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 in rural hospitals, the Japanese government recently introduced "regional caps" which restrict the total number of residents matched within each region of the country. To accommodate regional caps, the government modified the deferred acceptance mechanism in a particular manner. Motivated by this policy change, we study the design of matching markets under constraints on doctor distribution. This paper shows that the Japanese mechanism may result in avoidable inefficiency and instability and proposes a better mechanism that improves upon it in terms of efficiency and stability while respecting the regional caps.
JEL Classification Numbers: C70, D61, D63.
Keywords: medical residency matching, regional caps, the rural hospital theorem, stability, strategy-proofness, matching with contracts

Date: December 27, 2010.
We are grateful to Mustafa Oguz Afacan, Petér Biró, Erich Budish, Sylvain Chassang, Hisao Endo, Clayton Featherstone, Daniel Fragiadakis, Drew Fudenberg, Tadashi Hashimoto, John William Hatfield, Toshiaki Iizuka, Ryo Jinnai, Onur Kesten, Scott Duke Kominers, Hideo Konishi, Mihai Manea, Taisuke Matsubae, Aki Matsui, Yusuke Narita, Muriel Niederle, Parag Pathak, Al Roth, Dan Sasaki, Tayfun Sönmez, Satoru Takahashi, William Thomson, Alexis Akira Toda, Kentaro Tomoeda, Utku Ünver, Jun Wako, Alex Westkamp, Yosuke Yasuda, and seminar participants at Arizona State, Boston College, Caltech, Carnegie Mellon, Collegio Carlo Alberto, Harvard/MIT, Montreal, Paris School of Economics, Rice, Stanford, Texas A&M, Tohoku, Tokyo, USC, 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, Yoshiaki Kanno, Masataka Kawana, and Masaaki Nagano answered our questions about medical residency in Japan and introduced us to the relevant medical literature. Seung Hoon Lee and Pete Troyan provided excellent research assistance.

1. Introduction

The 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 sufficient numbers of doctors to meet their demands. For instance, a Washington Post article entitled "Shortage of Doctors Affects Rural U.S." describes a dire situation in the United States (Talbott, 2007):

> The government estimates that more than 35 million Americans live in underserved areas, and it would take 16,000 doctors to immediately fill that need, according to the American Medical Association.

Similar problems are present around the world. For example, one can easily find reports of doctor shortages in rural areas in the United Kingdom, India, Australia, and Thailand.[1]

One may wonder if the situation can be improved by appropriately designing a centralized matching mechanism for medical residents, an important part of labor supply for hospitals. However, the existing 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 recently become a hot political issue in Japan, where the deferred acceptance algorithm (Gale and Shapley, 1962) has placed around 8,000 doctors (mostly consisting of graduating medical students) in about 1,500 residency programs each year since 2003. In an attempt to increase the placement of residents in 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.

Motivated by this policy change, we study the design of matching markets under constraints on the doctor distribution. This paper shows that the current Japanese mechanism, which we call the Japan Residency Matching Program (JRMP) mechanism, may result in avoidable instability and inefficiency despite its resemblance to the deferred acceptance algorithm. We then propose an alternative mechanism that overcomes these

[1] Shallcross (2005), Alcoba (2009), Nambiar and Bavas (2010), and Wongruang (2010).

shortcomings while respecting the regional caps. 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 acceptance 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 nor doctor-pessimal nor hospital-pessimal stable matchings. Second, different stable matchings can leave different hospitals with unfilled positions, implying that the conclusion of 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 relate 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.[2] This situation can be analyzed by 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 students 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 formalized by Abdulkadiroğlu and Sönmez (2003). Lastly, the shortage of doctors in rural areas is

[2]The alleged imbalance of doctors in different specialties is regarded as a serious problem. Specialties such as radiology and dermatology are popular while others such as primary care and pediatrics are not.

a common problem around the globe. Countries mentioned above, such as the United States, the United Kingdom, and India, are just 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 abstract technical issues associated with regional caps 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 for that market. Improving the Japanese medical market is important by itself, which produces around 8,000 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 last paragraph. In this sense, our paper contributes to the general research agenda of market design, advocated by Roth (2002) for instance, that emphasizes the importance of addressing issues arising in practical allocation problems.

Related literature. This section discusses papers related to this study. The medical literature on doctor shortage and the Japanese situation is discussed in the next section.

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), and Hatfield and Milgrom (2005), 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, and thus there are few studies offering solutions to this problem. The current paper explores possible ways to offer some positive results.

Roth (1991) points out that some hospitals in the United Kingdom prefer to hire no more than one female doctor while offering multiple positions. Similarly, some schools (or school districts) desire certain diversity characteristics of their incoming classes such as ethnicity and academic performance (Abdulkadiroğlu and Sönmez, 2003; Abdulkadiroğlu, 2005; Ergin and Sönmez, 2006). Westkamp (2010) considers a college admission problem in which colleges have admission criteria based on trait-specific quotas. If one regards a region (instead of a hospital) as a single agent in our model, these models and ours appear similar in that an agent in both models has certain "preferences" over distributions more complex than responsive ones. However, the above models are different from ours. For instance, in our model, a distinction should be made between a matching of a doctor to one hospital in a region and a matching of the same doctor to a different hospital in the same region, but such a distinction cannot be even described in the former models. This distinction is essential in the context of residency matching because a doctor may have incentives to deviate by moving between hospitals within a single region. Thus results from these papers cannot be applied in this paper's environment.

Despite the above-mentioned difficulty, there is a way to make an association between our model and an existing model, namely the model of matching with contracts as defined by Hatfield and Milgrom (2005). More specifically, given a matching market with regional caps, one can define an associated matching model with contracts such that a stable allocation in the latter model induces a stable matching in the former. This correspondence allows us to show some of our results by using properties of the matching with contracts model established by Hatfield and Milgrom (2005), Hatfield and Kojima (2008, 2009), and Hatfield and Kominers (2009, 2010).[3] On the other hand, it is also worth noting that these models are still different. The reason is that certain types of blocks allowed in the matching model with contracts are considered infeasible in our context. Thus stable allocations in a matching model with contracts can induce only a subset of stable matchings in our model. For this reason, the structural properties of the set of stable matchings in our model are strikingly different from those in the matching model with contracts. For instance, a doctor-optimal stable allocation exists and the conclusion of the rural hospital theorem holds in their model but not in ours.[4]

Abraham, Irving, and Manlove (2007) study allocation of students to projects where a lecturer may offer multiple projects. Both projects and lecturers have capacity constraints. Sönmez and Ünver (2006) analyze a related model in the context of school choice in which there may be multiple school programs in a school building.[5] Their models are analogous to ours if we associate a lecturer and a project – and a school building and a school, respectively– in their models to a region and a hospital in our model, respectively. However, there are two notable differences. First, they assume that preferences of all projects provided by the same lecturer (school programs in the same building) are identical

[3]Note that residency matching and school choice with balance requirements mentioned in the last paragraph (Roth, 1991; Abdulkadiroğlu and Sönmez, 2003) can be modeled as special cases of this paper's model. A related issue appears in the National Resident Matching Program where a hospital may have multiple types of residency positions (Roth and Peranson, 1999; Niederle, 2007).

[4]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).

[5]Motivated by the matching system for higher education in Hungary, Biró, Fleiner, Irving, and Manlove (2010) extend these models to cases in which capacity constraints are imposed on a nested system of sets.

while such a restriction is not imposed in our model.[6] Second, the stability concepts employed in their models are different from ours, thus our results do not reduce to theirs even in their more specialized settings.

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 primarily about object allocation, and stability is not studied), 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.[7] 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 (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 motivated by an anti-trust lawsuit against the American medical resident matching clearinghouse are investigated by Bulow and Levin (2006), Kojima (2007), Konishi and Sapozhnikov (2008), Niederle (2007), and Niederle and Roth (2003). A classical resource allocation problem with multi-unit demand has attracted renewed attention in the context of practical course allocation at business schools as studied by Sönmez and Ünver (2010), Budish and Cantillon (2010), and Budish (2010). Initiated by Roth, Sönmez, and Ünver (2004, 2005, 2007), even the organ transplantation problem has become a subject of market design research in recent years. 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 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 weakly stable or efficient matching. Section 5 introduces and

[6] In our context, it is important to allow different hospitals in the same region to have different preferences because two hospitals rarely have identical preferences in practice.

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

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

In Japan, about 8,000 doctors and 1,500 residency programs participate in the matching process each year. This section describes how this process has evolved and how it has affected the debate on the 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.[8]

The 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 allocations of doctors against their will.[9] Describing the 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 cause many of the good properties of the deferred acceptance algorithm to fail.

Although the matching system was welcomed by many, it has 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 doctors to work for urban hospitals rather than rural hospitals, resulting in severe doctor shortages in rural areas. While

[8]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/).

[9]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.

there is no conclusive evidence supporting their claim, an empirical study by Toyabe (2009) finds that the geographical imbalance of doctors has increased in recent years according to several measures (the Gini coefficient, Atkinson index, and Theil index of the per-capita number of doctors across regions). By contrast, he also finds that the imbalance is lower when residents are excluded from the calculation. Based on these findings, he suggests that the matching system introduced in 2003 may have contributed to the widening regional imbalance of doctors.

To put such criticisms into context, we note that the 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, 2005b)[10] and many more who were allegedly underserved. One government official told one of the authors (personal communication) that the regional imbalance is one of the most important problems in the government's health care policy, together with financing health care cost. Popular media regularly report stories of doctor shortages, often in a very sensational tone.[11] There is evidence that the sufficient staffing of doctors in hospitals 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 may lead to 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 sum of the hospital capacities in a region exceeds its regional cap, then the capacity of each hospital is reduced to equalize the total capacity with the regional cap.[12] Then the deferred acceptance algorithm is

[10]A mui-chiku is defined by various criteria such as the ease of access to hospitals, the population, the regularity of clinic openings, and so forth (Ministry of Health, Labour and Welfare, 2005a).

[11]For instance, the Yomiuri Shimbun newspaper, with circulation of over 10,000,000, recently provoked a controversy by its article about the only doctor in Kamikoani-mura village, where 2,800 people live (Yomiuri Shimbun newspaper, 03/19/2010). Although the doctor, aged 65, took only 18 days off a year, she was persistently criticized by some "unreasonable demanding" patients. When she announced that she wanted to quit (which means that the village will be left with no doctor) because she was "exhausted," 600 signatures were collected in only 10 days, to change her mind.

[12] The capacity of a hospital is reduced proportionately to its original capacity in principle (subject to integrality constraints) although there are a number of fine adjustments and exceptions. This rule 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 can be considered exogenous.

implemented 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.

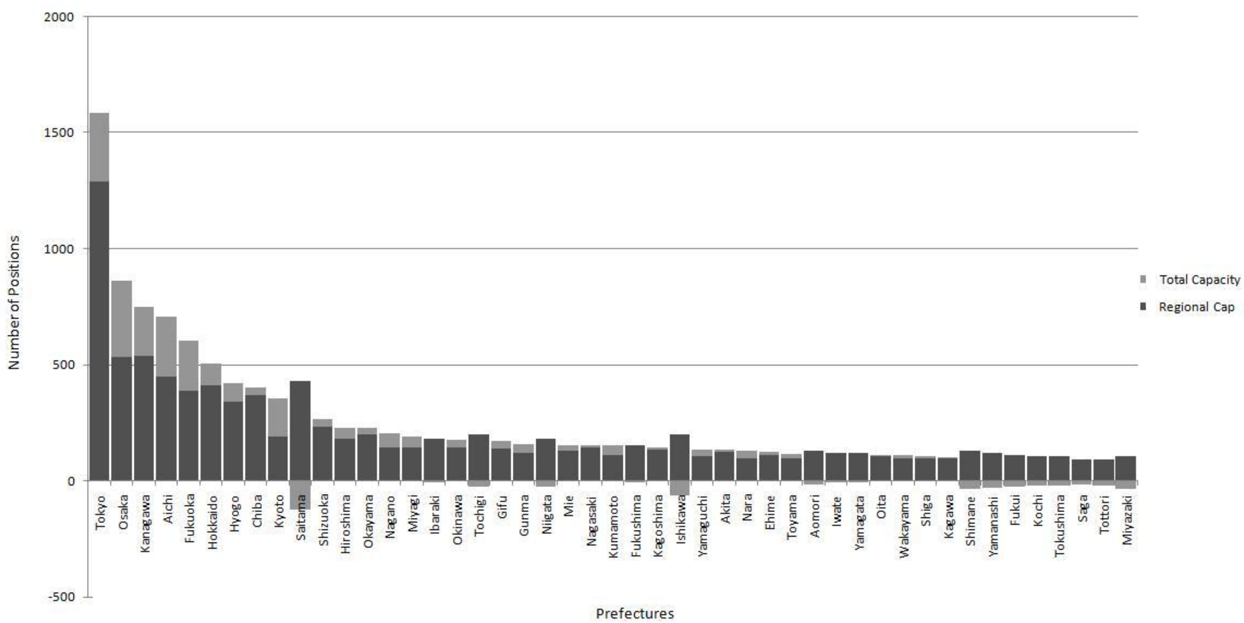

FIGURE 1. **Regional caps and total capacities.** For each prefecture, the total capacity is the sum of advertised positions in hospitals located in the prefecture in 2008. The regional caps are based on the government's plan in 2008 (Ministry of Health, Labour and Welfare, 2009a). Negative values of total capacities in some prefectures indicate the excess amount of regional caps beyond the advertised positions.

The magnitude of the regional caps is illustrated in Figure 1. Relatively large reductions are imposed on urban areas. For instance, hospitals in Tokyo and Osaka advertised 1,582 and 860 positions in 2008, respectively, but the government set the regional caps of 1,287 and 533, the largest reductions in the number of positions. The largest reduction in proportion is imposed on Kyoto, which offered 353 positions in 2008 but the number is dropped to 190, a reduction of about 46 percent. Indeed, the projected changes were so large that the government provided a temporary measure that limits per-year reductions

More specifically, the government decides the physical capacity of a hospital based on verifiable information such as the number of beds in it.

within a certain bound in the first years of operation, though the plan is to reach the planned regional cap eventually. In total, 34 out of 47 prefectures are given regional caps smaller than the numbers of advertised positions in 2008.

The new JRMP mechanism with regional caps was used in 2009 for the first time. The government claims that the change alleviated the 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, 2009b).[13] However, 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.[14] Among other things, a commonly expressed concern is that the current system with regional caps causes efficiency losses, for instance by preventing residents from learning their desired skills for practicing medical treatments. In the subsequent sections, we offer a theoretical framework to formally analyze these issues that arise in matching markets with regional caps.

3. Model

Let there be a finite set of doctors D and a finite set of hospitals H.[15] 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.

Doctor d is said to be **acceptable** to h if $d \succ_h \emptyset$.[16] Similarly, h is acceptable to d if $h \succ_d \emptyset$. It will turn out that only rankings of acceptable partners matter for our analysis,

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

[14] 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.

[15] We follow the convention in the literature to refer to a residency program as a "hospital."

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

so we often write only acceptable partners 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.

Each hospital $h \in H$ is endowed with a (physical) **capacity** q_h, which is a nonnegative integer. 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 $|D'| > q_h$.

In words, preference relation $\succ_h$ is responsive with a capacity if the ranking of a doctor (or keeping a position vacant) is independent of her colleagues, and any set of doctors exceeding its capacity is unacceptable. We assume that preferences of each hospital h are responsive with capacity q_h 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 $h \in H$, let $r(h)$ denote the region r such that $h \in H_r$. For each region $r \in R$, there is a **regional cap** q_r, which is a nonnegative integer.

A **matching** μ is a mapping that satisfies (i) $\mu_d \in H \cup \{\emptyset\}$ for all $d \in D$, (ii) $\mu_h \subseteq D$ for all $h \in H$, and (iii) for any $d \in D$ and $h \in H$, $\mu_d = h$ if and only if $d \in \mu_h$. That is, a matching simply specifies which doctor is assigned to which hospital (if any). A matching is **feasible** if $|\mu_r| \leq q_r$ for all $r \in R$, where $\mu_r = \cup_{h \in H_r} \mu_h$. In other words, feasibility requires that the regional cap for every region is satisfied. This requirement distinguishes the current environment from the standard model without regional caps: We allow for (though do not require) $q_r < \sum_{h \in H_r} q_h$, that is, the regional cap can be smaller than the sum of hospital capacities in the region.

Since regional caps are a 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 some $i \in D \cup H$.

To accommodate the regional caps, we introduce new stability concepts that generalize the standard notion. For that purpose, we first define two basic concepts. A matching μ

is **individually rational** if (i) for each $d \in D$, $\mu_d \succeq_d \emptyset$, and (ii) for each $h \in H$, $d \succeq_h \emptyset$ for all $d \in \mu_h$, and $|\mu_h| \leq q_h$. That is, no agent is matched with an unacceptable partner and each hospital's capacity is respected.

Given matching μ, a pair (d, h) of a doctor and a hospital is called a **blocking pair** if $h \succ_d \mu_d$ and either (i) $|\mu_h| < q_h$ and $d \succ_h \emptyset$, or (ii) $d \succ_h d'$ for some $d' \in \mu_h$. In words, a blocking pair is a pair of a doctor and a hospital who want to be matched with each other (possibly rejecting their partners in the prescribed matching) rather than following the proposed matching.

When there are no binding regional caps (in the sense that $q_r \geq \sum_{h \in H_r} q_h$ for every $r \in R$), a matching is said to be stable if it is individually rational and there is no blocking pair. Gale and Shapley (1962) show that there exists a stable matching in that setting. In the presence of binding regional caps, however, there may be no such matching that is feasible (in the sense that all regional caps are respected). Thus in some cases every feasible and individually rational matching may admit a blocking pair.

Given this observation, we define a weaker stability concept, in which certain types of blocking pairs are admitted. More specifically, whenever there is a blocking pair, we require that it is "caused" by the existence of regional caps. Recall that $r(h)$ is the region that h belongs to.

Definition 1. A matching μ is **weakly stable** if it is feasible, individually rational, and if (d, h) is a blocking pair then (i) $|\mu_{r(h)}| = q_{r(h)}$ and (ii) $d' \succ_h d$ for all doctors $d' \in \mu_h$.

As seen in the definition, only certain blocking pairs are admitted. More specifically, if doctor d and hospital h constitute a blocking pair then (i) the cap of hospital h's region is filled with doctors, and (ii) h prefers every currently matched doctor to d. If (d, h) is a blocking pair, condition (ii) implies that hospital h has a vacant position and desires to fill it with doctor d. Condition (i) is then motivated by the idea that such a blocking may be problematic in relation to feasibility because the number of doctors in the region already equals its regional cap. In this sense, weak stability requires that any blocking pair is "caused" by regional caps. Indeed, this concept reduces to the standard stability concept of Gale and Shapley (1962) if there are no binding regional caps.

The implicit idea behind the definition is that the government or some authority can interfere and prohibit a blocking pair to be formed if regional caps are an issue. Indeed, in Japan, participants seem to be effectively forced to accept the matching announced by the clearinghouse because a severe punishment is imposed on deviators.[17] One might then

[17]For example, violating hospitals can be excluded from participating in the matching mechanism in subsequent years (Japan Residency Matching Program, 2010).

wonder "If the government has the power to prohibit a blocking pair in certain cases, why doesn't it have the power to do so in all cases, so why do we care about stability in the first place?"

Our view is that even if the clearinghouse has power to enforce a matching (which may be the case in the Japanese residency match), an assignment that completely ignores participants' preferences would be undesirable. Indeed, as we discussed in Section 2, the introduction of a stable matching mechanism in this market was motivated by the criticism that the previous assignment system was "unfair" and "inefficient," rather than by a desire to prevent participants from circumventing the assignment by forming "blocking pairs."[18] In other words, we view minimizing blocking pairs as a normative criterion.[19] Given this observation, our weak stability captures the idea that it is desirable to minimize blocking pairs so that the only blocking pairs are "caused" by regional caps, which may be a legitimate reason to deny a blocking pair.[20]

A potential drawback of weak stability is that it allows for the existence of a blocking pair (d, h) such that the regional cap of $r(h)$, h's region, is full even if d is currently assigned to a hospital in $r(h)$ (that is, $\mu_d \in H_{r(h)}$). In practice, however, such a blocking pair may be a legitimate deviation because the total number of doctors matched within the region does not increase, thus the regional cap continues to be 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 match is denoted by $\varphi_i(\succ)$ for each $i \in D \cup H$.

A mechanism φ is said to be **strategy-proof** if there does not exist a preference profile $\succ$, an agent $i \in D \cup H$, and preferences $\succ'_i$ of agent i such that

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

[18]Another example of a labor market using a stable mechanism despite being heavily regulated is the labor market for junior academic positions in France (Haeringer and Iehle, 2010).

[19]"No justified envy" in the school choice literature corresponds to "no blocking pair" in our context, and it is viewed as a normative criterion.

[20]Another obvious normative criterion is (constrained) efficiency. Indeed, it will turn out that weak stability implies efficiency (Theorem 1). Thus weak stability has an additional normative appeal.

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.[21]

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).[22] 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 there does not exist a preference profile $\succ$, a doctor $d \in D$, and preferences $\succ'_d$ of doctor d such that

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

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 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

[21]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).

[22]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.

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).

The algorithm terminates at a step in which no rejection occurs. The algorithm always terminates in a finite number of steps. Gale and Shapley (1962) show that the resulting matching is stable in the standard matching model without any binding regional 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).[23] This result has been extended by many subsequent studies, suggesting that the incentive compatibility of the mechanism is quite robust and general.[24]

[23]Ergin (2002) defines a stronger version of group strategy-proofness. It requires that no group of doctors 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.

[24]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, there is an exogenously given (government-imposed) **target capacity** $\bar{q}_h \le q_h$ for each hospital h such that $\sum_{h \in H_r} \bar{q}_h \le q_r$ for each region $r \in R$.[25,26] 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. Note, however, that target capacities are not feasibility constraints by themselves: the goal of Japanese policy makers is to satisfy regional caps and target capacities were introduced to achieve that goal.

Although the mechanism is a variant of the deferred acceptance algorithm, the mechanism suffers from at least two problems. The first problem relates to stability: Despite the government's 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, $d_1, \ldots, d_{10}$. Preference profile $\succ$ is as follows:

$$\succ_{h_i}: d_1, d_2, \ldots, d_{10}, \quad \text{for } i = 1, 2,$$
$$\succ_{d_j}: h_1 \text{ if } j \le 3 \quad \text{and} \quad \succ_{d_j}: h_2 \text{ if } j \ge 4.$$

[25]Note that we allow the sum of target capacities to be strictly smaller than the regional cap. This is necessary if the sum of hospital capacities is strictly smaller than the regional cap; we allow this possibility even otherwise. All results, including (counter)examples, hold when we assume that the sum of target capacities is equal to the regional cap.

[26]In our model, $\bar{q}_h$ 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. As explained in footnote 12, however, the capacity can be considered exogenous in the Japanese context.

That is, three doctors prefer hospital h_1 to being unmatched (the option $\emptyset$) to hospital h_2, while the other seven doctors prefer hospital h_2 to being unmatched to hospital h_1. Let the target capacities be $\bar{q}_{h_1} = \bar{q}_{h_2} = 5$.[27]

At the first round of the JRMP 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} find h_1 unacceptable, they do not make further applications, so the algorithm terminates at this point. Hence the resulting matching μ is such that

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

This is not weakly stable: For example, hospital h_2 and doctor d_9 constitute a blocking pair while the regional cap for r is not binding. One may wonder whether the failure of weak stability depends on the assumption that some agents find some of potential partners unacceptable. However, a similar example can be constructed even if we require every agent finds every potential partner acceptable.[28]

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. 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 the regional cap is still respected, μ' is feasible. Moreover, every agent is weakly better off with doctors d_9 and d_{10} being strictly better off than at μ. Hence we conclude that the JRMP mechanism results in an inefficient matching in this example.[29]

[27] The specification of target capacities follows the formula used in Japan that we mentioned earlier.

[28] For instance, modify the market in the example by introducing another hospital h_3 in another region with regional cap two; let h_3 find every doctor acceptable and have two positions; d_1, d_2 and d_3 prefer h_1 to h_3 to h_2 to being unmatched, while all other doctors prefer h_2 to h_3 to h_1 to being unmatched (thus every doctor finds all hospitals acceptable). The resulting matching violates weak stability.

[29] In this example, not all hospitals are acceptable to all doctors. One may wonder whether this is an unrealistic assumption because doctors may be so willing to work that any hospital is acceptable. However, the example can be easily modified so that all hospitals are acceptable to all doctors while some doctors are unacceptable to some hospitals (which may be a natural assumption because, for instance,

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 this mechanism, we define and discuss the goals that we try to achieve with the mechanism.

5. Goal Setting: Stability Concepts and Strategy-Proofness

As discussed earlier, the concept of weak stability introduced in the previous section may be too weak. This is because it does not regard certain blocking pairs as legitimate deviations even if they 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 (constrained) efficient.*

When there is no regional cap (in which case weak stability reduces to the standard concept of stability), a matching is stable if and only if it is in the core, and any core outcome is efficient. Without regional caps, Theorem 1 follows straightforwardly from these facts. With regional caps, however, there is no obvious way to define an appropriate cooperative game or a core concept. Theorem 1 states that efficiency of weakly stable matchings still holds in our model.[30]

Now we formalize stability concepts that are stronger than weak stability defined in Section 3. The first notion presented below is meant to capture the idea that any blocking

typically a hospital only lists doctors who they interviewed). Also, in many markets doctors apply to only find a small subset of hospitals. In 2009, for instance, a doctor applied to only 3.3 hospitals on average (Japan Residency Matching Program, 2009a).

[30]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.

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. Recall that $r(h)$ is the region that hospital h belongs to.

Definition 2. A matching μ is **strongly stable** if it is feasible, individually rational, and if (d, h) is a blocking pair then (i) $|\mu_{r(h)}| = q_{r(h)}$, (ii) $d' \succ_h d$ for all doctors $d' \in \mu_h$, and (iii) $\mu_d \notin H_{r(h)}$.

The difference from weak stability defined in Definition 1 is an added condition (iii), "$\mu_d \notin H_{r(h)}$." Thus, a blocking pair such that the doctor in the pair moves between two hospitals in the same region should not exist. This is because such a movement keeps the total number of doctors in a region unchanged. The only blocking pair that can remain under this definition would actually violate the regional cap since condition (i) implies that the region's cap is currently binding, condition (ii) implies that the only blocking involves filling a vacant position, and condition (iii) implies that the doctor is not currently assigned in the hospital's region.

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:

$$\begin{aligned} \succ_{h_i} &: d \ \text{ for } i = 1, 2, \\ \succ_d &: h_1, h_2. \end{aligned}$$

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 the match of d and h_1 comprises the first choices of each other. Matching μ' is not strongly stable because (d, h_1) is a blocking pair and $\mu'_d = h_2 \in H_{r(h_1)}$ so the regional cap would not be violated.

The above example shows that strong stability is a strictly stronger concept than weak stability. Nonetheless, we will not pursue strong stability 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 preferences:

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

Matching μ such that $\mu_{h_1} = \{d_1\}$ and $\mu_{h_2} = \emptyset$ is weakly stable since h_1 is matched to its first choice and the regional cap is binding. Similarly μ' such that $\mu'_{h_1} = \emptyset$ and $\mu'_{h_2} = \{d_2\}$ is also weakly stable. It is easy to see that these are the only weakly stable matchings. However, neither μ nor μ' is strongly stable. To see that μ is not strongly stable, note that a pair (d_1, h_2) constitutes a blocking pair and $\mu_{d_1} = h_1 \in H_{r(h_2)}$ so the regional cap would not be violated. Similarly μ' is not strongly stable. Therefore, a strongly stable matching does not exist in this market.

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 reported preferences $\succ'_{d_1}$ 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 preferences when her true preferences are $\succ_{d_1}$.

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 preferences are $\succ_{d_2}$. 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 examples show 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 *any* blocking pair is regarded as a legitimate deviation as long as it does not violate a regional cap. To define an appropriate stability concept, we need to further restrict blocking pairs that are regarded as legitimate. We do so by using the notion of target capacity. More specifically, we now regard target capacities $(\bar{q}_h)_{h\in H}$ as reflecting certain distributional goals (though not feasibility constraints) and define a stability concept that respects target capacities as much as possible.[31]

Definition 3. A matching μ is **stable** if it is feasible, individually rational, and if (d, h) is a blocking pair then (i) $|\mu_{r(h)}| = q_{r(h)}$, (ii) $d' \succ_h d$ for all doctors $d' \in \mu_h$, and

[31]Depending on the distributional goals, target capacities can be set differently from those specified in the description of the JRMP mechanism. Given the absence of further information, we use the ones given in the JRMP mechanism in this paper.

(iii') either $\mu_d \notin H_{r(h)}$ or $|\mu'_h| - \bar{q}_h > |\mu'_{\mu_d}| - \bar{q}_{\mu_d}$,

where μ' is the matching such that $\mu'_d = h$ and $\mu'_{d'} = \mu_{d'}$ for all $d' \neq d$.

This concept is stronger than weak stability while weaker than strong stability. Conditions (i) and (ii) in the definition of weak stability are also required in stability, so stability is stronger than weak stability. Meanwhile stability is different from strong stability in that condition (iii) in strong stability is replaced by a condition (iii') and, since there are more possible cases in (iii') than in (iii), stability is weaker than strong stability.[32]

The first part of condition (iii'), $\mu_d \notin H_{r(h)}$, is identical to condition (iii) and addresses the case in which the deviating doctor is currently assigned outside the region of the deviating hospital. The second part declares that certain types of blocking pairs within a region (note that $\mu_d \in H_{r(h)}$ holds in the remaining case) are not regarded as legitimate deviations (recall that our interpretation of stability concepts is normative). To see this point, consider the inequality in condition (iii'),

$$|\mu'_h| - \bar{q}_h > |\mu'_{\mu_d}| - \bar{q}_{\mu_d}. \tag{5.1}$$

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, realizing a new matching μ'. 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 in the blocking pair "equalizes" the excesses over the target capacities compared to the current matching (that is, $|\mu_h| - \bar{q}_h < |\mu'_h| - \bar{q}_h \leq |\mu'_{\mu_d}| - \bar{q}_{\mu_d} < |\mu_{\mu_d}| - \bar{q}_{\mu_d}$), then such a movement should be regarded as a legitimate deviation. Thus, the only blocking pair within a region that can remain under this definition should satisfy condition (5.1).

We note that there may be other natural definitions of stability. For example, it may be desirable to entitle a hospital with capacity 20 to twice as many doctors over the target as a hospital with capacity 10. There may also be other criteria that are deemed desirable. To address this issue, in Section 7.3 and Appendix B we consider a class of stability concepts that includes the stability in Definition 3 as a special case and accommodates the above ideas.[33] For each stability notion in this class, we present a mechanism that

[32]For an example in which the three stability concepts – weak stability, stability, and strong stability – lead to different choices of matchings, consider Example 4 with the additional specification of a target capacity profile $(\bar{q}_{h_1}, \bar{q}_{h_2}) = (1, 0)$.

[33]In Appendix D we consider a stability concept stronger than the stability concepts in this class (while weaker than strong stability) and show that this concept suffers from the same types of drawbacks (as in Examples 4 and 5) as those for strong stability.

generates a stable matching. In the main part of this paper, we assume that the policy goal is expressed as in condition (5.1). However, this particular choice of a policy goal is *not* a necessary requirement for our analysis to work, as we will observe in Section 7.3 and Appendix B. We chose this condition because it is expositionally simple and appears to be a reasonable starting point. 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.

A natural question is whether a stable matching exists in every market. This question will be answered in the affirmative in the next section, where we propose an algorithm that always generates a stable matching.

6. A New Mechanism: Flexible Deferred Acceptance

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 an order of hospitals in region r: Denote $H_r = \{h_1, h_2, \dots, h_{|H_r|}\}$ and order h_i earlier than h_j if $i < j$. Given this order, 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 but have not been rejected by h so far. For each hospital $h \in H_r$, choose the $\bar{q}_h$ best acceptable doctors according to $\succ_h$ from D'_h if they exist, and otherwise choose all acceptable doctors in D'_h. Formally, for each $h \in H_r$ choose D''_h such that $D''_h \subset D'_h$, $|D''_h| = \min\{\bar{q}_h, |D'_h|\}$, and $d \succ_h d'$ for any $d \in D''_h$ and $d' \in D'_h \setminus D''_h$.

(b) Start with a tentative match D''_h for each hospital $h \in H_r$. Hospitals take turns to choose (one doctor at a time) the best remaining doctor in their current applicant pool. Repeat the procedure (starting with h_1, proceeding to $h_2, h_3, \dots$ and going back to h_1 after the last hospital) until the regional quota q_r is filled or the capacity of the hospital is filled or no doctor remains to be matched. Formally, let $\iota_i = 0$ for all $i \in \{1, 2, \ldots, |H_r|\}$. Let $i = 1$.

(i) If either the number of doctors already chosen by the region r as a whole equals q_r, or $\iota_i = 1$, then reject the doctors who were not chosen throughout this step and go back to Step 2.

(ii) 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}.

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

We define the **flexible deferred acceptance mechanism** to be a mechanism that produces, for each input, the matching at the termination of the above algorithm.[34]

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 the former lets hospitals fill their capacities "more flexibly" than the latter. To see this point, first observe that the way that hospitals choose doctors who applied in Step 4a 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 tentatively keep doctors beyond target capacities even if regional caps are not binding. This is addressed in Step 4b. In this step, hospitals in a region are allowed to keep more doctors than their target capacities 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 an ordering of hospitals. There are at least two reasons that this requirement may not cause problems such as conflicts among hospitals to get a "desirable position" in the order. The first is that, as we will discuss in Subsection 7.4, the effect of different ways of setting orders on the welfare of hospitals is ambiguous. More specifically, it may be the case that a hospital is better off being ordered later under some specification of preference profiles, while the opposite may be true under other specifications. Second, the flexible deferred acceptance

[34] We show in Theorem 2 that the algorithm stops in a finite number of steps.

algorithm can be modified without losing its desirable properties, by adding "Step 0" in which a particular ordering is chosen according to some probabilistic rule. The aforementioned problems can be resolved by, for example, choosing an order according to the uniform probability distribution.

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 violates both efficiency and weak stability, let alone stability. The flexible deferred acceptance algorithm selects a matching that is efficient and stable. Precisely, let doctors apply to hospitals in the specified order. For doctors d_1 to d_8, the algorithm does not proceed to Step 4b, as the number of doctors in each hospital is no larger than its target. When d_9 applies, doctors $d_1, \ldots, d_8$ are still matched to hospitals in Step 4a, and d_9 is matched to h_2 in Step 4b. In the same way, when d_{10} applies, doctors $d_1, \ldots, d_8$ are still matched to hospitals in Step 4a, and d_9 and d_{10} are matched to h_2 in Step 4b. Hence an efficient and stable matching results. Intuitively, the algorithm treats doctors' applications in a more flexible manner than in the JRMP algorithm. 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 the stability of the flexible deferred acceptance mechanism, recall that there is a sense in which hospitals fill their capacities "flexibly." More specifically, at each step of the algorithm hospitals can tentatively accept doctors beyond their 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.[35]

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 because, even if she is rejected, she will be able to apply to her second choice, and so

[35]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.

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 the latter model can be invoked to show the theorem. In fact, the proof shows that a more general result (Theorem 4) holds which can be applicable to the class of stability concepts mentioned in Section 7.3 and that the current model is indeed a special case of the general model (Propositions 5 and 6). Theorem 2 then follows as a corollary of these results.

Theorems 1 and 2 imply an appealing welfare property of the flexible deferred acceptance mechanism.

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. □

Recall that the JRMP mechanism does not necessarily produce an efficient matching. In light of this observation, Corollary 1 implies that the flexible deferred acceptance mechanism improves upon the JRMP mechanism not only in terms of stability but also in terms of efficiency.

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

Proposition 1. *If the number of doctors matched with $h \in H$ in the flexible deferred acceptance mechanism is strictly less than its target capacity, then for any $d \in D$ who are not matched with h, either d is unacceptable to h or d prefers its current match to h.*

Proof. Assume that d prefers h to her outcome under the flexible deferred acceptance mechanism. Then d has applied to h and was rejected under the flexible deferred acceptance algorithm. If the number of doctors matched with h in the flexible deferred acceptance mechanism is strictly less than its target capacity, then the number of doctors who have ever applied to h and are acceptable to h is strictly smaller than the target capacity of h. This implies that any doctor who applied to h and was rejected in the flexible deferred acceptance algorithm is unacceptable to h. In particular d is unacceptable, completing the proof. □

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 cap is already binding, thus this property is not implied by the fact that the outcome of the flexible deferred acceptance algorithm is stable.

7. Discussion

This section provides several discussions. In Subsection 7.1, we consider the rural hospital theorem of Roth (1986) and a related concept of the "match rate," the ratio of the number of all matched doctors to the total number of doctors (matched plus unmatched). Subsection 7.2 studies the existence issue of a side-optimal stable matching, that is, a matching that is preferred by all doctors or by all hospitals. Subsection 7.3 generalizes stability and the flexible deferred acceptance mechanism. Subsection 7.4 examines the welfare effect of different choices of target capacities and picking orders over hospitals in the flexible deferred acceptance mechanism, and Subsection 7.5 considers "floor constraints" instead of "ceiling constraints" (regional caps).

7.1. **The Rural Hospital Theorem and The Match Rate.**

In this subsection, we show that the conclusion of the rural hospital theorem does not hold in our environment. Motivated by this finding, we study how the flexible deferred acceptance mechanism works in terms of the match rate, that is, the proportion of the number of all matched doctors to the total number of doctors (matched plus unmatched).

7.1.1. *The Rural Hospital Theorem.* The rural hospital theorem (Roth, 1986) 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 conclusion of 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 matching μ while it does not do so in matching μ'. Also, d_1 is matched to a hospital in matching μ while unmatched in matching μ'. Hence both conclusions of the rural hospital theorem fail, even with the notion of strong stability. Since strong stability implies stability and weak stability, this example also shows that the conclusions of the rural hospital theorem fail with those stability concepts (analogously, all negative conclusions of this subsection and the next hold under any of these stability concepts).

One might suspect that, although the rural hospital theorem does 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 as a "rural region" for the sake of discussion. 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.

7.1.2. *The Match Rate.* Related to the rural hospital theorem is the notion of "match rate," which is the ratio of the number of all matched doctors to the total number of doctors (matched plus unmatched). 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.[36] This is perhaps because the match rate is an easy measure for participants to understand.[37]

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 deferred acceptance mechanism 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 $(\bar{q}_{h_1}, \bar{q}_{h_2}, \bar{q}_{h_3}) = (1, 0, 1)$. Then, the flexible deferred acceptance mechanism always selects a matching μ defined in Example 8. But

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

[37] 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.

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:

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

with everything else unchanged (thus hospitals h_1 and h_2 are in one region and h_3 is in the other, each region has a regional cap of one, and each hospital has capacity of one). Let $(\bar{q}_{h_1}, \bar{q}_{h_2}, \bar{q}_{h_3}) = (1, 0, 1)$. 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. 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 $\mu'_{d_2} \succ_{d_2} \mu_{d_2}$, 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, there are bounds on the match rates in the matchings produced by the flexible deferred acceptance mechanism. More specifically, the following comparison can be made with the JRMP mechanism as well as with the (unconstrained) deferred acceptance algorithm without regional caps:

Theorem 3. *For any preference profile,*

(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, which is a direct corollary of part (1), 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.[38]

Theorem 3 suggests that the flexible deferred acceptance mechanism matches reasonably many doctors. Characterizing stable mechanisms that achieve strategy-proofness for doctors and match "as many doctors as possible," as well as studying their relationship with the flexible deferred acceptance mechanism, is an interesting open question.

7.2. **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 7, and suppose that $(\bar{q}_{h_1}, \bar{q}_{h_2}) = (1, 0)$. There are two stable matchings, μ and μ' specified in Example 7, where only d_1 and h_1 are matched at μ while only d_2 and h_2 are matched at μ'. 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.

7.3. **A General Framework.** As mentioned in Section 5, the notion of stability is based on the idea that if the result of a move of a doctor within a region does not equalize the excesses over the target capacities compared to the current matching, it is not deemed as a legitimate deviation. We argued that this is not the only reasonable definition 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.

[38]For an example in which the deferred acceptance mechanism and the flexible deferred acceptance mechanism differ in terms of match rates, see Example 4 (with an arbitrary target capacity profile). For the flexible deferred accceptance mechanism and the JRMP mechanism, see Example 1.

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 is how to decide the picking order. One natural conjecture may be that choosing earlier (that is, having an earlier order in the flexible deferred acceptance algorithm) benefits a hospital. As we have mentioned earlier, this would be a problematic property: If choosing earlier benefits a 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 orders result in different stable matchings, thus the choice of an order does matter for the algorithm's outcome.

Example 11 (Ordering a hospital earlier may make it worse off)**.** Let there be hospitals h_1, h_2 and h_3 in region r_1, and h_4 in region r_2. Suppose that $(q_{h_1}, q_{h_2}, q_{h_3}, q_{h_4}) = (2, 1, 1, 1)$ and $(\bar{q}_{h_1}, \bar{q}_{h_2}, \bar{q}_{h_3}, \bar{q}_{h_4}) = (1, 0, 1, 1)$. The regional cap of r_1 is 2 and that for r_2 is 1. Preferences are

$$\succ_{h_1}: d_1, d_4, d_2, \quad \succ_{h_2}: d_3, \quad \succ_{h_3}: \text{arbitrary}, \quad \succ_{h_4}: d_2, d_1,$$
$$\succ_{d_1}: h_4, h_1, \quad \succ_{d_2}: h_1, h_4, \quad \succ_{d_3}: h_2, \quad \succ_{d_4}: h_1.$$

(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_4, 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_4, 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_4, displacing d_1 from h_4. Then d_1 applies to h_1. d_1 is accepted, displacing d_4 from h_1. The matching finalizes.

First, notice that hospital h_2 is better off in case (2). Thus being ordered 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). Thus being ordered later helps h_1 if she prefers $\{d_1\}$ to $\{d_2, d_4\}$. Therefore, the effect of a picking order 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. The following example shows that target monotonicity is not necessarily true.

Example 12 (Target monotonicity may fail)**.** Consider a market that is identical to the one in Example 11, except that the target of h_1 is now decreased to 0, with the order such that h_1 chooses before h_2.[39] Then h_1 is matched to $\{d_1\}$ under the flexible deferred acceptance mechanism. Therefore, if h_1 prefers $\{d_1\}$ to $\{d_2, d_4\}$, then h_1 is made better off when its target capacity is smaller.

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 if 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 the existence of an individually rational matching that respects floor constraints is not guaranteed. For example, if no doctor finds any hospital in a certain region to be acceptable, then satisfying a positive floor constraint for the region results in an individually irrational matching (doctors matched with hospitals in the region would just reject taking the job). Second, even if an individually rational matching exists, it is not clear whether a stable matching exists. In fact, an appropriate definition of stability in the presence of floor constraints is unclear.[40]

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

[39] When the target capacity of h_1 is decreased, the sum of the target capacities becomes strictly smaller than the regional cap (note that such a situation is allowed in our model). If one wishes to keep the sum equal to the regional cap, the example can be modified by increasing the target capacity of h_3 by 1, and the conclusion of the example continues to hold.

[40] A similar point is made in the context of school choice by Ehlers (2010).

acceptance mechanism. 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 in hospitals than the current mechanism.

With regional caps there may not necessarily exist a unique "right" notion of stability, and hence there may not necessarily exist a unique choice of the mechanism. 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 merits of imposing regional caps itself (except for a certain welfare result in Theorem 3). We took this approach because our model does not explicitly include patients or ethical concerns of the general populace, which may be underlying arguments for increasing doctors in rural areas. Similarly, we did not analyze other policies such as subsidies to incentivize residents to work in rural areas.[41] 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.[42] That is, as regional caps seem to be a strong political reality, we believe that it is important to take them as given and 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, approximate incentive compatibility similar to these papers may hold.

[41]This is not because subsidies are not important. In fact, subsidies are used to attract residents to rural areas in many countries such as the United States and Japan. However, there are political pressures to restrict the use of subsidies in the Japanese medical market. Beginning in 2011, for instance, the government will reduce subsidies to residency programs that pay annual salaries of more than 7,200,000 yen (about 85,000 U.S. dollars) to residents. In any case, our analysis is applicable given participants' preferences which reflect subsidies, thus our method can be employed on top of subsidies.

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

Second, studying more general constraint structures may be interesting. For instance, one could consider a hierarchy of regional caps, say one cap for a prefecture and one for each district within the prefecture. Or society may desire to regulate the total number of doctors practicing in certain specialties as well as in a region. One conjecture is that our results generalize as long as the constraint structure forms a hierarchy as analyzed by Milgrom (2009) and Budish, Che, Kojima, and Milgrom (2010). This paper focused on the simple setting of (one layer of) regional caps because that is the existing structure in the motivating problem of Japanese residency matching, but a generalization may become practically important if more complex constraints become politically possible in the future.

Third, it would be desirable to test how good the flexible deferred acceptance mechanism is relative to the JRMP mechanism. In a new project joint with Jun Wako, we have started talking with the matching organizers. Although the basic structure of the problem is as this paper has analyzed, specific details might need to be taken into account if our new design is to be used in practice. Simulating the performance of our mechanism based on actual data would be interesting as well.

Finally, it would be nice to study markets that have similar structures to the one in this paper. Markets mentioned in the Introduction are natural candidates for such a study. We expect some general insights will 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.

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.

Abraham, D. J., R. Irving, and D. Manlove (2007): "Two algorithms for the student-project allocation problem," *Journal of Discrete Algorithms*, 5, 73–90.

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/.

Biró, P., T. Fleiner, R. W. Irving, and D. F. Manlove (2010): "The College Admissions problem with lower and common quotas," *Theoretical Computer Science*, 411, 3136–3153.

Budish, E. (2010): "The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes," *Harvard University, Unpublished mimeo.*

Budish, E., and E. Cantillon (2010): "The Multi-Unit Assignment Problem: Theory and Evidence from Course Allocation at Harvard," Discussion paper, mimeo.

Budish, E., Y.-K. Che, F. Kojima, and P. R. Milgrom (2010): "Designing Random Allocation Mechanisms: Theory and Applications," 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. (2010): "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.

Ergin, H., and T. Sönmez (2006): "Games of School Choice under the Boston Mechanism," *Journal of Public Economics*, 90, 215–237.

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.

Haeringer, G., and V. Iehle (2010): "Knocking out the supply and sorting in centralized job markets," mimeo.

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 (2009a): "Result of the Residency Match," http://www.jrmp.jp/toukei/2009/2009toukei-5.pdf.

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

——— (2010): "Rules for Matching Participation (residency hospitals and university hospitals)," http://www.jrmp.jp/kiyaku-byouin.htm.

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

Kojima, F. (2007): "Matching and price competition: comment," *The American Economic Review*, 97(3), 1027–1031.

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.

Konishi, H., and M. Sapozhnikov (2008): "Decentralized matching markets with endogenous salaries," *Games and Economic Behavior*, 64(1), 193–218.

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.

——— (2004): *Putting Auction Theory to Work*. Cambridge University Press, Cambridge.

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

Ministry of Health, Labour and Welfare (2005a): "Proposed survey on no-doctor districts (in Japanese)," http://www.mhlw.go.jp/shingi/2005/01/s0124-11c.html.

——— (2005b): "Report on no-doctor districts (in Japanese)," http://www.mhlw.go.jp/toukei/list/76-16.html.

——— (2009a): "On opinions expressed in the public comment procedure (in Japanese)," http://www.mhlw.go.jp/shingi/2009/04/dl/s0423-8a.pdf.

——— (2009b): "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): "Relationship between wages and presence of a match in medical fellowships," *JAMA*, 290(9), 1153.

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., T. Sönmez, and U. Ünver (2004): "Kidney Exchange," *Quarterly Journal of Economics*, 119, 457–488.

——— (2005): "Pairwise Kidney Exchange," *Journal of Economic Theory*, 125, 151–188.

——— (2007): "Efficient Kidney Exchange: Coincidence of Wants in Markets with Compatibility-Based Preferences," *American Economic Review*, 97, 828–851.

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 (2006): "School Choice with Aggregate Capacity Constraints," mimeo.

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

——— (2010): "Course bidding at business schools," *International Economic Review*, 51(1), 99–123.

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.

Westkamp, A. (2010): "An Analysis of the German University Admissions System," mimeo.

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 weakly 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 weakly 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 weakly stable matching, $h \succ_d \mu_d$ implies one of the following (cases (1) and (2) correspond to a situation in which (d, h) is not a blocking pair of μ. Case (3) covers, by the definition of weak stability, the case in which (d, h) blocks μ):

(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 weakly 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 weakly 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 weakly 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 weakly 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 weakly stable, $\mu_{d''} \succ_{d''} h' = \mu'_{d''}$, contradicting the assumption that μ' Pareto dominates μ. $\square$

Appendix B. A general model

Let $\succeq_r$ be a weak ordering over nonnegative-valued integer vectors $W_r := \{w = (w_h)_{h\in H_r} | w_h \in \mathbb{Z}_+\}$. That is, $\succeq_r$ is a binary relation that is complete and transitive (but not necessarily antisymmetric). We write $w \succ_r w'$ if and only if $w \succeq_r w'$ holds but $w' \succeq_r w$ does not. Vectors such as w and w' are interpreted to be supplies of acceptable doctors to the hospitals in region r, but they only specify how many acceptable doctors apply to each hospital and no information is given as to who these doctors are. Given $\succeq_r$, a function $\tilde{\mathrm{Ch}}_r : W_r \to W_r$ is an **associated quasi choice rule** if $\tilde{\mathrm{Ch}}_r(w) \in \arg\max_{\succeq_r}\{w' | w' \le w\}$ for any non-negative integer vector $w = (w_h)_{h\in H_r}$.[43] We require that the quasi choice rule $\tilde{\mathrm{Ch}}_r$ be **consistent**, that is, $\tilde{\mathrm{Ch}}_r(w) \le w' \le w \Rightarrow \tilde{\mathrm{Ch}}_r(w') = \tilde{\mathrm{Ch}}_r(w)$. This condition requires that, if $\tilde{\mathrm{Ch}}_r(w)$ is chosen at w and the supply decreases to $w' \le w$ but $\tilde{\mathrm{Ch}}_r(w)$ is still available under w', then the same choice $\tilde{\mathrm{Ch}}_r(w)$ should be made under w' as well. Note that there may be more than one quasi choice rule associated with a given weak ordering $\succeq_r$ because the set $\arg\max_{\succeq_r}\{w' | w' \le w\}$ may not be a singleton for some $\succeq_r$ and w. Note also that there always exists a consistent quasi choice rule.[44] We assume that the regional preferences $\succeq_r$ satisfy the following mild regularity conditions:

(1) $w' \succ_r w$ if $w_h > q_h \ge w'_h$ for some $h \in H_r$ and $w'_{h'} = w_{h'}$ for all $h' \neq h$.
    This property says that the region desires no hospital to be forced to be assigned more doctors than its real capacity. This condition implies that, for any w, the component $[\tilde{\mathrm{Ch}}_r(w)]_h$ of $\tilde{\mathrm{Ch}}_r(w)$ for h satisfies $[\tilde{\mathrm{Ch}}_r(w)]_h \le q_h$ for each $h \in H_r$, that is, the capacity constraint for each hospital is respected by the (quasi) choice of the region.

(2) $w' \succ_r w$ if $\sum_{h\in H_r} w_h > q_r \ge \sum_{h\in H_r} w'_h$.
    This property simply says that region r prefers the total number of doctors in the region to be at most its regional cap. This condition implies that $\sum_{h\in H_r}(\tilde{\mathrm{Ch}}_r(w))_h \le q_r$ for any w, that is, the regional cap is respected by the (quasi) choice of the region.

(3) If $w' \lneq w \le q_{H_r} := (q_h)_{h\in H_r}$ and $\sum_{h\in H_r} w_h \le q_r$, then $w \succ_r w'$.

[43] For any two vectors $w = (w_h)_{h\in H_r}$ and $w' = (w'_h)_{h\in H_r}$, we write $w \le w'$ if and only if $w_h \le w'_h$ for all $h \in H_r$. We write $w \lneq w'$ if and only if $w \le w'$ and $w_h < w'_h$ for at least one $h \in H_r$. For any $W'_r \subseteq W_r$, $\arg\max_{\succeq_r} W'_r$ is the set of vectors $w \in W'_r$ such that $w \succeq_r w'$ for all $w' \in W'_r$.

[44] To see this point consider preferences $\succeq'_r$ such that $w \succ'_r w'$ if $w \succ_r w'$ and $w = w'$ if $w \succeq'_r w'$ and $w' \succeq'_r w$. The quasi choice rule that chooses (the unique element of) $\arg\max_{\succeq'_r}\{w' | w' \le w\}$ for each w is clearly consistent with $\succeq_r$.

This condition formalizes the idea that region r prefers to fill as many positions in hospitals in the region as possible so long as doing so does not lead to a 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 w, if there exists h such that $[\tilde{\mathrm{Ch}}_r(w)]_h < \min\{q_h, w_h\}$, then $\sum_{h' \in H_r} [\tilde{\mathrm{Ch}}_r(w)]_{h'} = q_r$. This captures the idea that the social planner should not waste caps allocated to the region: If some doctor is rejected 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.

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

$$w \le w' \Rightarrow \tilde{\mathrm{Ch}}_r(w) \ge \tilde{\mathrm{Ch}}_r(w') \wedge w,$$

that is,

$$w \le w' \Rightarrow [\tilde{\mathrm{Ch}}_r(w)]_h \ge \min\{[\tilde{\mathrm{Ch}}_r(w')]_h, w_h\} \text{ for every } h \in H_r. \tag{B.1}$$

This condition says that, when the supply of doctors is increased, the number of accepted doctors at a hospital can increase only when the hospital has accepted all acceptable doctors under the original supply profile. Formally, condition (B.1) is equivalent to

$$w \le w' \text{ and } [\tilde{\mathrm{Ch}}_r(w)]_h < [\tilde{\mathrm{Ch}}_r(w')]_h \Rightarrow [\tilde{\mathrm{Ch}}_r(w)]_h = w_h. \tag{B.2}$$

To see that condition (B.1) implies condition (B.2), suppose that $w \le w'$ and $[\tilde{\mathrm{Ch}}_r(w)]_h < [\tilde{\mathrm{Ch}}_r(w')]_h$. These assumptions and condition (B.1) imply $[\tilde{\mathrm{Ch}}_r(w)]_h \ge w_h$. Since $[\tilde{\mathrm{Ch}}_r(w)]_h \le w_h$ holds by the definition of $\tilde{\mathrm{Ch}}_r$, this implies $[\tilde{\mathrm{Ch}}_r(w)]_h = w_h$. To see that condition (B.2) implies condition (B.1), suppose that $w \le w'$. If $[\tilde{\mathrm{Ch}}_r(w)]_h \ge [\tilde{\mathrm{Ch}}_r(w')]_h$, the conclusion of (B.1) is trivially satisfied. If $[\tilde{\mathrm{Ch}}_r(w)]_h < [\tilde{\mathrm{Ch}}_r(w')]_h$, then condition (B.2) implies $[\tilde{\mathrm{Ch}}_r(w)]_h = w_h$, thus the conclusion of (B.1) is satisfied.

This definition of substitutability is analogous to *persistence* by Alkan and Gale (2003), who define the condition on a choice function in a slightly different context. While our definition is similar to substitutability as defined in standard matching models (see Chapter 6 of Roth and Sotomayor (1990) for instance), 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.

Given $(\succeq_r)_{r \in R}$, stability is defined as follows.

Definition 4. A matching μ is **stable** if it is feasible, individually rational, and if (d, h) is a blocking pair then (i) $|\mu_{r(h)}| = q_{r(h)}$, (ii) $d' \succ_h d$ for all doctors $d' \in \mu_h$, and

(iii') either $\mu_d \notin H_{r(h)}$ or $w \succeq_{r(h)} w'$,

where $w_{h'} = |\mu_{h'}|$ for all $h' \in H_{r(h)}$ and $w'_h = w_h + 1$, $w'_{\mu_d} = w_{\mu_d} - 1$ and $w'_{h'} = w_{h'}$ for all other $h' \in H_{r(h)}$.

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

The (Generalized) 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 $w = (w_h)_{h \in H_r}$ by
    (a) $w_{\bar{h}}$ is the number of doctors currently held at $\bar{h}$ plus one, and
    (b) w_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 $(\tilde{\text{Ch}}_r(w))_h$ 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 **(generalized) flexible deferred acceptance mechanism** to be a mechanism that produces, for each input, the matching given 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, instead of 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, that is, a doctor can sign at most one contract. For each doctor d, her preferences $\succ_d$ over $(\{d\} \times H) \cup \{\emptyset\}$ are given as follows.[45] 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 preferences $\succeq_r$ and its 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

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

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

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

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

In other words, contracts are substitutes if adding a contract to the choice set never induces a region to choose 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 region.

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

Proposition 2. *Suppose that $\succeq_r$ is substitutable. Then choice rule $\mathrm{Ch}_r(\cdot)$ defined above satisfies the substitutes condition and the law of aggregate demand.*

Proof. Fix a region $r \in R$. Let $X' \subseteq X$ 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

[45]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.

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 $d \succ_h \emptyset$.

(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 $\min\{[\tilde{\text{Ch}}_r(w')]_{h'}, w_{h'}\} = [\tilde{\text{Ch}}_r(w')]_{h'}$. Since $w' \geq w$ and condition (B.1) holds, this implies that

$$[\tilde{\text{Ch}}_r(w)]_{h'} \geq [\tilde{\text{Ch}}_r(w')]_{h'}. \tag{B.3}$$

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.3), 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 (B.2) imply $[\tilde{\text{Ch}}_r(w)]_h = w_h$. 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 $X = D \times H$ is said to be **individually rational** if (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) for any $r \in R$, $\text{Ch}_r(X') = X' \cap (D \times H_r)$.

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

(1) it is individually rational, 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 $(d, h) \in \mathrm{Ch}_r(X' \cup \{(d, h)\})$, 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) is a **block** of X'.

Given any individually rational set of contracts 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 individually rational X'.

Proposition 3. *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. Suppose that (d, h) is a blocking pair of μ. Denoting $r := r(h)$, by the definition of stability, it suffices to show that the following conditions (B.4) and (B.5) hold if $\mu_d \notin H_r$, and (B.4), (B.5) and (B.6) hold if $\mu_d \in H_r$:

$$|\mu_{H_r}| = q_r, \tag{B.4}$$

$$d' \succ_h d \text{ for all } d' \in \mu_h, \tag{B.5}$$

$$w \succeq_r w', \tag{B.6}$$

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.4) and (B.5) 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 by contradiction that condition (B.5) is violated, that is, $d \succ_h d'$ for some $d' \in \mu_h$. First, by consistency of $\tilde{\mathrm{Ch}}_r$, we have $[\tilde{\mathrm{Ch}}_r(w'')]_h \geq [\tilde{\mathrm{Ch}}_r(w)]_h$.[46] That

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

is, 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 $(d',h) \in X'$ imply that $(d,h) \in \text{Ch}_r(X' \cup (d,h))$.[47] 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 contracts, contradicting the assumption that X' is a stable allocation.

(2) Assume by contradiction that condition (B.4) is violated, so that $|\mu_{H_r}| \neq q_r$. Then, since $|\mu_{H_r}| \leq q_r$ by the construction of μ and the assumption that X' is individually rational, 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, $|\mu_h| < q_h$ because (d,h) is a blocking pair by assumption and (B.5) holds, 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 proposition suppose that $\mu_d \in H_r$ and by contradiction that (B.6) fails, that is, $w' \succ_r w$. Then it should be the case that $[\tilde{\text{Ch}}_r(w'')]_h = w''_h = w_h + 1 = |\mu_h| + 1$.[48] Also we have $|\mu_h| < q_h$ and hence $|\mu_h| + 1 \leq q_h$ and $d \succ_h \emptyset$, so

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

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

[48]To show this claim, assume by 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 $\tilde{\text{Ch}}_r(w'') \in \arg\max_{\succeq_r} \{w''' | w''' \leq w''\}$.

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. □

A **doctor-optimal stable allocation** in the matching model with contracts is a stable allocation that every doctor weakly prefers to every other stable allocation (Hatfield and Milgrom, 2005) We will show that the flexible deferred acceptance mechanism is "isomorphic" to the doctor-optimal stable mechanism in the associated matching model with contracts.

Proposition 4. *Suppose that $\succeq_r$ is substitutable for every $r \in R$. Then the doctor-optimal stable allocation in the associated matching model with contracts, X', exists. In the original model, the flexible deferred acceptance mechanism produces matching $\mu(X')$ in a finite number of steps.*

Proof. First observe that the doctor-optimal stable allocation in matching with contracts can be found by the cumulative offer process in a finite number of steps (Hatfield and Milgrom, 2005; Hatfield and Kojima, 2009). Then, we observe that each step of the flexible deferred acceptance algorithm corresponds to a step of the cumulative offer process, that is, at each step, if d proposes to h in flexible deferred acceptance algorithm, 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 a step of the flexible deferred acceptance algorithm corresponds to the set of contracts held by the region at the corresponding step of the cumulative offer process. □

Theorem 4. *Suppose that $\succeq_r$ is substitutable for every $r \in R$. Then 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.*

Proof. Propositions 3 and 4 imply that the flexible deferred acceptance algorithm finds a stable matching in a finite number of steps. Also, Propositions 2 and 4 imply that the flexible deferred acceptance mechanism is (group) strategy-proof for doctors, 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.** In this section we are going to establish Theorem 2 in the main text by showing that the stability concept in the main text can be rewritten by using a substitutable regional preferences.

Fix a region r. Given the target capacity profile $(\bar{q}_h)_{h\in H_r}$ and the vector $w \in W_r$, define the **ordered excess weight vector** $\eta(w) = (\eta_1(w), ..., \eta_{|H_r|}(w))$ 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 target capacities). 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 preferences $\succeq_r$ that compare the excess weights lexicographically. More specifically, let $\succeq_r$ be such that $w \succ_r w'$ if and only if there exists an index $i \in \{1, 2, \ldots, |H_r|\}$ such that $\eta_j(w) = \eta_j(w')$ for all $j < i$ and $\eta_i(w) > \eta_i(w')$. The associated weak regional preferences $\succeq_r$ are defined by $w \succeq_r w'$ if and only if $w \succ_r w'$ or $\eta(w) = \eta(w')$. We call such regional preferences **Rawlsian**.

Proposition 5. *Stability defined in the main text (Definition 3) is a special case of the general concept of stability in the Appendix (Definition 4) such that the regional preferences of each region are Rawlsian.*

Proof. Let μ be a matching and 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 $h' \in H_r \setminus \{h, \mu_d\}$. 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 $w_h+1-\bar{q}_h > w_{\mu_d}-1-\bar{q}_{\mu_d}$, which is equivalent to either $w_h - \bar{q}_h = w_{\mu_d} - 1 - \bar{q}_{\mu_d}$ or $w_h - \bar{q}_h \geq w_{\mu_d} - \bar{q}_{\mu_d}$. In the former case, obviously $\eta(w) = \eta(w')$, so $w \succeq_r w'$. In the latter case, $\{h' | w'_{h'} - \bar{q}_{h'} < |\mu_{\mu_d}| - \bar{q}_{\mu_d}\} = \{h' | w_{h'} - \bar{q}_{h'} < |\mu_{\mu_d}| - \bar{q}_{\mu_d}\} \cup \{\mu_d\}$, and $w_{h'} = w'_{h'}$ for all $h' \in \{h' | w_{h'} - \bar{q}_{h'} < |\mu_{\mu_d}| - \bar{q}_{\mu_d}\}$. Thus we obtain $w \succ_r w'$.

If $|\mu_h|+1-\bar{q}_h \leq |\mu_{\mu_d}|-1-\bar{q}_{\mu_d}$, then obviously $w' \succ_r w$. This completes the proof. □

Consider the (generalized) flexible deferred acceptance algorithm in a previous subsection. With the following quasi choice rule, this algorithm is equivalent to the flexible deferred acceptance algorithm in the main text: For each $w' \in W_r$,

$$\tilde{\text{Ch}}_r(w') = \max_{\substack{w = w^k \text{ for some } k \\ \sum_{h\in H_r} w_h \leq q_r}} w, \tag{B.7}$$

where $w^0 = (\min\{w'_h, \bar{q}_h\})_{h\in H_r}$ and $w^k \in W_r$ $(k = 1, 2, \ldots)$ is defined by

$$w^k_{h_j} = \min\{w'_{h_j}, q_{h_j}, w^{k-1}_{h_j} + \mathbb{I}_{j \equiv k \ (\text{mod } |H_r|)}\} \quad \text{for each } j = 1, 2, \ldots, |H_r|.$$

Proposition 6. *Rawlsian preferences are substitutable with the associated quasi choice rule (B.7) that corresponds to the flexible deferred acceptance algorithm in the main text.*

Proof. It is clear that the quasi choice rule $\tilde{\text{Ch}}_r$ defined in (B.7) satisfies the condition (B.1) for substitutability (as well as consistency and acceptance). Thus in the following, we will show that $\tilde{\mathrm{Ch}}_r$ indeed satisfies $\tilde{\mathrm{Ch}}_r(w) \in \arg\max_{\succeq_r}\{x|x \leq w\}$ for each w. Let $w' = \tilde{\mathrm{Ch}}_r(w)$. Assume by contradiction that $w' \notin \arg\max_{\succeq_r}\{x|x \leq w\}$ and consider an arbitrary $w'' \in \arg\max_{\succeq_r}\{x|x \leq w\}$. Then we have $w'' \succ_r w'$, so there exists i such that $\eta_j(w'') = \eta_j(w')$ for every $j < i$ and $\eta_i(w'') > \eta_i(w')$. Consider the following cases.

(1) Suppose $\sum_j \eta_j(w'') > \sum_j \eta_j(w')$. First note that $\sum_j \eta_j(w'') + \sum_h \bar{q}_h = \sum_h w''_h \leq q_r$ because $w'' \in \arg\max_{\succeq_r}\{x|x \leq w\}$. Thus $\sum_h w'_h = \sum_j \eta_j(w') + \sum_h \bar{q}_h < \sum_j \eta_j(w'') + \sum_h \bar{q}_h \leq q_r$. Moreover, the assumption implies that there exists a hospital h such that $w'_h < w''_h \leq \min\{q_h, w_h\}$. These properties contradict the construction of $\tilde{\mathrm{Ch}}_r$.

(2) Suppose $\sum_j \eta_j(w'') < \sum_j \eta_j(w')$. First note that $\sum_j \eta_j(w') + \sum_h \bar{q}_h = \sum_h w'_h \leq q_r$ by construction of $\tilde{\mathrm{Ch}}_r$. Thus $\sum_h w''_h = \sum_j \eta_j(w'') + \sum_h \bar{q}_h < \sum_j \eta_j(w') + \sum_h \bar{q}_h \leq q_r$. Moreover, the assumption implies that there exists a hospital h such that $w''_h < w'_h \leq \min\{q_h, w_h\}$. Then, w''' defined by $w'''_h = w''_h + 1$ and $w'''_{h'} = w''_{h'}$ for all $h' \neq h$ satisfies $w''' \leq w$ and $w''' \succ_r w''$, contradicting the assumption that $w'' \in \arg\max_{\succeq_r}\{x|x \leq w\}$.

(3) Suppose that $\sum_j \eta_j(w'') = \sum_j \eta_j(w')$. Then there exists some k such that $\eta_k(w'') < \eta_k(w')$. Let $l = \min\{k|\eta_k(w'') < \eta_k(w')\}$ 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 $\tilde{\mathrm{Ch}}_r$, this inequality holds only if $w'_h = \min\{q_h, w_h\}$, where h is an arbitrarily chosen hospital such that $w'_h - \bar{q}_h = \eta_i(w')$. Now it should be the case that $w''_h = \min\{q_h, w_h\}$ as well, because otherwise $w'' \notin \arg\max_{\succeq_r}\{x|x \leq w\}$.[49] Thus $w'_h = w''_h$. 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, we obtain $w'_h = w''_h$ for all h, that is, $w' = w''$. This is a contradiction to the assumption that $w' \notin \arg\max_{\succeq_r}\{x|x \leq w\}$ and $w'' \in \arg\max_{\succeq_r}\{x|x \leq w\}$.

The above cases complete the proof. □

Theorem 4 and Propositions 5 and 6 imply Theorem 2 in the main text.

[49]The proof that $w'' \notin \arg\max_{\succeq_r}\{x|x \leq w\}$ if $w''_h < \min\{q_h, w_h\}$ is as follows. Suppose that $w''_h < \min\{q_h, w_h\}$. Consider w''' defined by $w'''_h = w''_h + 1$, $w'''_{h'} = w''_{h'} - 1$ for some h' such that $w''_{h'} - \bar{q}_{h'} = \eta_i(w'')$, and $w'''_{h''} = w''_{h''}$ for all $h'' \in H_r \setminus \{h, h'\}$. Then we have $w'''_h - \bar{q}_h = w''_h - \bar{q}_h + 1 \leq w'_h - \bar{q}_h < w''_{h'} - \bar{q}_{h'}$, where the weak inequality follows because $w''_h < \min\{q_h, w_h\} = w'_h$. The strict inequality implies that $w'_h - \bar{q}_h \leq w''_{h'} - 1 - \bar{q}_{h'} = w'''_{h'} - \bar{q}_{h'}$. Hence $w'''_h - \bar{q}_h \leq w'''_{h'} - \bar{q}_{h'}$, which implies $w''' \succ_r w''$.

Appendix C. Proof of Theorem 3

Proof. **Part (1)** First note that the description of the deferred acceptance algorithm in the main text can be modified so that at each step t, each hospital chooses from all applications that have been made to it so far. We consider this (equivalent) version of the deferred acceptance algorithm in this proof.

Let μ and μ' be the matchings produced by the deferred acceptance mechanism and by the flexible deferred acceptance mechanism, respectively. Let $C_D(t)$ be the set of applications (pairs of a doctor and a hospital) that have been made up to and including step t of the deferred acceptance algorithm, and $C_F(t)$ be the corresponding set for the flexible deferred acceptance algorithm. Let T_D and T_F be the termination steps for the deferred acceptance algorithm and for the flexible deferred acceptance algorithm, respectively.

We first show that $C_D(T_D) \subseteq C_F(T_F)$. To see this, suppose the contrary, i.e., that $C_D(T_D) \not\subseteq C_F(T_F)$. Then there exists step t' such that $C_D(t) \subseteq C_F(T_F)$ for all $t < t'$ and $C_D(t') \not\subseteq C_F(T_F)$ holds. That is, t' is the first step such that an application not made in the flexible deferred acceptance algorithm is made in the deferred acceptance algorithm. Let h be the hospital that d applies to in this step. Notice that $h \succeq_d \mu_d$ and $\mu'_d \succ_d h$, hence it must be the case that $\mu'_d \succ_d \mu_d$. This implies that $\mu'_d \neq \emptyset$ and that d is rejected by μ'_d in some steps of the deferred acceptance algorithm. Let the first of such steps be t''. Since in the deferred acceptance algorithm doctors apply to hospitals in order of their preferences, $\mu'_d \succ_d \mu_d$ implies that $t'' < t'$, which in turn implies $C_D(t'') \subseteq C_F(T_F)$ by the definition of t'.

Now, we argue that the set of doctors accepted by μ'_d at step t'' of the deferred acceptance algorithm is a superset of the set of doctors accepted by μ'_d from the application pool $C_D(t'')$ (which is a subset of $C_F(T_F)$) at step T_F of the flexible deferred acceptance algorithm. To see this, note that if the same application pool $C_F(T_F)$ is given, the set of doctors accepted by μ'_d in the deferred acceptance algorithm is weakly larger than that of the flexible deferred acceptance algorithm, by the construction of these algorithms. Since in the deferred acceptance algorithm μ'_d accepts applications in order of its preferences, subtracting applications in $C_F(T_F) \setminus C_D(t'')$ does not shrink the set of doctors accepted by μ'_d within $C_D(t'')$ at step t'' of the deferred acceptance, which establishes our claim.

However, this contradicts our earlier conclusion that d is rejected by μ'_d at step t'' of the deferred acceptance algorithm while she is matched with μ'_d in the flexible deferred acceptance algorithm. Hence we conclude that $C_D(T_D) \subseteq C_F(T_F)$.

Now, since in the deferred acceptance algorithm each doctor d applies to hospitals in order of her preferences, μ_d is $\emptyset$ or the worst hospital for d in the set of hospitals associated with d in $C_D(T_D)$. Similarly, for each doctor d, μ'_d is the worst hospital for d in the set of hospitals associated with d in $C_F(T_F)$. If $\mu_d \neq \emptyset$, this and $C_D(T_D) \subseteq C_F(T_F)$ implies that $\mu_d \succeq_d \mu'_d$. If $\mu_d = \emptyset$, d has applied to all acceptable hospitals in the deferred acceptance algorithm. Thus $C_D(T_D) \subseteq C_F(T_F)$ implies that she has applied to all acceptable hospitals in the flexible deferred acceptance algorithm, too. Let h' be the worst acceptable hospital for d. Again, $C_D(T_D) \subseteq C_F(T_F)$ implies that all applications associated with h' in $C_D(T_D)$ is in $C_F(T_F)$. In particular, d's application to h' is in $C_F(T_F)$. Since in the deferred acceptance algorithm h' accepts applications in order of its preferences, subtracting applications in $C_F(T_F) \setminus C_D(T_D)$ does not shrink the set of doctors accepted by h' within $C_D(T_D)$ at step T_D of the deferred acceptance, d not being accepted by h' from $C_D(T_D)$ at step T_D of the deferred acceptance algorithm implies that she is not accepted by h' from $C_F(T_F)$ in step T_F of the flexible deferred acceptance algorithm either. But since we have shown that d's application to h' is in $C_F(T_F)$, this implies that in the flexible deferred acceptance algorithm d is rejected by h'. Because h' is the worst acceptable hospital for d and d's applications are made in order of her preferences, we conclude that $\mu'_d = \emptyset$, thus in particular $\mu_d \succeq_d \mu'_d$.

This shows that 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.

Our claim on the comparison between the flexible deferred acceptance mechanism and the JRMP mechanism can be proven in an analogous manner.

Part (2) The second part of the theorem's statement is an immediate corollary of the first. $\square$

Appendix D. Semi-strong stability

In the main text, we pointed out that a strongly stable matching may not exist. Then we weakened the requirement and introduced the stability concept. A natural question is whether a concept stronger than stability can be imposed. To investigate this issue, we define the following notion.

Definition 8. A matching μ is **semi-strongly stable** if it is feasible, individually rational, and if (d,h) is a blocking pair then (i) $|\mu_{r(h)}| = q_{r(h)}$, (ii) $d' \succ_h d$ for all doctors $d' \in \mu_h$, and (iii") either $\mu_d \notin H_{r(h)}$ or $|\mu_h| - \bar{q}_h \geq 0 \geq |\mu_{\mu_d}| - \bar{q}_{\mu_d}$.

The second part of condition (iii") says that a blocking pair (d,h) is not deemed as a legitimate deviation if doctor d is currently assigned in the region $r(h)$, the number of doctors matched with hospital μ_d is no more than its target, and that of hospital h is no less than its target. That is, a blocking pair that moves the distribution of doctors unambiguously away from the target capacity is not deemed to be a legitimate deviation. Note that some blocking pairs that are regarded as illegitimate deviations under stability are considered legitimate under this concept. For example, if hospital h_1 has the target capacity of 1 and $|\mu_{h_1}| = 10$, hospital h_2 has the target capacity of 5 and $|\mu_{h_2}| = 7$, and these two hospitals are in the same region, then a movement of a doctor from h_2 to a vacant position of h_1 is considered a legitimate deviation in semi-strong stability but not in stability.

Although semi-strong stability may seem to be an appropriate weakening of strong stability, unfortunately it has the same deficiency as strong stability: a semi-strongly stable matching does not necessary exist, and there exists no mechanism that is strategy-proof for doctors and selects a semi-strongly stable matching whenever there exists one.

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

Example 13 (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. The target capacities of hospitals are $(\bar{q}_{h_1}, \bar{q}_{h_2}, \bar{q}_{h_3}) = (0, 0, 1)$. We assume the following preference:

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

Matching μ such that $\mu_{h_1} = \{d_1\}$ and $\mu_{h_2} = \mu_{h_3} = \emptyset$ is stable. Similarly μ' such that $\mu'_{h_1} = \mu_{h_3} = \emptyset$ and $\mu'_{h_2} = \{d_2\}$ is also stable. It is easy to see that these are the only stable matchings. However, neither μ nor μ' is semi-strongly stable. To see that μ is not semi-strongly stable, note that a pair (d_1, h_2) constitutes a blocking pair and $\mu_{d_1} = h_1 \in H_{r(h_2)}$, and $|\mu_{h_1}| > \bar{q}_{h_1}$. Similarly μ' is not semi-strongly stable. Therefore, a semi-strongly stable matching does not exist in this market.

Note that Example 13 is similar to Example 4. In an analogous manner, we can easily modify Example 5 to construct an example in which there is no mechanism that is strategy-proof for doctors and finds a semi-strongly stable matching whenever there exists one.