A Dynamic Bargaining Framework for International Kidney Paired Exchange Programs *

Antonio Nicolò[†], Subhajit Pramanik[‡]

December 5, 2025

Abstract

This paper develops a theoretical framework for International Kidney Paired Exchange (IKPE) to address efficiency and fairness concerns in multicountry kidney exchanges. Drawing on a discrete version of the Kalai-Smorodinsky (KS) bargaining solution, we propose a mechanism that first maximizes the least-gained country's relative gain and then selects a matching that achieves the largest overall transplant benefit. The model features a dynamic weighting scheme that adjusts over time to compensate countries whose participation yields lower gains in the past rounds compared to other countries, ensuring an equitable distribution of cooperative gains. We prove that the proposed mechanism guarantees per-period Pareto optimality and individual rationality of the resulting allocation, as no country is worse off by cooperating than by acting alone. We also show that over multiple periods, dynamic weighting corrects historical imbalances, eventually converging countries' cumulative gains under mild stochastic assumptions. We further extend the framework to a dual-objective setting that simultaneously protects total transplants and access for hard-to-match recipients via a bi-criteria KS screen, retaining the per-period guarantees and long-run convergence.

Keywords: International kidney paired exchange, Kalai-Smorodinsky, dynamic weights.
JEL Codes: C78, D47.

*We greatly appreciate the comments of and discussions with David Delacrétaz. We thank the audience at the 14th Conference on Economic Design (Essex), the 20th Matching in Practice Workshop (Messina), PhD Day Edition IV (Verona) and at the seminars in Manchester and, Padova.

[†]University of Padova and University of Manchester. E-mail: antonio.nicolo@unipd.it
[‡]University of Padova. E-mail: subhajit.pramanik@phd.unipd.it

1 Introduction

The exchange of kidneys across incompatible donor-recipient pairs has become one of the most successful applications of market design in healthcare. Kidney exchange programs (KEPs) allow patients with willing but incompatible donors to swap donors with others in similar situations, creating mutually compatible transplant opportunities. In recent years, the possibility of international cooperation between national kidney exchange programs has attracted growing attention. When countries merge their pools of incompatible pairs, the joint pool becomes larger and more diverse, increasing the likelihood of finding compatible matches and thereby raising the total number of transplants performed.

Examples include the Scandinavian program (Andersson and Kratz (2020)), the South Alliance for Transplant (SAT) between Italy, Spain, and Portugal (Francisco et al. (2024)), and a recent pilot agreement between the United States and Italy.[1] These collaborations aim to broaden the pool of donor-recipient pairs and thereby increase the number of transplants.

Despite their promise, international programs remain underdeveloped and are often used primarily for pairs that are difficult to place domestically. A persistent challenge is cream skimming: countries prioritize easy matches within their own borders and leave only hard-to-match pairs for international exchange. As a result, the number of successful matches is reduced, and much of the potential benefit from pooling is lost.

Beyond cream skimming, IKPE faces additional obstacles: legal and logistical barriers related to cross-border organ transfers, the high costs of transporting kidneys, and differences in healthcare infrastructure across countries. Another major concern is the distribution of gains from cooperation. Countries with larger national pools may be less motivated to participate since they can secure sufficient matches domestically, while countries with smaller pools stand to benefit disproportionately. Without mechanisms to ensure equity, sustained collaboration is difficult to achieve.

Nevertheless, IKPE offers substantial potential to improve transplant opportunities for all participating countries. For such a program to be effective and sustainable in the long run, benefits must be distributed fairly across countries. A balanced approach is essential to maintain cooperation and ensure that each participant receives an equitable share of the gains from pooling.

While the efficiency gains from cooperation are well recognized, how to share these gains among participating countries remains a challenging and central question.

In this paper, we propose an allocation approach inspired by the Kalai-Smorodinsky (KS) bargaining solution, adapted to the discrete setting of kidney exchange matchings. Our main argument is that the KS framework offers a natural and compelling notion of fairness for cooperative problems of this kind. The idea of the KS solution is that each

[1] This program is managed by the Alliance for Paired Kidney Donation (link).

agent (or country, in our setting) should receive the same proportion of its maximum attainable benefit that is, the same fraction of the best possible outcome it could achieve through cooperation. This principle captures fairness in proportional gains, ensuring that no country is systematically disadvantaged relative to its potential contribution or benefit.

To situate our proposal, we note that the Nash and Kalai-Smorodinsky bargaining solutions are among the two most influential frameworks for determining fair outcomes in bargaining theory. Both emerge from a set of axioms that formalize principles of rationality and fairness in situations where two or more agents must agree on how to divide a feasible set of payoffs. The Nash solution maximizes the product of the agents' gains relative to a disagreement point, capturing a notion of compromise between competing interests. In contrast, the Kalai-Smorodinsky solution replaces Nash's axiom of independence of irrelevant alternatives with a monotonicity axiom, emphasizing equity in proportional achievement rather than compromise. In other words, while the Nash solution is compromise-oriented, the KS solution is equity-oriented.

This distinction is fundamental in our context. In international kidney exchange, the problem is not primarily one of conflicting interests to be reconciled, but rather one of sharing the benefits of cooperation in a fair and stable way. When countries agree to pool their incompatible donor-recipient pairs, they jointly expand the feasible set of transplants that can be performed. The resulting gains are cooperative surpluses, not conflicting claims. Therefore, an allocation principle that ensures proportional fairness where each country's benefit reflects its potential maximum within the cooperative arrangement is more appropriate than one based on compromise. The Kalai-Smorodinsky solution provides exactly this kind of fairness benchmark.

In practical terms, our contribution is to develop a discrete adaptation of the Kalai-Smorodinsky solution suitable for kidney exchange, where allocations are inherently indivisible and represented by integer-valued matchings. The proposed algorithm seeks outcomes that maintain the spirit of KS proportionality while ensuring feasibility and efficiency within the matching framework. By doing so, it promotes a balance between equity, efficiency, and incentive compatibility - three conditions essential for sustaining international cooperation in organ exchange.

Our paper makes two main contributions. First, it provides a conceptual justification for using equity-oriented bargaining principles to guide the allocation of benefits in international kidney exchange programs. Second, it offers an algorithm to compute a solution to the discrete and combinatorial environment of kidney exchange. Together, these contributions aim to strengthen the theoretical and practical foundations for fair and sustainable international kidney exchange systems.

We then extend the framework to the case in which each country pursues a dual objective: first, to maximize the number of its hard-to-match recipients who receive a transplant, and second, to maximize the total number of its recipients who are transplanted.

The remainder of this paper is organized as follows. Section 2 reviews the literature and specifies the gap where this paper lies. Section 3 introduces the notation and mechanism; Section 4 establishes its normative properties; Section 5 extends the framework to the case in which each country pursues a dual objective: giving priority to maximizing the number of hard-to-match recipients who receive a transplant, and conditional on this maximizing the total number of transplants; finally, Section 6 presents simulation evidence based on realistic assumptions about the underlying patient-donor population.

2 Related Literature

The modern design of kidney exchange programs originates from the seminal work of Roth et al. (2004, 2005), who formulated kidney exchange as a matching market and proposed algorithmic mechanisms based on the top trading cycles and integer programming approaches. Subsequent contributions by Ashlagi and Roth (2014) and others have analyzed the structure and efficiency of large exchange pools, providing theoretical foundations for national programs such as those operating in Europe.

In the United States, hospitals operate in a competitive environment in the market for organ transplantation, and several multi-hospital kidney exchange programs coexist. A substantial body of research has focused on how to reduce free-riding incentives and promote active participation in these programs. Hospitals typically have the option to join one or more kidney exchange networks, but they may also strategically withhold some of their patient-donor pairs. These strategic behaviors can undermine the efficiency of these multi-hospital programs and pose challenges for sustaining large-scale participation (Ashlagi and Roth (2012, 2014); Ashlagi et al. (2015); Toulis and Parkes (2015)). To address this issue, Agarwal et al. (2019) and Hajaj et al. (2015) proposed credit-based schemes, which were implemented in the National Kidney Registry, the largest US kidney exchange program. These schemes influence hospitals' incentives to submit pairs without altering the matching algorithm itself.

In Europe, International Kidney Exchange (IKPE) programs have been implemented between the Czech Republic and Austria (Böhmig et al. (2017)), as well as between Portugal, Spain, and Italy since 2018, and between Sweden, Norway, and Denmark in the Swedish-Danish Transplant Program (STEP), initiated by Sweden (Andersson and Kratz (2020)). The Italy-Portugal-Spain cooperation follows a stepwise structure: first, conducting national-level matches and then seeking international matches for the remaining donor-patient pairs.

Many of the participation and incentive issues observed in the United States also arise in cross-border collaborations in Europe. Ensuring that participating countries accept the proposed solutions is crucial for maintaining full participation in the IKPE. Without this, countries may exit the program. Therefore, long-term participation is essential for success in these programs. To address social equilibria, Carvalho and Lodi (2023) introduced a 2-round system with 2-way exchanges and provided a polynomial-time algorithm for computing a Nash equilibrium that maximizes the total number of transplants, improving on the results of Carvalho et al. (2017) for two countries. Similarly, Sun et al. (2021) considered 2-way exchanges and proposed a solution deemed fair if the minimum ratio of transplants across all countries was maximized.

In contrast to these approaches, Benedek et al. (2024, 2025) introduced the framework for partitioned matching games for IKPE. Their model adopts a credit-based adjustment rule (originating from Klimentova et al. (2016)): in each round, the algorithm first finds a maximum-cardinality matching and then chooses, among all maximum cardinality matchings, the one whose country-level outcomes lie closest to a predetermined target allocation.[2] Deviations from this target are carried forward as positive or negative credits, which shift next round's targets. Since fairness is implemented as a post-processing step: only after restricting attention to maximum-cardinality matchings, this framework cannot impose individual rationality as a hard constraint, guaranteeing each country its stand-alone benchmark may require accepting fewer transplants, which the model does not allow.

The ENCKEP simulator (Druzsin et al. (2024)) implements this partitioned-matching approach to evaluate real cross-border collaborations in Europe. Simulation studies (Druzsin et al. (2024); Matyasi and Biró (2024)) using the ENCKEP framework consistently show that fully pooling all countries' incompatible donor-recipient pairs into a single international graph yields substantially more transplants than the common two-step practice of running national exchanges first and sending only leftovers to the international pool. Across realistic multi-year scenarios, the joint or "borderless" policy outperforms the consecutive policy by roughly 10-15%, confirming that partial participation systematically sacrifices match efficiency. Subsequent work e.g., Biró et al. (2021) and the IP formulation in Mincu et al. (2021) extends this framework to accommodate heterogeneous national regulations and country-specific feasibility constraints.

Our contribution complements this literature by developing a fairness framework that differs in two conceptual respects. First, fairness and efficiency are embedded directly into the matching objective rather than introduced as an adjustment to the set of maximum matchings, allowing these goals to be treated as jointly determined rather than sequential. Second, the mechanism incorporates intertemporal considerations through a dynamic edge-weighting scheme that adjusts weights in response to cumulative country-level imbalances: countries that have received fewer transplants than other countries

[2]These target allocations, often represented as target vectors specifying the desired distribution of transplants across countries, evolve over time through accumulated credits.

obtain higher weights in subsequent periods, while those with higher outcomes receive lower weights. This formulation enables the mechanism to guarantee, in every period, that each country achieves at least its stand-alone outcome thereby satisfying individual rationality, while sustaining high overall performance across periods. By allowing the objective to internalize intertemporal fairness directly, the mechanism provides a unified approach to participation, equity, and efficiency without relying on exogenous target vectors or lexicographic tie-breaking.

To formalize this principle, we treat the international kidney exchange problem as a cooperative bargaining problem and use the Kalai and Smorodinsky (1975) (KS) solution as our fairness benchmark, a bargaining concept that proportionally distributes gains from cooperation. Embedding the KS criterion at each time period directly into the matching objective operationalizes the dynamic weighting scheme described above, ensuring that the mechanism maintains voluntary, self-enforcing cooperation while preventing any country from being systematically disadvantaged in the long run.

3 Setup

Consider a set $C = \{1, 2, \ldots, n\}$ of countries, each with a national pool of incompatible donor-recipient pairs who are willing to collaborate for creating a supra-national exchange program. Let c denote a generic country in C. We consider discrete time periods $t = 1, 2, \ldots, T$. At each time t, country c has a pool p_c^t of donor-recipient pairs; some pairs may arrive anew at time t, while others remain from previous periods if they were not matched in earlier rounds. We model the transplant problem at time t by a directed graph $G^t = (V^t, E^t)$:

- The node set V^t collects all incompatible donor-recipient pairs across all countries at time t.
- A directed edge $(v, v') \in E^t$ indicates that the donor of pair v can donate to the recipient of pair v', thus making a transplant from v to v'.

We allow edges to connect pairs from different countries, enabling international cooperation. Each edge $(v, v') \in E^t$ with v' belonging to country c is endowed with *non-zero positive weight* $w_c^t(v, v') > 0$. Here, the weight is associated with the ***recipient's*** country; $w_c^t(v, v')$ reflects the importance or priority assigned to a given transplant by the mechanism designer, relative to other possible transplants. At the start of the collaboration (time $t = 1$), all transplants are assumed to be equally valuable, $w_c^1(v, v') = 1$ for all $c \in C$. Later these weights are dynamically adjusted to capture the imbalance of gains between the countries in previous time periods.

Definition 1 (Matching)**.** *A* matching *at time t is a set of disjoint cycles $M^t \subseteq E^t$ in the directed graph G^t.*

Let $\mathcal{F}^t$ be the set of all matchings that can be performed at time t. Given a matching M^t let $\mathcal{S}_c^t$ be the set of recipients from country c who are transplanted at M^t and $X_c^t(M^t) \equiv |\mathcal{S}_c^t|$ as the **number** of recipients from country c who are transplanted at M^t. For brevity we will write $X_c^t(M^t)$ as X_c^t when no confusion arises.

Each country $c \in C$ is equipped with a preference relation $\succsim_c$ over the set $\mathcal{F}^t$. We assume that, in each period, every country aims to maximize the number of transplants received by patients within its national pool. Two remarks are important regarding the restrictions on countries' preferences. First, we assume that countries are indifferent to which specific patients receive an organ within their own national pool. In section 5, we consider an important extension in which countries prioritize patients who are harder to match, i.e., those with a lower probability of finding a compatible donor, over patients who are easier to match.

Second, we assume that countries aim to maximize the per period number of transplants, rather than the total number of transplants over multiple periods. This assumption is consistent with the current practice in most kidney exchange programs and can be justified in two ways. First, it is difficult to predict which patient-donor pairs will join the pool in the future. Second, and more importantly, it respects the normative principle that each patient should receive the best possible treatment and, consequently, should be transplanted as soon as this is the optimal option for them. Formally, for each country $c \in C$ we assume that $M^t \succsim_c \tilde{M}^t \quad \text{if and only if} \quad X_c^t \geq \tilde{X}_c^t$.

Let $G_c^t = (V_c^t, E_c^t)$ denote a subgraph of G^t restricted to pairs belonging to country c, M_c^t a corresponding matching, and let $d_c^t = \max\Big\{\text{number of vertices in a set of disjoint cycles } M_c^t \subseteq E_c^t\Big\}$ denote the number of transplants country c can perform *independently* at period t. Specifically, if country c *opts out* the international exchange program and only uses its *own* donor-recipient pairs, the d_c^t is the maximum number of transplants c can realize from its own pool alone, by forming disjoint set of cycles among its *intra-country* pairs.[3] We refer to d_c^t as the **disagreement option** for country c in period t.

For any matching $M^t \in \mathcal{F}^t$, the total number of transplants performed across all countries at time t is given by:

$$\mathbf{X}^t = \sum_{c \in C} X_c^t$$

This represents the aggregate number of transplants realized from the matching M^t.

Given a set of matchings $\mathcal{F}^t$ in a graph $G^t = (V^t, E^t)$, let B_c^t denote the largest number of transplants that country c could get at time t when *all* countries cooperate,

[3]Here, we only consider cycles. We do not consider any sort of chains as part of the matchings.

representing the most favourable outcome for country c. Formally,

$$B_c^t = \max_{M^t \in \mathcal{F}^t} \left\{ X_c^t(M^t) \, \forall \, M^t \in \mathcal{F}^t \right\},$$

where $\mathcal{F}^t$ is the set of *all matchings* at time t, and X_c^t denotes the number of transplants assigned to country c in that matching.

Taken together, a country's stand-alone payoff d_c^t and cooperative best B_c^t delimit its period-t surplus window; accordingly, in §3.1 we implement a *discrete* Kalai-Smorodinsky screen that targets proportional equity by maximizing the minimum realized share $\big(X_c^t - d_c^t\big)/\big(B_c^t - d_c^t\big)$ across countries before selecting a maximum-weight matching within the resulting feasible set.

3.1 The algorithm: DAKSA (Dynamically Adjusted Kalai-Smorodinsky Algorithm)

We adopt the idea of equitable distribution of transplants following *Kalai-Smorodinsky (KS)* bargaining solution (Kalai and Smorodinsky (1975)), but adapted for discrete settings as in Lahiri (2003). To identify a solution of an international kidney program for each period t we adopt a two-step procedure.

- **Step 1.t (Selecting equitable and individually rational allocations):** For each period t, and for each country c, let

$$r_c^t(M^t) = \frac{X_c^t(M^t) - d_c^t}{B_c^t - d_c^t}, \tag{1}$$

  if $B_c^t > d_c^t$, and $r_c^t(M^t) = 1$ otherwise[4]. Let $r_c^t(M^t)$ be the **relative gain country** c **gets at** M^t. Let r^t be the *highest-worst KS ratio* in period t

$$r^t \equiv \max_{M^t \in \mathcal{F}^t} \min_{c \in C} r_c^t(M^t). \tag{2}$$

  Intuitively, the ratio $r_c^t(M^t)$ measures the share of cooperative gain that country c secures under matching M^t. If a country receives only its disagreement outcome d_c^t, then $r_c^t(M^t) = 0$; if it attains its best possible cooperative outcome B_c^t, then $r_c^t(M^t) = 1$. Values between 0 and 1 indicate partial gains from cooperation. Thus, r^t serves as the fairness benchmark for the period: only matchings that ensure each country achieves at least this guaranteed share are considered in the next step.

- **Step 2.t (Maximizing Recipient-Weighted Efficiency):** Let $\mathcal{E}^t \subset \mathcal{F}^t$ be a set

[4] Countries with $B_c^t = d_c^t$ are treated as fully satisfied, with ratio 1.

of matchings in period t such that for each $M^t \in \mathcal{E}^t$,

$$\min_{c \in C} \Big(\frac{X_c^t(M^t) - d_c^t}{B_c^t - d_c^t} \Big) \quad \geq \quad r^t$$

By construction, $\mathcal{E}^t$ is non-empty. Among the set of matchings $\mathcal{E}^t$ the solution in period t is the *matching* (set of disjoint cycles) M^{*t} that *maximizes the total transplant weight*:

$$M^{*t} \equiv \arg\max_{M^t \in \mathcal{E}^t} \sum_{(v,v') \in M^t} w_c^t(v, v').$$

By construction $w_c^1(v, v') = 1$ for all pairs v' belonging to c and for all $c \in C$. For each $t > 1$, the weight updates as follows:

$$w_c^t(v, v') \;=\; w_c^{t-1}(v, v') + (1 - r_c^{t-1}(M^{*t-1})), \tag{3}$$

Rationale for a KS-style fairness screen: In each period, a country's stand-alone capability d_c^t (what it can achieve within its own pool) and its cooperative best B_c^t (what it could receive under full cooperation) jointly pin down the feasible interval of gains for that country. The KS ratios

$$r_c^t(M) \;=\; \frac{X_c^t(M) - d_c^t}{B_c^t - d_c^t}$$

measure the *share of cooperative surplus* actually realized by c under a matching M, i.e., how far $X_c^t(M)$ lies between these two benchmarks $[d_c^t,\, B_c^t]$. Framing fairness in terms of *proportional equity of gains* therefore aligns naturally with a KS logic: we protect disagreement payoffs while balancing access to the surplus generated by cooperation. Because kidneys are indivisible and the feasible set consists of discrete matchings, the continuous KS point is typically infeasible; we operationalize its spirit via a *discrete KS screen* that $\text{maxmin}_c\, r_c^t(M)$ in Step 1, then select a maximum-weight matching within the KS-feasible set in Step 2 so that no "free" transplants that preserve the minimum share are left on the table. This implementation preserves per-period individual rationality, enforces proportional equity in realized gains, and dovetails with our dynamic weights that correct intertemporal imbalances.[5]

Remark 1 (Expansion monotonicity of the discrete KS screen)**.** *If the feasible set of matchings expands from $\mathcal{F}^t$ to $\widetilde{\mathcal{F}}^t \supseteq \mathcal{F}^t$, then the guaranteed share weakly increases:* $\widetilde{r}^{\,t} \;\equiv\; \max_{M \in \widetilde{\mathcal{F}}^t} \min_c r_c^t(M) \;\geq\; r^t$.

[5] *Dual-objective extension:* The same rationale applies in component-wise *total transplants* and *hard-to-match* (§5) via a bi-criteria screen that enforces proportional equity along both dimensions before the weighted tie-break.

Remark 2 (Normalization and Comparability)**.** *Because each ratio $r_c^t(M)$ is defined using the country-specific disagreement (d_c^t) and best (B_c^t) benchmarks, it takes values in $[0, 1]$ and is already normalized by the feasible gain interval of country c. Hence, these ratios are directly comparable across countries without any additional scaling or transformation.*

Rationale for dynamic weights: Over multiple periods, certain countries might repeatedly receive fewer transplants because of the discrete nature of the matchings. To address this, the second step maximizes the weighted sum of transplants, where edge weights are dynamically adjusted to guarantee a fair distribution of the gains from cooperation.

In summary, the algorithm here aims to balance the goals of maximizing the total number of transplants and ensuring long-term participation and equitable distribution among countries. By incorporating recipient-weighted maximization and dynamic weight adjustments to address imbalances, the algorithm strives to create a fair and efficient kidney exchange system.

The two-step selection rule and its dynamic weight update can be expressed operationally as the pseudocode below; this is an implementation of the period-t definition given above:

Algorithm 1 DAKSA: Dynamically Adjusted Kalai-Smorodinsky Algorithm

Require: Countries $C = \{1, \dots, n\}$; horizon T; per-period graphs $G^t = (V^t, E^t)$ (with carryovers from period $t-1$); initial weights $w_c^1(v, v') \equiv 1$ for all edges whose recipient v' is in country c.
Ensure: A matching $M^{*t} \in \mathcal{F}^t$ for each t.

For all $c \in C$, set $r_c^0 \leftarrow 0$.
for $t = 1$ **to** T **do**
  Graph G^t is given or built from carryover $\cup$ new arrivals
  for each country $c \in C$ **do**
    Compute d_c^t = maximum transplants in c's standalone subgraph of G^t.
    Compute B_c^t = maximum transplants for c under full cooperation ($M \in \mathcal{F}^t$).
  end for
  if $t > 1$ **then** ▷ Dynamic weight update
    for each edge $(v \to v') \in E^t$ with recipient $v' \in c$ **do**
      $w_c^t(v, v') \leftarrow w_c^{t-1}(v, v') + \left(1 - r_c^{t-1}\right)$
    end for
  end if
  Step 1.t: $r^t \leftarrow \max_{M \in \mathcal{F}^t} \min_{c \in C} \ \rho_c^t(M)$ where $\rho_c^t(M) = \begin{cases} \frac{X_c^t(M) - d_c^t}{B_c^t - d_c^t}, & B_c^t > d_c^t \\ 1, & B_c^t = d_c^t \end{cases}$
  $\mathcal{E}^t \leftarrow \{\, M \in \mathcal{F}^t : \min_c \rho_c^t(M) \geq r^t \,\}$ ▷ KS-feasible set
  Step 2.t: $M^{*t} \in \arg\max_{M \in \mathcal{E}^t} \sum_{(v,v') \in M} w_c^t(v, v')$ ▷ Receiver $v' \in c$
  for each $c \in C$ **do** ▷ Record realized ratios for next update
    $X_c^t \leftarrow X_c^t(M^{*t}); \quad r_c^t \leftarrow \begin{cases} \frac{X_c^t - d_c^t}{B_c^t - d_c^t}, & B_c^t > d_c^t \\ 1, & B_c^t = d_c^t \end{cases}$
  end for
end for

To interpret the selected outcome in each period and to structure the properties that follow, we recall the standard efficiency and participation criteria used throughout the paper. We first formalise the notion of Pareto efficiency that underpins our welfare comparisons.

Definition 2 (Pareto Efficiency)**.** *A matching* M^t *is* Pareto efficient *if there exists no* $\tilde{M}^t \in \mathcal{F}^t$ *such that:*

1. $X^t_c(\tilde{M}^t) \;\geq\; X^t_c(M^t)$ *for* all c*, and*

2. $X^t_{c'}(\tilde{M}^t) \;>\; X^t_{c'}(M^t)$ *for* at least one $c' \in C$*.*

In other words, within each period t, no improvement can be made where at least one country is strictly better off without making another country worse off; i.e. the standard weak Pareto ordering with at least one strict improvement.
Next, we state the period-by-period participation guarantee relative to each country's standalone capacity.

Definition 3 (Individual Rationality)**.** *A matching* $M^t \in \mathcal{F}^t$ *is* individually rational *if*

$$X^t_c \;\geq\; d^t_c \quad \textit{for every country } c \in C.$$

That is, in each period t, each country c receives at least as many transplants as it could have achieved alone from its own pool.
To analyse mergers of pools and the monotonicity result, we also work with coalition-level feasibility and the associated maximal cooperative yield. Let C be the set of countries and $A_1, \cdots, A_k \subseteq C$ are disjoint coalition of countries. For any coalition $A_i \subseteq C$, define

$$\mathbf{X}^{*t}(A_i) \;=\; \max_{M^t \in \mathcal{F}^t(A_i)} \sum_{c \in A_i} X^t_c(M^t),$$

the maximal total transplants achievable when only countries in A_i participate, with $\mathcal{F}^t(A_i)$ the set of matchings restricted to A_i.
Using the coalitional yield notation, the monotonicity property we employ is captured by super-additivity, stated next. As presented in Kalai (2008),

Definition 4 (Super-additivity / Coalitional Monotonicity)**.** *A rule satisfies* super-additivity *(also known as* coalitional monotonicity*) if, for all disjoint coalitions* $A_i \subseteq C$*,*

$$\mathbf{X}^{*t}(A_1 \cup A_2 \cup \cdots A_k) \;\geq\; \mathbf{X}^{*t}(A_1) + \mathbf{X}^{*t}(A_2) + \cdots + \mathbf{X}^{*t}(A_k).$$

Taken together, the procedural rule and the definitions above set the terminology and framework for the analysis in later sections, where these properties will be examined in detail.

4 Model Properties and Discussion

Here, we show that the proposed solution satisfies several desirable normative properties. Throughout, M^{*t} denotes the matching selected by DAKSA in period t. We refer to the definitions in the section 3.

Using Definition 2, we check whether the period-t matching M^{*t} leaves any possible Pareto improvement unused and we observe the following:

Proposition 1 (Pareto Efficiency)**.** *M^{*t} is Pareto-efficient.*

Implication within the KS-feasible set: In the second step of each period (among all tie matching on $\max - \min$ ratio), the algorithm maximizes total weight and thus does *not* discard beneficial edges if they do not reduce the minimum ratio. Hence,

Corollary 1. *(Inclusion of Beneficial Cycles) Any cycle that increases the total number of transplants without reducing the* $\max - \min$ *(attained from Step 1 of the mechanism) will always be included in the selected matching. Thus, the mechanism never excludes beneficial transplant opportunities that preserve fairness.*

This corollary guarantees that within the KS-feasible set, once fairness is secured, the mechanism never leaves 'free' transplants on the table. If, after Step 1.t, there exist small national-level cycles (or other disjoint transplant opportunities) that can be executed without affecting any of the cycles supporting the $\max - \min$ ratio, the algorithm's design compels their inclusion in the final matching. This follows because Step 2.t always selects a maximum-weight matching within the KS-feasible set, and all such "extra" cycles contribute strictly positive weight while leaving the fairness constraint intact. Thus, no harmless transplant opportunity is ever omitted once fairness has been secured.

Now, we will look at the individual rationality property, as defined in Definition 3. That is, each country receives at least as many transplants as it could have obtained by withdrawing from international exchange and relying solely on its own pool. The following proposition states that the allocation identified by the algorithm is individually rational. The intuition of the proof is simple. Consider the trivial matching from "disagreement option" $M^t_{\text{Dis}} \in \mathcal{F}^t$ in which *each* country just uses its domestic pool and ignores international exchange, i.e. $X^t_c = d^t_c \; \forall \, c \in C$ (This is feasible simply by selecting no cross-border edges). Among the feasible matchings, there always exists the stand-alone solution (M^t_{Dis}) in which every country performs their exchanges among its pool of donor-recipient pairs. This matching is such that for each country $c \in C$, $r^t(M^t_{\text{Dis}}) = 0$. Consider any matching M^t such that at least one country c' gets a number of transplants strictly lower than $d^t_{c'}$, then $r^t_{c'}(M^t) < 0$, but then $M^t \notin \mathcal{E}^t$.

Proposition 2 (Individual Rationality)**.** *M^{*t} satisfies Individual rationality for all $c \in C$ and in every period t.*

Intertemporal remark: The period-wise *individual rationality* immediately yields the following dynamic implication. A country's outside option in any period t is the same as it was initially: run a domestic-only exchange. But the mechanism ensures *individual rationality*, at every time period. So, if a country does not join, it misses out on any cross-border cycles. Also, because your dynamic weighting *increases* for under-served countries only when they are actually in the pool, leaving the system forfeits those future weight boosts. Hence,

Corollary 2. *(No Incentive for Mid-Game Defection): Once a country c is in the program, it has no incentive to withdraw in a future period, because $X_c^{\tau}(M^{*\tau}) \geq X_c^{\tau}(M_{Dis}^{\tau})$ for any $\tau > t$.*

Now, we look at a coalition-level property captured by Definition 4. In many multi-country exchange settings, intuitively, when more countries pool their donor–recipient pairs, the set of feasible cross-border matches either stays the same or increases, never diminishing the total number of transplants. The proposition below formalizes this *monotonicity* (or super-additivity) property, showing that the grand coalition achieves at least as many transplants as any collection of disjoint sub-coalitions could achieve on their own. Consider a set C of countries participating in an International Kidney Paired Exchange (IKPE) program. For each coalition $A_i \subseteq N$, $\mathbf{X}^{*t}(A_i)$ is the maximum total number of transplants that coalition A_i can achieve when running the previously described algorithm *exclusively* among its members (i.e. ignoring all countries not in A_i).

Proposition 3 (Monotonicity)**.** *Let A_1 and A_2 be any two disjoint subsets of the set of countries C. Then the following* monotonicity *(or super-additivity) property holds: for each $t \geq 1$,*

$$\mathbf{X}^{*t}(A_1 \cup A_2) \;\; \geq \;\; \mathbf{X}^{*t}(A_1) \; + \; \mathbf{X}^{*t}(A_2),$$

So, merging two disjoint coalitions A_1 and A_2 never reduces the set of feasible cross-border donor–recipient matches; indeed, each subset's internal matches remain feasible, and new inter-subset edges can strictly increase total matches. Consequently, no sub-coalition can do better by splitting away from the grand coalition, ensuring that full participation is weakly optimal in terms of **total transplants**. The preceding argument immediately extends to multiple disjoint coalitions, as follows:

Corollary 3. *(Super additivity) The above argument can be extended by induction to any number of disjoint coalitions. For three disjoint coalitions A_1, A_2, A_3, one can first merge A_1 and A_2 (not losing value), then merge the result with A_3, and so on. Formally, for any partition of the player set into coalitions $A_1, A_2, \ldots, A_k$ (disjoint and covering C), repeated application of the inequality gives:*

$$\mathbf{X}^{t}(A_1 \cup A_2 \cup \cdots \cup A_k) \; \geq \; \mathbf{X}^{t}(A_1) + \mathbf{X}^{t}(A_2) + \cdots + \mathbf{X}^{t}(A_k).$$

But $A_1 \cup \cdots \cup A_k = N$ (the grand coalition). Therefore, for every partition of countries,

$$\mathbf{X}^t(N) \;\geq\; \sum_{i=1}^{k} \mathbf{X}^t(A_i)\,.$$

The super-additivity property provides a scalability guarantee for the mechanism: as more countries join the exchange, the total number of transplants can only increase (or remain unchanged in the worst case). This ensures that expanding participation in the program cannot harm aggregate performance, reinforcing the appeal of the mechanism as the basis for a growing supra-national exchange network.

Incentives and Hiding Pairs: Our mechanism does not eliminate incentives for countries to strategically withhold (especially easy-to-match in this setup) pairs before submission. This limitation is fundamental: Atlamaz and Klaus (2007) establish that with additive preferences and only two agents, no allocation rule can simultaneously satisfy Pareto efficiency and immunity to hiding-type endowment manipulation.[6] Because the effect of withholding depends on the realized global compatibility graph and that graph is unknown until all countries submit, so the profitability of hiding is inherently unpredictable ex ante.[7] Full strategy-proofness would require sacrificing pareto efficiency or adding external pool verification.

4.1 Long-run Convergence: Egalitarian Solution

Now, we analyze the long-term behavior of the proposed IKPE model under dynamic weight adjustments. Given the stochastic nature of donor-recipient arrivals, we model the system using a probability space and define the arrival process as a sequence of random variables.

Our key result establishes that, under mild assumptions, the dynamic weighting mechanism ensures that no country remains persistently disadvantaged. Specifically, while short-term fluctuations may cause some countries to receive fewer transplants than others, the long-run cumulative transplant ratios of all countries converge to the same limit.

This fairness property follows from the periodic reweighting of edge priorities: whenever a country falls significantly behind, its transplant opportunities are gradually prioritized in future rounds. As a result, we prove that all countries' cumulative gains become arbitrarily close over an infinite horizon, ensuring equitable long-term outcomes.

[6] Related impossibilities appear in Sertel and Özkal-Sanver (2002), Postlewaite (1979); Ashlagi and Roth (2014) document strategic withholding of pairs by hospitals in kidney exchange.

[7] A country hiding easy pairs might (i) lower its disagreement point d_c^t, inflating its KS ratio, or (ii) remove edges that would have created cross-border cycles, reducing total matches. Which effect dominates depends on the realized graph structure.

We begin by formalizing the assumptions. Let Ω be a probability space, where each $\omega \in \Omega$ corresponds to a realization of donor-recipient arrivals across countries at a given time period, and let μ be a measure on it. In each period $t \in \mathrm{N}$, the newly arriving donor-recipient pairs (across all countries) is realized from a random variable

$$Y^t : \Omega \to \mathcal{P}$$

where $\mathcal{P}$ is the set of all possible "arrival profiles" (i.e., the configuration of new donor-recipient pairs for every country). We assume:

Assumption 1. *(i.i.d. or Stationary Distribution) The random variables $\{Y^t\}_{t\in\mathrm{N}}$ are identically distributed according to μ. Equivalently, for every Borel-measurable set $\mathcal{A} \subseteq \mathcal{P}$,*

$$\Pr[Y^t \in \mathcal{A}] = \mu(\mathcal{A}), \ \ \forall\, t.$$

Assumption 2 (Sufficient Cross-Compatibility)**.** *For each country $c \in C$ there exists a constant $\varepsilon_c > 0$ and a Borel set $E_c \subseteq \mathcal{P}$, such that whenever $Y^t(\omega) \in E_c$ for some $\omega \in \Omega$, there is at least one feasible cross-border cycle (or set of cycles) that can yield an additional transplant for country c, then*

1. ***Uniform positive probability.*** *For every period $t \in \mathbb{N}$ and every realised history H_{t-1} (i.e. the multiset of donor-recipient pairs left unmatched after period $t-1$),*

$$\Pr\big[Y^t \in E_c \,\big|\, H_{t-1}\big] \ \geq \ \varepsilon_c.$$

2. ***Existence of a fairness-preserving augmenting cycle.*** *Whenever $Y^t \in E_c$, let S_t be the compatibility graph obtained by adding Y^t to the leftover pool H_{t-1}. Then S_t contains at least one cycle C_t (or set of vertex-disjoint cycles) satisfying*

   (a) Benefit to c: *C_t provides at least one additional transplant to country c.*

   (b) Fairness-safe cycle: *Appending C_t (where C_t is vertex-disjoint from M) to any matching M with $\min_c r_c^t(M) = r^t$ yields a matching $\tilde{M}$ with positive* $\max - \min$ *ratio; hence, $\max\limits_{\tilde{M}\in\mathcal{F}^t} \min\limits_{c\in C} r_c^t(\tilde{M}) \geq 0$.*

Formally the rationale behind these assumptions are, on the subset of arrivals E_c, there exists a set of Edges $M_c \subseteq E^t$ in the global compatibility graph (formed by new plus old unmatched pairs) that would add at least one transplant for country c *without decreasing* the minimum ratio below the maximum *min*-ratio matching. Because $\mu(E_c) > 0$, such "beneficial-for-c" arrivals occur infinitely often *with positive probability* over the infinite horizon. In the presence of these assumptions, all countries' total "fraction of potential improvement" eventually become arbitrarily close. The dynamic weighting ensures countries that lag behind are gradually favoured in subsequent matchings (provided there are feasible edges for them). The strategy here is that the weights act as a mirror of cumulative shortfalls: lower past ratios r_c^t generate larger increments $(1 - r_c^t)$, raising future selection priority. Coupled with a positive probability of fairness-safe opportunities, this produces infinitely many corrective events that eliminate cumulative disparities. Formally,

Theorem 1. *(Long-Term Cumulative Ratio Convergence) Under assumptions* A1 and A2*, applying DAKSA with dynamic weight adjustment, the cumulative ratios* $\{R_c^T\}_{c\in C}$ *satisfy: for any pair of countries* c *and* c'

$$\lim_{T\to\infty} \max_{c,c'\in C} \left| R_c^T - R_{c'}^T \right| = 0, \quad \textit{almost surely,}$$

$$\textit{where,} \quad R_c^T = \frac{\sum_{t=0}^{T} (X_c^t(M^{*t}) - d_c^t)}{\sum_{t=0}^{T} (B_c^t - d_c^t)}$$

Equivalently, the ratios $\{R_c^T\}_{c\in C}$ *converge to the same limit for all countries, implying that no country remains persistently disadvantaged in the long run.*

The theorem says short-run imbalances need not persist: while r_c^t and X_c^t may fluctuate, the *cumulative* KS ratios $\{R_c^T\}$ coalesce almost surely, so no country is permanently disadvantaged. The statement concerns cumulative ratios (not per-period outcomes) under mild conditions: stationary/i.i.d. arrivals and a positive probability of fairness-safe augmenting cycles, and it accommodates carryover pools and bounded cycle lengths. Design-wise, the dynamic reweighting acts as a self-correcting equity mechanism: countries that fall behind are endogenously prioritized later, supporting long-run participation and program stability. The proof (Appendix §A.4) proceeds with a weight identity linking w_c^t to running KS ratios, a "mirror' relation comparing weights across countries, inclusion of fairness-safe cycles in any KS feasible maximum weight matching, and a probabilistic convergence argument framed via a submartingale type drift and the (conditional) second Borel-Cantelli lemma to guarantee infinitely many corrective events.

4.2 The Price of Fairness

We quantify the efficiency cost of implementing the outcome of the DAKSA algorithm in each period t by comparing the utilitarian benchmark to the KS-selected outcome. The index below measures, for each t, the relative shortfall in total transplants that is attributable to the fairness constraint, and we then bound this loss in terms of the KS ratios.

For each period $t = 1, \ldots, T$ let $X_{\max}^t = \max_{M^t\in\mathcal{F}^t} \sum_{c\in C} X_c^t(M^t)$ denote the maximum number of transplants that could be performed in period t and $X^t_{KS} = \sum_{c\in C} X^t_c(M^{*t})$, the total number of transplants in period t according to the DAKSA solution.

Definition 5. *Following Dickerson et al. (2014) we define the* price of fairness *in period t as the relative loss of total transplants incurred by imposing our fairness constraint:*

$$PoF_t = \frac{X^t_{\max} - X^t_{KS}}{X^t_{\max}}, \qquad 0 \le PoF_t \le 1. \tag{4}$$

We recall $r^t = \max_{M^t\in\mathcal{F}^t} \min_{c\in C} r^t_c(M^t)$, and let $\hat{r}^{\,t} = \min_{c\in C} r^t_c(M^{*t}) \ge r^t$.

Proposition 4 (Period t efficiency loss)**.** *For every period t*

$$PoF_t \le 1 - \frac{\sum_{c\in C} d^t_c + \hat{r}^{\,t} \sum_{c\in C}(B^t_c - d^t_c)}{\sum_{c\in C} B^t_c} \le 1 - \hat{r}^{\,t} \le 1 - r^t.$$

Proof. First we prove two simple lemmas.

Lemma 1 (Lower bound on the KS total)**.** *For every period t,*

$$X^t_{KS} \ge \sum_{c\in C}\big[d^t_c + \hat{r}^{\,t}\big(B^t_c - d^t_c\big)\big].$$

Proof. By definition of $\hat{r}^{\,t}$, $X^t_c(M^{*t}) \ge d^t_c + \hat{r}^{\,t}(B^t_c - d^t_c)$ for every c. Summing over countries yields the claim. □

Lemma 2 (Upper bound on the utilitarian total)**.** *For every period t,*

$$X^t_{\max} \le \sum_{c\in C} B^t_c. \tag{5}$$

Proof. No allocation can give a country more than its individual best B^t_c, hence the inequality holds after summing across countries. □

By Lemma 1 and Lemma 2 we get

$$X^t_{KS} \ge L \qquad \text{and,} \qquad X^t_{\max} \le U \tag{6}$$

with $L = \sum_c [d^t_c + \hat{r}^{\,t}(B^t_c - d^t_c)]$ and $U = \sum_c B^t_c$. Since $PoF_t = 1 - X^t_{KS}/X^t_{\max}$, substituting X^t_{KS} with L and $X^t_{\max}$ with U we obtain the first inequality. Since

$$\sum_c d^t_c + \hat{r}^{\,t} \sum_c (B^t_c - d^t_c) = \hat{r}^{\,t} \sum_c B^t_c + (1 - \hat{r}^{\,t}) \sum_c d^t_c \ge \hat{r}^{\,t} \sum_c B^t_c,$$

the middle inequality follows, and $\hat{r}^{\,t} \ge r^t$ gives the last inequality. □

5 Dual-Objective Fairness: Total and Hard-to-Match

From a policy point of view, it is not always enough to ensure fairness across countries solely in terms of total transplants; one must also account for the allocation of transplants to particularly disadvantaged patient groups. In kidney exchange, these are often recipients with *high Panel Reactive Antibody (PRA)* levels, who are substantially harder to match than others.[8] To address this, we extend our framework to incorporate a dual fairness objective covering both total transplants and hard-to-match transplants leading to a modified procedure we call **DAKSA-2**.

Some recipients are *hard-to-match* because they have high panel reactive antibody (PRA) levels; others are *easy-to-match* with low PRA. Following the clinical practice, we fix a threshold $\tau_H \in (0, 1)$ (e.g., $80\% \equiv 0.8$) and call a recipient *hard-to-match* if PRA $\geq \tau_H$ and *easy-to-match* otherwise.[9] This refinement is policy-relevant: it lets us evaluate equity *both* in the overall counts of transplants and in the subset serving immunologically disadvantaged patients.

Notation and benchmarks:

We assume that each country $c \in C$ has lexicographic preferences: between two matchings, country c prefers the one that yields the highest number of transplants; and among matchings that yield the same total number of transplants, c prefers the one that includes the largest number of hard-to-match transplants. Let the matching M^T be an optimal matching for country c at time t when country c can only match pairs in its own national pool. Let All definitions mirror those in the baseline model but now track both the *total* and the *hard-to-match* components from the recipient's side.

For each country c and period t, and for any matching M^t:

- $X_c^{t,T}(M^t)$: total number of recipients from c transplanted by M^t.
- $X_c^{t,H}(M^t)$: number of *hard-to-match* recipients from c transplanted by M^t (PRA $\geq \tau_H$).

The domestic (standalone) and cooperative benchmarks are defined *receiver-wise* in the same spirit as before, but with the hard-to-match counts taken from the corresponding total-optimal solutions:

$$d_c^{t,T} \;=\; \max_{M_c^t \in \mathcal{F}_c^t} X_c^{t,T}(M_c^t), \qquad d_c^{t,H} \;=\; \max\big\{\, X_c^{t,H}(M_c^t) : M_c^t \in \arg\max_{M_c^t \in \mathcal{F}_c^t} X_c^{t,T}(M_c^t) \,\big\},$$

[8] In kidney transplantation, the Panel Reactive Antibody (PRA) test is used to assess a potential recipient's level of sensitization to Human Leukocyte Antigens (HLA), which are expressed on most human cells. A higher PRA value reflects a greater proportion of HLA antibodies in the patient's blood, thereby increasing the risk of graft rejection and reducing the likelihood of identifying a compatible donor organ.

[9] Only the *recipient*'s PRA determines the type; the donor side plays no role in the classification.

$$B_c^{t,T} \;=\; \max_{M^t \in \mathcal{F}^t} X_c^{t,T}(M^t), \qquad B_c^{t,H} \;=\; \max\big\{\, X_c^{t,H}(M^t) : M^t \in \arg\max_{M^t \in \mathcal{F}^t} X_c^{t,T}(M^t) \,\big\}.$$

Thus, $d_c^{t,T}$ is the maximum total number of transplants that country c can achieve domestically, while $d_c^{t,H}$ is the maximum number of hard-to-match recipients served in a total-optimal domestic matching. Specifically among the domestic matchings that maximize the total number of transplants each country selects those that maximize the number of hard-to-match transplants that can be performed. Similarly, $B_c^{t,T}$ is the maximum total number of transplants c can achieve under full cooperation, and $B_c^{t,H}$ is the maximum induced number of hard-to-match recipients in a cooperative total-optimal matching.

Remark 3. *An alternative definition of the disagreement and best allocation benchmarks would be obtained by the dual lexicographic order: first select the set of matchings that maximize the number of hard-to-match transplants and then choose among them those that maximize the total number of transplants.*

For $K \in \{T, H\}$, define the period-t KS ratios

$$r_c^{t,K}(M^t) \;=\; \begin{cases} \dfrac{X_c^{t,K}(M^t) - d_c^{t,K}}{B_c^{t,K} - d_c^{t,K}} & \text{, if } B_c^{t,K} > d_c^{t,K}, \\ 1 & \text{, if } B_c^{t,K} = d_c^{t,K}. \end{cases}$$

For $K \in \{T, H\}$, $r_c^{t,K}(M^t) = 0$ corresponds to the domestic benchmark $d_c^{t,K}$ and $r_c^{t,K}(M^t) = 1$ corresponds to the cooperative best $B_c^{t,K}$; the ratio measures the fraction of the "gains" country c received at time period t from M^t.

Now, we adapt Individual Rationality (IR) and Pareto Efficiency (PE) from section 3 to reflect both the total and the hard-to-match dimensions. The bi-criteria analogues of Definition 2 and Definition 3 will be used to evaluate any period-t matching in this setup.

Definition 6 (Bi-criteria Individual Rationality)**.** *A matching M^t is* bi-criteria IR *if, for every $c \in C$,*

$$X_c^{t,T}(M^t) \;\geq\; d_c^{t,T} \quad \textit{and} \quad X_c^{t,H}(M^t) \;\geq\; d_c^{t,H}.$$

In other words, cooperation never hurts a country, not only in total transplants but also in the number delivered to its hard-to-match recipients.

Definition 7 (Bi-criteria Pareto Efficiency)**.** *A matching M^t is* bi-criteria PE *if there is no $\tilde{M}^t \in \mathcal{F}^t$ such that, for all $c \in C$,*

$$X_c^{t,T}(\tilde{M}^t) \;\geq\; X_c^{t,T}(M^t) \quad \textit{and} \quad X_c^{t,H}(\tilde{M}^t) \;\geq\; X_c^{t,H}(M^t),$$

with at least one strict inequality for some country.

No period-t reallocation can weakly improve *both* dimensions for *all* countries and strictly improve one country in at least one dimension.

Definition 6 guarantees that countries are never worse off in either metric when they participate; the Definition 7 rules out reallocations that appear "efficient" on totals while covertly reducing service to hard-to-match patients (or vice versa).

Algorithm DAKSA-2: bi-criteria KS screen + weighted tie-break

To accommodate the two objectives, we modify only the fairness screen (Step 1.t); the tie-break and weight update remain as in DAKSA.

- **Step 1.t ($\max - \min - \min$ KS screen).** Define the period benchmark

$$r^{*t} \;=\; \max_{M^t \in \mathcal{F}^t} \min\Big\{ \underbrace{\min_{c \in C} r_c^{t,T}(M^t)}_{\text{overall KS ratio}},\; \underbrace{\min_{c \in C} r_c^{t,H}(M^t)}_{\text{hard KS ratio}} \Big\}.$$

  Let the KS-feasible set be

$$\mathcal{E}^t \;=\; \Big\{ M^t \in \mathcal{F}^t \;:\; \min_c r_c^{t,T}(M^t) \;\geq\; r^{*t} \;\text{ and }\; \min_c r_c^{t,H}(M^t) \;\geq\; r^{*t} \Big\}.$$

- **Step 2.t (weighted tie-break).** Select

$$M^{*t} \in \arg\max_{M^t \in \mathcal{E}^t} \sum_{(v,v') \in M^t} w^t(v,v'),$$

  where $w^t(v,v') = w_c^t(v,v')$ if the recipient v' belongs to country c.

The dynamic edge weight procedure remains unchanged. Initialize $w_c^1(v,v') = 1$. For $t > 1$,

$$w_c^t(v,v') \;=\; w_c^{t-1}(v,v') + \big(1 - r_c^{t-1}(M^{*\,t-1})\big),$$

where $r_c^{t-1} := \min\{r_c^{t-1,T},\, r_c^{t-1,H}\}$ computed as in the worst of the two components. Using the worst component ratio in the weight update focuses priority on countries lagging in either objective, while the bi-criteria KS screen ensures per-period feasibility on both dimensions.

DAKSA-2 jointly detects fairness in total and hard-to-match dimensions (via the $\max - \min - \min$ criterion) and then maximizes weighted efficiency between bi-criteria feasible matches. The strengthened IR and PE ensure that participation protects countries on both metrics, while the dynamic weights continue to correct intertemporal imbalances.

Having established the modified fairness criteria and selection rule, we now summarise the procedure in pseudocode form. The following pseudocode (2) of the algorithm implements the bi-criteria $\max - \min - \min$ screening stage, followed by the same weighted tie-break and dynamic edge weight update used in the single-objective case, but with the weight increments driven by the worst of the two component ratios.

Algorithm 2 DAKSA-2: Dual-Objective

Require: Countries $C = \{1, \ldots, n\}$; horizon T; per-period graphs $G^t = (V^t, E^t)$ (with carryover if modeled); hard-type threshold τ_H; initial weights $w_c^1(v, v') \equiv 1$ for all edges whose recipient v' is in country c.

Ensure: A matching $M^{*t} \in \mathcal{F}^t$ for each t.

For all $c \in C$, set $r_c^{0,T} \leftarrow 0$, $r_c^{0,H} \leftarrow 0$, and $r_c^0 \leftarrow 0$.
for $t = 1$ **to** T **do**
  G^t given or built from carryover $\cup$ new arrivals; recipient type is hard if and only if PRA$(v') \geq \tau_H$
  for each country $c \in C$ **do**
    Compute $d_c^{t,T}$ = maximum transplant in c's standalone subgraph of G^t.
    Compute $d_c^{t,H}$ = hard-to-match recipients count among $d_c^{t,T}$
    Compute $B_c^{t,T}$ = maximum transplants for c under full cooperation ($M \in \mathcal{F}^t$)
    Compute $B_c^{t,H}$ = hard-to-match recipients count among $B_c^{t,T}$
  end for
  if $t > 1$ **then** ▷ Dynamic weight update (worst-component ratio)
    for each edge $(v \to v') \in E^t$ with recipient $v' \in c$ **do**
      $w_c^t(v, v') \leftarrow w_c^{t-1}(v, v') + \left(1 - r_c^{t-1}\right)$ where $r_c^{t-1} := \min\{r_c^{t-1,T}, r_c^{t-1,H}\}$
    end for
  end if
  Step 1.t: define

$$\rho_c^{t,K}(M) = \begin{cases} \dfrac{X_c^{t,K}(M) - d_c^{t,K}}{B_c^{t,K} - d_c^{t,K}}, & \text{if } B_c^{t,K} > d_c^{t,K}, \\ 1, & \text{if } B_c^{t,K} = d_c^{t,K}, \end{cases} \qquad K \in \{T, H\}.$$

  $r^{*t} \leftarrow \max_{M \in \mathcal{F}^t} \min\Big\{\min_{c \in C} \rho_c^{t,T}(M), \min_{c \in C} \rho_c^{t,H}(M)\Big\}$
  $\mathcal{E}^t \leftarrow \Big\{M \in \mathcal{F}^t : \min_c \rho_c^{t,T}(M) \geq r^{*t} \text{ and } \min_c \rho_c^{t,H}(M) \geq r^{*t}\Big\}$ ▷ bi-criteria KS-feasible set
  Step 2.t: $M^{*t} \in \arg\max_{M \in \mathcal{E}^t} \sum_{(v,v') \in M} w_c^t(v, v')$ ▷ $v' \in c$
  for each $c \in C$ **do** ▷ Record realized component ratios for next update
    $X_c^{t,T} \leftarrow X_c^{t,T}(M^{*t})$; $X_c^{t,H} \leftarrow X_c^{t,H}(M^{*t})$

$$r_c^{t,T} \leftarrow \begin{cases} \dfrac{X_c^{t,T} - d_c^{t,T}}{B_c^{t,T} - d_c^{t,T}}, & B_c^{t,T} > d_c^{t,T} \\ 1, & B_c^{t,T} = d_c^{t,T} \end{cases} ; \quad r_c^{t,H} \leftarrow \begin{cases} \dfrac{X_c^{t,H} - d_c^{t,H}}{B_c^{t,H} - d_c^{t,H}}, & B_c^{t,H} > d_c^{t,H} \\ 1, & B_c^{t,H} = d_c^{t,H} \end{cases}$$

    $r_c^t \leftarrow \min\{r_c^{t,T}, r_c^{t,H}\}$ *worst-component ratio for the next weight update*
  end for
end for

Per-period property: With Definition 7 and Definition 6 in hand, the period-t outcome selected by the algorithm satisfies both criteria:

Proposition 5 (Bi-criteria IR and PE)**.** *For every period t, the matching M^{*t} selected by the Algorithm 2 is (i) bi-criteria Individually Rational and (ii) bi-criteria Pareto efficient.*

Each period's outcome protects every country on both metrics (total and hard) relative to going alone, and there is no alternative matching that weakly improves both metrics for all countries while strictly improving at least one.

As in the baseline model, dynamic reweighting creates a corrective drift: countries that fall behind in either dimension receive higher priority later. Under mild arrival conditions, this yields an egalitarian limit in cumulative terms for both totals and hard-to-match recipients.

Theorem 2 (Long-Term Bi-Criteria Cumulative Ratio Convergence)**.** *Under assumptions* A1 *and* A2*, and with dynamic weights updated according to the worst-component ratio* $r_c^t = \min\{r_c^{t,T}, r_c^{t,H}\}$*, the DAKSA-2 satisfies*

$$\lim_{\mathcal{T}\to\infty} \max_{c,c'\in C}\left| R_c^{\mathcal{T},T} - R_{c'}^{\mathcal{T},T} \right| = 0 \quad \textit{almost surely,} \quad \textit{where} \quad R_c^{\mathcal{T},T} = \frac{\sum_{t=0}^{\mathcal{T}}\big(X_c^{t,T}(M^{*t}) - d_c^{t,T}\big)}{\sum_{t=0}^{\mathcal{T}}\big(B_c^{t,T} - d_c^{t,T}\big)}.$$

Moreover,

$$\lim_{\mathcal{T}\to\infty} \max_{c,c'\in C}\left| R_c^{\mathcal{T},H} - R_{c'}^{\mathcal{T},H} \right| = 0 \quad \textit{almost surely,} \quad \textit{where} \quad R_c^{\mathcal{T},H} = \frac{\sum_{t=0}^{\mathcal{T}}\big(X_c^{t,H}(M^{*t}) - d_c^{t,H}\big)}{\sum_{t=0}^{\mathcal{T}}\big(B_c^{t,H} - d_c^{t,H}\big)}.$$

Equivalently, the cumulative gain ratios converge to a common limit across all countries in both dimensions (total and hard-to-match), ensuring that no country remains persistently disadvantaged along either metric.

The statement concerns cumulative ratios $R_c^{\mathcal{T},(\cdot)}$ (not period by period outcomes $r_c^{t,(\cdot)}$) and holds under stationary/i.i.d. arrivals plus a positive probability of fairness-safe augmenting cycles that preserve the bi-criteria KS constraint.
The argument follows the same martingale-type convergence strategy as in the single-objective case (Theorem 1). A weight identity links w_c^t to the running worst-component ratios, and a mirror relation compares these across countries. Since the martingale difference sequence is now two-dimensional but uniformly bounded, we first apply the Azuma-Hoeffding inequality to control deviations, and then invoke the conditional second Borel-Cantelli lemma exactly as before to ensure infinitely many corrective events. This drift toward lagging countries drives convergence in both dimensions.

Price of Fairness (PoF) analysis in Bi-criteria Setup:

The definition for Price of Fairness (Definition 5) depends *only* on the total utilitarian optimum and the total realised count; these are unchanged by introducing the additional fairness constraint.

Corollary 4 (PoF bound under DAKSA-2)**.** *With the same* PoF_t *as in Definition 5, the bound in Proposition 4 holds verbatim*

$$PoF_t \;\leq\; 1 - \frac{\sum_c d_c^t + \hat{r}^t \sum_c (B_c^t - d_c^t)}{\sum_c B_c^t} \;\leq\; 1 - \hat{r}^t \;\leq\; 1 - r^t.$$

The intuition is identical to Proposition 4 because it only uses (i) a lower bound on X_{KS}^t via the realised min-ratio and (ii) the upper bound $X_{\max}^t \leq \sum_c B_c^t$. The extra hardlayer constraint only changes the value of the realised minimum ratio from the baseline $\hat{r}^t_{\mathrm{DAKSA}}$ to $\hat{r}^{\,t} = \min\{\min_c r^t_c, \min_c r^{t,H}_c\}$; the algebra of the bound remains identical and yields $PoF_t \le 1 - \hat{r}^{\,t}$. If one also tracks the hard-only price of fairness:

$$PoF^H_t = \frac{\sum_c B^{t,H}_c - \sum_c X^{t,H}_c(M^{*t})}{\sum_c B^{t,H}_c},$$

the same derivation using $B^{t,H}_c, d^{t,H}_c, X^{t,H}_c(M)$ yields $PoF^H_t \le 1 - \hat{r}^{\,t}$. Since the two-tier benchmark is obtained by solving the same optimisation problem as the one-tier benchmark but with additional constraints, its feasible set is a subset of the one-tier feasible set. As a result, the maximum attainable fairness ratio under the two-tier benchmark cannot be higher than under the one-tier benchmark. Nevertheless, our bound keeps the same form $PoF_t \le 1 - r^{\star\,t}$ (now with the two-tier $r^{\star t}$), hence the qualitative statement from earlier section "PoF is controlled by the slack $1 - r$", is preserved. Let $\Delta^t_c := \max\{1 - r^t_c(M^{*t}),\, 1 - r^{t,H}_c(M^{*t})\} \in [0, 1]$ be country c's composite shortfall at t. Updating recipient-side weights by

$$w^{t+1}_c(v, v') = w^t_c(v, v') + \Delta^t_c$$

strictly increases the relative priority of any country that underachieved in *either* metric. As in the baseline analysis, this monotone rule ensures the realised minimum ratio $\hat{r}^{\,t}$ is non-decreasing over time and bounded by 1; under the same mild regularity on arrivals and feasibility used in the one-tier case, $\hat{r}^{\,t} \to 1$. By Corollary 4, $PoF_t \le 1 - \hat{r}^{\,t} \to 0$. Thus the two-tier rule inherits (a) the same per-period functional bound and (b) the same long-run vanishing PoF guarantee. Because $PoF^H_t \le 1 - \hat{r}^{\,t}$ as well, the worst-case efficiency loss is controlled uniformly for the hard subpopulation and for the overall population, matching the baseline qualitative message.

6 Simulation Results

In this section we show simulations that lie outside the formal model. They serve to display the theoretical properties and to provide intuition on the speed and robustness of convergence in practice.

Single-Objective Simulation Design (DAKSA): To evaluate the dynamic performance of DAKSA, we create a synthetic international exchange environment that mirrors the stylised facts outlined in section 3. For every country the generator produces a list of incompatible donor-recipient pairs together with the ABO blood types of both donor and patient as well as PRA details.[10]

[10] All experiments are implemented in a Jupyter notebook and rely on the open-source kidney-exchange instance generator developed by Delorme et al. (2022).

Each simulation covers a horizon of $T = 100$ periods. At $t = 1$, a total of 150 incompatible donor-recipient pairs are randomly distributed across the three countries so as to create an ex-ante imbalance in pool sizes. In every subsequent period, exactly 150 *new* incompatible pairs arrive across the three countries. The data generator draws the blood type of each donor's and each patient's independently according to the empirical worldwide frequency vector.[11] Unmatched pairs remain in the pool and, therefore, carry over to the next period until they are transplanted.

At each period, we apply the two-step DAKSA procedure described in section 3: we first maximise the minimum relative gain r_c^t across countries (step 1) and then select, among KS-feasible matchings, a maximum-weight matching (step 2). Edge weights start at $w_c^1(v, v') = 1$ and are updated recursively through

$$w_c^{t+1}(v, v') = w_c^t(v, v') + \big(1 - r_c^t(M^t)\big),$$

so that countries that perform poorly in one round obtain higher priority in subsequent rounds. Consistent with clinical practice and the prevailing optimisation literature, feasible matchings are restricted to vertex-disjoint cycles of length at most three (no chains).

Dual-Objective Simulation Design (DAKSA-2): We replicate the same international environment for the dual-objective setting, but with a smaller, stress-test inflow: at each period $t = 1, \ldots, 100$, exactly 60 *new* incompatible pairs arrive and are randomly distributed across the three countries. To model the "hard-to-match" dimension, we classify a fixed fraction of newly arriving recipients in each period as highly sensitised (HS), defined by PRA (or cPRA) $\geq 80\%$; in the baseline we set this fraction to 30% of new arrivals, consistent with upper-range reported prevalence figures for highly sensitised patients.[12] As before, unmatched pairs carry over, feasible matchings are restricted to vertex-disjoint cycles of length at most three (no chains), and we run the algorithm period-by-period. The only change is the fairness screen: DAKSA-2 enforces a bi-criteria $\max - \min - \min$ constraint on *both* the total and HS counts, and among the bi-criteria-feasible matchings the algorithm selects a maximum-weight solution.

6.1 National gains and participation incentives

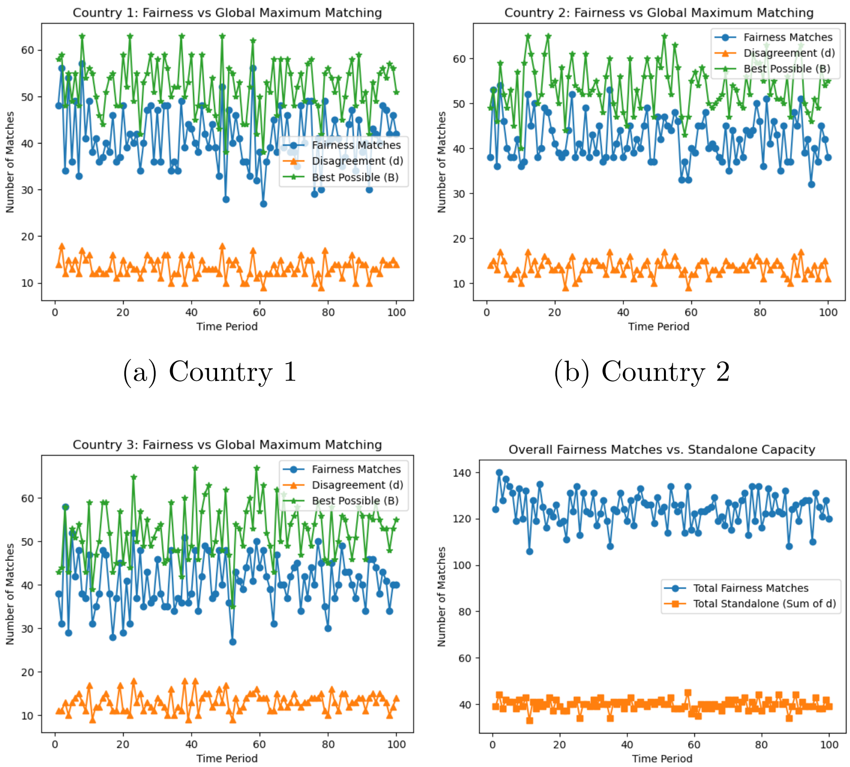
Figure 1a-1c (Green is B_c^t, Orange is d_c^t and Blue is the DAKSA output X_c^t) report period-by-period transplant counts for each country under **DAKSA**. For each country, the trajectory rises steadily, indicating that all participating countries achieve strictly

The data generator is available at `https://wpettersson.github.io/kidney-webapp/`.

[11] $(O, A, B, AB) = (0.47, 0.41, 0.09, 0.03)$

[12] See, for example, Mamode et al. (2022); Hart et al. (2021); Erdogmus et al. (2022) on the use of a cPRA$\geq$ 80% threshold and the prevalence of highly sensitised recipients. Higher values approaching 30% are reported in some specific cohorts and programmes (See for example, here).

positive gains relative to their autarky benchmark. At the aggregate level, Figure 1d (Blue is $\sum_{c\in C} X_c^t$ and Orange is $\sum_{c\in C} d_c^t$) shows that total transplants increase sharply over time. Thus, the fairness screen embedded in DAKSA does not reduce efficiency; rather, it accompanies a substantial improvement in overall welfare, consistent with the mechanism's theoretical properties.[13]



(a) Country 1 (b) Country 2

(c) Country 3 (d) Total Efficiency gain

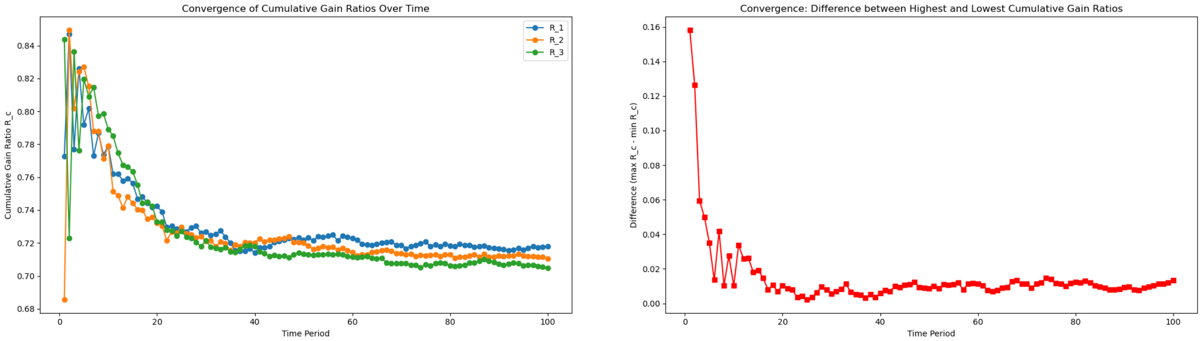
Figure 1: Match plots and Efficiency gain across countries.

6.2 Cumulative Convergence

Under **DAKSA**, Figure 2a reports the cumulative gain ratio R_c^t for each country. Although the three lines do not coincide point-wise, they draw steadily closer as time progresses. To measure this convergence more sharply, Figure 2b shows in each period the maximum vertical distance between any two of the R_c^t curves. The gap narrows monotonically during the first forty periods and remains below 0.02 thereafter, indicating that the mechanism quickly limits cross-country disparities. Under **DAKSA-2** the cumulative gain ratios for both total ($R_c^{t,T}$) and hard-to-match ($R_c^{t,H}$) recipients (Appendix subsubsection A.7.2) display the same qualitative pattern as under DAKSA: the country-level

[13] The detailed country-level plots for **DAKSA-2** are reported in Appendix subsubsection A.7.1; the patterns are qualitatively identical to the single-objective case. The qualitative patterns closely parallel those under DAKSA: all countries obtain sustained improvements relative to autarky, and the hard-to-match subgroup benefits from explicit protection.

trajectories steadily converge, and no country remains persistently disadvantaged along either dimension.



(a) Cumulative Gain of the countries at each time period

(b) Maximum difference between any two points at each time period.

Figure 2: Plots of Cumulative Gain (R_c^T) under DAKSA

Robustness across replications: To assess sensitivity to stochastic variation in arrivals, we repeat the 100-period experiment across multiple independent arrival streams. In the single-objective case we run 45 independent replications, and in the dual-objective case 50 replications, each with a fresh sequence of arrivals drawn from the same distribution. All reported curves in this section are averages across these replications, which stabilises the period-by-period statistics and allows meaningful inspection of convergence patterns. The resulting plot Figure 3 shows that, in every replication, the maximum gap in cumulative gain ratios contracts rapidly (to the order of 0.01 by 100^{th} time period here), consistent with the asymptotic equity guarantee in Theorem 1. The corresponding DAKSA-2 robustness results are reported in Appendix subsubsection A.7.3 and display the same pattern for both total and hard-to-match recipients.[14]

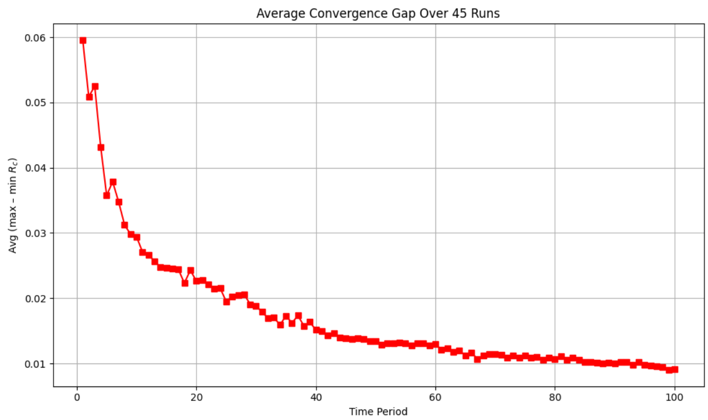
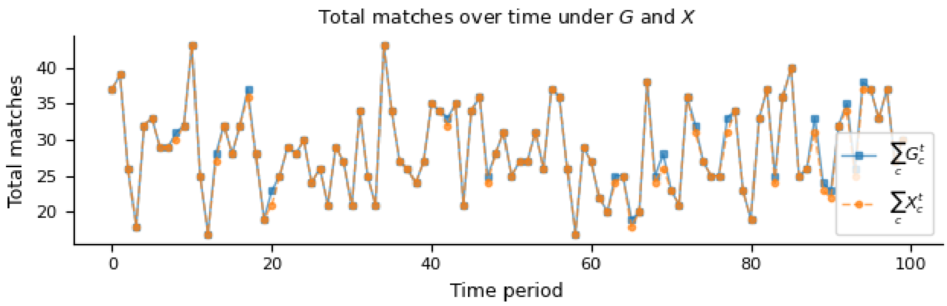


Figure 3: Average cumulative gain across 45 simulations. The convergence is consistent with Theorem 1.

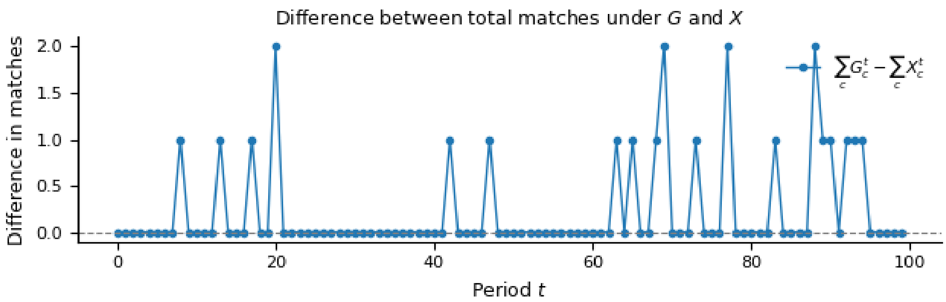
[14]We also simulate in Appendix subsubsection A.7.4 a similar environment but with unequal populations of donor-recipient pairs (approximately in the proportions 0.50, 0.35, and 0.15 among Country 1, Country 2, and Country 3, respectively).

6.3 Price of Fairness

To quantify the efficiency cost of enforcing per-period equity, we compare DAKSA with an efficiency-first benchmark that maximises total number of transplants in each time period. As shown in Figure 4, the two trajectories are almost indistinguishable: in the vast majority of periods DAKSA attains the *same* total number of transplants as the maximum-cardinality solution. When differences occur, they are typically of size one, and never exceed two, over the 100-period horizon. Aggregating across all periods[15], the total number of transplants under DAKSA falls short of the unconstrained benchmark by 23 matches in this specific case, or roughly 0.76% reduction from the global max across 100 time periods.[16]



(a) PoF - Total Max vs DAKSA



(b) PoF - Difference of Total Max and DAKSA

Figure 4: Comparison of PoF metrics using Total Max and DAKSA

7 Conclusion

This paper developed a framework for International Kidney Paired Exchange that integrates a Kalai-Smorodinsky fairness criterion with dynamic, feedback-driven edge weights. In its baseline form (DAKSA), the algorithm guarantees per-period individual rationality

[15] Total number of transplants is 2876 for Total max and 2853 for DAKSA.

[16] Due to local computational constraints, we performed this particular analysis with a reduced number of donor-recipient pairs: 60 incompatible pairs join the pool at each time period. All other settings remain the same.

and Pareto efficiency, while super-additivity ensures scalability as more countries participate. We further established a long-run egalitarian property: cumulative gain ratios converge across participants, preventing any country from being persistently disadvantaged. A bound on the Price of Fairness characterizes the efficiency cost of these guarantees.

We then extended the framework to a dual-objective version (DAKSA-2) that simultaneously protects overall transplant numbers and access for hard-to-match recipients. This extension preserves per-period fairness and efficiency while ensuring long-run convergence in both dimensions. Simulation experiments with realistic arrival processes confirm the theoretical predictions: national gains increase, disparities shrink over time, hard-to-match patients benefit from explicit protection, and efficiency losses remain small and stable. From a design perspective, the mechanism also ensures that once fairness constraints are satisfied, all additional beneficial (e.g., national) cycles are incorporated, while dynamic weights create a self-correcting priority system for lagging countries.

Future research can extend this framework along several dimensions. Strategic considerations such as entry and exit incentives or selective reporting of pairs remain unexplored. Alternative fairness objectives, including severity-weighted or pediatric priorities, could be embedded within the bargaining framework. Extending the model to incorporate chains and refining computational methods for large-scale deployment are further promising directions. Together, these avenues will deepen our understanding of how fairness and efficiency can be jointly sustained in large-scale international kidney exchange programs.

References

Agarwal, N., Ashlagi, I., Azevedo, E., Featherstone, C. R., and Karaduman, Ö. (2019). Market failure in kidney exchange. *American Economic Review*, 109(11):4026–4070.

Aliprantis, C. D. and Border, K. C. (2006). *Infinite dimensional analysis: a hitchhiker's guide*. Springer.

Andersson, T. and Kratz, J. (2020). Pairwise kidney exchange over the blood group barrier. *The Review of Economic Studies*, 87(3):1091–1133.

Ashlagi, I., Fischer, F., Kash, I. A., and Procaccia, A. D. (2015). Mix and match: A strategyproof mechanism for multi-hospital kidney exchange. *Games and Economic Behavior*, 91:284–296.

Ashlagi, I. and Roth, A. E. (2012). New challenges in multihospital kidney exchange. *American Economic Review*, 102(3):354–359.

Ashlagi, I. and Roth, A. E. (2014). Free riding and participation in large scale, multi-hospital kidney exchange. *Theoretical Economics*, 9(3):817–863.

Atlamaz, M. and Klaus, B. (2007). Manipulation via endowments in exchange markets with indivisible goods. *Social Choice and Welfare*, 28(1):1–18.

Azuma, K. (1967). Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal, Second Series*, 19(3):357–367.

Benedek, M., Biró, P., Kern, W., Pálvölgyi, D., and Paulusma, D. (2025). Partitioned matching games for international kidney exchange. *Mathematical Programming*, pages 1–36.

Benedek, M., Biró, P., Paulusma, D., and Ye, X. (2024). Computing balanced solutions for large international kidney exchange schemes. *Autonomous Agents and Multi-Agent Systems*, 38(1):1–41.

Biró, P., Van de Klundert, J., Manlove, D., Pettersson, W., Andersson, T., Burnapp, L., Chromy, P., Delgado, P., Dworczak, P., Haase, B., et al. (2021). Modelling and optimisation in european kidney exchange programmes. *European Journal of Operational Research*, 291(2):447–456.

Böhmig, G. A., Fronek, J., Slavcev, A., Fischer, G. F., Berlakovich, G., and Viklicky, O. (2017). Czech-austrian kidney paired donation: first european cross-border living donor kidney exchange. *Transplant International*, 30(6):638–639.

Carvalho, M. and Lodi, A. (2023). A theoretical and computational equilibria analysis of a multi-player kidney exchange program. *European Journal of Operational Research*, 305(1):373–385.

Carvalho, M., Lodi, A., Pedroso, J. P., and Viana, A. (2017). Nash equilibria in the two-player kidney exchange game. *Mathematical Programming*, 161:389–417.

Delorme, M., García, S., Gondzio, J., Kalcsics, J., Manlove, D., Pettersson, W., and Trimble, J. (2022). Improved instance generation for kidney exchange programmes. *Computers & Operations Research*, 141:105707.

Dickerson, J. P., Procaccia, A. D., and Sandholm, T. (2014). Price of fairness in kidney exchange. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 1013–1020.

Druzsin, K., Biró, P., Klimentova, X., and Fleiner, R. (2024). Performance evaluation of national and international kidney exchange programmes with the enckep simulator. *Central European Journal of Operations Research*, pages 1–21.

Erdogmus, S., Çelebi, Z. K., Turgut, D., Sayin, B., Ozdemir, F. N., Çolak, T., and Haberal, M. (2022). Sensitization status of patients on the deceased donor kidney transplant waiting list: A single-center experience. *Turkish Journal Nephrology*.

Francisco, J. T., Carvalho, R., Freitas, J., Coimbra, M. T., Vilela, S., Almeida, M., Tafulo, S., Pereira, R. C., Bolotinha, C., Ivo, M., et al. (2024). International kidney paired donation–the experience of a single center. *Brazilian Journal of Transplantation*, 26:e3423.

Freedman, D. A. (1975). On tail probabilities for martingales. *the Annals of Probability*, pages 100–118.

Hajaj, C., Dickerson, J., Hassidim, A., Sandholm, T., and Sarne, D. (2015). Strategy-proof and efficient kidney exchange using a credit mechanism. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29.

Hart, A., Lentine, K., Smith, J., Miller, J., Skeans, M., Prentice, M., Robinson, A., Foutz, J., Booker, S., Israni, A., et al. (2021). Optn/srtr 2019 annual data report: kidney. *American Journal of Transplantation*, 21:21–137.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30.

Kalai, E. (2008). Games in coalitional form. In *The New Palgrave Dictionary of Economics*, pages 1–13. Springer.

Kalai, E. and Smorodinsky, M. (1975). Other solutions to nash's bargaining problem. *Econometrica*, 43(3):513–518.

Klimentova, X., Viana, A., and Ševčovič, D. (2016). Fairness in static and dynamic kidney exchange programs. *Computers & Operations Research*, 73:1–16.

Lahiri, S. (2003). Axiomatic characterization of the nash and kalai-smorodinsky solutions for discrete bargaining problems. *Pure Mathematics and Applications*, 14(3):207–220.

Mamode, N., Bestard, O., Claas, F., Furian, L., Griffin, S., Legendre, C., Pengel, L., and Naesens, M. (2022). European guideline for the management of kidney transplant patients with hla antibodies: by the european society for organ transplantation working group. *Transplant International*, 35:10511.

Matyasi, L. and Biró, P. (2024). Testing re-optimisation strategies in international kidney exchange programmes by the enckep simulator. *Central European Journal of Operations Research*, 32(4):989–1011.

Mincu, R.-S., Biró, P., Gyetvai, M., Popa, A., and Verma, U. (2021). Ip solutions for international kidney exchange programmes. *Central European Journal of Operations Research*, 29:403–423.

Postlewaite, A. (1979). Manipulation via endowments. *The review of economic studies*, 46(2):255–262.

Roth, A., Sönmez, T., and Unver, U. (2004). Kidney exchange. *The Quarterly Journal of Economics*, 119(2):457–488.

Roth, A. E., Sönmez, T., and Ünver, M. U. (2005). Pairwise kidney exchange. *Journal of Economic Theory*, 125(2):151–188.

Sertel, M. R. and Özkal-Sanver, I. (2002). Manipulability of the men-(women-) optimal matching rule via endowments. *Mathematical Social Sciences*, 44(1):65–83.

Sun, Z., Todo, T., and Walsh, T. (2021). Fair pairwise exchange among groups. In Zhou, Z.-H., editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 419–425. International Joint Conferences on Artificial Intelligence Organization. Main Track.

Toulis, P. and Parkes, D. C. (2015). Design and analysis of multi-hospital kidney exchange mechanisms using random graphs. *Games and Economic Behavior*, 91:360–382.

A Appendix

A.1 Proof of Proposition 1

Proof. We prove that M^{*t} is Pareto efficient by contradiction. Suppose M^{*t} is not Pareto efficient. Then, there exists some $\tilde{M}^t \in \mathcal{F}^t$ such that $\tilde{M}^t$ is a Pareto improvement over M^{*t}. We show that this contradicts the two-step selection of M^{*t}.

Since $\tilde{X}_c^t \geq X_c^t$for all $c \in C$, we have

$$r_c^t(\tilde{M}^t) \;=\; \frac{\tilde{X}_c^t - d_c^t}{B_c^t - d_c^t} \;\geq\; \frac{X_c^t - d_c^t}{B_c^t - d_c^t} \;=\; r_c^t(M^t).$$

Since $\tilde{X}_{c'}^t > X_{c'}^t$ for at least one $c' \in C$, that country's ratio strictly increases, $r_{c'}^t(\tilde{M}^t) > r_{c'}^t(M^t)$. Thus

$$\min_{c \in C} r_c^t(\tilde{M}^t) \;\geq\; \min_{c \in C} r_c^t(M^t), \;\; \text{with at least one } c' \text{ having } r_{c'}^t(\tilde{M}^t) > r_{c'}^t(M^t).$$

Hence either

$$\min_{c \in C} r_c^t(\tilde{M}^t) > \min_{c \in C} r_c^t(M^t) \quad \text{or} \quad \min_{c \in C} r_c^t(\tilde{M}^t) = \min_{c \in C} r_c^t(M^{*t}) \text{ but with at least one ratio strictly larger.}$$

Hence the comparison of minimum ratios yields two possibilities:

- **Case A:** $\min_{c \in C} r_c^t(\tilde{M}^t) > \min_{c \in C} r_c^t(M^{*t})$

  This contradicts **Step 1.t** of the algorithm: M^t was chosen to *maximize* $\min_{c \in C} r_c^t(M^t)$. So there can be no $\tilde{M}^t \in \mathcal{F}^t$ that yields a strictly bigger min-ratio.

- **Case B:** $\min_{c \in C} r_c^t(\tilde{M}^t) = \min_{c \in C} r_c^t(M^{*t})$

  In this case, $\tilde{M}^t \in \mathcal{E}^t$. However, by assumption, $\tilde{M}^t$ guarantees a strictly higher number of transplants to at least one country c' and *does not reduce* transplants for any other country. Therefore it follows:

  $$\sum_{(v,v') \in \tilde{M}^t} w_t(v, v') \;>\; \sum_{(v,v') \in M^{*t}} w_t(v, v').$$

  which contradicts **Step 2.t** of the procedure.

Either way, $\tilde{M}^t$ cannot exist. We have shown that any matching $\tilde{M}^t$ that tries to Pareto-improve upon M^{*t} leads to a contradiction with the model's two-step procedure. Hence M^{*t} **cannot** be Pareto dominated by any other matching.

□

A.2 Proof of Proposition 2

Proof. Suppose, for contradiction, that the matching M^{*t} is not individually rational. Then there exists a country c' with

$$X^t_{c'} \; < \; d^t_{c'}$$

Hence

$$r^t_{c'}(M^{*t}) < 0$$

because the numerator is negative and the denominator $(B^t_{c'} - d^t_{c'})$ is non negative by assumption.[17] Consequently,

$$\min_{c \in C} \; r^t_c(M^{*t}) \; \leq \; r^t_{c'}(M^{*t}) \; < \; 0.$$

For that matching, since for all $c \in C$, $r^t_c(M^t_{\text{Dis}}) = 0$, then $r^t(M^t_{\text{Dis}}) \; = \; 0$.
It follows that

$$\min_{c \in C} r^t_c(M^{*t}) < 0 = \min_{c \in C} r^t_c(M^t_{\text{Dis}}).$$

But this contradicts the fact that M^{*t} maximizes $\min_{c \in C} r^t_c$ in **Step 1.t** of the algorithm.

□

A.3 Proof of Proposition 3

Proof. For any coalition $A \subseteq C$ let

$$r^t_A \; := \; \max_{M \in \mathcal{F}^t(A)} \; \min_{c \in A} r^t_c(M)$$

be the Kalai-Smorodinsky benchmark ratio when only the countries in A cooperate, and let

$$\mathcal{E}^t_A := \left\{ M \in \mathcal{F}^t(A) \; : \; \min_{c \in A} r^t_c(M) \geq r^t_A \right\}$$

be the set selected after *Step 1.t of the algorithm.* For any matching M, $W(M) := \sum_{e \in M} w_t(e)$ for its total weight in period t. Because $\mathcal{F}^t(A_1)$ and $\mathcal{F}^t(A_2)$ are both subsets of $\mathcal{F}^t(A_1 \cup A_2)$, maximising the minimum ratio over a *larger* index set can only reduce (or leave unchanged) the optimum. Hence

$$r^t_{A_1 \cup A_2} \; \leq \; \min\{ \, r^t_{A_1}, r^t_{A_2} \}. \tag{7}$$

Let $M_1 := M^{*t}(A_1)$ and $M_2 := M^{*t}(A_2)$. Because A_1 and A_2 share no patient-donor pairs, the edge sets M_1 and M_2 are vertex-disjoint, so their union $M_{12} := M_1 \cup M_2$ is a

[17] This necessarily concerns a country with $B^t_{c'} > d^t_{c'}$, since when $B^t_c = d^t_c$, we set $r^t_c = 1$.

matching in $\mathcal{F}^t(A_1 \cup A_2)$. For every country in A_1 the ratio achieved under M_{12} equals the ratio it already enjoyed in M_1, and similarly for A_2. Therefore

$$\min_{c \in A_1 \cup A_2} r_c^t(M_{12}) \;=\; \min\{\, r_{A_1}^t, r_{A_2}^t \} \;\overset{(7)}{\geq}\; r_{A_1 \cup A_2}^t .$$

Consequently $M_{12} \in \mathcal{E}_{A_1 \cup A_2}^t$ it passes the Step 1.t fairness screen when the two coalitions merge. Now, every transplant edge carries a strictly positive weight (initialised at 1 and never decreased), so

$$W(M_{12}) \;=\; \sum_{e \in M_{12}} w_t(e) \;\geq\; |M_{12}| \;=\; \mathbf{X}^{*t}(A_1) + \mathbf{X}^{*t}(A_2).$$

Step 2.t of the algorithm, run on $A_1 \cup A_2$, chooses

$$M^{*t}(A_1 \cup A_2) \;=\; \arg \max_{M \in \mathcal{E}_{A_1 \cup A_2}^t} W(M),$$

so it attains *at least* the total weight of M_{12} and therefore at least as many transplants:[18]

$$\mathbf{X}^{*t}(A_1 \cup A_2) = \sum_{e \in M^{*t}(A_1 \cup A_2)} 1 \;\;\geq\;\; |M_{12}| = \mathbf{X}^{*t}(A_1) + \mathbf{X}^{*t}(A_2).$$

□

A.4 Proof of Theorem 1

Proof. Let C be the finite set of countries, $|C| = n \geq 2$. For each period $t \in \mathbb{N}$:

- $X_c^t(M)$ - transplants country c receives under matching M.
- d_c^t - disagreement (stand-alone) number of transplants.
- B_c^t - period upper bound when all countries cooperate.
- $r_c^t(M) := \dfrac{X_c^t(M) - d_c^t}{B_c^t - d_c^t}$ - *period KS-ratio* of c.
- M^{*t} - matching chosen by DAKSA.
- $w_c^t(e)$ - weight attached to edge e whose recipient is in c.

Weights evolve according to

$$w_c^1(e) = 1, \qquad w_c^{t+1}(e) = w_c^t(e) + \big(1 - r_c^t(M^{*t})\big). \tag{8}$$

[18]Summing the constant 1 over the edge set simply counts the edges, so $\sum_{e \in M} 1 = |M|$. Because each transplant corresponds to one edge, the cardinality $|M|$ equals the total number of transplants.

Define the *cumulative KS-ratio*

$$R_c^T := \frac{\sum_{t=1}^T \big(X_c^t(M^{*t}) - d_c^t\big)}{\sum_{t=1}^T (B_c^t - d_c^t)}, \qquad T \geq 1.$$

Denominators are positive and uniformly bounded above by $D := \max_{c,t}(B_c^t - d_c^t) < \infty$.

Lemma 3 (Closed-form weight identity). *For every country c and period $t \geq 0$, the total weight on any edge incident to c satisfies*

$$w_c^{t+1} = (t+1) - \sum_{s=1}^{t} r_c^s(M^{*s}). \tag{9}$$

Proof. Induction on t. Base $t = 0$ is $w_c^1 = 1 = (0+1) - 0$. If the formula holds at t, then by the update rule (3)

$$w_c^{t+2} = \Big[(t+1) - \sum_{s=1}^{t} r_c^s\Big] + \big(1 - r_c^{t+1}\big) = (t+2) - \sum_{s=1}^{t+1} r_c^s.$$

□

Lemma 4 (Weight-ratio mirror). *For any $c, c' \in C$ and $t \geq 1$*

$$w_{c'}^t - w_c^t = \sum_{s=1}^{t-1} \big(r_c^s - r_{c'}^s\big). \tag{10}$$

Consequently, if $R_c^{t-1} \geq R_{c'}^{t-1} + \frac{1}{D}$ then $w_c^t \leq w_{c'}^t$.

Proof. Subtract the identities of Lemma 3 for c and c' and drop the $(t+1)$ term. For the implication, note that $1/D$ upper-bounds the denominators in $R_\cdot^{t-1}$. The detailed derivations are as follows:

Lemma 3 tells us that for every country k and each $t \geq 1$

$$w_k^t \;=\; t - \sum_{s=1}^{t-1} r_k^s(M^{*s}). \tag{11}$$

Recall that the original identity is $w_k^{t+1} = (t+1) - \sum_{s=1}^{t} r_k^s$; re-indexing by replacing t with $t-1$ gives (Equation 11). Now subtracting the weights for c' and c

$$\begin{aligned} w_{c'}^t - w_c^t &= \Big[t - \sum_{s=1}^{t-1} r_{c'}^s\Big] \;-\; \Big[t - \sum_{s=1}^{t-1} r_c^s\Big] \\ &= \sum_{s=1}^{t-1} \big(r_c^s - r_{c'}^s\big), \end{aligned} \tag{12}$$

which is exactly the identity Equation 12 claimed in the lemma. From a $1/D$ ratio lead to a weight ordering: Write each cumulative KS-ratio at time $t-1$ as

$$R_k^{t-1} = \frac{N_k^{t-1}}{D_k^{t-1}}, \quad N_k^{t-1} := \sum_{s=1}^{t-1}(X_k^s - d_k^s), \; D_k^{t-1} := \sum_{s=1}^{t-1}(B_k^s - d_k^s). \tag{13}$$

Because every term $B_k^s - d_k^s$ is bounded above by D, $D_k^{t-1} \leq (t-1)D$ for all k. Now assume the stated gap in cumulative ratios:

$$R_c^{t-1} \geq R_{c'}^{t-1} + \frac{1}{D}.$$

Multiply by $(t-1)D$ (a common upper bound on the denominators):

$$(t-1)D\,R_c^{t-1} \geq (t-1)D\,R_{c'}^{t-1} + 1.$$

Since $D_k^{t-1} \leq (t-1)D$, this implies that

$$N_c^{t-1} \geq N_{c'}^{t-1} + 1. \tag{14}$$

But

$$N_k^{t-1} = \sum_{s=1}^{t-1} r_k^s(B_k^s - d_k^s) \leq D\sum_{s=1}^{t-1} r_k^s, \tag{15}$$

so dividing Equation 14 by D yields

$$\sum_{s=1}^{t-1} r_c^s \geq \sum_{s=1}^{t-1} r_{c'}^s + \frac{1}{D}. \tag{16}$$

Now, plugging Equation 16 into the weight difference: Inserting Equation 16 into identity Equation 12:

$$w_{c'}^t - w_c^t = \sum_{s=1}^{t-1}(r_c^s - r_{c'}^s) \geq \frac{1}{D} > 0.$$

Hence, $w_c^t \leq w_{c'}^t$, proving the second statement of the lemma. □

Lemma 5 (Execution of helpful cycles)**.** *Fix t and let c be a country with R_c^t minimal among all countries. If a cycle C_t satisfying Assumption 2(b) exists, then C_t is included in* every *maximum-weight matching that meets the period-t KS constraint.*

Proof. c is laggard, which means c has the lowest R_c^t. Because c is the laggard, $R_c^t \leq R_{c'}^t$ for all $c' \neq c$. By Lemma 4 and the $1/D$ bound there is T such that for all $t \geq T$ we also have $w_c^t \geq w_{c'}^t$ for every $c' \neq c$. Adding C_t: (i) keeps the min KS-ratio feasible, and (ii) adds strictly positive weight (at least one edge of weight w_c^t). Therefore any feasible matching omitting C_t has strictly lower total weight, so Step 2.t selects a matching containing C_t. □

Lemma 6 (Infinitely many corrective cycles). *For each country c, the event $A_t(c) := \{c$ is the laggard at t and a cycle C_t as in Lemma 5 exists$\}$ occurs infinitely often almost surely.*

Proof. When c is the laggard, Assumption 2(a)-(b) imply $\Pr\left[A_t(c) \mid H_{t-1}\right] \geq \theta_c$. The conditional Borel-Cantelli lemma (Freedman (1975)) then guarantees $A_t(c)$ i.o. □

Proposition 6 (Two-country case). *If $|C| = 2$ then $|R_1^T - R_2^T| \xrightarrow{a.s.} 0$ as $T \to \infty$.*

Proof. Let $\Delta_T := R_1^T - R_2^T$ and $\delta := 1/D > 0$. Whenever $A_t(1)$ (resp. $A_t(2)$) occurs, Lemma 5 ensures that the algorithm gives country 1 (resp. 2) at least one extra transplant, so $|\Delta_t|$ drops by at least δ. Lemma 6 guarantees such drops happen infinitely often, hence $|\Delta_T| \to 0$. □

Now, extending the same intuition to n-countries: we want to show that, under Assumption 1 and Assumption 2,

$$\max_{c,c' \in C} |R_c^T - R_{c'}^T| \xrightarrow{a.s.} 0.$$

For this, let's define the potential (ratio spread) $\Phi_T := \max_c R_c^T - \min_c R_c^T \geq 0$. Let $c_{\min}(T)$ be a country attaining the minimum at time T. Whenever $A_t\big(c_{\min}(t)\big)$ occurs, Lemmas 5-6 imply $R_{c_{\min}(t)}$ jumps up by at least $\delta = 1/D$, so Φ_t decreases by at least δ. Because each country is the minimum infinitely often and each of such periods sees $A_t(c)$ with conditional probability $\theta_c > 0$, the conditional Borel-Cantelli lemma ensures Φ_t is reduced by δ infinitely many times. Being bounded below by 0, Φ_t must converge to 0 almost surely.

Equation 9 shows that *weights are equal to a common additive trend minus the running sum of KS ratios*. If a country lags in ratios, its weights automatically increase relative to countries with higher cumulative gains. Assumption 2 guarantees that helpful compatibility opportunities arrive often enough and Lemma 5 ensures that the algorithm always takes advantage of them once the weight reversal occurs. Hence, the share of the country with the lowest cumulative gain is absorbed and the maximum-min gap collapses.

□

A.5 Proof of Proposition 5

Proof. The proof follows similar strategy like the other proofs of Pareto Efficiency and Individual Rationality.

(IR): The disagreement matching in which each country c uses only its domestic pool attains $X_c^{t,T} = d_c^{t,T}$ and $X_c^{t,H} = d_c^{t,H}$ for all c, hence $\min_c r_c^{t,T} = 0$ and $\min_c r_c^{t,H} = 0$ are feasible. Therefore $r^{*t} \geq 0$, and by Step 1.t every $M \in \mathcal{E}^t$ satisfies $\min_c r_c^{t,K}(M) \geq r^{*t} \geq 0$ for $K \in \{T, H\}$. Unpacking the ratios yields $X_c^{t,K}(M) \geq d_c^{t,K}$ for all c and K, in particular for M^{*t}.

(PE): Suppose, toward a contradiction, that there exists $\tilde{M}^t \in \mathcal{F}^t$ such that for all c, $X_c^{t,T}(\tilde{M}^t) \geq X_c^{t,T}(M^{*t})$ and $X_c^{t,H}(\tilde{M}^t) \geq X_c^{t,H}(M^{*t})$, with at least one strict inequality in one component for some country. Then for $K \in \{T, H\}$ we have $r_c^{t,K}(\tilde{M}^t) \geq r_c^{t,K}(M^{*t})$ for all c, and strict inequality for some c in at least one K. Hence, either the minimum ratio in some component strictly increases (contradicting Step 1.t), or both component wise minima are equal to r^{*t} and $\tilde{M}^t \in \mathcal{E}^t$ while (because all edge weights are strictly positive and $\sum_c X_c^{t,T}$ equals the number of transplant edges) $\sum_{e \in \tilde{M}^t} w^t(e) > \sum_{e \in M^{*t}} w^t(e)$, contradicting the Step 2.t tie-break. Thus no such $\tilde{M}^t$ exists. □

A.6 Proof of Theorem 2

Proof. Let $(\Omega, \mathcal{F}, \Pr)$ is a probability space. For each period $t \in \mathbb{N}$, let $Y^t : \Omega \to \mathcal{P}$ be the arrival profile; the compatibility graph G^t is the measurable function of *(carryovers,* Y^t*)* defined in the model. Given (G^t, w^t), the period-t selection rule (DAKSA-2) chooses a measurable selector

$$M^{*t} \in \arg\max_{M \in \mathcal{E}^t} \sum_{(v,v') \in M} w^t(v, v'),$$

where $\mathcal{E}^t$ is the bi-criteria KS-feasible set. Weights update predictably via $w^{t+1} = w^t + \Delta^t$, with Δ^t a measurable function of (G^t, M^{*t}).
Define the natural filtration

$$\mathcal{H}_t := \sigma\big(Y^1, \dots, Y^t,\ M^{*1}, \dots, M^{*t}\big), \qquad t \geq 0, \quad \mathcal{H}_0 := \{\emptyset, \Omega\}.$$

Then G^t, M^{*t}, Δ^t, and w^{t+1} are $\mathcal{H}_t$ measurable, while w^t is $\mathcal{H}_{t-1}$ measurable (predictable).[19] Under Assumption 1, $\{Y^t\}$ are i.i.d., so Y^{t+1} is independent of $\mathcal{H}_t$ for every t.[20]

Lemma 7 (Weight–shortfall identity)**.** *Let* $r_c^{t,K} := r_c^{t,K}(M^{*t})$*,* $r_c^t := \min\{r_c^{t,T}, r_c^{t,H}\}$*, and* $\Delta_c^t := 1 - r_c^t \in [0, 1]$*. If* $w_c^1 \equiv 1$ *and* $w_c^{t+1} = w_c^t + \Delta_c^t$ *on recipient-side edges, then for all* $t \geq 1$*,*

$$w_c^t \;=\; 1 + \sum_{s=1}^{t-1} \Delta_c^s \qquad \text{and} \qquad w_c^t - w_{c'}^t \;=\; \sum_{s=1}^{t-1} \big(\Delta_c^s - \Delta_{c'}^s\big).$$

[19]This predictability is all we need to justify conditional expectations and martingale tools below.
[20]If arg max is set-valued, a measurable selector exists by the Kuratowski-Ryll-Nardzewski theorem; see Aliprantis and Border (2006).

Proof. Induction on t yields $w_c^t = 1 + \sum_{s=1}^{t-1} \Delta_c^s$. Subtract the identities for c and c'. $\square$

Lemma 8 (Fairness-safe augmentations are adopted)**.** *Fix t. If $Y^t \in E_{c,K}$ from Assumption 2 (read componentwise), then there exists a vertex-disjoint $C_t^{c,K}$ that (i) yields at least one additional K-type recipient for c and (ii) preserves bi-criteria KS feasibility when appended to any $M \in \mathcal{E}^t$. Any maximizer of $\sum_{(v,v')\in M} w^t(v,v')$ over $\mathcal{E}^t$ must include $C_t^{c,K}$; in particular M^{*t} includes $C_t^{c,K}$.*

Proof. By Assumption 2, such $C_t^{c,K}$ exists and is fairness-safe. Every edge in $C_t^{c,K}$ targets a recipient in c. Because DAKSA-2 assigns a *country-uniform* weight to all edges whose recipient lies in c (they all start at 1 and receive the same additive increments Δ_c^s each period), we have $w_c^t(v,v') \equiv w_c^t$ for all such edges. Let $L := |C_t^{c,K}| \geq 1$. Then

$$\sum_{(v,v')\in C_t^{c,K}} w_c^t(v,v') \;=\; L\, w_c^t \;\geq\; w_c^t \;>\; 0, \tag{17}$$

where $w_c^t > 0$ follows from $w_c^1 \equiv 1$ and $\Delta_c^s \in [0,1]$. Vertex-disjointness ensures no edge of M is removed, and fairness-safety guarantees $M \cup C_t^{c,K} \in \mathcal{E}^t$. Formally, for any $M \in \mathcal{E}^t$ define $M' := M \cup C_t^{c,K} \in \mathcal{E}^t$; then

$$\sum_{(v,v')\in M'} w^t(v,v') \;=\; \sum_{(v,v')\in M} w^t(v,v') \;+\; \sum_{(v,v')\in C_t^{c,K}} w_c^t(v,v') \;>\; \sum_{(v,v')\in M} w^t(v,v'),$$

contradicting maximality if $C_t^{c,K} \not\subseteq M$. Hence every maximizer must include $C_t^{c,K}$. $\square$

Lemma 9 (Linear frequency of fairness-safe opportunities)**.** *For each (c,K) there exists $\varepsilon_{c,K} > 0$ such that, almost surely,*

$$\sum_{t=1}^{N} \mathbf{1}\{Y^t \in E_{c,K}\} \;\geq\; \tfrac{1}{2}\,\varepsilon_{c,K}\,N \quad \textit{for all sufficiently large } N.$$

Proof. Let $I_t^{c,K} := \mathbf{1}\{Y^t \in E_{c,K}\}$. Assumption 2 gives $\mathbb{E}[I_t^{c,K} \mid \mathcal{H}_{t-1}] \geq \varepsilon_{c,K} > 0$ uniformly in t. Define the bounded-difference martingale $S_N^{c,K} := \sum_{t=1}^{N} \left(I_t^{c,K} - \mathbb{E}[I_t^{c,K} \mid \mathcal{H}_{t-1}]\right)$ with $|S_N^{c,K} - S_{N-1}^{c,K}| \leq 1$. By Azuma-Hoeffding (Azuma (1967), Hoeffding (1963)), $\Pr(S_N^{c,K} \leq -x) \leq \exp(-2x^2/N)$. Taking $x = \frac{1}{2}\varepsilon_{c,K} N$ and applying a conditional Borel-Cantelli (Freedman (1975)) yields the almost-sure linear lower bound[21]. $\square$

Lemma 10 (Denominator growth)**.** *If $\Pr(B_c^{t,K} - d_c^{t,K} > 0) > 0$ under Assumption 1, then $D_c^K(\mathcal{T}) := \sum_{t=1}^{\mathcal{T}} (B_c^{t,K} - d_c^{t,K})$ satisfies $D_c^K(\mathcal{T}) = \Theta(\mathcal{T})$ almost surely.*

[21] Azuma-Hoeffding supplies concentration uniformly conditional on $\mathcal{H}_{t-1}$; Freedman's conditional Borel-Cantelli converts conditional expectations into a.s. linear occurrence.

Proof. The sequence $\{B_c^{t,K} - d_c^{t,K}\}$ is i.i.d., nonnegative, and bounded (bounded cycle length). By the Strong Law, $D_c^K(\mathcal{T})/\mathcal{T} \to \mathbb{E}[B_c^{1,K} - d_c^{1,K}] > 0$ a.s. □

For $K \in \{T, H\}$ define

$$N_c^K(\mathcal{T}) := \sum_{t=1}^{\mathcal{T}} \bigl(X_c^{t,K}(M^{*t}) - d_c^{t,K}\bigr), \qquad R_c^{\mathcal{T},K} := \frac{N_c^K(\mathcal{T})}{D_c^K(\mathcal{T})}.$$

By Lemma 10, there exist constants $0 < m_K \le M_K < \infty$ and a random $T_0(\omega)$ such that for all $\mathcal{T} \ge T_0$ and all c,

$$m_K\,\mathcal{T} \;\le\; D_c^K(\mathcal{T}) \;\le\; M_K\,\mathcal{T}. \tag{18}$$

By Lemma 8, on every event $\{Y^t \in E_{c,K}\}$ the selected M^{*t} includes a fairness-safe $C_t^{c,K}$, increasing N_c^K by at least 1 and not decreasing any $N_{c'}^K$ (vertex-disjointness and feasibility). Define the directed shortfall

$$G_{c,c'}^K(t) := \max\{0,\ N_{c'}^K(t) - N_c^K(t)\}.$$

When country c receives a corrective K-type augmentation at t, N_c^K increases by at least 1, hence $G_{c,c'}^K$ weakly decreases for every c' and, whenever $G_{c,c'}^K(t^-) > 0$, it drops by at least 1. Thus, at such t,

$$N_c^K - N_{c'}^K \quad \text{decreases by at least 1} \quad \text{for all } c' \ne c. \tag{19}$$

By Lemma 9, for each (c, K) the corrective event $\{Y^t \in E_{c,K}\}$ occurs with linear frequency; by Lemma 10, each $D_c^K(\mathcal{T})$ grows linearly a.s. Suppose, toward a contradiction, that there exist c, c' and $\eta > 0$ and an infinite subsequence $\{\mathcal{T}_m\}$ with $R_c^{\mathcal{T}_m,K} - R_{c'}^{\mathcal{T}_m,K} \ge \eta$. Without loss, label the pair at time $\mathcal{T}_m$ so that ℓ is the leader and ι the laggard: $R_\ell^{\mathcal{T}_m,K} - R_\iota^{\mathcal{T}_m,K} \ge \eta$. Equivalently,

$$N_\ell^K(\mathcal{T}_m) D_\iota^K(\mathcal{T}_m) - N_\iota^K(\mathcal{T}_m) D_\ell^K(\mathcal{T}_m) \;\ge\; \eta\, D_\ell^K(\mathcal{T}_m)\, D_\iota^K(\mathcal{T}_m). \tag{20}$$

Using Equation 18, the right-hand side is bounded below by $\eta\, m_K^2\, \mathcal{T}_m^2$ for all large m. Consider the "potential" $\Phi_{\ell,\iota}^K(t) := N_\ell^K(t) D_\iota^K(t) - N_\iota^K(t) D_\ell^K(t)$. When ι receives a corrective K-type augmentation at period t, N_ι^K increases by 1 while D_ℓ^K and D_ι^K are unchanged at t, so $\Phi_{\ell,\iota}^K$ decreases by exactly $D_\ell^K(t)$. By Lemma 9, such events occur with linear frequency, and by Equation 18 we have $D_\ell^K(t) \ge m_K\, t$ eventually. Hence the cumulative decrease in $\Phi_{\ell,\iota}^K$ up to time $\mathcal{T}$ is at least $c_K \mathcal{T}^2$ for some $c_K > 0$. But Equation 20 asserts that $\Phi_{\ell,\iota}^K(\mathcal{T}_m) \ge \eta\, m_K^2\, \mathcal{T}_m^2$ along the subsequence, while the preceding paragraph shows that (due to the laggard's linear stream of corrective events) $\Phi_{\ell,\iota}^K$ must be driven down by at least $c_K \mathcal{T}_m^2$ along the same horizon. Choosing $c_K > \eta m_K^2/2$ (possible by Lemma 9 and Equation 18) yields a contradiction. Therefore, $\max_{c,c'} |R_c^{\mathcal{T},K} - R_{c'}^{\mathcal{T},K}| \to 0$ almost surely. Then

$$N_c^K(\mathcal{T}_m) - N_{c'}^K(\mathcal{T}_m) \;\geq\; \eta\, D_*^K(\mathcal{T}_m), \quad D_*^K(\mathcal{T}) := \min_{d\in C} D_d^K(\mathcal{T}) = \Theta(\mathcal{T}),$$

which contradicts the linear cumulative reductions implied by Equation 19 and Lemma 9. Hence,

$$\lim_{\mathcal{T}\to\infty} \max_{c,c'\in C} \left| R_c^{\mathcal{T},K} - R_{c'}^{\mathcal{T},K} \right| = 0 \qquad \text{a.s., for each } K \in \{T, H\}.$$

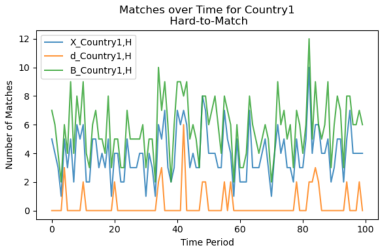
Taking $K = T$ and $K = H$ yields the two displays in the theorem. Equivalently, cumulative gain ratios equalize across countries in both dimensions[22]. □

[22] By Lemma 7, the worst-component update $\Delta_c^t = 1 - \min\{r_c^{t,T}, r_c^{t,H}\}$ induces persistent equity drift toward any lagging component, aligning the tie-break with fairness-safe augmentations.

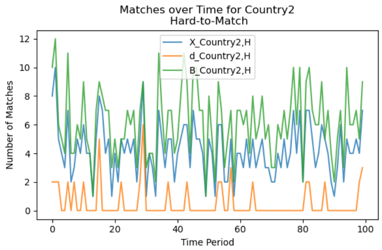
A.7 Additional Plots

A.7.1 Country-Level Plots for DAKSA-2

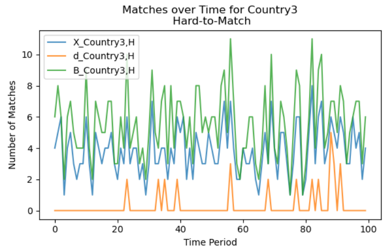
Figure 5a-5f display period-by-period outcomes for both total and hard-to-match recipients. The qualitative patterns closely parallel those under DAKSA: all countries obtain sustained improvements relative to autarky, and the hard-to-match subgroup benefits from explicit protection.



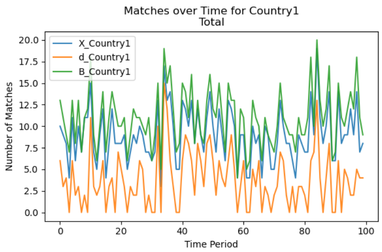
(a) Country 1 - Hard-to-Match



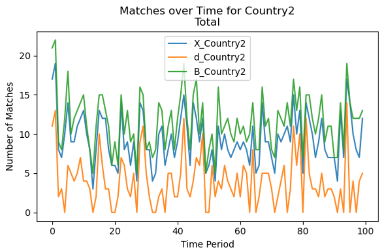
(b) Country 2 - Hard-to-Match



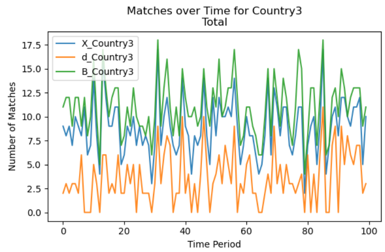
(c) Country 3 - Hard-to-Match



(d) Country 1 - Total



(e) Country 2 - Total

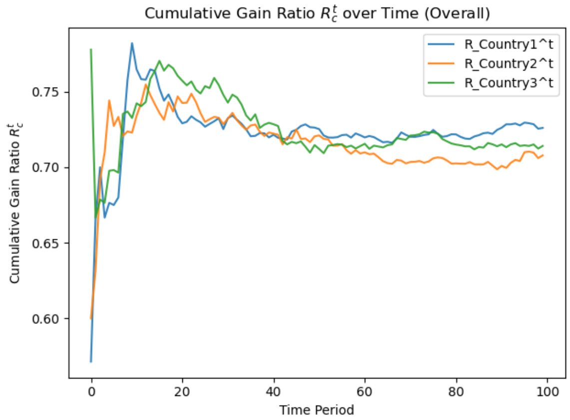


(f) Country 3 - Total

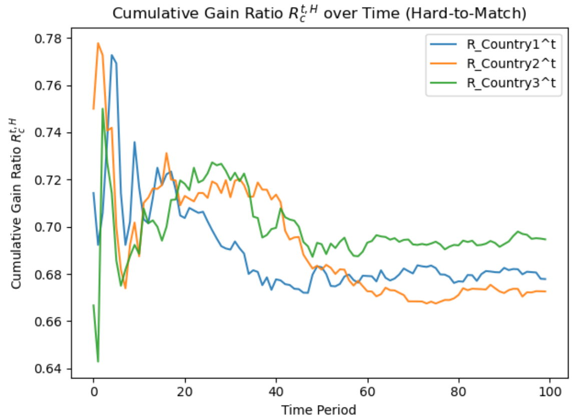
Figure 5: Country-level matches under DAKSA-2, shown separately for hard-to-match recipients (top row) and total recipients (bottom row). Blue line: realised matches X_c^t; orange: standalone benchmark d_c^t; green: cooperative best B_c^t.

A.7.2 Cumulative-Convergence Plots for DAKSA-2

For completeness, we report the full set of cumulative-ratio and maximum-gap plots for DAKSA-2. Figure 6a and Figure 6b track the cumulative gain ratios $R_c^{t,T}$ and $R_c^{t,H}$ for each country. As in the single-objective case, both the total and hard-to-match cumulative gain ratios converge across countries, and in each replication the maximum gap shrinks rapidly over time. These patterns are consistent with the bi-criteria convergence guarantee in Theorem 2.

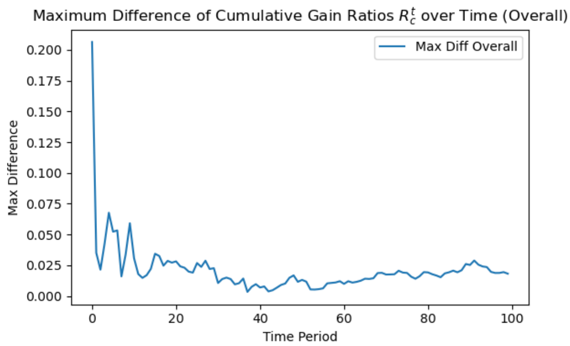


(a) Cumulative ratios $R_c^{t,T}$ (Total)

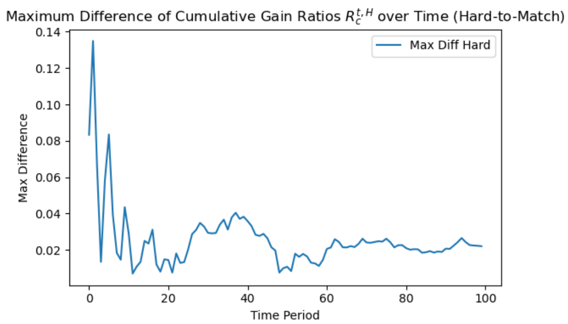


(b) Cumulative ratios $R_c^{t,H}$ (Hard-to-Match)

Figure 6: Cumulative gain ratios under DAKSA-2. Both total and hard-to-match dimensions converge across countries, confirming long-run equity.



(a) Maximum gap across $R_c^{t,T}$ (Total)

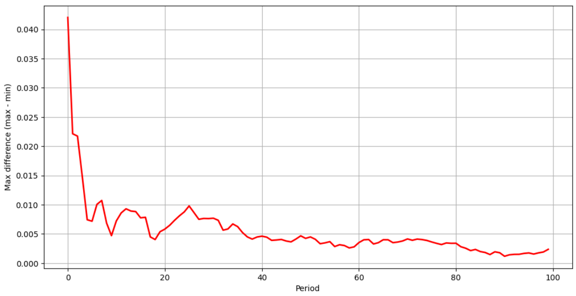


(b) Maximum gap across $R_c^{t,H}$ (Hard-to-Match)

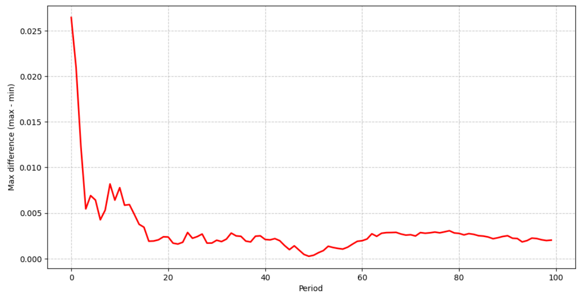
Figure 7: Convergence of cumulative gain ratios under DAKSA-2. Dispersion shrinks rapidly in both total and hard-to-match dimensions.

A.7.3 Robustness Results for DAKSA-2

For completeness, we report the robustness results for DAKSA-2. Across 50 independent arrival streams, both the total and hard-to-match cumulative gain ratios converge, and the maximum gap between countries contracts steadily over time to the order of 0.005 by the 100^{th} time period. These patterns mirror those observed under DAKSA and provide empirical support for the bi-criteria convergence guarantee in Theorem 2.



(a) Total

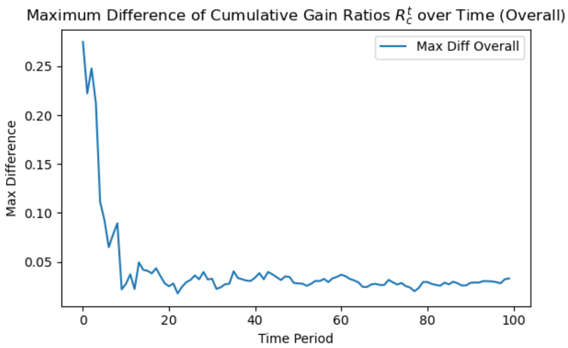


(b) Hard-to-Match

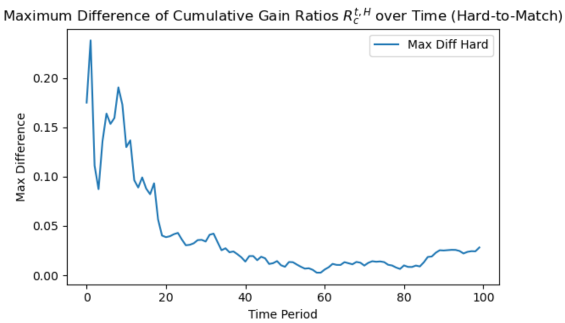
Figure 8: Maximum difference of average cumulative gain ratios at each round under DAKSA-2 after 50 rounds of simulations.

A.7.4 Unequal Pool Size Simulation across Countries under DAKSA-2

In addition to the baseline specification with equal expected pool sizes, we also simulate a variant with unequal populations of donor–recipient pairs across countries. Specifically, new arrivals are assigned so that, in expectation, approximately 50%, 35%, and 15% of pairs belong to Country 1, Country 2, and Country 3, respectively. The qualitative patterns remain consistent with the main results discussed in the paper.

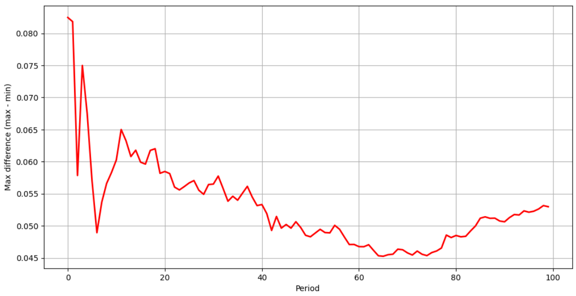


(a) Maximum gap across $R_c^{t,T}$ (Total)

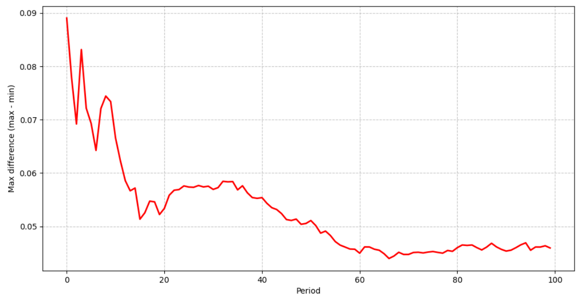


(b) Maximum gap across $R_c^{t,H}$ (Hard-to-Match)

Figure 9: Maximum difference of cumulative gain ratios (R_c^t) at each round under DAKSA-2. Convergence of cumulative gain ratios under DAKSA-2. Dispersion shrinks rapidly in both total and hard-to-match dimensions.



(a) Maximum gap across $R_c^{t,T}$ (Total)



(b) Maximum gap across $R_c^{t,H}$ (Hard-to-Match)

Figure 10: Maximum difference of average cumulative gain ratios at each round under DAKSA-2 after 50 rounds of simulations with unequal proportion of donor-recipient pairs. Both total and hard-to-match dimensions converge across countries, confirming long-run equity, consistent with Theorem 2.