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 March 23, 2023

# 1. Introduction

Allogenic stem cell transplantation is one of the treatments with the lowest recurrence rate for blood diseases such as leukemia. In this treatment method, (1) diseased cells and healthy hematopoietic stem cells are simultaneously killed by anticancer drugs and radiation, and (2) healthy hematopoietic stem cells donated by another person are transplanted. One requirement for bone marrow transplantation (hematopoietic stem cell transplantation) is that the donor’s white blood cell type, called HLA, matches the patient’s HLA.[[1]](#footnote-20) While the matching probability between two random people is less than 1%, the most likely HLA match is between siblings, with a probability of approximately 30%. The matching probability between parents and children is very low. If there is no matched donor among blood relatives, the patient must seek a donor from an unrelated person. In Japan, the patient generally seeks unrelated donors through the Japan Marrow Donor Program (JMDP). However, coordination by the JMDP results in a long transplant process, and only 60% of registered patients receive a transplant (Hirakawa et al. 2018).

There are two types of policies for the donor pool to increase patient survival. The first type is to increase the number of potential donors, which increases the matching probability. Compared to when the patient matches fewer than four donors, when the patient matches more than 200 donors, the patient’s transplant rate increases from 45% to 74% (Hirakawa et al. 2018). However, while the number of potential donors for JMDP has roughly doubled between 2000 and 2015, the probability of first-time matching has only increased by about 5% (Takanashi 2016).[[2]](#footnote-21) In other words, increasing the pool size would be ineffective because the marginal benefit of increasing the number of potential donors is small.

The second type is to increase the proportion of people in the donor pool who are less likely to refuse to donate. This type of measure contributes to the quality of the donor pool. Hirakawa et al. (2018) shows that many coordinations are interrupted before their first step, the confirmatory typing, due to the donor side reasons (see the [Section 2.1](#sec-background) section for the coordination flow). Therefore, creating a pool with many potential donors who are willing to donate would increase the transplant rate. The marginal benefit of measures that improve the quality of the donor pool should be higher than that of measures that increase the size of the donor pool.

Then, this study examines the effectiveness of information provision as one policy improving the quality of the donor pool. When a person registered with the JMDP becomes a potential donor for a patient, the potential donor receives a compatibility notice from the JMDP. The potential donors who respond to the notification by indicating a willingness to donate then take coordination for transplantation. We added a message to the compatibility notice based on information published by the JMDP and conducted a field experiment from September 2021 to February 2022 to test the effect of the added message. We asked the JMDP to provide data on the coordination process (including whether or not they responded to the notification and their willingness to donate) for those who received the compatibility notice during the experimental period and analyze the effect of providing information.

There are two intervention messages. One is information that there are few potential donors per patient registered in the JMDP. Stem cell donation through JMDP is a public good since multiple potential donors per patient take coordination simultaneously. Thus, potential donors have the incentive to free-ride. The first message prevents potential donors from free-riding behavior caused by the over-estimated expectation about the number of other ongoing potential donors. Another message is that only about half of the patients registered with the JMDP receive transplants. This information aims to increase the value of the altruistic behavior of stem cell donation. Our field experiment shows that information about the small number of potential donors per patient encourages young males to reply to the compatibility notice and indicate their willingness to donate.

# 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 |

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-31).

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 | | | |  |
| --- | --- | --- | --- | --- | --- |
| vars | 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 | 0.362 |
| Age | 38.376 | 38.121 | 37.448 | 37.978 | 0.067 |
| Number of past coordinations | 1.609 | 1.589 | 1.625 | 1.563 | 0.451 |
| Number of listed hospitals | 0.476 | 0.490 | 0.487 | 0.485 | 0.798 |
| Number of hospitals listed with PBSC collection | 0.162 | 0.167 | 0.166 | 0.164 | 0.804 |
| Number of hospitals listed with BM collection | 0.246 | 0.256 | 0.254 | 0.251 | 0.669 |

**?@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 = ).

Since potential donors cannot choose the experimental group, i.e., the experimental group is exogenous, a simple difference in means can identify the average treatment effect. However, this leads to bias because age and the number of weeks and months of the assignment are not perfectly balanced across experimental groups. Therefore, in addition to the difference-in-mean test, we estimate the following linear probability model for individual who received the compatibility notice in a week of month .

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.

# 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. 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 response rate is % in the control arm. 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) |
| Constant | 0.8835\*\*\* | 0.9510\*\*\* | 0.5533\*\*\* | 0.6463\*\*\* | 0.3303\*\*\* | 0.3047\*\*\* |
|  | (0.0065) | (0.0190) | (0.0125) | (0.0277) | (0.0089) | (0.0210) |
| Treatment B | 0.0105 | 0.0130\*\* | 0.0218 | 0.0203\* | -0.0113 | -0.0073 |
|  | (0.0075) | (0.0061) | (0.0135) | (0.0123) | (0.0094) | (0.0075) |
| Treatment C | -0.0057 | 0.0006 | -0.0053 | -0.0052 | -0.0004 | 0.0057 |
|  | (0.0081) | (0.0053) | (0.0151) | (0.0111) | (0.0150) | (0.0093) |
| Treatment D | 0.0058 | 0.0063 | 0.0046 | 0.0063 | 0.0012 | 0.0000 |
|  | (0.0074) | (0.0051) | (0.0166) | (0.0097) | (0.0138) | (0.0065) |
| 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. Standard errors clustered by experimental weeks are reported in parentheses. 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, prefecture dummies, month dummies, and week dummies. | | | | | | |

**?@tbl-full-reply** are 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. Column (1) shows that the experimental arm B, which sends the probability message, increases the reply rate by 1 percentage point or %. Controlling for individual attributes and other factors (column (2)) does not significantly change the size of the effect, but it is statistically significant at the 5% level. This effect stems from an increase in replies with positive intentions. Columns (3) and (5) show that experimental arm B increases the replay rate with positive intentions by 2 percentage points or % while it decreases the reply rate with negative intentions by 1 percentage point or %. Their statistical significance is weak. In addition, experimental arms C and D have smaller effect sizes than experimental arm B and are statistically non-significant. We estimated a logit model as a robustness analysis ((**full-reply-logit?**)). The results are consistent with the above table for the direction of the effect but are statistically non-significant. In sum, our intervention has an insignificant impact on replies overall.

|  |
| --- |
| 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 cluster standard errors by experimental weeks. 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, prefecture dummies, 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. The results show that experimental groups B and D, which add at least the probability message, increase the reply rate for young males by 6 percentage points or % (the baseline reply rate is %).

The effect of experimental arm B with the addition of only probability messages results from an increase in the reply rate with positive intention. Among males under 30, experimental arm B increases the reply rate with positive intention by 10 percentage points or % (baseline rate of %) but decreases the reply rate with negative intention by 3 percentage points or % (baseline rate of %). These are statistically significant effects. On the other hand, experimental group D, which adds the probability and patient messages, increases the reply rate with positive and negative intentions by the same amount. In particular, experimental arm D increases the reply rate with negative intention by 3 percentage points or %, which is statistically significant.

Among the other gender and age groups (especially men and women over 30), our intervention has no significant effect on replies or intentions. However, among women under 30, experimental group D reduces the reply rate with positive intention by percentage points or %. Thus, the message that potential donors per patient are few increases young men’s willingness to donate and encourages them to reply to the compatibility notice. 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.

## 3.2 Robustness to Heterogenous Effect

We perform robust checks of the heterogeneity of message effects. First, we compute p-values using the wild cluster bootstrap method (Cameron, Gelbach, and Miller 2008). When the number of clusters is small, the cluster standard errors over-reject the null hypothesis. To correct this, we use the wild cluster bootstrap method. **?@tbl-young-reply** in Appendix is estimation results, restricting the sample to women under 30 (columns (1)–(3)) and men under 30 (columns (4)–(6)). The results show that even with the wild bootstrap method, the probability message (experimental arm B) increases the reply rate with positive intentions among males under 30.

|  |
| --- |
| Figure 3: Distribution of Individual Treatment Effects on Intention Estimated by Random Causal Forest. |

**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 | -0.004 | 0.048 | 0.929 | 0.125 | 0.040 | 0.002 |
| C | 0.041 | 0.048 | 0.393 | 0.050 | 0.041 | 0.224 |
| D | -0.060 | 0.048 | 0.214 | 0.045 | 0.040 | 0.253 |
| Age: More than or equal to 30 | | | | | | |
| B | -0.008 | 0.029 | 0.787 | 0.021 | 0.022 | 0.323 |
| C | -0.017 | 0.030 | 0.568 | -0.009 | 0.023 | 0.696 |
| D | 0.008 | 0.029 | 0.784 | 0.023 | 0.022 | 0.311 |

We then test for heterogenous effects by random causal forest (Davis and Heller 2017; Athey and Wager 2019). This method divides the sample by combinations of covariates and estimates treatment effects within each subsample. This yields predicted values of individual treatment effects, which is a function of the covariates given by the researcher. The combination of covariates creating subsamples is determined to minimize the mean squared error of the estimated heterogeneous treatment effect. This minimization is equivalent to maximizing the variance of the estimated treatment effect (Athey and Imbens 2016).

**?@fig-int-rcf-cate** is the distribution of the predicted treatment effect of each experimental arm on replies with positive intention. Notably, 90% of men under 30 respond positively to the probability message. The random causal forest shows that the average treatment effect of experimental arm B among young males is percentage points, which is statistically significant (**?@tbl-int-rcf-cate**). The estimated treatment effect is almost consistent with the estimates obtained in [Figure 2](#fig-subsample-reply).

We also confirm that our intervention has a statistically non-significant effect for the other gender-age groups. This result is because there exist roughly equal numbers of people who react positively and negatively to the intervention. Then, what characteristics make people more likely to have a positive response? We examined the differences in average personal attributes of those who are positively affected and those who are not within each gender-age group (**?@tbl-covariate-int-rcf**). In the group of men and women over 30, those who are older and have more past coordination experience are more likely to respond positively to the intervention message. In the group of women over 30, those with more hospitals were more likely to react positively to the intervention message. On the other hand, in the group of men over 30 years old, those with fewer hospitals were more likely to respond positively to the intervention message. In the group of women under 30, younger people were more likely to respond positively to the intervention message. Moreover, those with more previous coordination experience are more likely to respond positively to experimental arms C and D, and those with more hospitals are more likely to respond positively to intervention group D.

**?(caption)**

## 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. |

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. |

[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\*\*\* | 0.1531\*\*\* | 0.0779\*\*\* | 0.0472\*\* | 0.0687\*\*\* | 0.0318 | 0.0574\*\*\* | 0.0272\* |
|  | (0.0115) | (0.0295) | (0.0065) | (0.0224) | (0.0062) | (0.0197) | (0.0052) | (0.0164) |
| Treatment B | 0.0325\*\* | 0.0351\*\*\* | 0.0051 | 0.0049 | 0.0059 | 0.0051 | 0.0040 | 0.0032 |
|  | (0.0141) | (0.0076) | (0.0076) | (0.0049) | (0.0074) | (0.0042) | (0.0060) | (0.0044) |
| Treatment C | 0.0146 | 0.0145\* | 0.0010 | -0.0016 | 0.0024 | -0.0007 | 0.0016 | -0.0011 |
|  | (0.0163) | (0.0087) | (0.0102) | (0.0058) | (0.0086) | (0.0047) | (0.0076) | (0.0050) |
| Treatment D | 0.0260 | 0.0318\*\*\* | 0.0084 | 0.0097\*\* | 0.0099 | 0.0101\* | 0.0030 | 0.0034 |
|  | (0.0161) | (0.0081) | (0.0093) | (0.0049) | (0.0102) | (0.0052) | (0.0095) | (0.0058) |
| 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. Standard errors clustered by experimental weeks are reported in parentheses. 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, prefecture dummies, month dummies, and week dummies. | | | | | | | | |

**?@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 cluster standard errors by experimental weeks. 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, prefecture dummies, 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.

**?(caption)**

# 4. 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 | 1.143 | 1.093 | 1.091 | 0.950 | 0.967 |
|  | [0.940, 1.315] | [0.958, 1.364] | [0.982, 1.216] | [0.975, 1.221] | [0.848, 1.063] | [0.860, 1.088] |
| Treatment C | 0.947 | 1.003 | 0.979 | 0.978 | 0.998 | 1.028 |
|  | [0.801, 1.120] | [0.837, 1.201] | [0.878, 1.092] | [0.869, 1.100] | [0.889, 1.120] | [0.909, 1.162] |
| Treatment D | 1.060 | 1.062 | 1.019 | 1.027 | 1.005 | 0.999 |
|  | [0.893, 1.257] | [0.886, 1.272] | [0.914, 1.136] | [0.915, 1.153] | [0.896, 1.128] | [0.886, 1.127] |
| Covariates |  | X |  | X |  | X |
| Num.Obs. | 10985 | 10985 | 10985 | 10985 | 10985 | 10985 |
| Log.Lik. | -3884.517 | -3629.432 | -7534.803 | -7254.947 | -6945.023 | -6818.753 |

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

|  | CT | | Candidate | | Consent | | Donation | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Treatment B | 1.189 | 1.219 | 1.071 | 1.075 | 1.093 | 1.090 | 1.074 | 1.069 |
|  | [1.049, 1.348] | [1.068, 1.390] | [0.860, 1.334] | [0.853, 1.354] | [0.867, 1.378] | [0.854, 1.391] | [0.833, 1.385] | [0.818, 1.396] |
| Treatment C | 1.083 | 1.090 | 1.014 | 0.985 | 1.038 | 1.006 | 1.030 | 0.998 |
|  | [0.951, 1.233] | [0.947, 1.253] | [0.808, 1.272] | [0.773, 1.256] | [0.816, 1.319] | [0.778, 1.300] | [0.792, 1.339] | [0.754, 1.321] |
| Treatment D | 1.150 | 1.195 | 1.119 | 1.148 | 1.157 | 1.171 | 1.055 | 1.071 |
|  | [1.011, 1.308] | [1.043, 1.370] | [0.895, 1.397] | [0.909, 1.449] | [0.915, 1.463] | [0.916, 1.497] | [0.812, 1.371] | [0.815, 1.406] |
| Covariates |  | X |  | X |  | X |  | X |
| Num.Obs. | 10435 | 10435 | 8587 | 8587 | 8558 | 8558 | 8441 | 8441 |
| Log.Lik. | -5909.753 | -5678.471 | -2427.295 | -2300.960 | -2243.901 | -2120.905 | -1906.131 | -1808.195 |

**Table** **:** Linear Probability Model of Reply and Intentions among Young Males and Females

|  | Young females | | | Young males | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | Reply | Positive | Negative | Reply | Positive | Negative |
|  | (1) | (2) | (3) | (4) | (5) | (6) |
| Constant | 0.9520\*\*\* | 0.5007\*\*\* | 0.4513\*\*\* | 0.8259\*\*\* | 0.4350\*\*\* | 0.3909\*\*\* |
|  | (0.0705) | (0.1284) | (0.0976) | (0.0578) | (0.0907) | (0.0837) |
| Treatment B | -0.0119 | -0.0225 | 0.0106 | 0.0609\*\*\* | 0.0979\*\*\* | -0.0369\*\* |
|  | (0.0215) | (0.0363) | (0.0336) | (0.0227) | (0.0295) | (0.0148) |
| Treatment C | 0.0183 | 0.0073 | 0.0111 | 0.0229 | 0.0233 | -0.0004 |
|  | (0.0198) | (0.0297) | (0.0299) | (0.0205) | (0.0198) | (0.0136) |
| Treatment D | -0.0344\* | -0.0616\*\* | 0.0271 | 0.0693\*\*\* | 0.0381 | 0.0312\*\* |
|  | (0.0198) | (0.0245) | (0.0271) | (0.0246) | (0.0236) | (0.0149) |
| Num.Obs. | 1127 | 1127 | 1127 | 1556 | 1556 | 1556 |
| Covariates | X | X | X | X | X | X |
| B = 0 | 0.686 | 0.632 | 0.887 | 0.128 | 0.063 | 0.135 |
| D = 0 | 0.265 | 0.132 | 0.506 | 0.087 | 0.264 | 0.179 |

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

|  | CT | Candidate | Consent | Donation |
| --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) |
| Constant | 0.0133 | 0.0084 | -0.0039 | -0.0051 |
|  | (0.0800) | (0.0983) | (0.0846) | (0.0737) |
| Treatment B | 0.0828\*\* | 0.0412\*\* | 0.0382\*\* | 0.0462\*\*\* |
|  | (0.0336) | (0.0190) | (0.0159) | (0.0158) |
| Treatment C | -0.0048 | -0.0037 | 0.0026 | 0.0110 |
|  | (0.0236) | (0.0129) | (0.0116) | (0.0113) |
| Treatment D | -0.0154 | 0.0151 | 0.0222\* | 0.0236\*\* |
|  | (0.0196) | (0.0136) | (0.0135) | (0.0118) |
| Num.Obs. | 1499 | 1275 | 1269 | 1253 |
| Covariates | X | X | X | X |
| B = 0 | 0.162 | 0.326 | 0.355 | 0.174 |

**Table** **:** Balance Test of Covariates Grouped By Predicted Treatment Effect

|  | Message B | | | Message C | | | Message D | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Negative | Positive | P-value | Negative | Positive | P-value | Negative | Positive | P-value |
| Female× Age<30 | | | | | | | | | |
| age | 27.02 | 23.43 | 0.00 | 26.43 | 23.49 | 0.00 | 24.96 | 24.41 | 0.00 |
| coordinate | 1.31 | 1.24 | 0.06 | 1.21 | 1.31 | 0.01 | 1.23 | 1.34 | 0.00 |
| hospital\_per\_area | 0.58 | 0.57 | 0.92 | 0.50 | 0.63 | 0.00 | 0.51 | 0.69 | 0.00 |
| PB\_per\_area | 0.19 | 0.21 | 0.16 | 0.18 | 0.22 | 0.02 | 0.17 | 0.26 | 0.00 |
| BM\_per\_area | 0.30 | 0.31 | 0.67 | 0.26 | 0.34 | 0.00 | 0.25 | 0.41 | 0.00 |
| Female× 30≤Age | | | | | | | | | |
| age | 41.03 | 42.94 | 0.00 | 39.61 | 45.12 | 0.00 | 40.18 | 43.39 | 0.00 |
| coordinate | 1.53 | 1.52 | 0.90 | 1.51 | 1.55 | 0.23 | 1.41 | 1.61 | 0.00 |
| hospital\_per\_area | 0.45 | 0.53 | 0.00 | 0.48 | 0.50 | 0.23 | 0.46 | 0.51 | 0.03 |
| PB\_per\_area | 0.15 | 0.18 | 0.00 | 0.15 | 0.19 | 0.00 | 0.16 | 0.17 | 0.16 |
| BM\_per\_area | 0.23 | 0.28 | 0.00 | 0.25 | 0.27 | 0.07 | 0.25 | 0.26 | 0.19 |
| Male× Age<30 | | | | | | | | | |
| age | 27.92 | 24.09 | 0.00 | 25.77 | 23.86 | 0.00 | 25.01 | 24.08 | 0.00 |
| coordinate | 1.51 | 1.30 | 0.00 | 1.41 | 1.28 | 0.00 | 1.23 | 1.40 | 0.00 |
| hospital\_per\_area | 0.53 | 0.52 | 0.93 | 0.45 | 0.56 | 0.00 | 0.45 | 0.58 | 0.00 |
| PB\_per\_area | 0.17 | 0.18 | 0.78 | 0.15 | 0.19 | 0.00 | 0.14 | 0.21 | 0.00 |
| BM\_per\_area | 0.29 | 0.27 | 0.60 | 0.22 | 0.29 | 0.00 | 0.21 | 0.32 | 0.00 |
| Male× 30≤Age | | | | | | | | | |
| age | 42.34 | 42.53 | 0.33 | 40.59 | 44.92 | 0.00 | 41.10 | 43.30 | 0.00 |
| coordinate | 1.87 | 1.74 | 0.00 | 1.75 | 1.85 | 0.00 | 1.63 | 1.89 | 0.00 |
| hospital\_per\_area | 0.39 | 0.49 | 0.00 | 0.58 | 0.29 | 0.00 | 0.36 | 0.51 | 0.00 |
| PB\_per\_area | 0.13 | 0.16 | 0.00 | 0.19 | 0.10 | 0.00 | 0.12 | 0.17 | 0.00 |
| BM\_per\_area | 0.20 | 0.25 | 0.00 | 0.31 | 0.14 | 0.00 | 0.18 | 0.26 | 0.00 |

**?(caption)**

# References

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, 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.

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>.

1. Recently, transplants between blood relatives with semi-matched HLA, called haploidentical transplants, have become widespread. In addition, the transplantation of blood cells contained in the umbilical cord and placenta connecting mother and child (cord blood transplant) has also become popular. Unlike bone marrow transplants, cord blood transplants can be performed even without the requirement of a perfect HLA match. [↑](#footnote-ref-20)
2. This is because the probability of a new member with a rare HLA type is low. [↑](#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-31)