Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Assigning Doctors to Rural Hospitals: Redesigning Colombia's Compulsory Social Service

Paula Jaramillo
Universidad de Los Andes
Cagatay Kayi
cagatay.kayi@urosario.edu.co
Universidad del Rosario
Santiago Velez-Ferro
Wayfair

Research Article

Keywords: matching, school choice, deferred acceptance, priorities, strategy-proofness, market design

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11323290/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Assigning Doctors to Rural Hospitals: Redesigning Colombia's Compulsory Social Service*

Paula Jaramillo† Çağatay Kayı‡ and Santiago Vélez-Ferro§

October, 2026

Abstract

We study many-to-one matching problems where doctors have strict preferences over hospitals, hospitals have strict preferences over sets of doctors, and the law assigns priorities to hospitals over doctors. A mechanism assigns a matching between doctors and hospitals. We are interested in mechanisms that respect priority, are efficient, are immune to strategic behavior, and minimize unfilled positions. We show that the lottery-based mechanisms used in Colombia in 2012–2013 satisfy none of these desirable properties. We then consider three mechanisms proposed in the literature: the deferred acceptance mechanism, the immediate acceptance mechanism (also known as the Boston mechanism), and the top trading cycle mechanism, and we explore different versions of *efficiency*, namely *doctor-efficiency* and *priority-constrained efficiency*.

We find that all three mechanisms are *priority-constrained efficient* and *minimize unfilled positions*. The deferred acceptance mechanism is the only one of the three that *respects priority*, and it is *strategy-proof*, although it is neither *doctor-efficient* nor *efficient*. The top trading cycle mechanism and the immediate acceptance mechanism are *doctor-efficient* but do not *respect priority*; the top trading cycle mechanism is *strategy-proof* while the immediate acceptance mechanism is not. Finally, under the deferred acceptance and top trading cycle mechanisms—but not under the immediate acceptance mechanism—no doctor can obtain exemption from the compulsory service by misreporting.

Keywords: matching; school choice; deferred acceptance; priorities; strategy-proofness; market design.

JEL Classification: C78; D47; D63; I18.

*We thank seminar participants at the 13th Meeting of the Society for Social Choice and Welfare and at Universidad del Rosario for helpful discussions and suggestions. This paper is based in part on a chapter of Vélez-Ferro's (2019) Ph.D. dissertation.

†Universidad de Los Andes, Facultad de Economía, Bogotá, Colombia.

‡*Corresponding author.* Universidad del Rosario, School of Economics and Business, Bogotá, Colombia; e-mail: cagatay.kayi@urosario.edu.co

§Wayfair, Boston, MA, USA.

1 Introduction

This paper studies two-sided many-to-one matching markets where doctors have strict preferences over hospitals, hospitals have strict preferences over sets of doctors, and the law assigns priorities to hospitals over doctors.

There are many real-life two-sided many-to-one matching markets. Two important instances are the national resident matching program and the student assignment system to public schools. The former is an entry-level matching market for hospitals and medical school graduates in the United States where doctors have strict preferences over hospitals and hospitals have strict preferences over sets of doctors. The latter is the public school match where students have strict preferences over schools and the law assigns priorities to schools over students. The design of such centralized mechanisms matters. Clearinghouses that produced unstable outcomes were abandoned; those that produced stable ones survived (Roth, 1991).

Many real-life matching markets employ centralized mechanisms to match *agents* to positions, and the only information the matchmaker requires from each participating agent is a ranking of the other side of the market. Moreover, we assume that the hospitals' quotas (i.e., the numbers of available positions) are commonly known. In practice, agents may submit only ordered lists of individual partners, and we restrict attention to mechanisms that ask for no more than that.

An assignment between doctors and hospitals is called a *matching*. A *mechanism* assigns a matching to each problem. Next, we describe desirable properties of mechanisms. First, we are interested in mechanisms that *respect priority*, i.e., for each problem, the mechanism assigns a matching such that there is no doctor-hospital pair such that they are not mates in the matching but the doctor would prefer to be matched to the hospital and the hospital has unfilled positions or is matched to another doctor who has a lower priority for the hospital.

Second, we are interested in a set of properties related to Pareto-efficiency. A mechanism is *doctor-efficient* if for each problem, the mechanism assigns a matching such that there is no other matching that each doctor finds at least as desirable and at least one doctor strictly prefers. A mechanism is *efficient* if for each problem, the mechanism assigns a matching such that there is no other matching that each agent finds at least as desirable and at least one agent strictly prefers. A mechanism is *priority-constrained efficient* if for each problem, the mechanism assigns a matching such that there is no priority-respecting matching that each agent finds at least as desirable and at least one agent strictly prefers.

The third property is *strategy-proofness*, i.e., no doctor should benefit from misrepresenting her preferences. The last property is *minimizing unfilled positions*, i.e., for each problem, the mechanism assigns a matching such that there is no doctor-hospital pair such that the hospital has unfilled positions but the doctor is exempt.

Since participation is compulsory and exemption arises only as a residual, a professional may wish to misreport in order to obtain *no* assignment rather than a better one—the manipulation observed in Colombia in 2012 (Section 1.1). The classical strategy-proofness results compare the partners an agent can obtain and are silent about this possibility, so we treat it as a separate question and answer it in Proposition 9.

1.1 Compulsory Social Service Allocation in Colombia

Origins and the Case for Compulsion

The Compulsory Social Service Allocation (CSS) was created to address the persistent inequality in the distribution of healthcare workers across Colombia's regions, particularly the chronic shortage of professionals in rural and marginalized urban areas. It was introduced in 1949 for physicians, extended to dentists in 1951, and to medical laboratory scientists and nurses in 1971. Participation lasts one year; the allocation is decided and implemented by the Ministry of Health on a quarterly basis. Colombia is not alone: Bolivia, Costa Rica, Ecuador, Honduras, Nicaragua, Mexico, Peru, and El Salvador all operate similar compulsory programs, with Mexico the first to do so in 1936.

The theoretical motivation is grounded in the matching literature. Roth (1986) documents that under any stable allocation the set of unfilled hospital positions is invariant—the *Rural Hospital Theorem*—and concludes that the maldistribution of physicians to rural areas "seems unlikely to be changed by any system that does not involve some element of compulsion, or some change in the relative numbers of available positions." Colombia's CSS is a direct policy response: by making participation compulsory, the Ministry sidesteps the constraints that stability imposes on rural allocations.

Structure of the CSS

Four health professions participate: Medical Laboratory Science, Nursing, Medicine, and Dentistry. Professional certification—the legal right to practice—is contingent on completing the CSS, giving the Ministry strong enforcement leverage. Three types of positions are available: rural (at least 70 % of the total), marginalized urban (at most 25 %), and research (at most 5 %). The first two are allocated through the centralized mechanism; research positions are assigned directly by the research entities involved.

Colombian law and Ministry regulations impose a three-tier priority structure. The first tier (law) grants geographic priority to indigenous professionals at hospitals in indigenous regions and to *raizal* professionals at hospitals in the Archipelago of San Andrés, Providencia, and Santa Catalina. The second tier (Ministry regulations) grants personal-circumstance priority to pregnant professionals, parents of small children, professionals with impairments or disabilities, and those requiring special medical treatment. The third tier (hospital preferences) gives precedence to professionals who graduated from or were born in the same state as the hospital. Within each tier, ties are broken by lottery. In our model, P_h therefore stands for the third-tier criteria together with the realized lottery draw; since the draw does not depend on what professionals report, it yields a strict order over doctors that is part of the data of the problem. Most professionals qualify only for a third-tier priority, making the effective structure quite coarse.

Unlike in most clearinghouses, high priority is not always welcome. Participants outnumber positions in most professions, so high priority raises the chance of being assigned, which some professionals would rather avoid.

The compulsory character of the CSS is enforced through two policy constraints. First, professionals cannot report an outside option: the algorithm treats all hospitals in the constructed ranking

as acceptable. Second, a professional who rejects an assigned position is barred from the next two quarterly allocations (a nine-month waiting penalty). Despite this, roughly 3–6 % of professionals reject their allocation in practice. Simulation evidence underscores why compulsion matters: under voluntary participation, roughly 115 physicians—about 30 % of participants—would have gone unassigned, and raising wages is an expensive and only partial substitute. An average increase of roughly 600 USD per position *per month*—about twice the Colombian monthly minimum wage at the time—would have induced only 35 additional physicians to be allocated, still leaving some 82 unassigned despite ample available positions. Notably, the 115 physicians who would go unassigned under voluntary participation are close in number to the 133 who strategized to avoid allocation under the 2012 mechanism, suggesting that the manipulation was driven by genuine preferences for exemption rather than by confusion about the mechanism (Vélez-Ferro, 2019).

Earlier Mechanisms and Their Failures

The Ministry allocated positions through a state-level lottery: each professional enrolled in one of Colombia's 32 states (*departamentos*), and a lottery assigned the enrolled professionals to hospitals in that state; those not assigned were exempt from the CSS. Because professionals could choose their state and any leftover professional was exempt, groups coordinated to flood low-capacity states and seek exemption. In July 2012, for instance, 135 physicians enrolled in a state with only two available positions, guaranteeing exemption for 133 of them; in April and June 2012, groups of 74 and 225 physicians applied for states offering only 6 and 7 positions, respectively. In the July 2012 allocation, nearly 600 physicians were exempted from service, while approximately 300 hospital positions remained unfilled.

In January 2013, the Ministry introduced enrollment caps (at most twice the available positions per state, on a first-come first-served basis). While this reduced the scale of gaming, exempted professionals still outnumbered unfilled positions, demonstrating that the incentive problem was not resolved.

Figures 1 and 2 document these failures. Figure 1 shows that across every quarterly allocation in 2012 and 2013 there were simultaneously exempted professionals *and* unfilled positions—a clear indication of both inefficiency and manipulation.

Figure 2 shows the over-enrollment in low-capacity states that coordinated manipulation produces.

The 2014 Reform: Adoption of the Deferred Acceptance Algorithm

Three failures led the Ministry of Health to redesign the mechanism in 2014: unfilled positions, strategic manipulation, and the disregard of legal priorities. The theoretical results in this paper provide the formal justification: among the mechanisms we study, $DA^{\succeq}$ is the only one that simultaneously respects priority, is priority-constrained efficient, is strategy-proof, and minimizes unfilled positions—the four properties that the Colombian setting requires. Since October 2014, the allocation has been determined using a version of Gale and Shapley's (1962) doctor-proposing deferred acceptance algorithm incorporating the priority structure described above.

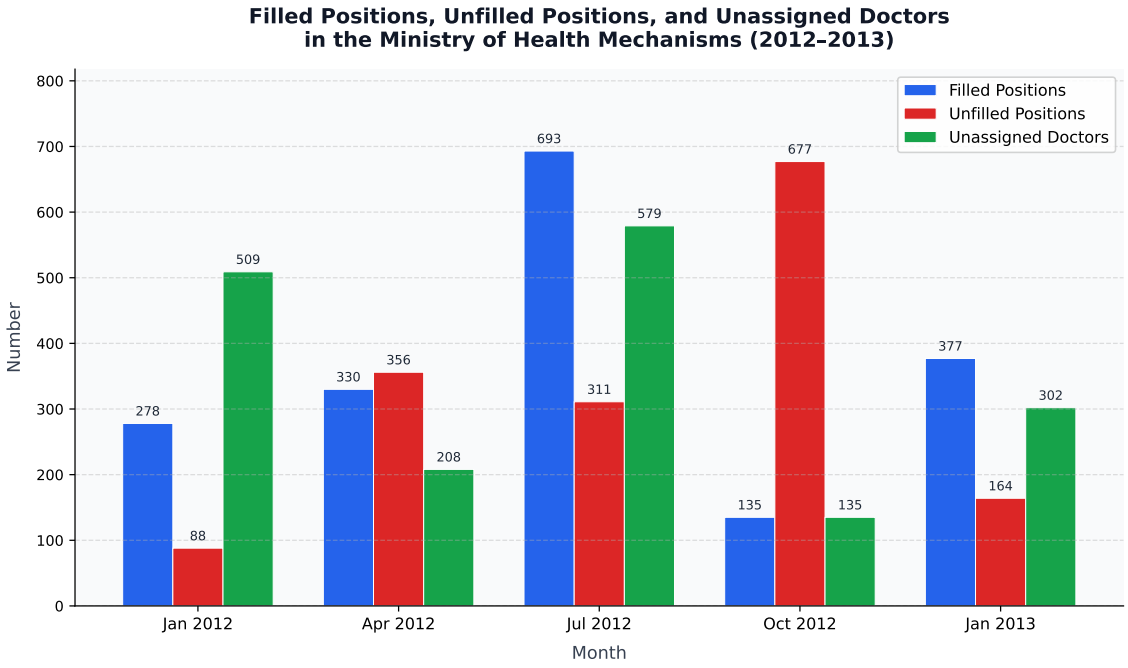


Figure 1: Filled positions, unfilled positions, and unassigned doctors in the mechanisms used by the Ministry of Health in 2012 and 2013 (Vélez-Ferro, 2019). Unassigned doctors are those exempted from the service.

Under this mechanism, professionals report up to five preferred hospitals directly and submit a ranking of the states (*departamentos*) in which they would be willing to serve. The Ministry constructs a complete preference ranking for each professional by placing the five reported hospitals at the top, followed by all hospitals in each reported state (in the order ranked), excluding those already listed.

Table 1 reports the scale of the first centralized allocation (October 2014). Congestion varied sharply by profession: medicine was the only profession with a surplus of positions (roughly 2.7 per participant), while dentistry was the most congested market, with 6.5 participants per position. Medicine is also the profession in which the coordinated exemption-seeking described above took place, and the one in which positions went unfilled while professionals were exempted (Vélez-Ferro, 2019).

Profession	Participants	Positions	Hospitals	Participants per position
Medical Laboratory Science	194	83	81	2.34
Nursing	828	194	148	4.27
Medicine	386	1,025	544	0.38
Dentistry	708	109	97	6.50

Table 1: Participants, available positions, and participating hospitals by profession in the first centralized allocation, October 2014 (Vélez-Ferro, 2019).

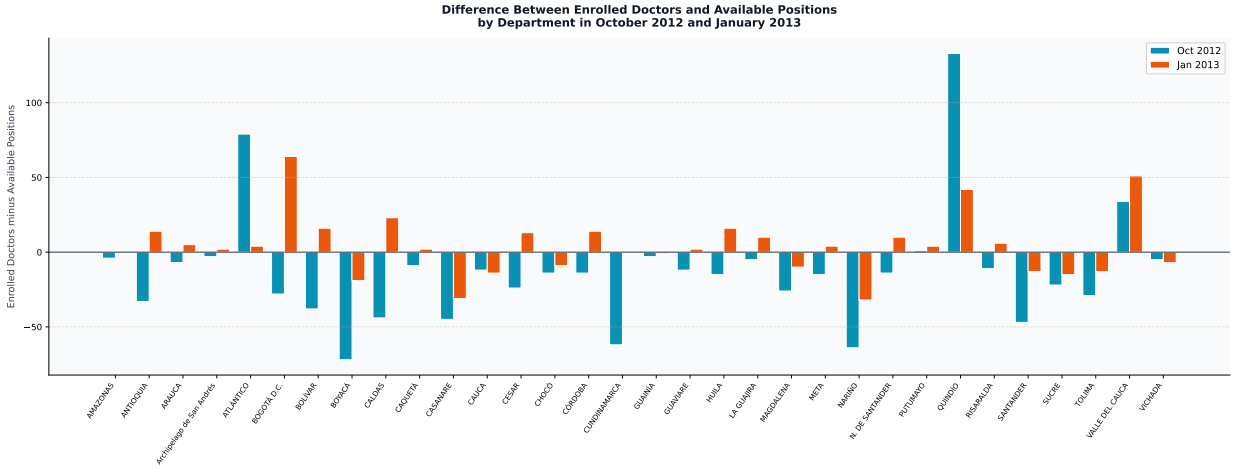


Figure 2: The difference between the number of enrolled doctors and the available positions in each state of Colombia in October 2012 and January 2013 (Vélez-Ferro, 2019).

After the transition, unfilled positions were minimized, the coordinated enrollment described above disappeared, and legal priorities were applied for the first time. Empirical analysis of the 2014 data confirms these gains and shows that welfare improvements of 8–12 % could be achieved by refining how ties within priority groups are broken—for instance, favoring professionals who directly reported a hospital over those for whom it appeared via a state ranking (Vélez-Ferro, 2019).

1.2 Related Literature

Our setting is a school choice problem in the sense of Abdulkadiroğlu and Sönmez (2003): hospitals are objects with legally mandated priorities, and we evaluate the same three mechanisms they analyze. It departs from that framework in two ways. First, hospital preferences enter the welfare criteria. They encode the policy criteria the Ministry uses within legal priority classes: place of birth and place of graduation. Second, and more consequentially, participation is compulsory. In school choice and in the entry-level labor markets studied since Gale and Shapley (1962) and Roth (1984), agents misreport to get a better partner. Strategy-proofness rules that out. In Colombia, a professional may misreport to get *no* partner. The classical incentive results—Dubins and Freedman (1981) and Roth (1982) for deferred acceptance, Roth (1982b) and Abdulkadiroğlu and Sönmez (2003) for top trading cycles—compare the partners an agent can obtain and are silent on this margin. Proposition 9 answers this question, and Remark 2 shows that the immediate acceptance mechanism does not. Our proof rests on Roth's (1986) Rural Hospital Theorem, which says that the same positions go unfilled under every stable matching. In Colombia's case that invariance is the problem itself: no voluntary system will staff the rural hospitals the law is trying to staff.

Coarse legal priorities make ties pervasive, which ordinarily forces a choice between constrained efficiency and strategy-proofness (Erdil and Ergin, 2008; Abdulkadiroğlu, Pathak, and Roth, 2009). Proposition 4 avoids that trade-off. What our results require of the tie-breaking rule is only that it be strict and independent of the professionals' reports; the third-tier criteria and the lottery draw

that settles remaining ties both satisfy this. The classical results then apply directly, conditional on the realized order. The trade-off returns for any rule built from the reports themselves (Footnote 5 and Section 4).

Finally, the paper documents a reform. Colombia's 2012–2013 lottery mechanisms failed in a way the theory predicts, and in 2014 the Ministry replaced them with the deferred acceptance mechanism we study here. Vélez-Ferro (2019) provides the empirical record of both regimes; we explain why the earlier mechanisms failed and why the chosen one succeeds. Roth's (1991) account of British clearinghouses, where markets with unstable outcomes were abandoned and those with stable outcomes survived, is the closest precedent for this kind of natural experiment.

1.3 Overview of the Results

To analyze the CSS allocation of medical professionals in Colombia, we study a combination of many-to-one matching and school choice problems, where each professional can work for at most one hospital and each hospital can hire at most the number of positions it offers. Each professional has a strict preference over hospitals and being exempt; each hospital has strict preferences over sets of professionals and being unmatched; and the Colombian law or the Ministry of Health assigns priorities to hospitals over professionals. Because hospitals in this setting are not strategic agents—the Ministry imposes their priority orders—the problem is effectively one-sided, in the spirit of school choice (Abdulkadiroğlu and Sönmez, 2003).

For the CSS, we consider Gale and Shapley's (1962) deferred acceptance algorithm incorporating both the reported preferences of professionals and the three-tier priority structure dictated by Colombian law and Ministry regulations. This mechanism resolves all three problems identified above. First, it incorporates priorities systematically, so that professionals with disabilities, mothers with small children, and professionals of indigenous origin are guaranteed precedence at the hospitals where they have legal priority. Second, because the deferred acceptance algorithm is strategy-proof, professionals have dominant-strategy incentives to report their true preferences—eliminating the manipulation documented in 2012. Third, the mechanism minimizes unfilled positions: whenever the total number of available positions does not exceed the number of participants, no position is left unfilled.

We also consider the immediate acceptance mechanism (the so-called Boston mechanism) and the top trading cycle mechanism. We find that all three mechanisms are *priority-constrained efficient* and *minimize unfilled positions*. We show that the deferred acceptance mechanism always assigns a matching that *respects priority*, although it is neither *doctor-efficient* nor *efficient*. On the other hand, the top trading cycle mechanism and the immediate acceptance mechanism are *doctor-efficient*, but do not *respect priority*. As expected, the deferred acceptance mechanism and the top trading cycle mechanism are *strategy-proof* but the immediate acceptance mechanism is not. Moreover, under the deferred acceptance mechanism and the top trading cycle mechanism, no doctor can obtain exemption by misreporting her preferences, whereas the immediate acceptance mechanism permits exactly the exemption-seeking behavior observed in 2012.

The remainder of the paper is organized as follows. In Section 2 we describe the model. In Section 3 we present our results. Section 4 concludes.

2 Model

There are two finite and disjoint sets of agents: a set of **doctors** $\boldsymbol{D}$ and a set of **hospitals** $\boldsymbol{H}$. Let $\boldsymbol{I} = \boldsymbol{D} \cup \boldsymbol{H}$ be the set of agents. We denote a generic doctor, hospital, and agent by d, h, and i, respectively. For each hospital h, there is an integer **quota** $\boldsymbol{q_h} \geq 1$ representing the number of positions it offers. Each doctor d can work for at most one hospital and each hospital h can hire at most q_h doctors. Let $q = (q_h)_{h \in H}$. For each $i \in I$, the set of potential partners of agent i is denoted by $\boldsymbol{N_i}$. If $i \in D$, $N_i = H$; if $i \in H$, $N_i = D$.

Each doctor d has a complete, transitive, and strict **preference relation** $\boldsymbol{P_d}$ over the hospitals and the prospect of being "exempt", denoted by $\emptyset$. For $h, h' \in H \cup \{\emptyset\}$, we write $h \, P_d \, h'$ if doctor d prefers h to h' ($h \neq h'$), and $h \, R_d \, h'$ if d finds h at least as good as h'. Note that for each doctor, each hospital is acceptable, but a doctor may prefer being exempt to working for a hospital.[1] Let $\boldsymbol{P_D} = (P_d)_{d \in D}$ be the preference profile for the doctors.

Let $h \in H$. A subset of doctors $D' \subseteq D$ is **feasible** for hospital h if $|D'| \leq q_h$. Let $\boldsymbol{\mathcal{F}(D, q_h)} = \{D' \subseteq D : |D'| \leq q_h\}$ denote the collection of feasible subsets of doctors for hospital h. The element $\emptyset \in \mathcal{F}(D, q_h)$ denotes "being unmatched". Each hospital h has a complete, transitive, and strict **preference relation** $\boldsymbol{\mathscr{P}_h}$ over $\mathcal{F}(D, q_h)$. Also, for each hospital, each feasible non-empty subset of doctors is "acceptable", i.e., for each $D' \in \mathcal{F}(D, q_h) \setminus \{\emptyset\}$, $D' \, \mathscr{P}_h \, \emptyset$. For $D', D'' \in \mathcal{F}(D, q_h)$ we write $D' \, \mathscr{P}_h \, D''$ if hospital h prefers D' to D'' ($D' \neq D''$), and $D' \, \mathscr{R}_h \, D''$ if hospital h finds D' at least as good as D''. Let $\boldsymbol{\mathscr{P}_H} = (\mathscr{P}_h)_{h \in H}$ be the preference profile for the hospitals.

Let $\boldsymbol{P_h}$ be the restriction of the preference relation $\mathscr{P}_h$ to $\{\{d\} \mid d \in D\} \cup \{\emptyset\}$, i.e., individual doctors in D and being unmatched. For each $d, d' \in D \cup \{\emptyset\}$, we write $d \, P_h \, d'$ if $\{d\} \, \mathscr{P}_h \, \{d'\}$ and $d \, R_h \, d'$ if $\{d\} \, \mathscr{R}_h \, \{d'\}$.[2] Since each doctor is acceptable, for each $d \in D$ we have $d \, P_h \, \emptyset$. Let $\boldsymbol{P_H} = (P_h)_{h \in H}$.

We assume that for each hospital $h \in H$, $\mathscr{P}_h$ is **responsive**[3], or more precisely, a responsive extension of P_h, such that:

(r1) as long as a hospital's quota is not reached, it prefers to fill a position with a doctor rather than leaving it unfilled, and

(r2) if faced with two sets of potential doctors that differ only in one doctor, the hospital prefers the set containing the more preferred doctor.

[1] We distinguish the prospect of being *exempt* ($\emptyset$), which releases the doctor from the compulsory social service, from the prospect of "being unmatched" in the sense of Roth (1984), whose convention denotes the outside option of agent d by d itself. Rejecting an assigned position carries severe consequences: the professional is barred from the following two allocation rounds and, in the extreme case, does not receive the license to practice medicine. We therefore assume that every hospital is acceptable to every doctor, so that being unmatched is never an outcome: each doctor is either assigned to a hospital or exempt. A doctor may nevertheless prefer exemption to service at some hospitals, as the enrollment behavior described in Section 1.1 indicates. Because the service is compulsory, however, a doctor cannot apply for exemption: she reports a ranking of hospitals only, and the position of $\emptyset$ in P_d is never reported. Whether exemption can nonetheless be obtained strategically is the subject of Proposition 9.

[2] With some abuse of notation, we often write x for a singleton $\{x\}$.

[3] See Roth (1985) and Roth and Sotomayor (1989) for a discussion of this assumption.

Formally, for each $D' \in \mathcal{F}(D, q_h)$:
(r1) if $d \in D \setminus D'$ and $|D'| < q_h$, then $(D' \cup \{d\}) \mathscr{P}_h D'$; and
(r2) if $d \in D \setminus D'$ and $d' \in D'$, then $((D' \setminus \{d'\}) \cup \{d\}) \mathscr{P}_h D'$ if and only if $d P_h d'$.

Each hospital $h \in H$ has a complete and transitive **priority relation $\succeq_h$** over D (not necessarily strict). Let $\succ_h$ and $\sim_h$ denote the asymmetric and symmetric parts of $\succeq_h$, respectively: if $d \succ_h d'$, then d has strictly higher priority than d' at h; if $d \sim_h d'$, we say that d and d' are *tied* in priority at h. The priority relation of each hospital is given by law, and hence priority supersedes preference; preferences are used only to break ties in priority.

Since the law may treat several doctors as equals, we work throughout with the **priority-augmented order $\succ'_h$**, defined for each $h \in H$ by

$$d \succ'_h d' \quad \text{if and only if} \quad \text{either } d \succ_h d', \text{ or } d \sim_h d' \text{ and } d P_h d'.$$

Since $\succeq_h$ is complete and P_h is strict, $\succ'_h$ is a strict order on D; it refines $\succeq_h$, in the sense that $d \succ_h d'$ implies $d \succ'_h d'$. Note that $\succ'_h$ depends only on the primitives $\succeq_h$ and P_h, and not on the doctors' reports. We write $\succ'_H = (\succ'_h)_{h \in H}$.

A **problem** is given by $(P_D, P_H, \succeq_H, q)$. Let $\mathbb{P}$ be a set of problems. A **matching** is a function μ on $D \cup H$ such that:

(1) each doctor is either matched to exactly one hospital or exempt, i.e., for each $d \in D$, either $\mu(d) \in H$ or $\mu(d) = \emptyset$;

(2) each hospital is matched to a feasible set of doctors, i.e., for each $h \in H$, $\mu(h) \in \mathcal{F}(D, q_h)$; and

(3) a doctor is matched to a hospital if and only if the hospital is matched to the doctor, i.e., for each $d \in D$ and $h \in H$, $\mu(d) = h$ if and only if $d \in \mu(h)$.

Let μ be a matching, $d \in D$, and $h \in H$. If $\mu(d) = h$ or equivalently $d \in \mu(h)$, we say d and h are *mates* in μ.[4] Let $\mathcal{M}(P_D, P_H, \succeq_H, q)$ be the set of matchings for the market $(P_D, P_H, \succeq_H, q)$.

A matching *respects priority* if there is no doctor-hospital pair such that they are not mates but the doctor would prefer the hospital and the hospital either has unfilled positions or is matched to another doctor who has lower priority for the hospital (with the hospital's preference breaking ties in priority). Formally, a matching μ **respects priority** if there are no $d \in D$ and $h \in H$ such that $[h P_d \mu(d)]$ and $[|\mu(h)| < q_h$ or there is $d' \in \mu(h)$ with $d \succ'_h d']$. Let $\mathcal{R}(P_D, P_H, \succeq_H, q)$ be the set of matchings that respect priority.[5]

[4] With some abuse of notation, we write $j \in \mu(i)$ for both cases: for a hospital h, $\mu(h) \subseteq D$; for a doctor d, we identify the singleton $\{\mu(d)\}$ with $\mu(d)$.

[5] Since $\succ'_h$ refines $\succeq_h$, a matching that respects priority in this sense never assigns a position to a doctor while a doctor with strictly higher legal priority prefers that hospital; the legal priority structure is therefore fully respected. The augmented order additionally settles the claims of doctors whom the law ranks as equals, using the hospital's own preference, as is standard in school choice with coarse priorities. If instead one defines $\mathcal{R}$ using the weak order $\succeq_h$ directly, the set of priority-respecting matchings expands and $DA^{\succeq}$ may fail priority-constrained efficiency, in the manner of Erdil and Ergin (2008); recovering constrained efficiency in that case is incompatible with strategy-proofness (Abdulkadiroğlu, Pathak, and Roth, 2009).

A matching is *doctor-efficient* if there is no other matching that all doctors weakly prefer and at least one doctor strictly prefers. A matching is *efficient* if there is no other matching that all agents weakly prefer and at least one agent strictly prefers. A matching is *priority-constrained efficient* if there is no priority-respecting matching that all agents weakly prefer and at least one agent strictly prefers. Formally:

A matching μ is **efficient** if there is no $\mu' \in \mathcal{M}(P_D, P_H, \succeq_H, q)$ such that for each $i \in I$, $\mu'(i)\, R_i\, \mu(i)$ and there is $i \in I$, $\mu'(i)\, P_i\, \mu(i)$.

A matching μ is **doctor-efficient** if there is no $\mu' \in \mathcal{M}(P_D, P_H, \succeq_H, q)$ such that for each $d \in D$, $\mu'(d)\, R_d\, \mu(d)$ and there is $d \in D$, $\mu'(d)\, P_d\, \mu(d)$.

A matching μ is **priority-constrained efficient** if there is no $\mu' \in \mathcal{R}(P_D, P_H, \succeq_H, q)$ such that for each $i \in I$, $\mu'(i)\, R_i\, \mu(i)$ and there is $i \in I$, $\mu'(i)\, P_i\, \mu(i)$.

Let $\mathcal{DE}$, $\mathcal{E}$, and $\mathcal{PCE}$ denote the sets of doctor-efficient, efficient, and priority-constrained efficient matchings, respectively.

Remark 1. Since the social service is compulsory, a doctor's possible preference for exemption plays no role in our normative analysis. Throughout the paper, in the definitions of *doctor-efficiency*, *efficiency*, and *priority-constrained efficiency*, welfare comparisons for each doctor d are made with respect to the restriction of P_d to H; a doctor who is exempt under two matchings is indifferent between them. Equivalently, we assume $h\, P_d\, \emptyset$ for each $d \in D$ and each $h \in H$ whenever efficiency properties are evaluated, and this convention is maintained in all examples below, except where a doctor's preference for exemption is itself the object of study (Proposition 5 and Remark 2). Doctors' preferences for exemption matter only for the manipulation incentives generated by M^{2012} and M^{2013} (Section 1.1).

Many real-life centralized matching markets employ mechanisms that only ask for preferences over individual partners. Since the social service is compulsory, doctors cannot report whether they wish to be exempt. Hence, for each doctor d, the mechanism focuses on the restriction of P_d to H. We also assume that the hospitals' quotas are commonly known.[6] Throughout the paper, we focus on this class of mechanisms. A **mechanism** $\boldsymbol{\varphi}$ assigns a matching to each problem. We denote agent i's match $\varphi(P_D, P_H, \succeq_H, q)(i)$ by $\varphi_i(P_D, P_H, \succeq_H, q)$, or simply by φ_i when the problem is clear from context.

Next, we describe desirable properties of mechanisms. The first is *respecting priority*, i.e., for each problem, the mechanism assigns a matching that respects priority. Formally, a mechanism φ **respects priority** if for each problem $(P_D, P_H, \succeq_H, q) \in \mathbb{P}$, we have $\varphi(P_D, P_H, \succeq_H, q) \in \mathcal{R}(P_D, P_H, \succeq_H, q)$.

The next set of properties is related to Pareto-efficiency. Formally, a mechanism φ is **efficient** if for each problem $(P_D, P_H, \succeq_H, q) \in \mathbb{P}$, we have $\varphi(P_D, P_H, \succeq_H, q) \in \mathcal{E}(P_D, P_H, \succeq_H, q)$. A mechanism φ is **doctor-efficient** if for each problem $(P_D, P_H, \succeq_H, q) \in \mathbb{P}$, we have $\varphi(P_D, P_H, \succeq_H, q) \in \mathcal{DE}(P_D, P_H, \succeq_H, q)$. A mechanism φ is **priority-constrained efficient** if for each problem $(P_D, P_H, \succeq_H, q) \in \mathbb{P}$, we have $\varphi(P_D, P_H, \succeq_H, q) \in \mathcal{PCE}(P_D, P_H, \succeq_H, q)$.

[6] In contrast to Sönmez (1997), we assume that quotas cannot be manipulated.

The next property is *strategy-proofness*, i.e., no doctor should benefit from misrepresenting her preferences. Formally, a mechanism φ is **strategy-proof** if for each problem $(P_D, P_H, \succeq_H, q) \in \mathbb{P}$, each $d \in D$, and each P'_d, we have

$$\varphi(P_d, P_{-d}, P_H, \succeq_H, q)(d) \; R_d \; \varphi(P'_d, P_{-d}, P_H, \succeq_H, q)(d).$$

The last property is *minimizing unfilled positions*, i.e., for each problem, the mechanism assigns a matching such that there is no doctor-hospital pair where the hospital has unfilled positions but the doctor is exempt. Formally, a mechanism φ **minimizes unfilled positions** if for each problem $(P_D, P_H, \succeq_H, q) \in \mathbb{P}$, there are no $d \in D$ and $h \in H$ such that $|\varphi_h(P_D, P_H, \succeq_H, q)| < q_h$ and $\varphi_d(P_D, P_H, \succeq_H, q) \notin H$.

Next, we define the mechanisms employed in many real-life problems. We start with the mechanisms used in Colombia by the Ministry of Health. To facilitate comparison, we make two changes: first, we describe the mechanisms in terms of hospitals rather than states (*departamentos*); second, we describe both mechanisms as *if* they incorporated priorities, which allows us to show that doing so is still insufficient to satisfy *respecting priority*.

Algorithm 1. [Ministry of Health – 2012, M^{2012}]

Step 1: Each doctor enrolls in a hospital.

Step 2: At each hospital, a lottery assigns doctors to positions. Doctors not assigned to any hospital are exempt from the compulsory social service.

For each problem $(P_D, P_H, \succeq_H, q) \in \mathbb{P}$, let $M^{2012}(P_D, P_H, \succeq_H, q)$ be the matching resulting from Algorithm 1.

The above mechanism was modified in January 2013 by the Ministry of Health.

Algorithm 2. [Ministry of Health – 2013, M^{2013}]

Step 1: Each doctor enrolls in a hospital. However, the number of enrolled doctors at each hospital cannot exceed twice the total available positions in that hospital. Enrollment is on a first-come first-served basis.

Step 2: At each hospital, a lottery assigns doctors to positions. Doctors not assigned to any hospital are exempt from the compulsory social service.

For each problem $(P_D, P_H, \succeq_H, q) \in \mathbb{P}$, let $M^{2013}(P_D, P_H, \succeq_H, q)$ be the matching resulting from Algorithm 2. Unlike the mechanisms introduced below, M^{2012} and M^{2013} elicit a single enrollment rather than a ranking. To evaluate them within our framework, we assume that each doctor enrolls at the top-ranked hospital of her reported ranking. Since both mechanisms involve lotteries, we say that M^{2012} (respectively, M^{2013}) fails a property if the property fails for some realization of the lottery.

The next mechanism is Deferred Acceptance with priority ($DA^{\succeq}$), based on Gale and Shapley's (1962) deferred acceptance algorithm.[7]

Algorithm 3. [Deferred Acceptance with priority, $DA^{\succeq}$]

Step 1: Each doctor proposes to her preferred hospital. Each hospital tentatively assigns positions to its proposers one at a time following the augmented priority order $\succ'_h$. All remaining proposers are rejected.

$\vdots$

Step k: Each doctor rejected in Step $k-1$ proposes to the next hospital in her preference list. Each hospital considers the doctors currently tentatively assigned (from all prior steps) together with its new proposers from Step k. It tentatively assigns positions following the augmented priority order $\succ'_h$. All remaining proposers are rejected. A doctor rejected by all hospitals is exempt from the compulsory social service.

The algorithm terminates when each doctor is assigned to a hospital or exempt from the compulsory social service. For each problem $(P_D, P_H, \succeq_H, q) \in \mathbb{P}$, let $DA^{\succeq}(P_D, P_H, \succeq_H, q)$ be the matching resulting from Algorithm 3.

The next mechanism is Immediate Acceptance (IA), also known as the Boston mechanism (Abdulkadiroğlu and Sönmez, 2003).

Algorithm 4. [Immediate Acceptance, IA]

Step 1: Each doctor proposes to her preferred hospital. Each hospital *permanently* assigns positions to its proposers one at a time following the augmented priority order $\succ'_h$. All remaining proposers are rejected.

$\vdots$

Step k: Each doctor rejected in Step $k-1$ proposes to the next hospital in her preference list. Each hospital permanently assigns its remaining unfilled positions to its proposers one at a time following the augmented priority order $\succ'_h$. All remaining proposers are rejected. A doctor rejected by all hospitals is exempt from the compulsory social service.

The algorithm terminates when each doctor is assigned to a hospital or exempt from the compulsory social service. For each problem $(P_D, P_H, \succeq_H, q) \in \mathbb{P}$, let $IA(P_D, P_H, \succeq_H, q)$ be the matching resulting from Algorithm 4.

Finally, we introduce the Top Trading Cycle mechanism introduced by David Gale in Shapley and Scarf (1974). Abdulkadiroğlu and Sönmez (2003) adapt this mechanism for problems with priorities; we extend it to problems with preferences and priorities.

Algorithm 5. [Top Trading Cycles, TTC]

[7] See Roth (2008) for an account of the history and applications of the deferred acceptance algorithm.

Step 1: Assign a counter to each hospital tracking remaining unfilled positions; initially set to the hospital's quota. Each doctor points to her preferred hospital. Each hospital points to the doctor with the highest priority (if two doctors are tied in priority, the hospital's strict preference breaks the tie). Since the sets of doctors and hospitals are finite, there is at least one cycle. (A *cycle* is an ordered list of distinct hospitals and distinct doctors $(d_1, h_1, d_2, \ldots, d_k, h_k)$ where d_1 points to h_1, h_1 points to d_2, ..., d_k points to h_k, and h_k points to d_1.) Moreover, each doctor and each hospital can be part of at most one cycle. Every doctor in a cycle is assigned a position at the hospital she points to and is removed. The counter of each hospital in a cycle is reduced by one; if it reaches zero, the hospital is removed. Counters of other hospitals remain unchanged.

$\vdots$

Step k: Each remaining doctor points to her preferred hospital among the remaining ones; each remaining hospital points to the doctor with the highest priority among the remaining ones (with hospital preferences breaking ties in priority). There is a cycle. Every doctor in a cycle is assigned a position at the hospital she points to and is removed. The counter of each hospital in a cycle is reduced by one; if it reaches zero, the hospital is removed. Counters of other hospitals remain unchanged. A doctor never part of any cycle is exempt from the compulsory social service.

The algorithm terminates when each doctor is assigned to a hospital or exempt from the compulsory social service. For each problem $(P_D, P_H, \succeq_H, q) \in \mathbb{P}$, let $TTC(P_D, P_H, \succeq_H, q)$ be the matching resulting from Algorithm 5.

3 Results

First, we present our results about the set of matchings. Figure 3 summarizes the relationships between sets of matchings.

Proposition 1. *The set of doctor-efficient matchings is a subset of the set of efficient matchings, i.e., for each problem* $(P_D, P_H, \succeq_H, q) \in \mathbb{P}$, $\mathcal{DE}(P_D, P_H, \succeq_H, q) \subseteq \mathcal{E}(P_D, P_H, \succeq_H, q)$. *However, the converse is not true in general.*

Proof. First, we show that any doctor-efficient matching is efficient. Let $(P_D, P_H, \succeq_H, q) \in \mathbb{P}$ be a problem and $\mu \in \mathcal{DE}(P_D, P_H, \succeq_H, q)$ be a doctor-efficient matching. Assume for contradiction that μ is not efficient. Then there is $\mu' \in \mathcal{M}(P_D, P_H, \succeq_H, q)$ such that for each $i \in I$, $\mu'(i)\, R_i\, \mu(i)$ and there is $i \in I$, $\mu'(i)\, P_i\, \mu(i)$. In particular, for each $d \in D$, $\mu'(d)\, R_d\, \mu(d)$. If $i = d \in D$, then $\mu'(d)\, R_d\, \mu(d)$ and $\mu'(d)\, P_d\, \mu(d)$, immediately contradicting doctor-efficiency. If $i = h \in H$, then $\mu'(h) \neq \mu(h)$, so by the consistency condition (3) of the matching definition there is $d \in D$ with $\mu'(d) \neq \mu(d)$. Since $\mu'(d)\, R_d\, \mu(d)$ and preferences are strict, $\mu'(d)\, P_d\, \mu(d)$, contradicting doctor-efficiency. Hence $\mu \in \mathcal{E}(P_D, P_H, \succeq_H, q)$.

The converse is not true. To see this, consider a problem with 2 doctors, 2 hospitals, and preferences and priorities given in Table 2. Each hospital h has quota $q_h = 1$. Since the priorities

in this and the following examples are strict, the augmented order coincides with the priority order: for each $h \in H$, we have $\succ'_h = \succ_h$.

Doctors		Hospitals			
P_{d_1}	P_{d_2}	$\succ_{h_1}$	$\succ_{h_2}$	P_{h_1}	P_{h_2}
h_1	h_2	d_1	d_2	d_2	d_1
h_2	h_1	d_2	d_1	d_1	d_2

Table 2: Preferences P and priorities $\succ$ in Proposition 1.

Note that $\mu = \begin{pmatrix} h_1 & h_2 \\ d_2 & d_1 \end{pmatrix}$ is *efficient* but not *doctor-efficient.* □

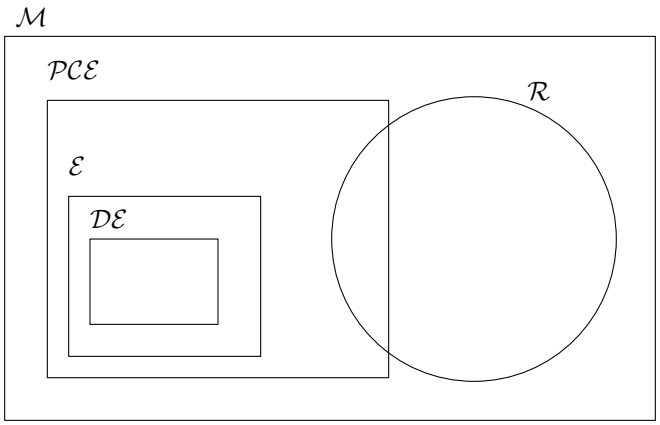


Figure 3: The relationships of the set of matchings.

Proposition 2. *The set of* efficient *matchings is a subset of the set of* priority-constrained efficient *matchings, i.e., for each problem* $(P_D, P_H, \succeq_H, q) \in \mathbb{P}$*,*

$$\mathcal{E}(P_D, P_H, \succeq_H, q) \subseteq \mathcal{PCE}(P_D, P_H, \succeq_H, q).$$

However, the converse is not true.

Proof. First, we show that any efficient matching is priority-constrained efficient. Let $(P_D, P_H, \succeq_H, q) \in \mathbb{P}$ be a problem and $\mu \in \mathcal{E}(P_D, P_H, \succeq_H, q)$ be an efficient matching. Assume for contradiction that μ is not priority-constrained efficient. Then there is $\mu' \in \mathcal{R}(P_D, P_H, \succeq_H, q)$ such that for each $i \in I$, $\mu'(i)\, R_i\, \mu(i)$ and there is $i \in I$, $\mu'(i)\, P_i\, \mu(i)$. Since $\mathcal{R}(P_D, P_H, \succeq_H, q) \subseteq \mathcal{M}(P_D, P_H, \succeq_H, q)$, this contradicts the fact that μ is efficient. Hence $\mu \in \mathcal{PCE}(P_D, P_H, \succeq_H, q)$.

The converse is not true. To see this, consider a problem with 3 doctors, 3 hospitals, and preferences and priorities given in Table 3. Each hospital h has quota $q_h = 1$.

Doctors			Hospitals					
P_{d_1}	P_{d_2}	P_{d_3}	$\succ_{h_1}$	$\succ_{h_2}$	$\succ_{h_3}$	P_{h_1}	P_{h_2}	P_{h_3}
h_2	h_1	h_1	d_1	d_2	d_2	d_2	d_1	d_3
h_1	h_2	h_2	d_3	d_1	d_1	d_3	d_2	d_2
h_3	h_3	h_3	d_2	d_3	d_3	d_1	d_3	d_1

Table 3: Preferences P and priorities $\succ$ in Proposition 2.

By Remark 1, any matching that leaves a doctor exempt is dominated by a matching that additionally assigns her to an unfilled hospital, so it suffices to consider the matchings in which every doctor is matched:

$$\mu^1 = \begin{pmatrix} h_1 & h_2 & h_3 \\ d_2 & d_1 & d_3 \end{pmatrix}, \quad \mu^2 = \begin{pmatrix} h_1 & h_2 & h_3 \\ d_3 & d_1 & d_2 \end{pmatrix}, \quad \mu^3 = \begin{pmatrix} h_1 & h_2 & h_3 \\ d_3 & d_2 & d_1 \end{pmatrix},$$

$$\mu^4 = \begin{pmatrix} h_1 & h_2 & h_3 \\ d_2 & d_3 & d_1 \end{pmatrix}, \quad \mu^5 = \begin{pmatrix} h_1 & h_2 & h_3 \\ d_1 & d_3 & d_2 \end{pmatrix}, \quad \mu^6 = \begin{pmatrix} h_1 & h_2 & h_3 \\ d_1 & d_2 & d_3 \end{pmatrix}.$$

Note that under μ^6, doctor d_1 receives h_1 (second choice), d_2 receives h_2 (second choice), and d_3 receives h_3 (third choice). Matching μ^1 makes d_1 and d_2 strictly better off, so $\mu^6 \notin \mathcal{E}(P_D, P_H, \succeq_H, q)$. To see that $\mu^6 \in \mathcal{PCE}(P_D, P_H, \succeq_H, q)$, note that the only matchings that make no doctor worse off than under μ^6 are μ^6 itself and μ^1 (any other full matching moves some doctor below her assignment under μ^6), and μ^1 does not respect priority (Proposition 3). Hence, no priority-respecting matching dominates μ^6. □

Proposition 3. *The intersection of the set of* doctor-efficient *matchings and the set of matchings that respect priority may be empty. Similarly, the intersection of the set of* efficient *matchings and the set of matchings that respect priority may be empty.*

Proof. Consider the problem in Proposition 2. Figure 4 illustrates the relationships between the sets of matchings for this problem. Under μ^1, no alternative matching makes all doctors at least as well off while making one strictly better off, so μ^1 is doctor-efficient. An analogous argument applies to μ^2, μ^3, and μ^4. For μ^3 and μ^4, note that any chain of improvements must eventually displace a doctor from her first-choice hospital, which makes her strictly worse off. The matching μ^1 does not respect priority because $d_3 \succ_{h_1} d_2$ yet d_3 prefers h_1 to h_3, and h_1 is matched to d_2. The same holds for μ^2, μ^3, and μ^4. Note also that μ^5 and μ^6 are *not* efficient: μ^1 dominates μ^6 and μ^2 dominates μ^5. Since doctor-efficient implies efficient (Proposition 1), μ^1, μ^2, μ^3, and μ^4 are efficient. Therefore, $\mathcal{E}(P_D, P_H, \succeq_H, q) \cap \mathcal{R}(P_D, P_H, \succeq_H, q) = \emptyset$. □

Proposition 4. *The intersection of the set of* priority-constrained efficient *matchings and the set*

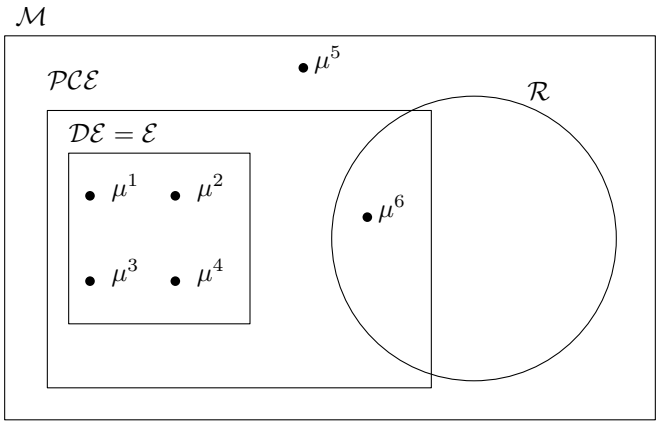


Figure 4: The relationships of the set of matchings for the problem in Proposition 2.

of matchings that respect priority is non-empty for every problem, i.e., for each $(P_D, P_H, \succeq_H, q) \in \mathbb{P}$,

$$\mathcal{PCE}(P_D, P_H, \succeq_H, q) \cap \mathcal{R}(P_D, P_H, \succeq_H, q) \neq \emptyset.$$

Moreover, $DA^{\succeq}(P_D, P_H, \succeq_H, q)$ *always belongs to this intersection. However,* $\mathcal{R}$ *is not always a subset of* $\mathcal{PCE}$.

Proof. $DA^{\succeq}$ *respects priority.* Suppose not. Then there exist $(P_D, P_H, \succeq_H, q)$, $d \in D$, and $h \in H$ such that $h\, P_d\, DA_d^{\succeq}(P_D, P_H, \succeq_H, q)$ and either $|DA_h^{\succeq}(P_D, P_H, \succeq_H, q)| < q_h$ or there is $d' \in DA_h^{\succeq}(P_D, P_H, \succeq_H, q)$ with $d \succ'_h d'$. Since d prefers h to her assignment, d proposed to h at some step and was rejected. A hospital with an unfilled position never rejects a proposer, so at that step h was full, holding q_h doctors all above d in $\succ'_h$. Because a hospital's tentative assignment only improves in $\succ'_h$ over subsequent steps, every $d' \in DA_h^{\succeq}(P_D, P_H, \succeq_H, q)$ satisfies $d' \succ'_h d$, and $|DA_h^{\succeq}(P_D, P_H, \succeq_H, q)| = q_h$—contradicting both cases.

$DA^{\succeq}$ *is priority-constrained efficient.* Let $\mu := DA^{\succeq}(P_D, P_H, \succeq_H, q)$. By Remark 1, every doctor prefers every hospital to exemption when welfare is evaluated, and every hospital finds every doctor acceptable. Consider the college-admissions market in which each hospital h ranks individual doctors by the strict order $\succ'_h$, with quota q_h and a responsive extension. In this market individual rationality is vacuous, so a matching is stable if and only if it respects priority; that is, $\mathcal{R}(P_D, P_H, \succeq_H, q)$ is exactly the set of stable matchings. Since μ is the outcome of the doctor-proposing deferred acceptance algorithm run on $\succ'_H$, every doctor finds μ at least as desirable as any stable matching (Gale and Shapley, 1962; Roth and Sotomayor, 1990).

Now suppose, for contradiction, that μ is not priority-constrained efficient: there is $\mu' \in \mathcal{R}(P_D, P_H, \succeq_H, q)$ with $\mu'(i)\, R_i\, \mu(i)$ for each $i \in I$ and $\mu'(i)\, P_i\, \mu(i)$ for some $i \in I$. For each doctor d we have $\mu'(d)\, R_d\, \mu(d)$ by assumption and $\mu(d)\, R_d\, \mu'(d)$ by doctor-optimality; since preferences are strict, $\mu'(d) = \mu(d)$ for each $d \in D$. By condition (3) of the definition of a matching, μ' and μ then assign the same set of doctors to each hospital, so $\mu' = \mu$—contradicting $\mu'(i)\, P_i\, \mu(i)$ for some i. Hence $\mu \in \mathcal{PCE}(P_D, P_H, \succeq_H, q)$, and together with the first part, $\mathcal{PCE} \cap \mathcal{R} \neq \emptyset$.

$\mathcal{R}(P_D, P_H, \succeq_H, q)$ is not always a subset of $\mathcal{PCE}(P_D, P_H, \succeq_H, q)$. Consider the problem in Table 4 with 3 doctors, 3 hospitals, and $q_h = 1$ for all h.

Doctors			Hospitals					
P_{d_1}	P_{d_2}	P_{d_3}	$\succ_{h_1}$	$\succ_{h_2}$	$\succ_{h_3}$	P_{h_1}	P_{h_2}	P_{h_3}
h_1	h_2	h_3	d_2	d_3	d_1	d_1	d_2	d_3
h_2	h_3	h_1	d_3	d_1	d_2	d_3	d_1	d_2
h_3	h_1	h_2	d_1	d_2	d_3	d_2	d_3	d_1

Table 4: Preferences P and priorities $\succ$ in Proposition 4.

As in Proposition 2, it suffices to consider the matchings in which every doctor is matched:

$$\nu^1 = \begin{pmatrix} h_1 & h_2 & h_3 \\ d_2 & d_3 & d_1 \end{pmatrix}, \quad \nu^2 = \begin{pmatrix} h_1 & h_2 & h_3 \\ d_1 & d_2 & d_3 \end{pmatrix}, \quad \nu^3 = \begin{pmatrix} h_1 & h_2 & h_3 \\ d_3 & d_1 & d_2 \end{pmatrix}$$

$$\nu^4 = \begin{pmatrix} h_1 & h_2 & h_3 \\ d_1 & d_3 & d_2 \end{pmatrix}, \quad \nu^5 = \begin{pmatrix} h_1 & h_2 & h_3 \\ d_2 & d_1 & d_3 \end{pmatrix}, \quad \nu^6 = \begin{pmatrix} h_1 & h_2 & h_3 \\ d_3 & d_2 & d_1 \end{pmatrix}.$$

One can verify that ν^1, ν^2, and ν^3 respect priority. To check ν^1: the only potential blockers are doctors who prefer a hospital where a lower-priority doctor is seated. Doctor d_1 prefers h_1 (seated: d_2 with $\succ_{h_1}$: $d_2 \succ d_3 \succ d_1$), so $d_2 \succ_{h_1} d_1$ and no violation occurs. All other pairs are checked similarly. The matchings ν^2 and ν^3 respect priority by analogous inspection of the priority columns in Table 4. However, ν^1 and ν^3 are not priority-constrained efficient: under ν^2 every agent *strictly* prefers its assignment (ν^2 gives each doctor her first-choice hospital and each hospital its most preferred doctor). The remaining matchings do not respect priority: the pair (d_3, h_1) blocks ν^4, the pair (d_2, h_3) blocks ν^5, and the pair (d_1, h_2) blocks ν^6. Figure 5 illustrates the relationships between the sets of matchings for this problem. □

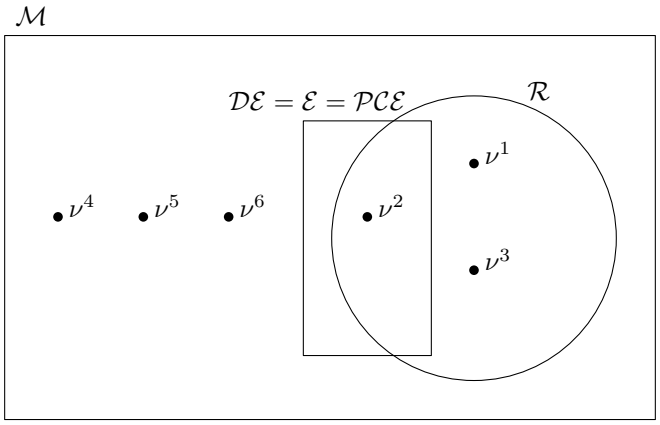


Figure 5: The relationships of the set of matchings in Proposition 4.

Next, we present our results regarding the properties of the mechanisms.

Proposition 5. *The mechanisms M^{2012} and M^{2013} are not* doctor-efficient*, not* efficient*, not* priority-constrained efficient*, do not* respect priority*, are not* strategy-proof*, and do not* minimize unfilled positions*.*

Proof. Consider a problem with 2 doctors, 2 hospitals, and preferences and priorities given in Table 5. Each hospital h has quota $q_h = 1$. We assume both hospitals share the same priority order $\succ_{h_1} = \succ_{h_2} = \succ_h$.

Doctors		Hospitals		
P_{d_1}	P_{d_2}	$\succ_h$	P_{h_1}	P_{h_2}
$\boxed{h_1}$	h_1	d_1	$\boxed{d_1}$	d_1
h_2	h_2	d_2	d_2	d_2
$\emptyset$	$\boxed{\emptyset}$		$\emptyset$	$\boxed{\emptyset}$

Table 5: Preferences P and priorities $\succ$ in Proposition 5.

By the convention introduced after Algorithm 2, both doctors enroll at their top-ranked hospital, h_1. Suppose the lottery at h_1 selects d_1 (the boxed entries in Table 5), so that M^{2012} and M^{2013} yield the matching:

$$\mu = \begin{pmatrix} h_1 & h_2 & \emptyset \\ d_1 & \emptyset & d_2 \end{pmatrix}.$$

Not priority-constrained efficient: There is a matching $\mu' = \begin{pmatrix} h_1 & h_2 \\ d_1 & d_2 \end{pmatrix}$ that respects priority and makes each agent at least as well off, with d_2 strictly preferring μ'. Hence M^{2012} and M^{2013} are not priority-constrained efficient, doctor-efficient, or efficient.
Do not respect priority: d_2 prefers h_2 to being exempt, and h_2 has an unfilled position.
Not strategy-proof: Suppose d_2 reports $P'_{d_2} = h_2, h_1$. Then M^{2012} and M^{2013} yield $\mu' = \begin{pmatrix} h_1 & h_2 \\ d_1 & d_2 \end{pmatrix}$. Since $\mu'(d_2)\, P_{d_2}\, \mu(d_2)$, the deviation P'_{d_2} is profitable for d_2.
Do not minimize unfilled positions: h_2 is unmatched ($|\mu(h_2)| = 0 < 1 = q_{h_2}$) while d_2 is also exempt.
Not immune to exemption-seeking: Under M^{2012} and M^{2013}, a doctor who ranks a hospital whose enrollment exceeds its quota raises the probability that she is exempt; in Table 5, doctor d_2 is exempt under the realization above but is assigned under the report $P'_{d_2} = h_2, h_1$. Hence a doctor's report affects whether she is exempt. □

Proposition 6. *The mechanisms IA and TTC do not* respect priority*.*

Proof. Consider the problem in Proposition 2. We first compute $IA(P_D, P_H, \succeq_H, q)$. In Step 1, d_1 proposes to h_2 and d_2, d_3 propose to h_1. Hospital h_2 permanently assigns its position to d_1, its only proposer. Hospital h_1, whose priority order is $d_1 \succ'_{h_1} d_3 \succ'_{h_1} d_2$, permanently assigns its position to

d_3 and rejects d_2. In Step 2, d_2 proposes to h_2, which is full, and is rejected. In Step 3, d_2 proposes to h_3 and is assigned. Hence $IA(P_D, P_H, \succeq_H, q) = \mu^2 = \left(\begin{smallmatrix} h_1 & h_2 & h_3 \\ d_3 & d_1 & d_2 \end{smallmatrix}\right)$. This matching does not respect priority: doctor d_2 prefers h_2 to $\mu^2(d_2) = h_3$, hospital h_2 is full with d_1, and $d_2 \succ'_{h_2} d_1$.

Next, we compute $TTC(P_D, P_H, \succeq_H, q)$. In Step 1, d_1 points to h_2 while d_2 and d_3 point to h_1; hospitals h_1, h_2, and h_3 point to d_1, d_2, and d_2, respectively. The cycle (d_1, h_2, d_2, h_1) forms: d_1 is assigned to h_2 and d_2 to h_1, and both hospitals are removed. In Step 2, d_3 points to h_3, which points to d_3, and d_3 is assigned to h_3. Hence $TTC(P_D, P_H, \succeq_H, q) = \mu^1 = \left(\begin{smallmatrix} h_1 & h_2 & h_3 \\ d_2 & d_1 & d_3 \end{smallmatrix}\right)$. This matching does not respect priority: doctor d_3 prefers h_1 to $\mu^1(d_3) = h_3$, hospital h_1 is full with d_2, and $d_3 \succ'_{h_1} d_2$. □

Proposition 7. *The mechanisms IA and TTC are* doctor-efficient. *Hence, they are also* efficient *and* priority-constrained efficient.

Proof. By construction, IA and TTC are the immediate acceptance and top trading cycles mechanisms applied to the strict priority profile $\succ'_H$, and by Remark 1, doctors' welfare is evaluated with respect to their rankings of H, with every hospital preferred to exemption. For this environment, Abdulkadiroğlu and Sönmez (2003) show that both mechanisms select a matching that is Pareto efficient for the doctors; hence IA and TTC are *doctor-efficient*. By Proposition 1, every doctor-efficient matching is *efficient*, and by Proposition 2, every efficient matching is *priority-constrained efficient*. □

Proposition 8. *The mechanism $DA^{\succeq}$ is neither* doctor-efficient *nor* efficient.

Proof. Consider the problem in Proposition 2. In Step 1 of the deferred acceptance algorithm, d_1 proposes to h_2 while d_2 and d_3 propose to h_1; hospital h_1 tentatively holds d_3 and rejects d_2 (since $d_1 \succ'_{h_1} d_3 \succ'_{h_1} d_2$). In Step 2, d_2 proposes to h_2, which holds d_2 and rejects d_1 (since $d_2 \succ'_{h_2} d_1$). In Step 3, d_1 proposes to h_1, which holds d_1 and rejects d_3. In Step 4, d_3 proposes to h_2 and is rejected (since $d_2 \succ'_{h_2} d_3$). In Step 5, d_3 proposes to h_3 and is held, and the algorithm terminates with $DA^{\succeq}(P_D, P_H, \succeq_H, q) = \mu^6$. As shown in Proposition 3, μ^1 dominates μ^6: doctors d_1 and d_2 receive their first-choice hospitals and d_3's assignment is unchanged, while h_1 and h_2 receive their most preferred doctors and h_3's assignment is unchanged. Hence μ^6 is neither doctor-efficient nor efficient. □

Under M^{2012} and M^{2013}, a professional who preferred exemption to service could secure it by enrolling in a congested state, and groups of professionals coordinated to do exactly that (Section 1.1). *Immunity to exemption-seeking* means that no doctor can change whether she is exempt by misreporting her preferences. Since a doctor reports only a ranking of hospitals and never reports whether she would rather be exempt, the question is whether the same loophole survives under the mechanisms we study: can a doctor obtain exemption by permuting her reported ranking? The next proposition answers this in the negative for $DA^{\succeq}$ and TTC.

Proposition 9. *Let $(P_D, P_H, \succeq_H, q) \in \mathbb{P}$ be a problem and $d \in D$ be a doctor. Under $DA^{\succeq}$ and under TTC, whether d is exempt does not depend on d's reported ranking of H, holding the*

reports of the other doctors fixed. Consequently, no doctor can obtain exemption by misreporting her preferences.

Proof. $DA^{\succeq}$. Let Q_d and Q'_d be two rankings of H and let μ and ν be the corresponding outcomes. Suppose $\nu(d) = \emptyset$. Since d ranks every hospital, she proposed to each of them and was rejected. A hospital rejects a proposer only while holding q_h doctors that are all above her in $\succ'_h$, and its tentative assignment only improves in $\succ'_h$ over subsequent steps. Hence, under ν, every hospital h is full and every $d' \in \nu(h)$ satisfies $d' \succ'_h d$.

Consider the market in which each hospital h ranks doctors by $\succ'_h$ with quota q_h, all doctors are acceptable to all hospitals, and doctor d reports Q_d while the others report as given. The matching ν is stable in this market: a pair (d, h) cannot block, because h is full with doctors above d in $\succ'_h$—a fact that does not involve d's ranking at all—and no pair among the other agents can block, because their reports and assignments are the same as in the market where ν was computed. The matching μ is stable in this market as well, being its deferred acceptance outcome. By the Rural Hospital Theorem (Roth, 1986), the same set of doctors is matched in every stable matching of a given market; since d is unmatched under ν, she is unmatched under μ. Exchanging the roles of Q_d and Q'_d gives the converse.

TTC. Suppose d is matched under Q_d, and let t be the step at which she is removed and g the hospital pointing to her in that cycle. Run the algorithm under Q'_d and suppose, for contradiction, that d is never part of a cycle. At any configuration of remaining agents, the cycles that do not contain d depend only on $\succ'_H$ and on the other doctors' reports, so they are identical in the two runs. Since d belongs to no cycle under Q'_d, and to none before step t under Q_d, the two runs execute the same cycles up to step t and reach the same configuration there. In particular g points to d at step t under Q'_d, and continues to do so at every later step: all doctors above d in $\succ'_g$ have been removed, removals are permanent, and d remains by assumption. Every cycle containing g therefore contains d; as d is in no cycle, g is in no cycle, its counter never decreases, and g is never removed. But while doctors and hospitals both remain the pointing graph contains a cycle, so agents keep being removed until one side is exhausted; since d is never removed, every hospital is eventually removed—contradicting that g is not. Hence d is matched under Q'_d as well. □

Remark 2. *Proposition 9 does not extend to IA.* Under the immediate acceptance mechanism, a doctor can obtain exemption by misreporting. Let $D = \{d_1, d_2, d_3\}$, $H = \{h_1, h_2\}$, $q_{h_1} = q_{h_2} = 1$, with augmented priorities $\succ'_{h_1} : d_1, d_3, d_2$ and $\succ'_{h_2} : d_2, d_1, d_3$, and preferences

$$P_{d_1} : h_2,\, h_1,\, \emptyset; \qquad P_{d_2} : h_1,\, h_2,\, \emptyset; \qquad P_{d_3} : \emptyset,\, h_1,\, h_2.$$

Under truthful reporting, in Step 1 doctor d_1 is assigned to h_2, and d_2 and d_3 apply to h_1, which permanently assigns its position to d_3; in Step 2, d_2 applies to the full hospital h_2 and is exempt. The outcome is

$$\begin{pmatrix} h_1 & h_2 & \emptyset \\ d_3 & d_1 & d_2 \end{pmatrix}.$$

Now, let d_3 report $Q_{d_3} : h_2,\, h_1$. In Step 1, d_1 and d_3 apply to h_2, which assigns its position to

d_1, while d_2 applies to h_1 and is assigned; in Step 2, d_3 applies to the full hospital h_1 and is exempt. The outcome is

$$\begin{pmatrix} h_1 & h_2 & \emptyset \\ d_2 & d_1 & d_3 \end{pmatrix}.$$

Since $\emptyset \, P_{d_3} \, h_1$, the deviation is profitable, and it yields exemption from the compulsory social service—the same outcome professionals obtained by manipulation under the 2012 and 2013 mechanisms.

Proposition 10. *The mechanism $DA^{\succeq}$ is* strategy-proof.

Proof. Let $(P_D, P_H, \succeq_H, q) \in \mathbb{P}$ be a problem, $d \in D$ be a doctor, and P'_d be a report; only the restrictions of P_d and P'_d to H enter the mechanism, and both are complete rankings of H. Let $\mu := DA^{\succeq}(P_d, P_{-d}, P_H, \succeq_H, q)$ and $\mu' := DA^{\succeq}(P'_d, P_{-d}, P_H, \succeq_H, q)$. By Proposition 9, $\mu(d) = \emptyset$ if and only if $\mu'(d) = \emptyset$; in that case $\mu(d) \, R_d \, \mu'(d)$ holds trivially, so suppose $\mu(d), \mu'(d) \in H$. By construction, $DA^{\succeq}$ is the doctor-proposing deferred acceptance mechanism of the market in which each hospital h ranks individual doctors by the strict order $\succ'_h$—which does not depend on the doctors' reports—has quota q_h, and finds every doctor acceptable. In that market, no doctor can obtain a hospital she strictly prefers by submitting a ranking other than her true one (Dubins and Freedman, 1981; Roth, 1982; see Roth, 1985, for the many-to-one case with responsive preferences). Since $\mu(d)$ and $\mu'(d)$ are both hospitals, it follows that $\mu(d) \, R_d \, \mu'(d)$. □

Proposition 11. *The mechanism TTC is* strategy-proof.

Proof. Let $(P_D, P_H, \succeq_H, q) \in \mathbb{P}$ be a problem, $d \in D$ be a doctor, and P'_d be a report. Let $\mu := TTC(P_d, P_{-d}, P_H, \succeq_H, q)$ and $\mu' := TTC(P'_d, P_{-d}, P_H, \succeq_H, q)$. By Proposition 9, $\mu(d) = \emptyset$ if and only if $\mu'(d) = \emptyset$, in which case there is nothing to show; so suppose $\mu(d), \mu'(d) \in H$. By construction, TTC is the top trading cycles mechanism with counters of Abdulkadiroğlu and Sönmez (2003) applied to the strict priority profile $\succ'_H$, which does not depend on the doctors' reports; for that mechanism, no doctor can obtain a hospital she strictly prefers by submitting a ranking other than her true one. Since $\mu(d)$ and $\mu'(d)$ are both hospitals, $\mu(d) \, R_d \, \mu'(d)$. □

Proposition 12. *The mechanism IA is not* strategy-proof.

Proof. Consider the problem in Proposition 2. As computed in the proof of Proposition 6, $IA(P_D, P_H, \succeq_H, q) = \mu^2$. Suppose d_2 reports $P'_{d_2} = h_2, h_1, h_3$. The resulting outcome is $\mu^3 = IA(P'_{d_2}, P_{D \setminus \{d_2\}}, P_H, \succeq_H, q)$. Since $\mu^3(d_2) \, P_{d_2} \, \mu^2(d_2)$, the deviation P'_{d_2} is profitable for d_2, violating strategy-proofness. □

Proposition 13. *The mechanisms $DA^{\succeq}$, IA, and TTC* minimize unfilled positions.

Proof. Suppose not. Then there exist $(P_D, P_H, \succeq_H, q)$, $h \in H$ with $|\varphi_h(P_D, P_H, \succeq_H, q)| < q_h$, and $d \in D$ with $\varphi_d(P_D, P_H, \succeq_H, q) = \emptyset$.

- $DA^{\succeq}$: Since $\varphi_d(P_D, P_H, \succeq_H, q) = \emptyset$, doctor d proposed to and was rejected by every hospital; in particular, h rejected d at some step. A hospital with an unfilled position never rejects a proposer, so h was full at that step; and since a hospital's set of tentatively assigned doctors never shrinks over the steps, h is full at termination, i.e., $|\varphi_h(P_D, P_H, \succeq_H, q)| = q_h$—contradiction.
- IA: The same argument applies, with *permanent* in place of *tentative* assignments.
- TTC: Since $\varphi_d(P_D, P_H, \succeq_H, q) = \emptyset$, doctor d is never part of a cycle, so d remains until the algorithm terminates. The algorithm terminates only when every doctor or every hospital has been removed; since d remains, every hospital is removed. A hospital is removed only when its counter reaches zero, i.e., when all of its positions are filled. Hence $|\varphi_h(P_D, P_H, \succeq_H, q)| = q_h$—contradiction.

□

Table 6 summarizes all results.

Property	M^{2012}	M^{2013}	$DA^{\succeq}$	IA	TTC
Doctor-efficiency	−	−	−	+	+
Efficiency	−	−	−	+	+
Priority-constrained efficiency	−	−	+	+	+
Respecting priority	−	−	+	−	−
Strategy-proofness	−	−	+	−	+
Minimizing unfilled positions	−	−	+	+	+
Immunity to exemption-seeking	−	−	+	−	+

Table 6: Summary of the results. A + (−) indicates that the mechanism satisfies (does not satisfy) the property.

4 Conclusion

Our results provide a theoretical foundation for the reform of the compulsory social service allocation in Colombia. We show that the lottery-based mechanisms used by the Ministry of Health in 2012 and 2013 fail every desirable property we consider: they are not *priority-constrained efficient*, they are not *strategy-proof*, they do not *respect priority*, and they do not *minimize unfilled positions*. The documented failures were not merely theoretical: in the July 2012 allocation, nearly 600 physicians were exempted from service even though approximately 300 hospital positions remained unfilled, and coordinated gaming was pervasive across multiple quarterly rounds.

Among the three standard mechanisms we examine, the deferred acceptance algorithm with priority, $DA^{\succeq}$, is the only one that simultaneously respects priority, is priority-constrained efficient, is strategy-proof, and minimizes unfilled positions. These properties are especially important in the Colombian context, where priority groups—indigenous professionals, *raizal* professionals, mothers

with small children, and professionals with disabilities—must be protected by law, and where unfilled rural positions represent a direct failure to extend healthcare access to underserved populations. Empirical estimates from the first centralized allocation (October 2014) confirm that adopting $DA^{\succeq}$ eliminated strategic manipulation and produced significant welfare gains (Vélez-Ferro, 2019).

A further advantage of $DA^{\succeq}$ over the mechanisms used in 2012 and 2013 is that it forecloses the strategy that generated the congestion documented above: under $DA^{\succeq}$—and under TTC, but not under IA—a professional who would rather be exempt cannot bring exemption about by reordering her reported list (Proposition 9). Whether a professional serves is determined by the priorities and by the other professionals' reports, not by her own report.

The immediate acceptance (Boston) mechanism and the top trading cycle mechanism achieve doctor-efficiency, which $DA^{\succeq}$ does not. However, they do so at the cost of failing to respect priority and, in the case of IA, at the cost of strategy-proofness. Given that priority compliance is a legal requirement in Colombia, these trade-offs weigh in favor of $DA^{\succeq}$.

Our analysis takes the tie-breaking rule as given. What our results require of it is only that it be strict and independent of the professionals' reports. Vélez-Ferro (2019) finds that breaking ties in favor of directly reported hospitals, rather than those implied by a ranking of states, would raise welfare by 8–12 % and cut rejection rates to 2–5 %. Our results are invariant to which strict refinement of $\succeq_h$ is adopted, provided the refinement is determined by the primitives rather than by the professionals' reports. The properties established in Section 3 depend on $\succ'_h$ only through its strictness and its independence of P_D. The Ministry may therefore calibrate the tie-breaking rule on empirical grounds without sacrificing them. A rule that favors directly reported hospitals is different. It makes the order depend on the reports, and the incentive results would have to be proved again. This is the familiar tension between efficiency and strategy-proofness under coarse priorities (Erdil and Ergin, 2008; Abdulkadiroğlu, Pathak, and Roth, 2009).

A key empirical lesson from the 2014 data is that the compulsory feature of the CSS is indispensable: simulation evidence indicates that roughly 115 physicians—about 30 % of participants—would have gone unassigned under voluntary participation, and that raising wages is an expensive and only partial substitute. An average monthly increase of roughly 600 USD per position, twice the monthly minimum wage of the time, would have induced only 35 additional physicians to be allocated (Vélez-Ferro, 2019). Our theoretical results therefore support both the choice of $DA^{\succeq}$ and the institutional design surrounding it. Since October 2014, the Ministry of Health has been using the deferred acceptance algorithm with priority, $DA^{\succeq}$, as the basis for the Colombian compulsory social service allocation.

References

[1] A. Abdulkadiroğlu, P.A. Pathak, and A.E. Roth (2009). Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match. *American Economic Review* 99(5): 1954–1978. doi:10.1257/aer.99.5.1954.

[2] A. Abdulkadiroğlu and T. Sönmez (2003). School Choice: A Mechanism Design Approach. *American Economic Review* 93(3): 729–747. doi:10.1257/000282803322157061.

[3] L.E. Dubins and D.A. Freedman (1981). Machiavelli and the Gale-Shapley Algorithm. *American Mathematical Monthly* 88(7): 485–494. doi:10.1080/00029890.1981.11995301.

[4] A. Erdil and H. Ergin (2008). What's the Matter with Tie-Breaking? Improving Efficiency in School Choice. *American Economic Review* 98(3): 669–689. doi:10.1257/aer.98.3.669.

[5] D. Gale and L.S. Shapley (1962). College Admissions and the Stability of Marriage. *American Mathematical Monthly* 69(1): 9–15. doi:10.2307/2312726.

[6] A.E. Roth (1982). The Economics of Matching: Stability and Incentives. *Mathematics of Operations Research* 7(4): 617–628. doi:10.1287/moor.7.4.617.

[7] A.E. Roth (1982b). Incentive Compatibility in a Market with Indivisible Goods. *Economics Letters* 9(2): 127–132. doi:10.1016/0165-1765(82)90003-9.

[8] A.E. Roth (1984). The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory. *Journal of Political Economy* 92(6): 991–1016. doi:10.1086/261272.

[9] A.E. Roth (1985). The Hospital Admission Problem is not Equivalent to the Marriage Problem. *Journal of Economic Theory* 36(2): 277–288. doi:10.1016/0022-0531(85)90106-1.

[10] A.E. Roth (1986). On the Allocation of Residents to Rural Hospitals: A General Property of Two-Sided Matching Markets. *Econometrica* 54(2): 425–427. doi:10.2307/1913160.

[11] A.E. Roth (1991). A Natural Experiment in the Organization of Entry-Level Labor Markets: Regional Markets for New Physicians and Surgeons in the United Kingdom. *American Economic Review* 81(3): 415–440.

[12] A.E. Roth (2008). Deferred Acceptance Algorithms: History, Theory, Practice, and Open Questions. *International Journal of Game Theory* 36(3): 537–569. doi:10.1007/s00182-008-0117-6.

[13] A.E. Roth and M.A.O. Sotomayor (1989). The College Admissions Problem Revisited. *Econometrica* 57(3): 559–570. doi:10.2307/1911052.

[14] A.E. Roth and M.A.O. Sotomayor (1990). *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Econometric Society Monograph Series. Cambridge University Press.

[15] L. Shapley and H. Scarf (1974). On Cores and Indivisibility. *Journal of Mathematical Economics* 1: 23–37. doi:10.1016/0304-4068(74)90033-0.

[16] T. Sönmez (1997). Manipulation via Capacities in Two-Sided Matching Markets. *Journal of Economic Theory* 77(1): 197–204. doi:10.1006/jeth.1997.2316.

[17] S. Vélez-Ferro (2019). *Essays on Centralized Market Allocations*. Ph.D. Dissertation, University of Maryland, College Park.