

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

Hiroki Kato<sup>† a</sup>, Fumio Ohtake<sup>b</sup>, Saiko Kurosawa<sup>c</sup>, Kazuhiro Yoshiuchi<sup>d</sup>, and Takahiro  
Fukuda<sup>e</sup>

<sup>a</sup>*Graduate School of Economics, Osaka University*

<sup>b</sup>*Center for Infectious Disease Education and Research (CiDER), Osaka University*

<sup>c</sup>*Department of Oncology, Ina Central Hospital*

<sup>d</sup>*Graduate School of Medicine, Tokyo University*

<sup>e</sup>*Department of Hematopoietic Stem Cell Transplantation, National Cancer Center Hospital*

Last updated on January 14, 2024

## Abstract

Only half of the patients registered with the Japan Marrow Donor Program receive allogeneic hematopoietic stem cell transplantation because much of the transplant coordination is interrupted due to the reluctance of registered donors to donate. In collaboration with the Japan Marrow Donor Program, we conducted a field experiment to test an information provision intervention for increasing registered donors' willingness to donate. We found that information about the low number of potential donor matches per patient increased the willingness to donate by 25% among men in their 20s. We also found that knowing that early coordination increases a patient's transplant rate encourages women in their 20s to respond early. These results suggest that providing information affects only certain genders and ages and encourages behavioral change among younger donors with better transplant outcomes.

*JEL classification:* D64, D90, H41, I11

---

\*In conducting this study, we would like to thank the Japan Marrow Donor Program Office for managing of the field experiment and providing us with the data. This study was conducted with the approval of the institutional review board of the Graduate School of Economics, Osaka University (approval number: R030305-2) and the Japan Marrow Donor Program (approval number: JMDP2021-04). Funding: This work was supported by the Japan Society for the Promotion of Science [grant number 20H05632 (F., Ohtake, H., Kato)] and the Ministry of Health, Labour and Welfare [grant number 19FF1001 (F. Ohtake, S., Kurosawa, T., Fukuda)].

<sup>†</sup>Present Address: 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan (Hitotsubashi Institute for Advanced Study, Hitotsubashi University). E-mail address: [hkato.econ@r.hit-u.ac.jp](mailto:hkato.econ@r.hit-u.ac.jp)

*Keywords:* Japan Marrow Donor Program, field experiment, text message, information provision, free-riding, present bias

## 1 Introduction

Allogeneic hematopoietic stem cell transplantation (HSCT) is one of the treatments with the lowest relapse rates for leukemia and other blood diseases. In this treatment, (1) anticancer drugs and radiation simultaneously kill tumor cells and healthy hematopoietic stem cells, and (2) healthy hematopoietic stem cells donated by others are transplanted. Transplantation requires that the donor's white blood cell type, called HLA, match the patient's HLA.<sup>1</sup> Whereas the probability of a match between two randomly selected individuals is less than 1%, the probability of a match between siblings is the highest at approximately 30%. The probability of a match between parents and their children is also low.

If there is no match among relatives, patients must seek a nonrelative donor. In Japan, patients usually seek nonrelative donors through the Japan Marrow Donor Program (JMDP). However, coordination through the JMDP takes a long time, and only 60% of registered patients receive transplants (Hirakawa et al., 2018). Therefore, it is important to shorten the time to transplantation and increase the transplantation rate in registered patients.

There are two types of donor pool policies that increase transplantation rates. The first includes policies that increase the number of potential donors to increase the probability of a match. Compared with patients matched with fewer than four donors, the transplantation rate increased from 45% to 74% for patients matched with more than 200 donors (Hirakawa et al., 2018). However, according to Takanashi (2016), the probability of a first-time match increased by only 5%, even though the number of potential donors nearly doubled between 2000 and 2015.<sup>2</sup> The marginal benefit of increasing the number of potential donors is small.

The second intervention increases the proportion of potential donors willing to donate and improves the quality of the donor pool. Hirakawa et al. (2018) found that many coordinations (73% of those conducted in 2004–2013) were interrupted before the first process of coordination, confirmatory typing, for donor-related reasons (including poor health). Younger donors are less likely

---

<sup>1</sup>In recent years, transplantation between close relatives with semi-matched HLA, known as haploidentical stem cell transplantation, has become more common. In addition, transplantation of blood cells from the umbilical cord or placenta that connects mother and child (cord blood transplantation) has also increased. Unlike bone marrow transplantation, cord blood transplantation can be performed even if the HLA is not a perfect match. In Japan, bone marrow (or peripheral blood) transplants between unrelated individuals (the focus of this study), will account for 20% of all transplants performed in FY2021 (The Japanese Data Center for Hematopoietic Cell Transplantation/The Japanese Society for Transplantation and Cellular Therapy, 2022).

<sup>2</sup>This is because the probability of a new donor with a rare HLA type is low.

to interrupt coordination for health reasons and more likely to interrupt coordination for other personal reasons, such as lack of motivation.<sup>3</sup> This is an issue not only in the JMDP but also in marrow donor programs in other countries (Haylock et al., 2022). Given the better transplant outcomes for younger men than for other genders and ages (for example, Kollman et al., 2016), interventions to increase the willingness to donate (especially among younger men) are more important than increasing the donor pool size.

Therefore, this study examines the effect of providing information that increases willingness to donate as one of the measures to improve the quality of the donor pool. When a potential donor registered with the JMDP is matched to a specific patient, the matched donor receives a compatibility notice from the JMDP. Matched donors who respond to the notice by indicating their willingness to donate are coordinated for transplantation. We added two new messages to the compatibility notice based on information published by the JMDP and conducted a field experiment in collaboration with the JMDP to test the effects of the additional messages.

The first message indicates that the number of HLA-compatible donors per patient is low. If there are other potential donors with the same HLA type in the pool, one's donation can be substituted for that compatible donor. In addition, multiple donors (up to 10) can be coordinated with a single patient simultaneously. Thus, transplantation through JMDP is public goods. Donors who gain utility from the patient's survival will be more reluctant to donate the more common the HLA. In other words, JMDP-mediated transplantation faces a standard "free-ride" problem (Bergstrom et al., 2009).

This problem is essentially the same as the volunteer dilemma, in which public goods are produced by the cooperative behavior of only one person. In the volunteer dilemma, the theory predicts that the probability of cooperative behavior by even one person decreases with the group size, and this hypothesis has been confirmed by laboratory experiments (Diekmann, 1985, 1986; Franzen, 1999; Davis and Heller, 2017). Additionally, an interview study of previously matched donors (Kurosawa et al., 2022) found that those with low donation intentions felt that they were "one of several donors," implying that the fact that their donation could be substituted by others discouraged them from donating. An important difference from the volunteer dilemma is that potential donors in the JMDP cannot know their own HLA type or the number of matched donors simultaneously; therefore, they cannot accurately estimate the size of the group. Consequently, potential

---

<sup>3</sup>While 15% of males men in their 20s were suspended due to donor health (history, back pain, undergoing treatment, etc.), 41% were suspended for reasons other than donor health (inability to contact, unavailability, etc.). In addition, 6% were interrupted due to lack of family consent, a prerequisite for transplantation. The remaining coordinations were either suspended due to patients or reached transplantation.

donors may overestimate the number of possible substitutes and may be reluctant to donate. The first message aims to correct the behavior resulting from this misperception.

The second message presents information that early coordination would increase a patient's transplant rate. This message is intended to prevent time-inconsistent donors from delaying their response to the compatibility notice. Time inconsistency is caused by present bias, a key finding of behavioral economics (Laibson, 1997; O'Donoghue and Rabin, 2001).<sup>4</sup> Time-inconsistent donors delay responding, even though they believe in advance that they should respond immediately. The second message shows that the utility of transplantation decreases over time (regardless of the individual's time preference) and aims to change the behavior of the time-inconsistent donor.

Four experimental arms were created to test the effects of the above messages. The first experimental arm sent a compatibility notice without the above two messages to the matched donors (control group). The second and third experimental arms added one of the above messages to the compatibility notice. In the fourth experimental arm, both messages were added to the compatibility notice.<sup>5</sup> We conducted a field experiment with 11,154 matched donors who received compatibility notices between September 2021 and February 2022. Participants were assigned to the experimental arms using weekly cluster randomization. We received coordination data at the end of June 2022, in collaboration with JMDP, to test the effects of the messages.

The experiment found that the intervention messages affected only younger donors with better transplant outcomes. This study has two main findings. First, providing only the information that there are fewer HLA-matched donors per patient increases the willingness to donate and the probability of reaching confirmatory typing among men in their 20s. However, this message has no statistically significant effect on other sex/genders or age groups. Second, providing information that early coordination would increase a patient's transplant rate does not affect the overall response rate for women in their 20s but does increase responses within four days of sending a compatibility notice.<sup>6</sup> In other words, this information shortens the number of days to respond rather than encouraging response behavior itself. This effect was not observed in the other sexes or age groups.

This study suggests that information can increase the willingness of potential donors to donate and provides practical insights into marrow donor programs worldwide, including the JMDP. Similar to the JMDP, the German-based international marrow donor program, DKMS, and the U.S. mar-

---

<sup>4</sup>Present bias is a phenomenon in which the time discount rate decreases over time. This can cause people to choose low current benefits even if they believe in advance that high future benefits are desirable.

<sup>5</sup>This experimental arm is designed to test the hypothesis that the simultaneous addition of two intervention messages may cause the matched donor to receive excessive information and suffer from cognitive overload.

<sup>6</sup>JMDP recommends a response to the compatibility notice within seven days.

row donor program, NMDP, have steadily increased their enrollment, but have faced challenges in keeping enrollees motivated and achieving coordination (Switzer et al., 1999, 2004; Haylock et al., 2022). Studies have examined the effectiveness of donor leave laws (Lacetera et al., 2014) and DKMS’s efforts to maintain donor motivation (Haylock et al., 2022).

Switzer et al. (2018) applied an intervention to a message sent by the NMDP when they asked matched potential donors to donate. Their intervention was a message that said, “Based on the information we currently have, you are in the unique position of likely being a perfect match for this patient.” This message was delivered over the phone to the potential donor whose HLA was a perfect match. Their experiment was not a fully randomized controlled trial and showed that this novel message did not increase the number of coordination. Although our intervention is very similar to that in this study, we show that the message effect is quite heterogeneous, affecting only certain genders and ages.

In addition, this study contributes to the economic study of costly prosocial behavior such as blood donation. Although stem cell transplantation and blood donation have some similarities, some differences also exist. In the case of blood donation, the time of intention to donate and the time of actual donation are the same, while in the case of stem cell transplantation, the two time points are different. Therefore, maintaining the willingness of potential donors is a challenge for marrow donor programs. Similar to stem cell transplantation, blood donation can be considered public goods. Wildman and Hollingsworth (2009) used data on blood donors only and showed that free-riding does not occur. In contrast, our study shows that information about the low number of HLA-compatible donors per patient increases the willingness of men in their 20s to donate, suggesting that they overestimate the number of HLA-compatible donors and engage in free-riding.

The remainder of this study is organized as follows: Section 2 provides an overview of the coordination process in the JMDP and details the field experiments. Section 3 presents the results, and Section 4 provides a discussion and conclusions.

## **2 Field Experiment**

### **2.1 Background: Coordination Process of JMDP**

To better understand the timing of our interventions and our data, we outline the coordination process leading up to the donation of stem cells by potential donors enrolled in the JMDP. First, when a potential donor is matched with a patient enrolled in the JMDP, the JMDP office sends the

donor the compatibility notice requesting a stem cell donation.<sup>7</sup> The matched donor completes a questionnaire and responds to the compatibility notice, indicating their willingness to donate.

Coordination of the transplant begins. The matched donor undergoes confirmatory typing within approximately one month. The coordinator explains the details of the donation process and asks matched donors and their families about their willingness to donate. Matched donors can choose between two collection methods (bone marrow or peripheral blood stem cell collection). In addition, the coordinating physician conducts an interview, medical examination, and blood draws to test for infection and blood type. These tests are performed to determine whether matched donors meet the criteria established by the JMDP.

Patients can be matched with up to 10 compatible donors at one time. The patient's physician selects the most appropriate candidate from matched donors who have undergone confirmatory typing. Importantly, the matched donor does not have access to any information about the matched patient (e.g., the number of other available matched donors). Nor can the matched donor obtain this information from the coordinator or the coordinating physician.

The matched donor, who is selected as the best donor, must give final consent after being informed by the coordinator and coordinating physician. Simultaneously, a representative of the donor's family must consent to the donation. Subsequently, the selected donor cannot change their mind. After the final consent is given, the selected donor admits themselves to hospital for approximately one week to undergo preoperative examinations and preparation for the donation. The donor then undergoes a surgical procedure to collect stem cells. The time from the confirmatory typing to collection is approximately 3 to 4 months.

## 2.2 Experimental Design

Our experiment intervenes in the content of the compatibility notice in which JMDP requests stem cell donation from a matched donor. The standard compatibility notice is as follows:

We inform you that your HLA type (white blood cell type) matches that of a patient on our registry and you have been selected as one of our potential donors. We are contacting you to ask if you would like to undergo further testing and interviews in preparation for donation. Please read the enclosed materials carefully, consider whether coordination is possible, and return the form *within 7 days* of receiving this information. [Insert intervention messages here] If we proceed with the coordination after receiving your

---

<sup>7</sup>At the same time, the JMDP office sends a social networking message to the matched donors informing them that JMDP has sent the compatibility notice.

return, a coordinator will contact you by phone to discuss the details of your request.

The donor should respond to the compatibility notice within 7 days. JMDP also encloses a handbook that describes the coordination process outlined in the previous subsection as well as the medical questionnaire and donor consent form.

We added two messages (a probability message and an early coordination message) to the compatibility notice to facilitate coordination.<sup>8</sup> The Probability message is as follows:

The number of registered donors whose HLA type matches that of a single registered patient is one in hundreds to tens of thousands. We hope you understand that while we may find more than one potential donor, it is not a large number.

This message emphasizes the low number of matched donors per registered patient. Transplantation is public goods in the sense that, if there are other potential donors with the same HLA type in the pool, one's donation can be substituted for that compatible donor. Thus, as predicted by the volunteer dilemma, the more common the HLA type, the more reluctant the donor will be to donate.

However, in the context of stem cell transplantation, the expectation of the group size influences the decision because a matched donor cannot know the exact size of the group. Therefore, the higher the donor's expectation regarding the number of potential donors with the same HLA type, the more reluctant they will be to donate. If a donor's beliefs are too high, the probability message discourages free-riding and increases willingness to donate by adjusting beliefs downward. The opposite effect may occur. If the donor's beliefs are underestimated, the probability message may induce free-riding and reduce donation intentions by adjusting beliefs upward.

The Early Coordination message is as follows:

About 60% of patients can receive a transplant through the Japan Marrow Donor Program. The earlier a donor can be found to donate bone marrow, the higher that percentage can be.

This message is expected to work for time-inconsistent matched donors. Time-inconsistent donors may think it is optimal to respond immediately in advance but delay their response. This

---

<sup>8</sup>In designing our intervention messages, we have been careful to avoid putting undue psychological pressure on potential donors. Specifically, first, we avoid language that sounds like an appeal. Second, we only use information that is publicly available from the JMDP. In addition, the risks of transplantation are explained in the usual way. The intervention message was approved by the Institutional Review Board of the Graduate School of Economics of Osaka University and the JMDP.

Table 1: 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

Notes: See Table 2 for detail intervention of each experimental arm.  
Control group is experimental arm A.

message suggests that early coordination increases the rate of transplantation for the patient and that the utility of transplantation decays over time (regardless of the individual’s time preference). Time-inconsistent donors who read this message may recognize that delaying a response does not increase utility and may change their decision to respond immediately. Thus, this message is expected to increase the response rate over a short period.

Four experimental arms were established to estimate the effects of two intervention messages. Experimental arm A received a standard compatibility notice with no intervention messages (control group). Experimental arms B and C received notices with probability messages and early coordination messages, respectively. Experimental arm D received a notice with two intervention messages added simultaneously. This experimental arm was designed to test the negative effects of the cognitive load caused by information overload.

The participants in the experiment were 11,154 matched donors who received a compatibility notice between September 2021 and February 2022. To maintain randomness to the best of the JMDP office’s abilities, we assigned experimental arms using weekly cluster randomization. The experimental arms were designed to be balanced across weeks and months as much as possible. Table 1 summarizes the assignment schedule. Before conducting the experiment, we obtained approval from the institutional review board of the Graduate School of Economics, Osaka University (approval number: R030305) and JMDP (approval number: JMDP2021-04).

### 2.3 Data and Empirical Strategy

JMDP provided coordination data at the end of June 2022. The unit of observation is the experimental participant. For individual characteristics, the data recorded gender, age, number of coordination experiences, and prefecture-level residence area. About During the coordination process, the data recorded whether each stage (response to compatibility notice, confirmatory typing, candidate selection, final consent, and collection) was reached. We used these variables as outcome variables. In addition, for responses to the notice, the data recorded the number of days required to



Table 2: Summary of Field Experiment

	Experimental Arms				F-test, p-value
	A	B	C	D	
<b>A. Interventions</b>					
Standard notification	X	X	X	X	
Probability message		X		X	
Early Coordination message			X	X	
<b>B. Sample Size</b>					
N	2535	3053	2726	2735	
<b>C. Balance Test</b>					
Male (= 1)	0.624	0.633	0.631	0.609	0.231
Age	38.376	38.121	37.448	37.978	0.004
Number of past coordinations	1.609	1.589	1.625	1.563	0.130
Number of listed hospitals	0.476	0.490	0.487	0.485	0.835
Number of hospitals listed with PBSC collection	0.162	0.167	0.166	0.164	0.838
Number of hospitals listed with BM collection	0.246	0.256	0.254	0.251	0.741

Notes: For balance test, we regress a covariate on treatment dummies and test a null hypothesis that all coefficients are zero. We use the robust standard errors for statistical inference.

respond and the willingness to donate. If coordination was interrupted, the reasons for the interruption were recorded in three categories (patient reasons, donor non-health reasons, and donor health reasons). The analysis uses 11,049 matched donors living in Japan whose coordination (including interruptions) was completed.<sup>9</sup>

For additional data, we used a list of medical institutions published on the Internet by the JMDP.<sup>10</sup> This list includes complete addresses, availability of bone marrow collection (BM collection), and availability of peripheral blood stem cell collection (PBSC collection). We aggregated this list at the prefecture level, calculated the number of hospitals per 10 square kilometers, and merged it with the coordination data using the prefecture as the merge key. We consider this variable to be the traveling cost of coordination and donation.

Table 2 summarizes the field experiment. Panel A shows the intervention for each experimental arm and Panel B shows the sample size for each experimental arm. Panel C shows a balanced test of whether randomization was successful. The assignment of the experimental arms is approximately random because there is no average difference between the experimental arms for any variable except for age. However, age may vary across groups. The average age of the experimental arm C is approximately one year younger than that of the control group.

Because the assignment of experimental arms should be independent of the potential outcomes (the outcome variable that would be observed if an experimental arm is assigned) conditioned

<sup>9</sup>One matched donor lived abroad. There were 104 matched donors with ongoing coordination at the time of data provision. The proportion of matched donors with ongoing matching is balanced across the experimental arms (F-test, p-value = 0.383).

<sup>10</sup><https://www.jmdp.or.jp/hospitals/view2/> (access date: August 4, 2022)

on a predetermined variable, we can identify the average treatment effect by the difference in means across experimental arms conditioned on the predetermined variable. Thus, we estimate the following linear probability model for individual  $i$  who received a compatibility notice in week  $w$  of month  $m$ .

$$Y_{imw} = \beta_1 \cdot B_{mw} + \beta_2 \cdot C_{mw} + \beta_3 \cdot D_{mw} + X_i' \gamma + \lambda_m + \theta_w + u_{imw}, \quad (1)$$

where  $X_i$  is a vector of individual characteristics. We add month and week dummy variables  $\lambda_m$  and  $\theta_w$  to control for common shocks in a given time period. The parameters of interest are  $(\beta_1, \beta_2, \beta_3)$ . When fixed effects are added, there seems to be no cause for the generation of correlations within the clusters (experimental weeks) of unobservable elements  $u_{imw}$ . Thus, we use robust standard errors for statistical inference.<sup>11</sup>

### 3 Experimental Results

#### 3.1 Effects on Reply and Intention

First, we estimate the message effect on the responses that are most likely to reflect the intention to donate. The outcome variable for responses is a dummy variable that takes the value of one if a donor responds to the compatibility notice, regardless of intention. The response rate in the control group (experimental arm A) is 88.35%.

We decompose the message effect on responses into two effects. The first is the effect on the responses associated with the intention to donate. The outcome variable is a dummy variable that takes one if a matched donor responded to the compatibility notice and indicated a willingness to donate. The second is the effect on responses without intention. The outcome variable is a dummy variable that takes one if a matched donor responded to the compatibility notice and did not indicate an intention to donate.

To estimate these effects, we code the outcome variables of non-responders as zero and include them in the analysis sample. Thus, the sum of the two outcome variables is always equal to a dummy variable for response, such that the sum of the effects on positive and negative intentions is the effect on response. In the control group, the response rate for positive intentions is 55.33%, whereas that for negative intentions is 33.03%. Thus, 62.63(= 55.33/88.35)% of respondents have the intention to donate.

---

<sup>11</sup> We conducted a regression analysis with cluster standard errors as a robustness check, confirming no change in the main results presented in this paper.

Table 3: Linear Probability Model of Reply and Intention

	Intention					
	Reply		Positive		Negative	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment B	0.010 (0.008)	0.014 (0.009)	0.022 (0.013)	0.021 (0.014)	−0.011 (0.013)	−0.007 (0.013)
Treatment C	−0.006 (0.009)	0.003 (0.009)	−0.005 (0.014)	−0.003 (0.014)	0.000 (0.013)	0.005 (0.014)
Treatment D	0.006 (0.009)	0.006 (0.009)	0.005 (0.014)	0.006 (0.014)	0.001 (0.013)	0.000 (0.013)
Control average	0.8835	0.8835	0.5533	0.5533	0.3303	0.3303
Covariates	X		X		X	
Num.Obs.	10 985	10 985	10 985	10 985	10 985	10 985

Notes: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The robust standard errors are in parentheses. Covariates are gender, (demeaned) age, its squared term, 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.

Coordination may be interrupted for patient reasons before a matched donor responds. Given that this happens independently of the matched donor’s intentions, we exclude it from our analysis sample (0.5–0.7% in each experimental arm).

Table 3 presents the estimation results. The odd columns include only the experimental arm dummy as an explanatory variable, which corresponds to a simple difference in means. The even columns control for individual attributes, month, and week fixed effects, in addition to the experimental arm dummies. The estimated results show that all experimental arms, with and without covariates, have no statistically significant effect on responses and intentions. As an alternative method, we estimate a logit model (see Table 8 in the Appendix) and obtain similar results (95% confidence intervals for odds ratios include 1).

Next, to test for heterogeneity in message effects, we divide the sample into four subsets by gender and age (under 30 or not) and estimate the message effects in each subset. Figure 1 presents the coefficient plots. The results show that experimental arms B and D, which include the probability message, increase the response rate of men in their 20s by approximately 6 percentage points (8.06% increase, since the control mean is 74.4%), which is statistically significant at the 5% level. In particular, experimental arm B, to which only the probability message was added, increased the number of responses with positive intentions by approximately 10 percentage points (25.91% increase, because the control mean is 38.6%), which is also statistically significant. However, for the other gender and age groups, the treatment messages have no statistically significant effect on responses and intentions.

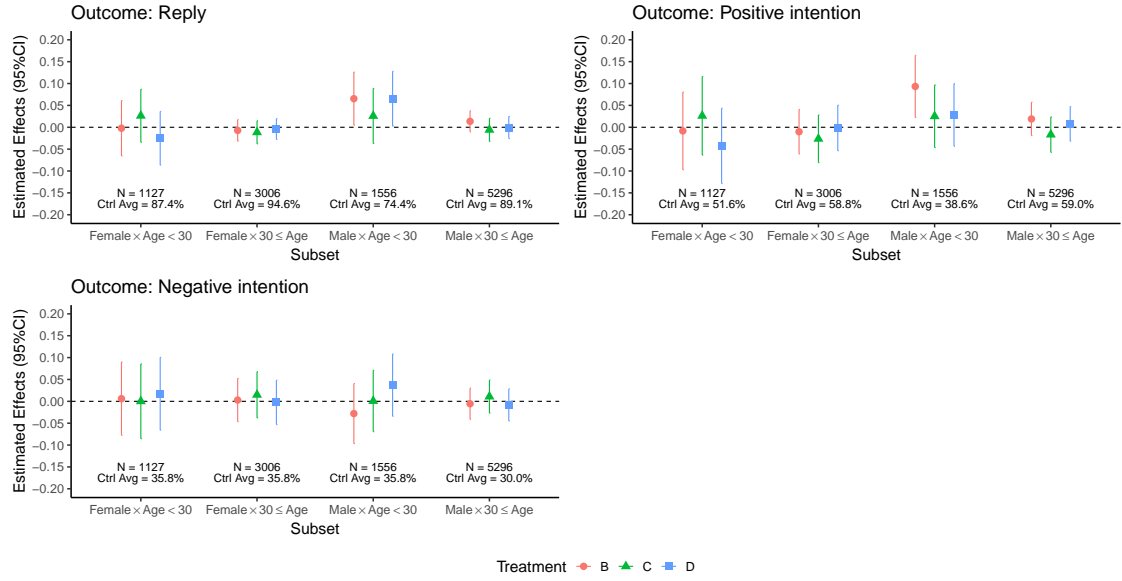


Figure 1: Effect on Reply and Intentions by Gender and Age Group. Notes: These plots show the average effect (and associated 95 percent confidential interval) on each outcome by gender and age group. We use robust standard errors. We control number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, number of hospitals with BM collection per 10 square kilometers, month and week dummies.

In the control group, approximately 50% ( $= 38.6/74.4$ ) of men in their 20s who responded to the compatibility notice were willing to donate, which is lower than the willingness rate of respondents in the other gender and age groups.<sup>12</sup> Experimental arm B (probability message) raises the intention rate of male respondents in their 20s from 50% to 60% ( $= (38.6 + 10)/(74.4 + 6)$ ). In addition, young men have better transplant outcomes than men of other sexes and ages. Taken together, we can say that the probability message improves coordination efficiency in the sense that it encourages behavioral change (initiation of coordination) in desirable donors.

### 3.2 Random Causal Forest

To further explore the heterogeneity of message effects, we use random causal forests (RCF), which allows us to test the heterogeneous treatment effects in a fully nonparametric manner (Athey and Imbens, 2016; Wager and Athey, 2018). The method is based on a regression tree algorithm and estimates the average treatment effect conditional on the observed characteristics. This method has been used in various contexts, including labor (Davis and Heller, 2017), education (Carlana et al., 2022), and energy conservation (Murakami et al., 2022).

<sup>12</sup>For women in their 20s, 59 ( $= 51.6/87.4$ )%; for men over 30, 66 ( $= 59.0/89.1$ )%; for women over 30, 62 ( $= 58.8/94.6$ )%.

The algorithm splits the sample into two subsamples (leaves) using one of the covariates  $X_j$ . Specifically, the algorithm divides the sample into samples for which  $X_j \leq x$  and  $X_j > x$ . The regression tree determines a specific threshold  $x$  to minimize the mean squared error of the outcome variable. Given that the RCF aims to estimate the average treatment effect of a leaf (conditional average treatment effect or CATE), the threshold  $x$  is set to minimize the expected mean squared error of the predicted treatment effect. This minimization is achieved by increasing the variance in the conditional mean treatment effect across leaves (heterogeneity) and decreasing the variance within leaves. This splitting process is repeated for each leaf until the termination condition is reached. RCF predicts the conditional mean treatment effect on terminal leaves.

A disadvantage of the regression tree algorithm is that the variance in the predictions increases (overfitting), which reduces the accuracy of the predictions. To prevent this, RCF introduces an algorithm called the ensemble method. It creates thousands of subsamples and grows a tree for each subsample. The final result is the average of the predictions from thousands of trees.<sup>13</sup>

RCF predicts the treatment effect from the observed characteristics. In other words, it can predict the treatment effects based on individual characteristics, even for individuals who did not receive the intervention. This method assumes that treatment assignment is conditionally independent of the potential outcome variable. In our study, this method could be used because the experimental arm was approximately randomly assigned. This subsection focuses on the effects on intentions. The covariates used in the RCF are gender, age, past coordination experience, and the density of hospitals in the area of residence.

Figure 2 illustrates the distribution of predicted treatment effects for each experimental group by gender and age. The predicted treatment effects vary considerably depending on age and gender. Even when conditioned on gender and age, the predicted treatment effects exhibit significant heterogeneity. Notably, the effect of experimental group B is positive for men in their early 20s.<sup>14</sup> As a result, the average treatment effect of experimental arm B for men in their 20s is 12.3 percentage points, which is close to the results of the subsample analysis shown in Figure 1 (see Table 10 in the Appendix). This effect is statistically significantly higher than the CATE for women of the same age group ( $-0.3$  percentage points) and men aged 30 and above (1.6 percentage points), as the 95% confidence intervals for the difference in effects do not include zero.

<sup>13</sup>For technical details, see Athey and Imbens (2016) and Wager and Athey (2018). This study also uses a technique called generalized random forests, which treats RCF as a special case; however, the intuition remains the same. See Athey and Wager (2019) and Athey et al. (2019) for details on generalized random forests.

<sup>14</sup>Experimental arm B also exhibits a positive impact of a similar magnitude on some men aged 30 to 40, comparable to the positive effect observed in men in their early 20s. Such individuals tend to have relatively more coordination experience and reside in regions with a higher concentration of hospitals (see Table 11 in the Appendix).

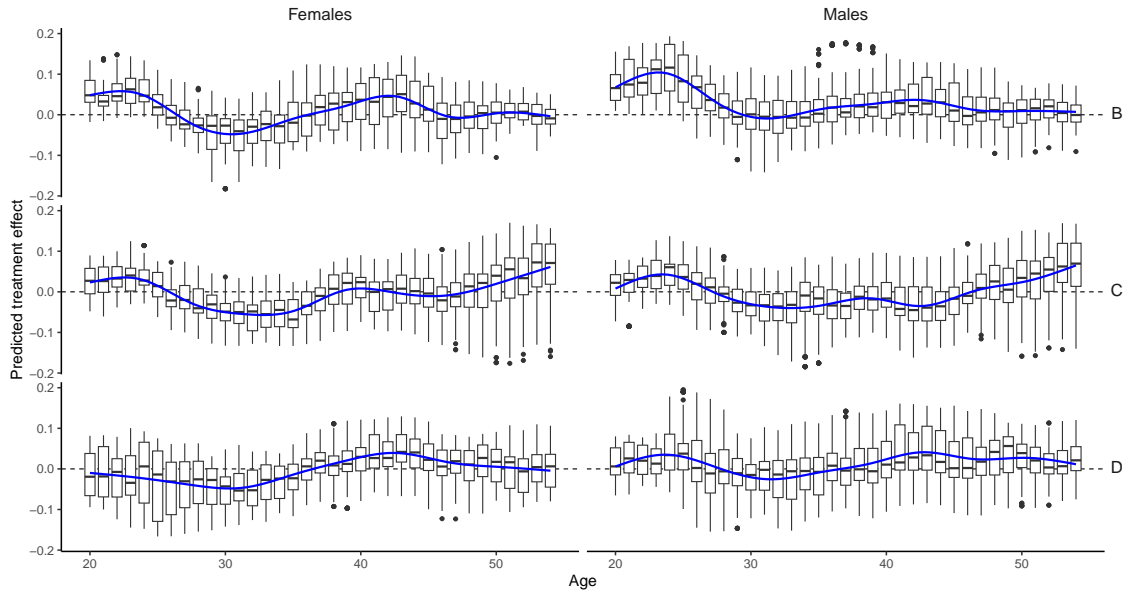


Figure 2: Boxplot of Predicted Treatment Effects by Treatment, Gender and Age. Notes: Fitted line represents GAM smoothing. Covariates are gender, age, number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, and number of hospitals with BM collection per 10 square kilometers.

From the perspective of the distribution and CATE, the effect of experimental arm D on men in their 20s, containing the probability message, is lower than the effect of experimental arm B. Since experimental arm D simultaneously presents the early coordination message, young men may have discounted the probability message due to information overload. Table 4 presents the results of regressing the predicted effects of experimental arm D on the predicted effects of experimental arm B and C. If donors fully understand the information, the coefficients, excluding the intercept, should be 1 (the effect of experimental arm D equals the sum of the effects of experimental arms B and C). However, the F-test rejects this null hypothesis, suggesting that donors may not be accurately evaluating the information. Furthermore, the impact of experimental arm C is larger than that of experimental arm B, indicating that young men may prioritize the early coordination message over the probability message. Therefore, by simultaneously providing both messages, young male donors may have made decisions emphasizing early coordination due to information overload.<sup>15</sup>

The heterogeneity of treatment effects indicates that there is a mixture of individuals experiencing positive and negative effects (see Figure 9, upper left, upper right, and lower left panels in the Appendix). Therefore, a targeting policy that tailors treatment to individuals may be more

<sup>15</sup>Similarly, young women also prioritize the early coordination message in decision-making. In contrast, men and women aged 30 and above prioritize the probability message in decision-making.

Table 4: Correlation of Predicted Treatment Effects

	Treatment effect of D			
	Female		Male	
	Age < 30	30 ≤ Age	Age < 30	30 ≤ Age
	(1)	(2)	(3)	(4)
(Intercept)	−0.031*** (0.002)	0.007*** (0.001)	0.005** (0.002)	0.008*** (0.001)
Treatment B	−0.036 (0.036)	0.396*** (0.014)	−0.061** (0.026)	0.646*** (0.016)
Treatment C	0.755*** (0.034)	0.325*** (0.013)	0.884*** (0.030)	0.274*** (0.011)
<i>F-test, p-value</i>				
H0: $B = 1$ & $C = 1$	< 0.001	< 0.001	< 0.001	< 0.001
H0: $B = C$	< 0.001	0.001	< 0.001	< 0.001
Num.Obs.	1127	3006	1556	5296

Notes: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The robust standard errors are in parentheses. Outcome variable is treatment effect of experimental arm D predicted by RCF. Explanatory variables are treatment effects of experimental arm B and C predicted by RCF.

Table 5: Uniform Assignment Policy and Targeting Policy

	Uniform		Targeting	
	Mean	S.D.	Mean	S.D.
B	0.0203	0.0565	0.0331	0.0416
C	-0.0032	0.0597	0.0222	0.0331
D	0.0085	0.0613	0.0285	0.0389
Optimum			0.0518	0.0455

Notes: The table presents the simple average and standard deviation of treatment effects for a particular policy. The uniform policy allocates a specific treatment to all individuals. The targeting policy allocates a treatment only to individuals for whom the treatment effect is positive, assigning the control (experimental arm A) to others. The optimal targeting policy provides the most effective treatment to each individual. When the effects of experimental arms B, C, and D are negative, this policy assigns the control.

Table 6: Average Characteristics of Optimal Treatment

	Optimal treatment			
	A	B	C	D
Male (= 1)	0.55	0.65	0.59	0.67
Age	36.73	34.44	43.03	39.50
Number of past coordinations	1.45	1.53	1.68	1.71
Number of listed hospitals	0.47	0.44	0.39	0.62
Number of hospitals listed with PBSC collection	0.15	0.15	0.13	0.22
Number of hospitals listed with BM collection	0.24	0.23	0.20	0.33
N	1961	3806	2271	2947

Notes: The table illustrates the average attributes of individuals assigned to a specific treatment under the optimal targeting policy.

effective than a policy that uniformly applies the treatment. Considering a targeting policy that assigns treatment only to individuals with a positive treatment effect, while assigning the control (experimental arm A) to others, Table 5 shows that this targeting policy enhances the average effect compared to a uniform assignment policy. For instance, the targeting policy for experimental arm B increases the average effect by 63%.

Since an individual may experience negative effects from one treatment but positive effects from another, a targeting policy that administers the most effective treatment to each individual (optimum targeting policy) may be more effective than a targeting policy that only treats individuals with a positive treatment effect.<sup>16</sup> Indeed, Table 5 demonstrates that optimal targeting is the most effective policy. Compared to a policy that uniformly assigns experimental arm B, the optimum targeting policy increases the average effect by 155%. Compared to a targeting policy that assigns experimental arm B to individuals with a positive effect, the optimum targeting policy increases the average effect by 133%. Table 6 compares the characteristics of average individual for each experimental arm in the optimum targeting policy. The allocation ratios for each experimental group in the optimum targeting policy are as follows: 18% for experimental arm A, 35% for experimental arm B, 20% for experimental arm C, and 27% for experimental arm D. Average individual assigned to group B is relatively man, young, having little coordination experience, and residing in areas with many hospitals (major cities like Tokyo). This result further suggests the effectiveness of group B for young men.



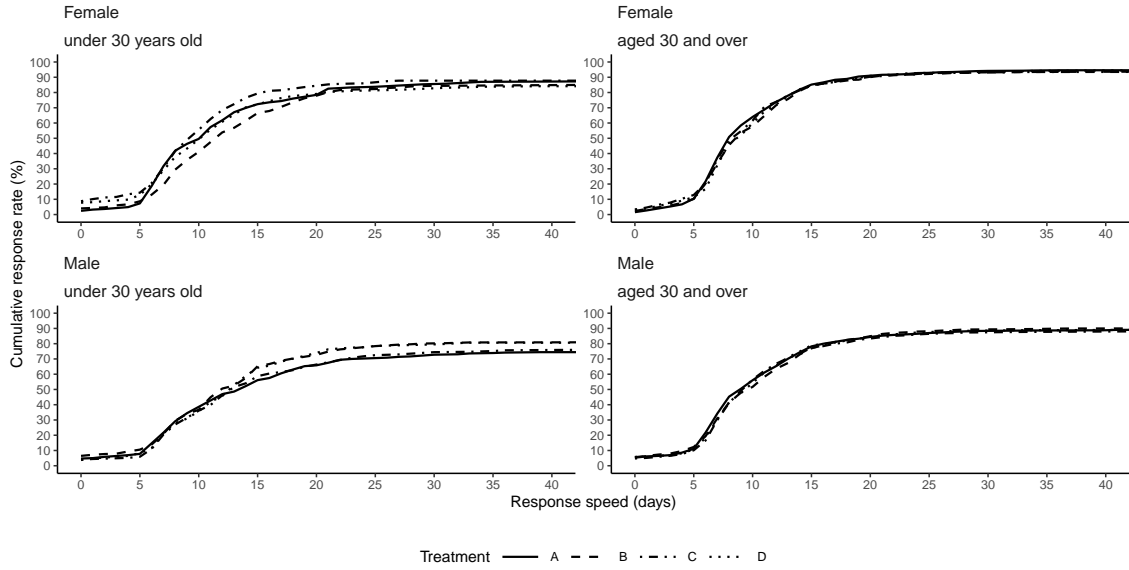


Figure 3: Cumulative Response Rates by Gender and Age Group.

### 3.3 Response Speed to Notification

The early coordination message states that early coordination increases the patient’s transplantation rate. Therefore, this message may encourage early responses to compatibility notice. Figure 3 shows the cumulative response rate for the 40 days after the compatibility notice was sent. Given that the effect on responses is heterogeneous by gender and age, as shown in the previous results, we also split the subsample by gender and age and focus on the heterogeneity of the effect on quick responses.

While there are no significant differences in the pattern of cumulative response rates for men and women over 30 years, there are some notable results for men and women under 30 years.<sup>17</sup> In the group of men in their 20s, 15 days after the compatibility notice was sent, the responses of experimental arms B and D began to increase relative to those of the controls. Consequently, the cumulative response rate of the two experimental arms is statistically significantly higher than that of the control arm (Figure 4).<sup>18</sup> Given that the compatibility notice recommends a response within 7 days, a response 15 days after the mailing of the compatibility notice cannot be considered an early response. Thus, although experimental arms B and D increased the ultimate response rate among men in their 20s, they did not encourage early responses. This result is natural be-

<sup>16</sup>In the optimum targeting policy, the control group (experimental arm A) is allocated when the treatment effects for other experimental arms are negative.

<sup>17</sup>For men and women over 30 years of age, the cumulative response rate for each day differs little between the experimental arms, except for certain periods, and is not statistically significant. See Figure 7 and 8 in Appendix.

<sup>18</sup>In the regression analysis, we created a dummy variable that takes 1 if the potential donor responded within  $d$  days as the outcome variable. If the potential donor responded after  $d$  days or did not respond, the outcome variable is 0.

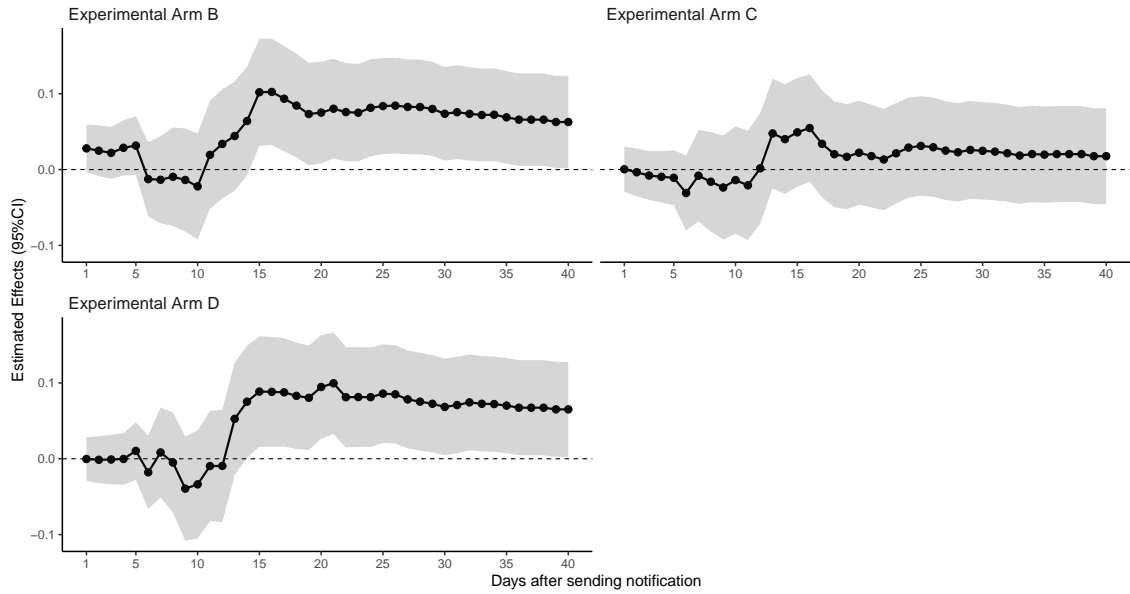


Figure 4: Effect on Reply within Specific Days after Sending Notification among Males Less than 30. Notes: These plots show the average effect (and associated 95 percent confidential interval) on cumulative responses on specific days. We use robust standard errors. We control number of past coordinations, 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.

cause the probability message, the intervention in experimental arms B and D, is not intended to encourage early responses. In addition, the trend in experimental arm C, where only Early Coordination messages are added, is almost the same as that in the control, so it cannot be said that Early Coordination messages encourage early responses in this group.

In the 20s women group, the responses of experimental arms C and D, with the addition of the early coordination message, increased compared with the control within four days of sending the compatibility message. As a result, the cumulative response rates of the two experimental arms are statistically significantly higher than that of the control arm (Figure 5). In addition, 10 days after the compatibility notice is sent, the responses of experimental arm C begin to temporarily increase relative to the control, but eventually cease to be significantly different from the cumulative response rate of the control. These results suggest that early coordination messages encourage females in their 20s to respond quickly.<sup>19</sup>

<sup>19</sup>We decompose the effect of message C on responses within four days (early responses) into two parts in terms of intentions and find that it increases positive and negative intentions to the same extent.

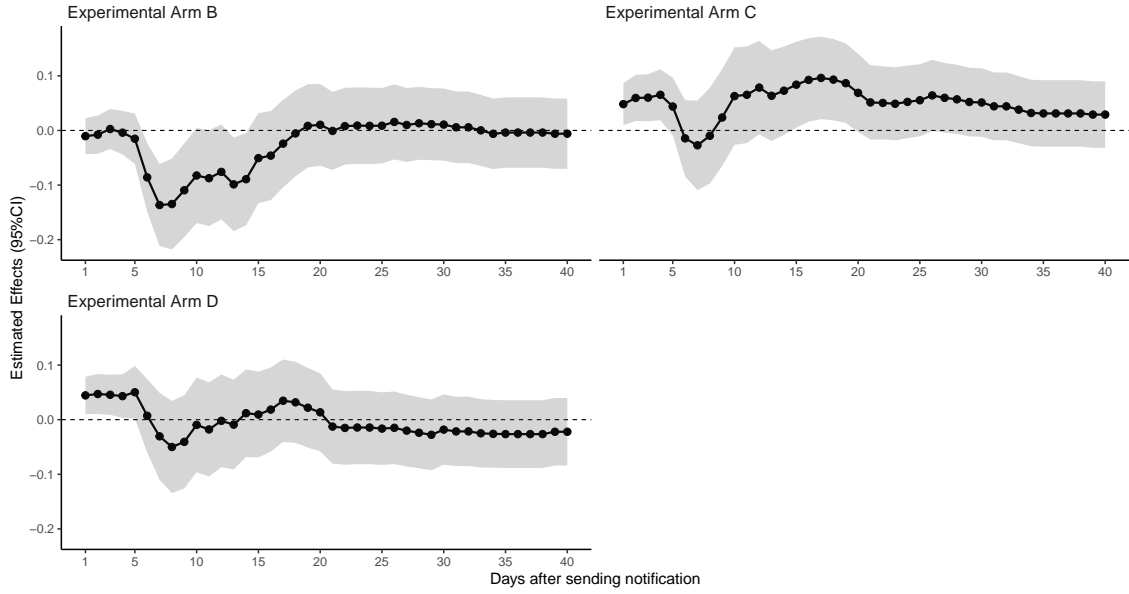


Figure 5: Effect on Reply within Specific Days after Sending Notification among Females Less than 30. Notes: These plots show the average effect (and associated 95 percent confidential interval) on cumulative responses on specific days. We use robust standard errors. We control number of past coordinations, 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.

### 3.4 Effects on the Coordination Process

Finally, we examine the impact of messages on each step of the coordination process. As explained in Section 2.1, the coordination process comprises four stages: confirmatory typing, candidate selection, final consent, and collection. As the outcome variable, we use a dummy variable that takes the value of 1 if a matched donor has reached each stage.

As in the analysis in Section 3.1, we exclude samples in which coordination appears to have been interrupted, independent of the matched donor willingness. When estimating the effect of confirmatory typing, we exclude cases of interruption due to patient-related reasons. Given that the patient's physician is likely to select a healthier matched donor at the time of candidate selection, cases of interruption of donor health reasons after candidate selection are considered to have occurred independently of the donor's willingness. Thus, when estimating the effects on candidate selection, final consent, and donation, we exclude samples interrupted for patient and samples interrupted for donor health reasons after candidate selection. Note that we should be cautious in our interpretation, because sample exclusion alone may not completely eliminate physician decision-making. In the control group, 24% of eligible donors underwent confirmatory testing, 8% became candidates, and 6% ultimately donated.

Table 7: Linear Probability Model of Coordination

	CT		Candidate		Consent		Donation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment B	0.033*** (0.012)	0.033*** (0.012)	0.005 (0.008)	0.003 (0.008)	0.006 (0.008)	0.004 (0.008)	0.004 (0.007)	0.002 (0.007)
Treatment C	0.015 (0.012)	0.014 (0.013)	0.001 (0.008)	−0.002 (0.009)	0.002 (0.008)	−0.001 (0.008)	0.002 (0.007)	−0.002 (0.008)
Treatment D	0.026** (0.012)	0.030** (0.012)	0.008 (0.009)	0.009 (0.009)	0.010 (0.008)	0.010 (0.008)	0.003 (0.007)	0.003 (0.008)
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.	10 435	10 435	8587	8587	8558	8558	8441	8441

Notes: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The robust standard errors are in parentheses. Covariates are gender, (demeaned) age, its squared term, 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.

Table 7 lists the results of the full-sample estimation. The results show that experimental arms B and D, which include probability messages, increase the rate of confirmatory typing by approximately 3 percentage points, which is a statistically significant effect. This effect is greater than that of the responses. This is because the likelihood of coordination interruption for donor reasons, including health reasons, is lower in experimental arms B and D than in the control group. Thus, although experimental arms B and D do not increase the overall number of people willing to donate, they maintain their intention to donate, which contributes to a reduction in coordination dropouts.

The effect of experimental arms B and D on the stage after candidate selection is less than 1 percentage point and not statistically significant. These findings should be interpreted with caution. Given that our intervention did not affect demand (number of patients), if our intervention had increased the number of people who reached confirmatory typing, it should have increased the number of people who were not selected as candidates for exogenous reasons.

In experimental arm B and the control group, the numbers of potential donors who reached confirmatory typing were 774 and 564, respectively. Therefore, experimental arm B would have increased the number of confirmatory typings by 210. The number of potential donors who were not selected as candidates for patient or donor health reasons was 556 in the experimental arm B and 385 in the control group. This means that experimental arm B has 171 more people who were not selected as candidates for exogenous reasons. Thus, the estimates include not only the effect of our intervention but also the effect of the demand for stem cell transplants.

Figure 6 divides the sample into four subsets by gender and age (under 30 or not) and estimates

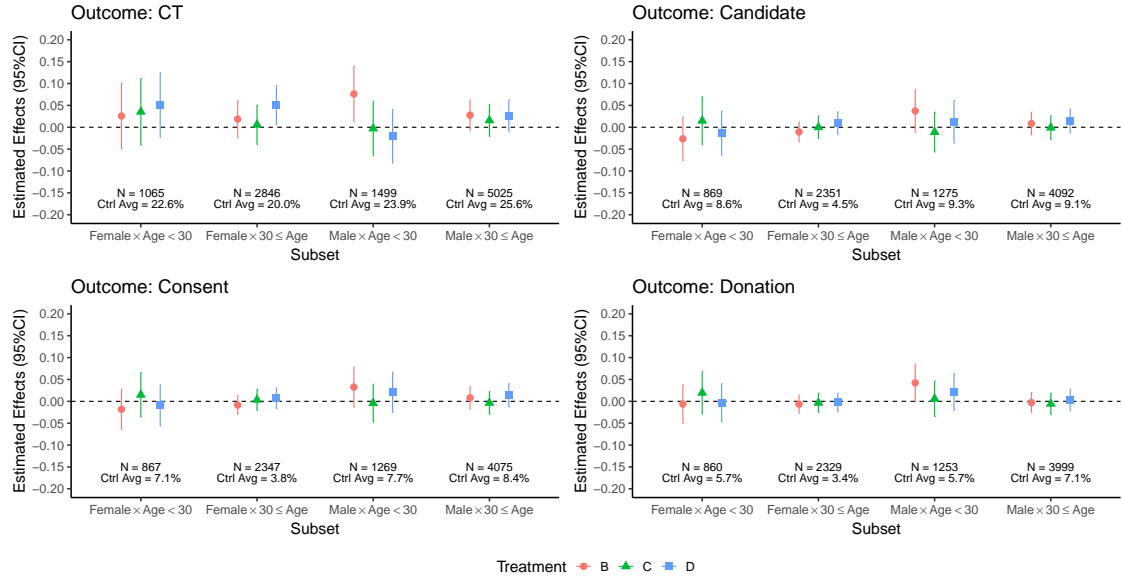


Figure 6: Effect on Coordination by Gender and Age Group. Note: These plots show the average effect (and associated 95 percent confidential interval) on each outcome by gender and age group. We use robust standard errors. We control number of past coordinations, number of hospitals per 10 square kilometers, 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.

the message effect in each subset. The results show that experimental arm B increases the proportion of men in their 20s who reach confirmatory typing by approximately 8 percentage points, which is statistically significant. Thus, it is possible that experimental arm B not only increases the willingness of men in their 20s to donate but also maintains it.

Experimental arm B may increase donations among men in their 20s by 4 percentage points, which is statistically significant at the 10% level. However, the effect on candidate selection and final consent is not statistically significant. As noted earlier, this effect may reflect not only the effect of our intervention but also the demand for stem cell transplants. If younger males have better transplant outcomes, the demand for stem cell transplants may be higher in this generation than in other genders and ages. The effect on donations may include a demand effect. In addition, experimental arm C, which encouraged early responses from women in their 20s, had no statistically significant effect on the coordination of this generation.

## 4 Discussion and Conclusions

This study examined the effects of providing information to increase the willingness to donate among potential donors enrolled in the JMDP. The results showed that information about a low

number of HLA-matched donors per patient (probability message) increased the willingness to donate among men in their 20s by approximately 10 percentage points. However, when we simultaneously provide the information that early coordination increases a patient's transplant rate (early coordination message), the positive effect of the probability message disappears because they give more weight to the early coordination message in their decision-making. In addition, this message increased the rate of achieving confirmatory typing by approximately 8 percentage points and may increase the transplant rate by 5 percentage points. In terms of numbers, the Probability message increased the number of transplanted donors from 71 to 121. Thus, information about HLA matching of others would increase the efficiency of coordination in the sense that the message encourages donors with good transplant performance to proceed with coordination.

These results suggest that men in their 20s overestimate the number of HLA-matched donors and engage in free-riding behavior. There are two possible reasons for the statistically insignificant effect of the probability message on the other genders and ages. First, compared to men in their 20s, others may have correctly estimated the number of HLA-matched donors. In this case, the probability message should not affect the decisions of potential donors.

The second possibility is that the altruistic preferences of men in their 20s differ from those of other genders and age groups. Economic studies suggest that there are two main types of motives for altruistic behavior: warm glow, in which one gains utility from one's altruistic behavior; and pure altruism, in which one gains utility from the results of altruistic behavior, such as the production of public goods (Andreoni, 1990). Those with relatively stronger warm glow were less likely to engage in free-riding behavior because they were less concerned about the actions of others. Thus, compared to men in their 20s, others may have a warm glow preference as the main driver of altruistic behavior. In short, the heterogeneity in probability messages can be explained by differences in beliefs or motivations.

The early coordination message had no effect on the overall response rate among women in their 20s, but had a positive effect on shorter responses (4 days or less). This suggests that it shortens the timing of replies rather than encouraging response behavior itself. Owing to our data limitations, we cannot test whether this message shortens the time to transplantation for patients; however, the information might contribute to a shorter coordination period.

The lack of effect for other genders and ages may be due to the possibility that others already have this information. Alternatively, women in their 20s may have a stronger degree of present bias and a greater tendency toward delayed behavior than other generations. The heterogeneity of the message effect could be explained by differences in information possession or time preferences.

Although this study identifies the causal effects of information provision through field experiments, it is limited by the fact that the data and experimental design do not allow for the identification of the mechanisms described above, which may be an issue for future work. This study has several practical implications. As discussed in Section 1, maintaining the intention of potential donors is a challenge for bone marrow donor programs due to the time lag between intention and behavior. In particular, young donors with good transplant outcomes are likely to fail to maintain their intentions and drop out of the program. This study suggests that information provided by the bone marrow donor program office at the time of matching with a patient may promote behavioral changes in young potential donors and may be one measure to address this problem. However, some studies (for example, Switzer et al., 2018) have shown that information provision is ineffective, so it is expected that the effectiveness of our information provision will be tested in bone marrow donor programs in other countries.

## Appendix

Table 8: Logit Model of Reply and Intention

	Reply		Intention			
			Positive		Negative	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment B	1.11 [0.94, 1.31]	1.15 [0.96, 1.37]	1.09 [0.98, 1.22]	1.09 [0.98, 1.22]	0.95 [0.85, 1.06]	0.97 [0.86, 1.09]
Treatment C	0.95 [0.80, 1.12]	1.02 [0.86, 1.22]	0.98 [0.88, 1.09]	0.99 [0.88, 1.11]	1.00 [0.89, 1.12]	1.03 [0.91, 1.16]
Treatment D	1.06 [0.89, 1.26]	1.06 [0.89, 1.27]	1.02 [0.91, 1.14]	1.02 [0.91, 1.15]	1.01 [0.90, 1.13]	1.00 [0.89, 1.13]
Covariates	X		X		X	
Num.Obs.	10 985	10 985	10 985	10 985	10 985	10 985
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 in square brackets. Covariates are gender, (demeaned) age, its squared term, 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.

Table 9: Logit Model of Coordination

	CT		Candidate		Consent		Donation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment B	1.19 [1.05, 1.35]	1.20 [1.05, 1.36]	1.07 [0.86, 1.33]	1.04 [0.83, 1.31]	1.09 [0.87, 1.38]	1.06 [0.83, 1.35]	1.07 [0.83, 1.39]	1.04 [0.80, 1.35]
Treatment C	1.08 [0.95, 1.23]	1.08 [0.94, 1.24]	1.01 [0.81, 1.27]	0.97 [0.76, 1.23]	1.04 [0.82, 1.32]	0.99 [0.77, 1.27]	1.03 [0.79, 1.34]	0.98 [0.74, 1.29]
Treatment D	1.15 [1.01, 1.31]	1.18 [1.03, 1.35]	1.12 [0.90, 1.40]	1.12 [0.89, 1.41]	1.16 [0.92, 1.46]	1.16 [0.91, 1.47]	1.06 [0.81, 1.37]	1.05 [0.80, 1.38]
Covariates	X		X		X		X	
Num.Obs.	10 435	10 435	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 in square brackets. Covariates are gender, (demeaned) age, its squared term, 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.



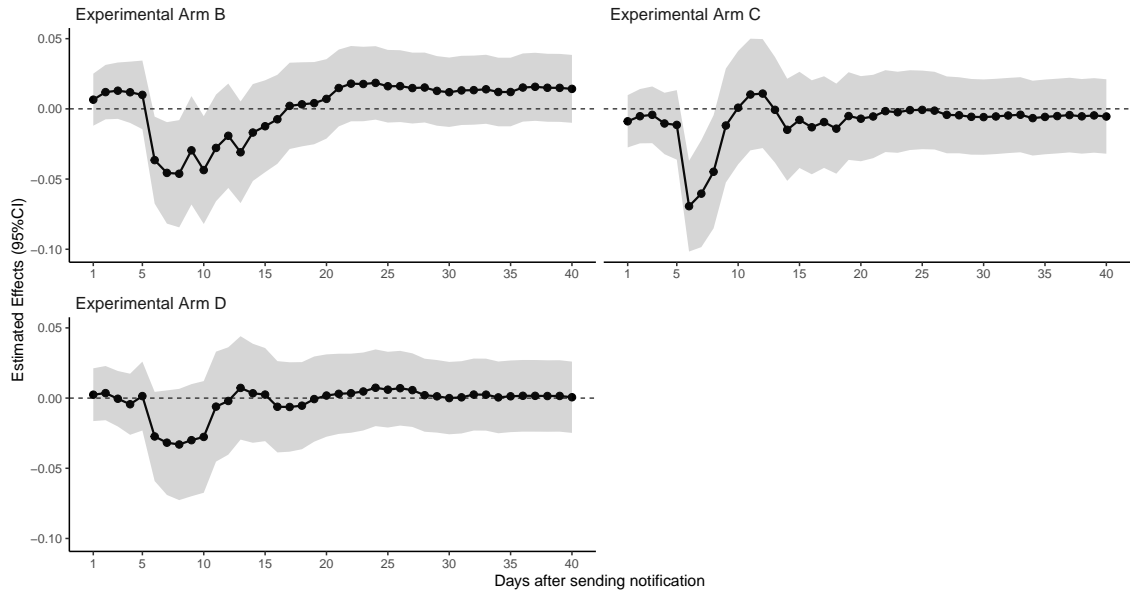


Figure 7: Effect on Reply within Specific Days after Sending Notification among Males More than 30. Notes: These plots show the average effect (and associated 95 percent confidential interval) on cumulative responses on specific day. We use robust standard errors. We control number of past coordinations, 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.

Table 10: Conditional Average Treatment Effect Estimated by RCF

	Female		Male	
	Age < 30	30 ≤ Age	Age < 30	30 ≤ Age
	(1)	(2)	(3)	(4)
B	-0.0026 (0.0476)	-0.0048 (0.0286)	0.1231*** (0.0405)	0.0160 (0.0214)
C	0.0354 (0.0480)	-0.0125 (0.0296)	0.0420 (0.0412)	-0.0146 (0.0225)
D	-0.0608 (0.0482)	0.0133 (0.0290)	0.0446 (0.0405)	0.0186 (0.0221)

Notes: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Standard errors are in parentheses. 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.

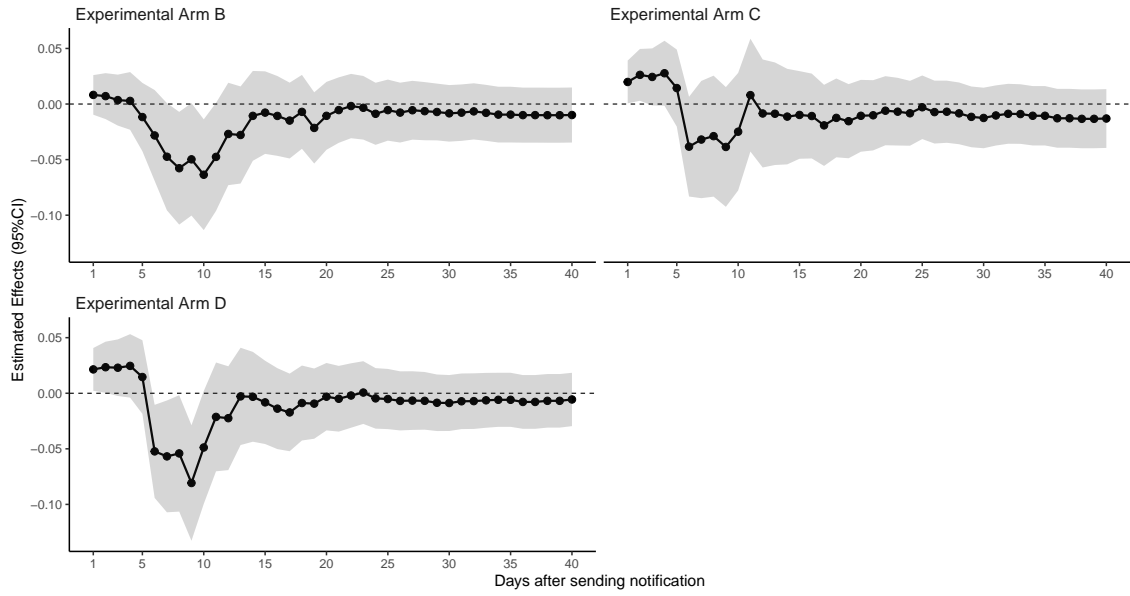


Figure 8: Effect on Reply within Specific Days after Sending Notification among Females More than 30. Notes: These plots show the average effect (and associated 95 percent confidential interval) on cumulative responses on specific day. We use robust standard errors. We control number of past coordinations, 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.

Table 11: Sample Characteristics of Predicted Treatment Effect of Arm B (Middle-aged Males)

	Predicted treatment effect		
	Effect $\leq 0.1$	$0.1 < \text{Effect}$	P-value
	(1)	(2)	(3)
Age	35.676	37.516	$< 0.001$
Number of past coordinations	1.648	2.813	$< 0.001$
Number of listed hospitals	0.393	1.526	$< 0.001$
Number of hospitals listed with PBSC collection	0.132	0.533	$< 0.001$
Number of hospitals listed with BM collection	0.197	0.909	$< 0.001$
N	1842	91	

Notes: Column (1) and (2) show average sample characteristics. Column (3) shows p-values of difference-in-means test. We use males with 30–39 age.

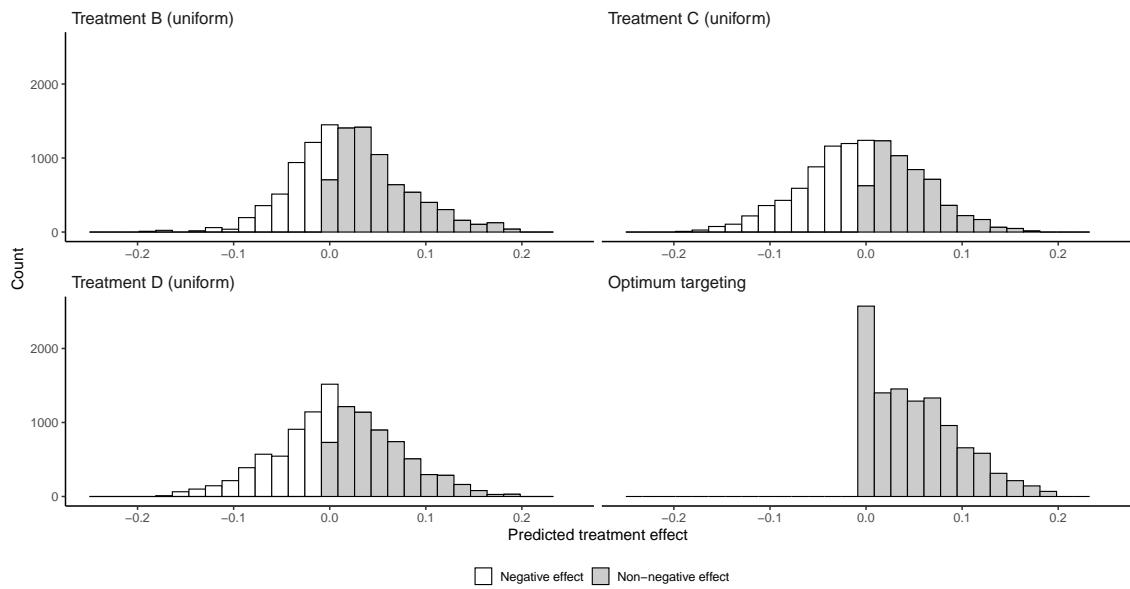


Figure 9: Distribution of Predicted Treatment Effect

## References

- Andreoni, J. (1990), ‘Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving’, *The Economic Journal* **100**(401), 464.
- Athey, S. and Imbens, G. (2016), ‘Recursive partitioning for heterogeneous causal effects’, *Proceedings of the National Academy of Sciences* **113**(27), 7353–7360.
- Athey, S., Tibshirani, J. and Wager, S. (2019), ‘Generalized random forests’, *The Annals of Statistics* **47**(2).
- Athey, S. and Wager, S. (2019), ‘Estimating Treatment Effects with Causal Forests: An Application’, *Observational Studies* **5**(2), 37–51.
- Bergstrom, T. C., Garratt, R. J. and Sheehan-Connor, D. (2009), ‘One Chance in a Million: Altruism and the Bone Marrow Registry’, *American Economic Review* **99**(4), 1309–1334.
- Carlana, M., La Ferrara, E. and Pinotti, P. (2022), ‘Goals and Gaps: Educational Careers of Immigrant Children’, *Econometrica* **90**(1), 1–29.
- Davis, J. M. and Heller, S. B. (2017), ‘Using Causal Forests to Predict Treatment Heterogeneity: An Application to Summer Jobs’, *American Economic Review* **107**(5), 546–550.
- Diekmann, A. (1985), ‘Volunteer’s Dilemma’, *Journal of Conflict Resolution* **29**(4), 605–610.
- Diekmann, A. (1986), Volunteer’s Dilemma. A Social Trap without a Dominant Strategy and some Empirical Results, in A. Diekmann and P. Mitter, eds, ‘Paradoxical Effects of Social Behavior’, Physica Heidelberg, Heidelberg, pp. 187–197.
- Franzen, A. (1999), The Volunteer’s Dilemma: Theoretical Models and Empirical Evidence, in M. Foddy, M. Smithson, S. Schneider and M. A. Hogg, eds, ‘Resolving Social Dilemmas: Dynamic, Structural, and Intergroup Aspects’, 1 edn, Psychology Press, New York.
- Haylock, M., Kampkötter, P., Macis, M., Sauter, J., Seitz, S., Slonim, R., Wiesen, D. and Schmidt, A. (2022), Improving the Availability of Unrelated Stem Cell Donors: Evidence from a Major Donor Registry, Technical Report 29857, National Bureau of Economic Research, Cambridge, MA.
- Hirakawa, T., Kurosawa, S., Tajima, K., Yamazaki, Y., Ikeda, N., Kojima, H., Tanaka, H., Kanamori, H., Miyamura, K., Kodera, Y., Fukuda, T. and Japan Marrow Donor Program (2018),

- ‘Japan Marrow Donor Program and its coordinating process: current situations’, *The Japanese Journal of Clinical Hematology* **59**(2), 153–160.
- Kollman, C., Spellman, S. R., Zhang, M.-J., Hassebroek, A., Anasetti, C., Antin, J. H., Champlin, R. E., Confer, D. L., DiPersio, J. F., Fernandez-Viña, M., Hartzman, R. J., Horowitz, M. M., Hurley, C. K., Karanes, C., Maiers, M., Mueller, C. R., Perales, M.-A., Setterholm, M., Woolfrey, A. E., Yu, N. and Eapen, M. (2016), ‘The effect of donor characteristics on survival after unrelated donor transplantation for hematologic malignancy’, *Blood* **127**(2), 260–267.
- Kurosawa, S., Tomine, Y., Yoshiuchi, K., Fukuda, T. and Japan Marrow Donor Program (2022), ‘An interview survey of psychosocial factors in unrelated stem cell donor candidates’, *Japanese Journal of Transplantation and Cellular Therapy* **11**(3), 169–176.
- Lacetera, N., Macis, M. and Stith, S. S. (2014), ‘Removing financial barriers to organ and bone marrow donation: The effect of leave and tax legislation in the U.S.’, *Journal of Health Economics* **33**, 43–56.
- Laibson, D. (1997), ‘Golden Eggs and Hyperbolic Discounting’, *The Quarterly Journal of Economics* **112**(2), 443–478.
- Murakami, K., Shimada, H., Ushifusa, Y. and Ida, T. (2022), ‘Heterogeneous Treatment Effects of Nudge and Rebate: Causal Machine Learning Field Experiment on Electricity Conservation’, *International Economic Review* p. iere.12589.
- O’Donoghue, T. and Rabin, M. (2001), ‘Choice and Procrastination’, *The Quarterly Journal of Economics* **116**(1), 121–160.
- Switzer, G., Dew, M., Stukas, A., Goycoolea, J., Hegland, J. and Simmons, R. (1999), ‘Factors associated with attrition from a national bone marrow registry’, *Bone Marrow Transplantation* **24**, 313–319.
- Switzer, G. E., Dew, M. A., Goycoolea, J. M., Myaskovsky, L., Abress, L. and Confer, D. L. (2004), ‘Attrition of Potential Bone Marrow Donors at Two Key Decision Points Leading to Donation’, *Transplantation* **77**(10), 1529–1534.
- Switzer, G. E., Macis, M., Fabi, R., Abress, L., Confer, D., Bruce, J., Howe, K., Fowler, S., McNulty, M., Pastorek, G. and Dew, M. A. (2018), ‘Providing Level-of-Match Information to Perfectly Matched Unrelated Stem Cell Donors: Evaluating Acceptability and Potential Changes in Donor Availability’, *Biology of Blood and Marrow Transplantation* **24**(10), 2110–2118.

Takanashi, M. (2016), ‘Marrow donor registration and cord blood banking: current issues’, *The Japanese Journal of Clinical Hematology* **57**(3), 271–277.

The Japanese Data Center for Hematopoietic Cell Transplantation/The Japanese Society for Transplantation and Cellular Therapy (2022), Hematopoietic Cell Transplantation in Japan. Annual Report of Nationwide Survey 2022., Technical report.

Wager, S. and Athey, S. (2018), ‘Estimation and Inference of Heterogeneous Treatment Effects using Random Forests’, *Journal of the American Statistical Association* **113**(523), 1228–1242.

Wildman, J. and Hollingsworth, B. (2009), ‘Blood donation and the nature of altruism’, *Journal of Health Economics* **28**(2), 492–503.