IMPROVING EFFICIENCY IN MATCHING MARKETS WITH REGIONAL CAPS: THE CASE OF THE JAPAN RESIDENCY MATCHING PROGRAM

YUICHIRO KAMADA AND FUHITO KOJIMA

Department of Economics, Harvard University
Cambridge, MA 02138
ykamada@fas.harvard.edu

Department of Economics, Stanford University
Stanford, CA 94305
fkojima@stanford.edu

Abstract. In an attempt to increase the placement of medical residents to rural hospitals, the Japanese government recently introduced "regional caps" which restrict the total number of residents matched within each region of the country. The government modified the deferred acceptance mechanism incorporating the regional caps. This paper shows that the current mechanism may result in avoidable inefficiency and instability and proposes a better mechanism that improves upon it in terms of efficiency and stability while meeting the regional caps. More broadly, the paper contributes to the general research agenda of matching and market design to address practical problems.
JEL Classification Numbers: C70, D61, D63.
Keywords: medical residency matching, regional caps, the rural hospital theorem, stability, strategy-proofness, matching with contracts

1. Introduction

Geographical distribution of medical doctors is a contentious issue in health care. One of the urgent problems is that many hospitals, especially those in rural areas, do not attract

Date: August 31, 2010.
We are grateful to Sylvain Chassang, Hisao Endo, Ryo Jinnai, Toshiaki Iizuka, Onur Kesten, Mihai Manea, Taisuke Matsubae, Aki Matsui, Andy McLennan, Yusuke Narita, Parag Pathak, Marcin Peski, Al Roth, Dan Sasaki, Satoru Takahashi, William Thomson, Kentaro Tomoeda, Jun Wako, Yosuke Yasuda, and seminar participants at Tohoku, Tokyo, Waseda, the 2010 Annual Meeting of the Japanese Economic Association, the 2010 SAET Conference, and the First Conference of the Chinese Game Theory and Experimental Economics Association for helpful comments. Doctors Keisuke Izumi and Masataka Kawana answered our questions about medical residency in Japan and introduced us to the relevant medical literature. Pete Troyan provided excellent research assistance.

sufficient numbers of medical residents to meet their demands. Describing the situation in the United States, Talbott (2007) reports that more than 35 million Americans live in underserved areas. Similar problems are present around the world. For instance, Shallcross (2005), Alcoba (2009), Nambiar and Bavas (2010), and Wongruang (2010) report doctor shortages in rural areas in the United Kingdom, India, Australia, and Thailand, respectively. Previous literature on stable matching suggests that a solution is elusive, as the rural hospital theorem (Roth, 1986) shows that any hospital that fails to fill all its positions in one stable matching is matched to an identical set of doctors in all stable matchings. This result implies that a hospital that cannot attract enough residents under one stable matching mechanism cannot increase the number of assigned residents no matter what other stable mechanism is used.

The shortage of residents in rural hospitals has become a hot political issue in Japan, where the deferred acceptance algorithm (Gale and Shapley, 1962) has placed around 8,000 graduating medical students to about 1,000 residency programs each year since 2003. In an attempt to increase the placement of residents to rural hospitals, the Japanese government recently introduced "regional caps" which, for each of the 47 prefectures that partition the country, restrict the total number of residents matched within the prefecture. The government modified the deferred acceptance algorithm incorporating the regional caps beginning in 2009 in an effort to attain its distributional goal.

This paper shows that the current Japanese mechanism may result in avoidable inefficiency and instability despite its resemblance to the deferred acceptance algorithm and proposes a better mechanism. More specifically, we first introduce concepts of stability and (constrained) efficiency that take regional caps into account. We point out that the current Japanese mechanism does not always produce a stable or efficient matching. We present a mechanism that we call the flexible deferred acceptancce mechanism, which finds a stable and efficient matching. We show that the mechanism is (group) strategy-proof for doctors, that is, telling the truth is a dominant strategy for each doctor (and even a coalition of doctors cannot jointly misreport preferences and benefit). The flexible deferred acceptance mechanism matches weakly more doctors to hospitals (in the sense of set inclusion) and makes every doctor weakly better off than the JRMP mechanism. These results suggest that replacing the current mechanism with the flexible deferred acceptance mechanism will improve the performance of the matching market.

We also find that the structural properties of the stable matchings with regional caps are strikingly different from those in the standard matching models. First, there does not necessarily exist a doctor-optimal stable matching (a stable matching unanimously preferred to every stable matching by all doctors). Neither do there exist hospital-optimal or doctor-pessimal or hospital-pessimal stable matchings. Second, different stable matchings can leave different hospitals with unfilled positions, implying that the rural hospital theorem fails in our context. Based on these observations, we investigate whether the government can design a reasonable mechanism that selects a particular stable matching based on its policy goals such as minimizing the number of unmatched doctors.

Although we closely connect our model to the Japanese residency matching market, the analysis is applicable to various other contexts in which similar mathematical structures arise. The first example is the allocation of residents across different medical specialties. In the United States, for instance, the association called Accreditation Council for Graduate Medical Education (ACGME) regulates the total number of residents in each specialty. The situation is isomorphic to our model in which medical specialties correspond to regions. Second, in some public school districts, multiple school programs often share one school building. In such a case, there is a natural bound on the total number of residents in these programs in addition to each program's capacity because of the building's physical size. This gives a mathematical structure isomorphic to the current model, suggesting that our analysis can be applied to the design of school choice mechanisms studied by Abdulkadiroğlu and Sönmez (2003). Lastly, the shortage of doctors in rural areas is a common problem around the globe. Countries mentioned above, such as the United States, the United Kingdom, and India, are but a few examples. If regional caps are imposed by a regulatory body such as a government, our analysis and mechanism would be directly applicable.

Let us emphasize that analyzing technical niceties associated with regional caps in the abstract is *not* the primary purpose of this paper. On the contrary, we study the market for Japanese medical residency in detail and offer practical solutions in that market. Improving the Japanese market is important by itself, which is one of the largest economies in the world producing over 8000 medical doctors each year. However, another point of this study is to provide a framework in which one can tackle problems arising in practical markets, which may prove useful in investigating other problems such as those which we have discussed in the previous paragraph. In that sense, this paper contributes to the general research agenda of matching and market design, advocated by Roth (2002) for instance, that emphasizes the importance of addressing issues arizing in real allocation problems.

Related literature. In the one-to-one matching setting, McVitie and Wilson (1970) show that a doctor or a hospital that is unmatched at one stable matching is unmatched in every stable matching. This is the first statement of the rural hospital theorem to our knowledge, and its variants and extensions have been established in increasingly general settings by Gale and Sotomayor (1985a,b), Roth (1984, 1986), Martinez, Masso, Neme, and Oviedo (2000), Hatfield and Milgrom (2005), and Hatfield and Kojima (2009), among others. As recent results are quite general, it seems that placing more doctors in rural areas has been believed to be a difficult (if not impossible) task, thus there are few studies offering solutions to this problem. The current paper explores possible ways to offer some positive results.

Roth (1991) reports that some hospitals in the United Kingdom prefer to hire no more than one female doctor while offering multiple positions (see Abdulkadiroğlu (2005) for a further analysis). Our flexible deferred acceptance mechanism has some resemblance to the deferred acceptance algorithm under such preferences. It turns out that Roth's model is mathematically unrelated from ours, but there is a way to associate our model to a framework of matching with contracts as defined by Hatfield and Milgrom (2005) by regarding a region (instead of a hospital) as a single agent. This correspondence allows us to show some of our results by exploiting properties of the matching with contracts model established by Hatfield and Milgrom (2005); Hatfield and Kojima (2008, 2009); Hatfield and Kominers (2009, 2010). However, these models are still different. For instance, a doctor-optimal stable matching exists and the rural hospital theorem holds in their model but not in ours.[1]

Milgrom (2009) and Budish, Che, Kojima, and Milgrom (2010) consider object allocation mechanisms with restrictions similar to the regional caps in our model. While their models are independent of ours (most notably, their analysis is not about two-sided matching but object allocation, and stability is not their primary concern), they share motivations with ours in that they consider flexible assignment in the face of complex constraints.

More broadly, this paper is part of a rapidly growing literature on matching market design. As advocated by Roth (2002), much of recent market design theory advanced by tackling problems arising in practical markets.[2] For instance, practical considerations in designing school choice mechanisms in Boston and New York City are discussed by Abdulkadiroğlu, Pathak, and Roth (2005, 2009) and Abdulkadiroğlu, Pathak, Roth, and Sönmez (2005, 2006). Abdulkadiroğlu, Che, and Yasuda (2008, 2009), Erdil and Ergin

[1]More specifically, the former result holds under the property called the substitute condition, and the latter under the substitute condition and another property called the law of aggregate demand or size (or cardinal) monotonicity (Alkan, 2002; Alkan and Gale, 2003).

[2]Literature on auction market design also emphasizes the importance of solving practical problems (see Milgrom (2000, 2004) for instance).

(2008), and Kesten (2009) analyze alternative mechanisms that may produce more efficient student placements than those that are currently used in New York City and Boston. Design issues of American market for physicians are investigated by Roth and Peranson (1999), Bulow and Levin (2006), and Niederle (2007). See Roth and Sotomayor (1990) for a comprehensive survey of the matching literature in the first three decades, and Roth (2007a) and Sönmez and Ünver (2008) for discussion of more recent studies.

The medical literature on doctor shortage in general and the Japanese situation in particular is discussed in the next section.

The rest of this paper proceeds as follows. Section 2 describes the Japanese residency matching market. In Section 3, we present the model of matching with regional caps and define weak stability and efficiency. We argue that weak stability is a mild requirement. Nonetheless, in Section 4 where we define the JRMP mechanism, we show that it does not necessarily produce a stable or efficient matching. Section 5 introduces and analyzes stronger stability concepts. In Section 6 we propose a new mechanism, the flexible deferred acceptance mechanism, and show that it produces a stable and efficient matching and is group strategy-proof. Section 7 discusses a number of further topics, and Section 8 concludes. Proofs are in the Appendix unless stated otherwise.

2. Residency Matching in Japan

This section describes residency matching in Japan and the issue of geographical distribution of residents. For further details of Japanese medical education written in English, see Teo (2007) and Kozu (2006). Also, information about the matching program written in Japanese is available at the websites of the government ministry and the matching organizer.[3]

Japanese residency matching started in 2003 as part of a comprehensive reform of the medical residency program. Prior to the reform, clinical departments in university hospitals, called *ikyoku*, had de facto authority to allocate doctors. The system was criticized because it was seen to have given clinical departments too much power and resulted in opaque, inefficient, and unfair allocation of doctors against their will.[4] Describing the

[3]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/).

[4]The criticism appears to have some justification. For instance, Niederle and Roth (2003) offer empirical evidence that a system without a centralized matching procedure reduces mobility and efficiency of resident allocation in the context of the U.S. gastroenterologist match.

situation, Onishi and Yoshida (2004) write "This clinical-department-centred system was often compared to the feudal hierarchy."

To cope with the above problem a new system, the Japan Residency Matching Program (JRMP), introduced a centralized matching procedure using the (doctor-proposing) deferred acceptance algorithm by Gale and Shapley (1962). Unlike its U.S. counterpart, the National Resident Matching Program (NRMP), the system has no "match variation" (Roth and Peranson, 1999) such as married couples, which would make many of good properties of the deferred acceptance algorithm fail.

Although the matching system was welcomed by many, it also received a lot of criticisms. This is because some hospitals, especially university hospitals in rural areas, felt that they attracted fewer residents under the new matching mechanism. They argued that the new system provided too much opportunity for students to choose urban hospitals over rural hospitals, resulting in severe doctor shortages in rural areas. While there is no conclusive evidence on the validity of their claim, an empirical study by Toyabe (2009) finds that several measures of geographical imbalance of doctors (Gini coefficients, Atkinson index, and Theil index of the per-capita number of doctors across regions) worsened in recent years, while these measures improve when residents are excluded from the calculation. Based on these findings, he suggests that the matching system from 2003 may have contributed to a widening regional imbalance of doctors.

To put such criticisms into context, we note that regional imbalance of doctors has been a long-standing and serious problem in Japan. As of 2004, there were over 160,000 people living in the so-called *mui-chiku*, which means "districts with no doctors" (Ministry of Health, Labour and Welfare, 2005) [5] and many more who were allegedly underserved. One government official told one of the authors (personal communication) that regional imbalance is one of the two most important problems in the government's health care policy, together with financing health care cost. There is evidence that the sufficient staffing of doctors in the hospital is positively correlated with the quality of medical care such as lower mortality (see Pronovost, Angus, Dorman, Robinson, Dremsizov, and Young (2002) for instance), thus the doctor shortage in rural areas appears to cause bad medical care.

In response to the criticisms against the matching mechanism, the Japanese government introduced a new system with regional caps beginning with the matching conducted in 2009. More specifically, a regional cap was imposed on the number of residents in each of the 47 prefectures that partition the country. If the total capacity demanded by hospitals

[5] A mui-chiku is defined as a district of roughly 4 kilometer radius in which at least 50 people live and have difficulty accessing health care.

in the region exceeds the regional cap, then the capacity of each hospital is reduced to equalize the total capacity with the regional cap.[6] Then it implements the deferred acceptance algorithm under the reduced capacities. We call this mechanism the Japan Residency Matching Program (JRMP) mechanism. The basic intuition behind this policy is that if residents are denied from urban hospitals because of the reduced capacities, then some of them will work for rural hospitals.

The new JRMP mechanism with regional caps was used in 2009 for the first time. The government claims that the change alleviated regional imbalance of residents: It reports that the proportion of residents matched to hospitals in rural areas has risen to 52.3 percent, an increase of one percentage point from the previous year (Ministry of Health, Labour and Welfare, 2009).[7] Meanwhile, there is mounting criticism to the JRMP mechanism as well. For instance, a number of governors of rural prefectures (see Tottori Prefecture (2009) for instance) and a student group (Association of Medical Students, 2009) have demanded that the government modify or abolish the JRMP mechanism with regional caps.[8] Among other things, a commonly expressed concern is that the current system with regional caps causes efficiency loss, preventing residents from learning important skills for practicing medical treatments. In the subsequent sections, we offer a theoretical framework to formally analyze these issues related to the regional cap and the existing JRMP mechanism.

3. Model

Let there be a set of doctors D and a set of hospitals H. Each doctor d has a strict preference relation $\succ_d$ over the set of hospitals and being unmatched (being unmatched is denoted by $\emptyset$). For any $h, h' \in H \cup \{\emptyset\}$, we write $h \succeq_d h'$ if and only if $h \succ_d h'$ or $h = h'$. Each hospital h has a strict preference relation $\succ_h$ over the set of subsets of doctors. For any $D', D'' \subseteq D$, we write $D' \succeq_h D''$ if and only if $D' \succ_h D''$ or $D' = D''$. We denote by $\succ = (\succ_i)_{i \in D \cup H}$ the preference profile of all doctors and hospitals.

[6]The capacity of a hospital is reduced proportionately to its original capacity in principle (subject to integrality constraints), but there are a number of fine adjustments and exceptions. If the total capacities demanded by hospitals in the region does not exceed the regional cap, then the capacities of hospitals in the regions are kept unchanged.

[7]Ministry of Health, Labour and Welfare (2009) defines "rural areas" as all prefectures except for 6 prefectures, Tokyo, Kyoto, Osaka, Kanagawa, Aichi, and Fukuoka, which have large cities.

[8]Interestingly, even regional governments in rural areas such as Tokushima and Tottori were opposed to the JRMP mechanism. They were worried that since the system reduces capacities of *each hospital* in the region, some of which could hire more residents, it can reduce the number of residents allocated in the regions even further. This feature - inflexibility of the way capacities are reduced - is one of the problems of the current JRMP mechanism, which we try to remedy by our alternative mechanism.

Doctor d is said to be **acceptable** to h if $d \succ_h \emptyset$.[9] Similarly, h is acceptable to d if $h \succ_d \emptyset$. Since only rankings of acceptable mates matter for our analysis, we often write only acceptable mates to denote preferences. For example,

$$\succ_d : h, h'$$

means that hospital h is the most preferred, h' is the second most preferred, and h and h' are the only acceptable hospitals under preferences $\succ_d$ of doctor d.

Given hospital $h \in H$ and nonnegative integer q_h, we say that preference relation $\succ_h$ is **responsive with capacity** q_h (Roth, 1985) if

(1) For any $D' \subseteq D$ with $|D'| \leq q_h$, $d \in D \setminus D'$ and $d' \in D'$, $D' \cup d \setminus d' \succeq_h D'$ if and only if $d \succeq_h d'$,
(2) For any $D' \subseteq D$ with $|D'| \leq q_h$ and $d' \in D'$, $D' \succeq_h D' \setminus d'$ if and only if $d' \succeq_h \emptyset$, and
(3) $\emptyset \succ_h D'$ for any $D' \subseteq D$ with $|D'| \geq q_h$.

In words, preference relation $\succ_h$ is responsive with a capacity if the ranking of a doctor is independent of her colleagues, and any set of doctors exceeding its capacity is unacceptable. We assume that preferences of all hospitals are responsive throughout the paper.

There is a finite set R which we call the set of **regions**. The set of hospitals H is partitioned into hospitals in different regions, that is, $H_r \cap H_{r'} = \emptyset$ if $r \neq r'$ and $H = \cup_{r \in R} H_r$, where H_r denotes the set of hospitals in region $r \in R$. For each region $r \in R$, there is a **regional cap** q_r, which is a nonnegative integer.

A **matching** μ is a mapping from D to $H \cup \{\emptyset\}$, with μ_d denoting a matching for doctor d. That is, a matching simply specifies the hospital where each doctor is assigned (if any). For each hospital $h \in H$, let $\mu_h = \{d \in D | \mu_d = h\}$ be the matching for h. A matching is **feasible** if $|\mu_{H_r}| \leq q_r$ for all $r \in R$, where $\mu_{H_r} = \cup_{h \in H_r} \mu_h$. In other words, feasibility requires that the regional cap for every region is satisfied. This requirement distinguishes the current environment from the standard model in the literature without regional caps.

Since regional caps are part of primitive of the environment, we consider a constrained efficiency concept. A feasible matching μ is **(constrained) efficient** if there is no feasible matching μ' such that $\mu'_i \succeq_i \mu_i$ for all $i \in D \cup H$ and $\mu'_i \succ_i \mu_i$ for at least one $i \in D \cup H$.

[9] We denote singleton set $\{x\}$ by x when there is no confusion.

To accommodate the regional caps, we introduce new stability concepts that generalize the standard notion. The first notion, which is the weakest of the stability conditions this paper introduces, is defined below.

Definition 1. A matching μ is **weakly stable** if it is feasible and

(1) **Individual rationality**: $\mu_d \succeq_d \emptyset$ for each $d \in D$; $d \succeq_h \emptyset$ for all $h \in H$ and $d \in \mu_h$.

(2) **No blocking pair**: If $h \succ_d \mu_d$, then one of the following holds.
   (a) $\emptyset \succ_h d$.
   (b) $|\mu_h| = q_h$ and $d' \succ_h d$ for all $d' \in \mu_h$.
   (c) $|\mu_{H_r}| = q_r$ for r such that $h \in H_r$ and $d' \succ_h d$ for all $d' \in \mu_h$.

The definition consists of two elements. First, a weakly stable matching must be *individually rational*. That is, no doctor is matched to an unacceptable hospital, and no hospital is matched to an unacceptable doctor. Second, weak stability requires that there be no *blocking pairs*. Intuitively, the condition requires that whenever a doctor prefers a hospital to her current partner, either the hospital is not interested in matching with the doctor or the regional cap prevents it from happening. More specifically, a blocking pair implicitly considered in the above definition is a pair of a doctor and a hospital such that the doctor prefers the hospital to her current partner, and either (i) the hospital prefers the doctor to one of the current partners or (ii) the doctor is acceptable to the hospital, the hospital has a vacant position, and adding one more doctor does not violate the regional cap of the region that the hospital belongs to. It can be readily verified that the conditions in item (2) of the definition is equivalent to the non-existence of such a blocking pair. The concept of weak stability reduces to the standard stability concept of Gale and Shapley (1962) (in which we drop condition (2c)) if $q_r > |D|$ for every $r \in R$.

This concept is rather weak because condition (2c) does not require that there be no pair of a doctor and a hospital who prefer each other to their current partners, if the hospital is in the same region as the hospital that the doctor is currently matched with. In practice, however, such a block may be a legitimate deviation and hence be required not to exist for a reasonable concept of stability, because forming such a block will not increase the total number of doctors matched within the region, thus keeping the regional cap respected. Example 3 in Section 5 makes this point explicit.

For this reason, we do not necessarily claim that weak stability is the most natural stability concept. In fact, we will introduce stronger concepts of stability later and analyze them to account for the issue discussed above. The main point of introducing weak stability for now is that, although this is a weak notion, we will later show that a matching produced by the current JRMP mechanism does not necessarily satisfy weak stability.

A **mechanism** φ is a function that maps preference profiles to matchings. The matching under φ at preference profile $\succ$ is denoted $\varphi(\succ)$ and agent i's matching is denoted by $\varphi_i(\succ)$ for each $i \in D \cup H$.

A mechanism φ is said to be **strategy-proof** if for all preference profiles $\succ$, all agents $i \in D \cup H$ and all preferences $\succ_i$ of agent i,

$$\varphi_i(\succ) \succeq_i \varphi_i(\succ_i, \succ_{-i}).$$

That is, no agent has an incentive to misreport her preferences under the mechanism. Strategy-proofness is regarded as a very important property for a mechanism to be successful.[10]

Unfortunately, however, there is no mechanism that produces a weakly stable matching for all possible preference profiles and is strategy-proof even in a market without regional caps, that is, $q_r > |D|$ for all $r \in R$ (Roth, 1982).[11] Given this limitation, we consider the following weakening of the concept requiring incentive compatibility only for doctors. A mechanism φ is said to be **strategy-proof for doctors** if for all preference profiles $\succ$, all doctors $d \in D$ and all preferences $\succ_d$ of doctor d,

$$\varphi_d(\succ) \succeq_d \varphi_d(\succ_d, \succ_{-d}).$$

A mechanism φ is said to be **group strategy-proof for doctors** if there is no preference profile $\succ$, a subset of doctors $D' \subseteq D$, and a preference profile $(\succ'_{d'})_{d' \in D'}$ of doctors in D' such that

$$\varphi_d((\succ'_{d'})_{d' \in D'}, (\succ_i)_{i \in D \cup H \setminus D'}) \succ_d \varphi_d(\succ) \text{ for all } d \in D'.$$

That is, no subset of doctors can jointly misreport their preferences to receive a strictly preferred outcome for every member of the coalition under the mechanism.

We do not necessarily regard (group) strategy-proofness for doctors as a minimum desirable property that our mechanism should satisfy (our criticism of the JRMP mechanism

[10]One good aspect of having strategy-proofness is that the matching authority can actually state it as the property of the algorithm to encourage doctors to reveal their true preferences. For example, the current webpage of the JRMP (last accessed on May 25, 2010, http://www.jrmp.jp/01-ryui.htm) states, as advice for doctors, that "If you list as your first choice a program which is not actually your first choice, the probability that you end up being matched with some hospital does not increase [...] the probability that you are matched with your actual first choice decreases." In the context of student placement in Boston, strategy-proofness was regarded as a desirable fairness property, in the sense that it provides equal access for children and parents with different degrees of sophistication to strategize (Pathak and Sonmez, 2008).

[11]Remember that a special case of our model in which $q_r > |D|$ for all $r \in R$ is the standard matching model with no binding regional caps.

in Section 4 does not hinge on (group) strategy-proofness), but it will turn out that the flexible deferred acceptance mechanism we propose in Section 6 does have this property.

As this paper analyzes the effect of regional caps in matching markets, it is useful to compare it with the standard matching model without regional caps. Gale and Shapley (1962) consider a matching model without any binding regional cap, which corresponds to a special case of our model in which $q_r > |D|$ for every $r \in R$. In that model, they propose the following **(doctor-proposing) deferred acceptance algorithm**:

- Step 1: Each doctor applies to her first choice hospital. Each hospital rejects the lowest-ranking doctors in excess of its capacity and all unacceptable doctors among those who applied to it, keeping the rest of the doctors temporarily (so doctors not rejected at this step may be rejected in later steps).

In general,

- Step t: Each doctor who was rejected in Step $(t-1)$ applies to her next highest choice (if any). Each hospital considers these doctors *and* doctors who are temporarily held from the previous step together, and rejects the lowest-ranking doctors in excess of its capacity and all unacceptable doctors, keeping the rest of the doctors temporarily (so doctors not rejected at this step may be rejected in later steps).

We denote by ϕ^D a mechanism that produces, for each input, the matching at the termination of the above algorithm.

The algorithm terminates at a step in which no rejection occurs. The algorithm always terminates in a finite number of steps. In their basic setting, Gale and Shapley (1962) show that the resulting matching is stable in the standard matching model without any binding reginal cap.

Even though there exists no strategy-proof mechanism that produces a stable matching for all possible inputs, the deferred acceptance mechanism is group strategy-proof for doctors (Dubins and Freedman, 1981; Roth, 1982).[12] The result has been extended by many subsequent researches, suggesting that the incentive compatibility of the mechanism is quite robust and general.[13]

[12]Ergin (2002) defines a stronger version of group strategy-proofness. It requires that no group of students can misreport preferences jointly and make some of its members strictly better off without making any of its members strictly worse off. He identifies a necessary and sufficient condition for the deferred acceptance mechanism to satisfy this version of group strategy-proofness.

[13]Researches generalizing (group) strategy-proofness of the mechanism include Abdulkadiroğlu (2005), Hatfield and Milgrom (2005), Martinez, Masso, Neme, and Oviedo (2004), Hatfield and Kojima (2008, 2009), and Hatfield and Kominers (2009, 2010).

4. The JRMP Mechanism and its Deficiency

In the JRMP mechanism, denoted φ^J, there is a government-imposed **target capacity** $\bar{q}_h \leq q_h$ for each hospital h such that $\sum_{h \in H_r} \bar{q}_h \leq q_r$ for each region $r \in R$. The **JRMP mechanism** is a rule that produces the matching resulting from the deferred acceptance algorithm except that, for each hospital h, it uses $\bar{q}_h$ instead of q_h as the hospital's capacity.

The JRMP mechanism is based on a simple idea: In order to satisfy regional caps, simply force hospitals to be matched to a smaller number of doctors than their real capacities, but otherwise use the standard deferred acceptance algorithm.

In our theoretical model we assume that $\bar{q}_h$ is exogenously given for each hospital h. In the current Japanese system, if the sum of the hospitals' capacities exceeds the regional cap, then the target $\bar{q}_h$ of each hospital h is set at an integer close to $\frac{q_r}{\sum_{h' \in H_r} q_{h'}} \cdot q_h$. That is, each hospital's target is (roughly) proportional to its capacity. This might suggest that hospitals have incentives to misreport their true capacities, but in Japan, the government regulates how many positions each hospital can offer so that the capacity could be considered exogenous.

Although the mechanism is a variant of the deferred acceptance algorithm, the mechanism suffers from at least two problems. The first problem is about stability: Despite its intention, the result of the JRMP mechanism is not necessarily weakly stable, as seen in the following example. The example also illustrates how the JRMP mechanism works.

Example 1 (JRMP mechanism does not necessarily produce a weakly stable matching)**.** There is one region r with regional cap $q_r = 10$, in which two hospitals, h_1 and h_2, reside. Each hospital h has a capacity of $q_h = 10$. Suppose that there are 10 doctors, $d_1, \dots, d_{10}$. Preference profile $\succ$ is as follows:

$$\succ_{h_i}: d_1, d_2, \dots, d_{10}, \quad \text{for } i = 1, 2;$$

$$\succ_{d_j}: h_1 \;\text{ if } j \leq 3 \quad \text{ and } \quad \succ_{d_j}: h_2 \;\text{ if } j \geq 4.$$

That is, three doctors prefer hospital h_1 to being unmatched to hospital h_2, while the other seven doctors prefer hospital h_1 to being unmatched to hospital h_2.

Following the current Japanese practice, set the target capacity at $\bar{q}_h = \frac{q_r}{q_{h_1}+q_{h_2}} \cdot q_h = \frac{10}{10+10} \cdot 10 = 5$ for each hospital h and consider the JRMP mechanism associated with this target profile. At the first round of the algorithm, doctors d_1, d_2 and d_3 apply to hospital h_1, and the rest of doctors apply to hospital h_2. Hospital h_1 does not reject anyone at this round, as the number of applicants is less than its target capacity, and all applicants are acceptable. Hospital h_2 rejects d_9 and d_{10} and accepts other applicants, because the number of applicants exceeds the *target capacity* (not the hospital's capacity itself!), and it prefers doctors with smaller indices (and all doctors are acceptable). Since d_9 and d_{10} prefer being unmatched to h_1, they do not make further applications, so the algorithm terminates at this point. Hence the resulting matching $\varphi^J(\succ)$ is such that $\varphi^J_{h_1}(\succ) = \{d_1, d_2, d_3\}$ and $\varphi^J_{h_2}(\succ) = \{d_4, d_5, d_6, d_7, d_8\}$. This is not weakly stable: For example, hospital h_2 and doctor d_9 constitute a blocking pair because, in particular, neither a hospital's capacity nor the regional cap is binding, so conditions (2b) and (2c) do not have a bite.

The second problem is about efficiency: The JRMP mechanism may result in an inefficient matching even in the constrained sense, as demonstrated in the following example.

Example 2 (JRMP mechanism does not necessarily produce an efficient matching)**.** Consider the same environment as in Example 1 again. Consider a matching μ defined by,

$$\mu_{h_1} = \{d_1, d_2, d_3\} \quad \text{and} \quad \mu_{h_2} = \{d_4, d_5, d_6, d_7, d_8, d_9, d_{10}\}.$$

Since each hospital's capacity and the regional cap are still respected, μ is feasible. Moreover, every agent is weakly better off with doctors d_9 and d_{10} strictly better off than at $\varphi^J(\succ)$. Hence we conclude that the JRMP mechanism results in an inefficient matching in this example.

The above two examples suggest that a problem of the JRMP mechanism is its lack of flexibility: The JRMP mechanism runs as if the target capacity is the actual capacity of hospitals, thus rejecting an application of a doctor to a hospital unnecessarily. The mechanism that we propose in Section 6 overcomes problems of both stability and inefficiency by, intuitively speaking, making the target capacities flexible. Before formally introducing the mechanism, we define and discuss our goal that we try to achieve by the mechanism.

5. Goal Setting: Stability Concepts and Strategy-Proofness

As discussed earlier, the concept of weak stability introduced in the previous section is rather weak. This is because it does not regard as legitimate certain hospital-doctor pairs that can be matched without violating the feasibility constraint related to regional caps. Then a natural question is: What is the "right" stability concept? In this section, we propose two stability concepts that are stronger than the one proposed in Section 3 and analyze their relevance and relationships. *The objective in this section is not to* *discuss technical details of these stability concepts per se, but to set an explicit goal for constructing a new algorithm, which we introduce in Section 6.*

Before defining and discussing the stability concepts, we demonstrate that the weak notion of stability does imply a desirable property, namely efficiency:

Theorem 1. *Any weakly stable matching is efficient.*

The fact that stability implies efficiency is well-known when there is no regional cap (recall that weak stability reduces to the standard concept of stability in a model without regional caps). In fact it is a straightforward implication of the fact that a matching is stable if and only if it is in the core and the fact that any outcome in the core is efficient. With regional caps, however, there is no obvious way to define a cooperative game, so the core is not well-defined. Theorem 1 states that efficiency of weakly stable matchings still holds in our model. To overcome the above difficulty, the proof presented in the Appendix shows this result directly rather than associating stability to the core in a cooperative game.

Now we formalize the stability concepts that are stronger than the weak stability as defined in Section 3. The first notion presented below is meant to capture the idea that any blocking pair that will not violate the regional cap should be considered legitimate, so the appropriate stability concept should require that no agents have incentives to form any such blocking pair.

Definition 2. A matching μ is **strongly stable** if it is feasible and

(1) **Individual rationality**: $\mu_d \succeq_d \emptyset$ for each $d \in D$; $d \succeq_h \emptyset$ for all $h \in H$ and $d \in \mu_h$.
(2) **No blocking pair**: If $h \succ_d \mu_d$, then one of the following holds.
   (a) $\emptyset \succ_h d$.
   (b) $|\mu_h| = q_h$ and $d' \succ_h d$ for all $d' \in \mu_h$.
   (c') $\mu_d \notin H_r$ and $|\mu_{H_r}| = q_r$ for r such that $h \in H_r$, and $d' \succ_h d$ for all $d' \in \mu_h$.

The difference from weak stability defined in Definition 1 is an added condition in (2c'), namely "$\mu_d \notin H_r$." That is, a blocking pair is considered legitimate when the doctor in the pair moves between two hospitals in the same region. Since such a movement keeps the total number of doctors in a region unchanged, what this definition says is that a blocking pair is deemed valid as long as the regional cap is satisfied.

To see the difference between weak stability and strong stability clearly, consider the following example.

Example 3 (Strong stability is strictly stronger than weak stability)**.** There is one region r with regional cap $q_r = 1$, in which two hospitals, h_1 and h_2, reside. Each hospital h has a capacity of $q_h = 1$. Suppose that there is only one doctor, d. Preferences are specified as follows:

$$\succ_{h_i}: \quad d \ \text{ for } i = 1, 2;$$
$$\succ_d: \qquad h_1, h_2.$$

First, note that there are two weakly stable matchings,

$$\mu = \begin{pmatrix} h_1 & h_2 \\ d & \emptyset \end{pmatrix},$$
$$\mu' = \begin{pmatrix} h_1 & h_2 \\ \emptyset & d \end{pmatrix}.$$

In each of matchings μ and μ', since the regional cap is binding, d is not allowed to change the partner. Moreover, since no one is unacceptable by anyone, any matching is individually rational. Thus both μ and μ' are weakly stable. By contrast, only μ is strongly stable: To check the strong stability of this matching, note just that in this matching, d is matched with her first choice, and h_1 is also matched with its first choice. To see that μ' is not strongly stable, notice that the pair (d, h_1) is a blocking pair because they can be paired without violating the regional cap constraint.

The above example shows that strong stability is a strictly stronger concept than weak stability. Nonetheless, we will not pursue to achieve strongly stable matchings when we construct an algorithm in Section 6. There are at least two reasons for this. The first reason is that a strongly stable matching does not necessarily exist. The following example demonstrates this point.

Example 4 (A strongly stable matching does not necessarily exist)**.** There is one region r with regional cap $q_r = 1$, in which two hospitals, h_1 and h_2, reside. Each hospital h has a capacity of $q_h = 1$. Suppose that there are two doctors, d_1 and d_2. We assume the following preference:

$$\succ_{h_1}: d_1, d_2, \qquad \succ_{h_2}: d_2, d_1;$$
$$\succ_{d_1}: h_2, h_1, \qquad \succ_{d_2}: h_1, h_2.$$

Note first that at most one doctor is matched in any feasible matching since the regional cap is one. A matching in which no doctor is matched is not strongly stable because any pair of a doctor and a hospital constitutes a blocking pair. A matching μ such that

$\mu_{h_1} = \{d_1\}$ is not strongly stable since a pair (d_1, h_2) is a blocking pair. A matching μ such that $\mu_{h_1} = \{d_2\}$ is not strongly stable because a pair (d_1, h_1) constitutes a blocking pair (h_1 can reject d_2 and be matched with d_1). By a symmetric argument, a matching μ such that $\mu_{h_2} = \{d_2\}$ or $\mu_{h_2} = \{d_1\}$ is not strongly stable. Therefore, a strongly stable matching does not exist in this example.

Even if a strongly stable matching does not always exist, can we try to achieve a weaker desideratum? More specifically, does there exist a mechanism that selects a strongly stable matching whenever there exists one? We show that such a mechanism does not exist if we also require certain incentive compatibility: There is no mechanism that selects a strongly stable matching whenever there exists one and is strategy-proof for doctors. This is the second reason that we do not attempt to achieve strong stability as a natural desideratum. To see this point consider the following example.

Example 5 (No mechanism that is strategy-proof for doctors selects a strongly stable matching whenever there exists one)**.** There is one region r with regional cap $q_r = 1$, in which two hospitals, h_1 and h_2, reside. Each hospital h has a capacity of $q_h = 1$. Suppose that there are two doctors, d_1 and d_2. We assume the following preferences:

$$\succ_{h_1}: d_1, d_2, \qquad \succ_{h_2}: d_2, d_1,$$
$$\succ_{d_1}: h_2, \qquad \succ_{d_2}: h_1.$$

In this market, there are two strongly stable matchings,

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

Now, suppose that a mechanism chooses μ under the above preference profile $\succ$. Then d_1 is unmatched. Consider a reported preference $\succ'_{d_1}$ of d_1,

$$\succ'_{d_1}: h_2, h_1.$$

Then μ' is a unique strongly stable matching, so the mechanism chooses μ' at $(\succ'_{d_1}, \succ_{-d_1}$). Doctor d_1 is better off at μ' than at μ since she is matched to h_2 at μ' while she is unmatched at μ. Hence, d_1 can profitably misreport her preference when her true preference is $\succ_{d_1}$.

If a mechanism chooses μ' under the above preference profile $\succ$, then by a symmetric argument, doctor d_2 can profitably misreport her preferences when her true preference is

$\succ_{d_2}$. Therefore there does not exist a mechanism that is strategy-proof for doctors and selects a strongly stable matching whenever there exists one.

The above example shows that a strongly stable matching need not exist, and there exists no mechanism that is strategy-proof for doctors and selects a strongly stable matching whenever there exists one. These results suggest that the concept of strong stability is not appropriate as our desideratum.

Although strong stability is "too strong" in the senses discussed above, it may still be desirable to have a notion stronger than weak stability. Strong stability is too strong because it allows any doctor-hospital pairs as a blocking pair as long as the pair does not violate a regional cap. One natural idea to restrict blocking pairs that are regarded as legitimate is to use the notion of target capacity. This leads us to the following concept.

Definition 3. A matching μ is **stable** with respect to a target capacity $(\bar{q}_h)_{h\in H}$ if it is feasible and

(1) **Individual rationality**: $\mu_d \succeq_d \emptyset$ for each $d \in D$; $d \succeq_h \emptyset$ for all $h \in H$ and $d \in \mu_h$.
(2) **No blocking pair**: If $h \succ_d \mu_d$, then one of the following holds.
  (a) $\emptyset \succ_h d$.
  (b) $|\mu_h| = q_h$ and $d' \succ_h d$ for all $d' \in \mu_h$.
  (c') $\mu_d \notin H_r$ and $|\mu_{H_r}| = q_r$ for r such that $h \in H_r$, and $d' \succ_h d$ for all $d' \in \mu_h$.
  (d') $\mu_d \in H_r$, $|\mu_{H_r}| = q_r$ for r such that $h \in H_r$, $|\mu_h| + 1 - \bar{q}_h > |\mu_{\mu_d}| - 1 - \bar{q}_{\mu_d}$, and $d' \succ_h d$ for all $d' \in \mu_h$.

This concept is stronger than weak stability while weaker than strong stability. Stability is different from strong stability in that condition (d') is added and, since there are more possible cases, stability is weaker than strong stability. Meanwhile, conditions (c') or (d') still imply condition (c) in the definition of weak stability, so stability is stronger than weak stability.

Condition (c') addresses the case in which the deviating doctor is not from the same region as the deviating hospital (note the condition $\mu_d \notin H_r$). Condition (d') declares that certain cases of movement of a doctor within a region (note the condition $\mu_d \in H_r$) is not regarded as legitimate blocks. Too see this point, consider the inequality in (d'),

$$|\mu_h| + 1 - \bar{q}_h > |\mu_{\mu_d}| - 1 - \bar{q}_{\mu_d}. \tag{5.1}$$

The left-hand side is the number of doctors matched to h in excess of its target $\bar{q}_h$ if d actually moves to h. The right hand side is the number of doctors matched to the original hospital μ_d in excess of its target $\bar{q}_{\mu_d}$ if d moves out of μ_d. This property says that such a movement will not decrease the imbalance of over-target numbers of matching across hospitals. Intuitively, if the movement of the doctor "equalizes" the excess over the target capacity than the current matching (that is, $|\mu_h| + 1 - \bar{q}_h \leq |\mu_{\mu_d}| - 1 - \bar{q}_{\mu_d}$), then such a movement should be regarded as a valid blocking pair. Thus, in order for the movement to be considered invalid, inequality (5.1) should hold.

We note that there may be other natural definitions of stability. For example, it may be natural to suppose that a hospital with capacity 20 is eligible for twice as many doctors over the target as a hospital with capacity 10. It will turn out in Section 7 that a variant of stability (which changes condition (d')) can accommodate this idea, and a variant of the mechanism that we propose in Section 6 can generate such a variant of stable matching. In the main part of this paper, we assume that the policy goal is expressed as condition (d'). However, this particular choice of the condition is *not* a necessary requirement for our analysis to work, as we will show in Section 7. We chose this condition for expositional clarity and it seems canonical in the absence of any additional information about the policy goal. The choice of a particular variant of stability should be in part the product of society's preferences, and we restrict ourselves to proposing solutions that are flexible enough to meet as wide a range of policy goals as possible.

In the next section, we propose an algorithm that always generates a stable matching.

6. The New Mechanism: The Flexible Deferred Acceptance Mechanism

We present a new mechanism that, for any given input, results in a stable matching. To do so, we first define the **flexible deferred acceptance algorithm:**

Assume that a target capacity profile $(\bar{q}_h)_{h \in H}$ is given as in the JRMP mechanism.

For each $r \in R$, specify indices over hospitals in region r so that $H_r = \{h_1, h_2, \dots, h_{|H_r|}\}$. Given this set of indices, consider the following algorithm.

(1) Begin with an empty matching, that is, a matching μ such that $\mu_d = \emptyset$ for all $d \in D$.

(2) Choose a doctor d who is currently not tentatively matched to any hospital and who has not applied to all acceptable hospitals yet. If such a doctor does not exist, then terminate the algorithm.

(3) Let d apply to the most preferred hospital $\bar{h}$ at $\succ_d$ among the hospitals that have not rejected d so far. Let r be the region such that $\bar{h} \in H_r$.

(4) (a) For each $h \in H_r$, let D'_h be the entire set of doctors who have applied to h so far. For each hospital $h \in H_r$, choose $\bar{q}_h$ best acceptable doctors according to

$\succ_h$ from D'_h if they exist, and otherwise choose all acceptable doctors related to h. Formally, for each $h \in H_r$ choose D'' such that

(i) $D'' \subset D'_h$,

(ii) $|D''| = \min\{\bar{q}_h, |D'_h|\}$,

(iii) $d \succ_h d'$ for any $d \in D''$ and $d' \in D'_h \setminus D''$.

(b) One by one, let each hospital in the region choose the best remaining doctor until the regional quota q_r is filled or the capacity of the hospital is filled or no doctor remains to be matched. Formally,

(i) Let $\iota_i = 0$ for all $i \in \{1, 2, \dots, |H_r|\}$. Let $i = 1$.

(ii) (A) If either the number of doctors already chosen by the region r as a whole equals q_r, or $\iota_i = 1$, then go back to Step 2.

(B) Otherwise, let h_i choose the most preferred (acceptable) doctor in D'_h at $\succ_h$ among the doctors that have not been chosen by h_i so far, if such a doctor exists and the number of doctors chosen by h_i so far is strictly smaller than q_{h_i}.

(C) If no new doctor was chosen at Step 4(b)iiB, then set $\iota_i = 1$. If a new doctor was chosen at Step 4(b)iiB, then set $\iota_j = 0$ for all $j \in \{1, 2, \dots, |H_r|\}$. If $i < |H_r|$ then increment i by one and if $i = |H_r|$ then set i to be 1 and go back to the beginning of Step 4(b)ii.

We define the **flexible deferred acceptance mechanism**, denoted φ^F, to be a mechanism that produces, for each input, the matching at the termination of the above algorithm.[14]

The flexible deferred acceptance mechanism is analogous to the deferred acceptance mechanism and the JRMP mechanism. What distinguishes the flexible deferred acceptance mechanism from the JRMP mechanism is that it lets hospitals fill their capacities "flexibly" than the latter. To see this point, first observe that the way that hospitals choose doctors who applied in (4)(a) is essentially identical to the one in the JRMP algorithm. As seen before, the JRMP may result in an inefficient and unstable matching because this step does not let hospitals to tentatively keep doctors beyond target capacities even if regional caps are not binding. This is addressed in step (4) (b). In that step, hospitals in a region are allowed to keep more doctors than their target capacities

[14]We show in Thorem 2 that the algorithm stops in finite steps.

if doing so keeps the regional caps respected. Thus there is a sense in which this algorithm corrects the deficiency of the JRMP mechanism while following closely the deferred acceptance algorithm.

In the flexible deferred acceptance algorithm, one needs to specify indices over hospitals. We will discuss in Subsection 7.4 the effect of different ways of setting indices on the welfare of hospitals.

The following example illustrates how the flexible deferred acceptance algorithm works.

Example 6 (The flexible deferred acceptance algorithm)**.** Consider the same example as in Example 1. Remember that the JRMP mechanism can produce a matching that violate both efficiency and weak stability, let alone stability. The flexible deferred acceptance algorithm selects a matching that is stable and efficient. Precisely, let doctors apply to hospitals in the order of their indices. For doctors d_1 to d_8, the algorithm does not go in to step (4)-(b), as the number of doctors in each hospital is no larger than its target. When d_9 applies, doctors $d_1, \dots, d_8$ are still matched to hospitals in step (4)-(a), and d_9 is matched to h_2 in step (4)-(b). In the same way, when d_{10} applies, doctors $d_1, \dots, d_8$ are still matched to hospitals in step (4)-(a), and d_9 and d_{10} are matched to h_2 in step (4)-(b). Hence a stable and efficient matching results. Intuitively, the algorithm makes the targeting more flexible. This is the idea behind the name "*flexible* deferred acceptance."

The following is the main result of this section.

Theorem 2. *The flexible deferred acceptance algorithm stops in finite steps. The mechanism produces a stable matching for any input and is group strategy-proof for doctors.*

To see an intuition for stability of the flexible deferred acceptance mechanism, recall that there is a sense in which hospitals fill their capacities "flexibly" in the flexible deferred acceptance algorithm. More specifically, at each step hospitals can tentatively accept doctors beyond the target capacities as long as the regional cap is not violated. Then the kind of rejection that causes instability in Example 1 does not occur in the flexible deferred acceptance algorithm. Thus an acceptable doctor is rejected from a preferred hospital either because there are enough better doctors in that hospital, or the regional quota is filled by other doctors. So such a doctor cannot form a blocking pair, suggesting that the resulting matching is stable.[15]

The intuition for strategy-proofness for doctors is similar to the one for the deferred acceptance mechanism. A doctor does not need to give up trying for her first choice

[15]The way that hospitals' capacities are filled after target capacties are filled ensures that no such blocking pair can "equalize" the distribution of doctors in excess of targets.

because, even if she is rejected, she will be able to apply to her second choice, and so forth. In other words, the "deferred" acceptance guarantees that she will be treated equally if she applies to a position later than others.

Although the above are rough intuitions of the results, the formal proof presented in Appendix B takes a different approach. It relates our model to the model of "(many-to-many) matching with contracts" (Hatfield and Milgrom, 2005). The basic idea of the proof is to regard each region as a consortium of hospitals that acts as one agent, and to define its choice function that selects a subset from any given collection of pairs (contracts) of a doctor and a hospital in the region. Once we successfully connect our model to the matching model with contracts, properties of that model can be invoked to show the theorem.

Corollary 1. *The flexible deferred acceptance mechanism produces an efficient matching for any input.*

Proof. By Theorem 2, the flexible deferred acceptance mechanism produces a stable matching. Since stability implies weak stability, the flexible deferred acceptance mechanism produces a weakly stable matching. By Theorem 1, weak stability implies efficiency, completing the proof. □

The matching generated by the flexible deferred acceptance mechanism satisfies the following additional property.

Theorem 3. *If the number of doctors matched with $h \in H$ in the flexible deferred acceptance mechanism is strictly less than its target, for any $d \in D$ who are not matched with h, either d is unacceptable by h or d prefers its current match to h.*

Hence, there exists no pair of a doctor and a hospital who want to deviate from the matching generated by the flexible deferred acceptance mechanism, if the number of doctors currently matched with the hospital is strictly less than its target. The conclusion of the theorem applies even if the regional capacity is already binding, thus this property is not implied by stability.

7. Discussion

This section provides several discussions that relate our model and results to existing theories. In Subsection 7.1, we show that there does not necessarily exist side-optiaml stable matchings, that is, matchings that are prefered by all doctors or by all hospitals. In Subsection 7.2, we consider the rural hospital theorem of Roth (1986) and show that the theorem does not hold in our environment. This subsection also discusses how the flexible deferred acceptance mechanism works in terms of the "match rate," the ratio of the number of doctors matched to some hospital to the total number of doctors. Subsection 7.3 considers the generalization of stability and the flexible deferred acceptance mechanism, and Subsection 7.4 examines the welfare effect of different choices of indices over hospitals in the flexible deferred acceptance mechanism and the target settings. Subsection 7.5 considers "floor constraints" instead of "ceiling constraints" (regional caps).

7.1. **Nonexistence of Side-Optimal Stable Matchings.** There does not necessarily exist a doctor-optimal stable matching (a stable matching unanimously preferred to every stable matching by all doctors). Neither does there exist a hospital-optimal stable matching. To see this point, consider the environment presented in Example 4, and assume that targets are $\bar{q}_{h_1} = \bar{q}_{h_2} = 0$. With this specification, there are two stable matchings,

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

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

Clearly, d_1 and h_1 strictly prefer μ to μ' while d_2 and h_2 strictly prefer μ' to μ. Thus there exists neither a doctor-optimal stable matching nor a hospital-optimal stable matching. Moreover, this example shows that there exists neither a doctor-pessimal stable matching nor a hospital-pessimal stable matching in general.

7.2. **The Rural Hospital Theorem and The Match Rate.** In this subsection, we examine the celebrated rural hospital theorem of Roth (1986). The theorem states that, in a matching model without regional caps, any hospital that fails to fill all its positions in one stable matching is matched to an identical set of doctors in all stable matchings. It also states that the set of unmatched doctors is identical across all stable matchings.

The theorem is of particular interest when we consider allocating a sufficient number of doctors to rural areas. Although the rural hospital theorem might suggest that increasing the number of doctors in a particular set of hospitals is impossible, the conclusion of the theorem does not necessarily hold in our context with regional caps, even with the most stringent concept of strong stability. The following example makes this point clear.

Example 7 (The rural hospital theorem does not hold)**.** There is one region r with regional cap $q_r = 1$, in which two hospitals, h_1 and h_2, reside. Each hospital h has a capacity of $q_h = 1$. Suppose that there are two doctors, d_1 and d_2. We assume the following preferences:

$$\succ_{h_1}: d_1, \quad \succ_{h_2}: d_2;$$

$$\succ_{d_1}: h_1, \quad \succ_{d_2}: h_2.$$

It is straightforward to check that there are two strongly stable matchings,

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

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

Notice that hospital h_1 fills its capacity in macthing μ while it does not do so in matching μ'. Also, d_1 is matched to a hospital in matching μ while it does not in matching μ'. Hence both conclusions of the rural hospital theorem fail, even with the notion of strong stability.

One might suspect that, although the rural hospital theorem dos not apply, it might be the case that each region attracts the same number of doctors in any strongly stable matchings. The following example shows that this is not true.

Example 8 (The number of doctors matched to hospitals in a rural region may be different in different strongly stable matchings)**.** We modify Example 7 by adding one more region r', which we interpret here for the sake of discussion as a "rural region." Region r' has the regional cap of $q_{r'} = 1$, and one hospital, h_3, resides in it. Suppose that h_3 has a capacity of $q_{h_3} = 1$. The preferences are modified as follows:

$$\succ_{h_1}: d_1, \quad \succ_{h_2}: d_2, \quad \succ_{h_3}: d_1;$$

$$\succ_{d_1}: h_1, h_3, \quad \succ_{d_2}: h_2.$$

It is straightforward to check that there are two strongly stable matchings,

$$\mu = \begin{pmatrix} h_1 & h_2 & h_3 & \emptyset \\ d_1 & \emptyset & \emptyset & d_2 \end{pmatrix},$$

$$\mu' = \begin{pmatrix} h_1 & h_2 & h_3 \\ \emptyset & d_2 & d_1 \end{pmatrix}.$$

Thus the hospital in rural region r' does not attract any doctors in matching μ, while it attracts one doctor in matching μ'.

Hence, when the number of doctors matched to hospitals in rural regions matters, the choice of a mechanism *is* an important issue, in the presence of regional caps.

Related to the rural hospital theorem is the notion of "match rate," which is the ratio of the number of doctors matched to some hospital to the total number of doctors. The match rate seems to be a measure that many people care about. For example, match rates are listed on the annual reports published by the NRMP and the JRMP.[16] This is perhaps because the match rate is an easy measure for participants to understand.[17]

Although it would be desirable if a mechanism could select a matching that has the maximum match rate, among the stable matchings, there exists no mechanism that always does so and is strategy-proof for doctors. In particular, our flexible deferred acceptance mechanism does not select a matching that has the maximum match rate among stable matchings. We first demonstrate in Example 9 that the flexible deferred acceptance mechanism does not always produce a stable matching with the maximal match rate. The second example, Example 10, shows that there does not exist a mechanism that is strategy-proof for doctors and always selects a matching with the maximum match rate among stable matchings.

Example 9 (The flexible DA does not necessarily select a matching with the highest match rate among stable matchings)**.** Take the same example as in Example 8. Also, let the target profile be $(\bar{q}_1, \bar{q}_2) = (1, 0)$. Then, the flexible deferred acceptance mechanism always selects a matching μ defined in Example 8. But this has a match rate of $1/2$, while the other matching, namely μ' defined in Example 8, has a match rate of 1.

It is an unfortunate fact that the flexible deferred acceptance mechanism does not necessarily maximize the match rate within stable matchings. A natural next question is whether there is any reasonable mechanism that can do so. The following example shows that the answer is negative in the sense that such a requirement is inconsistent with strategy-proofness.

Example 10 (No mechanism that is strategy-proof for doctors can always select a matching with the highest match rate among stable matchings)**.** Modify the environment in Example 8 as follows:

$$\succ_{h_1}: d_1, \quad \succ_{h_2}: d_2, \quad \succ_{h_3}: d_1, d_2;$$
$$\succ_{d_1}: h_1, h_3, \quad \succ_{d_2}: h_2, h_3,$$

[16]For instance, see National Resident Matching Market (2010) and Japan Residency Matching Program (2009).

[17]The ease of understanding may not be a persuasive reason for economic theorists to care about the match rates, but it seems to be a crucial issue for market designers. For a mechanism to work well in practice, it is essential that people are willing to participate in the mechanism. To this end, providing information in an accessible manner, as in the form of the match rates, seems to be of great importance.

with everything else unchanged. Let $\bar{q}_{h_1} = \bar{q}_{h_2} = \bar{q}_{h_3} = 0$. Notice that, given these preferences, there are two stable matchings, namely μ with $\mu_{d_1} = h_1$ and $\mu_{d_2} = h_3$, and μ' with $\mu'_{d_1} = h_3$ and $\mu'_{d_2} = h_2$. Take a mechanism that always selects a matching with the highest match rate among the stable matchings, if any. We show that this mechanism cannot be strategy-proof. Since both μ and μ' have match rate of 1, both can potentially be chosen by the mechanism. Suppose that the mechanism chooses μ. Then, doctor d_2 has an incentive to misreport her preferences: If she reports that hospital h_2 is the only acceptable match, then given the new profile of the preferences, the only stable matching that maximizes the match rate among stable matchings is μ'. Since $\mu' \succ_{d_2} \mu$, doctor d_2 indeed has an incentive to misreport. A symmetric argument can be made for the case in which the mechanism chooses μ' given the true preference profile. Hence, there does not exist a mechanism that is strategy-proof for doctors and always selects a matching with the highest match rate among stable matchings.

Despite the above negative results, we have the following result, which gives us bounds on the match rates in the matchings produced by the flexible deferred acceptance mechanism:

Theorem 4. *For any preference profile $\succ$,*

(1) *Each doctor $d \in D$ weakly prefers a matching produced by the deferred acceptance mechanism to the one produced by the flexible deferred acceptance mechanism to the one produced by the JRMP mechanism.*
(2) *If a doctor is unmatched in the deferred acceptance mechanism, she is unmatched in the flexible deferred acceptance mechanism. If a doctor is unmatched in the flexible deferred acceptance mechanism, she is unmatched in the JRMP mechanism.*

Notice that part (2) of the above result implies that the match rate is weakly higher in the deferred acceptance mechanism than in the flexible deferred acceptance mechanism, which in turn has a weakly higher match rate than the JRMP mechanism.

7.3. **More General Stability Concept and Algorithm.** As mentioned in Section 5, the notion of stability is based on the idea that if the result of a move of a doctor "unequalizes" the excess over the target capacities than the current matching, it is not deemed as a valid blocking pair. We argued that this is not a necessary choice of the concept as, for example, it may be natural to suppose that a hospital with capacity 20 is entitled to twice as many doctors (over the target) as a hospital with capacity 10. There may be other criteria, and a natural question is what kind of criteria can be accommodated in general.

Appendix B generalizes the concept of stability that takes this issue into account. We also propose a generalized version of the flexible deferred acceptance mechanism. We show that the generalized flexible deferred acceptance algorithm finds a stable matching as defined more generally, and it is group strategy-proof.

7.4. **Welfare Effects of Picking Orders and Targets.** The flexible deferred acceptance algorithm follows a certain picking order of hospitals in each region when there are some doctors remaining to be tentatively matched after hospitals have kept doctors up to their target capacities. One issue around the mechanism is how to decide the picking order. One natural conjecture may be that choosing earlier (that is, having a lower index in the flexible deferred acceptance mechanism) benefits a hospital. This would be a problematic property: If choosing earlier benefits the hospital, then how to order hospitals will be a sensitive policy issue to cope with because each hospital would have incentives to be granted an early picking order. Fortunately, the conjecture is not true, as shown in the following example. The example also shows that the different choices of indices result in different stable matchings, thus the choice of indices does matter for the algorithm's outcome.

Example 11. Let there be two hospitals, h_1 and h_2, in region r_1, and h_3 in region r_2. Suppose that $q_{h_1} = 2$, $\bar{q}_{h_1} = 1$, $q_{h_2} = q_{h_3} = 1$, and $\bar{q}_{h_2} = \bar{q}_{h_3} = 0$. Regional caps of r_1 is two and that for r_2 is one. Preferences are

$$\succ_{h_1}: d_1, d_4, d_2, \quad \succ_{h_2}: d_3, \quad \succ_{h_3}: d_2, d_1,$$

$$\succ_{d_1}: h_3, h_1, \quad \succ_{d_2}: h_1, h_3, \quad \succ_{d_3}: h_2, \quad \succ_{d_4}: h_1.$$

(1) Assume that h_1 is ordered earlier than h_2. In that case, in the flexible deferred acceptance mechanism, d_1 applies to h_3, d_2 and d_4 apply to h_1, and d_3 applies to h_2. d_2 and d_4 are accepted while d_3 is rejected. The matching finalizes.
(2) Assume that h_1 is ordered after h_2. In that case, in the flexible deferred acceptance mechanism, d_1 applies to h_3, d_2 and d_4 apply to h_1, and d_3 applies to h_2. But now d_2 is rejected while d_3 is accepted. Then d_2 applies to h_3, displacing d_1 from h_3. Then d_1 applies to h_1. d_1 is accepted.

First, notice that hospital h_2 is better off in case (2) than in case (1). Thus being indexed earlier helps h_2 in this example. However, if h_1 prefers $\{d_1\}$ to $\{d_2, d_4\}$ (which is consistent with the assumption that hospital preferences are responsive with capacities), then h_1 is also made better off in case (2) than in case (1). Thus being indexed later helps h_1 if she prefers $\{d_1\}$ to $\{d_2, d_4\}$. Therefore, the effect of picking indices on hospitals' welfare is not monotone.

A related concern is about what could be called "target monotonicity." That is, keeping everything else constant, does an increase of the target of a hospital make it better off under the flexible deferred acceptance mechanism? If so, then hospitals would have strong incentives to influence policy makers to give them large targets.

Consider the market that is identical to the one in Example 11, except that the target of h_1 is now decreased to 0, with the indices such that h_1 chooses before h_2. Then h_1 is matched to $\{d_1\}$ under the flexible deferred acceptance mechanism. Therefore, if h_1 prefers $\{d_1\}$ to $\{d_4, d_2\}$, then h_1 is made better off when it does not choose earlier than h_2.

7.5. **Floor Constraints.** The present paper offers a practical solution for the Japanese resident matching problem with regional caps. However, the regional cap may not be an ultimate objective per se, but a means to allocate medical residents "evenly" to different areas. Setting a cap –a ceiling constraint on the number of residents in a region– is an obvious approach to this desideratum, but there may be other possible regulations. For example, one might wonder setting floor constraints, as opposed to cap constraints, would be an easier and more direct solution. However, there are reasons that floor constraints may be difficult to use. First, even an existence of a "feasible matching" is not guaranteed. For example, if no doctor finds a hospital in a rural area to be acceptable, then setting a positive floor constraints on the number of doctors matched with hospitals in the area results in an individually irrational matching (doctors matched with hospitals in the rural area would just reject taking the job). Second, even if an individually rational matching exists, it is not clear whether a stable matching exists.[18]

8. Conclusion

This paper showed that the current matching mechanism used in Japan may result in avoidable inefficiency and instability despite its similarity to the celebrated deferred acceptance algorithm. We proposed a new mechanism, called the flexible deferred acceptance mechanism. This mechanism is (group) strategy-proof, generates a stable and efficient matching, and places more doctors to hospitals than the current mechanism.

With regional caps there may not necessarily exist a unique "right" notion of stability concept, and hence there may not necessarily exist the unique choice of the mechanism.

[18]A similar point is made in the context of school choice by Ehlers (2007).

The choice would depend on the government's welfare and distributional goals, and there is room for the government to select a particular stable matching based on such goals. We hope that this paper serves as a basis for achieving such goals and, more broadly, that it contributes to the general agenda of matching/market design theory to address specific issues arising in practical problems.

We intentionally refrained from judging the merit of imposing regional caps itself (except for a certain welfare result in Theorem 4). We took this approach because our model does not explicitly include patients or ethical concern by general populace, which may be underlying arguments for increasing doctors in rural areas. This issue might be relevant to some degree, but is outside the scope of this paper. Instead, we took an approach in the new tradition of market design research, in which one regards constraints such as fairness and repugnance as requirements to be respected and offers solutions consistent with them.[19] In a similar spirit, we did not analyze other policies such as subsidies to incentivize residents to work in rural area.[20] However, regional caps seem to stay as a political reality, and so we believe that it is important to take them as given and try to provide a practical solution.

The paper opens new avenues for further research topics. First, as mentioned before, strategy-proofness for every agent including hospitals is impossible even without regional caps if we also require stability. However, truthtelling is an approximately optimal strategy under the deferred acceptance mechanism in large markets under some assumptions (Roth and Peranson, 1999; Immorlica and Mahdian, 2005; Kojima and Pathak, 2009). Although such an analysis requires a much more specialized model structure than what this paper has and is outside the scope of this paper, a conjecture is that approximate incentive compatibility similar to these papers may hold under some conditions.

Second, it would be desirable to obtain the actual data to test how well the flexible deferred acceptance mechanism does relative to the JRMP matchings. We are planning to work on this as a future research topic.

Third, it would be nice to look at markets that have similar structures to the one we studied in this paper. We expect some general insights may carry over to such settings, while market-specific details should be carefully taken into account when we consider different markets in different political or cultural environments.

[19] This approach is eloquently advocated by Roth (2007b).

[20] This is not because subsidies are not important. In fact, subsidy is used to attract residents to rural areas is many countries such as the United States and Japan.

References

Abdulkadiroğlu, A. (2005): "College admissions with affirmative action," *International Journal of Game Theory*, 33(4), 535–549.

Abdulkadiroğlu, A., Y.-K. Che, and Y. Yasuda (2008): "Expanding Choice in School Choice," Economic Research Initiatives at Duke Research Paper No. 20.

——— (2009): "Resolving Conflicting Preferences in School Choice: the Boston Mechanism Reconsidered," *American Economic Review*, forthcoming.

Abdulkadiroğlu, A., P. A. Pathak, and A. E. Roth (2005): "The New York City High School Match," *American Economic Review Papers and Proceedings*, 95, 364–367.

——— (2009): "Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match," *American Economic Review*, 99, 1954–1978.

Abdulkadiroğlu, A., P. A. Pathak, A. E. Roth, and T. Sönmez (2005): "The Boston Public School Match," *American Economic Review Papers and Proceedings*, 95, 368–372.

——— (2006): "Changing the Boston School Choice Mechanism: Strategy-proofness as Equal Access," Harvard University and Boston College, unpublished mimeo.

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

Alcoba, N. (2009): "Hitting the road to alleviate Indias rural doctor shortage," *the Canadian Medical Association Journal*, http://www.cmaj.ca/cgi/reprint/180/10/E34.pdf.

Alkan, A. (2002): "A class of multipartner matching markets with a strong lattice structure," *Economic Theory*, 19(4), 737–746.

Alkan, A., and D. Gale (2003): "Stable Schedule Matching Under Revealed Preferences," *Journal of Economic Theory*, 84, 73–94.

Association of Medical Students (2009): "Public comment on the proposed revision to the residency match (in Japanese)," http://students.umin.jp/.

Budish, E., Y.-K. Che, F. Kojima, and P. R. Milgrom (2010): "Implementing Random Assignments: A Generalization of the Birkhov-von Neumann Theorem," mimeo.

Bulow, J., and J. Levin (2006): "Matching and Price Competition," *American Economic Review*, 96, 652–668.

Dubins, L. E., and D. A. Freedman (1981): "Machiavelli and the Gale-Shapley algorithm," *American Mathematical Monthly*, 88, 485–494.

Ehlers, L. (2007): "Controlled School Choice," Unpublished mimeo.

Erdil, A., and H. Ergin (2008): "What's the Matter with Tie-Breaking? Improving Efficiency in School Choice," *American Economic Review*, 98, 669–689.

Ergin, H. (2002): "Efficient Resource Allocation on the Basis of Priorities," *Econometrica*, 70, 2489–2498.

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

Gale, D., and M. A. O. Sotomayor (1985a): "Ms Machiavelli and the Stable matching problem," *American Mathematical Monthly*, 92, 261–268.

——— (1985b): "Some Remarks on the Stable Matching Problem," *Discrete Applied Mathematics*, 11, 223–232.

Hatfield, J. W., and F. Kojima (2008): "Group Incentive Compatibility for Matching with Contracts," *Games and Economic Behavior*, forthcoming.

——— (2009): "Substitutes and Stability for Matching with Contracts," forthcoming, *Journal of Economic Theory*.

Hatfield, J. W., and S. D. Kominers (2009): "Contract Design and Stability in Matching Markets," mimeo.

——— (2010): "Matching in Networks with Bilateral Contracts," mimeo.

Hatfield, J. W., and P. Milgrom (2005): "Matching with Contracts," *American Economic Review*, 95, 913–935.

Immorlica, N., and M. Mahdian (2005): "Marriage, Honesty, and Stability," *SODA*, pp. 53–62.

Japan Residency Matching Program (2009): "Result of the Residency Match," http://www.jrmp.jp/koho/2009/2009press.htm.

Kesten, O. (2009): "School Choice with Consent," *Quarterly Journal of Economics*, forthcoming.

Kojima, F., and M. Manea (2009): "Axioms for Deferred Acceptance," forthcoming, *Econometrica*.

Kojima, F., and P. Pathak (2009): "Incentives and stability in large two-sided matching markets," *American Economic Review*, 99(3), 608–627.

Kozu, T. (2006): "Medical education in Japan," *Academic Medicine*, 81(12), 1069–1075.

Martinez, R., J. Masso, A. Neme, and J. Oviedo (2000): "Single agents and the set of many-to-one stable matchings," *Journal of Economic Theory*, 91, 91–105.

——— (2004): "On group strategy-proof mechanisms for a many-to-one matching model," *International Journal of Game Theory*, 33, 115–128.

McVitie, D. G., and L. Wilson (1970): "Stable Marriage Assignments for Unequal Sets," *BIT*, 10, 295–309.

Milgrom, P. (2000): "Putting Auction Theory to Work: The Simulteneous Ascending Auction," *Journal of Political Economy*, 108(2), 245–272.

——— (2009): "Assignment messages and exchanges," *American Economic Journal: Microeconomics*, 1(2), 95–113.

Milgrom, P. R. (2004): *Putting Auction Theory to Work*. Cambridge University Press, Cambridge.

Ministry of Health, Labour and Welfare (2005): "Report on no-doctor districts (in Japanese)," http://www.mhlw.go.jp/toukei/list/76-16.html.

——— (2009): "On the outcome of the residency match (in Japanese)," http://www.mhlw.go.jp/topics/bukyoku/isei/rinsyo/matching/dl/091030-1a.pdf.

Nambiar, M., and J. Bavas (2010): "Rudd plan won't fix rural doctor shortage: RDA," http://www.abc.net.au/news/stories/2010/03/16/2846827.htm.

National Resident Matching Market (2010): "Results and Data, 2010 Main Residency Match," http://www.nrmp.org/data/resultsanddata2010.pdf.

Niederle, M. (2007): "Competitive wages in a match with ordered contracts," *The American Economic Review*, 97(5), 1957–1969.

Niederle, M., and A. Roth (2003): "Unraveling reduces mobility in a labor market: Gastroenterology with and without a centralized match," *Journal of Political Economy*, 111(6).

Onishi, H., and I. Yoshida (2004): "Rapid change in Japanese medical education," *Medical Teacher*, 26(5), 403–408.

Pathak, P., and T. Sonmez (2008): "Leveling the playing field: Sincere and sophisticated players in the Boston mechanism," *The American Economic Review*, 98(4), 1636–1652.

Pronovost, P., D. Angus, T. Dorman, K. Robinson, T. Dremsizov, and T. Young (2002): "Physician staffing patterns and clinical outcomes in critically ill patients: a systematic review," *JAMA, the Journal of the American Medical Association*, 288(17), 2151.

Roth, A. E. (1982): "The Economics of Matching: Stability and Incentives," *Mathematics of Operations Research*, 7, 617–628.

——— (1984): "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory," *Journal of Political Economy*, 92, 991–1016.

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

——— (1986): "On the allocation of residents to rural hospitals: a general property of two-sided matching markets," *Econometrica*, 54(2), 425–427.

——— (1991): "A Natural Experiment in the Organization of Entry Level Labor Markets: Regional Markets for New Physicians and Surgeons in the U.K.," *American Economic Review*, 81, 415–440.

——— (2002): "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics, Fisher-Schultz Lecture," *Econometrica*, 70, 1341–1378.

——— (2007a): "Deferred Acceptance Algorithms: History, Theory, Practice and Open Questions," forthcoming, *International Journal of Game Theory.*

——— (2007b): "Repugnance as a Constraint on Markets," *The Journal of Economic Perspectives*, pp. 37–58.

Roth, A. E., and E. Peranson (1999): "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design," *American Economic Review*, 89, 748–780.

Roth, A. E., and M. A. O. Sotomayor (1990): *Two-sided Matching: a Study in Game-theoretic Modeling and Analysis.* Econometric Society monographs, Cambridge.

Shallcross, T. (2005): "General Medicine in a Remote & Rural Hospital: Problems and Proposals a personal view," http://www.rcpe.ac.uk/publications/rarm/shalcross.pdf.

Sönmez, T., and M. Ünver (2008): "Matching, Allocation, and Exchange of Discrete Resources," *Handbook of Social Economics, Elseview, forthcoming.*

Talbott, C. (2007): "Shortage of Doctors Affects Rural U.S.," *the Washington Post*, http://www.washingtonpost.com/wp-dyn/content/article/2007/07/21/AR2007072100432.html.

Teo, A. (2007): "The current state of medical education in Japan: a system under reform," *Medical education*, 41(3), 302–308.

Tottori Prefecture (2009): "Emergency request to the proposed revision to the residency match by four prefectural governors (in Japanese)," http://www.pref.tottori.lg.jp/dd.aspx?menuid=102291.

Toyabe, S. (2009): "Trend in geographic distribution of physicians in Japan," *International Journal for Equity in Health*, 8(1), 5.

Wongruang, P. (2010): "Diagnosis bleak for rural doctors," *the Bangkok Post*, http://www.bangkokpost.com/news/investigation/34445/diagnosis-bleak-for-rural-doctors.

Appendix A. Proof of Theorem 1

Proof. Let μ be a stable matching and assume, for contradiction, that μ is not efficient. Then there exists a feasible matching μ' that Pareto dominates μ, that is, there is a feasible matching μ' such that $\mu'_i \succeq_i \mu_i$ for all $i \in D \cup H$, with at least one being strict. Noting that matching is bilateral, this implies that there exists a doctor $d \in D$ with $\mu'_d \succ_d \mu_d$. Since μ is a stable matching, $\mu_d \succeq_d \emptyset$ and hence $\mu'_d \neq \emptyset$, so $\mu'_d \in H$. Denote $h = \mu'_d$. Since μ is a stable matching, $h \succ_d \mu_d$ implies one of the following:

(1) $\emptyset \succ_h d$.
(2) $|\mu_h| = q_h$ and $d' \succ_h d$ for all $d' \in \mu_h$.
(3) $|\mu_{H_r}| = q_r$ for r such that $h \in H_r$ and $d' \succ_h d$ for all $d' \in \mu_h$.

Suppose $\emptyset \succ_h d$. Then, if $|\mu_h| = q_h$, then there is a doctor $d'' \in \mu'_h \setminus \mu_h$ such that $d'' \succ_h d'$ for some $d' \in \mu_h$ (otherwise, by responsiveness of the preference of h, it follows that $\mu_h \succ_h \mu'_h$). Then, since μ is stable, $\mu_{d''} \succ_{d''} h = \mu'_{d''}$, contradicting the assumption that μ' Pareto dominates μ. If $|\mu_h| < q_h$, then there should be a doctor $d'' \in \mu'_h \setminus \mu_h$ such that $d'' \succ_h \emptyset$ (otherwise, by responsiveness of the preference of h, it follows that $\mu_h \succ_h \mu'_h$). Then, since μ is stable, $\mu_{d''} \succ_{d''} h = \mu'_{d''}$, contradicting the assumption that μ' Pareto dominates μ.

Suppose $|\mu_h| = q_h$ and $d' \succ_h d$ for all $d' \in \mu_h$. Then there should be a doctor $d'' \in \mu'_h \setminus \mu_h$ such that $d'' \succ_h d'$ for some $d' \in \mu_h$ (otherwise, by responsiveness of the preference of h, it follows that $\mu_h \succ_h \mu'_h$). Then, since μ is stable, $\mu_{d''} \succ_{d''} h = \mu'_{d''}$, contradicting the assumption that μ' Pareto dominates μ.

Suppose $|\mu_{H_r}| = q_r$ for r such that $h \in H_r$ and $d' \succ_h d$ for all $d' \in \mu_h$. Then, if $|\mu'_h| \leq |\mu_h|$, then there should be a doctor $d'' \in \mu'_h \setminus \mu_h$ such that $d'' \succ_h d'$ for some $d' \in \mu_h$ (otherwise, by responsiveness of the preference of h, it follows that $\mu_h \succ_h \mu'_h$). Then, since μ is stable, $\mu_{d''} \succ_{d''} h = \mu'_{d''}$, contradicting the assumption that μ' Pareto dominates μ. If $|\mu'_h| > |\mu_h|$, then since $|\mu_{H_r}| = q_r$, there exists a hospital $h' \in H_r$ with $|\mu'_{h'}| < |\mu_{h'}|$. This, since $\mu'_{h'} \succ_{h'} \mu_{h'}$ as μ' Pareto dominates μ, implies that there should be a doctor $d'' \in \mu'_{h'} \setminus \mu_{h'}$ such that $d'' \succ_{h'} d'$ for some $d' \in \mu_{h'}$ (otherwise, by responsiveness of the preference of h', it follows that $\mu_{h'} \succ_{h'} \mu'_{h'}$). Then, since μ is stable, $\mu_{d''} \succ_{d''} h' = \mu'_{d''}$, contradicting the assumption that μ' Pareto dominates μ. □

Appendix B. A general model

Let $\succeq_r$ be a weak ordering over nonnegative-valued integer vectors $W_r := \{w = (w_h)_{h \in H_r} | w_h \in \mathbb{Z}_+\}$. We write $w \succ_r w'$ if and only if $w \succeq_r w'$ holds but $w' \succeq_r w$ does not. That is, $\succeq_r$ is a binary relation that is complete and transitive (but not necessarily antisymmetric). Given $\succeq_r$, a function $\tilde{\mathrm{Ch}}_r : W_r \to W_r$ is an **associated quasi choice rule** if $\tilde{\mathrm{Ch}}_r(x) \in \arg\max_{\succeq_r}\{y|y \leq x\}$ for any non-negative integer vector $x = (x_h)_{h \in H_r}$.[21] Throughout we require that the quasi choice rule $\tilde{\mathrm{Ch}}_r$ be **consistent**, that is, $\mathrm{Ch}_r(x) \leq y \leq x \to \mathrm{Ch}_r(y) = \mathrm{Ch}_r(x)$. This is a mild condition that the choice is made in a consistent manner: If $\mathrm{Ch}_r(x)$ is chosen at x and the supply decreases to $y \leq x$ but $\mathrm{Ch}_r(x)$ is still available under y, then the same choice $\mathrm{Ch}_r(x)$ should be made under y as well. Note that there may be more than one quasi choice rule associated with a given weak ordering $\succeq_r$ because the set $\arg\max_{\succeq_r}\{y|y \leq x\}$ may not be a singleton for some $\succeq_r$ and x. Throught we assume that the regional preference $\succeq_r$ satisfies some regularity conditions as described below.

(1) (a) $w' \succ_r w$ if $w_h > q_h \geq w'_h$ for some $h \in H_r$ and $w'_{h'} = w_{h'}$ for all $h' \neq h$, and
(b) $w' \succ_r w$ if $\sum_{h \in H_r} w_h > q_r \geq \sum_{h \in H_r} w'_h$.
These properties are mild and simply say that the region's preference should be such that it prefers the total number of doctors in the region to be at most its regional cap and it desires no hospital to be forced to be assigned more doctors than its real capacity. This condition implies that, for any y, the component $[\tilde{\mathrm{Ch}}_r(y)]_h$ of $\tilde{\mathrm{Ch}}_r(y)$ for h satisfies $[\tilde{\mathrm{Ch}}_r(y)]_h \leq q_h$ for each $h \in H_r$, that is, the capacity constraint for each hospital is respected, and $\sum_{h \in H_r}(\tilde{\mathrm{Ch}}_r(y))_h \leq q_r$, that is, the regional cap is respected, in the (quasi) choice by the region.

(2) If $y \lneq x \leq q_{H_r} := (q_h)_{h \in H_r}$, and $\sum_{h \in H_r} x_h \leq q_r$, then $x \succ_r y$. This condition formalizes the idea that the region prefers to fill as many positions in hospitals in the region as possible so long as doing so does not lead to violation of the hospitals' real capacities or the regional cap. This requirement implies that any associated quasi choice rule is **acceptant** (Kojima and Manea, 2009), that is, for each x, if there exists h such that $[\mathrm{Ch}_r(x)]_h < \min\{q_h, x_h\}$, then $|\mathrm{Ch}_r(x)| = q_r$. This condition captures the idea that the social planner should not waste caps allocated to the region: If there exists some doctor who is not accepted by a hospital even though she is acceptable to the hospital and the hospital's capacity is not binding, then the regional cap should be binding. This property seems to be a minimal requirement.

[21]For any two vectors $x = (x_h)_{h \in H_r}$ and $y = (y_h)_{h \in H_r}$, we write $x \leq y$ if and only if $x_h \leq y_h$ for all $h \in H_r$. We write $x \lneq y$ if and only if $x \leq y$ and $x_h < y_h$ for at least one $h \in H_r$.

The weak ordering $\succeq_r$ is **substitutable** if there exists an associated quasi choice rule $\tilde{\text{Ch}}_r$ that satisfies

$$w \leq w' \Rightarrow \tilde{\text{Ch}}_r(w) \geq \tilde{\text{Ch}}_r(w') \wedge w, \tag{B.1}$$

or equivalently,

$$w \leq w' \Rightarrow [\tilde{\text{Ch}}_r(w)]_h \geq \min\{[\tilde{\text{Ch}}_r(w')]_h, w_h\} \text{ for every } h \in H_r. \tag{B.2}$$

This definition is based on *persistence* by Alkan and Gale (2003).

Vectors such as w and w' are interpreted to be supplies of doctors, but they only specify how many doctors apply to each hospital and no information is given as to who these doctors are. Intuitively, the substitutability condition says that the number of accepted doctors at a hospital can increase only when the hospital has accepted all acceptable doctors under the original supply profile: To see this point, suppose that $[\tilde{\text{Ch}}_r(w)]_h < [\tilde{\text{Ch}}_r(w')]_h$. This assumption and the inequality in (B.2) imply $[\tilde{\text{Ch}}_r(w)]_h \geq w_h$. Since $[\tilde{\text{Ch}}_r(w)]_h \leq w_h$ holds by the definition of $\tilde{\text{Ch}}_r$, this implies $[\tilde{\text{Ch}}_r(w)]_h = w_h$.

The above definition is analogous to substitutability as defined in standard matching models, but there are two differences: (i) it is now defined on a region as opposed to a hospital, and (ii) it is defined over vectors that only specify how many doctors apply to hospitals in the region, and it does not distinguish different doctors.

Definition 4. A matching μ is **stable** with respect to $(\succeq_r)_{r \in R}$ if it is feasible and

(1) **Individual rationality**: $\mu_d \succeq_d \emptyset$ for each $d \in D$; $d \succeq_h \emptyset$ for all $h \in H$ and $d \in \mu_h$.
(2) **No blocking pair**: If $h \succ_d \mu_d$, then one of the following holds.
  (a) $\emptyset \succ_h d$.
  (b) $|\mu_h| = q_h$ and $d' \succ_h d$ for all $d' \in \mu_h$.
  (c') $\mu_d \notin H_r$ and $|\mu_{H_r}| = q_r$ for r such that $h \in H_r$, and $d' \succ_h d$ for all $d' \in \mu_h$.
  (d') $\mu_d \in H_r$, $w \succeq_r w'$ for r such that $h \in H_r$ where $w_{h'} = |\mu_{h'}|$ for all $h' \in H_r$ and $w'_h = w_h + 1$, $w'_{\mu_d} = w_{\mu_d} - 1$ and $w_{h'} = w_{h'}$ for all other $h' \in H_r$, and $d' \succ_h d$ for all $d' \in \mu_h$.

Given the above properties, we can think of the following (generalized) flexible deferred acceptance algorithm:

The Flexible Deferred Acceptance Algorithm For each region r, fix an associated quasi choice rule $\tilde{\text{Ch}}_r$ which satisfies condition (B.1). Note that the assumption that $\succeq_r$ is substitutable assures the existence of such a quasi choice rule.

(1) Begin with an empty matching, that is, a matching μ such that $\mu_d = \emptyset$ for all $d \in D$.
(2) Choose a doctor d arbitrarily who is currently not tentatively matched to any hospital and who has not applied to all acceptable hospitals yet. If such a doctor does not exist, then terminate the algorithm.
(3) Let d apply to the most preferred hospital $\bar{h}$ at $\succ_d$ among the hospitals that have not rejected d so far. If d is unacceptable to $\bar{h}$, then reject this doctor and go back to Step 2. Otherwise, let r be the region such that $\bar{h} \in H_r$ and define vector $x = (x_h)_{h \in H_r}$ by
   (a) $x_{\bar{h}}$ is the number of doctors currently held at $\bar{h}$ plus one, and
   (b) x_h is the number of doctors currently held at h if $h \neq \bar{h}$.
(4) Each hospital $h \in H_r$ considers the new applicant d (if $h = \bar{h}$) *and* doctors who are temporarily held from the previous step together. It holds its $(\tilde{\mathrm{Ch}}_r(x))_h$ most preferred applicants among them temporarily and rejects the rest (so doctors held at this step may be rejected in later steps). Go back to Step 2.

We define the **flexible deferred acceptance mechanism**, denoted φ^F, to be a mechanism that produces, for each input, the matching at the termination of the above algorithm.

B.1. **Associated Matching Model with Contracts.** It is useful to relate our model to a (many-to-many) matching model with contracts (Hatfield and Milgrom, 2005). Let there be two types of agents, doctors in D and regions in R. Note that we regard a region, as opposed to a hospital, as an agent in this model. There is a set of contracts $X = D \times H$.

We assume that, for each doctor d, any set of contracts with cardinality two or more is unacceptable. For each doctor d, her preference $\succ_d$ over $(\{d\} \times H) \cup \{\emptyset\}$ is given as follows.[22] We assume $(d,h) \succ_d (d,h')$ in this model if and only if $h \succ_d h'$ in the original model, and $(d,h) \succ_d \emptyset$ in this model if and only if $h \succ_d \emptyset$ in the original model.

For each region $r \in R$, we assume that the region has a preference $\succeq_r$ with an associated choice rule $\mathrm{Ch}_r(\cdot)$ over all subsets of $D \times H_r$. For any $X' \subset D \times H_r$, let $w(X') := (w_h(X'))_{h \in H_r}$ be the vector such that $w_h(X') = |\{(d,h) \in X' | d \succ_h \emptyset\}|$. For each X', the chosen set of contracts $\mathrm{Ch}_r(X')$ is defined by

$$\mathrm{Ch}_r(X') = \bigcup_{h \in H_r} \left\{ (d,h) \in X' | \#\{d' \in D | (d',h) \in X', d' \succeq_h d\} \leq (\tilde{\mathrm{Ch}}_r(w(X')))_h \right\}.$$

[22] We abuse notation and use the same notation $\succ_d$ for preferences of doctor d both in the original model and in the associated model with contracts.

That is, each hospital $h \in H_r$ chooses its $(\tilde{\text{Ch}}_r(w(X')))_h$ most preferred contracts available under X'.

We extend the domain of the choice rule to the entire class of all subsets of $D \times H$ by setting $\text{Ch}_r(X') = \text{Ch}_r(\{(d,h) \in X' | h \in H_r\})$ for any $X' \subseteq D \times H$.

Definition 5 (Hatfield and Milgrom (2005))**.** Choice rule $\text{Ch}_r(\cdot)$ satisfies the **substitutes condition** if there does not exist contracts $x, x' \in X$ and a set of contracts $X' \subseteq X$ such that $x' \notin \text{Ch}_r(X' \cup \{x'\})$ and $x' \in \text{Ch}_r(X' \cup \{x, x'\})$.

In other words, contracts are substitutes if the addition of a contract to the choice set never induces a hospital to take a contract it previously rejected. Hatfield and Milgrom (2005) show that there exists a stable allocation (defined in Definition 7) when contracts are substitutes for every hospital.

Definition 6 (Hatfield and Milgrom (2005))**.** Choice rule $\text{Ch}_r(\cdot)$ satisfies the **law of aggregate demand** if for all $X' \subseteq X'' \subseteq X$, $|\text{Ch}_r(X')| \leq |\text{Ch}_r(X'')|$.

Theorem 5. *For each $r \in R$, choice rule $\text{Ch}_r(\cdot)$ defined above satisfies the substitutes condition and the law of aggregate demand.*

Proof. Fix a region $r \in R$. Let X' be a subset of contracts and $x = (d,h) \in X \setminus X'$ where $h \in H_r$. Let $w = w(X')$ and $w' = w(X' \cup x)$. To show that Ch_r satisfies the substitutes condition, we consider a number of cases as follows.

(1) Suppose that $\emptyset \succ_h d$. Then $w' = w$ and, for each $h' \in H_r$, the set of acceptable doctors available at $X' \cup x$ is identical to the one at X'. Therefore, by inspection of the definition of Ch_r, we have $\text{Ch}_r(X' \cup x) = \text{Ch}_r(X')$, satisfying the conclusion of the substitutes condition in this case.

(2) Suppose that $d \succ_h \emptyset$. In this case, since $w' \geq w$, substitutability of $\tilde{\text{Ch}}_r$ implies that $\tilde{\text{Ch}}_r(w) \geq \tilde{\text{Ch}}_r(w') \wedge w$ or, equivalently,

$$[\tilde{\text{Ch}}_r(w)]_{h'} \geq \min\{[\tilde{\text{Ch}}_r(w')]_{h'}, w_{h'}\} \text{ for every } h' \in H_r. \tag{B.3}$$

(a) Consider a hospital $h' \in H_r \setminus h$. Note that we have $w'_{h'} = w_{h'}$. This and the inequality $[\tilde{\text{Ch}}_r(w')]_{h'} \leq w'_{h'}$ (which always holds by the definition of $\tilde{\text{Ch}}_r$) imply that $[\tilde{\text{Ch}}_r(w')]_{h'} \leq w_{h'}$. Thus we obtain $\min\{[\tilde{\text{Ch}}_r(w')]_{h'}, w_{h'}\} = [\tilde{\text{Ch}}_r(w')]_{h'}$. This and inequality (B.3) imply that

$$[\tilde{\text{Ch}}_r(w)]_{h'} \geq [\tilde{\text{Ch}}_r(w')]_{h'}. \tag{B.4}$$

Also observe that the set $\{d' \in D|(d',h') \in X'\}$ is identical to $\{d' \in D|(d',h') \in X' \cup x\}$, that is, the sets of doctors that are available to hospital h' are identical under X' and $X' \cup x$. This fact, inequality (B.4), and the definition of Ch_r imply that if $x' = (d',h') \notin \text{Ch}_r(X')$, then $x' \notin \text{Ch}_r(X' \cup x)$, obtaining the conclusion for the substitute condition in this case.

(b) Consider hospital h.

(i) Suppose that $[\tilde{\text{Ch}}_r(w)]_h \geq [\tilde{\text{Ch}}_r(w')]_h$. In this case we follow an argument similar to (but slightly different from) Case (2a): Note that the set $\{d' \in D|(d',h) \in X'\}$ is a subset of $\{d' \in D|(d',h) \in X' \cup x\}$, that is, the set of doctors that are available to hospital h under X' is smaller than under $X' \cup x$. These properties and the definition of Ch_r imply that if $x' = (d',h) \in X' \setminus \text{Ch}_r(X')$, then $x' \in X' \setminus \text{Ch}_r(X' \cup x)$, obtaining the conclusion for the substitute condition in this case.

(ii) Suppose that $[\tilde{\text{Ch}}_r(w)]_h < [\tilde{\text{Ch}}_r(w')]_h$. This assumption and inequality (B.3) imply $[\tilde{\text{Ch}}_r(w)]_h \geq w_h$. Since $[\tilde{\text{Ch}}_r(w)]_h \leq w_h$ holds by the definition of $\tilde{\text{Ch}}_r$, this implies $[\tilde{\text{Ch}}_r(w)]_h = w_h$. Thus, by the definition of Ch_r, any contract $(d',h) \in X'$ such that $d' \succ_h \emptyset$ is in $\text{Ch}_r(X')$. Equivalently, if $x' = (d',h) \in X' \setminus \text{Ch}_r(X')$, then $\emptyset \succ_h d'$. Then, again by the definition of Ch_r, it follows that $x' \notin \text{Ch}_r(X' \cup x)$ for any contract $x' = (d',h) \in X' \setminus \text{Ch}_r(X')$. Thus we obtain the conclusion of the substitute condition in this case.

To show that Ch_r satisfies the law of aggregate demand, simply note that $\tilde{\text{Ch}}_r$ is acceptant by assumption. This leads to the desired conclusion. □

A subset X' of X is said to be an **allocation** if it is individually rational for each agent, that is, (1) for any $d \in D$, $|\{(d,h) \in X'|h \in H\}| \leq 1$, and if $(d,h) \in X'$ then $h \succ_d \emptyset$, and (2) $\text{Ch}_r(X') = X'$ for any $X' \subseteq D \times H_r$.

Definition 7. A set of contracts $X' \subseteq X$ is a **stable allocation** if it is an allocation and

(1) $\cup_{r \in R} \text{Ch}_r(X') = X'$, and
(2) there exists no region $r \in R$, hospital $h \in H_r$, and a doctor $d \in D$ such that $(d,h) \succ_d x$ and $(d,h) \in \text{Ch}_r(X' \cup \{(d,h)\})$, where x is the contract that d receives at X' if any and $\emptyset$ otherwise.

When condition (2) is violated by some (d,h), we say that (d,h) blocks X' or (d,h) is a block of X'.

Given any allocation X', define a **corresponding matching** $\mu(X')$ in the original model by setting $\mu_d(X') = h$ if and only if $(d,h) \in X'$ and $\mu_d(X') = \emptyset$ if and only if no contract associated with d is in X'. Since each doctor regards any set of contracts with cardinality of at least two as unacceptable, each doctor receives at most one contract at X' and hence $\mu(X')$ is well defined for any allocation X'.

Theorem 6. *If X' is a stable allocation in the associated model with contracts, then the corresponding matching $\mu(X')$ is a stable matching in the original model.*

Proof. Suppose that X' is a stable allocation in the associated model with contracts and denote $\mu := \mu(X')$. Individual rationality of μ is obvious from the construction of μ. To show that there is no blocking pair for μ, assume that $h \succ_d \mu_d$ for some $d \in D$ and $h \in H$. Further assume that $d \succ_h \emptyset$ and, moreover, $|\mu_h| < q_h$ or $d \succ_h d'$ for some $d' \in \mu_h$ in negation of conditions (a) and (b) of the definition of stability (Definition 4). Let r be a region such that $h \in H_r$. By the definition of stability, it suffices to show that the following conditions (B.5) and (B.6) hold if $\mu_d \notin H_r$, and (B.5), (B.6) and (B.7) hold if $\mu_d \in H_r$,

$$|\mu_{H_r}| = q_r, \tag{B.5}$$

$$d' \succ_h d \text{ for all } d' \in \mu_h, \tag{B.6}$$

$$w \succeq_r w', \tag{B.7}$$

where $w = (w_h)_{h \in H_r}$ is defined by $w_{h'} = |\mu_{h'}|$ for all $h' \in H_r$ while $w' = (w'_h)_{h \in H_r}$ is defined by $w'_h = w_h + 1$, $w'_{\mu_d} = w_{\mu_d} - 1$ (if $\mu_d \in H_r$) and $w_{h'} = w_{h'}$ for all other $h' \in H_r$.

Claim 1. *Conditions (B.5) and (B.6) hold (irrespectively of whether $\mu_d \in H_r$ or not).*

Proof. First note that the assumption that $h \succ_d \mu_d$ implies that $(d,h) \succ_d x$ where x denotes the (possibly empty) contract that d signs under X'. Let $w'' = (w''_h)_{h \in H_r}$ be defined by $w''_h = w_h + 1$ and $w''_{h'} = w_{h'}$ for all other $h' \in H_r$.

(1) Assume, for contradiction, that condition (B.6) is violated, that is, $d \succ_h d'$ for some $d' \in \mu_h$. First, by consistency of $\tilde{\text{Ch}}_r$, we have $[\tilde{\text{Ch}}_r(w'')]_h \geq [\tilde{\text{Ch}}_r(w)]_h$.[23] This implies that weakly more contracts involving h are signed at $X' \cup (d,h)$ than at X'. This property, together with the assumptions that $d \succ_h d'$ and that

[23]To show this claim, assume for contradiction that $[\tilde{\text{Ch}}_r(w'')]_h < [\tilde{\text{Ch}}_r(w)]_h$. Then, $[\tilde{\text{Ch}}_r(w'')]_h < [\tilde{\text{Ch}}_r(w)]_h \leq w_h$. Moreover, since $w''_{h'} = w_{h'}$ for every $h' \neq h$ by construction of w'', it follows that $[\tilde{\text{Ch}}_r(w'')]_{h'} \leq w''_{h'} = w_{h'}$. Combining these inequalities, we have that $\tilde{\text{Ch}}_r(w'') \leq w$. Also we have $w \leq w''$ by the definition of w'', so it follows that $\tilde{\text{Ch}}_r(w'') \leq w \leq w''$. Thus, by consistency of $\tilde{\text{Ch}}_r$, we obtain $\tilde{\text{Ch}}_r(w'') = \tilde{\text{Ch}}_r(w)$, a contradiction to the assumption $[\tilde{\text{Ch}}_r(w'')]_h < [\tilde{\text{Ch}}_r(w)]_h$.

$(d', h) \in X'$ imply that $(d, h) \in \text{Ch}_r(X' \cup (d, h))$.[24] Thus, together with the above-mentioned property that $(d, h) \succ_d x$, (d, h) is a block of X' in the associated model of matching with contract, contradicting the assumption that X' is a stable allocation.

(2) Assume, for contradiction, that condition (B.5) is violated, so that $|\mu_{H_r}| \neq q_h$. Then, since $|\mu_{H_r}| \leq q_r$ by the construction of $\mu = \mu(X')$ and the assumption that X' is an allocation, it follows that $|\mu_{H_r}| < q_r$. Then $(d, h) \in \text{Ch}_r(X' \cup (d, h))$ because,

 (a) $d \succ_h \emptyset$ by assumption,

 (b) since $\sum_{h \in H_r} w_h = \sum_{h \in H_r} |\mu_h| = |\mu_{H_r}| < q_r$, it follows that $\sum_{h \in H_r} w''_h = \sum_{h \in H_r} w_h + 1 \leq q_r$. Moreover, $|\mu_h| < q_h$ by assumption and (B.6), so $w''_h = |\mu_h| + 1 \leq q_h$. These properties and the assumption that $\tilde{\text{Ch}}_r$ is acceptant imply that $\tilde{\text{Ch}}_r(w'') = w''$. In particular, this implies that all contracts $(d', h) \in X' \cup (d, h)$ such that $d' \succ_h \emptyset$ is chosen at $\text{Ch}_r(X' \cup (d, h))$.

Thus, together with the above-mentioned property that $(d, h) \succ_d x$, (d, h) is a block of X' in the associated model of matching with contract, contradicting the assumption that X' is a stable allocation.

□

To finish the proof of the theorem suppose that $\mu_d \in H_r$ and, for contradiction, that (B.7) fails, that is, $w' \succ_r w$. Then it should be the case that $[\tilde{\text{Ch}}_r(w'')]_h = w''_h = w_h + 1 = |\mu_h| + 1$.[25] Also we have $|\mu_h| < q_h$ and hence $|\mu_h| + 1 \leq q_h$ and $d \succ_h \emptyset$, so

$$(d, h) \in \text{Ch}_r(X' \cup (d, h)).$$

This relationship, together with the assumption that $h \succ_d \mu_d$, and hence $(d, h) \succ_d x$, is a contradiction to the assumption that X' is stable in the associated model with contracts.

□

[24] The proof of this claim is as follows. $\text{Ch}_r(X')$ induces each hospital $h' \in H_r$ to select its $[\text{Ch}_r(X')]_{h'}$ most preferred contracts while $\text{Ch}_r(X' \cup (d, h))$ induces each hospital to select a weakly larger number $[\text{Ch}_r(X' \cup (d, h))]_{h'}$ of its most preferred contracts. Since (d', h) is selected as one of $[\text{Ch}_r(X')]_{h'}$ most preferred contracts for h at X' and $d \succ_h d'$, we conclude that (d, h) should be one of $[\text{Ch}_r(X' \cup (d, h))]_{h'} \geq [\text{Ch}_r(X')]_{h'}$ most preferred contracts at $X' \cup (d, h)$, thus selected at $X' \cup (d, h)$.

[25] To show this claim, assume for contradiction that $[\tilde{\text{Ch}}_r(w'')]_h \leq w_h$. Then, since $w''_{h'} = w_{h'}$ for any $h' \neq h$ by the definition of w'', it follows that $\tilde{\text{Ch}}_r(w'') \leq w \leq w''$. Thus by consistency of $\tilde{\text{Ch}}_r$, we obtain $\tilde{\text{Ch}}_r(w'') = \tilde{\text{Ch}}_r(w)$. But $\tilde{\text{Ch}}_r(w) = w$ because X' is a stable allocation in the associated model of matching with contracts, so $\tilde{\text{Ch}}_r(w'') = w$. This is a contradiction because $w' \leq w''$ and $w' \succ_r w$ while $\tilde{\text{Ch}}_r(w'') \in \arg\max_{\succeq_r} \{w''' | w''' \leq w''\}$.

Since $\mathrm{Ch}_r(\cdot)$ satisfies the substitutes condition for each r, there exists a **doctor-optimal (doctor-pessimal) stable allocation** in the matching model with contracts, that is, a stable allocation that every doctor weakly prefers to every other stable allocation (Hatfield and Milgrom, 2005). Moreover, if choice rules of all regions satisfy substitutes and the law of aggregate demand, then the doctor-optimal stable mechanism (the mechanism that produces the doctor-optimal stable allocation for any input) is group strategy-proof. In particular, the doctor-optimal stable mechanism is strategy-proof.

We will show that the flexible DA is "isomorphic" to the doctor-optimal stable mechanism in the model with contracts.

Theorem 7. *The doctor-optimal stable allocation in the matching model with contracts, X', exists. The relation $\mu(X') = \mu$ holds, where μ is the matching produced by the flexible DA.*

Proof. First observe that the doctor-optimal stable allocation in matching with contracts can be found by the cumulative offer process (Hatfield and Milgrom, 2005; Hatfield and Kojima, 2009). Then, we observe that each step of the flexible DA corresponds to a step of the cumulative offer process, that is, at each step, if d proposes to h in flexible DA, then at the same step of the cumulative offer process, contract (d, h) is proposed. Moreover, for each region, the set of doctors accepted for hospitals in the region at the step of flexible DA corresponds to the set of contracts held by the region in the cumulative offer process. $\square$

Theorems 6 and 7 imply that the flexible DA finds a stable matching. Also, Theorems 5 and 7 imply that the flexible DA is (group) strategy-proof, as the substitutes condition and the law of aggregate demand imply that any mechanism that selects the doctor-optimal stable allocation is (group) strategy-proof (Hatfield and Milgrom, 2005; Hatfield and Kojima, 2008; Hatfield and Kominers, 2010)

B.2. **Stability in The Main Text.** Given the target capacity profile $(\bar{q}_h)_h$ and the weight vector w, define the **ordered excess weight function** η by setting $\eta_i(w)$ to be the i'th lowest value (allowing repetition) of $\{w_h - \bar{q}_h | h \in H_r\}$ (we suppress dependence of η on $\bar{q}$). For example, if $w = (w_{h_1}, w_{h_2}, w_{h_3}, w_{h_4}) = (2, 4, 7, 2)$ and $(\bar{q}_{h_1}, \bar{q}_{h_2}, \bar{q}_{h_3}, \bar{q}_{h_4}) = (3, 2, 3, 0)$, then $\eta_1(w) = -1, \eta_2(w) = \eta_3(w) = 2, \eta_4(w) = 4$.

Consider the regional preference $\succeq_r$ that compares the excess weights lexicographically. More specifically, let $\succeq_r$ be such that $w \succ_r w'$ if and only if there exists i such that $\eta_j(w) = \eta_j(w')$ for all $j < i$ and $\eta_i(w) > \eta_i(w')$. We call this a **Rawlsian regional preference**.

Proposition 1. *Stability as defined in the main text is a special case of the general stability with respect to a Rawlsian regional preference.*

Proof. Let w be defined by $w_{h'} = |\mu_{h'}|$ for each $h' \in H_r$ and w' by $w'_h = w_h + 1$, $w'_{\mu_d} = w_{\mu_d} - 1$, and $w'_{h'} = w_{h'}$ for all other $h' \in H_r$. It suffices to show that $w \succeq_r w'$ if and only if $|\mu_h| + 1 - \bar{q}_h > |\mu_{\mu_d}| - 1 - \bar{q}_{\mu_d}$.

Suppose that $|\mu_h| + 1 - \bar{q}_h > |\mu_{\mu_d}| - 1 - \bar{q}_{\mu_d}$. This means that

$$w_h + 1 - \bar{q}_h > w_{\mu_{\mu_d}} - 1 - \bar{q}_{\mu_d}$$
$$\iff w_h - \bar{q}_h = w_{\mu_{\mu_d}} - 1 - \bar{q}_{\mu_d} \text{ or } w_h - \bar{q}_h \geq w_{\mu_{\mu_d} - \bar{q}_{\mu_d}}.$$

In the former case, obviously $\eta(w) = \eta(w')$, so $w \succeq_r w'$. In the latter case, $w \succ_r w'$ because $\eta_j(w) = \eta_j(w')$ for all $j < i$ and $\eta_i(w) = \eta_i(w') + 1 > \eta_i(w')$, where i is the index for μ_d at w.

Suppose that $|\mu_h| + 1 - \bar{q}_h \leq |\mu_{\mu_d}| - 1 - \bar{q}_{\mu_d}$. This obviously means that $w' \succ_r w$. $\square$

Proposition 2. *A Rawlsian preference is substitutable (and the choice rule described in the flexible deferred acceptance algorithm provides an associated choice rule).*

Proof. It is clear that the quasi choice rule described in the flexible deferred acceptance algorithm, denoted $\tilde{\mathrm{Ch}}_r$, satisfies the substitutes condition, consistency and the law of aggregate demand. Thus in the following, we will show that $\tilde{\mathrm{Ch}}_r$ indeed satisfies $\tilde{\mathrm{Ch}}_r(w) \in \arg\max_{\succeq_r}\{x \leq w\}$ for each w. Let $w' = \tilde{\mathrm{Ch}}_r(w)$. For contradiction, suppose that $\tilde{\mathrm{Ch}}_r(w) \notin \arg\max_{\succeq_r}\{x \leq w\}$ and consider an arbitrary $w'' \in \arg\max_{\succeq_r}\{x \leq w\}$. Then we have $w'' \succ_r w'$, so there exists i such that $\eta_j(w'') = \eta_j(w')$ for every $j < i$ and $\eta_i(w'') > \eta_i(w')$. Since $\sum_j \eta_j(w'') = \sum_j \eta_j(w') = q_r$, there exists some k such that $\eta_k(w'') < \eta_k(w')$. Let $l = \{k|\eta_k(w'') < \eta_k(w')\}$ be the smallest of such indices. Then since $l > i$, we have $\eta_i(w') < \eta_i(w'') \leq \eta_l(w'') < \eta_l(w')$. Thus it should be the case that $\eta_i(w') + 2 \leq \eta_l(w')$. By the construction of $\tilde{\mathrm{Ch}}_r$, that is possible only if $\eta_i(w') = w_h$, where h is the hospital corresponding to i at w'. Now it should be the case that $w''_h = w_h$ as well, because otherwise $w'' \notin \arg\max_{\succeq_r}\{x \leq w\}$. Now consider the modified vectors of both w' and w'' that delete the entries corresponding to h. All the properties described above hold for these new vectors. Proceeding inductively, at some step $\eta_i(w') + 2 \leq \eta_l(w')$ but $w'_h \neq w_h$. This is a contradiction to the construction of $\tilde{\mathrm{Ch}}_r$. $\square$

Appendix C. Proof of Theorem 4

Proof. First, suppose, to the contrary, that $\mu'_d \succ_d \mu_d$ for some $d \in D$. This implies that d is rejected by μ'_d in deferred accceptance at some step. Since d is acceptable by μ'_d (as μ'_d

is stable), at that step, there has already exist $q_{\mu'_d}$ doctors in hospital μ'_d, all of whom are preferred to d by μ'_d. This implies that there exists $d' \in \mu_{\mu'_d} \setminus \mu'_{\mu'_d}$. Take such d'. Notice that $\mu'_{d'} \succ_{d'} \mu_{d'}$.

Now, repeat the same argument by replacing d by d' to find another doctor d'' such that $\mu'_{d''} \succ_{d''} \mu_{d''}$. Continuing this process, we find an infinite sequence of doctors, $(d, d', d'', \dots)$, such that for each d^* in the sequence, $\mu'_{d^*} \succ_{d^*} \mu_{d^*}$ holds. But the finiteness of D implies that we can find at least two doctors such that these doctors appear infinitely often. Let the first doctor who appears twice in this sequence be $\hat{d}$. Let D' be the set of doctors who are within $\hat{d}$'s first appearance and the second appearance in the sequence, and doctor $\hat{d}$ herself. Let μ^* be the matching such that, starting from μ, each doctor $d^* \in D'$ is moved from μ_{d^*} to μ'_{d^*}. Notice that, relative to μ, μ^* would make all doctors weakly better off with some doctors being strictly better off. Notice also that all hospitals μ_{d^*} for $d^* \in D'$ have binding capacity constraints in μ'. Hence, stability of μ' (implied by the assumption that μ' is generated by the flexible deferred acceptance mechanism) implies that μ^* is stable in the context of no regional caps. But this contradicts the fact that the deferred acceptance mechanism is doctor-optimal.

$\mu' \succeq_d \mu''$ for all $d \in D$ is obvious, since the JRMP mechanism corresponds to deleting step (4)-(b) from the flexible deferred acceptance mechanism.

The second part of the theorem's statement is an immediate corollary of the first. □

Appendix D. Semi-strong stability

Definition 8. A matching μ is **semi-strongly stable** with respect to a target capacity $(\bar{q}_h)_{h\in H}$ if it is feasible and

(1) **Individual rationality**: $\mu_d \succeq_d \emptyset$ for each $d \in D$; $d \succeq_h \emptyset$ for all $h \in H$ and $d \in \mu_h$.

(2) **No blocking pair**: If $h \succ_d \mu_d$, then one of the following holds.
  - (a) $\emptyset \succ_h d$.
  - (b) $|\mu_h| = q_h$ and $d' \succ_h d$ for all $d' \in \mu_h$.
  - (c') $\mu_d \notin H_r$ and $|\mu_{H_r}| = q_r$ for r such that $h \in H_r$, and $d' \succ_h d$ for all $d' \in \mu_h$.
  - (d') $|\mu_h| \geq \bar{q}_h$, $|\mu_{\mu_d}| \leq \bar{q}_{\mu_d}$, $|\mu_{H_r}| = q_r$ for r such that $h \in H_r$, and $d' \succ_h d$ for all $d' \in \mu_h$.

Condition 2(d') says that a pair of a doctor d and a hospital h is not deemed as a valid blocking pair if the doctor is currently matched with a hospital h' in the same region as hospital h, the number of doctors matched with hospital h' is no more than its target and that of hospital h is no less than its target. That is, the pair that moves the matching unambiguously away from the target capacity, which is "not justifiable," is not deemed as a valid blocking pair.

Notice that when we argue what is deemed as a blocking pair and what is not, we implicitly and informally assume a notion of distance between two matchings. Although it may be "intuitively right" to argue that a blocking that moves a doctor from a hospital with less than its target capacity to a hospital with no less than its target capacity, it may not be clear if a blocking that moves a doctor from a hospital with no less than its target capacity to a hospital with no less than its target capacity should be forbidden. For example, if hospital h_1 has a target 1 and $|\mu_{h_1}| = 10$, hospital h_2 has a target 5 and $|\mu_{h_2}| = 5$, and these two hospitals are in the same region, then it is not clear if a blocking that moves a doctor from hospital h_1 to h_2 must be forbidden.[26]

Although the definition of semi-strong stability seems to be a natural first attempt to weaken the notion of strong stability, unfortunately it has the same deficiency as strong stability: It does not necessary exist, and there does not exist a mechanism that selects a semi-strongly stable matching whenever there exists one.

The following example shows that a semi-strongly stable matching may not exist.

Example 12. *semi-strongly stable matching may not exist.*

There is one region r with regional cap $q_r = 1$, in which three hospitals, h_1, h_2 and h_3, reside. Each hospital h has a capacity of $q_h = 1$. Suppose that there are two doctors, d_1 and d_2. Targets for hospitals are $(\bar{q}_{h_1}, \bar{q}_{h_2}, \bar{q}_{h_3}) = (0, 0, 1)$. We assume the following preference:

$$\succ_{h_1}: d_1, d_2, \qquad \succ_{h_2}: d_2, d_1, \qquad \succ_{h_3}: \text{ arbitrary};$$
$$\succ_{d_1}: h_2, h_1, \qquad \succ_{d_2}: h_1, h_2.$$

Note that in any feasible matching, at most 1 doctor is matched, because the regional cap is 1. Also, note that at least 1 doctor is matched in any stable matching, as otherwise any pair of a doctor and a hospital can constitute a blocking pair. Finally, no doctor is matched to h_3 in any semi-strongly stable matching because such a matching is individually irrational. Now, suppose that in a strongly stable matching, $\mu_{h_1} = \{d_1\}$. But this is impossible, as a pair (d_1, h_2) can constitute a blocking pair. Next, suppose that in a strongly stable matching, $\mu_{h_1} = \{d_2\}$. But again this is impossible, as a pair (d_1, h_1) can constitute a blocking pair (h_1 can reject d_2 to be paired with d_1). By a symmetric argument, $\mu_{h_2} = \{d_2\}$ and $\mu_{h_2} = \{d_1\}$ cannot be strongly stable. Therefore, a semi-strongly stable matching does not exist in this example.

[26]Note that, for this example to work, we need at least one more hospital in the same region, as the sum of the target capacity in a region must be no more than the regional cap.

Notice that, in Example 12 is very similar to Example 4. Specifically, the only modification is about the target: In Example 4, target capacities are not explicitly defined, as strong stability is defined independent of the specification of target capacities. In Example 12, one hospital that should have no effect on the resulting matching (as it is not acceptable by any doctors) is added, and all targets are placed on this hospital. Consequently, the existing hospitals get targets of only 0, so any number of doctors matched to these hospitals is no less than the target. Hence condition (2d') in the definition of semi-strong stability has no bite, which enables the same example to work here. In an analogous manner, we can easily modify Example 5 to construct an example in which it is not possible to construct a strategy-proof mechanism that always finds a semi-strongly stable matching, whenever there exists one.

The nonexistence of a strongly stable matching may justify our weaker stability concept. As we argued before, although the concept of weak stability is strong enough to rule out the JRMP mechanism, it is "too weak" to be ideal. The next natural idea to weaken the concept of strong stability would be to make the notion of "distance" from the target (as discussed when we defined semi-strong stability) more restrictive. In semi-strongly stable matching, for any pair of a doctor and a hospital, there should not exist a measure of distance from the target that justifies its moves. In the next definition, we change the order of the statements: Fix the measure of the distance first. Then, there should not exists a pair of a doctor and a hospital that want to deviate from the current matching.