Stability and Strategy-Proofness for Matching with Constraints: A Problem in the Japanese Medical Match and Its Solution

By Yuichiro Kamada and Fuhito Kojima*

Real matching markets are subject to constraints. In the United States, the association called Accreditation Council for Graduate Medical Education regulates the total number of medical residents in each specialty. In some public school districts, multiple school programs often share one school building, so there is a bound on the total number of students in these programs in addition to each program's capacity because of the building's physical size.[1] The Japanese government introduced a new medical matching system in 2009 that imposes a "regional cap" in each of its 47 prefectures, which regulates the total number of medical residents who can be matched in each region.

Based on Kamada and Kojima (2011), this paper studies matching markets with such constraints by examining the Japanese medical matching market with regional caps in a great detail.[2] Specifically, we argue that the new system introduced in 2009 (the JRMP mechanism[3]) needs a fix, and provide an alternative mechanism that does better. The new system was introduced as a response to the criticisms that the formerly-used mechanism, the deferred acceptance (DA) mechanism due to Gale and Shapley (1962), allocated too many doctors in urban areas, reducing the quality and quantity of medical services in rural areas. We will not argue that the JRMP mechanism should be abandoned in favor of the DA mechanism; rather we take the constraint seriously and try to achieve a better outcome *given the constraint.*

Doing this requires us to overcome (at least) two nontrivial steps. *First,* an appropriate notion of stability is not straightforward in the presence of regional caps. We will show that the seemingly most straightforward concept has a crucial problem, and propose an alternative one. *Second,* constructing a new mechanism is not a trivial task. We will show that the one that is often proposed to us has a problematic incentive property, and propose our mechanism which has a better incentive property. A more complete treatment is provided in a full-length paper by Kamada and Kojima (2011).

I. Many-to-One Matching Model with Regional Caps

Following the model introduced by Kamada and Kojima (2011), let there be a finite set of hospitals, H, and a finite set of doctors, D. Each hospital h is associated with its capacity, denoted $q_h > 0$. Each agent $i \in H \cup D$ is associated with strict responsive preferences, $\succ_i$.[4] A matching specifies who is matched with whom, that is, it is a mapping that satisfies $\mu_d \in H \cup \{\varnothing\}$ for all $d \in D$, $\mu_h \subseteq D$ for all $h \in H$, and $\mu_d = h$ if and only if $d \in \mu_h$. Here, $\varnothing$ denotes "being unmatched."

A **mechanism** is a function that maps

* Kamada: Department of Economics, Harvard University, Cambridge, MA 02138, ykamada@fas.harvard.edu. Kojima: Department of Economics, Stanford University, Stanford, CA 94305, fkojima@stanford.edu. We are grateful to Scott Kominers, Jacob Leshno, Parag Pathak, Al Roth, and Jun Wako for comments.

[1] See Abdulkadiroğlu and Sönmez (2003) for the introduction to school choice problems.

[2] Information about the matching program written in Japanese is available at the websites of the government ministry and the matching organizer. See the websites of the Ministry of Health, Labor and Welfare (http://www.mhlw.go.jp/topics/bukyoku/isei/rinsyo/) and the Japan Residency Matching Program (http://www.jrmp.jp/).

[3] JRMP is an abbreviation for Japan Residency Matching Program.

[4] Informally, preferences are responsive if the ranking of an agent is independent of her colleagues, and any set of agents exceeding the capacity is unacceptable (Roth 1985).

preference profiles to matchings. It is **strategy-proof for doctors** if reporting true preferences is a dominant strategy for every doctor.

Each hospital h belongs to exactly one **region** r. Let H_r denote the set of hospitals that belong to region r. Each region r is associated with a **regional cap** $q_r > 0$. We say that a matching is **feasible** if $\sum_{h \in H_r} |\mu_h| \leq q_r$ for each r. A matching μ is said to be **(constrained) efficient** if there is no feasible matching μ' such that $\mu_i \succeq_i \mu'_i$[5] for all $i \in D \cup H$ and $\mu'_i \succ_i \mu_i$ for some $i \in D \cup H$.

As mentioned in the introduction, the Japanese government introduced regional caps as a constraint in order to increase the placement of doctors in rural areas. The following numbers illustrate the significance of regional caps: hospitals in Tokyo, Osaka, and Kyoto advertised 1,582, 860, and 353 positions in 2008, respectively, but the government set the regional caps of 1,287, 533, and 190.[6] In total, 34 out of the 47 prefectures are given regional caps smaller than the numbers of advertised positions in 2008.

II. The JRMP Mechanism and Its Deficiency

The JRMP mechanism is a modification of the DA mechanism. As DA is not guaranteed to produce a feasible matching, the Japanese government has introduced a "target capacity" for each hospital. The target capacity for hospital h, $\bar{q}_h$, is an exogenously given number proportional to q_h and no more than q_h, with the property that $\sum_{h \in H_r} \bar{q}_h \leq q_r$ for each region r.[7] The JRMP mechanism produces a matching that is obtained by running the DA algorithm, regarding these target capacities as real capacities. By definition the resulting matching is feasible.

Unfortunately, the JRMP mechanism suffers from a number of drawbacks, as shown by the following example.

EXAMPLE 1: Consider a market with ten doctors, $d_1, \dots, d_{10}$ and two hospitals, h_1 and h_2 in a single region with regional cap 10. Suppose that d_1, d_2, and d_3 prefer h_1 to $\varnothing$ to h_2, while the remaining seven doctors prefer h_2 to $\varnothing$ to h_1. Each of the two hospitals is associated with the capacity of 10, and prefers d_k to d_{k+1} for $k = 1, \dots, 9$ and d_{10} to being unmatched. The target capacity given by the JRMP mechanism is 5 for each hospital, and the mechanism produces a matching μ such that $\mu_{h_1} = \{d_1, d_2, d_3\}$ and $\mu_{h_2} = \{d_4, \dots, d_8\}$.

Notice that in this resulting matching, only 8 doctors are matched in total, while the regional cap is 10. Also, h_2's capacity is not binding in this matching (as its capacity is 10 while the matched number is only 5, which is a target capacity exogenously given by the mechanism). This means that even if, say, d_9 is matched to h_2, the regional cap is not violated, while no one is worse off and d_9 and h_2 are strictly better off.

This suggests that the JRMP mechanism is not constrained efficient, and it lacks a certain kind of stability. We will be clear on what we mean by this "lack of stability" in the next section.

III. Stability Notions

The standard stability notion (defined for models without regional caps) requires individual rationality and the absence of blocking pairs: Matching μ is **individually rational** if $\mu_i \succ_i \mu'_i$ for any $i \in D \cup H$ and μ' such that $\mu'_i \subsetneq \mu_i$. Given μ, a pair (d, h) is said to be a **blocking pair** if $h \succ_d \mu_d$ and either (a) $|\mu_h| < q_h$ and $d \succ_h \varnothing$ or (b) $d \succ_h d'$ for some $d' \in \mu_h$.

This notion does not take regional caps into account, so in particular there exist cases in which no stable matching in the above sense is feasible, simply because all stable matchings violate the regional caps.

[5] $\succeq_i$ is the weak preferences associated with $\succ_i$.

[6] The changes under the government's plan were so large that it provided a temporary measure that limits per-year reductions within a certain bound in the first years of operation, although the plan is to reach the planned regional cap eventually.

[7] Formally, $\bar{q}_h$ for $h \in H_r$ is set to be equal to an integer close to $\frac{q_r}{\sum_{h' \in H_r} q_{h'}} \cdot q_h$ whenever $q_r < \sum_{h' \in H_r} q_{h'}$, and $\bar{q}_h = q_h$ otherwise.

However an assignment that completely ignores participants' preferences would be undesirable. This suggests that an appropriate weakening of the stability concept is hoped for.

We first show that a straightforward fix of the above stability concept still suffers from the existence issue. To see this point, we say that a matching μ is **strongly stable** if it is feasible, individually rational, and any blocking pair violates a regional cap. Formally, if (d, h) is a blocking pair where $h \in H_r$, then (i) $|\mu_r| = q_r$, (ii) $d' \succ_h d$ for all $d' \in \mu_h$, and (iii) $\mu_d \notin H_r$ must be satisfied.

The following example illustrates a drawback of the strong stability concept.

EXAMPLE 2: Consider the market with two doctors, d_1 and d_2, and two hospitals, h_1 and h_2 in a single region with regional cap 1. Their preferences have a cyclic form– every agent regards everyone acceptable, d_1's first choice is h_1, h_1's first choice is d_2, d_2's first choice is h_2, and h_2's first choice is d_1. In this market, there is no strongly stable matching. To see this, first note that, for a matching to be strongly stable, it must respect the regional cap, hence there is at most one doctor matched in this market. If no doctor is matched then, say, (d_1, h_1) can form a blocking pair which does not violate the regional cap. This means that at least one doctor must be matched. Since the problem is symmetric, we need to consider only two cases (a) $\mu_{d_1} = h_1$ and (b) $\mu_{d_1} = h_2$. Case (a) is not strongly stable, as h_1 would be better off by rejecting d_1 and hiring d_2 (to form a blocking pair (d_2, h_1)). This is the standard non-stability argument. On the other hand, case (b) is not strongly stable either, because d_1 can move within a region to h_1 to form a blocking pair (d_1, h_1). Notice that this type of blocking pair did not exist when we consider the standard matching markets without regional caps. In such a market, the unique stable matching is $\mu_{d_1} = h_2$ and $\mu_{d_2} = h_1$, hence d_1 could not move to h_1. However, in our context, d_1 can do so because the seat in h_1 is empty precisely due to the constraint that in the region at most one doctor can be matched.

The example suggests that we need to give up strong stability as our goal.[8] To establish a reasonable concept of stability, we propose a *stability under regional preferences* $\succeq_r$. For this purpose, we introduce the notion of regional preferences. The **regional preference relation** $\succeq_r$ of region r is a weak ordering over integer vectors specifying the *number* of allocated doctors in each hospital in r. If a vector w (i) satisfies feasibility and (ii) respects the regional cap, then we assume that r strictly prefers w to vector $w' \neq w$ if the latter fails either of (i) or (ii); or if w matches weakly more doctors to all hospitals in r than w'. We also assume that $\succeq_r$ is substitutable.[9]

A matching μ is said to be **stable under regional preferences** $\succeq_r$ if it is feasible, individually rational, and any blocking pair violates a regional cap and makes r weakly worse off. Formally, if (d, h) is a blocking pair such that $h \in H_r$, then (i) $|\mu_r| = q_r$, (ii) $d' \succ_h d$ for all $d' \in \mu_h$, and (iii') $\mu_d \notin H_r$ or the distribution of doctors under μ is preferred to that under μ' by $\succeq_r$, where μ' is the matching produced by satisfying a blocking pair (d, h). This concept is weaker than strong stability because condition (iii') is weaker than condition (iii) of strong stability while conditions (i) and (ii) of these concepts are identical.

One possibility for regional preferences is to prefer distributions of doctors that have "fewer gaps" from the target capacities. Another example would be to prefer to have "more equalized" numbers of doctors across hospitals in the region. Stability under regional preferences captures such desiderata. The way that regional preferences are determined could depend on the policy goal of the government or of the region.

[8]We can also show that no mechanism that finds a strongly stable matching whenever it exists is strategy-proof for doctors. See Kamada and Kojima (2011, Example 5).

[9]See Kamada and Kojima (2011) for the formal definition.

IV. The Iterated Deferred Acceptance Mechanism

As a solution to the issue raised by Example 1, we often encounter suggestions by government officials and matching theorists for a particular kind of mechanisms. The mechanism, named the **iterated deferred acceptance (iterated DA)** mechanism by Kamada and Kojima (2011), uses the following algorithm: This algorithm consists of finite steps of rounds. In round 1, DA is run regarding the target capacities as the real capacities. If the resulting matching fills all the target capacities, then the algorithm stops. Otherwise, the algorithm proceeds to round 2 after the target capacities are modified as follows: hospitals set their new target capacities equal to their matched numbers of doctors if they have vacant seats relative to their target capacities; these vacant seats are reallocated to other hospitals in the same region according to a certain pre-specified rule. In round 2, DA is run with these modified target capacities. If the resulting matching fills all the new target capacities then the algorithm stops and otherwise it continues. We do the same in all other rounds, with a restriction that once a hospital has reduced its target capacity then it never increases (and require that the algorithm stop if no further reallocation is possible).

As one might expect, this mechanism produces a (strongly) stable matching in Example 1. However, as shown by the following example borrowed from Kamada and Kojima (2011), this mechanism is not strategy-proof for doctors.

EXAMPLE 3: Consider a market with two doctors, d_1 and d_2, and two hospitals h_1 and h_2 in a single region with regional cap 2. Each doctor prefers h_1 to h_2 to being unmatched. Each hospital is associated with a capacity of 2 and a target capacity of 1, and prefers d_1 to d_2 to being unmatched. In this market, the iterated DA ends in one round, resulting in the matching with $\mu_{h_1} = \{d_1\}$ and $\mu_{h_2} = \{d_2\}$. Doctor d_2 has an incentive to misreport her preferences. For, if she reports that she prefers h_1 to being unmatched to h_2, then the iterated DA proceeds to the second round with one seat moving from h_2 to h_1, and in the second round the matching $\mu_{h_1} = \{d_1, d_2\}$ is realized and the algorithm stops. Hence the iterated DA mechanism is not strategy-proof for doctors.

V. The Flexible Deferred Acceptance Mechanism

The reason behind inefficiency and instability of the JRMP mechanism was its rigidity of target capacities. Thus we need some kind of flexibility with respect to capacities. The iterated DA is one such attempt, but unfortunately it modified DA in a wrong way so it was not strategy-proof for doctors.

As an alternative, Kamada and Kojima (2011) proposed a new mechanism. To do so, instead of matching between hospitals and doctors, they considered matching between regions and doctors where each region has complicated "preferences" that is induced by hospitals' preferences as well as the regional cap. To deal with such complicated preferences they utilized the theory of matching with contracts.[10] Using this idea, they constructed a mechanism that they call the flexible deferred acceptance (FDA) mechanism, and established the following: The FDA mechanism produces a constrained efficient and stable matching under any given regional preferences, and it is strategy-proof for doctors (Theorems 1 and 2 of Kamada and Kojima (2011)).

Thus the FDA mechanism satisfies the crucial desiderata under regional caps—stability under regional preferences and strategy-proofness.[11] The FDA mechanism is new and has not been employed in practice so far, but it may be a compelling design under constraints that often appear in

[10]See Hatfield and Milgrom (2005), Hatfield and Kojima (2009, 2010), and Hatfield and Kominers (2009, 2012).

[11]In Example 1, FDA selects the Pareto-efficient matching μ' such that $\mu'_{h_1} = \{d_1, d_2, d_3\}$ and $\mu'_{h_2} = \{d_4, \ldots, d_{10}\}$, which in particular is stable under regional preferences. In Example 3, FDA selects a matching μ' such that $\mu'_{h_1} = \{d_1, d_2\}$ and $\mu'_{h_2} = \emptyset$, thus both doctors are matched with their best choices. This in particular implies that strategy-proofness for doctors is not violated in this example.

applications. It would be desirable to compare the performance of the FDA mechanism with those of other mechanisms such as the DA and JRMP mechanisms based on real data.

REFERENCES

Abdulkadiroğlu, Attila and Tayfun Sönmez (2003): "School Choice: A Mechanism Design Approach," *American Economic Review*, 93, 729-747.

Gale, D., and L. S. Shapley (1962): "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, 69, 9-15.

Hatfield, John W. and Fuhito Kojima (2009): "Group Incentive Compatibility for Matching with Contracts," *Games and Economic Behavior*, 67, 745-749.

Hatfield, John W. and Fuhito Kojima (2010): "Substitutes and Stability for Matching with Contracts," *Journal of Economic Theory*, 145, 1704-1723.

Hatfield, John W. and Scott D. Kominers (2009): "Contract Design and Stability in matching markets," *mimeo.*

Hatfield, John W. and Scott D. Kominers (2012): "Matching in networks with Bilateral Contracts," *American Economic Journal: Microeconomics*, forthcoming.

Hatfield, John W. and Paul Milgrom (2005): "Matching with Contracts," *American Economic Review*, 95, 913-935.

Kamada, Yuichiro and Fuhito Kojima (2011): "Improving Efficiency in Matching Markets with Regional Caps: The Case with Japan Residency Matching Program," *mimeo.*

Kamada, Yuichiro, Fuhito Kojima, and Jun Wako (2011): "Matching Theory and Its Applications: 'Regional Imbalance' of Medical Residents and Its Solutions" (in Japanese), *Iryo-Keizai Kenkyu (Research in Health Economics)*, 23, 5-19.

Roth, Alvin E. (1985): "The College Admission Problem is not Equivalent to the Marriage Problem," *Journal of Economic Theory*, 36, 277-288.