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

# 1 Field Experiment

## 1.1 Background: Coordination Process of JMDP

フィールド実験の介入のタイミングとデータの理解を促すために、骨髄バンクに登録したドナー候補者が幹細胞を提供するに至るまでの工程を概観しておく。 はじめに、あるドナー候補者のHLAが骨髄バンクに登録した患者のHLAと一致すると、骨髄バンクの事務局はその候補者に幹細胞提供を依頼する適合通知を送付する[[1]](#footnote-20)。 ドナー候補者は適合通知に対して提供の意向と問診表を記入して、返信する。

提供を希望したドナー候補者はおよそ1カ月以内に確認検査を受ける。確認検査では、コーディネーターが提供方法（骨髄採取・末梢血幹細胞採取）を詳細に説明し、ドナー候補者とその家族の意向について調査する。 このとき、ドナー候補者は希望する採取方法を選べる。 また、調整医師が問診・診察・感染症の有無や血液型を調べるための一般血液検査を実施する。 この検査を通じて、ドナー候補者が、骨髄バンクが設定した基準を満たしているかどうかを調べる。

骨髄バンクに登録したドナーは同時に最大10人のドナー候補者とのコーディネーションを進められる。 患者の主治医は確認検査まで至った複数のドナー候補者の中から最も最適な候補者（第一候補者）を選ぶ。 重要な点であるが、ドナー候補者は自身とマッチした患者が何人のドナーとのコーディネーションを進めているかを知ることができない。 また、コーディネーターや調整医師もこの点を知らないので、ドナー候補者が他者からその情報を得ることもできない。

第一候補者となったドナーはコーディネーターと調整医師からの説明を再度受けて、最終的な意思決定をする（最終同意）。 このとき、ドナーの家族の代表者も採取に同意する必要がある。 また、最終同意後、ドナーは自分の意向を変えられない。 最終的な同意をした後、ドナーは約1週間の入院をして、術前検査と採取準備（貧血を防ぐための自己血採血）を受ける。 この後、ドナーは幹細胞を採取する処置を受ける。確認検査から採取までの期間はおよそ3～4カ月程度である。

## 1.2 Experimental Design

knitr::include\_graphics(here("image", "intervention.png"))

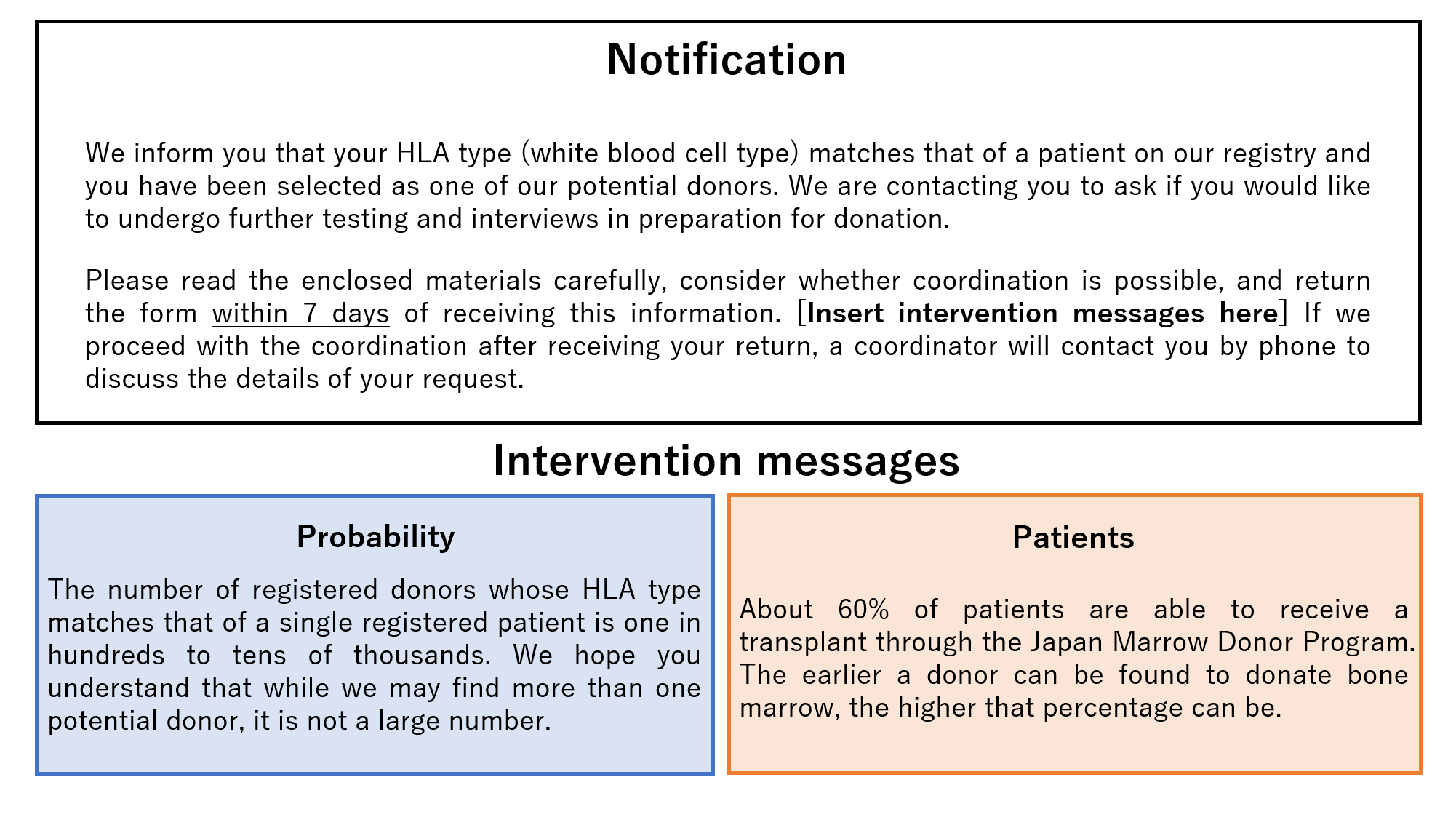


Figure 1.1: Intervention Messages

我々は、骨髄バンクがドナー候補者に幹細胞提供を依頼する適合通知の内容に介入を施した。上図に介入の内容を示す。 介入前の適合通知はドナー候補者として選ばれたこと、7日以内に提供意思と問診表を記入して返信してほしいことを記載している。 また、適合通知と併せて同封した資料の一つは先の小節で述べた採取までの流れを示したハンドブックである。

我々はコーディネーションの促進を目的として、適合通知に行動科学の知見に基づく二種類のメッセージを追加した（図1.1）[[2]](#footnote-25)。 確率メッセージは骨髄バンクに登録した患者とHLA型が一致するドナー登録者が少ないことを強調している。 これは同時に進行している他のドナー候補者の数に関する信念の過大評価を防ぎ、幹細胞提供の「ただ乗り」行動を抑制できる可能性がある。

患者メッセージは骨髄バンクを通して移植を受けられる患者が半数しかいないことを強調している。 これは自分が協力しなくても患者は助かるだろうという認識を修正し、ドナー候補者の利他性を刺激していることを目的としている。 さらに、提供に至るまでの期間を短くすることの重要性を強調することで、適合通知への早い返信を促している。

二種類の介入メッセージの効果を推定するために、我々は四つの実験群を設ける。 実験群Aは介入メッセージのない適合通知を送る（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患者情報メッセージを送る。 さらに、情報過多による認知負荷の負の影響を検証するために、二つの介入メッセージを両方加えた適合通知を送付する実験群Dも設ける。

flextable(schedule) %>%  
 align(j = -1, align = "center", part = "all")

Table 1.1: 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 |

我々は2021年9月から2022年2月にかけて骨髄バンクが適合通知を送付したドナー候補者11,154名をフィールド実験の対象とした。実験群の割り当ては骨髄バンク事務局の業務の無理のない範囲で週単位でクラスターランダム化した。このとき、週・月の固定効果を取り除けるように、実験群が週・月でバランスするように割り当てた。割り当てのスケジュールは表1.1にまとめている。

## 1.3 Data and Empirical Strategy

lm\_robust(  
 ongoing ~ treat,  
 data = rawdt,  
 cluster = RCTweek,  
 se\_type = "stata"  
) %>%  
{  
 f <- summary(.)$fstatistic[1]  
 numdf <- summary(.)$fstatistic[2]  
 dendf <- summary(.)$fstatistic[3]  
 p <- pf(f, numdf, dendf, lower.tail = FALSE)  
  
 sprintf("F-value = %1.3f; p-value = %1.3f", f, p)  
}

## [1] "F-value = 0.105; p-value = 0.956"

我々は2022年6月末に骨髄バンクが管理するコーディネーションデータの提供を受けた。観測単位はフィールド実験の対象であるドナー候補者である。個人属性として性別・年齢・過去のコーディネーション回数・居住地域（都道府県レベル）を記録している。コーディネーションの過程について、提供に至るまでの各工程（適合通知への返信・確認検査・第一候補者選定・最終同意・採取）に到達したかどうかを記録しており、これらをアウトカム変数として用いる。特に、適合通知の返信については、返信したかどうかに加えて、返信日数・提供意向についても記録している。コーディネーションが途中で中断した場合、その理由を三つのカテゴリー（患者理由・ドナーの健康以外の理由・ドナーの健康上の理由）で記録している。分析対象は国内在住でコーディネーションが完全に終了している11,049名とする[[3]](#footnote-27)。

追加的なデータとして、我々はJMDPがホームページ上で公開している施設リストを用いる。このデータは病院の住所に加えて、骨髄採取が可能かどうか、末梢血幹細胞採取が可能かどうかを含んでいる。我々はこのデータを都道府県レベルで集計して、10平方キロメートル当たりの病院の数を計算し、コーディネーションデータと都道府県をキーとして突合する。我々はこの変数をコーディネーションや提供の移動コストとみなす。

tbl <- use %>%  
 datasummary(  
 (`Male (=1)` = male) +  
 (`Age` = age) +  
 (`Number of past coordinations` = coordinate) +  
 (`Number of listed hospitals` = hospital\_per\_area) +  
 (`Number of hospitals listed with PBSC collection` = PB\_per\_area) +  
 (`Number of hospitals listed with BM collection` = BM\_per\_area) ~ mean \* treat,  
 data = .,  
 add\_rows = size,  
 add\_columns = balance\_test[c(5, 1, 3, 4, 6, 2), 5],  
 align = "lccccc",  
 output = "data.frame"  
 ) %>%  
 bind\_cols(  
 tibble(type = c(  
 rep("A. Interventions", 3),  
 rep("B. Sample Size", 1),  
 rep("C. Covariates, Balance Test", 6)  
 ))  
 )  
  
as\_grouped\_data(tbl, groups = "type") %>%  
 as\_flextable(hide\_grouplabel = TRUE) %>%  
 align(j = -1, align = "center", part = "all") %>%  
 padding(i = 2:4, j = 1, padding = 20) %>%  
 padding(i = 6, j = 1, padding = 20) %>%  
 padding(i = 8:13, j = 1, padding = 20) %>%  
 padding(i = 6, j = 1, padding = 20) %>%  
 width(j = 1, 2) %>%  
 fontsize(size = 9, part = "all")

Table 1.2: Overview of Field Experiment

|  | 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. Covariates, Balance Test | | | | | |
| Male (=1) | 0.62 | 0.63 | 0.63 | 0.61 | 0.36 |
| Age | 38.38 | 38.12 | 37.45 | 37.98 | 0.07 |
| Number of past coordinations | 1.61 | 1.59 | 1.62 | 1.56 | 0.45 |
| Number of listed hospitals | 0.48 | 0.49 | 0.49 | 0.49 | 0.80 |
| Number of hospitals listed with PBSC collection | 0.16 | 0.17 | 0.17 | 0.16 | 0.80 |
| Number of hospitals listed with BM collection | 0.25 | 0.26 | 0.25 | 0.25 | 0.67 |

1.2はフィールド実験を概観している。パネルAは各実験群の介入をまとめており、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は共変量のバランステストの結果を示しており、割り当てのランダム化が成功しているかどうかを検証している。ほとんどの共変量が群間で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。ただし、実験群C・Dのドナー候補者は実験群A・Bのドナー候補者よりも若く、群間でアンバランスな可能性がある（F-test, p-value = ）。

ドナー候補者は実験群を選択できない、すなわち、実験群は外生的であるので、単純な二群比較は平均処置効果を識別できる。ただし、年齢および割り当ての週と月の数が実験群間で完全にバランスしていないので、単純な二群比較はバイアスを伴う可能性がある。そこで、二群比較に加えて、我々は月の第週に適合通知を受け取った個人について以下の線形確率モデルを推定する、

ここで、は個人属性ベクトル、とはそれぞれ週・月のダミー変数である。 我々はランダム化のユニットである実験週レベルでクラスタした標準誤差を使用する。

# 2 Results: Effects on Reply and Intentions

はじめに、ドナー候補者の意向が最も現れる返信に対する効果を推定する。返信のアウトカム変数は提供意向に関わらず適合通知に返信したらならば1を取るダミー変数である。さらに、提供意向の変数を用いて、返信に対する効果を二つに分解する。第一に、正の意向を示して、適合通知に返信した効果である。アウトカム変数は適合通知に返信し、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。第二に、負の意向を示して、適合通知に返信した効果である。アウトカム変数は適合通知に返信し、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。提供の意向に関する二つのアウトカム変数の和は必ず返信のアウトカム変数と一致するので、二つの意向に対する効果の和は必ず返信に対する効果となる。コントロール（実験群A）の返信率は%である。正の意向を伴って返信した比率は%である一方で、負の意向を伴って返信した比率は%である。したがって、適合通知に返信したドナー候補者の%が提供したいと考えている。

返信をする以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。このケースはドナー候補者の意向と無関係に生じるコーディネーションの中断なので、分析のサンプルから除外する。各実験群の–%をこのケースで除外しており、その比率は群間でバランスしている。

## 2.1 Main Results

est\_stock %>%  
 pull(fit) %>%  
 setNames(paste0("(", seq\_len(length(.)), ")")) %>%  
 modelsummary(  
 coef\_map = c(  
 "(Intercept)" = "Constant",  
 "treatB" = "Treatment B",  
 "treatC" = "Treatment C",  
 "treatD" = "Treatment D"  
 ),  
 stars = c("\*\*\*" = .01, "\*\*" = .05, "\*" = .1),  
 fmt = 4,  
 gof\_omit = "R2|AIC|BIC|Log|Std|FE|se\_type",  
 add\_rows = add\_table,  
 ) %>%  
 flextable::add\_header\_row(  
 values = c("", "Reply", "Positive", "Negative"),  
 colwidths = c(1, 2, 2, 2)  
 ) %>%  
 flextable::add\_header\_row(  
 values = c("", "Intention"),  
 colwidths = c(3, 4)  
 ) %>%  
 align(j = -1, align = "center", part = "all") %>%  
 width(j = 1, 1) %>%  
 fontsize(size = 9, part = "all")

Table 2.1: Linear Probability Model of Replay and Intention

|  | | | Intention | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | Reply | | Positive | | Negative | |
|  | (1) | (2) | (3) | (4) | (5) | (6) |
| Constant | 0.8835\*\*\* | 0.9510\*\*\* | 0.5533\*\*\* | 0.6463\*\*\* | 0.3303\*\*\* | 0.3047\*\*\* |
|  | (0.0065) | (0.0190) | (0.0125) | (0.0277) | (0.0089) | (0.0210) |
| Treatment B | 0.0105 | 0.0130\*\* | 0.0218 | 0.0203\* | -0.0113 | -0.0073 |
|  | (0.0075) | (0.0061) | (0.0135) | (0.0123) | (0.0094) | (0.0075) |
| Treatment C | -0.0057 | 0.0006 | -0.0053 | -0.0052 | -0.0004 | 0.0057 |
|  | (0.0081) | (0.0053) | (0.0151) | (0.0111) | (0.0150) | (0.0093) |
| Treatment D | 0.0058 | 0.0063 | 0.0046 | 0.0063 | 0.0012 | 0.0000 |
|  | (0.0074) | (0.0051) | (0.0166) | (0.0097) | (0.0138) | (0.0065) |
| Control average | 0.8835 | 0.8835 | 0.5533 | 0.5533 | 0.3303 | 0.3303 |
| Covariates |  | X |  | X |  | X |
| Num.Obs. | 10985 | 10985 | 10985 | 10985 | 10985 | 10985 |

表2.1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のみを説明変数として用いているのに対して、偶数列は個人属性や月・週の固定効果も制御している。列(1)は、確率メッセージを加えた適合通知を送付する実験群Bが返信率をもしくは%高めていることを示している。個人属性などをコントロールする（列(2)）と、効果のサイズは大きく変化しないが、統計的に5%水準で有意となる。この効果は正の意向を伴った返信の増加に由来している。列(3)と(5)では、実験群Bが正の意向を伴った返信率をもしくは%高めている一方で、負の意向を伴った返信率をもしくは%下げている。ただし、これらの統計的な有意性は、個人属性などをコントロールしても、弱い。また、実験群CとDは実験群Bよりも効果のサイズが小さく、統計的に非有意である。我々は頑健性の分析としてロジットモデルを推定した（表3.1）。その結果は効果の方向について上表と整合的であるが、統計的に非有意である。総合すると、我々の介入は全体的に返信に大きな影響を与えていない。

est\_stock\_sub <- stock %>%  
 mutate(age\_less30 = if\_else(age < 30, 1, 0)) %>%  
 group\_by(outcome, male, age\_less30) %>%  
 nest() %>%  
 mutate(est = map(data, ~ lm\_robust(  
 update(mod$ctrl, . ~ . - male - age\_demean),  
 cluster = RCTweek,  
 se\_type = "stata",  
 data = .x  
 ))) %>%  
 mutate(  
 fit = map(est, tidy),  
 fit = map(fit, ~ subset(.x, str\_detect(term, "treat"))),  
 fit = map(fit, ~ dplyr::select(.x, -outcome)),  
 N = map\_chr(est, ~ paste("N =", nobs(.x))),  
 mean = map\_dbl(data, ~ with(subset(., treat == "A"), mean(value))),  
 mean = sprintf("Control avg = %1.3f", mean)  
 ) %>%  
 select(-data, -est) %>%  
 unnest(cols = fit) %>%  
 mutate(  
 pos = paste0(male, age\_less30),  
 pos = factor(  
 pos,  
 levels = c("01", "00", "11", "10"),  
 labels = c(  
 "Female\u00d7\nAge<30",  
 "Female\u00d7\n30\u2264Age",  
 "Male\u00d7\nAge<30",  
 "Male\u00d7\n30\u2264Age"  
 )  
 ),  
 term = str\_replace(term, "treat", ""),  
 term = factor(term, LETTERS[2:4])  
 )

## 3 coefficients not defined because the design matrix is rank deficient  
##   
## 3 coefficients not defined because the design matrix is rank deficient  
##   
## 3 coefficients not defined because the design matrix is rank deficient  
##   
## 3 coefficients not defined because the design matrix is rank deficient  
##   
## 3 coefficients not defined because the design matrix is rank deficient  
##   
## 3 coefficients not defined because the design matrix is rank deficient  
##   
## 3 coefficients not defined because the design matrix is rank deficient  
##   
## 3 coefficients not defined because the design matrix is rank deficient  
##   
## 3 coefficients not defined because the design matrix is rank deficient  
##   
## 3 coefficients not defined because the design matrix is rank deficient  
##   
## 3 coefficients not defined because the design matrix is rank deficient  
##   
## 3 coefficients not defined because the design matrix is rank deficient

text <- est\_stock\_sub %>%  
 select(male, age\_less30, pos, N, mean) %>%  
 distinct()

## Adding missing grouping variables: `outcome`

plot\_list <- unique(est\_stock\_sub$outcome) %>%  
 purrr::map(function(x) {  
 subset(est\_stock\_sub, outcome == x) %>%  
 ggplot(aes(x = pos, y = estimate)) +  
 geom\_hline(aes(yintercept = 0), linetype = 2) +  
 geom\_point(  
 aes(color = term, shape = term),  
 size = 3, position = position\_dodge(0.5)  
 ) +  
 geom\_errorbar(  
 aes(ymin = conf.low, ymax = conf.high, color = term),  
 position = position\_dodge(0.5),  
 width = 0  
 ) +  
 geom\_text(  
 aes(y = -0.125, label = N),  
 data = subset(text, outcome == x),  
 color = "black"  
 ) +  
 geom\_text(  
 aes(y = -0.15, label = mean),  
 data = subset(text, outcome == x),  
 color = "black"  
 ) +  
 scale\_y\_continuous(  
 breaks = seq(-0.2, 0.2, by = 0.05), limits = c(-0.17, 0.17)  
 ) +  
 labs(  
 title = paste("Outcome:", x),  
 x = "Subset",  
 y = "Estimated Effects (95%CI)",  
 color = "Treatment", shape = "Treatment"  
 ) +  
 simplegg()  
 })  
  
wrap\_plots(plot\_list, ncol = 2) +  
 plot\_layout(guides = "collect") &  
 theme(legend.position = "bottom")

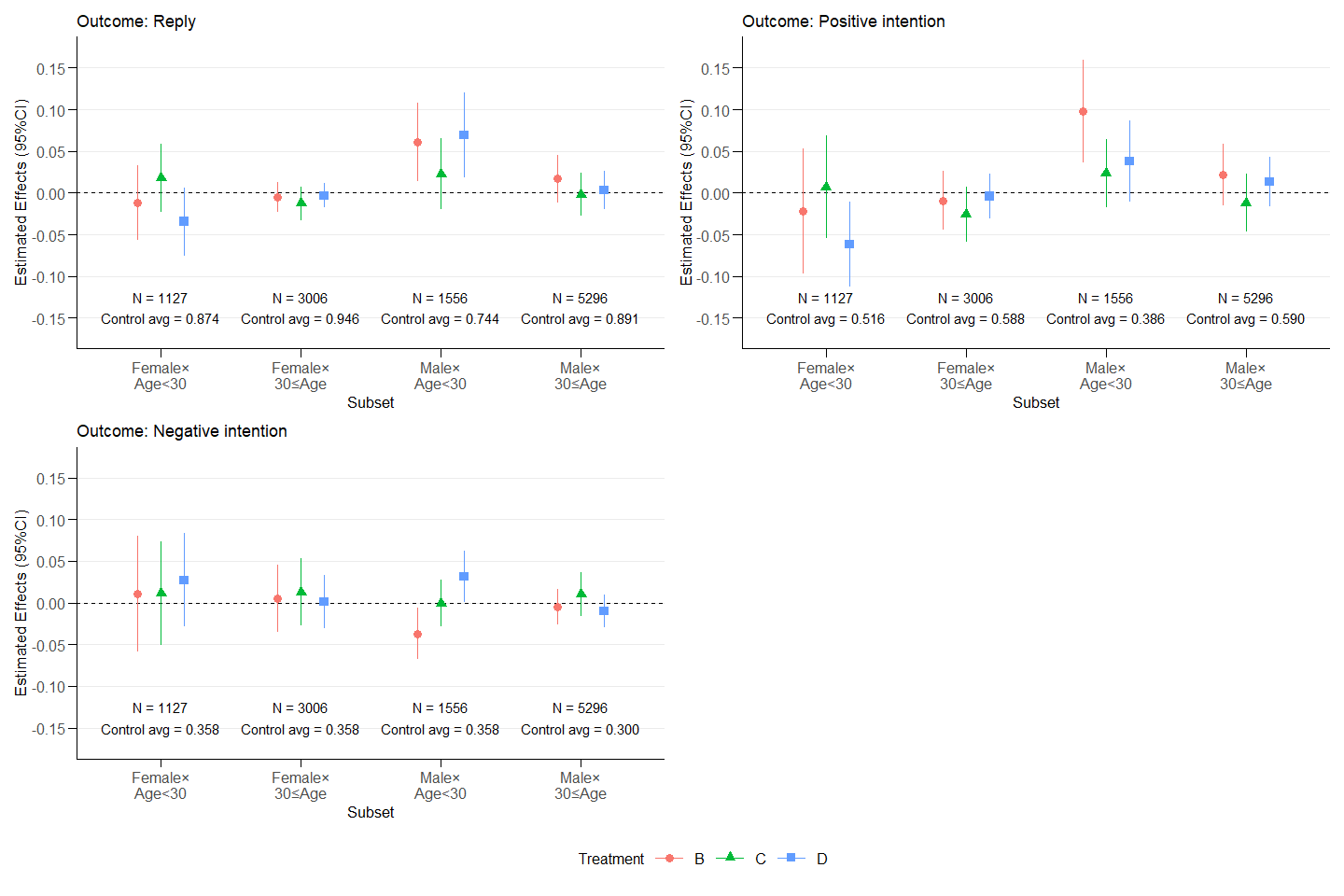


Figure 2.1: Effect on Reply and Intentions by Gender and Age Group. Note: These plots show the average effect (and associated 95% confidential interval) on each outcome by gender and age group. We cluster standard errors by experimental weeks. We control number of past coordinations, number of hospitals per 10 square kilometers, number of hospitals with PBSC collection per 10 square kilometers, number of hospitals with BM collection per 10 square kilometers, prefecture dummies, month dummies, and week dummies.

しかしながら、メッセージの効果は大きな異質性がある。図2.1は性別と年齢層（30歳未満かどうか）でサンプルを4分割して、各サブセットでメッセージの効果を推定している。その結果、確率メッセージを加えた実験群BとDは30歳未満の男性の返信率をもしくは%高めている（ベースラインの返信率は%）。とくに、確率メッセージだけを加えた実験群Bの効果は正の意向を伴う返信率の増加に起因している。このグループで、実験群Bは正の意向を伴う返信率をもしくは%高めている（ベースラインの比率は%）一方で、負の意向を伴う返信率をもしくは%下げている（ベースラインの比率は%）。これらは統計的に有意な効果である。他の性年代のグループ（とくに、30歳以上の男女）では、介入メッセージが返信や意向に大きな影響を与えていない[[4]](#footnote-33)。したがって、他のドナー候補者が少ないというメッセージは若年男性の提供意向を高め、適合通知への返信を促している。若年男性ドナーは移植成績が良いにも関わらず、提供意向が他の性年代よりも明らかに低いことを考慮すると、確率メッセージは移植コーディネーションの効率性を改善しているといえる。

# Appendix

# 3 Estimation of Logit Models

est\_stock %>%  
 pull(fit) %>%  
 setNames(paste0("(", seq\_len(length(.)), ")")) %>%  
 modelsummary(  
 estimate = "{or}",  
 statistic = "[{lower.or}, {upper.or}]",  
 coef\_map = c(  
 "treatB" = "Treatment B",  
 "treatC" = "Treatment C",  
 "treatD" = "Treatment D"  
 ),  
 add\_rows = add\_tables,  
 gof\_omit = "R2|AIC|BIC|RMSE|Std|FE|se\_type",  
 fmt = fmt\_sprintf("%.3f")  
 ) %>%  
 flextable::add\_header\_row(  
 values = c("", "Reply", "Positive", "Negative"),  
 colwidths = c(1, 2, 2, 2)  
 ) %>%  
 flextable::add\_header\_row(  
 values = c("", "Intention"),  
 colwidths = c(3, 4)  
 ) %>%  
 align(j = -1, align = "center", part = "all") %>%  
 width(j = 1, 1) %>%  
 fontsize(size = 9, part = "all")

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :  
## prediction from a rank-deficient fit may be misleading  
  
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :  
## prediction from a rank-deficient fit may be misleading  
  
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :  
## prediction from a rank-deficient fit may be misleading

Table 3.1: Logit Model of Reply and Intention

|  | | | Intention | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | Reply | | Positive | | Negative | |
|  | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment B | 1.112 | 1.143 | 1.093 | 1.091 | 0.950 | 0.967 |
|  | [0.940, 1.315] | [0.958, 1.364] | [0.982, 1.216] | [0.975, 1.221] | [0.848, 1.063] | [0.860, 1.088] |
| Treatment C | 0.947 | 1.003 | 0.979 | 0.978 | 0.998 | 1.028 |
|  | [0.801, 1.120] | [0.837, 1.201] | [0.878, 1.092] | [0.869, 1.100] | [0.889, 1.120] | [0.909, 1.162] |
| Treatment D | 1.060 | 1.062 | 1.019 | 1.027 | 1.005 | 0.999 |
|  | [0.893, 1.257] | [0.886, 1.272] | [0.914, 1.136] | [0.915, 1.153] | [0.896, 1.128] | [0.886, 1.127] |
| Covariates |  | X |  | X |  | X |
| Num.Obs. | 10985 | 10985 | 10985 | 10985 | 10985 | 10985 |
| Log.Lik. | -3884.517 | -3629.432 | -7534.803 | -7254.947 | -6945.023 | -6818.753 |

1. 適合通知を送付したことを伝えるSMSメッセージも送付している。 [↑](#footnote-ref-20)
2. メッセージの作成の際、我々はドナー候補者に過度なプレッシャーを与えないように適切な配慮をしている。第一に、嘆願調のようなメッセージを避けている。第二に、メッセージの作成に際して、骨髄バンクが公開している情報のみを使用している。第三に、これまでと同様に移植リスクに関する説明をしている。 [↑](#footnote-ref-25)
3. 海外に在住している人が1人いた。また、データ提供時点でコーディネーションが進行している人が104名いた。コーディネーションが進行している人の比率は実験群間でバランスしている（F-value, p-value = ） [↑](#footnote-ref-27)
4. ただし、30歳未満の女性で、実験群Dが統計的に有意に正の意向を伴う返信率をもしくは%下げている。 [↑](#footnote-ref-33)