Games and Economic Behavior 74 (2012) 299–310

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Transplant quality and patients' preferences in paired kidney exchange ☆

Antonio Nicoló [a], Carmelo Rodríguez-Álvarez [b,*]

[a] Dipartamento di Scienze Economiche "Marco Fanno", Università degli Studi di Padova, Via del Santo 33, 35123 Padova, Italy
[b] Facultad CC, Económicas y Empresariales, Campus de Somosaguas, Universidad Complutense de Madrid, 28223 Madrid, Spain

ARTICLE INFO

Article history:
Received 27 April 2010
Available online 27 May 2011

JEL classification:
C78
D02
D78
I10

Keywords:
Market design
Matching
Kidney exchange
Protective behavior

ABSTRACT

Paired Kidney Exchange (PKE) programs solve incompatibility problems of donor–patient pairs in living donor kidney transplantation by arranging exchanges of donors among several pairs. Further efficiency gains may emerge if the programs consider the quality of the matches between patients and donors. Limitations on the number of simultaneous required operations imply that every efficient PKE program introduces incentives for the patients to misreport how they rank the option of remaining in dialysis with respect to the available kidneys. Truthfully revealing their preferences is however, the unique protective (lexicographic maximin) strategy for patients under pairwise exchange maximizing PKE programs.

1. Introduction

The best treatment for end-stage renal disease is kidney transplantation. Kidneys available for transplantation may be obtained from deceased donors or from willing living donors. Unfortunately, a kidney of a living potential donor may be unsuitable for transplantation for a particular patient because the mismatch between donor and patient blood and tissues types would lead to the immediate rejection and loss of the graft.

The fact that dialysis treatment is very effective at preventing patients with end-stage renal disease from dying of their renal failure, and the possibility of living donation generate interesting new strategies to alleviate the (universal) shortage of kidneys. Two incompatible donor–patient pairs may be mutually compatible, and a swap of donors between the two pairs would result in two successful transplantations (*Paired Kidney Exchange*: PKE).[1] Further benefits can be obtained if kidney swaps involve three or more donor–patient pairs. South Korea, the Netherlands, the United Kingdom, and Spain

☆ We are grateful to Salvador Barberà, Bettina Klaus, Flip Klijn, Marco Mariotti, Jordi Massó, Alvin Roth, Jo Seldeslachts, the Associate Editor, and two anonymous referees for insightful comments and discussions. We thank seminar audiences at Alicante, Málaga, Vigo, Warwick, and York, as well as the participants of the 27th Arne Ryde Symposium 2009 at Lund, the ASSET Meeting 2008 at EUI, the BOMOPA Workshop at Bologna, the 14th Coalition Theory Network Workshop at Maastricht, and the 1st UECE Lisbon Meetings. We express our gratitude to Miguel Ángel Frutos and María Valentín for providing valuable information on the Spanish Paired Kidney Exchange Program. Previous versions of this work circulated under the title "Feasibility Constraints and Protective Behavior in Efficient Kidney Exchange". Nicoló acknowledges financial support from MIUR under grant 2007 EENEAX. Rodríguez-Álvarez gratefully acknowledges the financial support from the Ministerio de Educación y Ciencia through Programa Ramón y Cajal 2006, and Research Grant ECO2008-03674/ECON, the Fondo Social Europeo, the Consejería de Innovación Ciencia y Empresa (Junta de Andalucía) through Research Grant P09-SEJ-4941, and the Fundación Ramón Areces.
* Corresponding author. Fax: (+34) 91 394 2591.
E-mail addresses: antonio.nicolo@unipd.it (A. Nicoló), carmelor@ccee.ucm.es (C. Rodríguez-Álvarez).
[1] See Roth et al. (2004), Delmonico (2004), Delmonico et al. (2004), Segev et al. (2005a, 2005b), Spital (2004).

0899-8256/$ – see front matter
doi:10.1016/j.geb.2011.05.009

have developed national centralized PKE programs.[2] In the absence of a national coordinated program, in the United States different transplant centers have developed their own PKE programs.[3] Among the applied alternatives, it is worth noting the New England Paired Kidney Exchange Program (NEPKE) initiated in 2006. The NEPKE is based on the design of a kidney exchange clearinghouse proposed in Roth et al. (2005a, 2005b). According to the NEPKE approach, a Central Transplant Coordinator (CTC) uses kidney assignment rules that take into account the medical details of all patients and donors and proposes compatible exchanges among the pairs. A key issue when designing these rules is that they should not provide incentives for patients (or their doctors) to lie about medical details to improve their chances of getting a compatible kidney.

Many characteristics of the donor (age, health status) as well as matching characteristics of the donor–patient pair, (HLA antigen mismatches) determine the probability of long-term graft survival and the patient's life quality after the operation.[4] Up to now, the main objective of PKE programs has been to conduct as many transplants as possible. Match quality considerations have been subsidiary to that of finding compatible matches. However, it is generally accepted that better matches between donors and recipients would generate large efficiency gains via reductions of the graft rejection rates and to increases in life expectancy after transplantation. Moreover, efficiency gains could also be obtained by proposing low quality matches to patients who would prefer to receive a low quality kidney rather than staying on dialysis. Because Roth et al. (2005a, 2005b) assume that patients find all compatible kidneys as equally preferable, the NEPKE approach cannot address the issue of the quality of the match between donors and recipients.

The objective of this paper is to introduce the additional component of match quality to the mechanism design approach to PKE programs initiated by Roth et al. (2004, 2005a, 2005b). With this purpose in mind, we propose a PKE model that explicitly incorporates transplant quality decisions and patients' preferences. Specifically, as in the NEPKE, we assume that exchanges involving too many donor–patient pairs are unfeasible by logistic constraints, although exchanges are not initially restricted to be pairwise. In our model, patients' evaluation of available kidneys are based on measures of quality of the donor–patient match and on the induced graft survival probabilities. These features and probabilities are observable by CTC doctors and verifiable by means of clinical tests. In order to guarantee the participation of patients in the PKE program, CTC need to take into account the minimum probability of graft survival that each patient is willing to accept to undergo transplantation. That minimum probability depends on how the patient evaluates her quality of life undergoing dialysis, but also on her expectations about the quality of future pools of kidneys available for exchange, and her attitude toward risk and uncertainty. Thus, patients' preferences over available kidneys are known by the CTC, but how patients rank the option of remaining in dialysis is private information. With these assumptions, we show that under feasibility constraints on the number of simultaneous paired kidney exchanges truthful revelation of patients' preferences is not compatible with a weak version of efficiency.

In the light of the negative result that we obtain in the standard model, we propose an alternative approach that reconciles the notion that patients care basically about obtaining a compatible kidney with the evidence on the heterogeneity of compatible kidneys. By misreporting their preferences, for instance by raising the ranking of the dialysis option above some desirable kidneys, patients may end up receiving a better kidney, but they can also loose a beneficial transplant. Since PKE programs normally involve the coordination of nephrology services and patients at several hospitals, patients have little information about the remaining patients, and they can hardly predict the result of their strategic choices. In this scenario, patients are extremely concerned about the possibility of loosing a desirable kidney. Thus, we assume that patients choose their strategies so as to "protect" themselves from bad eventualities as much as possible. We capture this choice of "lexicographic maximin" strategies with the notion of "protective behavior" proposed by Barberà and Dutta (1982), Barberà and Jackson (1988) and later applied to college-admission problems by Barberà and Dutta (1995).[5] Under protective behavior, if kidney exchanges are restricted to involve only two donor–patient pairs, then a plethora of rules provide (strong) incentives for patients to report their true preferences. Among them, we single out the class of (ordinal) rules that satisfy Arrow's classical Axiom of Choice (Arrow, 1959). Interestingly, the national PKE programs implemented in Europe use kidney assignment rules that maximize the number of pairwise exchanges and satisfy the Axiom of Choice. Hence, introducing transplant quality and patients' preferences would not involve radical modifications of the already existent PKE protocols. This positive result vanishes if larger exchanges are admitted. Thus, our results highlight the limitations of multiple pair exchanges and provide strategic arguments beyond the logistical reasons for concentrating on pairwise exchanges.

At this point we compare our results with previous works in the kidney exchange literature. This article complements the literature initiated by Roth et al. (2004). In that paper, the authors assume that patients have heterogeneous preferences over compatible kidneys, but they do not take into account feasibility constraints in the number of operations required in the exchange. This issue explains the contrast between their possibility results and our initial impossibility results. Subsequent works, Roth et al. (2005a, 2005b), focus on a dichotomous domain of preferences where patients are indifferent among compatible kidneys. In Roth et al. (2007), these authors show that efficiency gains could be attained – and almost exhausted – if kidney exchanges among three donor–patient pairs were admitted. Hatfield (2005) shows that under arbitrary feasibility

[2] See Klerk et al. (2004, 2005), Park et al. (2004), http://www.organdonation.nhs.uk, and http://www.ont.es.

[3] The United Network for Transplant Sharing (UNOS) has recently started a National Kidney Paired Donation Pilot Program involving the cooperation of different regional networks. See http://optn.transplant.hrsa.gov/resources/kpdpp.asp and Ashlagi and Roth (2011).

[4] See Duquesnoy et al. (2003), Freeman (2007), Klerk et al. (2004, 2005), Kranenburg et al. (2004), Merion et al. (2005), Øien et al. (2007), Opelz (1997), Schnitzler et al. (1999), Su et al. (2004).

[5] See Klijn et al. (2010) for an experimental study of protective behavior in a school choice context.

constraints, for assignment rules that satisfy Arrow's Axiom of Choice revealing the true preferences and the complete set of potential donors is a dominant strategy for the patients. Recently, Ünver (2010) explicitly incorporates dynamic features and investigates the design of efficient exchange programs in a dynamic context. Sönmez and Ünver (2010) further investigate the introduction of compatible donor–patient pairs, and Yilmaz (2011) analyzes egalitarian issues in the NEPKE framework. The main difference with these papers relies on their assumption on the dichotomous domain. Since we explicitly introduce graft quality in the analysis, we operate on a richer family of assignment rules. In addition, we deal with an alternative informational setting where the only private information is patients' ranking of the dialysis option.

As regards as the literature on kidney transplantation, Su and Zenios (2005, 2006) study the problem of patient choice in the standard donation model when suboptimal kidneys are available for transplantation. Su and Zenios (2005) use simulations based on US data to estimate the effects of patient choice and find that patient choice to withdraw a suboptimal graft may reduce between 17 percent and 30 percent the number of transplantations. Su and Zenios (2006) propose a second best mechanism design approach to allocate deceased-donor kidneys. In their model patients face a trade-off between quality of the graft and time in the waiting list. These authors suggest the endogenous construction of different waiting lists for kidneys depending on their quality. In equilibrium, patients that are not willing to accept low quality kidneys stay in the waiting list for longer than less demanding patients. Both papers share our assumptions regarding patients' preferences and the fact that patients are endowed with private information regarding their reservation values. Our focus on static PKE complements the authors' analysis of the dynamic strategic aspects of patient choice in deceased-donor waiting lists.

The remainder of the paper is organized as follows. In Section 2, we outline the model of kidney allocation problems and basic notation. In Section 3, we introduce the concept of kidney assignment rules and some desirable conditions. In Section 4, we present an introductory impossibility result. In Section 5, we define the protective behavior and present the positive results. In Section 6, we conclude.

2. Kidney assignment problems

Consider a finite society consisting of a set $N = \{1, \ldots, n\}$ of patients ($n \geqslant 3$) who need a kidney for transplantation. Each patient $i \in N$ has only one potential donor and all kidneys come from living donors. Thus, $\Omega = \{\omega_1, \ldots, \omega_n\}$ denotes the set of kidneys available for transplantation, where ω_i refers to the kidney of patient i's donor.[6] To describe the situation of a patient who remains in dialysis, we denote by ω_0 the null kidney. We say that a patient who remains in dialysis is matched with ω_0.

The probability of graft survival when kidney $\omega \in \Omega$ is transplanted to patient i depends on objective medical criteria. Thus, the fitness of each available kidney to each patient is directly observable by the CTC. We denote by $p_i(\omega) \in [0, 1)$ the probability of graft survival when kidney ω is transplanted to patient i. We say that patient i and kidney ω are *incompatible*, if $p_i(\omega) = 0$. This reflects the possibility that patient i's body will reject the graft of kidney ω, because of blood-type or tissues incompatibility, or any other reason. We say that patient i and kidney ω are *compatible* if $p_i(\omega) > 0$. The *probability of graft survival matrix* $\mathbf{P} \in \mathbb{M}_{n\times n}$ is defined by $\mathbf{P}_{i,j} \equiv p_j(\omega_i)$ for each $i, j \in N$. The matrix $\mathbf{P}$ contains all the relevant medical information about patients and donors that is observable by the CTC. Without any loss of generality, for all $\omega, \omega' \in \Omega$ if $p_i(\omega) > 0$ then $p_i(\omega) \neq p_i(\omega')$. Moreover, in order to simplify notation, we assume that for each patient i, $p_i(\omega_i) = 0$.

Each patient i is equipped with a preference – a complete, reflexive, and transitive binary relation – $\succsim_i$ on $\Omega \cup \{\omega_0\}$. We denote by $\succ_i$ the associated strict preference relation and by $\sim_i$ the associated indifference relation. We refer to an n-tuples of preferences $\succsim = (\succsim_1, \ldots, \succsim_n)$ as a preference profile. For each patient i and each preference profile $\succsim$, $\succsim_{-i}$ denotes the complementary profile of preferences of all patients in $N \setminus \{i\}$.

Patients' preferences over kidneys are based on the information contained in the probability of graft survival matrix $\mathbf{P}$. Specifically, for each patient i, the preference $\succsim_i$ is consistent with $\mathbf{P}$ if

(i) for each pair $\omega, \omega' \in \Omega$, $p_i(\omega) \geqslant p_i(\omega') \Leftrightarrow \omega \succsim_i \omega'$,
(ii) for each ω such that $p_i(\omega) = 0$, $\omega_0 \succ_i \omega$,
(iii) there is no $\omega \in \Omega$ such that $\omega \sim_i \omega_0$.

By (i), when comparing two kidneys, every patient prefers the kidney that yields the highest probability of graft survival. By (ii), every patient prefers to remain in dialysis rather than receiving an incompatible kidney. Finally, by (iii), patients are never indifferent between receiving a specific kidney and remaining in dialysis. Given a probability of graft survival matrix $\mathbf{P}$, $\mathcal{R}_i^{\mathbf{P}}$ is the set of preferences which are consistent with $\mathbf{P}$ for patient i. Let $\mathcal{R}^{\mathbf{P}} \equiv \times_{i\in N} \mathcal{R}_i^{\mathbf{P}}$ and slightly abusing notation let $\mathcal{R}_{-i}^{\mathbf{P}} \equiv \times_{j\neq i} \mathcal{R}_j^{\mathbf{P}}$.

Note that patients' preferences are not completely determined by $\mathbf{P}$. Specifically, $\mathbf{P}$ does not determine the rank of ω_0 in each patient's preference. Hence, the minimum probability of graft survival such that a patient prefers to undergo a transplant rather than remaining in dialysis is private information of the patient.

A (*kidney exchange*) *problem* is a pair $(\mathbf{P}, \succsim)$ such that $\succsim \in \mathcal{R}^{\mathbf{P}}$.

---

[6] We briefly discuss the possibility of multiple donors in Section 5.

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

(i) for each $i, j \in N$, $i \neq j$ and each $\omega, \omega' \in \Omega$, 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 assigned to another patient, then the initial patient is not assigned the null kidney. For each patient i and each assignment a, we denote by a_i the kidney that patient i receives at a.

In every assignment, kidneys are allocated by forming exchange cycles of patient–donors pairs. In each exchange 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 pairs, all the kidneys must be reaped from the donors and transplanted to the patients simultaneously. Since hospitals face evident logistic restrictions and coordination problems, 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 $a_{j'} = \omega_i$,[7] or
(ii) $a_i = \omega_0$.

Clearly, for each assignment a, the partition π_a is unique and well defined. We define the *cardinality of a* as the $max_{p \in \pi_a} \#p$.[8]

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

For each $k \in \mathbb{N}$, $k \leqslant n$, 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 pairs are admitted. The assignment a is a *pairwise assignment* if $a \in \mathcal{A}^2$.

3. Kidney assignment rules

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

Every matrix $\mathbf{P}$ together with a rule $\varphi(\cdot)$ define a revelation mechanism. The revelation mechanism specifies a set of players (the patients), a set of strategies for each patient, the sets $\mathcal{R}_i^{\mathbf{P}}$, and an outcome function, $\varphi(\mathbf{P}, \cdot)$. Note that the mechanism $(\mathbf{P}, \varphi)$ falls short of defining a game in strategic form because $\mathbf{P}$ does not introduce all the information about patients' preferences.

Next, we present a formal definition of the standard conditions for desirable rules. The reader should keep in mind that all the conditions refer to a given observed matrix $\mathbf{P}$.

Individual rationality. For each $i \in N$ and each $\succsim \in \mathcal{R}^{\mathbf{P}}$, $\varphi_i(\succsim) \succsim_i \omega_0$.

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

Individual rationality takes into account a patient's right to refuse any transplant and remain in dialysis. Clearly, *k-efficiency* is the natural version of efficiency that considers 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 no 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 CTC may have all the information about the degree of compatibility and fitness among patients and available donors, there is a key piece of information that remains private for patients and must be elicited for public use, the minimum graft survival probability they require to undergo transplantation. The objective of

[7] Note that $j = j'$ and $i = j = j'$ are allowed.
[8] For each set S, $\#S$ refers to the number of elements of S.

this section is to analyze whether we can construct rules that provide incentives to patients to reveal their true preferences in the presence of feasibility constraints on the cardinality of the proposed assignments.

Strategy-proofness. For each $i \in N$, each $\succsim \in \mathcal{R}^{\mathbf{P}}$, and each $\succsim'_i \in \mathcal{R}^{\mathbf{P}}_i$,

$$\varphi_i(\succsim) \succsim_i \varphi_i(\succsim'_i, \succsim_{-i}).$$

Strategy-proofness implies that reporting the true preferences is a (weakly) dominant strategy for every patient in all the games compatible with the revelation mechanism $\mathbf{P}$ and φ. Note that *strategy-proofness* is weak in our framework because only the rank of ω_0 is private information.

The literature on the allocation of indivisible objects has extensively studied the problem of designing assignment rules that satisfy *individual rationality* and *strategy-proofness*.[9] When patients have strict preferences, there is a natural way to assign the available kidneys among the patients based on Gale's top trading cycle procedure. Given a problem, let every patient point to her preferred donor. A top trading cycle consists of a cycle 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 remaining in dialysis is her best available option.) Give each patient in a top trading cycle her preferred kidney, and remove them from the problem with her assigned kidney. Repeat the process until each patient receives a kidney. If preferences regarding compatible kidneys are strict and there is finite number of patients and kidneys, then the resulting assignment is unique. A top trading cycle may involve all patients. Hence, the induced rule satisfies *individual rationality*, *n-efficiency*, and *strategy-proofness* and but it violates *k-efficiency* for $k < n$.[10]

Our first result shows that feasibility constraints make it impossible to construct *efficient* rules that provide the right incentives to patients at every preference profile.

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

Proof. We study two cases. We first analyze the restriction to pairwise exchanges. Then, we provide the proof for $k \geqslant 3$. 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 multi-valued ($k \geqslant 3$).[11]

Assume, by way of contradiction, that 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 probability of graft survival matrix $\mathbf{P}$ such that its restriction to these patients and their donors' kidneys is:

$$\mathbf{P} = \begin{pmatrix} 0 & 0.5 & 0.9 \\ 0.9 & 0 & 0.5 \\ 0.5 & 0.9 & 0 \end{pmatrix},$$

and so that for each $i \in \{1, 2, 3\}$, and each $\omega \notin \{\omega_0, \omega_1, \omega_2, \omega_3\}$, $p_i(\omega) = 0$. Thus, for each $\succsim \in \mathcal{R}^{\mathbf{P}}$, $\omega_2 \succ_1 \omega_3 \succ_1 \omega_1$; $\omega_3 \succ_2 \omega_1 \succ_2 \omega_2$; and $\omega_1 \succ_3 \omega_2 \succ_3 \omega_3$. In order to simplify notation, let $N = \{1, 2, 3\}$. (By *individual rationality*, this is without loss of generality.)

Let $\succsim \in \mathcal{R}^{\mathbf{P}}$ be such that

$\succsim_1$	$\succsim_2$	$\succsim_3$
ω_2	ω_3	ω_1
ω_3	ω_1	ω_2
ω_0	ω_0	ω_0
...	...	...

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 $\succsim' \in \mathcal{R}^{\mathbf{P}}$ be such that $\omega_2 \succ'_1 \omega_0 \succ'_1 \omega_3$ and $\succsim'_{-1} = \succsim_{-1}$. By *strategy-proofness*, $\varphi_1(\succsim') \succsim'_1 \varphi_1(\succsim) = \omega_2$. Hence, $\varphi_1(\succsim') = \omega_2$. Finally, let $\succsim'' \in \mathcal{R}^{\mathbf{P}}$ be such that $\omega_3 \succ''_2 \omega_0 \succ''_2 \omega_1$ and $\succsim''_{-2} = \succsim'_{-2}$. By *individual rationality*, $\varphi_2(\succsim'') \in \{\omega_0, \omega_3\}$. By *strategy-proofness*, $\varphi_2(\succsim') = \omega_1 \succsim'_2 \varphi_2(\succsim'')$. Hence, $\varphi_2(\succsim'') = \omega_0$. Because, by *individual rationality*, $\varphi_1(\succsim'') \in \{\omega_0, \omega_2\}$, $\varphi_1(\succsim'') = \omega_0$. Therefore, $\varphi(\succsim'') = [(1, \omega_0), (2, \omega_0), (3, \omega_0)]$. Note that the assignment $a = [(1, \omega_0), (2, \omega_3), (3, \omega_2)]$ is 2-*feasible*, and $a_i \succsim''_i \varphi_i(\succsim'')$ for each $i \in N$ and $a_2 \succ''_2 \varphi_2(\succsim'')$. Then, φ violates 2-*efficiency*.

[9] The allocation of indivisible objects without transfers is known as the housing market (Gale and Shapley, 1962; Shapley and Scarf, 1974).

[10] See Roth and Postlewaite (1977), Roth (1982). Ma (1994) characterizes the rule that selects the (single-valued) core correspondence obtained by the top trading cycle as the unique rule that satisfies *strategy-proofness*, *individual rationality*, and *n-efficiency* in this setting.

[11] See Sönmez (1999).

Next, we analyze the general case. Let $k \geqslant 3$. Remember that $k < n$ and then there are at least $k+1$ patients. Assume, by way of contradiction, that there is a rule φ that satisfies *individual rationality*, *k-efficiency*, and *strategy-proofness* for every $\mathbf{P}$. Let the matrix $\mathbf{P}$ be such that for every $i = 1, \ldots, k+1$:

$$p_i(\omega_{i+1}) > p_i(\omega_{i+2}) > p_i(\omega), \quad \forall \omega \in \Omega \setminus \{\omega_{i+1}, \omega_{i+2}\} \text{ (modulo } k+1).$$

In order to simplify notation (by *individual rationality*, without loss of generality), let $N = \{1, \ldots, k+1\}$.

Let $\succsim \in \mathcal{R}^{\mathbf{P}}$ be such that:

$\succsim_1$	$\succsim_2$	$\cdots$	$\succsim_{k-1}$	$\succsim_k$	$\succsim_{k+1}$
ω_2	ω_3	$\ldots$	ω_k	ω_{k+1}	ω_1
ω_0	ω_0	$\ldots$	ω_0	ω_0	ω_2
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	ω_0
					$\ldots$

By *individual rationality*, either no object is assigned to any patient $1, \ldots, k+1$, or patient $k+1$ receives ω_2, patient 1 receives the null kidney, and every other patient i receives ω_{i+1} (the kidney of her next to the right neighbor). By *k-efficiency*:

$$\varphi(\succsim) = \begin{bmatrix} (1, \omega_0) & \\ (i, \omega_{i+1}) & \forall i = 2, \ldots, k \\ (k+1, \omega_2) & \end{bmatrix}.$$

Let $\succsim' \in \mathcal{R}^{\mathbf{P}}$ be such that for each $i \neq k-1$, $\succsim'_i = \succsim_i$, and

$\succsim'_1$	$\succsim'_2$	$\cdots$	$\succsim'_{k-1}$	$\succsim'_k$	$\succsim'_{k+1}$
ω_2	ω_3	$\ldots$	ω_k	ω_{k+1}	ω_1
ω_0	ω_0	$\ldots$	ω_{k+1}	ω_0	ω_2
$\ldots$	$\ldots$	$\ldots$	ω_0	$\ldots$	ω_0
			$\ldots$		$\ldots$

By *strategy-proofness*, $\varphi_{k-1}(\succsim') \succsim'_{k-1} \varphi_{k-1}(\succsim) = \omega_k$. Note that ω_k is patient $k-1$'s preferred kidney. Then, $\varphi_{k-1}(\succsim') = \omega_k$. By *k-efficiency* and *individual rationality*, $\varphi(\succsim') = \varphi(\succsim)$.

Let $\bar{\succsim} \in \mathcal{R}^{\mathbf{P}}$ be such that for each $i \neq k+1$, $\bar{\succsim}_i = \succsim'_i$, and

$\bar{\succsim}_1$	$\bar{\succsim}_2$	$\cdots$	$\bar{\succsim}_{k-1}$	$\bar{\succsim}_k$	$\bar{\succsim}_{k+1}$
ω_2	ω_3	$\ldots$	ω_k	ω_{k+1}	ω_1
ω_0	ω_0	$\ldots$	ω_{k+1}	ω_0	ω_0
$\ldots$	$\ldots$	$\ldots$	ω_0	$\ldots$	$\ldots$
			$\ldots$		

The same arguments we employed to determine $\varphi(\succsim)$ apply here to obtain:

$$\varphi(\bar{\succsim}) = \begin{bmatrix} (i, \omega_{i+1}) & \forall i < k-1 \\ (k-1, \omega_{k+1}) & \\ (k, \omega_0) & \\ (k+1, \omega_1) & \end{bmatrix}.$$

Note that $\omega_1 = \varphi_{k+1}(\bar{\succsim}) = \varphi(\bar{\succsim}_{k+1}, \succsim'_{-(k+1)}) \succ'_{k+1} \varphi_{k+1}(\succsim') = \omega_2$, which contradicts *strategy-proofness*. □

The previous impossibility result is robust to the introduction of weak preferences regarding kidneys. All we require is to admit the existence of two indifference classes for acceptable kidneys. Namely, for each patient i and each pair $\omega, \omega' \in \Omega$ there is a preference $\succsim_i$ such that $\omega \succ_i \omega' \succ_i \omega_0$. Hence, Theorem 1 contrasts with the positive results in the dichotomous domains of preferences by Roth et al. (2005a) and Hatfield (2005). Moreover, the result can be extended to settings where patients may have incomplete information (i.e., beliefs) about the preferences of the remaining patients. In a result that parallels the results of Roth (1989), we can prove that there are probability of graft survival matrix and sets of patients' beliefs about other patients' preferences such that there is no rule that satisfies *individual rationality*, *k-efficiency*, and (Bayesian) *incentive compatibility*.[12]

[12] A precise statement of this result can be found in the working paper version of the manuscript. In a recent paper, Villa and Patrone (2009) prove that the rule that maximizes the sum of the welfare of the patients is not incentive compatible.

5. Protective behavior in kidney exchange problems

The negative result presented in the previous section is particularly discouraging since we balance the enrichment of the preference domain with the increase in the information available to the CTC. Namely, patients' preferences regarding available kidneys are known by the CTC because they depend on measurable and verifiable features. In order to avoid the negative implications of Theorem 1, we investigate patients' incentives under a natural assumption on patients' behavior. PKE programs arrange exchanges a few times a year – ranging from every four weeks to every three months – potentially involving pools of hundreds of patients. In this scenario, patients are well aware of the vital decision they have to take, but hardly have any information about the preferences or the identity of the other participants in the program. The combination of these facts can certainly induce patients to be extremely concerned about the possibility of forfeiting available acceptable organs. Thus, patients waiting for a transplant may be strongly risk-averse, and they might prefer to choose their strategies so as to "protect" themselves from the worst eventuality as far as possible.[13] This "maximin" assumption is captured by the notion of "protective behavior" proposed by Barberà and Dutta (1982, 1995).

Consider the direct revelation mechanism defined by $(\mathbf{P}, \varphi)$.[14] For each patient i, each preference $\succsim_i \in \mathcal{R}_i^{\mathbf{P}}$, and each $\omega \in \Omega \cup \{\omega_0\}$, let:

$$c(\omega, \succsim_i) = \bigl\{\succsim'_{-i} \in \mathcal{R}_{-i}^{\mathbf{P}} \mid \varphi\bigl(\succsim_i, \succsim'_{-i}\bigr) \sim_i \omega\bigr\}.$$

Then, $c(\omega, \succsim_i)$ is the set of complementary profiles of preferences reported by patients $j \neq i$ under which patient i receives a kidney which is for her as good as kidney ω, when patient i reports preferences $\succsim_i$.

Protective domination. Given $\mathbf{P}$ and φ, for each patient i, each $\succsim_i \in \mathcal{R}_i^{\mathbf{P}}$, and each pair of strategies $\succsim'_i, \succsim''_i \in \mathcal{R}_i^{\mathbf{P}}$, $\succsim'_i$ *protectively dominates* $\succsim''_i$, denoted $\succsim'_i \mathrm{d}(\succsim_i) \succsim''_i$ if there exists $\omega \in \Omega \cup \{\omega_0\}$ such that:

(i) $c(\omega', \succsim'_i) \cap c(\omega'', \succsim''_i) = \emptyset$ for each $\omega', \omega'' \in \Omega \cup \{\omega_0\}$ with $\omega \succsim_i \omega'$ and $\omega'' \succ_i \omega'$.
(ii) $c(\omega, \succsim'_i) \subsetneq c(\omega, \succsim''_i)$.

For each patient i $\mathrm{D}(\succsim_i) \equiv \{\succsim'_i \in \mathcal{R}_i^{\mathbf{P}} \mid \text{there is no } \succsim'' \in \mathcal{R}_i^{\mathbf{P}} \text{ such that } \succsim'' \mathrm{d}(\succsim_i) \succsim'\}$ be the set of *protective strategies* of patient i.

Consider two strategies $\succsim'_i, \succsim''_i$ such that $\succsim' \mathrm{d}(\succsim) \succsim''$. Condition (i) implies that there is a kidney ω, such that patient i cannot improve by reporting $\succsim''$ instead of $\succsim'$ for all the complementary preference profiles in which the outcome she obtains reporting $\succsim'$ is not better than ω. Condition (ii) implies that there are complementary preference profiles for which $\succsim''_i$ yields ω (or something indifferent to ω) while that $\succsim'_i$ yields an outcome better than ω. Hence, if patient i wants to avoid "disasters", she prefers to report preference $\succsim'_i$ instead of $\succsim''_i$.

The protective domination relation is not complete, but it is transitive.[15] Thus, for each preference $\succsim_i \in \mathcal{R}_i^{\mathbf{P}}$, the set $\mathrm{D}(\succsim_i)$ is not empty. Moreover, because $\mathcal{R}_i^{\mathbf{P}}$ is finite, if there is a unique protective strategy, $\{\succsim'_i\} = \mathrm{D}(\succsim_i)$, then $\succsim'_i \mathrm{d}(\succsim_{-i}) \succsim''_i$ for each $\succsim''_i \in \mathcal{R}_i^{\mathbf{P}} \setminus \{\succsim'_i\}$.

We are interested in rules that provide incentives to the patients to reveal their true preferences. In this protective scenario, requiring to report the true ranking of the null kidney to be a protective strategy would not suffice to induce truthful revelation. Notice that, if there are several different protective strategies in addition to reporting the true preference, then it is not clear that truth-telling is an optimal strategy for the patients. This fact calls for a stronger requirement.

For each patient i, strategies $\succsim_i, \succsim'_i \in \mathcal{R}_i^{\mathbf{P}}$ are *equivalent* if for each $\succsim_{-i} \in \mathcal{R}_{-i}^{\mathbf{P}}$, $\varphi_i(\succsim_i, \succsim_{-i}) \sim_i \varphi_i(\succsim'_i, \succsim_{-i})$.

Truth-telling is the unique protective strategy for patient i if for each $\succsim_i \in \mathcal{R}_i^{\mathbf{P}}$,

$$\mathrm{D}(\succsim_i) = \bigl\{\succsim'_i \mid \succsim_i \text{ and } \succsim'_i \text{ are equivalent}\bigr\}.$$

Beyond its intuitive appeal, there is implementation justification for the notion of truth-telling as a unique protective strategy. In finite environments, Barberà and Dutta (1982, Theorem 1) show that the only rules such that outcome of the rule coincides with the result of the direct revelation game when all the patients choose a protective strategy, are those that guarantee that truth-telling is the unique protective strategy.

---

[13] According to anecdotal evidence from Italian and Spanish standard cadaveric donor kidney transplantation programs, patients' choices are consistent with extreme risk-aversion. For instance, the North Italian Transplant program (NITp) contemplates the possibility of using organs with higher risk of infectious diseases. (See http://www.nitp.org.) In the 2003–2008 period, the NITp has transplanted 970 kidneys of this type and only two patients had complications (which have been successfully treated in both cases). Nonetheless, most patients are very reluctant to enroll in the waiting list for this type of organs, and only after waiting various years for a standard donor they accept to join the program.

[14] Throughout this section, we assume that all the information available to the CTC, namely $\mathbf{P}$ and φ, is also available to the patients. The results hold if we assume that patients have information only about the set of mutually compatible exchanges and not the whole matrix $\mathbf{P}$. We delay the discussion on the incomplete information case to the end of this section.

[15] See Barberà and Jackson (1988).

For each patient i, let $\Omega_i(\mathbf{P},\varphi) = \{\omega \in \Omega \mid \text{for some} \succsim \in \mathcal{R}^{\mathbf{P}}, \varphi_i(\succsim) = \omega\}$ For each patient i and each preference $\succsim_i \in \mathcal{R}_i^{\mathbf{P}}$, we define i's *set of desirable kidneys* $\Omega^+(\succsim_i) \equiv \{\omega \in \Omega_i(\mathbf{P},\varphi) \mid \omega \succ_i \omega_0\}$. Hence, i's set of desirable kidneys includes all the kidneys that she can receive in a exchange in the range of the rule φ and that improve upon remaining in dialysis. Note that for each pair $\succsim_i, \succsim_i' \in \mathcal{R}_i^{\mathbf{P}}$, since the only difference between $\succsim_i$ and $\succsim_i'$ is the ranking of ω_0, either $\Omega^+(\succsim_i) \subseteq \Omega^+(\succsim_i')$ or $\Omega^+(\succsim_i') \subseteq \Omega^+(\succsim_i)$.

At this point, we introduce two conditions that turn out to be necessary for truth-telling to be a unique protective strategy.

Invariance. For each $i \in N$, and each pair of preference profiles $\succsim, \succsim' \in \mathcal{R}^{\mathbf{P}}$ with $\succsim_{-i} = \succsim_{-i}'$, $\Omega^+(\succsim_i) = \Omega^+(\succsim_i')$ implies $\varphi_i(\succsim) = \varphi_i(\succsim')$.

Weak consistency. For each $i \in N$, each $\succsim \in \mathcal{R}^{\mathbf{P}}$, and each $\succsim_i \in \mathcal{R}_i^{\mathbf{P}}$, if $\varphi_i(\succsim) = \omega_0$ and $\Omega^+(\succsim_i) \subset \Omega^+(\succsim_i')$ then $\varphi_i(\succsim_i', \succsim_{-i}) = \omega_0$.

Invariance requires that a patient is unaffected by a change in her preferences that only involves preference reversals with respect to assignments that are not in the range of the rule. If the set of desirable kidneys do not change, then both preferences are equivalent. *Weak consistency* is a convenient weakening of the Axiom of Choice for single-valued choice functions.[16] If a patient remains in dialysis when she reports $\succsim_i$, then she cannot receive a compatible kidney when she announces a preference such that the set of desirable kidneys shrinks. Note that if a rule satisfies *individual rationality*, then *weak consistency* applies the logic behind the Axiom of Choice only in situations in which each patient receives the worst possible outcome, ω_0.

Proposition 1. *For each probability of graft survival matrix* $\mathbf{P}$ *and each rule* φ*, if* φ *satisfies individual rationality and for each patient i truth-telling is the unique protective strategy, then* φ *satisfies invariance and weak consistency.*

Proof. We start with the proof of *invariance*. Consider a patient i, and preferences $\succsim_i, \succsim_i' \in \mathcal{R}_i^{\mathbf{P}}$ such that $\succsim_i \neq \succsim_i'$ and $\Omega^+(\succsim_i) = \Omega^+(\succsim_i')$. We prove that $\succsim_i$ and $\succsim_i'$ are equivalent strategies. Note that by *individual rationality*, for each $\omega \notin \Omega^+(\succsim_i)$, $c(\omega, \succsim_i) = c(\omega, \succsim_i') = \emptyset$. Let $\bar{\succsim}_{-i} \in c(\omega_0, \succsim)$. Assume to the contrary that $\bar{\succsim}_{-i} \notin c(\omega_0, \succsim')$. Then, $\succsim_i$ and $\succsim_i'$ are not equivalent. By *individual rationality*, $\varphi_i(\succsim_i', \bar{\succsim}_{-i}) \succ_i \omega_0$. By (i) of the definition of protective domination, $\succsim_i$ does not protectively dominates $\succsim_i'$ when i's true preference is $\succsim_i$, which contradicts truth-telling being the unique protective strategy. Thus, $c(\omega_0, \succsim_i) \subseteq c(\omega_0, \succsim_i')$. Since $\Omega^+(\succsim_i) = \Omega^+(\succsim_i')$, using the same reasoning, we obtain $c(\omega_0, \succsim_i') \subseteq c(\omega_0, \succsim_i)$. Hence, $c(\omega_0, \succsim_i) = c(\omega_0, \succsim_i')$. Next, let $\bar{\omega} \in \Omega^+(\succsim_i)$ be such that there is no $\omega' \in \Omega_i(\mathbf{P},\varphi)$ that satisfies $\bar{\omega} \succ_i \omega' \succ_i \omega_0$. We can repeat the arguments above to show that $c(\bar{\omega}, \succsim_i) = c(\bar{\omega}, \succsim_i')$, and iteratively to the remaining $\omega \in \Omega^+(\succsim_i)$. Hence, $\succsim_i$ and $\succsim_i'$ are equivalent and for each $\succsim_{-i} \in \mathcal{R}_{-i}^{\mathbf{P}}$, $\varphi_i(\succsim_i, \succsim_{-i}) = \varphi_i(\succsim_i', \succsim_{-i})$.

We conclude with the proof of *weak consistency*. For each patient i and each $\succsim_i \in \mathcal{R}_i^{\mathbf{P}}$, by *individual rationality*, for each $\omega \in \Omega$ such that $\omega_0 \succ_i \omega$, $c(\omega, \succsim_i) = \emptyset$. Moreover, there exists $\bar{\succsim}_{-i} \in \mathcal{R}_{-i}^{\mathbf{P}}$ such that for each $j \neq i$, $\Omega^+(\bar{\succsim}_j) = \emptyset$. Then, by *individual rationality*, for each $\succsim_i \in \mathcal{R}_i^{\mathbf{P}}$, $\varphi_i(\succsim_i, \bar{\succsim}_{-i}) = \omega_0$, and $c(\omega_0, \succsim_i) \neq \emptyset$. Assume to the contrary that φ does not satisfy *weak consistency*, then there are a patient i and a preference $\succsim_i' \in \mathcal{R}_i^{\mathbf{P}}$ such that $\Omega^+(\succsim_i) \subset \Omega^+(\succsim_i')$, $\varphi_i(\succsim) = \omega_0$, and $\varphi_i(\succsim_i', \succsim_{-i}) = \omega \neq \omega_0$. Note that, by *individual rationality*, $\omega \succ_i \omega_0$. Hence, by (i) of the definition of protective domination, $\succsim_i$ does not protectively dominates $\succsim_i'$. Moreover, $\succsim_i$ and $\succsim_i'$ are not equivalent strategies. These facts contradict that truth-telling is the unique protective strategy. □

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

Proposition 2. *For each probability of graft survival matrix* $\mathbf{P}$*, if the rule* φ *satisfies individual rationality,* 2*-efficiency, invariance, and weak consistency, then for each patient i truth-telling is the unique protective strategy.*

Proof. Fix a patient i and a preference $\succsim_i \in \mathcal{R}_i^{\mathbf{P}}$. By *invariance*, all preferences $\succsim_i' \in \mathcal{R}_i^{\mathbf{P}}$ such that $\Omega^+(\succsim_i') = \Omega^+(\succsim_i)$ are equivalent to $\succsim_i$. Thus, we have to prove that for each $\succsim_i' \in \mathcal{R}_i^{\mathbf{P}}$ with $\Omega^+(\succsim_i') \neq \Omega^+(\succsim_i)$, $\succsim_i \mathrm{d}(\succsim_i) \succsim_i'$. Let $\succsim_i' \in \mathcal{R}_i^{\mathbf{P}}$ be such that $\Omega^+(\succsim_i') \setminus \Omega^+(\succsim_i) \neq \emptyset$. By *individual rationality*, for all $\omega \in \Omega$ with $\omega_0 \succ_i' \omega$, $c(\omega, \succsim_i) = c(\omega, \succsim_i') = \emptyset$. Let $j \in N$ such that $\omega_j \in \Omega^+(\succsim_i') \setminus \Omega^+(\succsim_i)$, and let $\succsim_{-i} \in \mathcal{R}_{-i}^{\mathbf{P}}$ be such that $\omega_i \in \Omega^+(\succsim_j)$ and for each $k \notin \{i, j\}$, $\Omega^+(\succsim_k) = \emptyset$. By *individual rationality* and 2-*efficiency*, $\varphi_i(\succsim_i, \succsim_{-i}) = \omega_0$ and $\varphi_i(\succsim_i', \succsim_{-i}) = \omega_j$. Note that $\omega_0 \succ_i \omega_j$, and by *individual rationality*, $c(\omega_j, \succsim_i) = \emptyset$. Then, $c(\omega', \succsim_i) \cap c(\omega'', \succsim_i') = \emptyset$ for each $\omega', \omega'' \in \Omega \cup \{\omega_0\}$ such that $\omega_j \succsim_i \omega'$ and $\omega'' \succ_i \omega_j$ and $c(\omega_j, \succsim_i) \subsetneq c(\omega_j, \succsim_i')$, which suffices to prove that $\succsim_i \mathrm{d}(\succsim_i) \succsim_i'$. Finally, let $\succsim_i'' \in \mathcal{R}_i^{\mathbf{P}}$ be such that $\Omega^+(\succsim_i) \setminus \Omega^+(\succsim_i'') \neq \emptyset$. By

[16] See Arrow (1959) and Sen (1971). We follow Hatfield (2005) in the terminology.

individual rationality, for each $\bar{\omega} \in \Omega$ with $\omega_0 \succ_i \bar{\omega}$, $c(\bar{\omega}, \succsim_i) = c(\bar{\omega}, \succsim_i'') = \emptyset$. By *weak consistency*, $c(\omega_0, \succsim_i) \subseteq c(\omega_0, \succsim_i'')$. Let $\omega_j \in \Omega^+(\succsim_i) \setminus \Omega^+(\succsim_i'')$. Notice that $\omega_j \succ_i \omega_0$. Let $\succsim_{-i}^* \in \mathcal{R}_{-i}^{\mathbf{P}}$ such that for each $k \neq i, j$, $\Omega^+(\succsim_k^*) = \emptyset$, and $\omega_i \in \Omega^+(\succsim_j^*)$. By 2-*efficiency* $\varphi_i(\succsim_i, \succsim_{-i}^*) = \omega_j$. By *individual rationality*, $\varphi_i(\succsim_i'', \succsim_{-i}^*) = \omega_0$. Therefore, $c(\omega', \succsim_i) \cap c(\omega'', \succsim_i') = \emptyset$ for each $\omega', \omega'' \in \Omega \cup \{\omega_0\}$ such that $\omega_0 \succsim_i \omega'$, and $\omega'' \succ_i \omega_0$, and $c(\omega_0, \succsim_i) \subsetneq c(\omega_0, \succsim_i')$, and $\succsim_i \mathrm{d}(\succsim_{-i}) \succsim_i''$. $\square$

The results in Propositions 1–2 are clearly positive. Since *invariance* and *weak consistency* are fairly mild conditions, there is a large class of rules that provide incentives for patients to reveal their true preferences under protective behavior. Moreover, these results are in line with the results in Roth et al. (2005a) and Hatfield (2005) although we start from different assumptions regarding patients' preferences and strategic behavior. In order to illustrate the richness of this class of rules, we introduce additional notation.

For each $\succsim \in \mathcal{R}^{\mathbf{P}}$, let

$$\mathcal{IR}(\succsim) \equiv \{a \in \mathcal{A} \mid \text{for all } i \in N,\ a_i \succsim_i \omega_0\}$$

and

$$\mathcal{E}^2(\succsim) \equiv \left\{a \in \mathcal{A}^2 \;\middle|\; \text{there is no } a' \in \mathcal{A}^2;\ \text{for all } i \in N,\ a_i' \succsim_i a_i \text{ and } \exists j \in N, a_j' \succ_j a\right\}.$$

The set of efficient and individually rational pairwise assignments is immediately defined as $\mathcal{ER}^2(\succsim) \equiv \mathcal{IR}(\succsim) \cap \mathcal{E}^2(\succsim)$.

A rule φ is *pairwise rationalizable* if there is a complete linear order ρ on $\mathcal{A}^2$ such that for each $\succsim \in \mathcal{R}^{\mathbf{P}}$, $\varphi(\succsim)$ is the maximal element of ρ in $\mathcal{ER}^2(\succsim)$.

Pairwise rationalizable rules select the highest ranked assignment in the set of efficient and individually rational pairwise assignments according to a fixed linear order. Note that the linear order ρ on $\mathcal{A}^2$ that defines each pairwise rationalizable rule may depend on $\mathbf{P}$ but not on the specific preference profile reported by the patients. Hence, *pairwise rationalizable* rules satisfy *invariance*. Moreover, they clearly satisfy *weak consistency*. Hence, we obtain the following corollary from Propositions 1 and 2.

Corollary 1. *For each probability of graft survival matrix* $\mathbf{P}$, *if* φ *is a pairwise rationalizable rule then,* φ *satisfies individual rationality and* 2*-efficiency. Furthermore, truth-telling is the unique protective strategy for the patients.*

Example 1. For each patient i let $\sigma_i : [0, 1] \to \mathbb{R}$ be a non-decreasing function. For each $a \in \mathcal{A}$, the score of assignment a, is defined by

$$s(a) \equiv \sum_{i \in \{i' \in N:\ a_i \neq \omega_0\}} \sigma_i\bigl(p_i(a_i)\bigr).$$

Assume that $\mathbf{P}$ is such that for each $a, b \in \mathcal{A}^2$ $s(a) \neq s(b)$. Let the complete linear order $\rho^{\max}$ on $\mathcal{A}^2$ be such that for each $a, b \in \mathcal{A}^2$,

$$a\, \rho^{\max}\, b \quad \text{if} \quad \begin{array}{ll} \text{either} & \#\{i \in N \mid a_i \neq \omega_0\} > \#\{j \in N \mid b_j \neq \omega_0\}, \\ \text{or} & \#\{i \in N \mid a_i \neq \omega_0\} = \#\{j \in N \mid b_j \neq \omega_0\} \text{ and } s(a) > s(b). \end{array}$$

A rule $\tilde{\varphi}$ is a *exchange maximizing rule* if for each $\succsim \in \mathcal{R}^{\mathbf{P}}$, $\tilde{\varphi}(\succsim)$ is the maximal element of $\rho^{\max}$ in $\mathcal{ER}^2(\succsim)$.

It is worth to note that the kidney exchange programs implemented in several European countries employ protocols that resemble *exchange-maximizing* rules. The Dutch National Living Donor Kidney Exchange Program ranks assignments by focusing on patients' *ex ante* match probabilities of finding a compatible donor in the donor pool.[17] This would correspond to constant score functions σ_i. On the other hand, the recently initiated Spanish and British National Kidney Exchange Programs explicitly incorporate information based on the probabilities of graft survival.[18] Both programs, besides the ex ante match probability and other idiosyncratic variables, use score functions that depend on HLA-mismatch scores of the patients (which is emphasized in the UK) and the match between donor and patients ages (more relevant in Spain). European programs do not take into account patients' incentives and assume that patients accept any compatible exchange. However, by simply adjusting their allocation algorithms to take into account patients' preferences, their protocols would generate rules such that truth-telling is the unique protective strategy.

Proposition 2 shows that, when only pairwise exchanges are possible, under protective behavior *invariance* and *weak consistency* are almost sufficient for providing the right incentives to patients. Our next example shows however, that the restriction on pairwise exchanges is essential for the positive result. If larger cycles are possible, then there are matrices $\mathbf{P}$ for which truth-telling may fail to be a protective strategy.

[17] See Klerk et al. (2004, 2005), Keizer et al. (2005).

[18] See NHS Blood and Transplant (2009), Organización Nacional de Trasplantes (2009).

Example 2. Let $N = \{1, 2, 3\}$ and let

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 0.5 \\ 0.9 & 0 & 0.9 \\ 0 & 0.9 & 0 \end{pmatrix}.$$

The preference domain $\mathcal{R}^{\mathbf{P}}$ is defined by

$$\mathcal{R}_1^{\mathbf{P}} = \left\{ \begin{array}{cc} \succsim_1 & \succsim_1' \\ \hline \omega_2 & \omega_0 \\ \omega_0 & \dots \\ \dots & \dots \end{array} \right\}, \qquad \mathcal{R}_2^{\mathbf{P}} = \left\{ \begin{array}{cc} \succsim_2 & \succsim_2' \\ \hline \omega_3 & \omega_0 \\ \omega_0 & \dots \\ \dots & \dots \end{array} \right\}, \qquad \mathcal{R}_3^{\mathbf{P}} = \left\{ \begin{array}{ccc} \succsim_3 & \succsim_3' & \succsim_3'' \\ \omega_2 & \omega_2 & \omega_0 \\ \omega_1 & \omega_0 & \dots \\ \omega_0 & \dots & \dots \end{array} \right\}.$$

Let φ be a rule that satisfies *individual rationality* and 3-*efficiency*, and assume that $\varphi(\succsim) \in \mathcal{A}^3$. Thus $\varphi(\succsim) = [(1, \omega_2), (2, \omega_3), (3, \omega_1)]$. We can compute

$$\begin{aligned} c(\omega_0, \succsim_3) &= \{(\succsim_1, \succsim_2'), (\succsim_1', \succsim_2')\}, & c(\omega_0, \succsim_3') &= \{(\succsim_1, \succsim_2'), (\succsim_1', \succsim_2')\}, \\ c(\omega_1, \succsim_3) &= \{(\succsim_1, \succsim_2)\}, & c(\omega_1, \succsim_3') &= \emptyset, \\ c(\omega_2, \succsim_3) &= \{(\succsim_1', \succsim_2)\}, & c(\omega_2, \succsim_3') &= \{(\succsim_1, \succsim_2), (\succsim_1', \succsim_2)\}. \end{aligned}$$

Hence, $c(\omega_0, \succsim_3) = c(\omega_0, \succsim_3')$, and $c(\omega_1, \succsim_3') \subset c(\omega_1, \succsim_3)$. Therefore $\succsim_3' \mathrm{d}(\succsim_3) \succsim_3$, and reporting the true preferences is a dominated strategy for patient 3.

Example 2 shows that patients' private information introduces limitations on the welfare gains that can be obtained by multiple-way kidney exchanges. For rules that satisfy *individual rationality* and *k-efficiency*, it may be necessary to restrict the cardinality of the recommended exchanges to provide the right incentives for the patients. The problem affects any rule that maximizes the number of exchanges. Hence, Example 2 presents a strategic *rationale* for the focus on pairwise exchanges. In addition to the logistic and direct incentives problems described by Roth et al. (2005a), the restriction to pairwise exchanges may be necessary to obtain the correct information from patients in the protective behavior scenario.

Some final remarks are in order.

The results in this section depend crucially on information available to patients. In light of the proof of Proposition 2, it is clear that it is not necessary that the patients have perfect knowledge of the probability of graft survival matrix **P**. It suffices that the patients know the sets of compatible kidneys of the remaining patients. On the other hand, if each patient only has information about her own preferences, then the arguments in the proof of Proposition 2 would apply to prove a general version for arbitrary feasibility restrictions. In such an incomplete information framework, for each $k \leqslant n$, if a rule φ satisfies *individual rationality*, *k-efficiency* and *weak consistency*, then truth-telling is the unique protective strategy.

PKE programs involve the cooperation between different local transplant centers. Doctors at those centers have access to information about the kidney exchange process that is not available to the patients. Since patients usually follow the advice of their local doctors, doctors may induce the coordination among their patients. Thus, it is natural to investigate whether it is possible to find rules for which truth-telling is a unique protective strategy for groups of patients. It is immediate to find kidney exchange problems where the coordination among groups of patients is beneficial. For instance, let $N = \{1, 2, 3\}$,

$$\mathbf{P} = \begin{pmatrix} 0 & 0.9 & 0.5 \\ 0.9 & 0 & 0.75 \\ 0.3 & 0.3 & 0 \end{pmatrix},$$

and let $\succsim \in \mathcal{R}^{\mathbf{P}}$ be such that $\omega_2 \succ_1 \omega_3 \succ_1 \omega_0$ and $\omega_1 \succ_2 \omega_3 \succ_2 \omega_0$. For every rule that satisfies *individual rationality* and 2-*efficiency*, patients 1 and 2 would prefer to coordinate committing to report $\omega_2 \succ_1' \omega_0 \succ_1' \omega_3$ and $\omega_1 \succ_2' \omega_0 \succ_2' \omega_0$ rather than simply reporting their true preferences. Note that the coordination of the strategies has two effects. On the one hand, given the preference reported by the other patient, misreporting reduces (or eliminates) the possibility of being assigned the null kidney. On the other hand, both patients receive their best preferred kidney independently of the report of patient 3. Of course, the coordination among patients promoted by local doctors is more likely to be appear in decentralized programs.[19] In fact, under protective behavior, patients in local programs may prefer to arrange exchanges in local centers rather than enrolling in the national program. By doing so, they may avoid the possibility of not receiving a kidney even if the national program offers the possibility of finding a better match in the larger pool of donors.

Finally, in this article we have focused on situations in which each patient has only one possible donor. When patients can have multiple donors, it could be possible that a patient may have incentives to withdraw some of her possible donors if by doing so the assignment rule assigns her a better match. Again, with slight modifications of the arguments in the

[19] This is one of the major concerns on the introduction of the UNOS National Pilot PKE Program. See Ashlagi and Roth (2011). This problem is avoided in the centralized program recently implemented in Spain, by explicitly insisting on the *ex ante* anonymity of the protocol.

proof of Proposition 2, we can prove that rules that satisfy a version of Arrow's Axiom of Choice (defined over feasible assignments) are immune to such manipulations.

6. Concluding remarks

In this paper, we have proposed a framework that incorporates match quality and patients' preferences in the design of PKE programs. Since patients' welfare depends on the quality of the match with the potential donors, we assume that patients do not consider compatible kidneys as homogeneous. We show that limitations on the number of simultaneous required operations imply that every efficient PKE protocol introduces incentives for the patients to misreport the minimum match-quality that they would accept. In order to overcome these unsatisfactory results, we introduce alternative assumptions on patients' strategic behavior. Namely, we assume that patients waiting for a transplant follow a refinement of maximin behavior, the protective behavior (Barberà and Dutta, 1982). In this scenario, we obtain mixed results. If only exchanges involving two donor–patient pairs are admitted, then priority based rules implemented by different PKE programs around the world provide right incentives for the patients. If larger exchange chains are feasible but some restriction remains, the positive result unravels.

Regarding the relevance of match quality in transplant from living donors, we think it is a issue that has major welfare consequences. We stress in the manuscript that many National PKE programs recognize the heterogeneity in the quality of the compatible matches for each patient. We believe that transplant quality and patients' participation constraints are important issues for PKE program. On the one hand, the probability of performing successful exchanges depends on the size of the pool of patient-donor pairs participating to the program. Moreover, our analysis can be immediately extended to include patients with compatible donors. By allowing such donor–patient pairs to get a better match, a PKE program can dramatically increase the number of performed exchanges. (See Gentry et al., 2007; Roth et al., 2005b.) Therefore, efficiency losses that a PKE program that focuses on pairwise exchange may suffer with respect to a program that allows larger chains of exchanges may be overcome by the gains obtained by enlarging the pool of participants. On the other hand, a better match increases the life expectancy of the recipient and reduces the probability that this patient will need another operation in the future, lowering the pressure on the list of patients waiting for a transplant.

References

Arrow, K.J., 1959. Rational choice functions and orderings. Economica 26, 121–127.

Ashlagi, I., Roth, A.E., 2011. Individual rationality and participation in large scale, multi-hospital kidney exchange. Unpublished manuscript, MIT and Harvard Business School.

Barberà, S., Dutta, B., 1982. Implementability via protective equilibria. J. Math. Econ. 10, 49–65.

Barberà, S., Dutta, B., 1995. Protective behavior in matching models. Games Econ. Behav. 8 (2), 281–296.

Barberà, S., Jackson, M.O., 1988. Maximin, leximin, and the protective criterion: Characterizations and comparisons. J. Econ. Theory 46, 34–44.

Delmonico, F.L., 2004. Exchanging kidneys – advances in living-donor transplantation. N. Engl. J. Med. 350, 1812–1814.

Delmonico, F.L., Lipkowitz, G.S., Morrissey, P.E., Stoff, J.S., Himmelfarb, J., Harmon, W., Pavlakis, M., Mah, H., Goguen, J., Luskin, R., Milford, E., Chobanian, G.B.M., Bouthot, B., Lorber, M., Rohrer, R.J., 2004. Donor kidney exchanges. Amer. J. Transplantation 4, 1628–1634.

Duquesnoy, R.J., Takemoto, S., de Lange, P., Doxiadis, I.I.N., Schreuder, G.M.T., Persijn, G.G., Claas, F.J., 2003. HLAMatchmaker: A molecularly based algorithm for histocompatibility determination. III. Effect of matching at the HLA–A,B amino acid triplet level on kidney transplant survival. Transplantation 75 (6), 884–889.

Freeman, R.B., 2007. Survival benefit: Quality versus quantity and trade-offs in developing new renal allocation systems. Amer. J. Transplantation 7, 1043–1046.

Gale, D., Shapley, L., 1962. College admission and stability of marriage. Amer. Math. Monthly 69, 9–15.

Gentry, S.E., Segev, D.L., Simmerling, M., Montgomery, R.A., 2007. Expanded kidney paired donation through participation by compatible pairs. Amer. J. Transplantation 7, 99–107.

Hatfield, J., 2005. Pairwise kidney exchange: Comment. J. Econ. Theory 125, 189–193.

Keizer, K.M., Klerk, M.d., Haase-Kromwijk, B.J., Weimar, W., 2005. The Dutch algorithm for allocation in living donor kidney exchange. Transplantation Proc. 37, 589–591.

Klerk, M., Keizer, K.M., Weimar, W., 2004. Donor exchange for renal transplantation. N. Engl. J. Med. 351, 935.

Klerk, M., Keizer, K.M., Class, F.H., Wiltvliet, M., Haase-Kromwijk, B.J., Weimar, W., 2005. The Dutch national living donor kidney exchange program. Amer. J. Transplantation 5, 2302–2305.

Klijn, F., Pais, J., Vorsatz, M., 2010. Preference intensities and risk aversion in school choice: A laboratory experiment. Harvard Business School, working paper 10-093.

Kranenburg, L.W., Visak, T., Weimar, W., Zuidema, W., Klerk, M.d., Hilhorst, M., Passchier, J., Busschbach, J.N.M.I.J.J., 2004. Startling a crossover kidney transplantation program in the Netherlands: Ethical and psychological considerations. Transplantation 78, 194.

Ma, J., 1994. Strategy-proofness and the strict core in a market with indivisibilities. Int. J. Game Theory 23, 75–83.

Merion, R.M., Ashby, V.B., Wolfe, R.A., Distant, D.A., Hulbert-Shearon, T.E., Metzger, R.A., Ojo, A.O., Port, F.K., 2005. Deceased-donor characteristics and the survival benefit of kidney transplantation. J. Amer. Med. Assoc. 294 (21), 2726–2733.

NHS Blood and Transplant, 2009. National matching scheme for paired and pooled (kidney) donation. http://www.organdonation.nhs.uk/ukt/about_transplants/or_allocation/kidney_(renal)/living_donation/paired_donation_matching_scheme.jsp (accessed, March 2, 2011).

Øien, C.M., Reisæter, A.V., Leivestad, T., Dekker, F.W., Line, P.D., Os, I., 2007. Living donor kidney transplantation: The effects of donor age and gender on short- and long-term outcomes. Transplantation 83 (5), 600–606.

Opelz, G., 1997. Impact of HLA compatibility on survival of kidney transplants from unrelated live donors. Transplantation 64, 1473–1475.

Organización Nacional de Trasplantes, 2009. Programa Nacional de Donación Renal Cruzada en España. Technical report, ONT, Documentos de Consenso, http://www.ont.es/infesp/DocumentosDeConsenso/Programa_Donacion_Renal_Cruzada.pdf (accessed, March 2, 2011).

Park, K., Lee, J., Huh, K., Kim, S., Kim, Y.S., 2004. Exchange living-donor kidney transplantation: Diminution of donor organ shortage. Transplantation Proc. 36, 2949–2951.

Roth, A.E., 1982. Incentive compatibility in a market with indivisible goods. Econ. Letters 9, 127–132.
Roth, A.E., 1989. Two-sided matching with incomplete information about others' preferences. Games Econ. Behav. 1, 191–209.
Roth, A.E., Postlewaite, A., 1977. Weak versus strong domination in a market of indivisible goods. J. Math. Econ. 4, 131–137.
Roth, A.E., Sönmez, T., Ünver, M.U., 2004. Kidney exchange. Quart. J. Econ. 119, 457–488.
Roth, A.E., Sönmez, T., Ünver, M.U., 2005a. Pairwise kidney exchange. J. Econ. Theory 125, 151–188.
Roth, A.E., Sönmez, T., Ünver, M.U., 2005b. A kidney exchange clearinghouse in New England. Amer. Econ. Rev. Pap. Proc. 95, 376–380.
Roth, A.E., Sönmez, T., Ünver, M.U., 2007. Efficient kidney exchange: Coincidence of wants in a markets with compatibility preferences. Amer. Econ. Rev. 97 (3), 828–851.
Schnitzler, M.A., Hollenbeak, C.S., Cohen, D.S., Woodward, R.S., Lowell, J.A., Singer, G.G., Tesi, R.J., Howard, T.K., Mohanakumar, T., Brennan, D.C., 1999. The economic implications of HLA matching in cadaveric renal transplantation. N. Engl. J. Med. 341, 1440–1446.
Segev, D.L., Gentry, S.E., Melancon, J.K., Montgomery, R.A., 2005a. Characterization of waiting times in a simulation of kidney paired donation. Amer. J. Transplantation 5, 2448–2455.
Segev, D.L., Gentry, S.E., Warren, D.S., Rech, B., Montgomery, R.A., 2005b. Kidney paired donation and optimizing the use of living donor organs. J. Amer. Med. Assoc. 293, 1883–1890.
Sen, A.K., 1971. Choice functions and revealed preference. Rev. Econ. Stud. 38, 307–317.
Shapley, L., Scarf, H., 1974. On cores and indivisibility. J. Math. Econ. 1, 23–37.
Sönmez, T., 1999. Strategy-proofness and essentially single-valued cores. Econometrica 67, 677–689.
Sönmez, T., Ünver, M.U., 2010. Altruistic kidney exchange. Unpublished manuscript, Boston College.
Spital, A., 2004. Donor exchange for renal transplantation. N. Engl. J. Med. 351, 936.
Su, X., Zenios, S.A., 2005. Patient choice in kidney allocation: A sequential stochastic assignment model. Operations Res. 53, 443–455.
Su, X., Zenios, S.A., 2006. Recipient choice can address the efficiency-equity trade-off in kidney transplantation: A mechanism design model. Manage. Sci. 52, 1647–1660.
Su, X., Zenios, S.A., Chertow, G.M., 2004. Incorporating recipient choice in kidney transplantation. J. Amer. Soc. Nephrology 15, 1656–1663.
Ünver, M.U., 2010. Dynamic kidney exchange. Rev. Econ. Stud. 77, 372–414.
Villa, S., Patrone, F., 2009. Incentive compatibility in kidney exchange problems. Health Care Manage. Sci. 12, 351–362.
Yilmaz, Ö., 2011. Kidney exchange: An egalitarian mechanism. J. Econ. Theory 146, 592–618.