) consider that only pairwise exchanges between two couples of donor–patient pairs are feasible. In this environment, it is shown that priority mechanisms ensure that it is a dominant strategy for patients to truthfully reveal both the set of donors they can receive kidneys and the set of patients that their donor can donate a kidney to.⁷

Despite the evident success of the application of Roth et al. (2005a)’s approach in New England, some assumptions remain controversial. In Europe and in several areas of the US, it is generally accepted that patients and doctors do not consider all compatible kidneys as homogeneous. Many individual characteristics of the donor, like age, health status, as also matching characteristics of the couple donor–patient, like HLA mismatches, are statistically significant in determining the probability of long–term graft survival. Moreover they also affect the quality of life of transplanted patients because depending on

Zenios et al. (2001).

⁵Their proposal is actually been used by the New England Program of Paired Kidney Exchange since 2006. See www.nepke.org and Roth et al. (2005b) for additional details.

⁶For each couple involved in the exchange, a kidney must be reaped and another kidney must be implanted.

⁷Hatfield (2005) shows that the results are robust to any form of feasibility constraint. More recently, it has been shown that efficiency gains could be attained (and almost exhausted) if 3-way KPE and LPE were admitted,(Saidman et al., 2006; Roth et al., 2007), as well as the potential benefits of altruistic no–related (Samaritan) donors in LPE (Sönmez and Ünver, 2005; Roth et al., 2006).

the "quality of the organ" the type and cost of medical treatment after the operation vary.⁸ Hence, we propose an alternative approach to KPE that incorporates some important features of the European view on the kidney exchange. First, as Roth et al. (2005a) we assume that there exist feasibility constraints on the number of simultaneous operations (even if we do not initially restrict our attention to pairwise exchanges). However, we depart from the Roth et al. (2005a)'s model in two aspects.

- (i) Patients do not consider compatible kidneys homogeneous. Moreover, there is common information about patients' preferences over available kidneys because preferences over kidneys are determined by characteristics of patients and donors which are observable by doctors and verifiable by means of medical tests. Therefore, it is not necessary to elicit them from patients.⁹
- (ii) Patients have private information on the minimum quality –reservation value– that is acceptable for them. The choice of receiving a given kidney or continuing the dialysis treatment depends in fact on patient's eagerness to receive an organ instead of continuing the dialysis and waiting for a better kidney. Hence, it depends on how the patient subjectively evaluates the quality of life under dialysis, her expectations about the quality of future pools of kidneys available for exchange, and on his/her attitude towards risk and uncertainty.

According to our approach the quality of a donor–patient match can be measured according to some objective criterion. One example is the Lifetime Years From Transplant (LYFT) method. LYFT is an estimation of the difference between two potential remaining lifetimes, the lifetimes with and without transplant for each candidate on a waiting list and therefore is a metric of the efficacy of transplant.¹⁰ The assumption on the existence

⁸See Duquesnoy et al. (2003), Merion et al. (2005), Keizer et al. (2005), Klerk et al. (2004), Opelz (1997), Kranenburg et al. (2004), and Schnitzler et al. (1999).

⁹A possible problem is that, even if doctors know their own patients' preferences, they could be tempted to report strategically these information to the central planner in order to favour their own patients. This is in fact the justification for the mechanism design approach in Roth et al. (2004). However, this problem seems less relevant in the case of European continental countries, where Transplant Services are normally public and have more information and less coordination problems than those in US; therefore the assumption that doctors truthfully report information about their patients seems reasonable.

¹⁰See "Predicting the Life Years From Transplant (LYFT): Choosing a Metric", Scientific Registry for Transplant Recipients working paper, May 16, 2007 at www.unos.org.

of compatible kidneys with different quality and the possibility that a patient may reject a compatible kidney seems also to fit the existence of Extended Donor Criteria and the use of organs previously regarded as unsuitable, because improvements on immunosuppressive treatment show that although of lower quality, they present good probability of survival.¹¹ In this scenario, given the little information that patients may have about the remaining patients' reservation values, we investigate whether we can construct rules such that revealing the true reservation value is a dominant strategy for the patients. Unfortunately, truth-full revelation is not compatible with a weak version of efficiency.

In the light of the negative result that is obtained in the standard (revelation) model, we propose an alternative "behavioral" approach. Normally, KPE programs involve the coordination of nephrology services and patients of several hospitals. In this environment, it becomes natural to assume that patients have little or no information at all about the remaining patients involved in the program. Hence, patients may take into consideration the fact that, since they do not know the other patients' reservation values, by misreporting their own reservation values, they also could end up losing the possibility of a beneficial transplant or, on the contrary, receiving an undesirable kidney. In our specific environment, it is natural to assume that patients might prefer to choose their strategies so as to "protect" themselves from the worst eventuality as far as possible. We capture this "lexicographic maximin" behavior assumption with the notion of "protective behavior" proposed by Barberà and Dutta (1982) and later axiomatized by Barberà and Jackson (1988). Hence, with this assumption on patients' strategic behavior we try to encompass the notion that patients only care about obtaining a kidney in so far it is compatible with the evidence on the heterogeneity of compatible kidneys.

In this protective behavior scenario, we show that truth-ful revelation of patients reservation values can be attained despite the feasibility constraints. We show that if kidney exchanges are restricted to involve only pairs of donor-patient couples, then a plethora of rules provide (strong) incentives for the patients to report their true reservation values. Basically, any priority rule or a rule that maximizes a fixed ordering over the set of feasible and individually rational kidney assignments would work. Surprisingly, the positive result vanishes if larger exchanges are admitted. There are kidney allocation problems where if 3-way exchanges are possible, then truth-telling is a protectively dominated strategy.

¹¹See Su et al. (2004), Su and Zenios (2005).

Thus, in some sense, our results justify the possibility of introducing a pairwise kidney exchange in Europe, but also provide additional theoretical support beyond the logistic reasons for concentrating in the possibilities that arise in pairwise exchanges.

The remainder of the paper is organized as follows. In Section 2, we outline the model of kidney allocation problems and basic the notation and in Section 3, we introduce the concept of kidney allocation rule and some desirable conditions. In Section 4 we present the impossibility results. In Section 5, we define the protective behavior and the positive results. In Section 6, we conclude and discuss some lines of further research.

2 Kidney Allocation Problems

Consider a finite society consisting of a set $N = \{1, \dots, n\}$ of patients ($n > 3$) who need a kidney for transplantation. Each patient has a potential donor, and $\Omega = \{\omega_0, \omega_1, \dots, \omega_n\}$ denotes the set of kidneys available for transplantation. The kidney ω_0 refers to the situation in which a patient does not receive any kidney, while ω_i for each $i \neq 0$ refers to the kidney of patient i 's donor. We assume that each patient has only one potential donor and that there are not kidneys without living donor.¹²

Each patient i is equipped with a complete and transitive preference relation \succsim_i on Ω . Patients' preferences are based on rankings expressed, through objective, observable, and medical criteria that measure the fitness of each available kidney to each patient.¹³ We express patients' preferences by using numerical valuations over kidneys. For each $i, j \in N$, we denote by $v_i(\omega_j)$ represent i 's valuation of kidney ω_j . For each $i \in N$, $\omega, \omega' \in \Omega$, we say patient i considers kidney ω at least as good as kidney ω' – $\omega \succsim_i \omega'$ – if and only if $v_i(\omega) \geq v_i(\omega')$. Of course, given \succsim_i the associated strict preference relation \succ_i and the indifference relation \sim_i are defined in the standard way. We normalize in such a way that for each $i \in N$, and each $\omega \in \Omega \setminus \{\omega_0\}$, $v_i(\omega) \in [0, 1)$. If for some $i \in N$ and $\omega \in \Omega$ $v_i(\omega) = 0$, we say that patient i and kidney ω are *incompatible*. This possibility is meant to reflect the fact that patient i 's body will reject the graft of kidney ω , because of blood–type incompatibility, positive crossmatch, or any other reason. We say that patient

¹²Our main results go through if we introduce additional structure to the environment and admit multiple donors, or LPEs.

¹³These rankings may be based on the LYFT index –Life Years From Transplantation– or any other quality–efficiency criteria.

i and kidney ω are **compatible** if $v_i(\omega) > 0$. Without loss of generality, we assume that agents have strict preferences over compatible kidneys, and therefore for each $i \in N$ and each $\omega, \omega' \in \Omega$ if $v_i(\omega) \neq 0$ and $v_i(\omega') \neq 0$, then $v_i(\omega) \neq v_i(\omega')$. A **preference profile** is a matrix $\mathbf{P} \in \mathbb{M}_{n \times n}$ and it is defined by $P_{k,j} \equiv v_j(\omega_k)$ for each $i, j \in N$. Preference profiles contain all the information about patients' observable priority rankings.

Each patient i is also endowed with a reservation value $r_i \in (0, 1)$. We interpret r_i as the valuation that patient i assigns to receive ω_0 . Hence, $r_i \equiv v_i(\omega_0)$. Reservation values may incorporate patients' subjective valuation of being on dialysis and not receiving any kidney, as well as the endogenous expectation of receiving a new organ in the future from a new pool of donor–patient couples. We assume that for each patient i $r_i > 0$. Thus, patients always prefer to stay on dialysis rather than receiving an incompatible kidney. In order to be consistent with our assumption on strict preferences for compatible kidneys, we assume that patients are never indifferent between receiving a kidney and being maintained in the waiting list. Thus, for each i and each $\omega \neq \omega_0$, $r_i \neq v_i(\omega)$. Given a patient i and a preference profile \mathbf{P} , we denote by $R_i \equiv \{r_i \in (0, 1) \mid \forall \omega \in \Omega \setminus \{\omega_0\}, r_i \neq v_i(\omega)\}$ the set of i 's reservation values that are consistent with \mathbf{P} .¹⁴ Let $\mathcal{R} \equiv \times_{i \in N} R_i$. We call $\mathbf{r} \in \mathcal{R}$ patients' reservation values profile. For each $i \in N$ and each $\mathbf{r} \in \mathcal{R}$, \mathbf{r}_{-i} denotes the restriction of \mathbf{r} to the patients in $N \setminus \{i\}$.

A (**kidney exchange**) **problem** \mathbf{K} is a pair $\mathbf{K} = (\mathbf{P}, \mathbf{r})$.

An **assignment** a is an n -tuple of pairs $a = [(1, \omega), \dots, (n, \omega')]$ such that

- (i) for each $i, j \in N$, $i \neq j$ and each $\omega, \omega' \in \Omega \setminus \{\omega_0\}$, if $(i, \omega), (j, \omega') \in a$, then $\omega \neq \omega'$.
- (ii) if there are $i, j \in N$ such that $(i, \omega_j) \in a$, then $(j, \omega_0) \notin a$.

An assignment is an allocation of the available kidneys to the patients. By (i), an assignment never allocates the same kidney to two different patients, unless that kidney is the null kidney. By (ii), if the kidney of a patient's donor is allocated to another patient, then the initial patient is not allocated the null kidney. For each patient i and each assignment a , we denote by a_i the kidney that patient i receives at a .

¹⁴The reader should keep in mind that R_i depends on \mathbf{P} . We are abusing notation but since \mathbf{P} is always a primitive of the analysis, there will not be room for confusion in the arguments.

In every assignment, kidneys are allocated by forming exchange cycles of patient–donors couples. In each cycle, every patient receives a kidney from the donor of some patient in the cycle and simultaneously her donor’s kidney is transplanted to another patient in the cycle. In an exchange cycle among k couples, all the kidneys must be reaped from the donors and transplanted to the patients simultaneously. Moreover, the operations must be conducted in the same facility. If this constraints are not taken into account, it could be the case that once a donor’s kidney is transplanted to another patient, the donor of the recipient may reject to donate her kidney in order to avoid any clinical complication involved in the operation. This fact implies that an assignment among k couples involves $2k$ simultaneous operations. Since hospitals face evident logistic restrictions on the number of available operation rooms,¹⁵ we incorporate such constraints in our analysis through a narrower definition of feasible assignments.

For each assignment a , let π_a be the finest partition of the set of patients such that for each $p \in \pi_a$ and each $i \in p$:

- (i) either there are $j, j' \in p$, with $a_i = \omega_j$ and $a_{j'} = \omega_i$,¹⁶
- (ii) or $a_i = \omega_0$.

Clearly, for each assignment a the partition π_a is unique and well-defined. We define the **cardinality of a** as the $\max_{p \in \pi_a} \#p$.

The cardinality of an assignment refers to the size of the largest cycle formed in the assignment. Basically, it refers to the maximum number of simultaneous operations involved in an assignment. Of course, the concept of cardinality is crucial for our notion of feasibility.

For each $k \in \mathbb{N}$, $k \leq n$, we say that the assignment a is **k -feasible** if a ’s cardinality is not larger than k . Let \mathcal{A}^k be the set of all k -feasible assignments.

An interesting case of feasibility restrictions appears when only immediate exchanges between two couples are admitted. An assignment a is a *pairwise-exchange* assignment ($a \in \mathcal{A}^2$) if a satisfies that if for some $i, j \in N$ $(i, \omega_j) \in a$, then $(j, \omega_i) \in a$.

¹⁵That should go in the Introduction. Comment on Romanian 4s, and that 3 could be accepted as an exceptional upper bound.

¹⁶Note that $j = j'$ and $i = j = j'$ and then $a_i = \omega_i$ are allowed.

3 Kidney Assignment Rules

In this paper, we are interested in rules that select a (kidney assignment) for each (kidney exchange) problem. An (*assignment*) *rule* is a mapping φ that selects an assignment a for each problem $\mathbf{K} = (\mathbf{P}, \mathbf{r})$. For each patient i and each problem $\mathbf{K} = (\mathbf{P}, \mathbf{r})$, we denote by $\varphi_i(\mathbf{P}, \mathbf{r})$ the object assigned to i by φ at \mathbf{K} . As we take patients' preferences profile \mathbf{P} as given, whenever there is no room for confusion, we drop \mathbf{P} from the arguments and simply write $\varphi(\mathbf{r})$.

The assignment selected by a rule can be interpreted as an optimal recommendation that takes into account the preferences of the patients for the available kidneys and their reservation values and that tries to find a compromise between their (maybe conflicting) interests.

Next, we present formal definition of the standard conditions for desirable rules. Since we assume that patients' preferences over *all available kidneys are perfectly observable (and common knowledge)*, while the reservation values are private information for each patient, the reader should keep in mind that all the conditions refer to a given observed preference profile \mathbf{P} .

Individual Rationality For each $i \in N$ and each $\mathbf{r} \in \mathcal{R}$, $v_i(\varphi_i(\mathbf{r})) \geq \max\{r_i, v_i(\omega_i)\}$.

k -Efficiency For each $\mathbf{r} \in \mathcal{R}$, $\varphi(\mathbf{r}) \in \mathcal{A}^k$ and there is no $a \in \mathcal{A}^k$ such that for each $i \in N$ $a_i \succsim_i \varphi_i(\mathbf{r})$ and for some $j \in N$, $a_j \succ_j \varphi_j(\mathbf{P}, \mathbf{r})$.

Individual rationality is a minimal participation constraint which takes into account patient's right of refusing any transplant and receiving her donor's kidney. On the other hand, *k-efficiency* is the natural version of efficiency taking into account the feasibility restrictions on the cardinality of the assignments because at most $2k$ simultaneous operations can be carried out in the kidney exchange process. Of course, *n-efficiency* corresponds to the classical notion of (full) Pareto efficiency when there are not feasibility constraints.

4 Incentive-Compatibility and Feasibility Constraints

A central issue in the design of an optimal kidney exchange program is the use of all the relevant information in the assignment of the available kidneys. Although doctors may have all the information about the degree of compatibility and fitness among patients and kidneys that is imbedded in preference profiles, there is a key piece of information that remains private information for the patients and must be elicited for public use, their reservation values. The objective of this section is to analyze whether we can construct rules that provide incentives to the patients to reveal their true reservation values in the presence of feasibility constraints on the cardinality of the proposed assignments.

Before presenting the classical notion of truth-full revelation, and in order to provide a starker view of its implication, we introduce some additional definitions. Every preference profile \mathbf{P} together with a rule φ define a game form $\mathcal{G}(\mathbf{P}, \varphi)$. The game form specifies a set of players (the patients), a set of strategies for each patient (the sets R_i that are consistent with \mathbf{P}), and an outcome function ($\varphi(\mathbf{P}, \cdot)$). The game form $\mathcal{G}(\mathbf{P}, \varphi)$ fails short of being a game in normal (strategic) form because \mathbf{P} does not introduce the information about patients' preferences about the possibility of receiving the null kidney ω_0 . On the other hand, for every problem $\mathbf{K} = (\mathbf{P}, \mathbf{r})$ and each rule φ do define the *direct revelation game* $\Gamma(\mathbf{K}, \varphi)$. The direct revelation game $\Gamma(\mathbf{K}, \varphi)$ is the game where the set of players is the set of patients, the strategies for each patient i consists of announcing a reservation value in R_i , and for each profile of reservation values, the outcomes are defined by the rule φ , and patients preferences are defined by \mathbf{P} and \mathbf{r} .

Given that patients' valuations of available kidneys \mathbf{P} are public information, but patients' reservation values \mathbf{r} , we are interested in rules such that patients have incentives to reveal their true reservation values in all the direct revelation games that are induce the same game form $\mathcal{G}(\mathbf{P}, \varphi)$.

We are interested in a strong incentive compatibility property. Each patient knows her own reservation value, the observable preferences \mathbf{P} , and the rule φ . However, patients do not have information about the reservation values of the remaining patients. Moreover, it seems rather heroic to assume that patients may have beliefs about other patients reservation values, or correct (equilibrium) beliefs about their strategic behavior. In such incomplete information situations, it appears desirable to focus on rules that provide incentives to the patients to reveal their private information in all the direct revelation

games that induce the same game form $\mathcal{G}(\mathbf{P}, \varphi)$. Hence, reporting the true reservation value should be a (weakly) dominant strategy for every patient.

Strategy-proofness For each $i \in N$, each $\mathbf{r} \in \mathcal{R}$, and each $r'_i \in R_i$, $\varphi_i(\mathbf{r}) \succeq_i \varphi_i(r'_i, \mathbf{r}_{-i})$.

Note that *strategy-proofness* is weak in our framework because only reservation values are private information. The justification for the requirements of *strategy-proofness* is two-fold. On the normative side, if patients do not provide the correct reservation values, then the assignment selected by the rule may be based on incorrect information, and therefore it may represent a far from optimal recommendation to the society. On the positive side, in order to compute their best strategy in the direct revelation game, patients simply need to know their own reservation values.

The literature on the allocation of indivisible objects has extensively studied the problem of designing *strategy-proof* and *individual rational* assignment rules.¹⁷ When patients have strict preferences, there is a natural way to allocate the kidneys. We can simply use the Gale's top trading cycle procedure to allocate objects in markets with individual property rights. According to this procedure, given a kidney allocation problem, let every patient point to the donor with her favorite kidney. A top trading cycle consists of patients such that each patient in the cycle points to the donor of the next patient in the cycle. (A single patient may constitute a cycle, by pointing to herself if her donor's kidney is her best preferred kidney or if she thinks that no available kidney is acceptable and she prefers not to perform any operation). Since there is a finite number of patients and kidneys, for each problem there is at least one top trading cycle. Give each patient in a top trading cycle her best preferred kidney, and remove them from the problem with her assigned kidney. Repeat the process until each patient receives a kidney (maybe the null kidney). The resulting assignment is unique given that preferences over compatible kidneys are strict.¹⁸ Moreover, the induced rule satisfies *individual rationality*, *n-efficiency*, and *strategy-proofness*.¹⁹ However, in some problems a top trading cycle may involve all the patients and fail to select a k -feasible assignment for each $k < n$.

¹⁷See Gale and Shapley (1962); Shapley and Scarf (1974); Abdulkadiroğlu and Sönmez (1999).

¹⁸Patients will rather pick the null kidney rather than been assigned an incompatible kidney.

¹⁹See Roth and Postlewaite (1977); Roth (1982).

Example 1. Let $N = \{1, 2, 3, 4\}$. Consider the problem $\mathbf{K} = (\mathbf{P}, \mathbf{r})$ with

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 0 & 0.99 \\ 0.99 & 0 & 0 & 0 \\ 0 & 0.99 & 0 & 0 \\ 0 & 0.95 & 0.99 & 0 \end{pmatrix},$$

and for each $i \in N$, $r_i = 0.9$. In this problem, the top trading cycle procedure selects the assignment

$$\bar{a} = [(1, \omega_2)(2, \omega_3)(3, \omega_4)(4, \omega_1)].$$

Under \bar{a} , there is a top trading cycle that involves all the patients and every patient receives her best preferred kidney. However, rule that satisfies 3-efficiency must select the assignment a' such that

$$a' = [(1, \omega_2)(2, \omega_4)(3, \omega_0)(4, \omega_1)].$$

Clearly, for each $i \in N$, $\bar{a}_i \succ_i a'_i$ and $\bar{a}_2 \succ_2 a'_2$ but \bar{a} is not 3-feasible.

Our first result shows that feasibility constraints may make impossible to construct *efficient* rules that provide the right incentives to the patients at least at some preference profiles.

Theorem 1. For each $2 \leq k \leq n - 1$, there are \mathbf{P} such that no rule satisfies individual rationality, k -efficiency, and strategy-proofness.

Proof. We study separately two cases. We analyze first the restriction to pairwise exchanges. Then, we provide the proof for for $k \geq 3$. In both cases we exploit arguments similar to those employed in the literature of strategy-proof assignment rules in economies with indivisibilities where the core is empty ($k = 2$) or multivalued ($k \geq 3$).²⁰

Assume to the contrary there is a rule φ that satisfies *individual rationality*, *2-efficiency*, and *strategy-proofness* for every \mathbf{P} . Consider three patients $\{1, 2, 3\}$ and a preference profile \mathbf{P} such that its restriction to these patients and their donors' kidneys is:

$$\mathbf{P} = \begin{pmatrix} 0.5 & 0.75 & 0.99 \\ 0.99 & 0.5 & 0.75 \\ 0.75 & 0.99 & 0.5 \end{pmatrix},$$

²⁰See Sönmez (1999).

and so that for each $i \in \{1, 2, 3\}$, and each $\omega \notin \{\omega_0, \omega_1, \omega_2, \omega_3\}$, $v_i(\omega) = 0$. Thus, $\omega_2 \succ_1 \omega_3 \succ_1 \omega_1$, $\omega_3 \succ_2 \omega_1 \succ_2 \omega_2$, and $\omega_1 \succ_3 \omega_2 \succ_3 \omega_3$. By *individual rationality*, and in order to simplify notation, we can assume that $N = \{1, 2, 3\}$.

Let $\mathbf{r} = (r_1, r_2, r_3) = (0.6, 0.6, 0.6)$. By *individual rationality* and *2-efficiency*, φ selects an assignment in which two patients exchange their donors' kidneys while the remaining patient receives the null kidney. We assume without loss of generality that $\varphi(\mathbf{r}) = [(1, \omega_2), (2, \omega_1), (3, \omega_0)]$.

Next, let $\mathbf{r}' = (r'_1, r'_2, r'_3) = (0.9, 0.6, 0.6)$. By *strategy-proofness*, $\varphi_1(\mathbf{r}') = \omega_2$. Finally, let $\mathbf{r}'' = (r''_1, r''_2, r''_3) = (0.9, 0.9, 0.6)$. By *individual rationality* and *strategy-proofness*, $\varphi_2(\mathbf{r}) = \omega_0$. Then, $\varphi(\mathbf{r}'') = [(1, \omega_0), (2, \omega_0), (3, \omega_0)]$. Note that the assignment $a = [(1, \omega_0), (2, \omega_3), (3, \omega_2)]$ is *2-feasible*, and $a_i \succsim_i \varphi_i(\mathbf{r}'')$ for each $i \in N$ and $a_2 \succ_2 \varphi_2(\mathbf{r}'')$. Then, φ violates *2-efficiency*.

Next, we analyze the general case. Let $k \geq 3$. Remember that $k < n$ and then there are at least $k+1$ patients. Assume to the contrary there is a rule φ that satisfies *individual rationality*, *k-efficiency*, and *strategy-proofness* for every \mathbf{P} . Let the preference profile \mathbf{P} be such that for every $i = 1 \dots, k+1$:

$$v_i(\omega_{i+1}) > v_i(\omega_{i+2}) > v_i(\omega_i) > v_i(\omega) = 0, \quad \forall \omega \in \Omega \setminus \{\omega_0, \omega_i, \omega_{i+1}, \omega_{i+2}\}. \quad (\text{modulo } k+1).$$

By *individual rationality*, we can simplify notation and assume that $N = \{1, \dots, k+1\}$.

\succsim_1	\succsim_2	\dots	\succsim_{k-1}	\succsim_k	\succsim_{k+1}
ω_2	ω_3	\dots	ω_k	ω_{k+1}	ω_1
ω_3	ω_4	\dots	ω_{k+1}	ω_1	ω_2
ω_1	ω_2	\dots	ω_{k-1}	ω_k	ω_{k+1}
\dots	\dots	\dots	\dots	\dots	\dots

Table 1: \mathbf{P} : Theorem 1, Case $k \geq 3$.

Let $\mathbf{r} \in \mathcal{R}$ be such that:

$$\begin{aligned} v_i(\omega_{i+2}) < r_i < v_i(\omega_{i+1}) & \quad \text{for each } i \neq k+1 \\ v_{k+1}(\omega_{k+1}) < r_{k+1} < v_{k+1}(\omega_2) \end{aligned}$$

Note that, by *k-efficiency* and *individual rationality* either no object is assigned to any patient $1, \dots, k+1$, or patient $k+1$ receives ω_2 , patient 1 receives the null object,

and every other patient i receives ω_{i+1} (the kidney of her next to the right neighbor). By *k-efficiency*:

$$\varphi(\mathbf{r}) = \begin{bmatrix} (1, \omega_0), \\ (i, \omega_{i+1}), \quad \forall i = 2, \dots, k \\ (k+1, \omega_2) \end{bmatrix}.$$

Let $\mathbf{r}' \in \mathcal{R}$ be such that for each $i \neq k-1$, $r_i = r'_i$ and $v_{k-1}(\omega_{k-1}) < r_{k-1} < v_{k-1}(\omega_{k+1})$. By *strategy-proofness*, $\varphi_{k-1}(\mathbf{r}') \succeq_{k-1} \varphi_{k-1}(\mathbf{r}) = \omega_k$. Note that ω_k is patient $k-1$'s preferred kidney. Then, $\varphi_{k-1}(\mathbf{r}') = \omega_k$. By *k-efficiency* and *individual rationality*, $\varphi(\mathbf{r}) = \varphi(\mathbf{r}')$.

Let $\bar{\mathbf{r}} \in \mathcal{R}$ be such that for each $i \neq k+1$, $r'_i = \bar{r}_i$ and $v_{k+1}(\omega_2) < \bar{r}_{k+1} < v_{k+1}(\omega_1)$. The same arguments we employed to determine $\varphi(\mathbf{r})$ apply here to obtain:

$$\varphi(\bar{\mathbf{r}}) = \begin{bmatrix} (i, \omega_{i+1}) \text{ (modulo } k+1), \quad \forall i \notin \{k, k-1\} \\ (k-1, \omega_{k+1}), \\ (k, \omega_0) \end{bmatrix}.$$

Note that $\omega_1 = \varphi_{k+1}(\bar{\mathbf{r}}) = \varphi(\bar{r}'_{k+1}, \mathbf{r}'_{-(k+1)}) \succ_{k+1} \varphi_{k+1}(\mathbf{r}') = \omega_2$, which contradicts *strategy-proofness*. \square

Remark 1. *The previous impossibility result is robust to the introduction of weak preferences over kidneys. All we require is to admit the existence of two indifference classes for acceptable kidneys.*

5 Protective Behavior in Kidney Exchange Problems

The previous section presents a very negative result on designing allocation rules for kidney exchange problems when agents have heterogenous preferences over the set of compatible kidneys. This negative result is particularly discouraging since we balance the enrichment of preference domain with the increase in the information available to the planner. Namely, we make the assumption that patients' preferences over available kidneys are known by the planner because they depend on measurable and verifiable characteristics of patients and donors, like blood types of patients and donors, their age, health status, race, HLA mismatches, etc. Still, the fact that reservation values depend on unobservable patients' characteristics, put severe limitation on the properties that the allocation mechanism satisfies. Information on the patients' reservation values have to

be elicited and patients might be tempted to misreport such information to get better kidneys. However, if patients waiting for a transplant are strongly risk-averse, they might prefer to choose their strategies so as to "protect" themselves from the worst eventuality as far as possible. This "lexicographic maximin" behavior assumption is captured by the notion of "protective behavior".

Consider a problem $\mathbf{K} = (\mathbf{P}, \mathbf{r})$, a rule φ , and its associated direct revelation game $\Gamma(\mathbf{K}, \varphi)$. For each patient i , each $s_i \in R_i$, and each real number $k \in \mathbb{R}$, let:

$$c_i^{\Gamma(\mathbf{K}, \varphi)}(k, s_i) = \{\mathbf{s}_{-i} \in \mathcal{R}_{-i} \mid v_i(\varphi(\mathbf{P}, (s_i, \mathbf{s}_{-i}))) = k\}.$$

Then, $c_i^{\Gamma(\mathbf{K}, \varphi)}(k, s_i)$ is the set of restricted profiles of reservation values of the remaining patients under which i receives a kidney ω with $v_i(\omega) = k$ when i announces that her reservation value is s_i .

For each patient $i \in N$ and each problem \mathbf{K} , $s_i, s'_i \in R_i$, s_i **protectively dominates** s'_i (**at** $\Gamma(\mathbf{K}, \varphi)$), denoted $s_i \text{ d}(\mathbf{K}, \varphi) s'_i$ if there exists $k \in \mathbb{R}$ such that:

- (i) $c_i^{\Gamma(\mathbf{K}, \varphi)}(t, s_i) \cap c_i^{\Gamma(\mathbf{K}, \varphi)}(t', s'_i) = \emptyset$ for each $t \leq k$ and $t < t'$,
- (ii) $c_i^{\Gamma(\mathbf{K}, \varphi)}(k, s_i) \subset c_i^{\Gamma(\mathbf{K}, \varphi)}(k, s'_i)$.

Let $D_i(\mathbf{K}, \varphi) \equiv \{s_i \in S_i \mid \text{there is no } s'_i \in S_i \text{ with } s'_i \text{ d}(\mathbf{K}, \varphi) s_i\}$ be the set of **protective strategies** of patient i in the direct revelation game $\Gamma(\mathbf{K}, \varphi)$.

In order to compare two strategies according to this criterion an agent looks at the utility level of the worst outcome (say $l = \min_{\omega \in \Omega} v_i(\omega)$). Strategy s_i protectively dominates s'_i if two conditions hold. First, it never occurs that there exists a profile such that strategy s_i induces this minimum utility level and strategy s'_i induces a higher level of utility. Second, there are some profiles such that s'_i induces the minimum level of utility, while s_i induces a larger payoff for patient i . If the first condition holds but not the second because $c_i^{\Gamma(\mathbf{K}, \varphi)}(l, s_i) = c_i^{\Gamma(\mathbf{K}, \varphi)}(l, s'_i)$, then agent i considers how s_i and s'_i do perform respect to the next to the worst utility level.

Clearly, protective domination relation is not complete, but it is transitive. Thus, for each patient i , each problem \mathbf{K} , and each rule φ , the set $D_i(\mathbf{K}, \varphi)$ is not empty. Moreover, if there is a unique protective strategy, $\{s_i\} = D_i(\mathbf{K}, \varphi)$, then $s_i \text{ d}(\mathbf{K}, \varphi) s'_i$ for each $s'_i \in R_i \setminus \{s_i\}$.

Next, we introduce the precise notion of truth-full strategies under protective behavior. Consider a preference profile \mathbf{P} and a rule φ . Truth-telling requires that reporting the true reservation value is a protective strategy for the player at every direct revelation game generated by \mathbf{P} and φ . However, protective domination is not a complete relation. Thus, if there are several different protective strategies besides reporting the true value, then we could not say that truth-telling is an optimal strategy for the patients because, it is not clear that is less risky than the remaining protective strategies. This fact calls for a stronger requirement, in such a way that revealing the true reservation value should imply unequivocally less risk than the remaining strategies. In any case, our definition is not strong enough to preclude the existence of strategically equivalent strategies.

For each patient i , each problem $\mathbf{K} = (\mathbf{P}, \mathbf{r})$, and each rule φ , strategies $s_i, s'_i \in R_i$ are **equivalent at** $\Gamma(\mathbf{K}, \varphi)$ if for each $\mathbf{s}_{-i} \in \mathcal{R}_{-i}$, $v_i(\varphi((s_i, \mathbf{s}_{-i}))) = v_i(\varphi((s'_i, \mathbf{s}_{-i})))$.

Fix a preference profile \mathbf{P} and a rule φ . **Truth-telling is a protective strategy for patient i (at $\mathcal{G}(\mathbf{P}, \varphi)$)** if for each $\mathbf{r} \in \mathcal{R}$, $r_i \in D_i((\mathbf{P}, \mathbf{r}), \varphi)$.

Analogously, **truth-telling is the unique protective strategy for $i \in N$ (at $\mathcal{G}(\mathbf{P}, \varphi)$)** if for each $\mathbf{r} \in \mathcal{R}$, $r_i \in D_i((\mathbf{P}, \mathbf{r}), \varphi)$ and every $s_i \in D_i((\mathbf{P}, \mathbf{r}), \varphi)$ is equivalent to r_i .

The following example shows the difference between truth-telling as a unique protective strategy and *strategy-proofness*.

Example 2. Let ψ be the rule that maximizes the number of transplants obtained through pairwise exchanges. (Ties are broken according to patient 1's preferences). It is easy to see that there are problems where ψ does not satisfy strategy-proofness because patients may successfully manipulate at some profile in which they are getting a utility larger than their reservation value. Consider the problem $\mathbf{K} = (\mathbf{P}, \mathbf{r})$ such that

$$\mathbf{P} = \begin{pmatrix} 0.1 & 0.5 & 0.5 \\ 0.9 & 0.1 & 0.9 \\ 0.5 & 0.9 & 0.1 \end{pmatrix}.$$

and $\mathbf{r} = (0.2, 0.6, 0.2)$. Then $\psi(\mathbf{r}) = [(1, \omega_3), (2, \omega_0), (3, \omega_1)]$. If patient 3 announces $r'_3 = 0.6$, then $\psi(r'_3, \mathbf{r}_{-3}) = [(1, \omega_0), (2, \omega_3), (3, \omega_2)]$. Note that $\omega_2 \succ_3 \omega_1$, which contradicts

strategy-proofness. Nevertheless, $r_3 = 0.2$ is the unique (modulo equivalence) protectively dominant strategy at $\Gamma(\mathbf{K}, \psi)$. To check why to announce $r'_3 = 0.6$ when $r_3 = 0.2$ is protectively dominated by $r_3 = 0.2$ at $\Gamma(\mathbf{K}, \psi)$ consider the reservation values profile $\mathbf{r}' = (0.4, 0.92, 0.6)$. It follows that $\psi(\mathbf{r}') = [(1, \omega_0), (2, \omega_0), (3, \omega_0)]$ and $\psi(r_3, \mathbf{r}'_{-3}) = [(1, \omega_3), (2, \omega_0), (3, \omega_1)]$. Hence $\{(r_1 = 0.4, r_2 = 0.92)\} \in c_i^{\Gamma(\mathbf{K}, \psi)}(0.2, r'_3) \cap c_i^{\Gamma(\mathbf{K}, \psi)}(0.5, r_3)$ violating (i) of the definition of protective domination. It is possible to follow similar arguments with every other strategy $s_3 \in R_3$ to check that either s_3 is equivalent to r_3 or it is protectively dominated by r_3 . Finally note that, if patients preferences are dichotomous and it is assumed that patients are indifferent among every kidney ω such that $r_i < v_i(\omega)$, φ satisfies strategy-proofness.

In order to follow with the analysis of truth-telling as a protective strategy, we need to introduce additional notation and definitions. Consider a patient i and a preference profile \mathbf{P} . For each rule φ , the set of possible outcomes for patient i at $\mathcal{G}(\mathbf{P}, \varphi)$ is defined by:

$$\Omega_i(\mathbf{P}, \varphi) \equiv \{\omega \in \Omega \text{ such that there exists } \mathbf{r} \in \mathcal{R} \text{ with } \varphi_i(\mathbf{P}, \mathbf{r}) = \omega\}.$$

With the definition of $\Omega_i(\mathbf{P}, \varphi)$ at hand, for every problem $\mathbf{K} = (\mathbf{P}, \mathbf{r})$ the set of acceptable kidneys for i is defined as:

$$\Omega_i^+(\mathbf{K}, \varphi) \equiv \{\omega \in \Omega_i(\mathbf{P}, \varphi) \text{ such that } v_i(\omega) \geq \max\{v_i(\omega_i), r_i\}\}.$$

Of course, i 's set of unacceptable kidneys is analogously defined:

$$\Omega_i^-(\mathbf{K}, \varphi) \equiv \{\omega \in \Omega_i(\mathbf{P}, \varphi) \text{ such that } v_i(\omega) < \max\{v_i(\omega_i), r_i\}\}.$$

Of course, if φ satisfies *individual rationality*, then $\varphi_i(\mathbf{r}) \in \Omega_i^+(\mathbf{K}, \varphi) \neq \emptyset$. Finally let $\omega_i^+(\mathbf{K}, \varphi) \equiv \arg \min_{\omega \in \Omega_i^+(\mathbf{K}, \varphi) \setminus \{\omega_0\}} v_i(\omega)$ if $\Omega_i^+(\mathbf{K}, \varphi) \neq \emptyset$, $\omega_i^+(\mathbf{K}, \varphi) \equiv \emptyset$ otherwise; and $\omega_i^-(\mathbf{K}, \varphi) \equiv \arg \max_{\omega \in \Omega_i^-(\mathbf{K}, \varphi)} v_i(\omega)$ if $\Omega_i^-(\mathbf{K}, \varphi) \neq \emptyset$, $\omega_i^-(\mathbf{K}, \varphi) \equiv \emptyset$ otherwise. Thus, $\omega_i^+(\mathbf{K}, \varphi)$ is i 's worst acceptable kidney and $\omega_i^-(\mathbf{K}, \varphi)$ is i 's best unacceptable kidney. Note that if patient i 's reservation value is lower than her donor's kidney valuation ($r_i < v_i(\omega_i)$), then $\omega_i^+(\mathbf{K}, \varphi) = \omega_i$, and $\omega_i^-(\mathbf{K}, \varphi) = \emptyset$.

At this point, we introduce two weak conditions that turn out to be necessary for truth-telling being a unique protective strategy in our environment. Again both properties refer to a given preference profile \mathbf{P} .

Invariance. For each $i \in N$, and each pair $\mathbf{r}, \mathbf{r}' \in \mathcal{R}$ such that $\mathbf{r}_{-i} = \mathbf{r}'_{-i}$, if $\Omega_i^+(\mathbf{P}, \mathbf{r}, \varphi) = \Omega_i^+(\mathbf{P}, \mathbf{r}', \varphi)$, then $\varphi_i(\mathbf{P}, \mathbf{r}) = \varphi_i(\mathbf{P}, \mathbf{r}')$.

Weak Consistency. For each $i \in N$, each $\mathbf{r} \in \mathcal{R}$, and each $r_i \in R_i$, if $\varphi_i(\mathbf{r}) = \omega_0$ and $r_i < r'_i$, then $\varphi_i(\mathbf{r}_{-i}, r'_i) = \omega_0$.

Invariance requires that if a patient changes her reported reservation value, but this change does not affect her set of acceptable kidneys, then the patient receives a kidney with the same valuation than the original one. Note that *invariance* does not imply ordinality. A rule satisfying *invariance* may be responsive to the cardinal information of patients' reported reservation values. For instance, think of a serially dictatorial (priority) rule that always picks the best feasible allocation for a given patient, and then proceeds iteratively (serially breaking ties) according to a priority list that depends on the reservation value reported by that first patient in the list. *Weak Consistency* is a convenient weakening of the Axiom of Choice for single-valued choice functions.²¹ Note that if a rule satisfies *individual rationality*, then *weak consistency* applies the logic behind the Axiom of Choice only at situations where the patients receive the worst possible outcome. That is, to situations where the patient receives the null kidney and obtaining her reservation value. Simply, if a patient does not receive a kidney when she reports r_i , then she cannot be awarded a kidney when she raises her reservation value.

In the following proposition, we show that *invariance* and *weak consistency* are necessary for truth-telling being a unique protective strategy for any preference \mathbf{P} .

Proposition 1. *For each preference profile \mathbf{P} and each rule φ , if φ satisfies individual rationality and for each $i \in N$ truth telling is the unique protective strategy at $\mathcal{G}(\mathbf{P}, \varphi)$, then φ satisfies invariance and weak consistency.*

Proof. Fix $i \in N$ and $\mathbf{K} = (\mathbf{P}, \mathbf{r})$. We start with the proof of *invariance*.

Assume first that $r_i < v_i(\omega_i)$. In this case, $\Omega_i^-(\mathbf{K}, \varphi) = \emptyset$ and we need to prove that every $s_i \in R_i$ such that $s_i < v_i(\omega^+(\mathbf{K}, \varphi))$, is equivalent to r_i at $\Gamma(\mathbf{K}, \varphi)$. Let $s_i < v_i(\omega_i)$. By *individual rationality*, for each $\bar{\mathbf{s}}_{-i} \in \times_{j \neq i} \mathcal{R}_j$, $v_i(\varphi_i(s_i, \bar{\mathbf{s}}_{-i})) \geq v_i(\omega_i)$.

²¹See Arrow (1959) and Sen (1971), we follow Hatfield (2005) in the terminology.

Let $\bar{\mathbf{r}}_{-i} \in \times_{j \neq i} \mathcal{R}_j$ be such that $\varphi(r_i, \bar{\mathbf{r}}_{-i}) = \omega_i$. Because truth-telling is the unique protective strategy for patient i , $c_i^{\Gamma(\mathbf{K}, \varphi)}(v_i(\omega_i), r_i) \subseteq c_i^{\Gamma(\mathbf{K}, \varphi)}(v_i(\omega_i), s_i)$ and $\varphi(s_i, \bar{\mathbf{r}}_{-i}) = \omega_i$. Consider now the problem $\mathbf{K}' = (\mathbf{P}, \mathbf{r}')$ such that $r'_i = s_i$ and $\mathbf{r}'_{-i} = \bar{\mathbf{r}}_{-i}$. Repeating the previous argument, *individual rationality* and the fact that truth-telling is the unique protective strategy for i , imply that $c_i^{\Gamma(\mathbf{K}', \varphi)}(v_i(\omega_i), s_i) \subseteq c_i^{\Gamma(\mathbf{K}', \varphi)}(v_i(\omega_i), r_i)$. Note that given φ , both \mathbf{K} and \mathbf{K}' are associated to the same game form $-\mathcal{G}(\mathbf{P}, \mathbf{r})-$. Hence, $c_i^{\Gamma(\mathbf{K}, \varphi)}(v_i(\omega_i), r_i) = c_i^{\Gamma(\mathbf{K}', \varphi)}(v_i(\omega_i), r_i)$, $c_i^{\Gamma(\mathbf{K}, \varphi)}(v_i(\omega_i), s_i) = c_i^{\Gamma(\mathbf{K}', \varphi)}(v_i(\omega_i), s_i)$, and we obtain $c_i^{\Gamma(\mathbf{K}, \varphi)}(v_i(\omega_i), s_i) = c_i^{\Gamma(\mathbf{K}, \varphi)}(v_i(\omega_i), r_i)$. We can apply the same argument iteratively for each level $k > r_i$, to conclude that s_i and r_i are equivalent.

Next, assume that $r_i > v_i(\omega_i)$ and that $\Omega_i^-(\mathbf{K}, \varphi) \neq \emptyset$.²² Hence, we need to check that every $s_i \in \mathcal{R}_i$ such that

$$v_i(\omega^-(\mathbf{K}, \varphi)) < s_i < v_i(\omega^+(\mathbf{K}, \varphi)),$$

is equivalent to r_i . Let $s_i \in (v_i(\omega^-(\mathbf{K}, \varphi)), v_i(\omega^+(\mathbf{K}, \varphi)))$. By *individual rationality* and the definitions of $\omega_i^-(\mathbf{K}, \varphi)$ and $\omega_i^+(\mathbf{K}, \varphi)$, for each $\bar{\mathbf{s}}_{-i} \in \times_{j \neq i} \mathcal{R}_j$, $v_i(\varphi_i(s_i, \bar{\mathbf{s}}_{-i})) \geq r_i$. Let $\hat{\mathbf{r}}_{-i} \in \times_{j \neq i} \mathcal{R}_j$ be such that $\varphi_i(r_i, \hat{\mathbf{r}}_{-i}) = \omega_0$. Because truth-telling is the unique protective strategy for patient i , $c_i^{\Gamma(\mathbf{K}, \varphi)}(r_i, r_i) \subseteq c_i^{\Gamma(\mathbf{K}, \varphi)}(r_i, s'_i)$. and $\varphi(s_i, \hat{\mathbf{r}}_{-i}) = \omega_0$. Consider now the problem $\mathbf{K}'' = (\mathbf{P}, \mathbf{r}'')$ such that $r''_i = s_i$ and $\mathbf{r}''_{-i} = \hat{\mathbf{r}}_{-i}$. Repeating the argument of the previous paragraph, *individual rationality* and the fact that truth-telling is the unique protective strategy for i imply that $c_i^{\Gamma(\mathbf{K}'', \varphi)}(r_i, s'_i) \subseteq c_i^{\Gamma(\mathbf{K}'', \varphi)}(r_i, r_i)$. Note again that given φ , both \mathbf{K} and \mathbf{K}'' are associated to the same game form $-\mathcal{G}(\mathbf{P}, \mathbf{r})$. Hence, $c_i^{\Gamma(\mathbf{K}, \varphi)}(r_i, r_i) = c_i^{\Gamma(\mathbf{K}'', \varphi)}(s_i, r_i)$, $c_i^{\Gamma(\mathbf{K}, \varphi)}(r_i, s_i) = c_i^{\Gamma(\mathbf{K}'', \varphi)}(s_i, s_i)$, what yields $c_i^{\Gamma(\mathbf{K}, \varphi)}(r_i, s_i) = c_i^{\Gamma(\mathbf{K}, \varphi)}(r_i, r_i)$. We can apply the arguments of the previous paragraph iteratively for each level $k > r_i$, to conclude that s_i and r_i are equivalent. Finally note that if $r_i > v_i(\omega_i)$ and $\Omega_i^-(\mathbf{K}, \varphi) = \emptyset$, the arguments of this paragraph apply without changes to prove that every $s_i < v_i(\omega_i^+(\mathbf{K}, \varphi))$ is equivalent to s_i , which completes the proof of *invariance*.

We conclude with the proof of *weak consistency*. Note that by *individual rationality*, for each $k < \max\{r_i, v_i(\omega_i)\}$, $c^{\Gamma(\mathbf{K}, \varphi)}(k, r_i) = \emptyset$ and $c^{\Gamma(\mathbf{K}, \varphi)}(\max\{r_i, v_i(\omega_i)\}, r_i) \neq \emptyset$. Assume that φ does not satisfy *weak consistency*, then there are $i \in N$, and $r_i < r'_i$ such that $\varphi_i(\mathbf{r}) = \omega_0$ and $\varphi_i(r'_i, \mathbf{r}_{-i}) \neq \omega_0$. Note that, by *individual rationality*, $r_i > v_i(\omega_i)$. Then, there is $\mathbf{r}' \in \mathcal{R}$ with $r'_i > r_i$ such that $\varphi_i(r_i, \mathbf{r}'_{-i}) = \omega_0$ but $\varphi_i(\mathbf{r}') \neq \omega_0$. By

²²By *individual rationality*, this is always the case if $v_i(\omega_i) \neq 0$.

individual rationality, $\varphi_i(\mathbf{r}') \in \Omega_i^+(\mathbf{K}, \varphi) \setminus \{\omega_0\}$. Hence, by (i) of the definition of protective domination, r_i does not protectively dominates r'_i . Moreover, r_i and r'_i are not equivalent strategies at $\Gamma(\mathbf{K}, \varphi)$. These facts contradict that truth-telling is the unique protective strategy. \square

The next proposition shows that *invariance* and *weak consistency* are also sufficient if only pairwise exchanges are admitted.

Proposition 2. *For each preference profile \mathbf{P} , if the rule φ that satisfies individual rationality, 2-efficiency, invariance, and weak consistency, then for each patient i truth-telling is the unique protective strategy at $\mathcal{G}(\mathbf{P}, \varphi)$.*

Proof. Fix a patient $i \in N$ and a problem $\mathbf{K} = (\mathbf{P}, \mathbf{r})$. By *invariance*, we need only prove that r_i protectively dominates every strategy $s_i \in R_i$ such that $s_i \notin [\omega_i^-(\mathbf{K}, \varphi), \omega_i^+(\mathbf{K}, \varphi)]$ if $\omega_i^-(\mathbf{K}, \varphi) \neq \emptyset$, and every $s'_i \notin (0, \omega_i^+(\mathbf{K}, \varphi)]$ if $\omega_i^-(\mathbf{K}, \varphi) = \emptyset$.

Assume first that $r_i < v_i(\omega_i)$. Then, $\omega_i^+(\mathbf{K}, \varphi) = \omega_i$. Note that, by *individual rationality*, for each $t < v_i(\omega_i)$, $c_i^{\Gamma(\mathbf{K}, \varphi)}(t, r_i) = \emptyset$. Let $s_i \in R_i$ be such that $s_i > v_i(\omega_i) = \omega_i^+(\mathbf{K}, \varphi)$. Clearly, for each $t \leq r_i$ and each $t' > t$, $c_i^{\Gamma(\mathbf{K}, \varphi)}(t, r_i) \cap c_i^{\Gamma(\mathbf{K}, \varphi)}(t', s_i) = \emptyset$, which proves the (ii) of protective domination. Next, consider $\hat{\mathbf{r}} \in \mathcal{R}$ such that $\hat{r}_i = s_i$ and for each $j \neq i$ and each $\omega \in \Omega \setminus \{\omega_0\}$, $\hat{r}_j > v_j(\omega)$. By *individual rationality*, for each $j \in N$ $\varphi_j(\hat{\mathbf{r}}) = \omega_0$. Then, $c_i^{\Gamma(\mathbf{K}, \varphi)}(r_i, r_i) = \emptyset \subset c_i^{\Gamma(\mathbf{K}, \varphi)}(r_i, s_i) \neq \emptyset$, which proves condition (ii) of protective domination.

Next, assume that $r_i > v_i(\omega_i)$ and assume that $\omega_i^-(\mathbf{K}, \varphi) \neq \emptyset$. Let $s'_i \in R_i$ be such $s'_i < v_i(\omega_i^-(\mathbf{K}, \varphi))$. By *individual rationality*, $c_i^{\Gamma(\mathbf{K}, \varphi)}(t, r_i) = \emptyset$ for all $t \leq v_i(\omega_i^-(\mathbf{K}, \varphi))$. Let $j \in N$ be such that $\omega_i^-(\mathbf{K}, \varphi) = \omega_j$.²³ Consider the reservation values profile $\mathbf{r}' \in \mathcal{R}$ such that $r'_l = s'_l$, for each $l \notin \{i, j\}$ and each $\omega \in \Omega \setminus \{\omega_0\}$, $r'_l > v_l(\omega)$; and $v_j(\omega_i) > r'_j$. (Note that this is possible because $\omega_i^-(\mathbf{K}, \varphi) \in \Omega_i(\mathbf{P}, \varphi)$.) By *individual rationality* and *k-efficiency*, $\varphi_i(\mathbf{r}') = \omega_j$ and $\varphi_i(r_i, \mathbf{r}'_{-i}) = \omega_0$. Clearly, $v_i(\varphi_i(r_i, \mathbf{r}'_{-i})) \geq r_i > v_i(\varphi_i(\mathbf{r}'))$. Hence $c_i^{\Gamma(\mathbf{K}, \varphi)}(v_i(\varphi_i(\mathbf{r}')), r_i) \subset c_i^{\Gamma(\mathbf{K}, \varphi)}(v_i(\varphi_i(\mathbf{r}')), s'_i) \neq \emptyset$. This suffices to prove that r_i protectively dominates s'_i .

Finally, we have to check that if $r_i > v_i(\omega_i)$, r_i protectively dominates every $s''_i \in R_i$ such that $s''_i > v_i(\omega_i^+(\mathbf{K}, \varphi)) \geq r_i$. By *individual rationality*, for each $t < r_i$, $c_i^{\Gamma(\mathbf{K}, \varphi)}(t, r_i) =$

²³It is possible that $j = i$.

\emptyset . Note that for each $\mathbf{r}^* \in \mathcal{R}$ such that $r_i^* = r_i$ and $\varphi_i(\mathbf{r}_{-i}^*) = \omega_0$, and each $s_i > r_i$, by *weak consistency*, $\varphi_i(s_i, \mathbf{r}_{-i}^*) = \omega_0$. Then, for each $t' > r_i$ we have $c_i^{\Gamma(\mathbf{K}, \varphi)}(r_i, r_i) \cap c_i^{\Gamma(\mathbf{K}, \varphi)}(t', s_i'') = \emptyset$ which proves condition (i) of the definition of protective domination. Moreover, we also have that $c_i^{\Gamma(\mathbf{K}, \varphi)}(r_i, r_i) \subseteq c_i^{\Gamma(\mathbf{K}, \varphi)}(r_i, s_i'')$. In order to conclude the argument, we show that the previous inclusion is proper. Let $j \in N$ be such that $\omega^+(r_i) = \omega_j$. Consider the reservation values profile $\mathbf{r}_{-i}'' \in \mathcal{R}$ such that $r_i'' = s_i''$, for each $m \notin \{i, j\}$ and each $\omega \in \Omega \setminus \{\omega_0\}$, $v_m(\omega) < r_m$; and $v_j(\omega_j) > r_j''$. By *individual rationality* and *k-efficiency*, $\varphi_i(r_i, \mathbf{r}_{-i}'') = \omega_j$ and $\varphi_i(\mathbf{r}'') = \omega_0$. Hence $c_i^{\Gamma(\mathbf{K}, \varphi)}(r_i, r_i) \subset c_i^{\Gamma(\mathbf{K}, \varphi)}(r_i, r_i'')$ which proves condition (ii) of the definition of protective domination. \square

The results in Propositions 1–2 are clearly positive. Under protective behavior, there is a wide class of rules that provide incentives for the patients to reveal their true reservation values. This family includes every rule that maximizes a strict order over the set of efficient and individually rational pairwise assignments. Hence, rules that maximize the number of exchanges, or serial priority rules satisfy our axioms. In addition, the previous results are in line with the results in Roth et al. (2005a) and Hatfield (2005) despite we start from different assumptions on patients' preferences and behavior.

The previous result shows that *weak consistency* is almost sufficient for providing the right incentives to extremely protective patients if only pairwise exchanges are permitted. In the light of the arguments of the proof of Proposition 2, the result extends to preference profiles and rules in which for every patient every kidney in the set of possible outcomes can be obtained through a pairwise exchange.²⁴ However, this observation does not hold generally. The next proposition proves that if larger cycles are possible, truth-telling may fail to be a protective strategy for every *individual rationality* and *k-efficiency*. Our last result shows that the assumption on 2-feasibility is essential for our positive result. Thus, we find a strategic *rationale* for the restriction to pairwise exchanges beyond the logistic and direct incentives problems described by Roth et al. (2005a). Under protective behavior, the restriction to pairwise exchanges may be necessary in order to obtain the correct information from the patients.

Proposition 3. *Let φ be a rule that satisfies individual rationality and k-efficiency for some $k \geq 3$. There are preference profiles \mathbf{P} such that if for some $\mathbf{r} \in \mathcal{R}$ $\#\varphi(\mathbf{r}) > 2$ then truth-telling is a protectively dominated strategy in $\Gamma((\mathbf{P}, \mathbf{r}), \varphi)$*

²⁴Consider for instance the preference profile presented in the proof of the case $k = 2$ of Theorem 1.

Proof. The proof consists of the following counterexample. Let $N = \{1, 2, 3\}$ and consider a preference profile \mathbf{P} such that

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 0.75 \\ 0.9 & 0 & 0.9 \\ 0 & 0.9 & 0 \end{pmatrix}.$$

Consider a rule φ that satisfies *individual rationality* and *3-efficiency*. *Individual rationality* and *3-efficiency* imply that φ is defined according to Table 1 and Table 2.

$r_2 \setminus r_3$	$r_3 > 0.9$	$r_3 \in (0.75, 0.9)$	$r_3 < 0.75$
$r_2 > 0.9$	$(\omega_0, \omega_0, \omega_0)$	$(\omega_0, \omega_0, \omega_0)$	$(\omega_0, \omega_0, \omega_0)$
$r_2 < 0.9$	$(\omega_0, \omega_0, \omega_0)$	$(\omega_0, \omega_3, \omega_2)$	$(\omega_0, \omega_3, \omega_2)$

Table 2: $r_1 > 0.9$

$r_2 \setminus r_3$	$r_3 > 0.9$	$r_3 \in (0.75, 0.9)$	$r_3 < 0.75$
$r_2 > 0.9$	$(\omega_0, \omega_0, \omega_0)$	$(\omega_0, \omega_0, \omega_0)$	$(\omega_0, \omega, \omega_0)$
$r_2 < 0.9$	$(\omega_0, \omega_0, \omega_0)$	$(\omega_0, \omega_3, \omega_2)$	either $(\omega_0, \omega_3, \omega_2)$, or $(\omega_2, \omega_3, \omega_1)$

Table 3: $r_1 < 0.9$

Assume now that there is $\mathbf{r}' \in \mathcal{R}$ such that $\#\varphi(\mathbf{r}') = 3$. That is, there is \mathbf{r}' such that $\varphi(\mathbf{r}') = (\omega_2, \omega_3, \omega_1)$. Necessarily, $r'_1 < 0.9$, $r'_2 < 0.9$, and $r'_3 < 0.75$. Consider the problem $\mathbf{K}' = (N, \mathbf{P}, \mathbf{r}')$ and its associated direct revelation game $\Gamma(\mathbf{K}', \varphi)$. It is immediate that $c_3^{\Gamma(\mathbf{K}', \varphi)}(r'_3, r'_3) = c_3^{\Gamma(\mathbf{K}', \varphi)}(r'_3, 0.8) = \{(r_1, r_2) \mid r_2 > 0.9\}$. On the other hand,

$$c_3^{\Gamma(\mathbf{K}', \varphi)}(0.75, r'_3) \neq \emptyset, \quad c_3^{\Gamma(\mathbf{K}', \varphi)}(0.75, 0.8) = \emptyset,$$

Note that

$$\varphi_3(r'_1, r'_2, 0.8) = \omega_2 \succ_3 \omega_1 = \varphi_3(\mathbf{r}').$$

Then, $s_3 = 0.8$ protectively dominates r'_3 at the game $\Gamma(\mathbf{K}', \varphi)$. Note that, if her true reservation value is r'_3 but she reports $s'_3 = 0.8$, patient 3 is not taking any risk at the other patients' profiles for which she receives the null kidney. However, in the case of

receiving an acceptable kidney, by reporting $s_3 = 0.8$ she always gets her best preferred kidney. Interestingly, we have to move beyond the first round of comparisons between strategies to check domination. \square

6 Concluding Remarks

In this paper, we have proposed a model that retains the flavor of the formulation by Roth et al. (2005a) that has been adopted in the design of a kidney exchange clearing-house in New England, but it departs in two important aspects from their model. We assume that patients may have heterogeneous preferences over the set of compatible kidneys, enlarging the relevant domain of preferences, but at the same time, we assume that the social planner may avail with detailed information about patients' preferences. Our first batch of results shows the difficulties to fulfill different forms of incentive compatibility if there are restrictions on the cardinality of feasible exchanges, even under the assumptions that the set of available information is large. However, positive results are restored if patients are strongly averse to the risk of refusing a transplant of a compatible kidney. Namely, those rules which are strategy-proof and efficient in the dichotomous domain, still satisfy the desired properties when agents have heterogeneous preferences and adopt protective behavior and only pairwise exchanges are feasible. Interestingly, the difficulties return if larger exchanges are admitted. These results have strong policy implications. If our assumptions fit real life situations, then the efficiency gains of making possible cycles larger than pairwise exchanges can be overcome by the impossibility of eliciting truthful information from patients and the inefficiency that could derive from the strategic manipulation of revealed preferences. Since the cost of slackening the feasibility constraints are high (3-feasibility means having six operating rooms and six surgical teams available at the same time), then our work puts some doubts on the economic advantage of these investments for the healthcare service.

In order to conclude, we devote a few lines to sketch some open venues of further research. First of all the assumption on patients' (protective) behavior deserves to be tested by means of controlled questionnaires on the population of patients in the waiting lists. Second, incentive problems in kidney exchange environments have been studied on static model as ours. However, it is evident that kidney transplantation has a dynamic component that has been neglected (Su and Zenios 2005 is a remarkable exception). It

seems a promising line of new research the analysis of dynamic and strategic models where patients and kidneys are available sequentially and simultaneously living donation and kidney exchange are feasible procedures. In the light of the technical difficulties that appear in standard queue-management models, the analysis of protective behavior in such settings is a promising line of investigation.

References

- A. Abdulkadiroğlu and T. Sönmez. House allocation with existing tenants. *Journal of Economic Theory*, 88:233–260, 1999.
- A. Abdulkadiroğlu and T. Sönmez. School choice: A mechanism design approach. *American Economic Review*, 93-3:729–747, 2003.
- K. J. Arrow. Rational choice functions and orderings. *Economica*, 26:121–127, 1959.
- S. Barberà and B. Dutta. Implementability via protective equilibria. *Journal of Mathematical Economics*, 10:49–65, 1982.
- S. Barberà and M. O. Jackson. Maximin, leximin, and the protective criterion: characterizations and comparisons. *Journal of Economic Theory*, 46:34–44, 1988.
- F. L. Delmonico. Exchanging kidneys – advances in living-donor transplantation. *The New England Journal of Medicine*, 350:1812–1814, 2004.
- F. L. Delmonico, G. S. Lipkowitz, P. E. Morrissey, J. S. Stoff, J. Himmelfarb, W. Harmon, M. Pavlakis, H. Mah, J. Goguen, R. Luskin, E. Milford, G. B. M. Chobanian, B. Bouthot, M. Lorber, and R. J. Rohrer. Donor kidney exchanges. *American Journal of Transplantation*, 4:1628–1634, 2004.
- R. J. Duquesnoy, S. Takemoto, P. de Lange, I. I. N. Doxiadis, G. M. T. Schreuder, G. G. Persijn, and F. J. Claas. HLAMatchmaker: A molecularly based algorithm for histocompatibility determination. III. effect of matching at the HLA–A,B amino acid triplet level o kidney transplant survival. *Transplantation*, 75(6):884–889, 2003.
- D. Gale and L. Shapley. College admission and stability of marriage. *American Mathematical Monthly*, 69:9–15, 1962.

- J. W. Hatfield. Pairwise kidney exchange: Comment. *Journal of Economic Theory*, 125: 189–193, 2005.
- I. Kaplan, J. A. Houp, M. S. Leffell, J. M. Hart, and A. A. Zachary. A computer match program for paired and unconventional kidney exchanges. *American Journal of Transplantation*, 5:2306–2308, 2005.
- K. Keizer, M. de Klerk, B. Haase-Kromwijk, and W. Weimar. The dutch algorithm for allocation in living donor kidney exchange. *Transplantation Proceedings*, 37:589–591, 2005.
- M. d. Klerk, K. Keizer, and W. Weimar. Donor exchange for renal transplantation. *The New England Journal of Medicine*, 351:935, 2004.
- L. W. Kranenburg, T. Visak, W. Weimar, and et al. Startling a crossover kidney transplantation program in the Netherlands: ethical and psychological considerations. *Transplantation*, 78:194, 2004.
- R. M. Merion, V. B. Ashby, R. A. Wolfe, D. A. Distant, T. E. Hulbert-Shearon, R. A. Metzger, A. O. Ojo, and F. K. Port. Deceased–donor characteristics and the survival benefit of kidney transplantation. *Journal of American Medical Association*, 294(21): 2726–2733, 2005.
- G. Opelz. Impact of HLA compatibility on survival of kidney transplants from unrelated live donors. *Transplantation*, 64:1473–1475, 1997.
- A. Roth. The economist as engineer: Game theory, experimental economics, and computation as tools of design economics. *Econometrica*, 70:1341–1378, 2002.
- A. E. Roth. Incentive compatibility in a market with indivisible goods. *Economics Letters*, 9:127–132, 1982.
- A. E. Roth and A. Postlewaite. Weak versus strong domination in a market of indivisible goods. *Journal of Mathematical Economics*, 4:131–137, 1977.
- A. E. Roth, T. Sönmez, and M. U. Ünver. Kidney exchange. *Quarterly Journal of Economics*, 119:457–488, 2004.

- A. E. Roth, T. Sönmez, and M. U. Ünver. Pairwise kidney exchange. *Journal of Economic Theory*, 125:151–188, 2005a.
- A. E. Roth, T. Sönmez, and M. U. Ünver. A kidney exchange clearinghouse in New England. *American Economic Review, Papers and Proceedings*, 95:376–380, 2005b.
- A. E. Roth, T. Sönmez, M. U. Ünver, F. L. Delmonico, and S. L. Saidman. Utilizing list exchange and nondirected donations through “chain” paired kidney donations. *American Journal of Transplantation*, 6:2694–2705, 2006.
- A. E. Roth, T. Sönmez, and M. U. Ünver. Efficient kidney exchange: Coincidence of wants in a markets with compatibility preferences. *American Economic Review*, 97–3: 828–851, 2007.
- S. L. Saidman, A. E. Roth, T. Sönmez, M. U. Ünver, and F. L. Delmonico. Increasing the opportunity of live kidney donation by matching for two- and three-way exchanges. *Transplantation*, 81:773–782, 2006.
- M. A. Schnitzler, C. S. Hollenbeak, D. S. Cohen, R. S. Woodward, J. A. Lowell, G. G. Singer, R. J. Tesi, T. K. Howard, T. Mohanakumar, and D. C. Brennan. The economic implications of HLA mathcing in cadaveric renal transplantation. *The New England Journal of Medicine*, 341:1440–1446, 1999.
- D. L. Segev, S. E. Gentry, J. K. Melancon, and R. A. Montgomery. Characterization of waiting times in a simulation of kidney paired donation. *American Journal of Transplantation*, 5:2448–2455, 2005a.
- D. L. Segev, S. E. Gentry, D. S. Warren, B. Rech, and R. A. Montgomery. Kidney paired donation and optimizing the use of living donor organs. *Journal of American Medical Association*, 293:1883–1890, April 2005b.
- A. K. Sen. Choice functions and revealed preference. *Review of Economic Studies*, 38: 307–317, 1971.
- L. Shapley and H. Scarf. On cores and indivisibility. *Journal of Mathematical Economics*, 1:23–37, 1974.
- T. Sönmez. Strategy-proofness and essentially single-valued cores. *Econometrica*, 67: 677–689, 1999.

- T. Sönmez and M. U. Ünver. Kidney exchange with good samaritan donors: A characterization, 2005. Unpublished Manuscript, Boston College and University of Pittsburgh.
- A. Spital. Donor exchange for renal transplantation. *The New England Journal of Medicine*, 351:936, 2004.
- X. Su and S. A. Zenios. Recipient choice can address the efficiency–equity trade–off in kidney transplantation: A mechanism design model. *forthcoming, Management Science*, 2005.
- X. Su, S. A. Zenios, and G. M. Chertow. Incorporating recipient choice in kidney transplantation. *Journal of the American Society of Nephrology*, 15:1656–1663, 2004.
- R. Wilson. Architecture of power markets. *Econometrica*, 70:1299–1340, 2002.
- S. A. Zenios, E. Woodle, and L. Ross. *Primum non nocere*: avoiding harm to vulnerable waitlist candidates in an indirect kidney exchange. *Transplantation*, 72:648–654, 2001.