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

Hiroki Kato (Hitostubashi Institute for Advanced Study, Hitotsubashi University, Tokyo, Japan)

Fumio Ohtake (Center for Infectious Disease Education and Research (CiDER), Osaka University, Osaka, Japan)

Saiko Kurosawa (Department of Oncology, Ina Central Hospital, Nagano, Japan)

Kazuhiro Yoshiuchi (Graduate School of Medicine, Tokyo University, Tokyo, Japan)

Takahiro Fukuda (Department of Hematopoietic Stem Cell Transplantation, National Cancer Center Hospital, Tokyo, Japan)

Last updated on June 02, 2023

# 1 Introduction

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

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

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

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

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

# 2 Field Experiment

## 2.1 Background: Coordination Process of JMDP

フィールド実験の介入のタイミングの理解を促すために、JMDPに登録した潜在ドナーによる幹細胞の提供までのコーディネーション過程を概観しておく。はじめに、潜在ドナーがJMDPに登録した患者とマッチすると、JMDP事務局は潜在ドナーに幹細胞提供を依頼する適合通知（compatibility notice）を送付する[[3]](#footnote-23)。潜在ドナーは問診票を記入し、提供意思の有無を示して、適合通知に返信する。

その後、移植に向けたコーディネーションが始まる。潜在ドナーはおよそ1カ月以内に確認検査（confirmatory typing）を受ける。この段階では、コーディネーターは提供手順の詳細（骨髄採取もしくは末梢血幹細胞採取）を説明し、潜在ドナーとその家族に提供の意向を調査する。潜在ドナーは二つの採取方法を選べる。加えて、コーディネーター医師が問診、検診、感染症の有無と血液型を調べる採血をする。これらのテストは潜在ドナーがJMDPの定める基準に満たしているかどうかを検査する。

患者は同時に最大10人の潜在ドナーとのコーディネーションを受けられる。患者の医師は確認検査を受けた潜在ドナーから最も適切な候補者を選ぶ。重要なこととして、潜在ドナーはマッチしている患者の情報（他の潜在ドナーと何人マッチしているかなど）を一切知ることができない。また、潜在ドナーがコーディネーターやコーディネーター医師からそのような情報を得ることもできない。

最も適切な候補者として選ばれた潜在ドナーはコーディネーターやコーディネーター医師からの説明を受けた後に、最終同意をする必要がある。同時に、ドナーの家族の代表者も採取に同意しなければならない。ただし、これ以降、潜在ドナーは意向を変えられない。最終同意後、潜在ドナーは術前検査や採取の準備を受けるために1週間程度の入院をする。この後、ドナーは幹細胞採取の手術を受ける。確認検査から採取までの時間はおよそ3–4カ月である。

## 2.2 Experimental Design



Figure 2.1: Intervention Messages

我々の実験はJMDPが潜在ドナーに幹細胞提供を依頼する適合通知の内容に介入を施す。図 2.1 は我々の介入を示す。適合通知は提供ドナーが7日以内に返信するべきことを注記している。また、JMDPは適合通知と一緒に前小節で概観したコーディネーション過程を記述したハンドブックを同封している。

我々はコーディネーションを促進するために適合通知に二つのメッセージを加えた[[4]](#footnote-29)。確率メッセージはJMDPに登録した患者あたりの潜在ドナーの数が少ないことを強調している。一人以上の潜在ドナーが一人の患者とマッチするので、幹細胞提供は公共財の性質を有している。標準的な経済学理論は潜在ドナーがただ乗りするインセンティブを持っており、提供のやる気を削いでしまうことを提示している。また、貢献者の数が多いほど、ただ乗りするインセンティブは強くなるはずである。これまでの実証研究はこの仮説に対して混在したエビデンスを示している。我々は潜在ドナーが他の潜在ドナーの数について過度に多く見積もっており、標準的な経済学理論に従ったただ乗り行動を取っていると考えた。よって、確率メッセージは他の潜在ドナーの数の信念を下方修正することを通じて、提供意向を高めることを目的としている。

第二のメッセージはEarly Coordinationメッセージである。このメッセージは登録患者の半数程度しか移植を受けられないことを強調して、潜在ドナーの利他性を刺激している。さらに、このメッセージは適切な潜在ドナーが早く見つかるほど、患者の生存率が高くなることを伝えている。

我々は二つの介入メッセージの効果を推定するために四つの実験群を設けた。実験群Aは介入メッセージなしの標準的な適合通知を送付した（コントロール群）。実験群BとCはそれぞれ確率メッセージとEarly Coordinationメッセージを追加した適合通知を送付した。実験群Dは二つの介入メッセージを同時に加えた適合通知を送付した。この実験群は情報過多による認知負荷の負の影響を検証する目的で設けた。

Table 2.1: 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 |

実験参加者は2021年9月から2022年2月の間に適合通知を受け取った11,154名の潜在ドナーである。JMDP事務局に無理のない範囲でランダムネスを保つために、我々は実験群を週単位のクラスターランダム化で割り当てた。このとき、可能な限り週と月で実験群がバランスするように設計している。表 2.1 に割り当てスケジュールをまとめている。

## 2.3 Data and Empirical Strategy

我々はJMDPの協力のもと、コーディネーションデータを2022年6月末に受け取った。観測単位は実験参加者である。個人属性として、データは性別・年齢・コーディネーション経験回数・都道府県レベルの在住地区を記録している。コーディネーション過程として、データは各段階（適合通知への返信・確認検査・候補者選定・最終同意・採取）に到達したかどうかを記録している。我々はこれらの変数をアウトカム変数として用いる。特に、適合通知への返信について、データは返信日数と提供意向を追加的に記録している。コーディネーションが中断したならば、データは中断理由を三つのカテゴリ（患者側の理由・健康以外のドナー理由・ドナーの健康理由）で記録している。分析では、日本在住かつコーディネーションが（中断を含めて）終了している11,049名の潜在ドナーを用いる[[5]](#footnote-32)。

追加的なデータとして、我々はJMDPがウェブ上で公開している医療機関リストを用いる。このリストは完全な住所・骨髄採取（BM collection）が可能かどうか・末梢血幹細胞採取（PBSC collection）が可能かどうかを含んでいる。我々はこのリストを都道府県レベルで集約し、10平方キロメートルあたりの病院数を計算し、それをコーディネーションデータと都道府県をマージキーにして突合した。我々はこの変数をコーディネーションや提供のトラベルコストとして考慮する。

Table 2.2: Overview of Field Experiment

|  | Experimental Arms | | | |  |
| --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | F-test, p-value |
| A. Interventions | | | | | |
| Standard notification | X | X | X | X |  |
| Probability message |  | X |  | X |  |
| Patients message |  |  | X | X |  |
| B. Sample Size | | | | | |
| N | 2535 | 3053 | 2726 | 2735 |  |
| C. Balance Test | | | | | |
| Male (=1) | 0.624 | 0.633 | 0.631 | 0.609 | 0.362 |
| 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 |
| Notes: Balance test regresses a covariate on treatment dummies and test a null hypothesis that all coefficients are zero. We use the clustered standard error with the CR2 adjustment for inference (cluster unit is an experimental week). | | | | | |

表 2.2 はフィールド実験の概要である。パネルAは各実験群の介入を示しており、パネルBは各実験群のサンプルサイズを示している。パネルCはランダム化が成功しているかどうかを検証するバランステストである。多くの変数が実験群間で平均的にバランスしているので、実験群の割り当てはおおよそランダムである。

我々の関心のあるパラメータは母集団全体の平均処置効果であり、実験群間の平均値の差で識別できる。それに加えて、我々は月・週固定効果を取り除き、共変量を加えることで推論の確度を高めるために、線形確率モデルを推定する。月の週に適合通知を受け取った個人についての推定モデルは以下の通りである。

ここでは個人属性ベクトル、とはそれぞれ月および週のダミー変数である。

# 3 Experimental Results

## 3.1 Effects on Reply and Intention

初めに、潜在ドナーの意向を最も反映していると考えられる返信への介入効果を推定する。返信のアウトカム変数は提供の意向に関わらず適合通知に返信したならば1を取るダミー変数である。コントロール群（実験群A）の返信率は%である。

我々は返信へのメッセージ効果を二つに分解する。第一に、提供意向を伴った返信への効果である。アウトカム変数は潜在ドナーが適合通知に返信し、かつ提供意思を示したならば1を取るダミー変数である。第二に、提供意向を伴わない返信に対する効果である。アウトカム変数は潜在ドナーが適合通知に返信し、かつ提供意思を示していないならば1を取るダミー変数である。ただし、これらの効果を推定するとき、我々は未返信者のアウトカム変数を0とコーディングして、サンプルに含める。二つのアウトカム変数の和は必ず返信のダミー変数となるので、正の意向と負の意向に対する効果の和は返信に対する効果となる。コントロール群では、正の意向を伴う返信率は%である一方で、負の意向を伴う返信率は%であった。したがって、返信者の%は提供に積極的である。

潜在ドナーが返信する前に、患者側の都合でコーディネーションが中断してしまう可能性がある。これは潜在ドナーの意向と独立に生じるので、我々は各実験群でそのようなケースに該当する–%を分析サンプルから除外している。ただし、このケースは実験群間でバランスしている（F-value, p-value = 0.643）。

Table 3.1: Linear Probability Model of Reply and Intention

|  | | | Intention | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | Reply | | Positive | | Negative | |
|  | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment B | 0.0105 | 0.0139\*\* | 0.0218 | 0.0210 | -0.0113 | -0.0071 |
|  | (0.0075) | (0.0059) | (0.0135) | (0.0123) | (0.0094) | (0.0076) |
| Treatment C | -0.0057 | 0.0026 | -0.0053 | -0.0029 | -0.0004 | 0.0054 |
|  | (0.0081) | (0.0047) | (0.0151) | (0.0112) | (0.0150) | (0.0094) |
| Treatment D | 0.0058 | 0.0062 | 0.0046 | 0.0058 | 0.0012 | 0.0004 |
|  | (0.0074) | (0.0052) | (0.0166) | (0.0101) | (0.0138) | (0.0069) |
| Control average | 0.8835 | 0.8835 | 0.5533 | 0.5533 | 0.3303 | 0.3303 |
| Covariates |  | X |  | X |  | X |
| Num.Obs. | 10985 | 10985 | 10985 | 10985 | 10985 | 10985 |
| Notes: \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01. The clustered standard errors with the CR2 adjustment are reported in parenetheses (cluster unit is an experimental week). Covariates are gender, squared polynomial of (demeaned) age, number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, number of hospitals with BM collection per 10 square kilometers, month dummies, and week dummies. | | | | | | |

表 3.1 は線形確率モデルの推定結果である。奇数列は実験群ダミーだけを説明変数に入れており、単純な平均値の差に対応している。偶数列は実験群ダミーに加えて、個人属性・月および週の固定効果をコントロールしている。推定結果はすべての実験群は、共変量の有無に関わらず、返信や意向に対して統計的に有意な効果を持っていない。我々は代替手法としてロジットモデルを推定（表 3.4 ）し、同様の結果を得ている（オッズ比の95%信頼区間が1を含んでいる）。

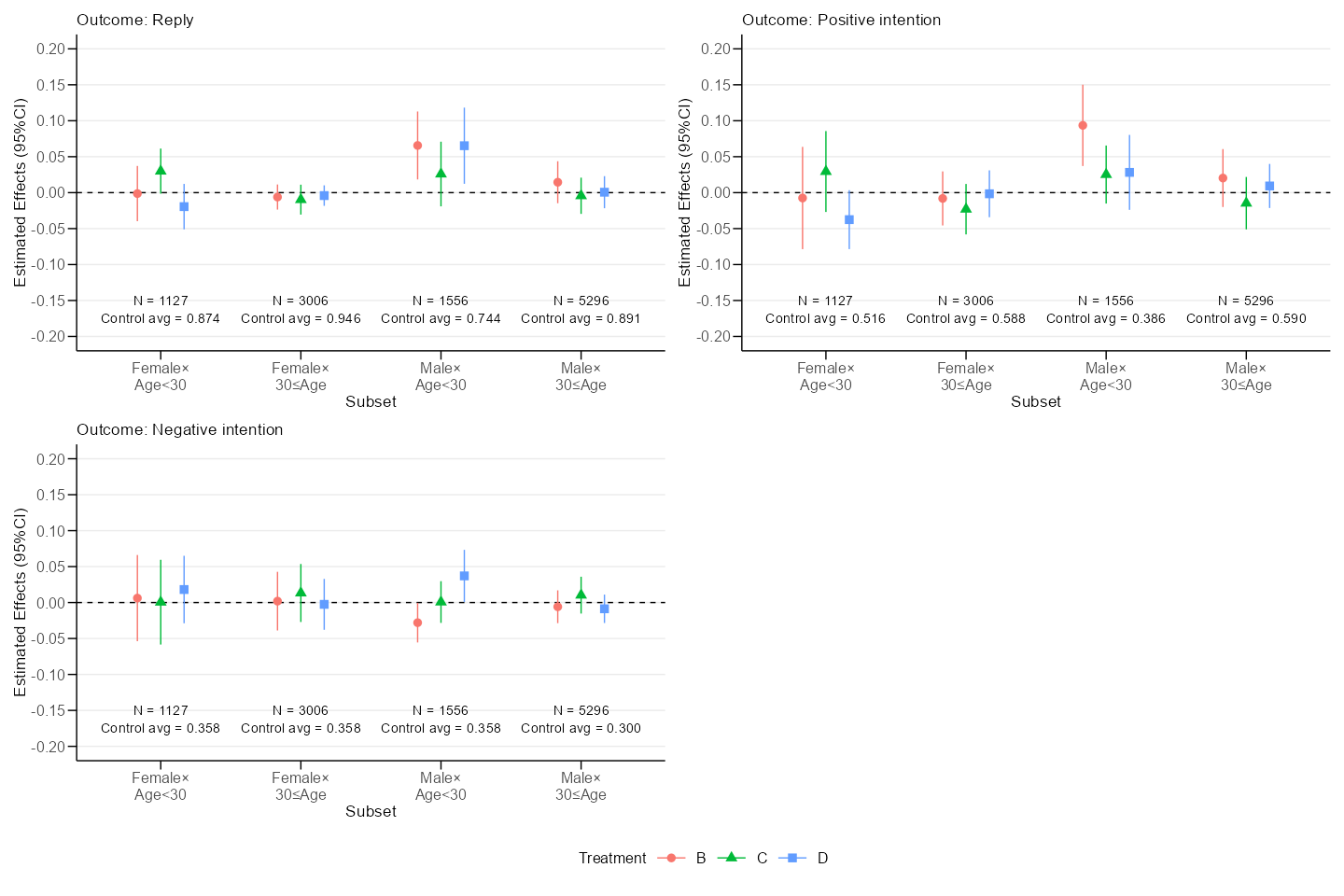


Figure 3.1: Effect on Reply and Intentions by Gender and Age Group. Note: These plots show the average effect (and associated 95% confidential interval) on each outcome by gender and age group. We use clustered standard errors with the CR2 adjustment (cluster unit is experimental week). We control number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, number of hospitals with BM collection per 10 square kilometers, month dummies, and week dummies.

メッセージ効果の異質性を検証するために、我々はサンプルを性別と年齢層（30歳未満か否か）で4つのサブセットに分割し、各サブセットでメッセージ効果を推定する。図 3.1 は係数プロットである。結果として、若い男性グループで、確率メッセージのみを加えた実験群Bが正の意向を伴う返信を約10ポイント（コントロール平均が%なので、%増加）増やしており、これは統計的に有意である。実験群Bはこのグループにおける提供意向を伴わない返信率を下げているので、全体の返信率に対する効果は6ポイントであり、正の意向を伴う返信に対する効果よりも小さい。他の性・年代と比較して、移植成績が良いにも関わらず若い男性の提供意欲は低いことを考慮すれば、確率メッセージはコーディネーションの効率性を改善しているといえる。ただし、他の性・年代グループで、我々の介入は返信や意向に対して統計的に有意な効果をもたない。

## 3.2 Random Causal Forest

介入効果の異質性を深堀りするために、我々は完全にノンパラメトリックに介入効果の異質性を検証できるrandom causal forest (RCF)を用いる(Athey and Imbens 2016; Wager and Athey 2018)。この手法は回帰木アルゴリズムに基づいており、観察可能な特徴で条件づけた平均トリートメント効果を推定する。この手法は労働(Davis and Heller 2017)・教育(Carlana, La Ferrara, and Pinotti 2022)・節電(Murakami et al. 2022)などの広い文脈で用いられている。

このアルゴリズムは研究者によって与えられた共変量の一つを用いて、サンプルを二つのサブサンプル（リーフ）に分割する。具体的に、このアルゴリズムはであるサンプルとであるサンプルに分割する。回帰木はアウトカム変数の平均二乗誤差（誤差はあるリーフに属するサンプルのアウトカムの平均値と観察されるアウトカムの差である）を最小にするように特定の閾値を決める。RCFはあるリーフに属する平均処置効果（条件付き平均処置効果）を推定することが目的なので、予測処置効果の期待平均二乗誤差を最小にするように閾値を決める。この最小化はリーフ間の条件付き平均効果の分散（異質性）を高めかつ、リーフ内の分散を抑えることと同じである。これを終端条件に達するまで各リーフで二つのリーフに分けるプロセスを繰り返す。RCFは終端のリーフにおける条件付き平均処置効果を予測する。

回帰木アルゴリズムの欠点は予測値の分散が大きく（過学習）なり、予測精度が落ちることである。これを防ぐために、RCFはアンサンブル法と呼ばれるアルゴリズムを導入している。これは全データから重複なしのサブサンプルを数千個作り、各サブサンプルで回帰木を育てる。最終的な結果は数千個の回帰木の予測値の平均とする[[6]](#footnote-42)。

大雑把に言えば、RCFは観察される個人属性で介入の処置効果を予測する。すなわち、介入を受けていない人についても、その人の属性に基づいて処置効果を予測できる。ただし、この手法は処置の割り当てが潜在成果変数と条件つきにランダムであることを仮定してなければならない。実験群はおおよそランダムに割り当てられているので、我々はこの手法を用いることができる。この小節では、意向を伴う返信への効果に着目する。RCFに用いる共変量は性別・年齢・過去のコーディネーション経験・在住地域の病院の密度である。

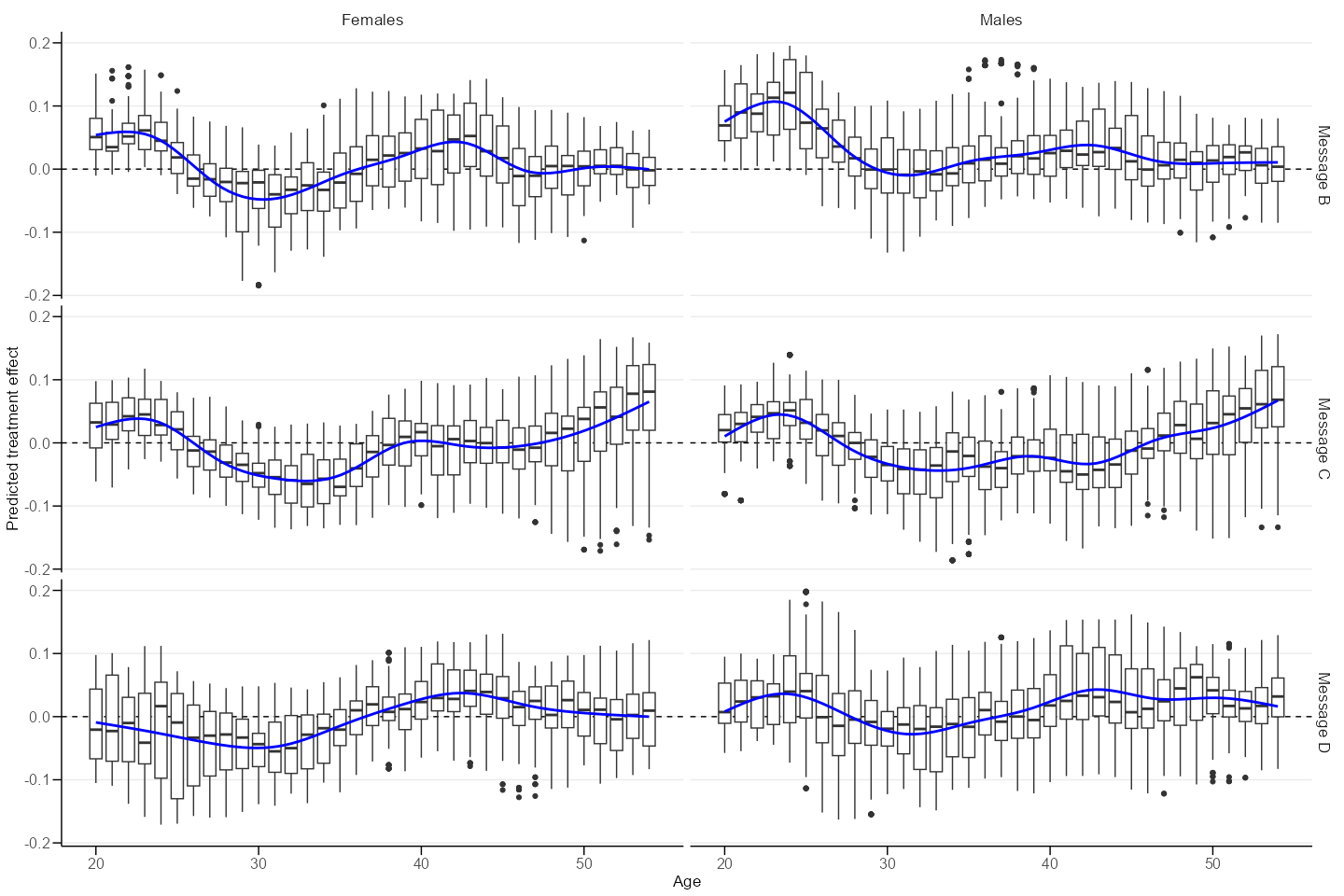


Figure 3.2: Boxplot of Predicted Treatment Effects by Gender and Age. Notes: Blue fitted line represents GAM smoothing. Covariates are gender, age, number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers and number of hospitals with BM collection per 10 square kilometers.

図3.2は意向を伴う返信に対する各実験群の予測トリートメント効果の分布を年齢別にプロットしている。注目するべきこととして、25歳以下の男性のすべてに対して、実験群Bの効果は正である。結果として、RCFは30歳以下の男性における実験群Bの平均効果は12.1ポイントであり、図??で示した線形確率モデルのサブサンプル分析の結果と近い値である（表3.8参照）。また、30–40代男性の一部に対して、実験群Bの効果は10ポイント以上とである。そのような男性は相対的に病院の数が多い地域に住んでおり、コーディネーションや提供のトラベリングコストが低いことがわかっている。

Early coordinationメッセージのみを加えた実験群Cの効果は大きく異質的である。30代前半の女性に対して、ほとんどのトリートメント効果は負である。40代後半以上の男女に対して、処置効果の中央値は正であるが広く分布している。驚くべきことに、この年齢層の男女において、実験群Cが正の効果を持つ人は相対的に病院の数が少ない（トラベリングコストが高い）地域に住んでいることがわかっている。

Table 3.2: Correlation of Predicted Treatment Effects

|  | Treatment effect D | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Females | | Males | | Females | | Males | |
|  | Age<30 | 30≤Age | Age<30 | 30≤Age | Age<30 | 30≤Age | Age<30 | 30≤Age |
|  | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Constant | -0.0385\*\*\* | 0.0075\*\*\* | -0.0143\*\*\* | 0.0141\*\*\* | -0.0354\*\*\* | 0.0074\*\*\* | 0.0061\*\*\* | 0.0101\*\*\* |
|  | (0.0019) | (0.0007) | (0.0022) | (0.0007) | (0.0018) | (0.0008) | (0.0021) | (0.0007) |
|  | [-0.0422, -0.0348] | [0.0061, 0.0090] | [-0.0186, -0.0099] | [0.0128, 0.0154] | [-0.0388, -0.0319] | [0.0059, 0.0090] | [0.0019, 0.0103] | [0.0087, 0.0114] |
| B + C | 0.3110\*\*\* | 0.3553\*\*\* | 0.3565\*\*\* | 0.4221\*\*\* |  |  |  |  |
|  | (0.0211) | (0.0073) | (0.0189) | (0.0071) |  |  |  |  |
|  | [0.2695, 0.3524] | [0.3409, 0.3697] | [0.3195, 0.3936] | [0.4081, 0.4361] |  |  |  |  |
| B |  |  |  |  | -0.0118 | 0.3633\*\*\* | -0.0936\*\*\* | 0.6048\*\*\* |
|  |  |  |  |  | (0.0349) | (0.0134) | (0.0269) | (0.0157) |
|  |  |  |  |  | [-0.0802, 0.0567] | [0.3370, 0.3896] | [-0.1464, -0.0408] | [0.5740, 0.6355] |
| C |  |  |  |  | 0.7012\*\*\* | 0.3479\*\*\* | 0.9537\*\*\* | 0.3080\*\*\* |
|  |  |  |  |  | (0.0365) | (0.0126) | (0.0305) | (0.0105) |
|  |  |  |  |  | [0.6296, 0.7727] | [0.3232, 0.3726] | [0.8939, 1.0135] | [0.2874, 0.3285] |
| Num.Obs. | 1127 | 3006 | 1556 | 5296 | 1127 | 3006 | 1556 | 5296 |
| Notes: \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01. Robust standard errors are reported in parentheses. Square brackets show 95 percent confidential intervals. | | | | | | | | |

実験群Dは確率メッセージ（実験群B）とearly Coordinationメッセージ（実験群C）の両方を加えている。したがって、もし潜在ドナーが実験群Dの情報を完全に理解しているならば、実験群Dの効果は実験群BとCの効果の合計となるはずである。この点を検証するために、表??の列(1)–(4)は実験群Dの処置効果を実験群BとCの効果の合計で回帰している。どの性・年代でも実験群BとCの効果の合計が1ポイント増えると、実験群Dの効果は0.3–0.4ポイント増加する。95%信頼区間は1を含まないので、この増分は1ポイントより統計的に有意に低い。この結果は潜在ドナーが実験群Dの情報を完全に理解しておらず、情報過多による認知負荷になっていること示唆している。

完全に情報を受け取っていない潜在ドナーは確率メッセージもしくはearly coordinationメッセージのどちらかを重視するか、二つの情報を同程度割り引いて評価しているだろう。この点を検証するために、表??の列(4)–(8)は実験群Dの処置効果を実験群Bの効果と実験群Cの効果で回帰している。その結果、若年層の男女では、実験群CとDの偏相関は実験群BとDの偏相関よりも強い。すなわち、彼らは確率メッセージよりもearly coordinationメッセージを重視している。したがって、確率メッセージを提供しているにも関わらず、実験群Dが30歳未満の男性に効果がないのである。

興味深いことに、高齢層の男女では、実験群CとDの偏相関は実験群BとDの偏相関よりも弱い。すなわち、彼らはearly coordinationメッセージよりも確率メッセージを重視している。したがって、40代後半以上の男女における実験群Cの効果の中央値が8ポイント程度であるにも関わらず、実験群Dの処置効果の中央値はほとんどゼロである。

## 3.3 Response Speed to Notification

Early Coordinationメッセージはコーディネーション期間が長くなると、患者の生存確率を下げてしまうという事実を提供している。したがって、このメッセージは適合通知への早期返信を促している可能性がある。この可能性を検証するために、我々は特定の期間以内に返信する確率への効果を線形確率モデルで推定する。アウトカム変数は潜在ドナーが日以内に返信したならば1を取るダミー変数である。日以降に返信した・返信しなかった場合、アウトカム変数は0となる。の値を十分に大きくすれば、この分析は第3.1節の結果と一致する。さらに、第3.1節と同様に、我々はドナーの意向の視点から二つの効果に分解する。一つは提供意向を伴う日以内の返信への効果である。アウトカムは日以内に返信し、かつ提供の意向を示したならば1を取るダミー変数である。もう一つは提供意向を伴わない日以内の返信への効果である。アウトカムは日以内に返信し、かつ提供の意向を示さなければ1を取るダミー変数である。日以降に返信した・返信しなかった場合、これらのアウトカムも0とコーディングされる。我々はに焦点を当てる。

コントロール群では、返信者の平均的な返信日数は10日である。また、正もしくは負の意向を伴う返信の平均返信日数はそれぞれ9日・12日である。最長の返信日数は46日である。加えて、適合通知に目安として記載されている7日以内の返信は33%であり、正の意向を伴う7日以内の返信は23%である（未返信者を含む）。

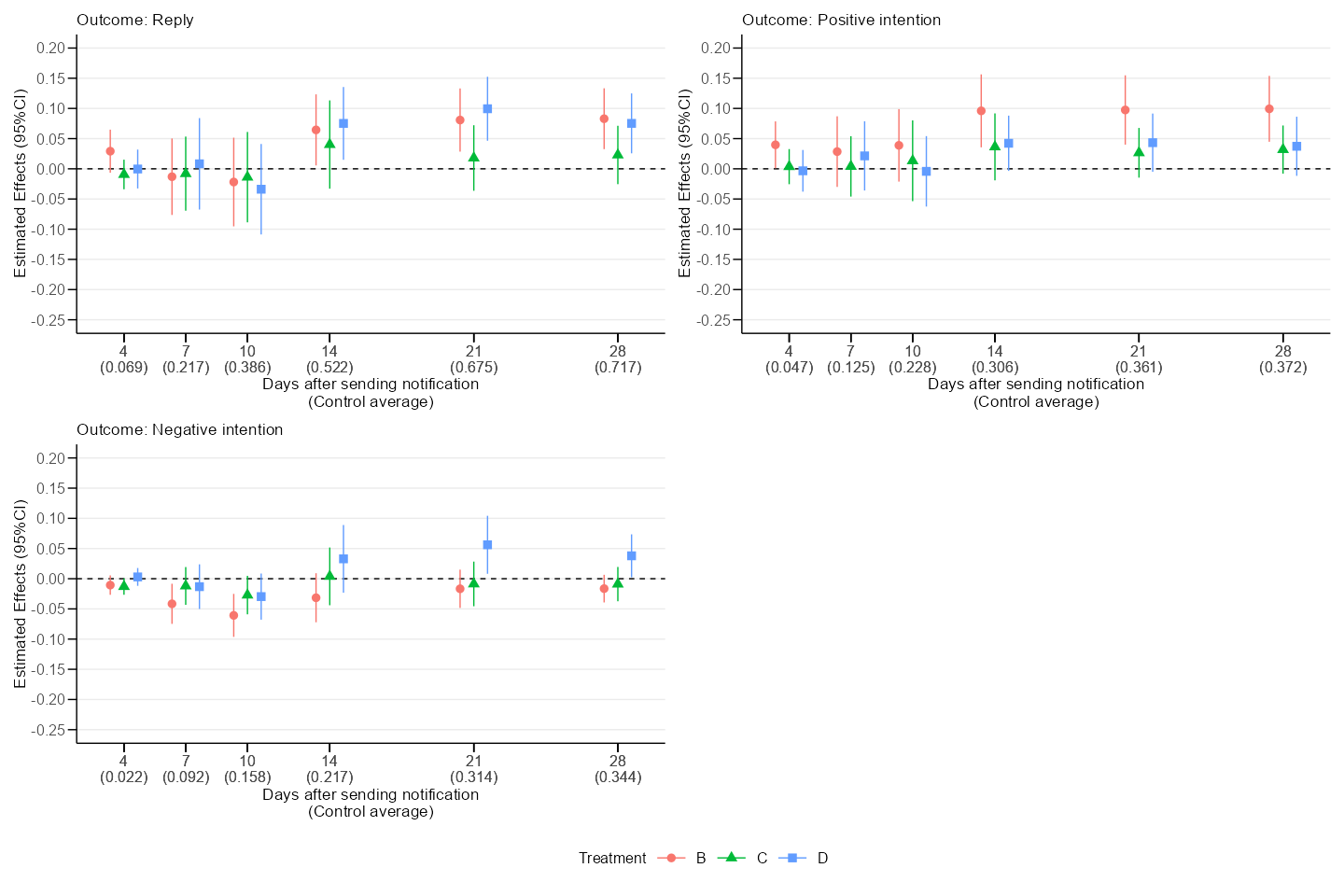


Figure 3.3: Effect on Reply within Specific Days after Sending Notification among Males Less than 30. Notes: These plots show the average effect (and associated 95% confidential interval) on each outcome. We use clustered standard errors with the CR2 adjustment (cluster unit is experimental week). We control number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, number of hospitals with BM collection per 10 square kilometers, month dummies, and week dummies.

先の小節で見たように返信や意向への効果は異質的であるので、この小節は性・年代でサブサンプルに分割して、返信日数への効果の異質性に注目する。図3.3は30歳未満の男性の結果である。このサブグループでは、我々は確率メッセージのみを加えた実験群Bが提供意向を伴う全体の返信率を10ポイント程度高めていたことをすでに明らかにしている。図3.3はその効果が14日以内の返信から生じていることを示唆している。実験群Bは提供意向を伴う10日以内の返信を（統計的に有意ではないが）5ポイント程度高めており、14日以内の返信を10ポイント程度高めている。すなわち、実験群Bは提供意向を伴う10日以内の返信を5ポイント、提供意向を伴う10–14日以内の返信を5ポイント高め、これらが全体の返信率を10ポイント高めていることにつながっている。また、このグループでは、早期返信を促している可能性のある実験群C（Early Coordinationメッセージ）が短期間の返信を高めているということはなかった。

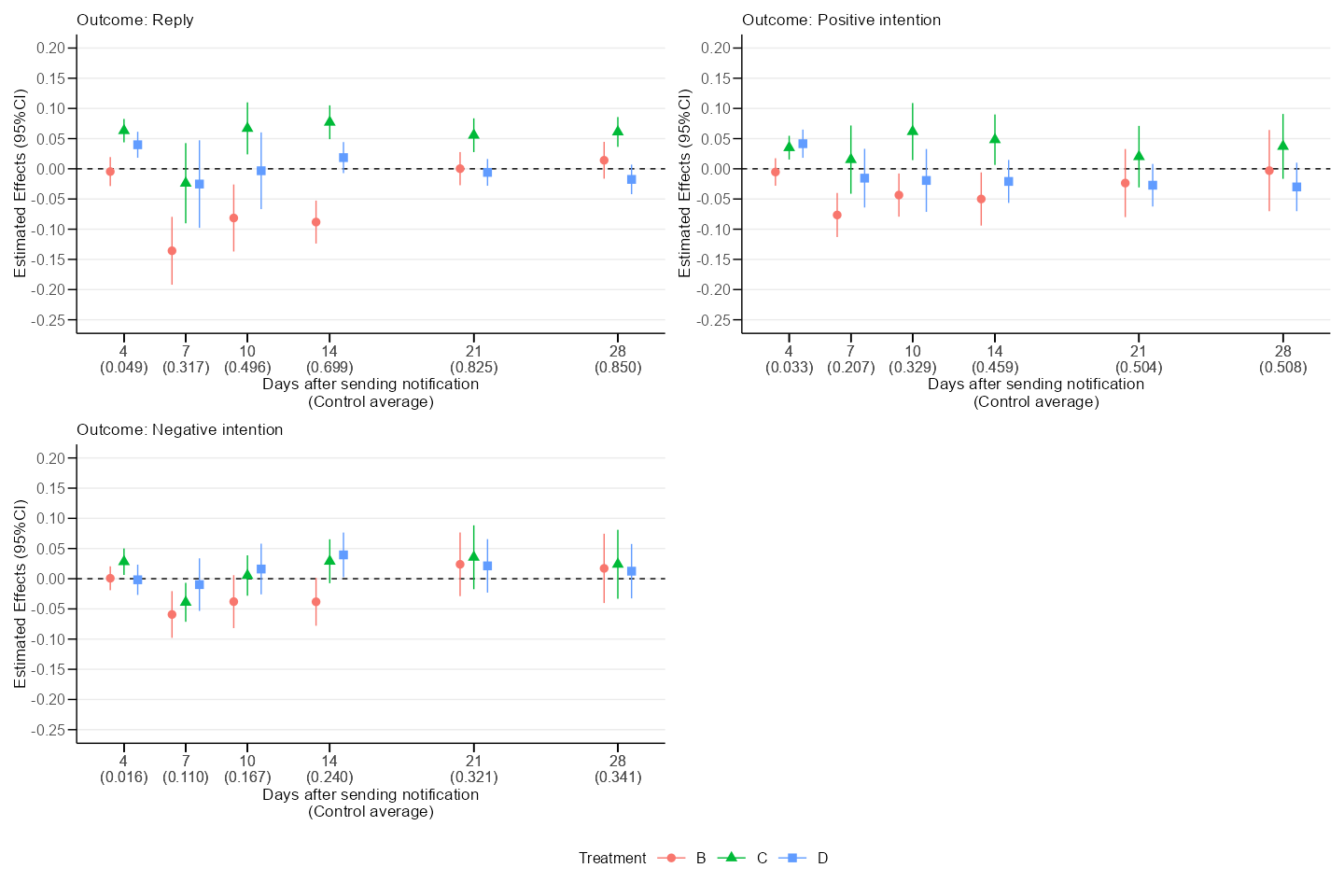


Figure 3.4: Effect on Reply within Specific Days after Sending Notification among Females Less than 30. Notes: These plots show the average effect (and associated 95% confidential interval) on each outcome. We use clustered standard errors with the CR2 adjustment (cluster unit is experimental week). We control number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, number of hospitals with BM collection per 10 square kilometers, month dummies, and week dummies.

しかしながら、実験群Cは30歳未満の女性の早期返信を促している可能性がある。図3.3は30歳未満の女性の結果である。図3.1より実験群Cは全体の返信率に統計的に有意な影響を与えていない。一方で、この実験群は4日以内の返信率を6ポイント程度高めており、これは5%水準で有意である。

## 3.4 Effects on the Coordination Process

Table 3.3: Linear Probability Model of Coordination Process

|  | CT | | Candidate | | Consent | | Donation | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Constant | 0.2350\*\*\* | 0.0797\*\*\* | 0.0779\*\*\* | 0.0079 | 0.0687\*\*\* | -0.0011 | 0.0574\*\*\* | 0.0055 |
|  | (0.0115) | (0.0152) | (0.0065) | (0.0099) | (0.0062) | (0.0092) | (0.0052) | (0.0072) |
| Treatment B | 0.0325\*\* | 0.0326\*\*\* | 0.0051 | 0.0033 | 0.0059 | 0.0037 | 0.0040 | 0.0020 |
|  | (0.0141) | (0.0073) | (0.0076) | (0.0050) | (0.0074) | (0.0041) | (0.0060) | (0.0043) |
| Treatment C | 0.0146 | 0.0136 | 0.0010 | -0.0023 | 0.0024 | -0.0011 | 0.0016 | -0.0017 |
|  | (0.0163) | (0.0087) | (0.0102) | (0.0057) | (0.0086) | (0.0044) | (0.0076) | (0.0047) |
| Treatment D | 0.0260 | 0.0299\*\*\* | 0.0084 | 0.0087\* | 0.0099 | 0.0098\* | 0.0030 | 0.0029 |
|  | (0.0161) | (0.0077) | (0.0093) | (0.0046) | (0.0102) | (0.0048) | (0.0095) | (0.0053) |
| Control average | 0.2350 | 0.2350 | 0.0779 | 0.0779 | 0.0687 | 0.0687 | 0.0574 | 0.0574 |
| Covariates |  | X |  | X |  | X |  | X |
| Num.Obs. | 10435 | 10435 | 8587 | 8587 | 8558 | 8558 | 8441 | 8441 |
| Notes: \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01. The clustered standard errors with the CR2 adjustment are reported in parenetheses (cluster unit is experimental weeks). Covariates are gender, squared polynomial of (demeaned) age, number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, number of hospitals with BM collection per 10 square kilometers, month dummies, and week dummies. | | | | | | | | |

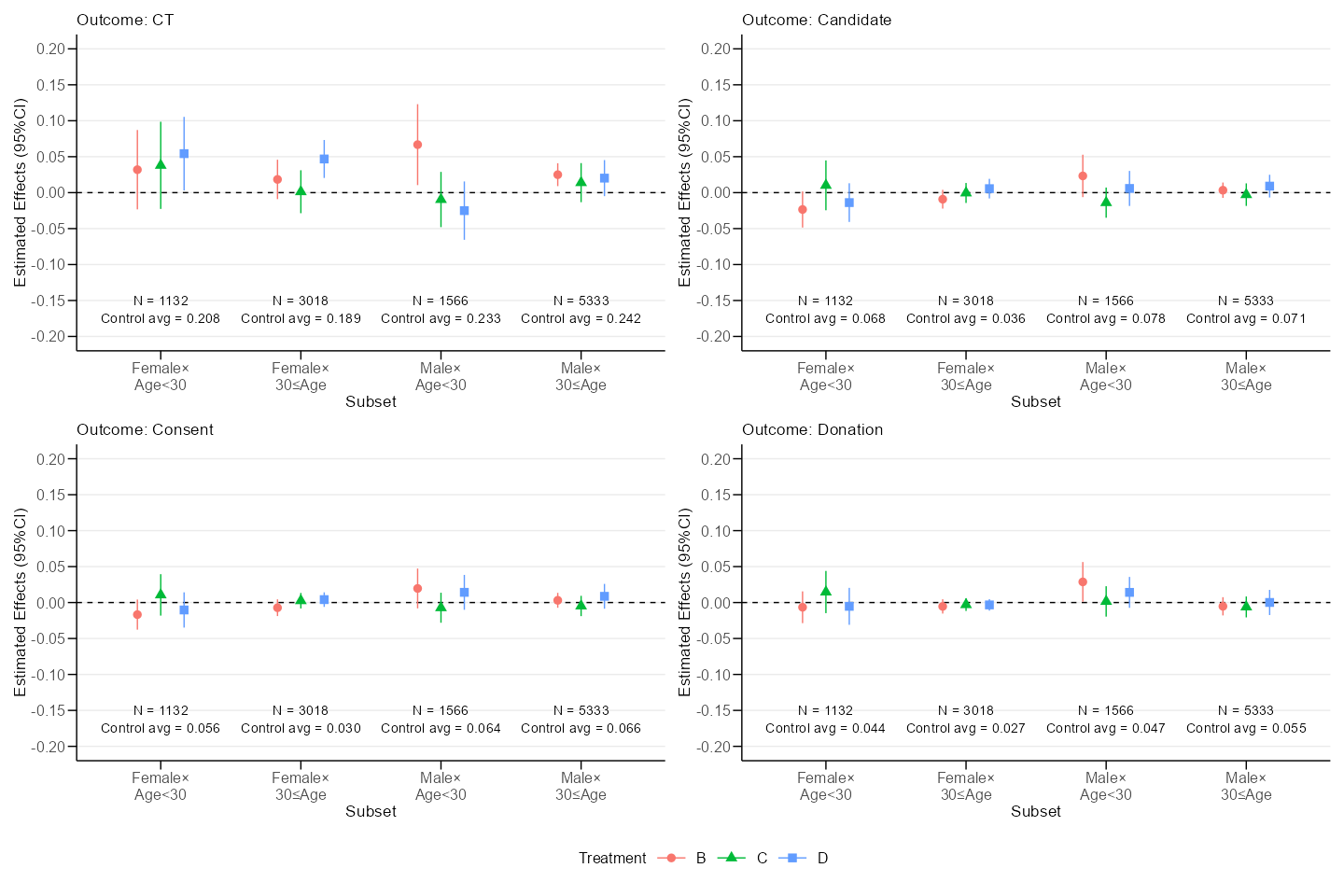


Figure 3.5: Effect on Coordination by Gender and Age Group. Note: These plots show the average effect (and associated 95% confidential interval) on each outcome by gender and age group. We use clustered standard errors with the CR2 adjustment (cluster unit is experimental week). We control number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, number of hospitals with BM collection per 10 square kilometers, month dummies, and week dummies.

# Appendix

Table 3.4: Logit Model of Reply and Intention

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

Table 3.5: Logit Model of Coordination

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

Table 3.6: Linear Probability Model of Reply and Intentions for Males Less Then 30 Years

|  | | | Intention | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | Reply | | Positive | | Negative | |
|  | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment B | 0.0661\*\* | 0.0656\*\*\* | 0.0935\*\* | 0.0936\*\*\* | -0.0274 | -0.0280\*\* |
|  | (0.0237) | (0.0228) | (0.0346) | (0.0273) | (0.0182) | (0.0133) |
| Treatment C | 0.0197 | 0.0259 | 0.0218 | 0.0252 | -0.0021 | 0.0007 |
|  | (0.0258) | (0.0217) | (0.0314) | (0.0195) | (0.0230) | (0.0140) |
| Treatment D | 0.0620\*\* | 0.0653\*\* | 0.0198 | 0.0282 | 0.0422 | 0.0371\*\* |
|  | (0.0260) | (0.0256) | (0.0435) | (0.0252) | (0.0299) | (0.0175) |
| Control average | 0.7821 | 0.7821 | 0.4216 | 0.4216 | 0.3605 | 0.3605 |
| Covariates |  | X |  | X |  | X |
| Bootstrap, p-value |  |  |  |  |  |  |
| B = 0 | 0.0380 | 0.0900 | 0.0510 | 0.0150 | 0.2720 | 0.1840 |
| D = 0 | 0.0440 | 0.1000 | 0.6860 | 0.4350 | 0.2470 | 0.1740 |
| Num.Obs. | 1556 | 1556 | 1556 | 1556 | 1556 | 1556 |
| Notes: \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01. The clustered standard errors with the CR2 adjustment are reported in parenetheses (cluster unit is experimental weeks). Covariates are number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, number of hospitals with BM collection per 10 square kilometers, month dummies, and week dummies. Bootstrap p-value is based on the wild cluster bootstrap-t with a null hypothesis imposed, which is proposed by Cameron et al. (2008). | | | | | | |

Table 3.7: Linear Probability Model of Coordination among Males

|  | CT | | Candidate | | Consent | | Donation | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Treatment B | 0.0513 | 0.0668\*\* | 0.0203 | 0.0233 | 0.0174 | 0.0196 | 0.0269\*\* | 0.0287\*\* |
|  | (0.0307) | (0.0272) | (0.0169) | (0.0143) | (0.0139) | (0.0134) | (0.0129) | (0.0134) |
| Treatment C | -0.0254 | -0.0096 | -0.0166 | -0.0139 | -0.0076 | -0.0072 | 0.0017 | 0.0016 |
|  | (0.0318) | (0.0186) | (0.0129) | (0.0102) | (0.0114) | (0.0101) | (0.0122) | (0.0102) |
| Treatment D | -0.0252 | -0.0250 | 0.0075 | 0.0058 | 0.0161 | 0.0142 | 0.0141 | 0.0141 |
|  | (0.0342) | (0.0196) | (0.0152) | (0.0118) | (0.0155) | (0.0117) | (0.0130) | (0.0104) |
| Control average | 0.2337 | 0.2337 | 0.0805 | 0.0805 | 0.0702 | 0.0702 | 0.0581 | 0.0581 |
| Covariates |  | X |  | X |  | X |  | X |
| Bootstrap, p-value |  |  |  |  |  |  |  |  |
| B = 0 | 0.3140 | 0.1080 | 0.5140 | 0.2830 | 0.4440 | 0.3360 | 0.1400 | 0.2220 |
| Num.Obs. | 1566 | 1566 | 1566 | 1566 | 1566 | 1566 | 1566 | 1566 |
| Notes: \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01. The clustered standard errors with the CR2 adjustment are reported in parenetheses (cluster unit is experimental weeks). Covariates are number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, number of hospitals with BM collection per 10 square kilometers, month dummies, and week dummies. Bootstrap p-value is based on the wild cluster bootstrap-t with a null hypothesis imposed, which is proposed by Cameron et al. (2008). | | | | | | | | |

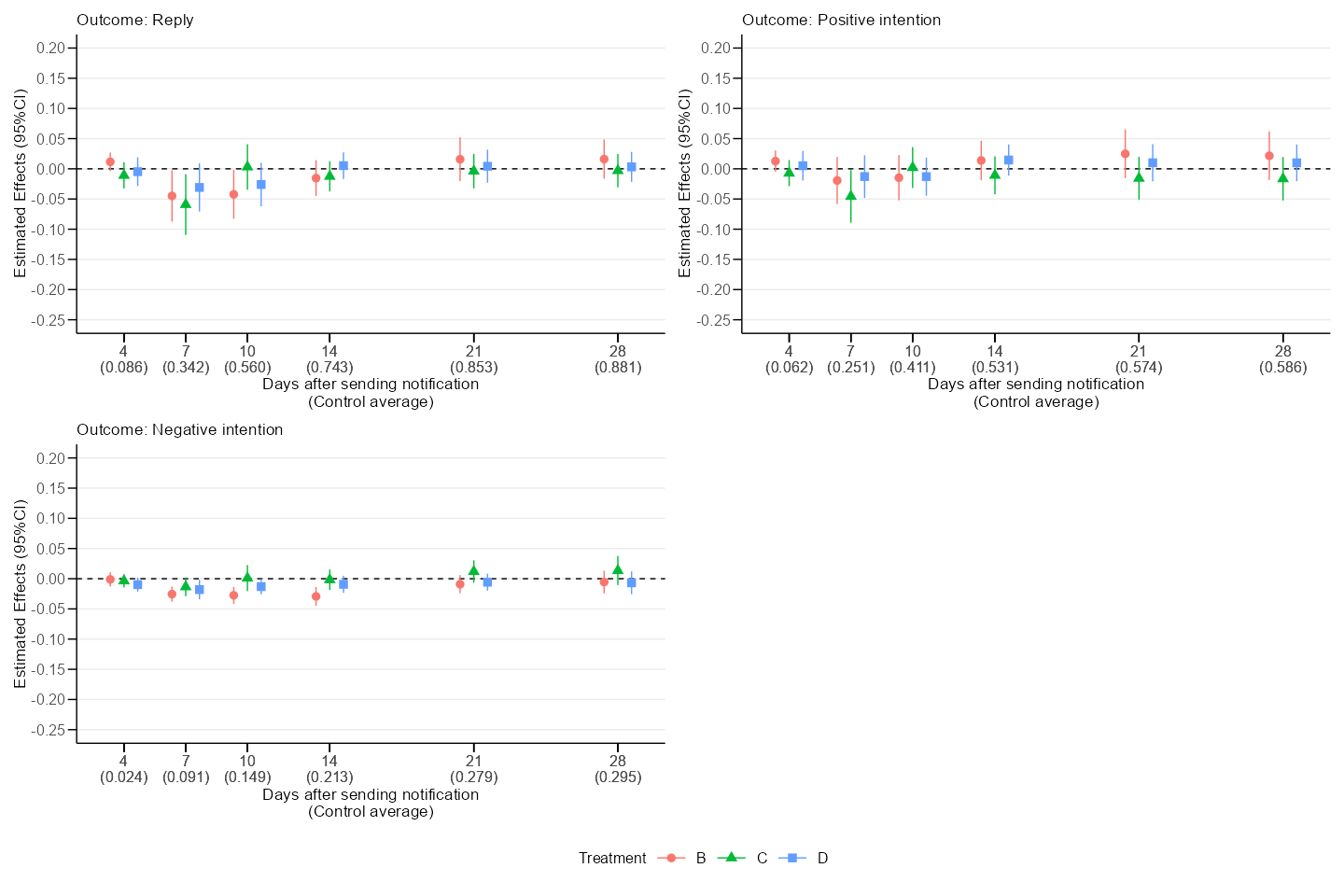


Figure 3.6: Effect on Reply within Specific Days after Sending Notification among Males More than 30. Notes: These plots show the average effect (and associated 95% confidential interval) on each outcome. We use clustered standard errors with the CR2 adjustment (cluster unit is experimental week). We control number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, number of hospitals with BM collection per 10 square kilometers, month dummies, and week dummies.

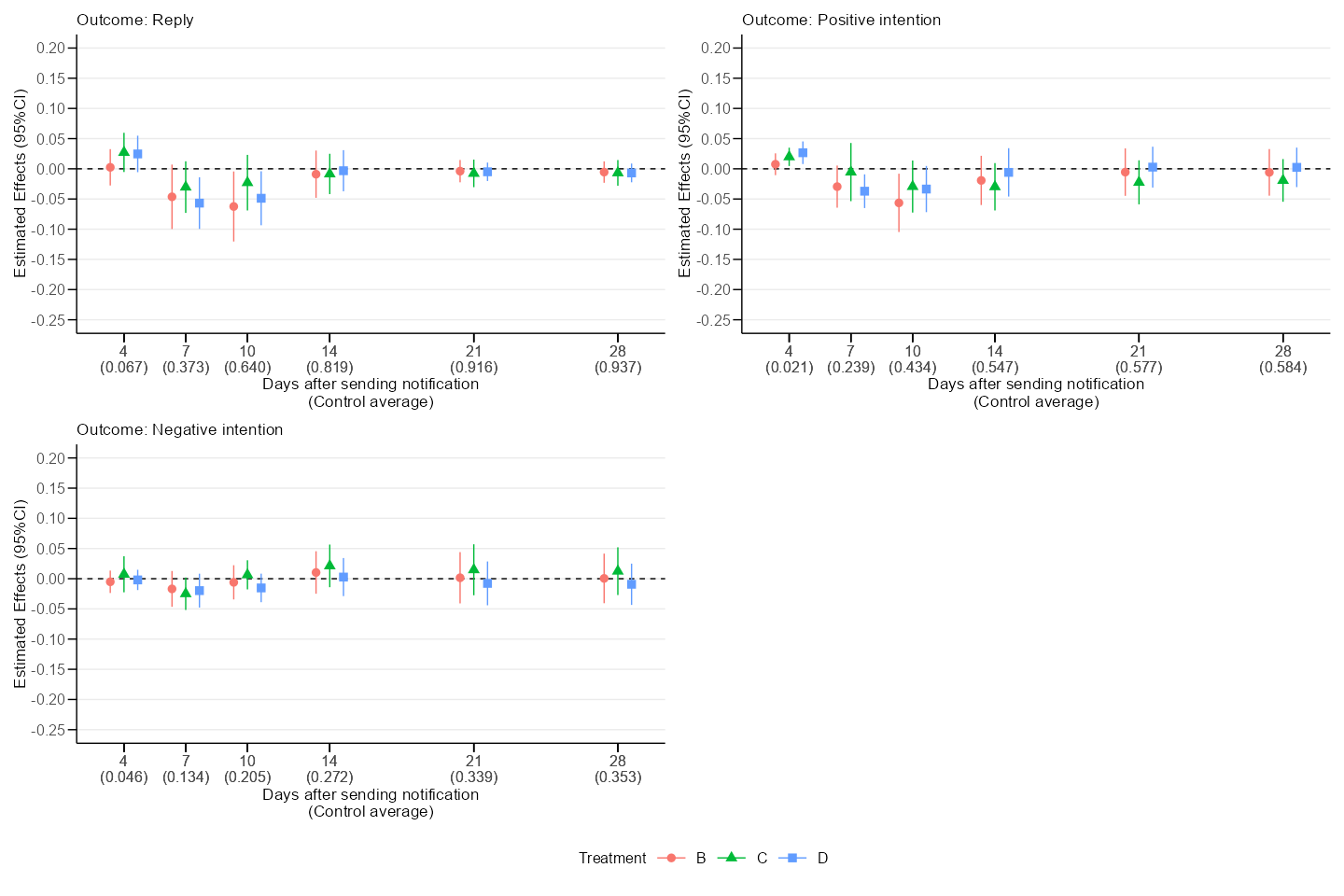


Figure 3.7: Effect on Reply within Specific Days after Sending Notification among Females More than 30. Notes: These plots show the average effect (and associated 95% confidential interval) on each outcome. We use clustered standard errors with the CR2 adjustment (cluster unit is experimental week). We control number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, number of hospitals with BM collection per 10 square kilometers, month dummies, and week dummies.

Table 3.8: Conditional Average Treatment Effect Estimated by RCF

|  | Females | | | Males | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | Estimate | S.E. | P-value | Estimate | S.E. | P-value |
| Age: Less than 30 | | | | | | |
| B | -0.006 | 0.048 | 0.899 | 0.124 | 0.040 | 0.002 |
| C | 0.047 | 0.048 | 0.328 | 0.049 | 0.041 | 0.229 |
| D | -0.057 | 0.048 | 0.236 | 0.044 | 0.040 | 0.263 |
| Age: More than or equal to 30 | | | | | | |
| B | -0.006 | 0.029 | 0.825 | 0.019 | 0.021 | 0.382 |
| C | -0.014 | 0.030 | 0.629 | -0.011 | 0.023 | 0.634 |
| D | 0.014 | 0.029 | 0.628 | 0.021 | 0.022 | 0.350 |
| Notes: See Athey and Wager (2019) for estimation method of conditional average treatment effect (CATE). Since these estimates are asymptotically normal, we calculate z-score under the null hypothesis that CATE is zero, and obtain p-value. | | | | | | |

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>.

Carlana, Michela, Eliana La Ferrara, and Paolo Pinotti. 2022. “Goals and Gaps: Educational Careers of Immigrant Children.” *Econometrica* 90 (1): 1–29. <https://doi.org/10.3982/ECTA17458>.

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.

Murakami, Kayo, Hideki Shimada, Yoshiaki Ushifusa, and Takanori Ida. 2022. “HETEROGENEOUS TREATMENT EFFECTS OF NUDGE AND REBATE: CAUSAL MACHINE LEARNING IN A FIELD EXPERIMENT ON ELECTRICITY CONSERVATION.” *International Economic Review*, June, iere.12589. <https://doi.org/10.1111/iere.12589>.

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

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

1. 近年、ハプロ移植と呼ばれる、半合致したHLAを持つ近親者間での移植が広がりつつある。加えて、母子をつなぐさい帯や胎盤に含まれる血液細胞の移植（さい帯血移植）も同様に人気になっている。骨髄移植と異なり、HLAが完全に合致していなくても、さい帯血移植は実施できる。 [↑](#footnote-ref-20)
2. これは珍しいHLAの型を持つ新しいドナーの確率が低いからである。 [↑](#footnote-ref-21)
3. JMDPは同時に潜在ドナーにJMDPが適合通知を送付したことを知らせるSNSメッセージを送付する。 [↑](#footnote-ref-23)
4. 介入メッセージを作成したとき、我々は潜在ドナーに対して過度な心理的プレッシャーを与えないように配慮している。具体的には、第一に、嘆願調のような表現を避けている。第二に、JMDPが公開している情報のみを用いている。また、移植リスクは通常通り説明している。 [↑](#footnote-ref-29)
5. 一人が海外に在住していた。また、データ提供時点でコーディネーションが進行中であった潜在ドナーは104名いた。コーディネーションが進行中の潜在ドナーの比率は実験群間でバランスしている（F-test, p-value = ）。 [↑](#footnote-ref-32)
6. 技術的な詳細は Athey and Imbens (2016) ・ Wager and Athey (2018) を参照せよ。また、本研究ではRCFを特殊ケースとして扱う一般化ランダムフォレストと呼ばれる手法を用いているが、直観は変わらない。一般化ランダムフォレストの詳細は Athey and Wager (2019) ・ (**Athey2019a?**) を参照せよ。 [↑](#footnote-ref-42)