A conditional match rate anomaly and an order change pressure in residency matching programs*

Munetomo Ando† Nihon University
Minoru Kitahara‡ Osaka Metropolitan University

April 28, 2022

Abstract

In the medical residency matching markets in the U.S. and Japan, we observe that an applicant's probability of being matched to the first listed program is disproportionately higher than that to the second listed conditional on that the applicant is rejected by the first, while the conditional probabilities are relatively stable after that. In addition, several experts have pointed out that participating programs sometimes exert pressure on applicants to influence the creation of their lists.

In this paper, we first show that the latter can explain the former, considering a specialty of first listing. Then, we use the data to find both substantial pressure and welfare loss, assuming a simple reduced-form acceptance and forcing process. In addition, we consider a probabilistic permutation of the submitted list to prevent pressure-induced list reordering and find in the data that the benefits of the intervention outweigh the associated efficiency losses.

JEL Classification numbers: C78, D47

Keywords: Random permutation, ranking change pressure, residency matching.

1 Introduction

In many countries, when a medical student passes the Medical Licensing Examination and becomes a doctor, he or she needs to attend a residency program. A centralized matching method is currently used to determine the training program for medical students.[1] In the U.S., the National Resident Marching Program (NRMP) is the central clearinghouse, and in Japan, Japan Residency Matching Program (JRMP) plays the role.

Each applicant decides his or her preference for the residency programs for which he or she has completed the necessary procedures and then submits the rank order list to the clearinghouse.[2] Each residency program also submits its preference list and the maximum number of residents it will accept. After that, the clearinghouse determines the matching outcome according to an algorithm and reports the results for the participants.

The clearinghouses in the U.S. and Japan use a modified version of the deferred acceptance algorithm (DAA) as the matching procedure. The original DAA provides a stable matching, and the students-proposing DAA is strategy-proof for students in the sense that truth-telling is a weakly dominant strategy. In contrast, residency programs generally do not satisfy strategy-proofness, but truthful reporting has been shown to be approximately optimal in sufficiently large markets (Kojima and Pathak 2009).

*This work was supported by JSPS KAKENHI Grant Numbers JP17K03776, JP18K01513 and JP19K01542.

†E-mail: ando.munetomo@nihon-u.ac.jp.

‡E-mail: kitahara@econ.osaka-cu.ac.jp.

[1] Roth (2003) provides a brief overview of the history of residency matching.

[2] The average number of programs on an applicant's list is 9.83 in the U.S. and 3.25 in Japan. The former is a weighted average of 11.22 (matched) and 4.21 (unmatched). Note that these figures are for 2019 (see https://www.nrmp.org/wp-content/uploads/2021/08/Impact-of-Length-of-ROL-on-Match-Results-2021.pdf and https://www.jrmp2.jp/toukei/2019/2019toukei-5.pdf).

Then, do applicants actually make the list following to their own preferences? Rees-Jones (2018), based on the data from 579 respondents from 23 U.S. medical schools in 2012, reported that 83% of respondents made their lists according to their preferences, but 16% submitted something different from their preferences in a way that was not just a mistake.[3]

So what could be the reason for the difference between the applicant's preference and the submitted list? One possible explanation is that the residency program is exerting pressure on the applicant to change the list. Of course, the NRMP has a rule against putting pressure on applicants. The match communication code of conduct requires the director of the program to commit to the five clauses, including the following:

- **Respecting an applicant's right to privacy and confidentiality**
  Program directors and other interviewers may freely express their interest in a candidate, but they shall not ask an applicant to disclose the names, specialties, geographic location, or other identifying information about programs to which the applicant has or may apply.

- **Discouraging unnecessary post-interview communication**
  Program directors shall not solicit or require post-interview communication from applicants, nor shall program directors engage in post-interview communication that is disingenuous for the purpose of influencing applicants' ranking preferences.[4]

However, some programs do not comply with this code of conduct, and applicants are pressured to change their rank order lists. For example, Yarris et al. (2010) reported post-interview communication in emergency medicine in the U.S. based on a questionnaire survey. They found that 89% of respondents were contacted by the programs after the interview but before the list was submitted. This practice may violate the rules that prohibit unnecessary communication after the interview.[5] Moreover, Grimm et al. (2016) reported that some residents changed their rankings on the submission list because they felt pressured in post-interview communication.

In Japan, due to the nature of the matching algorithm, residency programs should not ask applicants about their order of preference, although the JRMP does not have such an explicit prohibition. In practice, however, a survey of students who participated in the 2018 residency match included the question "Have you ever been asked about your preference ranking during an interview?" To this, 62% of respondents answered "yes."[6]

So how exactly do residency programs put pressure on applicants? These are guesses, but for example, a program might ask an applicant, "We want you to put our program first on your list," or ask, "Is our program your first choice?"

In fact, if the program were to put pressure on applicants, asking them to list the program in the first place would be efficient. If the applicant actually lists the program as the first place, and the program lists the applicant at the top of the list, then the match between the two is guaranteed.[7] Another advantage is that the promise to write the program first on the list is easy to keep. This is because if an applicant promises to write the program in the first place, but the matching does not succeed, it will surely be discovered that the applicant has broken the promise.[8]

We want to quantify how much of the change in the ranking of the list due to pressure exists. To do this, we first identify whether applicants have a higher probability of matching with the program listed first on the list than with any other program.

Figures 1 and 2 show a conditional match rate anomaly. Figures 1 and 2 tell us that an applicant's match probability with the program listed first on the list is consistently high. In the U.S., an applicant has approximately a 50% probability of being matched with the first program on the list. In Japan, the probability is

[3]See Rees-Jones and Skowronek (2018) for an experiment in preference misrepresentation.

[4]See https://www.nrmp.org/communication-code-of-conduct/

[5]See Sbicca et al. (2012) for a detailed explanation of the code violations.

[6]See https://jrmp2.s3-ap-northeast-1.amazonaws.com/question/2018gakusei.pdf

[7]Grimm et al. (2016) reported that "[i]n our survey, only 5.2% (14 of 268) of program directors reported that they always or usually move applicants up their rank order lists after the applicant promises to rank their program No. 1."

[8]Interestingly, in Grimm et al. (2016)'s survey, "52.6% (141 of 268) of program directors reported that at least once a year 1 or more applicants falsely claim they are ranking their program No. 1."

approximately 75%.[9] In contrast, the conditional probability that an applicant will be matched with the second through fourth programs on the list is relatively low and stable. These probabilities are conditional on being rejected by all programs listed higher.[10] This data is consistent with the story that some programs are pressuring applicants to change their lists.

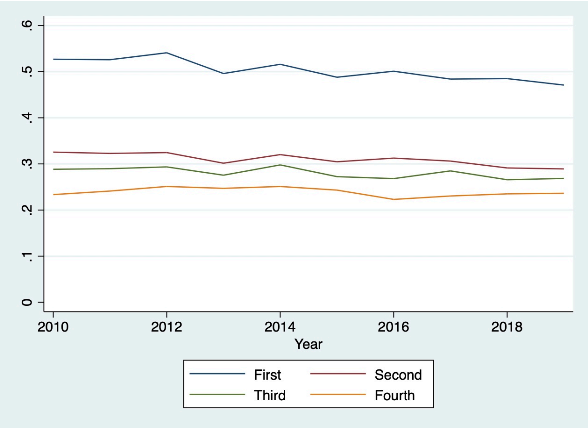


Figure 1: The U.S.

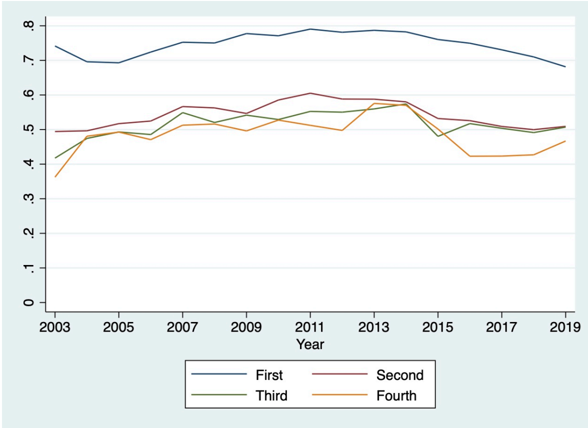


Figure 2: Japan

If the programs pressure applicants and the applicants actually change their lists, does this result in a loss of efficiency? The match may change if an applicant submits a list that differs from the true preference order due to pressure. For example, if an applicant had submitted a list based on the true preference order, he or she would have been matched with the second program, but the actual match would have been the fourth program. We refer to this as the number of the rank losses caused by the forced list change to measure the welfare loss, which is two.

Of course, a high probability of a forced list change does not necessarily imply a large welfare loss. In particular, if all acceptant programs put pressure to list themselves first, then, as long as the applicant chooses the best one among them as the first, the match outcome does not change (Remark 1). Therefore, it is necessary to empirically show what percentage of applicants are actually under pressure and how much welfare loss occurs. Then, the U.S. and Japanese data imply substantial forces and welfare loss by assuming a simple reduced-form acceptance and forcing process (Section 3).[11]

How can such losses be reduced? In the first place, an applicant changes his or her list according to pressure

[9] Figures 1 and 2 show the probabilities in the past 10 years through 2019. In these figures, we have not used data beyond 2020 because COVID-19 may have changed the way programs interview applicants in the residency matching. Since the current matching system in Japan started in 2003, Figure 2 uses data from 2003 (rather than 2010).

[10] Both in Japan and the U.S., how many applicants are matched to their which rank programs are available. Moreover, in Japan, how many applicants are unmatched with which length of their lists are also available. Then, we can calculate the acceptance rate of a program with a list rank conditional on that the applicant has been rejected by all higher rank programs as

$$\text{the } k\text{th rate} = \frac{\#\text{matched to the } k\text{th}}{\#\text{applicants} - \sum_{k'<k}(\#\text{unmatch with the list length of } k' + \#\text{matched to the } k'\text{th})}$$

for Japan, and approximately for the U.S. as

$$\text{the } k\text{th rate} \approx \frac{\#\text{matched to the } k\text{th}}{\#\text{applicants} - \sum_{k'<k}\#\text{matched to the } k'\text{th}},$$

as long as the number of unmatch applicants with their list lengths shorter than k is so small. Note that for the match probability with the second and subsequent programs on the list, it is necessary to take into account that some applicants may not have enough programs on the list to calculate the probability correctly. Therefore, when we derived the probabilities for Japan, we properly dealt with this issue in our calculations. However, we did not take into account it when calculating the probabilities for the U.S. due to data limitations.

[11] For simplicity, we assume the following reduced-form process. The applicants are ex-ante homogeneous. The programs simultaneously decide whether the applicants are acceptable with an exogenous acceptance probability, a. We also assume that each acceptant program exerts pressure on the applicant with an exogenous probability, e. The resident who receives effective pressure from at least one program will choose the best program and place it first on the list.

because if he or she does not keep his or her promise and is not matched with the program, he or she will be identified as a betrayer. And suppose it becomes clear that he or she has broken a promise. In that case, the applicant may be disadvantaged as an untrustworthy person in his or her subsequent activities in the medical community.

A possible solution is an introducing a random permutation of the applicants' submitted lists before running the matching algorithm. Specifically, we consider a random exchange of the first and the second with a certain probability, ε (Section 2.1.5). Under such interventions, even if the pressured program could not match the applicant, it would not be possible to judge with 100% certainty whether the applicant had betrayed the program or not. Moreover, with a sufficiently large exchange probability, the certainty of the applicant's betrayal falls below a required level (Remark 2). Consequently, the applicant can submit the list based on the true preferences order on the programs, and the welfare loss caused by the pressure disappears.

On the other hand, the random exchange itself causes an associated welfare loss, which makes determining whether this intervention is profitable or not an empirical issue. Again with our simplified model, we show that the benefits of this proposed intervention outweigh the associated losses with data from the U.S. and Japan. In fact, in Japan, we show that expected rank loss can be reduced to about 1/4 when the significance level is set to 15% (again in Section 3).

The rest of the paper is organized as follows: Section 2 describes the model. Section 3 describes the results of empirical evaluations. Section 4 concludes.

2 Model

We consider a representative doctor to be matched with at most one residency program in I. Here $I = \{1,\dots,L\}$ with $L \geq 2$ is the set of programs that the doctor has applied to and found to be acceptable, where program k is the kth most preferred one for the doctor. Thus, we implicitly assume that there is no tie in ranking.

2.1 Process

2.1.1 Matching

Given the doctor's list of programs $\boldsymbol{l} = (l_1,\dots,l_L) \in \{(l'_1,\dots,l'_L) \in I^L | l_k \neq l_{k'} \text{ for } k' \neq k\}$, where program l_k is the program listed kth in the list, and the set of programs $A \subset I$ which would accept the doctor, the doctor is matched with program $\mu(\boldsymbol{l}, A) = l_k$ if (i) $l_k \in A$ and (ii) $l_{k'} \notin A$ for all $k' < k$, and is unmatched if $A = \emptyset$.

Note that the elements of set A generally depend on other doctors' lists. However, we ignore the possibility in our model. In this sense, our model is just a simplified reduced-form one.

We also note that the doctor's list $\boldsymbol{l}$ is not necessarily equal to his or her preferences, $I = \{1,\dots,L\}$. This is because, as we will see below, the doctor may change the submitting list if he or she feels pressured by the programs. Moreover, as we will study later, the list $\boldsymbol{l}$ may further be different from the submitted list, because interventions implemented by the clearinghouse may result in a random permutation of the list.[12]

2.1.2 Order change pressure

Given the set of programs which the doctor is under (effective) pressure to list first, $F \subset \{1,\dots,L\}$, we let the doctor's submitted list under the pressure structure F, $\lambda(F) = (\lambda_1(F),\dots,\lambda_L(F))$, be such that (i) $\lambda_1(F) = i$ if $i \in F$ and $i < i'$ for all $i' \in F \setminus \{i\}$, and (ii) $\lambda(F) = (1,\dots,L)$ if $F = \emptyset$. That is, the doctor chooses to list first the most preferred program among pressuring ones, and submits just the true preference if there is no pressure.

Note that our assumption on the doctor's choice among pressuring programs is most optimistic about the welfare loss caused by the list change, since in general a less preferable pressuring program may be chosen to be listed first. In this sense, our welfare loss estimate derived later is just the lower bound, so that it may underestimate the value of removing such pressures.

[12] In summary, there are three kinds of orders for a particular applicant: the order by true preference (Section 2.1.1), the order in the list submitted to the clearinghouse (Section 2.1.2), and the order in the list used by the matching algorithm after probabilistic permutation (Section 2.1.5).

2.1.3 Pressure and acceptance

Let $e \in (0,1)$ be an independent and an identical probability that each program that is going to accept the doctor will exert effective pressure on that doctor. Also, the probability that a program that does not accept the doctor will exert pressure is zero. In summary, we denote $\tilde{F}^e(A)$ for the pressure structure, that is $i \in \tilde{F}^e(A)$ with probability $e \in (0,1)$ if $i \in A$, and 0 if $i \notin A$, independently.

Note that, by letting the probability be zero if the program i is not the element of the set A, we ignore the possibility that a program uncertain of the acceptance may succeed in put an effective pressure on the doctor.[13] By the identicalness and independence, we also ignore that the decisions of the doctor and a program regarding the pressure in general depend on how the doctor evaluates the program and whether other programs also put pressure.

2.1.4 Acceptance

We assume that the probability that a residency program will accept the doctor is $a \in (0,1)$, and the acceptance decisions of individual programs are independent. We denote as $\tilde{A}^a$ the set of programs that accept the doctor, that is, $i \in \tilde{A}^a$ with probability $a \in (0,1)$, independently.

By the identicalness, we ignore the possibility of correlation between the preference of a doctor and the priority for each program, such as that caused by the fitness of the doctor for each program.[14] By the independence, we also ignore the possibility of some doctor generally being prioritized over another doctor, such as that caused by the former having higher ability than the latter.

Then, by letting $\tilde{\lambda}^{a,e} = (\tilde{\lambda}_1^{a,e}, \cdots, \tilde{\lambda}_L^{a,e}) \equiv \lambda(\tilde{F}^e(\tilde{A}^a))$ for notational simplicity, the event that a program i is the kth in the submitted list is represented by $i = \tilde{\lambda}_k^{a,e}$.

2.1.5 Random exchange

We will also consider adding a random permutation of the doctor's submitted list $\boldsymbol{l} = (l_1, l_2, l_3, \ldots, l_L)$ before running the matching algorithm. In particular, for $\varepsilon \in [0, 1/2]$, let $\tilde{\pi}^\varepsilon(\boldsymbol{l}) = (l_2, l_1, l_3, \ldots, l_L)$ with probability ε and $\boldsymbol{l}$ with probability $1 - \varepsilon$.[15] That is, we will consider a random exchange of the first and the second, with the probability of $\varepsilon > 0$,[16] and then, given a submitted list $\boldsymbol{l}$ and an acceptance structure A, the resulting matching becomes $\tilde{\mu}^\varepsilon(\boldsymbol{l}, A) \equiv \mu(\tilde{\pi}^\varepsilon(\boldsymbol{l}), A)$. Note that if $\varepsilon = 0$, then they are the original list and procedure, $\tilde{\pi}^0(\boldsymbol{l}) = \boldsymbol{l}$ and hence $\tilde{\mu}^0(\boldsymbol{l}, A) = \mu(\boldsymbol{l}, A)$.

2.2 Key statistics and remarks

2.2.1 Conditional match rates

Let $P_k^{a,e}$ denote the kth *conditional match rate*, the probability to be matched with the kth program conditional on that the doctor has already been rejected by all programs listed higher, that is, $P_1^{a,e} \equiv \Pr\left(\tilde{\lambda}_1^{a,e} \in \tilde{A}^a\right)$, and

[13] Taking such a possibility into account may be important, given that it is pointed out (see, e.g., Doolittle (2017)) that an applicant may misunderstand a message from a program which itself has no intention to put such a pressure. We can introduce and quantify this possibility by instead letting $i \in \tilde{F}^e(A)$ with probability $\underline{e} \in (0, e]$ if $i \notin A$. Then, the possibility of misunderstanding is captured by the decline in the conditional match rates from the second to the third, since it becomes possible that even after the rejection by the first, there remains a program which both puts pressure on and would accept the doctor, which cannot be if $\underline{e} = 0$, as discussed below.

[14] In fact, by assuming an extreme correlation distribution, we can reproduce the conditional match rate pattern without resorting to the pressure process, which may actually be plausible to explain just one observation. However, the pattern is found both in the Japanese and the U.S. markets and also across time, with so different values that considering that the parameters of the extreme distribution happens to be adjusted appropriately for each case may be unreasonable.

[15] This paper examines a simple method of changing the doctor's submitted list. Naturally, there can be other effective ways to change the list. It is also important to consider the measures that the programs can take against such intervention methods. For example, a program could not only ask a doctor to write it first on the list, but also request that the second and lower places on the list be left blank.

[16] For the cases where the exchange probability e is greater than 1/2, the doctor can deal with it by replacing the first and the second positions on his or her list in advance. Therefore, it is sufficient to consider cases where the probability is less than 1/2.

for all $k \in \{2,\dots,L\}$,

$$P_k^{a,e} \equiv \Pr\left(\tilde{\lambda}_k^{a,e} \in \tilde{A}^a \middle| \tilde{\lambda}_{k'}^{a,e} \notin \tilde{A}^a \text{ for all } k' < k\right).$$

Then, we show that the conditional match rate first drops and then becomes constant, as observed in the Japanese and the U.S. markets.

Proposition. $P_1^{a,e} > P_2^{a,e} = \cdots = P_L^{a,e}$.

Proof. We show that $P_1^{a,e} = 1-(1-ae)^L\frac{1-a}{1-ae}$, which is larger than $1-(1-a\times 0)\frac{1-a}{1-a\times 1} = a$, and for all $k \in \{2,\dots,L\}$, $P_k^{a,e} = \frac{a-ae}{1-ae}$, which is smaller than $\frac{a-a\times 0}{1-a\times 0} = a$. For notational simplicity, let $\tilde{F}^{a,e} \equiv \tilde{F}^e\left(\tilde{A}^a\right)$.

For the former, note that $\tilde{\lambda}_1^{a,e} \in \tilde{A}^a$ if $\tilde{F}^{a,e} \neq \emptyset$, since by $\Pr\left(i \in \tilde{F}^e(A) | i \notin A\right) = 0$, $i \in \tilde{F}^e(A)$ implies $i \in A$, and that $\tilde{\lambda}^{a,e} = (1,\dots,L)$ if $\tilde{F}^{a,e} = \emptyset$. Thus, $P_1^{a,e} = 1-\Pr\left(\tilde{\lambda}_1^{a,e} \notin \tilde{A}^a\right)$ is equal to

$$1-\Pr(\tilde{F}^{a,e}=\emptyset)\Pr(1\notin\tilde{A}^a|\tilde{F}^{a,e}=\emptyset) = 1-\Pr(\tilde{F}^{a,e}=\emptyset)\Pr(1\notin\tilde{A}^a|1\notin\tilde{F}^{a,e}),$$

where we use the conditional independence for the last equality.

For the latter, by $\Pr\left(\tilde{\lambda}_1^{a,e} \notin \tilde{A}^a \middle| \tilde{F}^{a,e} \neq \emptyset\right) = 0$, $P_k^{a,e} = \frac{\Pr\left(\tilde{\lambda}_{k'}^{a,e}\notin\tilde{A}^a \text{ for all } k'<k \text{ but } \tilde{\lambda}_k^{a,e}\in\tilde{A}^a\right)}{\Pr\left(\tilde{\lambda}_{k'}^{a,e}\notin\tilde{A}^a \text{ for all } k'<k\right)}$ is equal to

$$\frac{\Pr\left(\tilde{\lambda}_{k'}^{a,e} \notin \tilde{A}^a \text{ for all } k'<k \text{ but } \tilde{\lambda}_k^{a,e} \in \tilde{A}^a \middle| \tilde{F}^{a,e}=\emptyset\right)}{\Pr\left(\tilde{\lambda}_{k'}^{a,e} \notin \tilde{A}^a \text{ for all } k'<k \middle| \tilde{F}^{a,e}=\emptyset\right)},$$

which is equal to, since $\tilde{\lambda}^{a,e} = (1,\dots,L)$ if $\tilde{F}^{a,e} = \emptyset$,

$$\begin{aligned}
&\frac{\Pr\left(k' \notin \tilde{A}^a \text{ for all } k'<k \text{ but } k \in \tilde{A}^a \middle| \tilde{F}^{a,e}=\emptyset\right)}{\Pr\left(k' \notin \tilde{A}^a \text{ for all } k'<k \middle| \tilde{F}^{a,e}=\emptyset\right)} \\
&= \frac{\Pr\left(k' \notin \tilde{A}^a \text{ for all } k'<k \text{ but } k \in \tilde{A}^a\right)\Pr\left(\tilde{F}^{a,e}=\emptyset \middle| k' \notin \tilde{A}^a \text{ for all } k'<k \text{ but } k \in \tilde{A}^a\right)}{\Pr\left(k' \notin \tilde{A}^a \text{ for all } k'<k\right)\Pr\left(\tilde{F}^{a,e}=\emptyset \middle| k' \notin \tilde{A}^a \text{ for all } k'<k\right)} \\
&= \frac{\Pr\left(k \in \tilde{A}^a\right)\Pr\left(k \notin \tilde{F}^{a,e} \middle| k \in \tilde{A}^a\right)}{\Pr\left(k \notin \tilde{F}^{a,e}\right)},
\end{aligned}$$

where we use the conditional independence for the last equality. □

That is, since programs which put the pressures are also those which would accept the doctor, and the doctor lists first a program among who puts the pressure if there is at least one such a program, the fact that the program listed first rejects the doctor implies that all remaining programs are not those which both put the effective pressures and would accept the doctor, and that the submitted list is unchanged from the original one,[17] which uniformly reduce no less than the second conditional match rates, from $\Pr(i \in A)$ to $\Pr(i \in A | i \notin \tilde{F}^e(A))$. Moreover, the cases where the most preferred program would not accept the doctor but there is one which puts the pressure are added to the first match cases, increasing the probability from $\Pr(i \in A)$ to $1-\Pr\left(\tilde{F}^{a,e}=\emptyset\right)\Pr\left(1\notin\tilde{A}^a|1\notin\tilde{F}^{a,e}\right)$.

Note that, given L, we can recover a and e from the values of $P_1^{a,e}$ and $P_2^{a,e}$. In particular, since $1-P_2^{a,e} = 1-\frac{a-ae}{1-ae} = \frac{1-a}{1-ae}$, $1-ae = \left(\frac{1-P_1^{a,e}}{\frac{1-a}{1-ae}}\right)^{1/L} = \left(\frac{1-P_1^{a,e}}{1-P_2^{a,e}}\right)^{1/L}$, and hence

$$a = 1-(1-ae)(1-P_2^{a,e}) = 1-\left(\frac{1-P_1^{a,e}}{1-P_2^{a,e}}\right)^{1/L}(1-P_2^{a,e}), \tag{1}$$

[17] The critical assumption is that programs who would not accept the doctor cannot put the effective pressure. See also footnote 13.

by which

$$e = \frac{1-(1-ae)}{a} = \frac{1-\left(\frac{1-P_1^{a,e}}{1-P_2^{a,e}}\right)^{1/L}}{1-\left(\frac{1-P_1^{a,e}}{1-P_2^{a,e}}\right)^{1/L}(1-P_2^{a,e})}. \tag{2}$$

We will use them to empirically evaluate the Japanese and the U.S. markets in Section 3.

2.2.2 Rank loss

We discuss the rank loss caused by the pressure. The rank loss refers to how the actual matching deviates from that under the true preferences.

Let $RL^{a,e}$ denote the average rank loss caused by the forced list change, that is,

$$RL^{a,e} \equiv E\left[\mu\left(\tilde{\lambda}^{a,e},\tilde{A}^a\right) - \mu\left((1,\dots,L),\tilde{A}^a\right)\right],$$

where we let $\mu\left(\tilde{\lambda}^{a,e},\tilde{A}^a\right) - \mu\left((1,\dots,L),\tilde{A}^a\right) = 0$ if $\tilde{A}^a = \emptyset$.[18]

We use this as a measure of the welfare loss. Note that then, many points regarding welfare concern are ignored. For example, the importance of one rank gain may be different depending on rank levels, which may be incorporated by introducing some non-linear function to evaluate each rank as $u(i)$. Moreover, the existence of pressure itself may be stressful (see, e.g., Doolittle (2017)), even if the resulting match is unchanged, for which some evaluation of $\tilde{F}^{a,e}$ itself, such as $-c|\tilde{F}^{a,e}|$ with $c>0$ being the constant stress cost, is necessary.

It is remarkable that while the size of the first drop increases as the pressure probability (e) increases, the average rank loss, which first rises from zero, vanishes as it becomes almost sure.

Remark 1. *While $P_1^{a,e} - P_2^{a,e}$ is increasing in e, $RL^{a,e} > \lim_{e\downarrow 0} RL^{a,e} = \lim_{e\uparrow 1} RL^{a,e} = 0$.*

Proof. The former part simply follows from that $P_1^{a,e} = 1-(1-ae)^L\frac{1-a}{1-ae}$ is increasing, and $P_2^{a,e} = \frac{a-ae}{1-ae}$ is decreasing in e.

For the latter, note that $RL^{a,e}$ is equal to

$$\sum_{i=1}^{L-1} \Pr\left(\min_{i'\in\tilde{A}^a} i' = i\right) \Pr\left(\min_{i'\in\tilde{F}^{a,e}} i' > i \,\middle|\, \min_{i'\in\tilde{A}^a} i' = i\right) E\left[\min_{i'\in\tilde{F}^{a,e}} i' - i \,\middle|\, \min_{i'\in\tilde{F}^{a,e}} i' > \min_{i'\in\tilde{A}^a} i' = i\right],$$

the second term of which is equal to, by the conditional independence,

$$\Pr\left(i\notin\tilde{F}^{a,e}\,\middle|\, i\in\tilde{A}^a\right)\left(1-\prod_{i'>i}\Pr\left(i'\notin\tilde{F}^{a,e}\right)\right) = (1-e)(1-(1-ae)^{L-i}),$$

which converges, by $i<L$, to 0 as e approaches either 0 or 1. □

That is, given that the pressure probability conditional on the acceptance is almost 1, no pressure from a program implies almost surely that the program will not accept the doctor anyway, and hence lowering its rank in the list does not matter. Moreover, as long as the doctor chooses the best among the pressured programs as the first listed, the match outcome does not change.

Thus, a large drop in the conditional match rate as observed in the Japanese and the U.S. markets could be compatible with a negligible rank loss. Therefore, it is still an empirical issue whether the pressure so substantial as to cause such a large observational gap also causes a non-negligible welfare loss, which we will evaluate in the empirical part.

---

[18]Note that $\mu\left(\lambda(\tilde{F}^e(A),A\right) = \emptyset$ if and only if $\mu\left((1,\dots,L),A\right) = \emptyset$, which is equivalent to $A = \emptyset$.

2.2.3 Type I errors

In the following, we examine the possibility that a program that pressures a doctor to write itself first on the doctor's submitted list may be able to determine whether the doctor actually did write itself first.

Let $Q^{a,e,\varepsilon}_{i\to k}$ denote the probability of program i not to be matched with the doctor conditional on that the program would accept the doctor and the doctor follows the pressure to list the program kth, for which we further specify that the orders among others would remain the same in the changed list. That is, let $\tilde{\lambda}^{a,e}_{i\to k}$ be such that $\tilde{\lambda}^{a,e}_{i\to k,k}=i$ and for all $k',k''\in\left\{k'''\in\{1,\ldots,L\}\,\middle|\,\lambda^{a,e}_{k'''}\neq i\right\}$, $\tilde{\lambda}^{a,e}_{i\to k,k'}>\tilde{\lambda}^{a,e}_{i\to k,k''}$ if and only if $\tilde{\lambda}^{a,e}_{k'}>\tilde{\lambda}^{a,e}_{k''}$. Then,

$$Q^{a,e,\varepsilon}_{i\to k}=\Pr\left(\tilde{\mu}^{\varepsilon}\left(\tilde{\lambda}^{a,e}_{i\to k},\tilde{A}^{a}\right)\neq i\,\middle|\,i\in\tilde{F}^{a,e}\cap\tilde{A}^{a}\right).$$

Note that $Q^{a,e,0}_{i\to k}$ is the probability under the original procedure, with no addition of permutation.

We consider that with a significance level $\alpha>0$, conditional on that the program is not matched with the doctor, the hypothesis that the doctor follows the pressure to list kth a program that would accept the doctor is rejected if $Q^{a,e,\varepsilon}_{i\to k}\leq\alpha$, and this rejection possibility makes the doctor follow the pressure. Then, we see the specialty of first listing under the original procedure ($\varepsilon=0$), that the hypothesis rejection is possible with any small significance level, or the zero type I error, while the necessary significance level is bounded away from zero for all other ranks.

Remark 2. $Q^{a,e,0}_{i\to 1}=0\leq\alpha$ for any α. Moreover, for all $k\geq 2$, $Q^{a,e,0}_{i\to k}\geq\underline{Q}^{a,e}\equiv 1-(1-a)(1-ae)^{L-2}$.

Proof. By $\tilde{\mu}^0=\mu$, $\tilde{\lambda}^{a,e}_{i\to 1,1}=i$ and $i\in\tilde{A}^a$ imply $\tilde{\mu}^0(\tilde{\lambda}^{a,e}_{i\to 1},\tilde{A}^a)=i$, which proves the former. Moreover, for the latter, for $k\geq 2$, $Q^{a,e,0}_{i\to k}$ is equal to

$$\begin{aligned}&1-\Pr\left(\tilde{A}^a\cap\{1,\ldots,k-\mathbf{1}_{i\geq k}\}\setminus\{i\}=\emptyset\right)\Pr\left(\tilde{F}^{a,e}\cap\{k+\mathbf{1}_{i<k},\ldots,L\}\setminus\{i\}=\emptyset\,\middle|\,\tilde{A}^a\cap\{1,\ldots,k-\mathbf{1}_{i\geq k}\}\setminus\{i\}=\emptyset\right)\\&=1-(1-a)^{k-1}(1-ae)^{L-k},\end{aligned}\tag{3}$$

where we use the conditional independence for the last equality.[19] □

That is, since the program would accept the doctor, as long as the doctor truly listed the program first, the program would be certainly matched with the doctor. On the other hand, for other ranks, even if the doctor follows the listing, the unmatch may occur because the doctor may match another program listed higher, the probability of which is enhanced by the existence of first listing pressure which makes programs listed higher more probable to accept the doctor. Of course, it may be negligible if the lower bound itself is positive but so small as standard significance levels, such as 10%, which will be checked after the estimation of a and e.

Then, we consider introducing a random exchange of the first and second in the submitted list with a probability small but sufficient to negate the testing. Given the following monotonicities, for the complete negation (in the sense of any k and e), it is necessary and sufficient letting the exchange probability ε so large that the type I error for the first listing given that other programs do not exercise any pressuring is larger than the significance level α, $Q^{a,0,\varepsilon}_{i\to 1}\equiv\lim_{e\to 0}Q^{a,e,\varepsilon}_{i\to 1}>\alpha$.

Remark 3. $Q^{a,e,\varepsilon}_{i\to k}\geq Q^{a,0,\varepsilon}_{i\to 1}$ *irrespective of e and k. Moreover,* $Q^{a,0,\varepsilon}_{i\to 1}$ *is increasing in* ε*.*

Proof. For the former, since $(\tilde{\pi}^{\varepsilon}_1(\boldsymbol{l}),\tilde{\pi}^{\varepsilon}_2(\boldsymbol{l}))=(l_1,l_2)$ with probability ε and (l_2,l_1) with probability $1-\varepsilon$, $Q^{a,e,\varepsilon}_{i\to 1}=\varepsilon Q^{a,e,0}_{i\to 2}+(1-\varepsilon)Q^{a,e,0}_{i\to 1}$ and $Q^{a,e,\varepsilon}_{i\to 2}=\varepsilon Q^{a,e,0}_{i\to 1}+(1-\varepsilon)Q^{a,e,0}_{i\to 2}$. Moreover, since $\{\tilde{\pi}^{\varepsilon}_1(\boldsymbol{l}),\ldots,\tilde{\pi}^{\varepsilon}_{k-1}(\boldsymbol{l})\}=\{l_1,\ldots,l_{k-1}\}$ for $k>2$, $Q^{a,e,\varepsilon}_{i\to k}=Q^{a,e,0}_{i\to k}$ for $k>2$. Then, since by (3), $Q^{a,e,0}_{i\to k}\geq Q^{a,e,0}_{i\to 2}\geq Q^{a,e,0}_{i\to 1}$ for $k>2$, by $\varepsilon\leq 1/2$, $Q^{a,e,\varepsilon}_{i\to k}\geq Q^{a,e,\varepsilon}_{i\to 1}$ for $k\geq 2$. Moreover, again by (3), $Q^{a,e,\varepsilon}_{i\to 1}\geq Q^{a,0,\varepsilon}_{i\to 1}$.

For the latter, by (3), $Q^{a,0,\varepsilon}_{i\to 1}=1-\varepsilon(1-a)-(1-\varepsilon)=a\varepsilon$. □

[19] Here $\mathbf{1}_x$ is the indicator function that is eqaul to 1 if x holds; 0 otherwise.

At the same time, since the exchange itself causes a welfare loss, even if the negation is achieved, the associated loss by the permutation intervention,

$$PRL^{a,\varepsilon} \equiv E\left[\tilde{\mu}^{\varepsilon}\left((1,\ldots,L),\tilde{A}^{a}\right) - \mu\left((1,\ldots,L),\tilde{A}^{a}\right)\right],$$

may also be substantial.

Remark 4. *$PRL^{a,\varepsilon}$ is also increasing in ε.*

Proof. Note that $PRL^{a,\varepsilon} = (2-1)\Pr\left(\{1,2\} \subset \tilde{A}^{a}\right)\Pr\left(\tilde{\pi}_1^{\varepsilon}((1,\ldots,L)) = 2\right) = a^2\varepsilon$. □

Moreover, while the loss to yield a given type I error is smaller with a smaller acceptance rate, a smaller rate also more restricts the possibility of the negation.

Remark 5. *$Q_{i\to 1}^{a,0,\varepsilon} = \alpha$ for some ε if and only if $\alpha \leq a/2$, which is increasing in a, while $PRL^{a,\varepsilon}$ with ε solving $Q_{i\to 1}^{a,0,\varepsilon} = \alpha$ is also increasing in a.*

Proof. For the former, recall $Q_{i\to 1}^{a,0,\varepsilon} = a\varepsilon$, which also implies the latter by $PRL^{a,\varepsilon} = a^2\varepsilon$. □

Thus, whether in the first place, a sufficiently large type error can be achieved, and if so, with so allowable associated loss also remains the empirical issue.

3 Empirical evaluation

For each market, we set (P_1, P_2, L),[20] derive (a,e) from (1) and (2) with them, and calculate key statistics as the following table, where RL_{rand}^{a} denotes the rank loss if the matched program is randomly chosen from those which would accept the doctor,

$$RL_{\text{rand}}^{a} = \left(1 - \Pr\left(\tilde{A}^{a} = \emptyset\right)\right)\left(\frac{L+1}{2} - E\left[\mu\left((1,\ldots,L),\tilde{A}^{a}\right)\middle|\tilde{A}^{a} \neq \emptyset\right]\right),$$

and PRL_{α}^{a} denotes the rank loss in order to attain the type I error α, i.e., with $a\varepsilon = \alpha$,

$$PRL_{\alpha}^{a} = PRL^{a,\alpha/a}.$$

Market	P_1	P_2	L	a	e	RL	RL_{rand}	$\underline{Q}$	$PRL_{0.15}$	QL
Japan	0.7	0.5	4	0.5599	0.2141	0.3425	0.8373	0.6591	0.0840	0.0685
U.S.	0.5	0.3	10	0.3232	0.1024	0.7740	2.5588	0.4829	0.0485	0.0704

Thus, the rank loss reaches into more than 30% ($0.7740/0.2558 \approx 0.302$) of what the efficient assignment adds to the random assignment. Moreover, we can achieve the type I error of 15% ($0.3232/2 \approx 0.162$), with the associated loss less than a quarter ($0.0840/0.3425 \approx 0.245$) of the improvement by the negation.

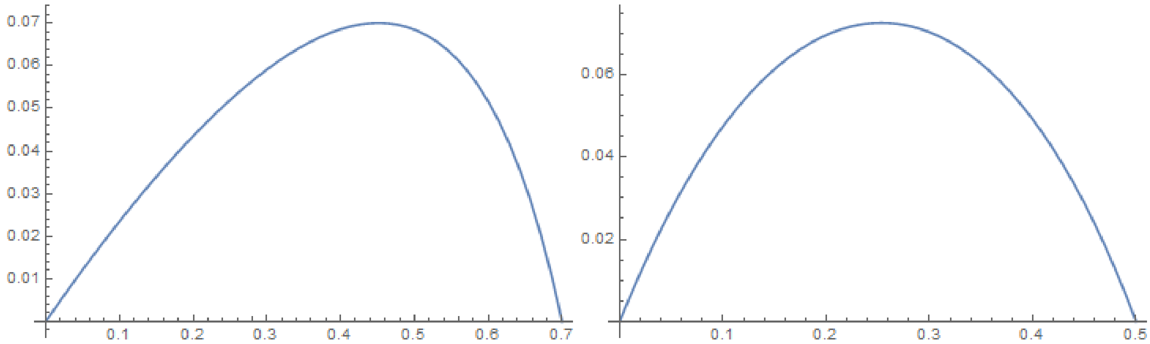
For the rank loss evaluation, consider that the match value of each program for the doctor is independently drawn from an identical continuous distribution, and the preference order is determined according to these values. Then, as is easily calculated, one rank difference is translated into $1/(L+1)$ quantile difference, that is, the expected quantile loss is

$$QL^{a,e} = \frac{RL^{a,e}}{L+1},$$

which for each market is also reported in the table, where we find that once evaluated in quantile, the estimated losses are very similar. Thus, on average, around 7 percent point quantile is lost in both markets.

We also plot how the estimated quantile loss would be if P_2 took different values, as follows. Somewhat interestingly, the actual P_2 attains almost highest value in both markets.

[20] There is a much larger room for the choice of L since it is very heterogeneous in the data, while we also find that the results are fairly robust to its reasonable variation.



QL (vertical) by P_2 (horizontal) for Japan (left) and U.S. (right)

Finally, it may also be interesting that the estimated pressure probability is much smaller ($0.1024 \ll 0.2141$) in the U.S. market, where the matching authority explicitly prohibits such conduct.

4 Concluding remarks

In this paper, we studied a conditional match rate anomaly and the reason why this occurs. First, we showed that such an anomaly could be explained by the possibility that a program planning to accept an applicant forces the applicant to list it first. Then, under a simple reduced-form acceptance and forcing process, we demonstrate that the U.S. and Japanese data imply substantial forces and welfare loss.

Finally, we have suggested one idea that could alleviate this problem, that is introducing a random permutation of the applicants' submitted list before running the matching algorithm and showed that the introduction pays if the intervention makes the pressure ineffective.

This paper showed that the probability of a program exerting pressure on an applicant is approximately twice as high in Japan (0.2141) as in the U.S. (0.1024). This is an interesting difference, which may be due to the weak (or non-existent) direct regulation by the communication code of conduct in Japan.

The method of estimating the pressure probabilities described in this paper may be used for the matching markets other than the residency matching. In recent years, the matching algorithms have been utilized in many environments. For example, Google uses algorithms in determining worker assignments.[21] When the matching algorithm is used within a company in this way, the players may try to influence the matching results by putting pressure on each other. We need to do further research on the inefficiencies created by pressure and how to solve them.

References

[1] Doolittle, B.D. (2017) "Communication after the residency interview: A call for professionalism," *NEJM Knowledge+*, (https://knowledgeplus.nejm.org/blog/communication-after-residency-interview/).

[2] Grimm, L. J., Avery, C. S., and Maxfield, C. M. (2016) "Residency postinterview communications: More harm than good?" *Journal of Graduate Medical Education*, 8(1): 7–9.

[3] Kojima, F., and Pathak, P. A. (2009) "Incentives and stability in large two-sided matching markets." *American Economic Review*, 99 (3): 608–27.

[4] Rees-Jones, A. (2018) "Suboptimal behavior in strategy-proof mechanisms: Evidence from the residency match," *Games and Economic Behavior*, 108, 317–330.

[5] Rees-Jones, A. and Skowronek, A. (2018) "An experimental investigation of preference misrepresentation in the residency match," *Proceedings of the National Academy of Sciences*, 115 (45): 11471–11476.

[21] See for detail "Google's algorithm-powered internal job marketplace" (https://rework.withgoogle.com/blog/googles-algorithm-powered-internal-job-marketplace/).

[6] Roth A.E. (2003) “The origins, history, and design of the resident match,” *JAMA*. 289(7): 909-–912. doi:10.1001/jama.289.7.909 (https://jamanetwork.com/journals/jama/fullarticle/195998).

[7] Sbicca, J. A., Gordon, K., and Takahashi, S. (2012) “National resident matching program violations,” *AMA Journal of Ethics*, 14(12): 932–936.

[8] Yarris, L.M., Delorio, N.M., and Gaines, S.S. (2010) “Emergency medicine residency applicants’ perceptions about being contacted after interview day,” *Western Journal of Emergency Medicine*, 11(5): 474–478.