Only You: A Field Experiment of Text Message to Prevent Free-riding in Japan Marrow Donor Program

Hiroki Kato

Fumio Ohtake

Saiko Kurosawa

Kazuhiro Yoshiuchi

Takahiro Fukuda

Last Updated on May 8, 2023

# 1. Introduction

同種造血幹細胞移植は白血病などの血液病に対する最も再発率の低い治療法の一つである。この治療法では、(1)抗がん剤や放射線によって異常な細胞と健康な造血幹細胞を同時に殺し、(2)他者から提供された健康な造血幹細胞を移植する。骨髄移植（造血幹細胞移植）における要件は、ドナーのHLAと呼ばれる白血球の型が患者のHLAと一致していることである[[1]](#footnote-20)。ランダムに選ばれた二人のマッチング確率は1%未満である一方で、兄弟姉妹間で一致する可能性が最も高く、その確率は30%程度である。また、親子間のマッチング確率はかなり低い。親族の中に適合するドナーがいなければ、患者は近親者以外からドナーを探さなければならない。日本において、患者は一般的に日本骨髄バンク（JMDP）を介して非近親者のドナーを探すことになる。しかしながら、JMDPによるコーディネーションは移植に至るまで長時間を要し、登録患者の60%のみが移植を受けられない (Hirakawa et al. 2018)。

患者の生存確率を高めるためのドナープールに対する介入は二種類ある。第一に、潜在的なドナーの人数を増やして、マッチング確率を高めるような政策である。4人未満のドナーと合致する患者と比較して、200人以上のドナーと合致する患者の移植率は45%から74%に上昇する (Hirakawa et al. 2018)。しかしながら、2000年から2015年にかけて潜在的なドナーの数は約二倍に増えたにもかかわらず、初回マッチング確率は5%程度しか増えていない [Takanashi (2016)][[2]](#footnote-21)。したがって、潜在的なドナーを増やすことの限界便益は小さいので、プールの規模を拡大することは非効率的であるだろう。

第二の政策はドナープールに提供を断らない潜在的なドナーの比率を高めることである。この種の政策はドナープールの質に貢献するものである。Hirakawa et al. (2018) は多くの移植コーディネーションはドナー側の都合により、コーディネーションの第一過程である確認検査の前に中断していることを示している（コーディネーションの過程は [Section 2.1](#sec-background) を参照）。したがって、提供の意思の強い多くの潜在ドナーでできたプールを作ることは移植率を高めるだろう。また、ドナープールの質を改善する政策の限界便益はドナープールの規模を拡大する政策のそれよりも高いだろう。

そこで、本研究はドナープールの質を改善する政策の一つとして情報提供の効果を検証する。JMDPに登録した人がある患者の潜在的なドナーとなると、潜在ドナーはJMDPから適合通知を受け取る。そして、提供意思を示して適合通知に返信する潜在ドナーは移植のためのコーディネーションを受ける。我々はJMDPによって公開されている情報に基づいて適合通知に新規のメッセージを加えて、追加メッセージの効果を検証するフィールド実験を2021年9月から2022年2月にかけて実施した。介入メッセージは二つある。第一に、患者一人あたりの潜在的なドナーの数が少ないという情報（確率情報）である。患者一人につき複数の潜在的なドナーが同時にコーディネーションを受けるので、JMDPを通じた幹細胞移植は公共財の性質を持つ。したがって、同時にコーディネーションを受けている潜在的なドナーが多ければ多いほど、フリーライドするインセンティブが強くなる。第一のメッセージはコーディネーション進行中のドナーの数に関する過剰な期待によって生じるただ乗り行動を妨げることを目的としている。もう一つのメッセージはJMDPに登録した患者の半分しか移植を受けられないというもの（利他情報）である。この情報は幹細胞提供という利他行動の価値を高めることを目的としている。また、このメッセージは候補となるドナーが早く見つけられるほど、移植率を高められることも明記して、適合通知に早く返信してもらうことを目的としている。

我々はJMDPに実験機関に適合通知を受け取った人に関するコーディネーション過程のデータの提供を依頼し、情報を提供することの効果を分析した。我々のフィールド実験は情報提供の効果は異質的であることを示している。確率情報は移植成績がよいにもかかわらず適合通知への返信率が低い若年層の男性に効果的であるが、利他情報は若年男性に効果的でなかった。さらに、二つの情報を同時に提供する介入も、利他情報に強く影響を受けてしまうために、効果的でなかった。

# 2. Field Experiment

## 2.1 Background: Coordination Process of JMDP

To promote an understanding of the timing of interventions in our field experiment, we provide an overview of the coordination process leading up to the donation of stem cells by a potential donor registered with the JMDP. First, when a potential donor matches a patient registered with the JMDP, the JMDP office sends the potential donor the compatibility notice requesting stem cell donation.[[3]](#footnote-23) The potential donor responds to the notification by filling out a medical questionnaire and indicating the willingness to donate.

After that, the coordination for the transplant starts. The potential donor undergoes the confirmatory typing within approximately one month. In this step, the coordinator explains in detail the donation procedure (bone marrow collection and peripheral blood stem cell collection) and investigates the intentions of the potential donor and family. At this time, the potential donor can choose the collection method. In addition, the coordinating physician conducts an interview, medical examination, and general blood tests to determine the presence of infection and blood type. These tests examine whether the potential donor meets the criteria set by the JMDP.

Patients can proceed with coordination with up to 10 potential donors at the same time. The patient’s physician selects the most suitable candidate from potential donors who take the confirmatory typing. Importantly, the potential donor does not know how many other donors the matched patient coordinates with. Nor can a potential donor obtain this information from the coordinator or coordinating physician.

The candidate needs to make a final consent after receiving explanations from the coordinator and coordinating physician. At this time, a representative of the donor’s family must also consent to the collection. In addition, the candidate cannot change his/her intention after that. After giving final consent, the first candidate is hospitalized for approximately one week to take preoperative testing and prepare for collection. After this, the donor undergoes a procedure to collect the stem cells. The time between the confirmation test and collection is approximately 3-4 months.

## 2.2 Experimental Design

|  |
| --- |
| Figure 1: Intervention Messages |

Our intervention is the content of the compatibility notice in which the JMDP requests stem cell donations from potential donors. The [Figure 1](#fig-intervention) shows our intervention. Before the intervention, the compatibility notice states that the potential donor should return within seven days indicating to donate and completing a medical questionnaire. One of the materials enclosed along with the notification is a handbook that describes the coordination process, which we overview in the previous subsection.

We added two messages to the compatibility notice to facilitate coordination ([Figure 1](#fig-intervention)).[[4]](#footnote-29) The probability message highlights the small number of potential donors per patient registered with the JMDP. This message prevents overestimation of beliefs about the number of other ongoing potential donors and may discourage free-riding behavior in stem cell donation.

The patient message emphasizes that only half of the patients can receive a transplant through the JMDP. This message aims to correct the perception that patients are saved without the cooperation and to stimulate altruism in potential donors. Furthermore, emphasizing the importance of shortening the coordination time encourages early replies to the compatibility notice.

We made four experimental arms to estimate the effects of the two intervention messages. Experimental arm A sends the standard compatibility notice without intervention messages (control group). Experimental arms B and C receive the probability message and the patient message, respectively. In addition, to test the negative effect of cognitive load due to information overload, we also made an experimental arm D adding two intervention messages to one compatibility notice.

**Table** **:** Assignment Schedule

| week | Sep 21 | Oct 21 | Nov 21 | Dec 21 | Jan 22 | Feb 22 |
| --- | --- | --- | --- | --- | --- | --- |
| Week 1 | B | C | C | D | B | A |
| Week 2 | D | B | A | A | C | B |
| Week 3 | A | D | B | C | D | C |
| Week 4 | C | A | D | B | A | D |

**?(caption)**

Our experimental subjects are 11,154 potential donors to whom the JMDP sent the compatibility notice between September 2021 and February 2022. The assignment of experimental groups is cluster-randomized based on a week because we preserve as much randomness as feasible by the JMDP office. We assigned four experimental groups to balance on week and month to remove week/month fixed effects. We summarize the assignment schedule in **?@tbl-assignment**.

## 2.3 Data and Empirical Strategy

We received the coordination data maintained by the JMDP at the end of June 2022. The observation unit is the experimental subject. As individual attributes, the data record gender, age, number of past coordination, and region of residence (at the prefectural level). For the coordination process, the data records whether or not to reach each step (reply to the compatibility notice, the confirmatory typing, the candidate selection, the final consent, and the collection). We use these variables as outcome variables. In particular, for replies to the compatibility notice, the data additionally record the number of days of reply and the intention to donate. If the coordination was interrupted, the data record the reason for the interruption in three categories (patient reasons, reasons other than donor’s health, and donor’s health reasons). In analysis, we use 11,049 individuals who reside in Japan and whose coordination has been completed.[[5]](#footnote-32)

As additional data, we use the list of hospitals published by the JMDP on its website. In addition to hospital addresses, this data includes whether bone marrow (BM) collection and peripheral blood stem cell (PBSC) collection are available. We aggregate this data at the prefectural level, calculate the number of hospitals per 10 square kilometers, and merge it and the coordination data using the prefecture as a merge key. We consider this variable as the traveling cost of coordination and donation.

**Table** **:** Overview of Field Experiment

|  | Experimental Arms | | | |  |
| --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | F-test, p-value |
| A. Interventions | | | | | |
| Standard notification | X | X | X | X |  |
| Probability message |  | X |  | X |  |
| Patients message |  |  | X | X |  |
| B. Sample Size | | | | | |
| N | 2535 | 3053 | 2726 | 2735 |  |
| C. Balance Test | | | | | |
| Male (=1) | 0.624 | 0.633 | 0.631 | 0.609 | 0.424 |
| Age | 38.376 | 38.121 | 37.448 | 37.978 | 0.096 |
| Number of past coordinations | 1.609 | 1.589 | 1.625 | 1.563 | 0.515 |
| Number of listed hospitals | 0.476 | 0.490 | 0.487 | 0.485 | 0.826 |
| Number of hospitals listed with PBSC collection | 0.162 | 0.167 | 0.166 | 0.164 | 0.831 |
| Number of hospitals listed with BM collection | 0.246 | 0.256 | 0.254 | 0.251 | 0.711 |
| Notes: Balance test regresses a covariate on treatment dummies and test a null hypothesis that all coefficients are zero. We use the clustered standard error with the CR2 adjustment for inference (cluster unit is an experimental week). | | | | | |

**?(caption)**

**?@tbl-experiment-summary** summarizes the field experiments. Panel A shows the interventions for each experimental group, and Panel B shows the sample size for each experimental group. Panel C shows the balance test, which examines whether the randomization is successful. Since most covariates are well-balanced across groups, the clustered assignment is approximately random. However, potential donors in experimental groups C and D may be younger than those in experimental groups A and B (F-test, p-value = ).

Our target parameter is the overall population average treatment effect, which can be identified by the difference in means due to exogenous treatment assignment. In addition to the difference in means, we estimate a linear probability model to remove month/week fixed effects and to increase the precision of inference by adding covariates. Our estimation model for individual who received the compatibility notice in a week of month is as follows:

where is the individual attribute vector and and are the week and month dummy variables, respectively. We use standard errors clustered at the experimental week level, a unit of randomization.

For inference, we use standard errors clustered by experimental week, which is the unit of randomization. If the distribution of the treatment variable is skewed (the skewness of the binary variable indicating experimental arm B is approximately 1), the standard clustered standard errors will be small, which leads to an excessively low coverage probability of the confidence interval. Therefore, following the recommendation of Imbens and Kolesár (2016), we adjust and report standard errors and degrees of freedom by CR2.[[6]](#footnote-34)

# 3. Experimental Results

## 3.1 Effects on Reply and Intention

To begin, we estimate the effect of our interventions on reply, which is the most likely manifestation of the potential donor’s intention. The outcome variable for replying is a dummy variable that takes one if the potential donor replies to the compatibility notice regardless of donor intention. The response rate is % in the control arm (experimental arm A).

We decompose the message effect on response to the notification into two parts. The first is the effect on replying with a positive intention. For estimation, the outcome variable is a dummy variable that takes one if the potential donor replied to the notice and indicated the willingness to donate. The second is the effect on replying with a negative intention. The outcome variable is a dummy variable that takes one if the potential donor replied to the notice and did not indicate the willingness to donate. When estimating these effects, we include non-responders in the sample, coding their outcomes as 0. Since the sum of the two outcome variables for the intention to donate is always equal to the dummy variable for the reply, the sum of the effects on the positive and negative intentions is always the effect on replying. The proportion of replying with positive intention is %, while the proportion of replying with negative intention is % in the control arm. Thus, % of potential donors who replied to the compatibility notice were willing to donate.

The coordination may be interrupted by the patient before potential donors reply. Since this case occurs independently of the potential donor’s intention, we exclude it from the sample in our analysis. We exclude –% of each experimental group, and the proportions are well balanced across groups (F-test, p-value=).

**Table** **:** Linear Probability Model of Replay and Intention

|  | | | Intention | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | Reply | | Positive | | Negative | |
|  | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment B | 0.0105 | 0.0139 | 0.0218 | 0.0210 | -0.0113 | -0.0071 |
|  | (0.0082) | (0.0088) | (0.0146) | (0.0172) | (0.0103) | (0.0103) |
| Treatment C | -0.0057 | 0.0026 | -0.0053 | -0.0029 | -0.0004 | 0.0054 |
|  | (0.0087) | (0.0063) | (0.0163) | (0.0152) | (0.0162) | (0.0126) |
| Treatment D | 0.0058 | 0.0062 | 0.0046 | 0.0058 | 0.0012 | 0.0004 |
|  | (0.0079) | (0.0073) | (0.0180) | (0.0139) | (0.0151) | (0.0094) |
| Control average | 0.8835 | 0.8835 | 0.5533 | 0.5533 | 0.3303 | 0.3303 |
| Covariates |  | X |  | X |  | X |
| Num.Obs. | 10985 | 10985 | 10985 | 10985 | 10985 | 10985 |
| Notes: \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01. The clustered standard errors with the CR2 adjustment are reported in parenetheses (cluster unit is an experimental week). Covariates are gender, squared polynomial of (demeaned) age, number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, number of hospitals with BM collection per 10 square kilometers, month dummies, and week dummies. | | | | | | |

**?(caption)**

**?@tbl-full-reply** is the estimation results of the linear probability model. Odd columns use only the experimental arm dummies as an explanatory variable, whereas even columns additionally control for individual attributes and month and week fixed effects. The results show that all experimental arms, with or without covariates or month/week fixed effects, have no statistically significant effects on overall replies or replies with positive or negative intentions. We estimated a logit model as robustness ( (**full-reply-logit?**) ) and obtained the same results (95 percent confidential interval of odds ratio includes 1).

|  |
| --- |
| Figure 2: Effect on Reply and Intentions by Gender and Age Group. Note: These plots show the average effect (and associated 95% confidential interval) on each outcome by gender and age group. We use clustered standard errors with the CR2 adjustment (cluster unit is experimental week). We control number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, number of hospitals with BM collection per 10 square kilometers, month dummies, and week dummies. |

Next, to test the heterogeneity of message effects, we divide the sample into four subsets by gender and age group (less than 30 or not) and estimate the message effects in each sub-group. [Figure 2](#fig-subsample-reply) is the coefficient plot. As a result, experimental arm B, which added only the probability message, statistically significantly increases the reply rates with the intention to donate by about 10 percentage points or % (the control average is %) for young men. Since experimental arm B reduces the rate of replies without donating intentions, the effect of this group on the overall reply rate is 6 percentage points (statistically significant at 10% level), which is smaller than its effect on replies with donating intentions.[[7]](#footnote-42) Given that young male donors have lower donation intentions than the other gender-age group despite their better transplant outcomes, the probability message improves the efficiency of transplant coordination. Note that, among the other gender and age groups, our intervention has no significant effect on replies or intentions.

## 3.2 Random Causal Forest

To dig deeper into the heterogeneity of intervention effects, we use a fully nonparametric method to test for heterogeneous effects called random causal forest (RCF) (Davis and Heller 2017; Athey and Wager 2019). This method divides the sample by combinations of covariates and estimates treatment effects within each subsample. A set of subsamples is determined to maximize the variance of the treatment effect (Athey and Imbens 2016). This method produces predicted treatment effects, which is a function of the covariates.[[8]](#footnote-44)

|  |
| --- |
| Figure 3: Boxplot of Predicted Treatment Effects by Gender and Age. Notes: Blue fitted line represents GAM smoothing. Covariates are gender, age, number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers and number of hospitals with BM collection per 10 square kilometers. |

[Figure 3](#fig-rcf-int-boxplot) plots the predicted treatment effect on replies with positive intention for each experimental group against age. Notably, for all men under age 25, the effect for experimental group B is positive. As a result, the RCF shows the average treatment effect of 12.1 percentage points for experimental group B among men under age 30, which is similar to the results of the subsample analysis of the linear regression analysis presented in [Figure 2](#fig-subsample-reply) (see **?@tbl-rcf-int-cate**). For a small fraction of men aged 30~40, the effect of experimental group B is more than 10 percentage points. We find that such men live in places with relatively more hospitals and have a lower traveling cost of coordination and donation.

The effects of experimental group C, with only altruistic messages added, are largely heterogeneous. For women in their early 30s, most treatment effects are negative. For men and women in their late 40s and older, the median value of treatment effect is positive but widely distributed. Surprisingly, we find that, for men and women in this age group, those for whom experimental group C has a positive effect live in areas with relatively few hospitals (high traveling costs).

**Table** **:** Correlation of Predicted Tratment Effects

|  | Females | | Males | | Females | | Males | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Age<30 | 30≤Age | Age<30 | 30≤Age | Age<30 | 30≤Age | Age<30 | 30≤Age |
|  | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Constant | -0.0359\*\*\* | 0.0060\*\*\* | -0.0162\*\*\* | 0.0128\*\*\* | -0.0325\*\*\* | 0.0055\*\*\* | 0.0038\* | 0.0078\*\*\* |
|  | (0.0018) | (0.0007) | (0.0021) | (0.0007) | (0.0017) | (0.0008) | (0.0021) | (0.0007) |
|  | [-0.0395, -0.0324] | [0.0046, 0.0074] | [-0.0203, -0.0120] | [0.0116, 0.0141] | [-0.0360, -0.0291] | [0.0040, 0.0069] | [-0.0002, 0.0079] | [0.0066, 0.0091] |
| B + C | 0.3245\*\*\* | 0.3578\*\*\* | 0.3819\*\*\* | 0.4164\*\*\* |  |  |  |  |
|  | (0.0195) | (0.0069) | (0.0182) | (0.0069) |  |  |  |  |
|  | [0.2862, 0.3628] | [0.3442, 0.3714] | [0.3462, 0.4175] | [0.4029, 0.4299] |  |  |  |  |
| B |  |  |  |  | 0.0120 | 0.4084\*\*\* | -0.0444\* | 0.6450\*\*\* |
|  |  |  |  |  | (0.0363) | (0.0135) | (0.0254) | (0.0150) |
|  |  |  |  |  | [-0.0592, 0.0833] | [0.3820, 0.4348] | [-0.0942, 0.0054] | [0.6156, 0.6745] |
| C |  |  |  |  | 0.6862\*\*\* | 0.3126\*\*\* | 0.9343\*\*\* | 0.2756\*\*\* |
|  |  |  |  |  | (0.0370) | (0.0123) | (0.0293) | (0.0101) |
|  |  |  |  |  | [0.6135, 0.7589] | [0.2885, 0.3368] | [0.8768, 0.9918] | [0.2557, 0.2955] |
| Num.Obs. | 1132 | 3018 | 1566 | 5333 | 1132 | 3018 | 1566 | 5333 |
| Notes: \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01. Robust standard errors are reported in parentheses. Square brackets show 95 percent confidential intervals. | | | | | | | | |

**?(caption)**

Experimental groups B and D contain probability messages. However, for young men, despite the positive effect of experimental group B on replies with intentions, experimental group D has no effect. To test this point, we analyze the correlation of treatment effects for each experimental group as predicted by the RCF.

If the participants fully understood the information in the message D, then the effect of experimental arm D should be the sum of the effects of experimental arms B and C. Columns (1)–(4) of **?@tbl-rcf-int-cate-corr** regress the predicted treatment effect of D on the sum of the predicted treatment effects of B and C. If we cannot reject that the coefficient on this explanatory variable is 1, the potential donor would have perfectly received the information from the message in experimental arm D. However, for all gender and age groups, the coefficient is around –, allowing us to reject the null hypothesis that the coefficient is 1. In other words, potential donors do not receive exactly the information from experimental group D.

Donors who do not receive accurate information may focus on either the probability or the altruistic message, or they may discount the information in two messages to the same degree. To test this point, columns (4)–(8) of **?@tbl-rcf-int-cate-corr** regress the predicted treatment effect in D on the predicted treatment effect in B and C. The results show that for young men and women, the correlation between the treatment effects of B and D is statistically non-significant, while the positive correlation between the predicted treatment effects of C and D is statistically significant. In other words, young men and women value the altruistic message rather than the probability message when the compatibility notice presents both two messages simultaneously. This is the reason why experimental arm D, which includes probability messages, has no effect on young men.

Interestingly, when we add the two messages simultaneously to the compatibility notice, men and women over 30 years of age give more weight to the probability message than to the altruistic message. For men and women in their late 40s and older, the altruistic message (experimental arm C) is, on average, more effective than the probability message (experimental arm B). However, because they place more importance on the probability message in treatment D, the treatment effect for this arm is close to zero on average.

## 3.3 Response Speed to Notification

Since our data include the number of reply days, we examine the effect over time and estimate the effect on the probability , which is a probability of replying within days, in a linear probability model. In this model, the outcome variable is a dummy taking one if the potential donor replied to the compatibility notice within days. If the value of is sufficiently large, the effect on is consistent with the effect on replies shown in the previous subsection. Also, as in the previous subsection, we decompose the effect on into two effects in terms of the donor’s intention. For positive intention, the outcome is a dummy variable taking one if the potential donor replied within days and indicated the intention to donate. For negative intention, the outcome variable is a dummy variable that takes one if the potential donor replied within days and did not indicate the intention to donate.

|  |
| --- |
| Figure 4: Effect on Reply within Specific Days after Sending Notification among Males Less than 30. Notes: These plots show the average effect (and associated 95% confidential interval) on each outcome. We use clustered standard errors with the CR2 adjustment (cluster unit is experimental week). We control number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, number of hospitals with BM collection per 10 square kilometers, month dummies, and week dummies. |

Since the effects of our intervention messages are heterogeneous, we also focus on heterogeneity in this subsection. [Figure 4](#fig-young-male-speed) restricts the sample to men under 30. The JMDP specifies seven days as a suggested reply time in the compatibility notice. In the control group of this subsample, 21% of potential donors replied within seven days. The earlier the reply date, the more likely it is that those who replied to the notification are willing to donate. For example, 68% of those who responded within four days wanted to donate, and 51% of those who responded within 28 days had the willingness to donate. In other words, the slower a potential donor replies, the less likely a potential donor is to be willing to donate.

In this gender-age group, we have already found that experimental arm B, which added only the probability message, increases the reply rate with positive intention. This effect is primarily due to the behavior of potential donors who take longer than 10 days to reply. Experimental arm B does not have statistically significant effects on the reply rate within 10 days and the reply rate with positive intentions. Experimental arm B has statistically significant effects on the reply rate within 14–28 days and the reply rate with positive intentions. In addition, a notable result is that experimental arm B reduces the probability of replying with negative intentions within 7–10 days.

|  |
| --- |
| Figure 5: Effect on Reply within Specific Days after Sending Notification among Females Less than 30. Notes: These plots show the average effect (and associated 95% confidential interval) on each outcome. We use clustered standard errors with the CR2 adjustment (cluster unit is experimental week). We control number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, number of hospitals with BM collection per 10 square kilometers, month dummies, and week dummies. |

[Figure 5](#fig-young-female-speed) restricts the sample to women under 30. We already found that experimental arm C, adding only the patient information, has no statistically significant effect on the overall reply rate and intention. However, when focusing on short-term behavior, experimental arm C statistically significantly increases the reply rate within four days and the reply rate within 10–28 days by 5 percentage points. The positive effect on the reply rate within 4 days results from equal increases in positive and negative intentions. However, the positive effect on the reply rate within 10–29 days is mainly due to an increase in reply rate with a negative intention. However, the statistical significance of the effect of arm C on intentions is weak. Surprisingly, experimental group B, which is effective for young men, reduced the positive intention and the reply rate within 7–14 days.

## 3.4 Effects on the Coordination Process

Finally, we examine the impact on each step in the coordination process after replying to the compatibility notice. As explained in [Section 3.4](#sec-process), the coordination process has four stages: confirmatory typing (CT), candidate selection, final consent, and donation. We use as an outcome variable a dummy variable taking one if a potential donor has reached each step. Also, as in the analysis in [Section 3.1](#sec-reply), we exclude samples that are likely to have been terminated independently of the potential donor’s intentions. When estimating the effect on CT, we exclude cases of interruptions due to the patient. Since the patient’s physician can select the healthiest potential donor as a candidate at the candidate selection stage, when estimating the effect on the process after candidate selection, we exclude cases of interruptions due to the donor’s health reasons or patient. However, we must be careful in interpreting the effect on the process after candidate selection, since sample exclusion may not completely remove the physician’s decision. In the control group, 23% of potential donors took CT, 8% of potential donors became candidates, and 6% of potential donors ultimately donated.

**Table** **:** Linear Probability Model of Coordination Process.

|  | CT | | Candidate | | Consent | | Donation | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Constant | 0.2350\*\*\* | 0.0797\*\*\* | 0.0779\*\*\* | 0.0079 | 0.0687\*\*\* | -0.0011 | 0.0574\*\*\* | 0.0055 |
|  | (0.0126) | (0.0174) | (0.0070) | (0.0117) | (0.0068) | (0.0103) | (0.0056) | (0.0081) |
| Treatment B | 0.0325\* | 0.0326\*\* | 0.0051 | 0.0033 | 0.0059 | 0.0037 | 0.0040 | 0.0020 |
|  | (0.0152) | (0.0100) | (0.0082) | (0.0069) | (0.0080) | (0.0059) | (0.0065) | (0.0064) |
| Treatment C | 0.0146 | 0.0136 | 0.0010 | -0.0023 | 0.0024 | -0.0011 | 0.0016 | -0.0017 |
|  | (0.0177) | (0.0120) | (0.0110) | (0.0078) | (0.0093) | (0.0061) | (0.0083) | (0.0065) |
| Treatment D | 0.0260 | 0.0299\*\* | 0.0084 | 0.0087 | 0.0099 | 0.0098 | 0.0030 | 0.0029 |
|  | (0.0174) | (0.0105) | (0.0100) | (0.0064) | (0.0111) | (0.0067) | (0.0103) | (0.0074) |
| Control average | 0.2350 | 0.2350 | 0.0779 | 0.0779 | 0.0687 | 0.0687 | 0.0574 | 0.0574 |
| Covariates |  | X |  | X |  | X |  | X |
| Num.Obs. | 10435 | 10435 | 8587 | 8587 | 8558 | 8558 | 8441 | 8441 |
| Notes: \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01. The clustered standard errors with the CR2 adjustment are reported in parenetheses (cluster unit is experimental weeks). Covariates are gender, squared polynomial of (demeaned) age, number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, number of hospitals with BM collection per 10 square kilometers, month dummies, and week dummies. | | | | | | | | |

**?(caption)**

**?@tbl-tbl-full-coordination** is the estimated result of the linear probability model. The results show that experimental arm B, adding only the probability message, increases the probability of reaching the CT by 3 percentage points or 13%, which is statistically significant. Controlling for covariates, experimental arm D, adding both the probability message and the patient message, also increases the probability of taking the CT by 3 percentage points or 13%, which is statistically significant. The results so far are also obtained by estimating with a logit model **?@tbl-full-coordination-logit**. Experimental arm D also increases the probability of candidate selection and reaching a final consent, which is weakly statistically significant and is not statistically significant in the estimation with the logit model.

|  |
| --- |
| Figure 6: Effect on Coordination by Gender and Age Group. Note: These plots show the average effect (and associated 95% confidential interval) on each outcome by gender and age group. We use clustered standard errors with the CR2 adjustment (cluster unit is experimental week). We control number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, number of hospitals with BM collection per 10 square kilometers, month dummies, and week dummies. |

[Figure 6](#fig-subsample-coordination) is a coefficient plot of the subsample analysis divided by gender and age. The results show that experimental arm B increases the probability of taking the CT for males under age 30. Surprisingly, this experimental arm also statistically significantly increases the donation probability for this gender-age group. However, we have not been able to distinguish whether this is an effect of potential donor intent or simply reflects the tendency of patient physicians to select younger men with better transplant records. Furthermore, when computing p-values using the wild cluster bootstrap method, we cannot reject the null hypothesis for a positive effect of experimental arm B. Thus, these results are not robust. Experimental arm B also statistically significantly increases the probability of CT for men over 30 years old. Experimental arm D statistically significantly increases the probability of CT for men and women over 30 years old.

# 4. Discussion

我々は返信の意思決定に関する意思決定モデルを構築し、情報提供介入のメカニズムを整理する。ドナー候補者は0期から期の間にいつ返信をするのかを決める。ある期の効用はとし、ここでとはそれぞれ提供の（期待）価値とコストである。また、とは時間割引因子であり、提供価値と提供コストで異なることを許容する。また、すべての期でが負となるとき、ドナー候補者は適合通知に返信しない[[9]](#footnote-67)。この節では、理論モデルから得られる示唆のみを提示し、詳細な導出を補論に示す。

|  |
| --- |
| Figure 7: Simulation of Optimal Response Speed. Note: We assume , , and . |

理論モデルの結果は以下のようにまとめられる。第一に、提供価値が低いドナー候補者は返信しないことを選択するのが最適である。返信の有無を決める提供価値の基準は時間割引因子に依存する。の値が十分に高い（）、すなわち「いつでも助けられる」と考える人について提供価値の基準はの値が低い（）、すなわち「今でないと助けられない」と考える人よりも低い。第二に、「今でないと助けられない」と考える人の最適行動は返信しないか、すぐに返信するのどちらかである。第三に、「いつでも助けられる」と考える人の最適な返信時期は提供価値・コスト・二つの時間割引因子に依存する。とくに、提供価値の増加は返信時期を早める。また、提供価値が十分に高い（）とき、提供価値の時間割引因子の上昇は返信時期を遅くする。[Figure 7](#fig-theory-simulation) は、、としたときの最適な返信時期を示したヒートマップであり、以上の三つの結果を視覚的に示している。

我々の介入は情報提供を通じて提供価値（）と提供価値の時間割引因子（）を変える。確率メッセージは幹細胞提供の公共財の性質によって生じるただ乗り行動ができないことを強調することによって提供価値を高めている。また、利他メッセージは患者の半数しか移植を受けられないことを強調することによって提供価値を高めている。さらに、利他メッセージは早い返信が患者の生存確率に寄与することを強調することで提供価値の時間割引因子を下げている。すなわち、利他メッセージは提供価値を高めると同時に、その価値の経時的な減衰の程度も高めている。

したがって、意思決定モデルは確率メッセージが返信行動を促進するだけでなく、その時期も早めていることを示唆している。確率メッセージは若年男性の返信率と返信時の提供意向を高めている。したがって、若年男性の多くは「今でないと助けられない」考えており（）、その価値を小さく見積もっている（）と考えられる。もしそうであれば、理論モデルより、確率メッセージは若年男性が未返信から即時返信へと行動を変えていると考えられる。事実、限界的であるもの、確率メッセージは若年男性について4日以内の返信率を高めている。

一方で、利他メッセージは提供価値を高めるとともに、提供価値の時間割引因子を下げているので、返信行動を促す効果と返信行動を阻害する効果が混在している。利他メッセージは若年女性の返信行動を統計的に有意な影響を与えていない。したがって、若年女性の多くは提供価値を低く見積もっている可能性がある。もしそうであれば、利他メッセージは割引因子減少によって「いつでも助けられる」と考えている若年女性の返信行動を阻害している一方で、提供価値増加によって「今でないと助けられない」と考えている若年女性の即時返信を促進している。したがって、総合的に、利他メッセージは若年女性の返信行動に影響を与えていないのだろう。また、利他メッセージは若年女性の4日以内の返信を促進しており、平均的な返信日数も縮めている。これは利他メッセージが若年女性の即時返信を促している可能性の証左となる。

# 5. Conclusion

Our field experiment revealed that adding information to the compatibility notice, which informs that becoming a potential donor, affects the donor’s willingness to donate. In particular, the information that there were fewer potential donors per patient increases the willingness of young men to donate. Examining why the effects were heterogeneous is a topic for future work.

# Appendix

**Table** **:** Logit Model of Reply and Intention

|  | | | Intention | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | Reply | | Positive | | Negative | |
|  | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment B | 1.112 | 1.149 | 1.093 | 1.092 | 0.950 | 0.969 |
|  | [0.940, 1.315] | [0.965, 1.367] | [0.982, 1.216] | [0.977, 1.220] | [0.848, 1.063] | [0.863, 1.089] |
| Treatment C | 0.947 | 1.023 | 0.979 | 0.988 | 0.998 | 1.026 |
|  | [0.801, 1.120] | [0.856, 1.222] | [0.878, 1.092] | [0.879, 1.109] | [0.889, 1.120] | [0.909, 1.160] |
| Treatment D | 1.060 | 1.061 | 1.019 | 1.024 | 1.005 | 1.002 |
|  | [0.893, 1.257] | [0.887, 1.268] | [0.914, 1.136] | [0.914, 1.148] | [0.896, 1.128] | [0.889, 1.129] |
| Covariates |  | X |  | X |  | X |
| Num.Obs. | 10985 | 10985 | 10985 | 10985 | 10985 | 10985 |
| Log.Lik. | -3884.517 | -3712.289 | -7534.803 | -7364.638 | -6945.023 | -6869.968 |
| Notes: We show odds ratios and associated 95 percent confidential intervals. Covariates are gender, squared polynomial of (demeaned) age, number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, number of hospitals with BM collection per 10 square kilometers, month dummies, and week dummies. | | | | | | |

**?(caption)**

**Table** **:** Logit Model of Coordination

|  | CT | | Candidate | | Consent | | Donation | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Treatment B | 1.189 | 1.197 | 1.071 | 1.042 | 1.093 | 1.059 | 1.074 | 1.036 |
|  | [1.049, 1.348] | [1.050, 1.363] | [0.860, 1.334] | [0.829, 1.310] | [0.867, 1.378] | [0.831, 1.348] | [0.833, 1.385] | [0.795, 1.351] |
| Treatment C | 1.083 | 1.081 | 1.014 | 0.967 | 1.038 | 0.988 | 1.030 | 0.976 |
|  | [0.951, 1.233] | [0.942, 1.242] | [0.808, 1.272] | [0.760, 1.229] | [0.816, 1.319] | [0.766, 1.274] | [0.792, 1.339] | [0.739, 1.288] |
| Treatment D | 1.150 | 1.179 | 1.119 | 1.123 | 1.157 | 1.156 | 1.055 | 1.053 |
|  | [1.011, 1.308] | [1.031, 1.349] | [0.895, 1.397] | [0.892, 1.415] | [0.915, 1.463] | [0.907, 1.475] | [0.812, 1.371] | [0.804, 1.380] |
| Covariates |  | X |  | X |  | X |  | X |
| Num.Obs. | 10435 | 10435 | 8587 | 8587 | 8558 | 8558 | 8441 | 8441 |
| Log.Lik. | -5909.753 | -5764.480 | -2427.295 | -2349.439 | -2243.901 | -2168.120 | -1906.131 | -1851.371 |
| Notes: We show odds ratios and associated 95 percent confidential intervals. Covariates are gender, squared polynomial of (demeaned) age, number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, number of hospitals with BM collection per 10 square kilometers, month dummies, and week dummies. | | | | | | | | |

**?(caption)**

**Table** **:** Linear Probability Model of Reply and Intentions for Males Less Then 30 Years

|  | | | Intention | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | Reply | | Positive | | Negative | |
|  | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment B | 0.0661\*\* | 0.0656\* | 0.0935\*\* | 0.0936\*\* | -0.0274 | -0.0280 |
|  | (0.0255) | (0.0326) | (0.0372) | (0.0394) | (0.0196) | (0.0197) |
| Treatment C | 0.0197 | 0.0259 | 0.0218 | 0.0252 | -0.0021 | 0.0007 |
|  | (0.0281) | (0.0318) | (0.0340) | (0.0280) | (0.0249) | (0.0201) |
| Treatment D | 0.0620\* | 0.0653 | 0.0198 | 0.0282 | 0.0422 | 0.0371 |
|  | (0.0280) | (0.0357) | (0.0470) | (0.0359) | (0.0322) | (0.0233) |
| Control average | 0.7821 | 0.7821 | 0.4216 | 0.4216 | 0.3605 | 0.3605 |
| Covariates |  | X |  | X |  | X |
| Bootstrap p-value |  |  |  |  |  |  |
| B = 0 | 0.0380 | 0.0820 | 0.0510 | 0.0440 | 0.2720 | 0.1840 |
| D = 0 | 0.0510 | 0.1040 | 0.6880 | 0.4350 | 0.2240 | 0.1740 |
| Num.Obs. | 1556 | 1556 | 1556 | 1556 | 1556 | 1556 |
| Notes: \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01. The clustered standard errors with the CR2 adjustment are reported in parenetheses (cluster unit is experimental weeks). Covariates are number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, number of hospitals with BM collection per 10 square kilometers, month dummies, and week dummies. Bootstrap p-value is based on the wild cluster bootstrap-t with a null hypothesis imposed, which is proposed by Cameron et al. (2008). | | | | | | |

**?(caption)**

**Table** **:** Linear Probability Model of Coordination among Young Males

|  | CT | Candidate | Consent | Donation |
| --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) |
| Constant | -0.0556 | -0.0492 | -0.0350 | -0.0062 |
|  | (0.0555) | (0.0433) | (0.0421) | (0.0357) |
| Treatment B | 0.0764\* | 0.0381 | 0.0331 | 0.0426 |
|  | (0.0393) | (0.0233) | (0.0230) | (0.0234) |
| Treatment C | -0.0029 | -0.0103 | -0.0036 | 0.0061 |
|  | (0.0281) | (0.0147) | (0.0156) | (0.0165) |
| Treatment D | -0.0207 | 0.0126 | 0.0211 | 0.0215 |
|  | (0.0278) | (0.0182) | (0.0187) | (0.0172) |
| Num.Obs. | 1499 | 1275 | 1269 | 1253 |
| Covariates | X | X | X | X |
| Bootstrap p-value |  |  |  |  |
| B = 0 | 0.089 | 0.145 | 0.217 | 0.120 |
| Notes: \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01. The clustered standard errors with the CR2 adjustment are reported in parenetheses (cluster unit is experimental weeks). Covariates are number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, number of hospitals with BM collection per 10 square kilometers, month dummies, and week dummies. Bootstrap p-value is based on the wild cluster bootstrap-t with a null hypothesis imposed, which is proposed by Cameron et al. (2008). | | | | |

**?(caption)**

**Table** **:** Conditional Average Treatment Effect Estimated by RCF.

|  | Females | | | Males | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | Estimate | S.E. | P-value | Estimate | S.E. | P-value |
| Age: Less than 30 | | | | | | |
| B | 0.001 | 0.048 | 0.989 | 0.124 | 0.040 | 0.002 |
| C | 0.046 | 0.048 | 0.343 | 0.041 | 0.041 | 0.318 |
| D | -0.051 | 0.049 | 0.291 | 0.049 | 0.040 | 0.226 |
| Age: More than or equal to 30 | | | | | | |
| B | -0.003 | 0.029 | 0.916 | 0.020 | 0.021 | 0.350 |
| C | -0.010 | 0.030 | 0.731 | -0.008 | 0.023 | 0.721 |
| D | 0.013 | 0.029 | 0.646 | 0.024 | 0.022 | 0.283 |
| Notes: See Athey and Wager (2019) for estimation method of conditional average treatment effect (CATE). Since these estimates are asymptotically normal, we calculate z-score under the null hypothesis that CATE is zero, and obtain p-value. | | | | | | |

**?(caption)**

**?(caption)**

**?(caption)**

**?(caption)**

# References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge. 2023. “When Should You Adjust Standard Errors for Clustering?” *The Quarterly Journal of Economics* 138 (1): 1–35. <https://doi.org/10.1093/qje/qjac038>.

Athey, Susan, and Guido Imbens. 2016. “Recursive Partitioning for Heterogeneous Causal Effects.” *Proceedings of the National Academy of Sciences* 113 (27): 7353–60. <https://doi.org/10.1073/pnas.1510489113>.

Athey, Susan, Julie Tibshirani, and Stefan Wager. 2019. “Generalized Random Forests.” *The Annals of Statistics* 47 (2). <https://doi.org/10.1214/18-AOS1709>.

Athey, Susan, and Stefan Wager. 2019. “Estimating Treatment Effects with Causal Forests: An Application.” *Observational Studies* 5 (2): 37–51. <https://doi.org/10.1353/obs.2019.0001>.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller. 2008. “BOOTSTRAP-BASED IMPROVEMENTS FOR INFERENCE WITH CLUSTERED ERRORS.” *The Review of Economics and Statistics* 90 (3): 414–27.

Davis, Jonathan M. V., and Sara B. Heller. 2017. “Using Causal Forests to Predict Treatment Heterogeneity: An Application to Summer Jobs.” *American Economic Review* 107 (5): 546–50. <https://doi.org/10.1257/aer.p20171000>.

Hirakawa, Tsuneaki, Saiko Kurosawa, Kinuko Tajima, Yusuke Yamazaki, Nami Ikeda, Hiroto Kojima, Hidenori Tanaka, et al. 2018. “Japan Marrow Donor Program and its coordinating process: current situations.” *The Japanese Journal of Clinical Hematology* 59 (2): 153–60.

Imbens, Guido W., and Michal Kolesár. 2016. “Robust Standard Errors in Small Samples: Some Practical Advice.” *Review of Economics and Statistics* 98 (4): 701–12. <https://doi.org/10.1162/REST_a_00552>.

Takanashi, Minoko. 2016. “Marrow donor registration and cord blood banking: current issues.” *The Japanese Journal of Clinical Hematology* 57 (3): 271–77. <https://doi.org/10.11406/rinketsu.57.271>.

Wager, Stefan, and Susan Athey. 2018. “Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests.” *Journal of the American Statistical Association* 113 (523): 1228–42. <https://doi.org/10.1080/01621459.2017.1319839>.

1. 近年、ハプロ移植と呼ばれる、半合致したHLAを持つ近親者間での移植が広がりつつある。加えて、母子をつなぐさい帯や胎盤に含まれる血液細胞の移植（さい帯血移植）も同様に人気になっている。骨髄移植と異なり、HLAが完全に合致していなくても、さい帯血移植は実施できる。 [↑](#footnote-ref-20)
2. これは珍しいHLAの型を持つ新しいドナーの確率が低いからである。 [↑](#footnote-ref-21)
3. JMDP also sends the potential donor an SMS message informing that the JMDP sends the compatibility notice. [↑](#footnote-ref-23)
4. When having made intervention messages, we have taken appropriate care not to put undue pressure on potential donors. First, we avoid messages that sound like a plea. Second, we use only publicly available information from the JMDP. Third, we explain the risks of transplantation as usual. [↑](#footnote-ref-29)
5. one person resided abroad. There were also 104 persons for whom coordination was in progress when we received the data. The proportion of people with ongoing coordination is well balanced across experimental arms (F-value, p-value = ). [↑](#footnote-ref-32)
6. If the proportion of clusters sampled from the population is large and the variance of the within-cluster average treatment effect is large, the true cluster standard errors can be extremely conservative (Abadie et al. 2023). The CR2 adjustment can remove bias in the estimation of cluster standard errors but cannot correct for bias in the true cluster standard errors. Therefore, it should be interpreted with caution. [↑](#footnote-ref-34)
7. Instead of cluster standard errors with the CR2 adjustment, we compute p-values by the wild cluster bootstrap method (Cameron, Gelbach, and Miller 2008) and find similar results. See **?@tbl-wildbs-youngmen-reply**. [↑](#footnote-ref-42)
8. See Wager and Athey (2018) and Athey, Tibshirani, and Wager (2019) for technical and implementation details. [↑](#footnote-ref-44)
9. 単純化のために、この意思決定モデルは返信する人が必ず提供意向を持っていることを暗に仮定している。 [↑](#footnote-ref-67)