NBER WORKING PAPER SERIES

# NUDGING THE NUDGER: A FIELD EXPERIMENT ON THE EFFECT OF PERFORMANCE FEEDBACK TO SERVICE AGENTS ON INCREASING ORGAN DONOR REGISTRATIONS

Julian House
Nicola Lacetera
Mario Macis
Nina Mazar

Working Paper 30547
http://www.nber.org/papers/w30547

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2022

We are thankful to the Crabtree Foundation for financial support, and to Service Ontario and Trillium Gift of Life Network for their collaboration. We thank participants to seminars at Carnegie Mellon University, the Data Colada seminar series, INSEAD, IESE, Yale SOM's Spring Decision Making Symposium, the University of Minnesota, the University of Toronto and the Society for Judgment and Decision Making Conference in 2022 for their feedback. The study received approval by the Research Ethics Board of the University of Toronto (Protocol no. 32650), and was pre-registered at the AEA RCT Registry (No. 0001974). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Nudging the Nudger: A Field Experiment on the Effect of Performance Feedback to Service Agents on Increasing Organ Donor Registrations

Julian House, Nicola Lacetera, Mario Macis, and Nina Mazar

NBER Working Paper No. 30547
October 2022
JEL No. C93,D90,H41,I10,J45,M50

## ABSTRACT

We conducted a randomized controlled trial involving nearly 700 customer-service representatives (CSRs) in a Canadian government service agency to study whether providing CSRs with performance feedback with or without peer comparison affected their subsequent organ donor registration rates. Despite having no tie to remuneration or promotion, the provision of individual performance feedback three times over one year resulted in a 25% increase in daily signups, compared to otherwise similar encouragement and reminders. Adding benchmark information that compared CSRs performance to average and top peer performance did not further enhance this effect. Registrations increased more among CSRs whose performance was already above average, and there was no negative effect on lower-performing CSRs. A post-intervention survey showed that CSRs found the information included in the treatments helpful and encouraging. However, performance feedback without benchmark information increased perceived pressure to perform.

Julian House
Behavioural Insights Unit,
Treasury Board Secretariat,
Government of Ontario
Toronto, ON M7A 1Y7
Canada
Julian.House@ontario.ca

Mario Macis
Johns Hopkins University
Carey Business School
100 International Drive, Office 1333
Baltimore, MD 21202
and IZA
and also NBER
mmacis@jhu.edu

Nicola Lacetera
University of Toronto
Institute for Management and Innovation
3359 Mississauga Road, Room KN 235
Mississauga, ON L5L 1C6
and NBER
nicola.lacetera@utoronto.ca

Nina Mazar
Boston University
Questrom School of Business
595 Commonwealth Ave
Boston, MA 02215
nina@ninamazar.com

# 1. Introduction

A shortage of organs for transplantation exists in most countries around the world, resulting in large medical costs (e.g., dialysis treatment) and untold human suffering as people wait to receive an effective medical procedure. This imbalance between supply and demand occurs despite the broad social support and positive attitude that the donation of organs (as well as of blood and tissue) enjoys virtually everywhere and the expressed intention of most people to consider donating their organs upon death.[1] Many countries have regulatory provisions and engage in educational and other activities to further promote and facilitate donations. These include presumed consent, prompted choice, public service announcements, campaigns, and the possibility to sign up to an organ donor registry in multiple contexts, for example, while applying for government-issued documents such as a driver's license.

In this paper, we report on a randomized controlled trial (RCT) conducted in the province of Ontario, Canada, where an organ donation shortage persists despite recent improvements in the number of registered organ donors (registrations increased from 24% of eligible adults in 2014 to 35% in 2020; Trillium Gift of Life Network, 2021). As a result, every three days, one resident of Ontario dies waiting for an organ transplant (https://www.beadonor.ca/). We collaborated with ServiceOntario, an agency of the Ministry of Government and Consumer Services. The agency provides a single point of contact for most government services in the province (e.g., driver and vehicle licensing, license plate stickers,[2] public health insurance registration, and business licensing). Most of the organ registrations in Ontario (pre-Covid-19 pandemic: 85%, Trillium Gift of Life Network, 2017) have occurred with customers visiting ServiceOntario centers in person.[3] Because of their unique role, ServiceOntario customer-service representatives (CSRs) are ideally positioned to promote organ donor registrations when interacting with citizens. Indeed, operational policy instructs them to implement a prompted-choice procedure with all customers. But are CSRs registering as many customers as possible, and if not, can they be encouraged to improve? Intermediaries between the giver and the beneficiaries are often crucial in promoting prosocial activities, but evidence on their behaviors and motivations is scant. Our

[1] For example, over 90 percent of Ontarians, 95 percent of Americans, and 84 percent of Germans are in favor of organ donation (status: April 5, 2022: https://www.giftoflife.on.ca/en/; https://www.donornetworkwest.org; https://www.bundesregierung.de/).

[2] License plate stickers are now no longer issued (they are renewed online for free), but they were during the time that our study covers).

[3] Ontario residents can also join the organ donor registry online at BeaDonor.ca.

study considers the role of these third parties in motivating organ donations. There is substantial variation in signup rates across ServiceOntario CSRs (Robitaille et al., 2021). Due to a host of factors, including limited mental bandwidth, time pressure, insufficient salience of organ donor registration, a desire to avoid confrontation, as well as the absence of any material incentives for CSRs to improve their signup rates, some CSRs may fail to solicit customers consistently, or they may be less than convincing when they do. Conversely, other CSRs may be especially effective thanks to their intrinsic motivation, communication skills, or other individual traits.

We focus on the potential role of information and, more specifically, feedback about one's performance as a motivator to increase registration rates. CSRs in government-operated ServiceOntario offices (28% of all offices in the province; Trillium Gift of Life Network 2017) do not receive economic compensation for, nor are they evaluated based upon their organ donor registration activities. Moreover, they are plausibly unaware of their actual performance on this activity and how it compares with their colleagues. At the same time, ServiceOntario could readily provide performance feedback for that voluntary part of the CSRs' job and would presumably be interested in doing so, if beneficial.

Previous research has shown that receiving private feedback about one's *individual* performance can affect subsequent effort and outcomes (e.g., Bandiera et al. 2015). Knowledge (or perceptions) of the typical behavior of others can also affect one's behavior (Duflo and Saez 2002, Munshi and Myaux 2006). In particular, people respond to information about their performance *relative* to that of their peers (Alcott 2011, Croson et al. 2009), and the evidence indicates that this is not simply due to material benefits (Ball et al. 2001) or competitive preferences (Charness and Grosskopf 2001; Charness and Rabin 2002). In fact, studies have shown that people care about (and exert effort based on) their relative position ranking even when this does not produce extrinsic benefits such as financial rewards or social status (Charness et al. 2011, Tran and Zeckhouser 2012). These findings suggest that intrinsic motivation and self-image concerns may explain why information on rankings can affect people's behavior (Bénabou and Tirole 2006, Gneezy et al. 2011). However, the effect of this information is not necessarily positive. For instance, high performers might "relax" and low-performers "give up" when informed of their relative ranking, leading employees to reduce their performance upon receiving feedback (Bandiera et al. 2013; Allcott and Kessler 2019). Moreover, Reiff et al. (2022) have shown that peer comparison may negatively affect outcomes that are often not

measured, such as job satisfaction and burnout.[4] Overall, the existing theories and evidence do not provide precise predictions about the effect of motivating ServiceOntario's CSRs with performance feedback in the absence of extrinsic incentives.[5]

The key interventions in our experiment consisted of providing CSRs, three times over a span of one year (June 2017, January 2018 and June 2018), with information via e-mail about their individual organ-donor signup performance, with or without a regional benchmark. We then measured the effect of the interventions on CSRs' organ-donor registrations over the subsequent weeks and months. Specifically, we randomly assigned 694 CSRs to one of three groups: (1) a "*standard-reminder*" condition in which CSRs received a typical e-mail communication from ServiceOntario about the Trillium Gift Of Life Network, providing them with basic up-to-date organ donor statics, reminding them of the role that they, the CSRs play in encouraging people to join the registry, and with an appeal to help further in this mission; (2) an "*individual performance*" condition that, in addition to the standard message, included information on the CSR's individual organ donor signup performance over the past six months (absolute and per one-hundred customer interactions); and (3) a "*regional benchmark*" condition in which we added information about the performance of all CSRs operating in the region.

Overall, providing performance feedback resulted in 0.15 additional signups per CSR per day compared to encouragement alone, corresponding to a 25% increase over the baseline. There was no difference between individual performance feedback and individual performance feedback plus the regional benchmark. The effect was particularly pronounced in the few weeks immediately following receipt of a treatment e-mail, but a substantial effect persisted for several months. The increase in organ donor signups was concentrated among high-performance CSRs, and there were no effects (positive or negative) on organ donor signups among low-performance CSRs.

To obtain additional insights and help interpret the results, we conducted a post-intervention survey among treated CSRs. In particular, we were interested in learning whether participants paid attention to our intervention e-mails and whether the feedback interventions had any

[4] For a discussion of and additional references on how social comparisons in the workplace can create psychological costs, see Larkin, Pierce, and Gino (2012).

[5] Performance evaluations relative to peers may also result in sabotage and other types of unethical behavior (Edelman and Larkin, 2014; Charness, Masclet and Villeval 2014). However, ServiceOntario's customer representatives' efforts and tasks are largely independent of each other and independent of their peers. Thus, this possibility does not apply to our context.

psychological benefits (e.g., increased the meaning of a CSR's work) or costs (e.g., caused feelings of competition or psychological pressure). The responses revealed that CSRs did pay attention to the treatment e-mails and found the information useful. The survey also indicated that CSRs in the individual performance feedback condition without the benchmark reported experiencing greater pressure to perform. Similar to Reiff et al. (2022), we conclude that performance-feedback interventions can have psychological costs that should be considered when producing a comprehensive evaluation of these policies.[6]

In the next section, we describe the context and the experimental design. In Section 3 we describe the data and present our empirical findings, and in Section 4 we discuss the results and conclude.

# 2. Setting and research design

## 2.1 Context

According to the 2017 Annual Report of the Office of the Auditor General of Ontario, ServiceOntario's in-person centers processed about 25 million transactions annually, 20% of which were handled by government-*owned* offices. ServiceOntario employs several hundred customer service representatives (CSRs) in their 82 offices throughout the province. CSRs at government-owned service centers receive a fixed salary and do not earn a commission for the transactions they process.[7] CSRs regularly receive reminders, about two to four times a year, via e-mail to support organ donor registrations by asking customers whether they would like to register to be organ donors. In addition to not receiving commission pay for signing customers to the registry, this specific activity does not affect the CSRs' performance reviews.

[6] Reiff et al. (2022) evaluated the effects of providing performance feedback to primary care physicians; they find no effect on behavioral outcomes (preventive care measures recommended to patients) and a negative effect of peer comparison on job satisfaction and burnout. In addition to the context, our study differs from Reiff et al. (2022) in the nature of the intervention. In our study, we provide participants summary measures of their peers' performance (regional average and 80$^{th}$ percentile), whereas Reiff et al. (2022) provide the names of the top-25 performers.

[7] Although we do not know how CSRs at privately-owned service centers are paid, privately-owned service centers themselves receive commission for each transaction. Of the just under 300 service centers in Ontario, approximately 72% are privately-owned.

## 2.2 The field experiment

### 2.2.1 Treatment conditions

Our main study consisted of a randomized controlled trial that involved all government-owned ServiceOntario centers that were active at the time of implementation. The goal of our study was to assess whether different types of performance feedback information, though seemingly inconsequential (organ donor registration performance does not influence the performance evaluation or salary of CSRs), affected subsequent organ donor signups. To this aim, we designed the following experimental conditions:

- **"Reminder" (R)**: CSRs in this condition received an e-mail including basic statistics about organ donations in Ontario, a reminder of the role that ServiceOntario plays in adding individuals to the registry, and an appeal to CSRs to help further this mission and exert effort on that activity.
- **"Individual Feedback" (RIF)**: The e-mail had the same information as in Condition R plus the following additions: The number of customers the specific CSR served in the previous six months and how many customers the CSR signed up to the organ donor registry, in absolute terms and for every one-hundred customer interactions (the latter information was expressed both numerically and graphically).
- **"Benchmark" (RIFB)**: In addition to the same information as in Condition RIF, the e-mail included the regional average and 80th percentile for the number of signups per one-hundred-customer interactions in the previous six months (graphical).

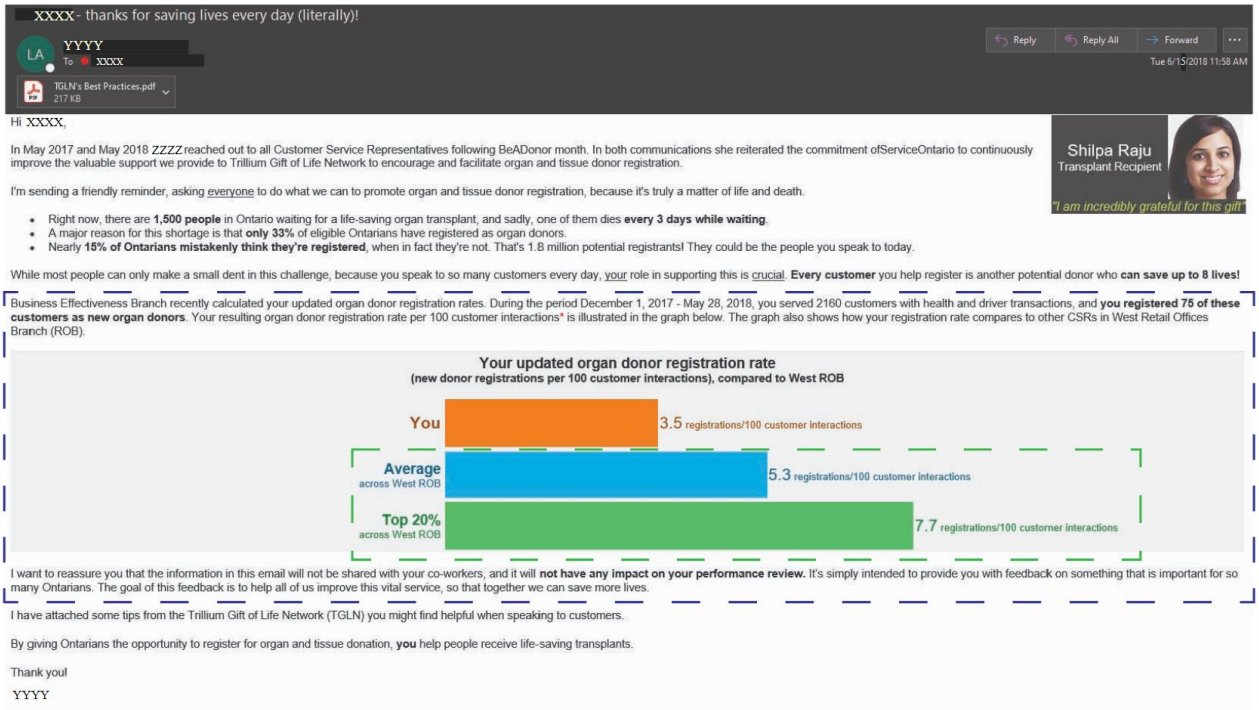
Figure 1 shows examples of the three e-mails. Each CSR received their assigned communication (held constant throughout) up to three times over the course of the study: on June 20, 2017, January 29, 2018, and June 15, 2018.[8]

Strictly speaking, the baseline reminder condition "R" is not a pure control because it does include a message that could affect performance in the activity of interest. We chose this design for the baseline condition for two reasons. First, this choice increases the study's external validity. As mentioned above, ServiceOntario typically sends e-mails to CSRs a few times a year

[8] About two weeks before each of our three intervention dates, all CSRs, regardless of their treatment condition, received an e-mail from a senior provincial government executive announcing that they would soon receive communication about the organ donor registry. The e-mail did not specify what kind of information and did not mention that the communication to come would be part of an experiment; its primary purpose was to increase the CSRs' attention toward, and likelihood of opening our e-mail interventions.

to encourage them to ask customers if they want to register their consent to be a donor.[9] Our "reminder" message was not qualitatively different from those routine e-mails. Therefore, CSRs likely did not perceive our reminder e-mail (or the "standard" part of the longer e-mails in conditions RIF and RIFB) as particularly unusual. Second, having a message in the baseline condition would allow us to identify the motivating impact of the presence and type of performance information, separately from simple reminder or salience effects.

**Figure 1: Sample of treatment emails sent to CSRs**




*Notes*: The figure reports a snapshot of the emails that the CSRs received, according to their assigned treatment condition. The content of the emails was the same for all CSRs expect for the individual feedback and benchmark statistics, shown in dashed blue and green boxes, respectively. The names of the email sender and receiver, as well as of the government executive who sent the is information emails a few weeks before each intervention date, are redacted, and the date on the top right of the figure corresponds to the day in which CSRs received the email for the third intervention.

### 2.2.2 Randomization

To minimize informational spillovers between CSRs in different conditions, we randomly assigned the conditions by office (i.e., all CRSs in any given office would receive the same experimental condition). We also stratified the randomization by the four regions in which ServiceOntario partitions the Province: North, East, West, and Center because these regions

[9] In fact, ServiceOntario requested that we not exclude any workers from any form of communication whatsoever.

present socio-economic and geographical differences. Condition assignment by office also complied with requests from our partner organization to maintain equality of treatment within a specific location. One challenge of this assignment strategy was that some CSRs work in more than one office; within an intervention wave, about 30% of CSRs did.[10] We chose to assign each of these multi-location workers to the office (and thus, condition) where they typically spent more time in the months immediately preceding the intervention. This deviation from full adherence to our design added "natural" variation, allowing us to control for office fixed effects in the econometric analyses (see Section 3 below).

When we assigned CSRs to experimental conditions, the dataset at our disposal (as of April 30, 2017) included 565 individual CSRs in 79 offices.[11] However, we have outcome data for a total of 82 offices. Thus, three offices were not used to assign conditions, although we subsequently observed CSRs working at those locations on some days.

### 2.2.3 Constructing signup performance measures

The computation of individual signup performance measures and region-specific benchmarks required several steps and assumptions, given some peculiarities of the internal data collection processes at ServiceOntario and the administrative data structure.

ServiceOntario classifies its various services into two broad categories: health-related and not health-related. The health-related services include issuing or renewing "health cards" (the document that certifies the right to access Ontario's free public health care system). The main non-health-related services are the issuance or renewal of driver's licenses and the provision of license plates; others include the issuance of Ontario photo cards (i.e., IDs that are not also a driver's license), birth certificates, etc. A CSR's interaction with a customer may include one or more of these services. At the end of the interaction, a CSR is supposed to ask customers whether they are willing to sign up for the Ontario Organ Donor Registry. ServiceOntario keeps track of all these activities that each CSR performs. However, it does so in two separate systems. For health-related services, CSR activities are recorded at the level of the single customer interaction. These data also include whether an interaction resulted in a new organ-donor registry

[10] Of these CSRs, about two thirds worked in two offices, and very few in more than three.

[11] There were 24 offices and 177 CSRs in Condition M (7.4 CSRs per office on average), 27 offices and 198 CSRs in Condition MF (7.3 CSRs per office) and 28 offices and 190 CSRs in Condition MFB (6.8 CSRs per office).

signup. For all non-health-related services, instead, the information is aggregated at the CSR-day level and stored in a separate database.

To assemble complete information on each CSR's daily activity, we had to merge these two data sources at the same level of aggregation. The first step of this procedure was to aggregate the customer interaction data for health-related activities at the CSR-day level. In particular, the count of entries for a given CSR on a given day provided us with the number of unique customers to whom they provided health-related services. Also, the total number of new organ-donor registry entries on each day measures a CSR's daily "absolute" signup performance.

Second, we merged these data, at the CSR-day level, with the data on non-health-related activities. For these activities, the data included the number of *services* (or transactions) per day but not the exact number of *customers* receiving non-health-related services; this is because a customer may need more than one service in a single visit. The data at our disposal and additional information we obtained from ServiceOntario indicate that, on average, a given interaction with a customer concerns about 1.3 health-related services. Therefore, we assumed the same for non-health services, divided the daily non-health services by 1.3 to measure unique customer interactions, and added these to the daily health-related interactions for a given CSR-day to obtain total daily unique customer interactions.

For each CSR, we then added up all the new organ donor registry signups that they made over the period preceding each of our three intervention waves. We also added up all the unique customer interactions to express the registry signup performance in terms of the overall activity of a CSR. We reported these figures in the e-mails sent to CSRs in conditions RIF and RIFB.

To calculate the regional benchmark statistics for the RIFB e-mails, we computed the averages and $80^{th}$ percentile of the total individual signups (per 100 overall interactions) in each of Ontario's four regions.

The number of organ donor signups per day includes some outliers. Some of these are implausibly large values which are likely reporting errors. Others may indicate that on a given day, a CSR did not engage in any face-to-face activities with customers but worked on "mail-in" (i.e., bulk) activities such as processing received organ donor consent forms in paper form via postal mail. For this reason, in the analyses below, we winsorized the number of signups and the total number of daily customer interactions at the $99.9^{th}$ percentile.

### 2.2.4 Data quality and execution challenges

In querying and extracting the administrative data for the period before our first intervention wave, staff at our partner organization incurred a procedural error that led to misreporting the number of total transactions (i.e. services) for some CSRs (but not the number of signups per day). Specifically, the original data transfer, based on which we computed the individual and regional signup performance measures, did not include certain transactions. Consequently, in the first of our three e-mail waves, CSRs in conditions RIF and RIFB saw incorrect values for their number of signups per hundred transactions and, limited to Condition RIFB, the regional benchmark. However, the differences were in the order of a few decimal points, and it is unlikely that CSRs were aware of the number of signups they made in the previous six months per hundred transactions to that level of precision. Thus, it is also unlikely that the discrepancies substantively affected the behavior of the CSRs who only received information on their performance (condition RIF). One concern remains about the CSRs in condition RIFB. Because of the misreporting, some CSRs read that their performance was above/below the regional average or 80$^{th}$ percentile when that was not actually the case. This 'mismatch' occurred for 58 CSRs. That is, 58 out of 224 CSRs in the RIFB condition received incorrect information about their standing relative to the regional average or 80$^{th}$ percentile in the first intervention wave. Reassuringly, as we show below when describing our findings, excluding these CSRs for the period between the first and the second intervention wave (5,206 CSR-day observations) does not affect the estimates meaningfully.

### 2.2.5 Survey of CSRs

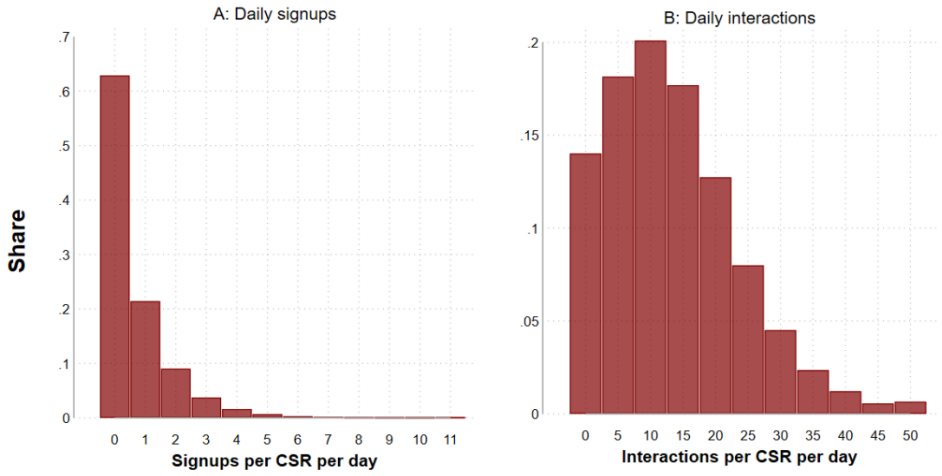
Following the RCT, we administered a survey to the CSRs on January 15, 2019. The survey's goals were to obtain individual CSR characteristics (e.g., age, gender, and tenure), investigate their motivations, and gauge whether they had paid attention to the intervention e-mails in the previous eighteen months and how they perceived the messages. In particular, we measured whether receiving the messages created pressure to perform. All CSRs who were active at the time of the survey and who were part of the field experiment received an e-mail inviting them to complete the survey, and 283 completed it (40.8% overall completion rate; 36.7%, 36.9%, and 48.2% in conditions R, RIF and RIFB respectively).

# 3. Data and Findings

## 3.1 Descriptive statistics and balance tests

The full dataset includes CSR-day-level information from November 1, 2016 through April 30, 2019, amounting to 295,884 observations. As a result of winsorization, the sample on which we ran our analyses excludes about 30,000 observations. The final dataset for analysis thus includes 265,475 observations on 694 CSRs, operating in 82 offices on 745 distinct days. The average (winsorized) new customer signups to the organ donor registry (in what follows, we refer to that as "signups") and total interactions per day over the entire period of observation are 0.65 (range: 0-11) and 13.4 (range: 0-50.7), respectively (Figure 2). The sample is balanced between conditions in the pre-intervention period (from November 2016 through June 15, 2017). At the individual CSR level, the F statistics for the joint significance of differences in the share of women, tenure, average signups, and average total transactions per day were not statistically significant.

**Figure 2: Distribution of the numbers of daily signups and customer interactions per CSR over the entire period of observation**




*Notes*: Notes: the figures report the empirical distribution of signups (panel A) and customer interactions (panel B) per day and CSR between November 1 2016 and April 30 2019. The figures exclude the 0.1th percentile of highest values for daily signups and interaction (values greater than 11 and 50.7, respectively). We rounded the values of daily transactions to the closest integer.

Comparing conditions pairwise, the differences in average daily interactions are significant at the 5% level between conditions RIF and RIFB. At the office level, which we used to assign CSRs to conditions, there are no (joint or pairwise) significant differences in signups, interactions, and average number of CSRs in each office.

We also have demographic information for the 283 CSRs who completed our survey in January 2019. The respondents' average age was 44 years (SD = 11, Median = 46), and about 85% of them identified as women, which is similar to the age (45.8) and gender distribution (85.2% female) of the entire population of CSRs according to ServiceOntario's HR records. These CSRs were almost equally distributed among those with less than 4 years of work at ServiceOntario, between 4 and 10 years, and more than 10 years. There were no differences in these characteristics across the experimental conditions.

## 3.2 Econometric specification

Equation (1) below shows the main econometric model that we estimate:

$$\begin{aligned} Y_{cd} = \alpha_0 &+ I_{MF}\alpha_{MF} + I_{MFB}\alpha_{MFB} + I_M[I_{t=1}\,\beta_{M1} + I_{t=2}\beta_{M2} + I_{t=3}\beta_{M3}] \\ &+ I_{MF}[I_{t=1}\beta_{MF1} + I_{t=2}\beta_{MF2} + I_{t=3}\beta_{MF3}] \\ &+ I_{MFB}[I_{t=1}\beta_{MFB1} + I_{t=2}\beta_{MFB2} + I_{t=3}\beta_{MFB3}] + \gamma X_{cd} + \varepsilon_{cd}. \end{aligned} \tag{1}$$

*Y*, the outcome variable, is either the number of signups by CSR *c* on day *d*, or a binary indicator for whether a CSR *c* made at least one new signup to the organ registry on day *d*. The variables $I_M$, $I_{MF}$, $I_{MFB}$ are binary indicators for whether a CSR was in conditions R, RIF or RIFB, respectively (value of one if they were, and zero if they were not). $I_{t=1}$, $I_{t=2}$ and $I_{t=3}$ take a value of one if an observation is in the period after the first, second, or third intervention wave, respectively, and zero in any other period. Therefore, the estimates of the "$\beta$" parameters indicate the average differences between the number of signups by CSRs in a given condition and post-intervention period, and the signups of CSRs in the same condition in the pre-intervention period. For example, the estimate $\hat{\beta}_{MF2}$ represents the average difference in daily signups between the period after the second intervention wave (and before the third) and the period before the first intervention wave for CSRs in condition RIF. Linear combinations of the parameters provide other treatment effects of interest. Within a given *condition* (e.g., condition RIFB), the difference between two "$\beta$" estimates represents the differential impact of a treatment in a given period as compared to the pre-intervention period – a "difference in difference" within

a condition; for example, $\hat{\beta}_{MFB3} - \hat{\beta}_{MFB2}$ estimates the differential impact of the second and third intervention waves for condition RIFB with respect to the period before the first intervention wave for that same condition. Within a given post-intervention *period*, we can establish the differential treatment effect between conditions by taking the difference between parameter estimates for a given period and different conditions. For instance, $\hat{\beta}_{MFB1} - \hat{\beta}_{MF1}$ estimates how condition RIFB changed signups in the first post-intervention period compared to the pre-intervention period, relative to the same change for condition RIF – a within-period, between-condition difference-in-differences.

If we take condition R as the reference case, a natural exercise is to measure the differential impact of the feedback conditions with respect to the encouragement-email reminder. In a more fine-grained distinction, we split the pre- and post-intervention periods in intervals of about sixty to seventy days each. By looking at shorter sub-periods separately, we can assess if any effect was higher immediately after the reception of the intervention e-mails or stable throughout an intervention period.

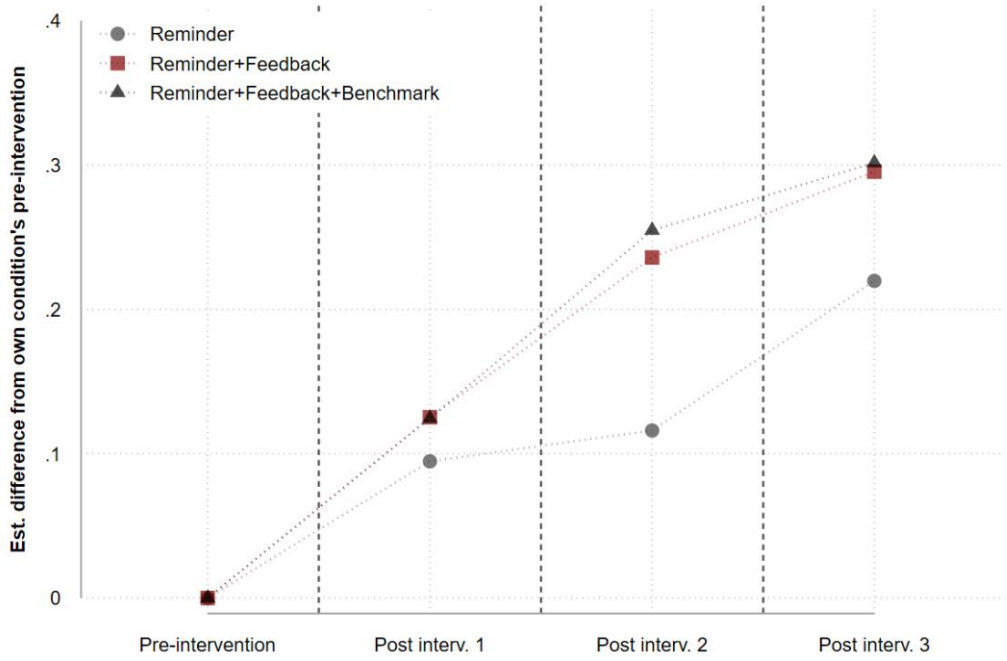
The vector $X_{cd}$ represents control variables. The data do not include many details about each CSR or their offices, but we have some relevant information that might make our estimates of interest more reliable. First, the number of signups and the likelihood of signing up at least one customer plausibly correlate positively with the activity level of a CSR in a given day. Therefore, we add as a control the total number of customer interactions a CSR had on a given day. The number of daily interactions may indicate the productivity of a CSR. However, for the most part, the daily volume depends on factors beyond the control of CSRs such as the number of hours worked in a day and the haphazard assignment to particular clients or types of services. We also derive a measure of CSR experience within our sample: the number of days of activity since we begin to observe a given CSR in our data. All models include month-year fixed effects to account for common time and seasonal trends. In additional specifications, we include CSR-level fixed effects to further control for unobserved individual differences. Although the assignment to experimental conditions was at the office level, some CSRs worked in more than one office during the study period while keeping the same condition assignment throughout. This variation allows us to add office-level fixed effects in some specifications. Finally, we account for possible correlation among observations for each CSR and within offices in a given intervention period

by estimating two-way cluster-robust standard errors (at the CSR- and the office-intervention period level).[12]

## 3.3 Main findings

Figure 3 reports the nine estimated "$\beta$" coefficients from equation (1) above. The values are from column (2) of Table 1 below.

**Figure 3: Effects of Reminder, Individual feedback, and Regional performance benchmark communications on the number of daily signups in each intervention period**




Notes: The figure reports the estimated average changes in daily organ donor signup per CSR after each of the three interventions, compared to the average daily signups in the pre-intervention periods for each condition (normalized to zero). Each intervention consisted in an e-mail whose content differed according to the experimental conditions to which a CSR was randomly assigned. The e-mails were sent on June 20, 2017 (1st intervention), January 29, 2018 (2nd intervention), and June 15, 2018 (3rd intervention). The values in the graphs correspond to the estimates reported in Column (2) of Table 1.

Daily individual signups increased significantly in all conditions compared to the pre-intervention average performance. The increase from one condition's own pre-intervention level was higher for conditions RIF and RIFB than R, especially after the second and third intervention waves. This suggests that receiving information about one's performance, with or without benchmarks, enhanced the signup activity of the CSRs.

[12] We use the Stata command reghdfe to allow for two-way standard error clustering.

**Table 1: Effects of Mail, Individual feedback, and Regional performance benchmark communications on the number of daily signups, in each intervention period: Main regression estimates**

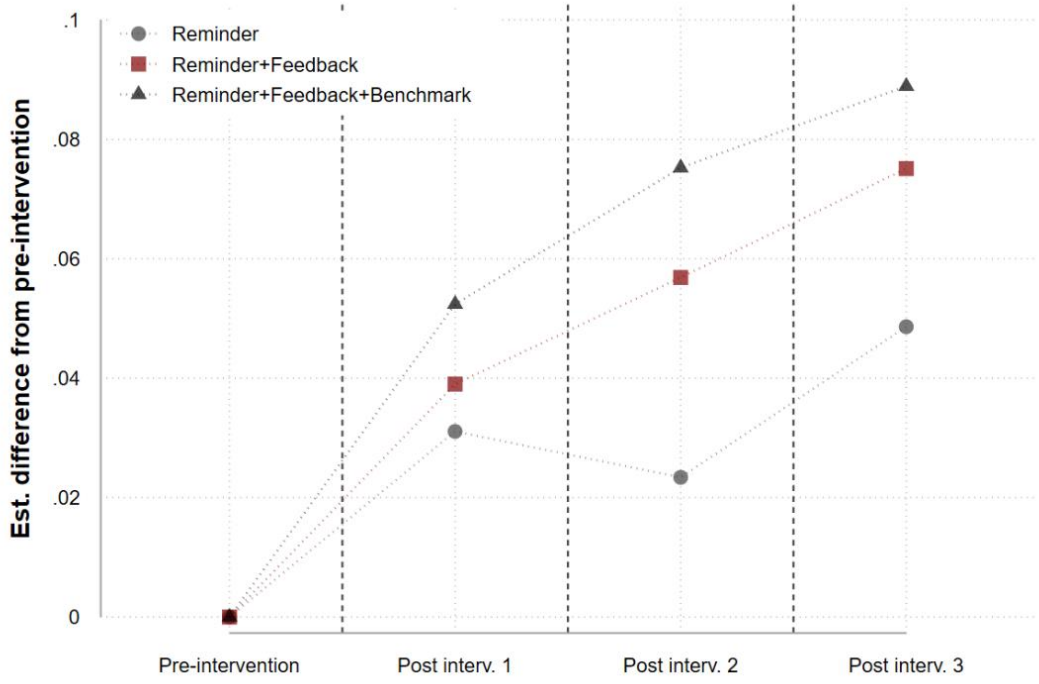
| Outcome variable: | Daily signups | | | *I* (Daily signups>0) |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| RIF | -0.050 | | -0.048 | |
| | (0.087) | | (0.071) | |
| RIFB | -0.114** | | -0.109* | |
| | (0.058) | | (0.058) | |
| R:1st int. | 0.107* | 0.095** | 0.109*** | 0.031** |
| | (0.058) | (0.039) | (0.032) | (0.014) |
| R:2nd int. | 0.131** | 0.116** | 0.141*** | 0.023 |
| | (0.060) | (0.054) | (0.051) | (0.018) |
| R:3rd int. | 0.254*** | 0.220*** | 0.265*** | 0.049** |
| | (0.073) | (0.068) | (0.060) | (0.021) |
| RIF:1st int. | 0.153 | 0.125** | 0.152*** | 0.039** |
| | (0.139) | (0.054) | (0.056) | (0.017) |
| RIF:2nd int. | 0.292** | 0.236*** | 0.282*** | 0.057** |
| | (0.120) | (0.064) | (0.058) | (0.023) |
| RIF: 3rd int. | 0.376*** | 0.295*** | 0.366*** | 0.075*** |
| | (0.114) | (0.076) | (0.069) | (0.025) |
| RIFB:1st int. | 0.129* | 0.124*** | 0.122*** | 0.052*** |
| | (0.070) | (0.044) | (0.033) | (0.013) |
| RIFB:2nd int. | 0.304*** | 0.255*** | 0.301*** | 0.075*** |
| | (0.083) | (0.060) | (0.052) | (0.019) |
| RIFB:3rd int. | 0.378*** | 0.302*** | 0.363*** | 0.089*** |
| | (0.083) | (0.066) | (0.064) | (0.022) |
| Constant | 0.424*** | -0.206* | 0.362*** | -0.044 |
| | (0.083) | (0.116) | (0.094) | (0.042) |
| CSR fixed effects | | x | | x |
| Office fixed effects | | | x | |
| RIF:1st int.-R:1st int. | 0.046 | 0.031 | 0.043 | 0.008 |
| RIFB:1st int.-R:1st int. | 0.022 | 0.030 | 0.014 | 0.021 |
| RIFB:1st int.-RIF:1st int. | -0.024 | -0.001 | -0.030 | 0.013 |
| RIF:2nd int.-R:2nd int. | 0.161 | 0.120** | 0.141*** | 0.335* |
| RIFB:2nd int.-R:2nd int. | 0.172** | 0.139*** | 0.160*** | 0.519*** |
| RIFB:2nd int.-RIF:2nd int. | 0.011 | 0.019 | 0.018 | 0.183 |
| RIF:3rd int.-R:3rd int. | 0.123 | 0.076 | 0.101*** | 0.265 |
| RIFB:3rd int.-R:3rd int. | .0.124* | 0.082** | 0.098*** | 0.403** |
| RIFB:3rd int.-RIF:3rd int. | 0.002 | 0.006 | -0.003 | 0.014 |
| Observations | 265,475 | 265,475 | 265,475 | 265,475 |
| R-squared | 0.145 | 0.297 | 0.173 | 0.253 |

Notes: The table reports estimates from OLS regressions where the unit of observation is a CSR on a given day in which that CSR is active. The regressors are binary indicators for experimental conditions and interactions between experimental conditions and interventions. R indicates the reminder e-mail condition, RIF the reminder + individual feedback condition, and RIFB the reminder + individual feedback + regional benchmark condition. The estimated parameter on a given interaction term (e.g., RIFB: 2nd int.) represents the estimated difference in daily signups between the period that the interaction term identifies, and the pre-intervention period for the same condition. The bottom part of the table reports relevant differences between estimated parameters. The regressions include variables that measure the number of days a CSR was present in the sample at any given date, its square, the total number of interactions of a CSR in a given day, and year-month fixed effects. Standard errors are clustered both at the CSR level and at the level of office-intervention period. * p<0.1, ** p<0.05, *** p<0.01.

This effect required some time, and more than one intervention, to show. The extra, statistically significant increases for conditions RIF and RIFB over condition R are, in the second and the third period, about 0.13-0.15 signups per day. With a pre-intervention overall average of 0.6 daily signups as the reference, this represents an increase of about 25% compared to providing basic information and encouragement.

Figure 4 displays the estimated treatment effects from the specification in column (4) of Table 1, where the outcome variable is a binary indicator for having signed up at least one customer in a given day. The patterns from Figures 3 and 4 suggest that the intervention had an impact both on the extensive margin (more CSRs signing customers to the organ donor registry) and on the intensive margin (a higher number of signups per CSR).

**Figure 4: Effects of Reminder, Individual feedback and Regional performance benchmark communications on the share of positive daily signups, in each intervention period**




Notes: The figure reports the estimated average changes in the likelihood of signing up at least one customer in a given day per CSR after each of the three interventions, compared to the average daily signups in the pre-intervention periods for each single condition (normalized to zero). Each intervention consisted in an e-mail whose content differed according to the experimental conditions to which a CSR was randomly assigned. The e-mails were sent on June 20 2017 (1st intervention), January 29 2018, (2nd intervention) and June 15 2018 (3rd intervention). The values in the graphs correspond to the estimates reported in Column (4) of Table 1.

## 3.4 Additional analyses and robustness checks

Tables 2 and 3 report the estimates from additional analyses to investigate other potential effects of the interventions and assess the robustness of the estimates from our main specification. The parameter estimates in Table 2 are from the model described in equation (1) but with different left-hand-side variables. In columns (1) and (2), the outcome variable is the ratio between daily signups and daily customer interactions (multiplied by 100). This is an alternative way to control for the overall activity of a CSR. The estimates in column (1) are from a model without the number of daily interactions among the regressors, whereas those in column (2) are from a model that also includes daily interactions on the right-hand side; the estimates of interest are very similar. The estimated treatment effects show the same patterns (in size and statistical significance) as those in Table 1 above. In column (3), the estimates are from a model where the outcome is the number of total daily *transactions*; as mentioned above, a CSR may provide more than one service (transaction) to the same client. One concern is that our various treatments may negatively affect the overall activity of a CSR because, for example, they might spend more time talking to customers about the organ donor registry in an attempt to sign them up. The estimates suggest that this substitution or "crowd out" effect did not occur.

In Columns (1) through (5) of Table 3, the estimates are from regressions where we either excluded part of the sample or controlled for additional variables. First, we dropped the observations pertaining to CSRs whose performance, as described in Section 2.2.4 above, was miscalculated in a way that ended up assigning them to the wrong side of the two regional benchmarks in the first intervention wave. Second, we restricted the sample to only the CSRs who answered the post-intervention survey. Third, we added an indicator variable for CSR-day observations in which the data report zero transactions and also ran the analyses excluding these observations from the sample. A report of zero transactions may indicate a coding error in the ServiceOntario system or that a CSR was active on a given day but not in direct customer-facing tasks. Finally, we limited the sample to CSRs who never worked on mail-in registrations. This restriction is another way to isolate observations with implausibly high reported daily signups. Column (6) reports estimates from a Poisson model given the discrete-count nature of our primary outcome variable.[13] All columns show results similar to those in Table 1.

[13] We employ, specifically, a Poisson pseudo-likelihood regression with multiple levels of fixed effects (see Correia, Guimarães, Zylkin [2019a, b]. Stata command: ppmlhdfe).

**Table 2: Effects of Reminder, Individual feedback, and Regional performance benchmark communications on the number of daily signups in each intervention period: Alternative outcomes**

| Outcome variable: | 100*Daily signups/interactions | | Daily transactions |
|---|---|---|---|
| | (1) | (2) | (4) |
| R:1st int. | 0.762*** | 0.769*** | 0.071 |
| | (0.273) | (0.28) | (0.648) |
| R:2nd int. | 1.045** | 1.056** | 0.254 |
| | (0.433) | (0.450) | (1.004) |
| R:3rd int. | 1.318*** | 1.331*** | 0.302 |
| | (0.480) | (0.496) | (1.143) |
| RIF:1st int. | 0.991*** | 0.981** | -0.418 |
| | (0.370) | (0.387) | (1.001) |
| RIF:2nd int. | 1.584*** | 1.551*** | -1.119 |
| | (0.483) | (0.506) | (1.288) |
| RIF: 3rd int. | 1.720*** | 1.735*** | 0.190 |
| | (0.533) | (0.558) | (1.347) |
| RIFB:1st int. | 1.118*** | 1.123*** | 0.075 |
| | (0.274) | (0.284) | (0.669) |
| RIFB:2nd int. | 2.158*** | 2.168*** | 0.188 |
| | (0.448) | (0.461) | (0.956) |
| RIFB:3rd int. | 2.280*** | 2.330*** | 1.272 |
| | (0.475) | (0.495) | (1.212) |
| Constant | 0.925 | 2.106** | 31.825*** |
| | (0.874) | (0.883) | (1.535) |
| CSR fixed effects | x | x | x |
| RIF:1st int.-R:1st int. | 0.229 | 0.213 | -0.489 |
| RIFB:1st int.-R:1st int. | 0.356 | 0.354 | 0.003 |
| RIFB:1st int.-RIF:1st int. | 0.127 | 0.142 | 0.492 |
| RIF:2nd int.-R:2nd int. | 0.539* | 0.495 | -1.373 |
| RIFB:2nd int.-R:2nd int. | 1.113*** | 1.112*** | -0.652 |
| RIFB:2nd int.-RIF:2nd int. | 0.574* | 0.617* | 1.308 |
| RIF:3rd int.-R:3rd int. | 0.401 | 0.404 | -0.113 |
| RIFB:3rd int.-R:3rd int. | 0.962*** | 0.999*** | 0.070 |
| RIFB:3rd int.-RIF:3rd int. | 0.560* | 0.594* | 1.083 |
| Observations | 265,475 | 265,475 | 265,475 |
| R-squared | 0.120 | 0.118 | 0.359 |

Notes: The table reports estimates from OLS regressions where the unit of observation is a CSR on a given day in which that CSR is active. The regressors are binary indicators for experimental conditions and interactions between experimental conditions and interventions. R indicates the reminder e-mail condition, RIF the reminder + individual feedback condition, and RIFB the reminder + individual feedback + regional benchmark condition. The estimated parameter on a given interaction term (e.g., RIFB: 2nd int.) represents the estimated difference in daily signups between the period that the interaction term identifies, and the pre-intervention period for that condition. The bottom part of the table reports relevant differences between estimated parameters. The regressions include variables that measure the number of days a CSR was present in the sample at any given date, its square, the total number of interactions of a CSR in a given day, and year-month fixed effects. Standard errors are clustered both at the CSR level and at the level of office-intervention period. * p<0.1, ** p<0.05, *** p<0.01.

**Table 3: Effects of Reminder, Individual feedback, and Regional performance benchmark communications on the number of daily signups in each intervention period: Robustness to subsample and econometric specifications**

| Outcome variable: | Daily signups | | | | | |
|---|---|---|---|---|---|---|
| Sample: | Exclude performance bechmark mismatches | CSRs who answered the survey | Full | Exclude CSR observation with no customer interactions | Exclude CSRs with any mailin | Full |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| R:1st int. | 0.088** | 0.073 | 0.096** | 0.092** | 0.115** | 0.091*** |
| | (0.038) | (0.054) | (0.038) | (0.039) | (0.045) | (0.027) |
| R:2nd int. | 0.111** | 0.111 | 0.117** | 0.113** | 0.184*** | 0.113*** |
| | (0.054) | (0.076) | (0.052) | (0.054) | (0.062) | (0.043) |
| R:3rd int. | 0.215*** | 0.204** | 0.224*** | 0.234*** | 0.278*** | 0.217*** |
| | (0.068) | (0.084) | (0.067) | (0.071) | (0.087) | (0.053) |
| RIF:1st int. | 0.134** | 0.119* | 0.126** | 0.130*** | 0.163*** | 0.114*** |
| | (0.056) | (0.062) | (0.051) | (0.049) | (0.054) | (0.038) |
| RIF:2nd int. | 0.245*** | 0.221*** | 0.242*** | 0.279*** | 0.301*** | 0.231*** |
| | (0.066) | (0.079) | (0.062) | (0.065) | (0.075) | (0.053) |
| RIF: 3rd int. | 0.305*** | 0.326*** | 0.300*** | 0.334*** | 0.354*** | 0.300*** |
| | (0.077) | (0.088) | (0.073) | (0.077) | (0.093) | (0.062) |
| RIFB:1st int. | 0.125*** | 0.117* | 0.122*** | 0.122*** | 0.127** | 0.140*** |
| | (0.045) | (0.060) | (0.044) | (0.045) | (0.053) | (0.033) |
| RIFB:2nd int. | 0.254*** | 0.260*** | 0.251*** | 0.262*** | 0.266*** | 0.260*** |
| | (0.063) | (0.078) | (0.059) | (0.063) | (0.062) | (0.051) |
| RIFB:3rd int. | 0.302*** | 0.291*** | 0.300*** | 0.330*** | 0.313*** | 0.324*** |
| | (0.067) | (0.081) | (0.064) | (0.068) | (0.081) | (0.056) |
| No interactions in a given day | | | -0.189*** | | | |
| | | | (0.031) | | | |
| Constant | -0.186 | -0.332** | -0.135 | -0.131 | -0.263 | |
| | (0.120) | (0.165) | (0.115) | (0.125) | (0.164) | |
| CSR fixed effects | x | x | x | x | x | x |
| Specification | Linear | Linear | Linear | Linear | Linear | Poisson |
| Observations | 259,330 | 120,609 | 265,475 | 242,923 | 134,953 | 264,911 |
| R-squared | 0.299 | 0.288 | 0.298 | 0.287 | 0.339 | |

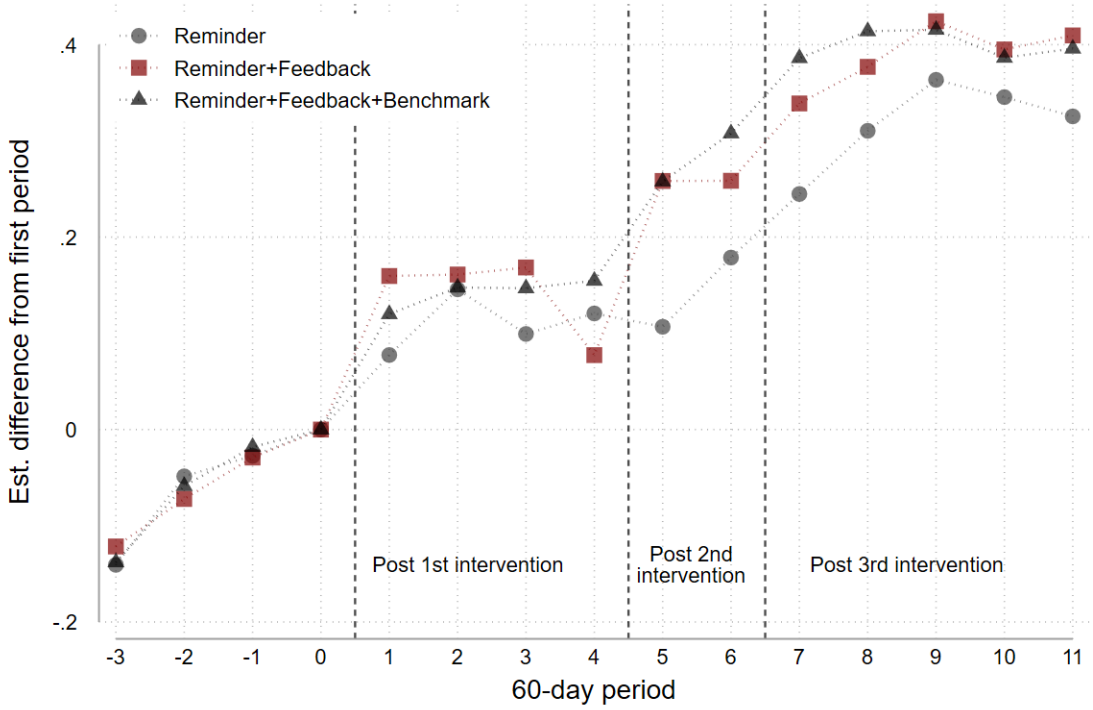
Notes: Columns (1) through (5) reports estimates from OLS regressions where the unit of observation is a CSR on a given day in which that CSR is active. Column (6) reported the estimated marginal effects from a Poisson regression. The regressors are binary indicators for experimental conditions and interactions between experimental conditions and interventions. R indicates the reminder e-mail condition, RIF the reminder + individual feedback condition, and RIFB the reminder + individual feedback + regional benchmark condition. The estimated parameter on a given interaction term (e.g., RIFB: 2nd int.) represents the estimated difference in daily signups between the period that the interaction term identifies, and the pre-intervention period for the same condition. The regressions include variables that measure the number of days a CSR was present in the sample at any given date, its square, the total number of interactions of a CSR in a given day, and year-month fixed effects. Standard errors are clustered both at the CSR level and at the level of office-intervention period. * p<0.1, ** p<0.05, *** p<0.01.

## 3.5 The dynamics of the treatment effects

Figure 5 shows the estimated treatment effects from a specification with fourteen subperiods instead of four. Each time interval is between sixty and seventy days (Table A1 in the appendix

reports the complete set of regression estimates). The estimates corresponding to points 1, 5 and 7 on the x-axis indicate the average performance change between the first (approximately) two months after the 1st, 2nd, and 3rd intervention dates, respectively, and the two months before those dates.

**Figure 5: Effects of Reminder, Individual feedback, and Regional performance benchmark communications on the number of daily signups by sub-periods within each intervention**

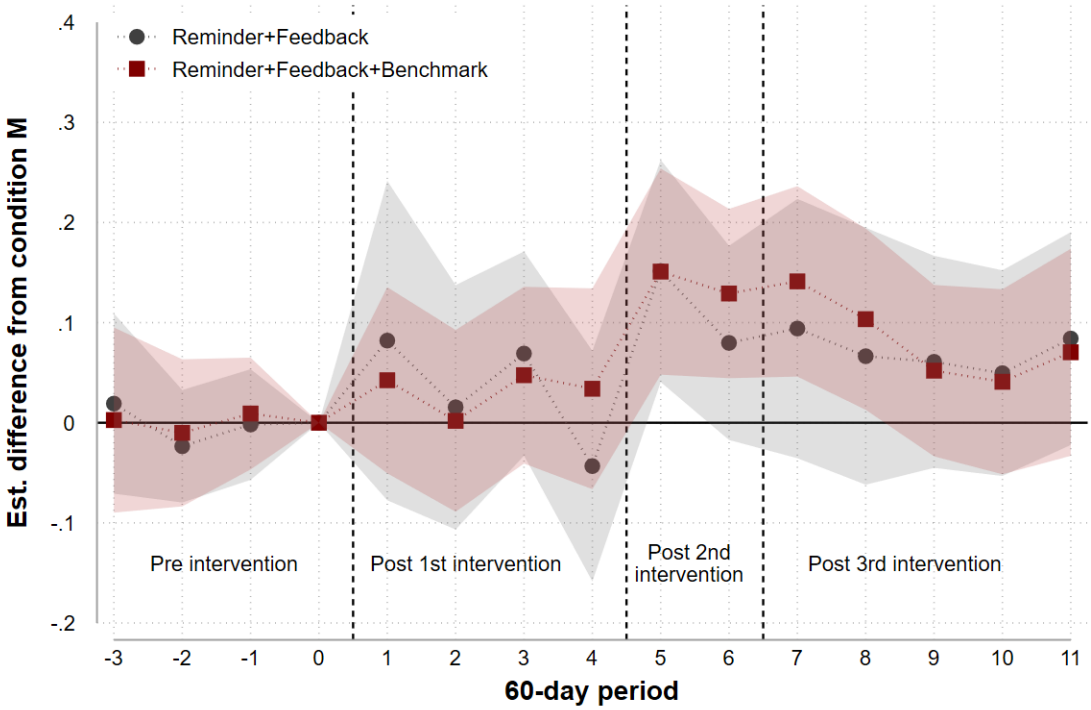



Notes: The figure reports the estimated average changes in daily organ donor signup per CSR in sub-periods of sixty to seventy days within each post-intervention period, compared to the average daily signups in subperiod immediately preceding the first wave of treatment (subperiod 0 on the x-axis), for each single condition. The estimate in correspondence of point 1 (5, 7) on the x-axis indicates the average performance change between the first approximately two months after the first (second, third) intervention date, and the two months before. The values at point 5 represent the estimate difference between the first two months after the second intervention Each intervention consisted in an e-mail whose content differed according to the experimental conditions to which a CSR was randomly assigned. The e-mails were sent on June 20, 2017 (1st intervention), January 29, 2018, (2nd intervention) and June 15, 2018 (3rd intervention).

The response to the e-mails concentrated mainly in the few weeks after reception. This was especially the case for those including feedback information (RIF and RIFB) after the second intervention wave. These increases were followed by tapering, but signup rates remained higher than baseline values. To better gauge the differential impact of the two feedback-based

conditions over the simple e-mail reminder, in terms of size, significance, and the time when it occurred, Figure 6 reports, for each subperiod, the "differences-in-differences" for conditions RIF and RIFB relative to condition R, again setting the two months immediately before the first intervention wave as the reference, as in an event study. The graph shows more explicitly when the treatments were particularly effective. The estimates from these additional tests are comparable to those from our main specifications.

**Figure 6: Effects of Reminder, Individual feedback, and Regional performance benchmark communications on the number of daily signups, by sub-periods within each intervention: differences from the Mail condition**



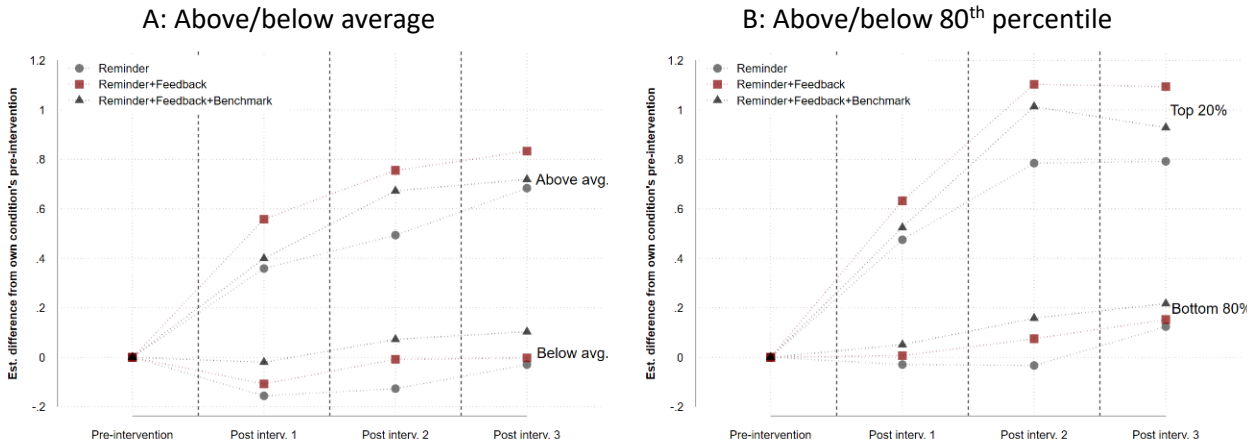

Notes: The figures reports the difference-in-differences estimates of the effect of the reminder + individual feedback and the reminder + individual feedback + regional benchmark conditions on daily signups per CSR with respect to the Mail condition, in sub-periods of sixty to seventy days within each post-intervention period, where the reference is the subperiod immediately preceding the first wave of treatment (subperiod 0 on the x-axis), for each single condition. Shaded areas represents 95% confidence intervals. Each intervention consisted in an e-mail whose content differed according to the experimental conditions to which a CSR was randomly assigned. The e-mails were sent on June 20 2017 (1st intervention), January 29 2018, (2nd intervention) and June 15 2018 (3rd intervention).

## 3.6 Heterogeneous effects by prior performance

One recurring question in the literature on the effects of performance feedback is whether this information is equally effective on the whole distribution of prior outcomes or only on

individuals in particular intervals of the performance distribution. The evidence from previous research is mixed and context-specific. As shown in Figure 7 and Table 5, in our context, encouragement (R) and the two types of feedback (RIF and RIFB) were effective almost exclusively on individuals with higher-than-average performance levels. Panels A and B of Figure 7 report the parameter estimates from the same model whose estimates are in columns (1)-(2) and (3)-(4) of Table 5, respectively.

**Figure 7: Effects of Reminder, Individual feedback, and Regional performance benchmark communications on the number of daily signups, by signup performance, in each intervention period**




Notes: The figures reports the estimated average changes in daily organ donor signup per CSR after each of the three interventions, compared to the average daily signups in the pre-intervention periods for each single condition. In panel A, the estimates per condition are separate between CSRs whose performance in the period immediately before a given intervention was above the regional average (Table 5, column 2), and those with a performance below average (Table 5, column 1). In Panel B, the separation is between the CSRs with pre-intervention performance in the top 20% (Table 5, column 4), and those with performance in the bottom 80% in a given region (Table 5, column 3). Each intervention consisted in sending an e-mail whose content differed according to the experimental conditions to which a CSR was randomly assigned. The e-mails were sent on June 15 2017 (1st intervention), January 29 2018, (2nd intervention) and June 20 2018 (3rd intervention).

The positive effects of the intervention e-mails were concentrated on the high performers, with little difference between those above average and the smaller sample of those in the top 20% within a region. The response of the CSRs in the R condition was, again, less strong than in the other conditions. Observing the regional benchmark, however, did not have a large additional effect over just observing one's individual past signup performance. The absence of a negative effect of feedback provision among below-average CSRs is reassuring, as it suggests that

relatively low-performing CSRs did not respond by further reducing their effort (this could have happened if, for example, low-performing CSRs had been discouraged by the feedback).

**Table 5: Effects of Reminder, Individual feedback, and Regional performance benchmark communications on the number of daily signups in each intervention period: Heterogeneous effects**

| Outcome variable: | Daily signups | | | |
|---|---|---|---|---|
| Sample: | Signup performance below average | Signup performance above average | Bottom 80% signup performance | Top 20% signup performance |
| | (1) | (2) | (3) | (4) |
| RIF | -0.047 | -0.052 | -0.049 | -0.052 |
| | (0.087) | (0.088) | (0.087) | (0.088) |
| RIFB | -0.128** | -0.100 | -0.121** | -0.104* |
| | (0.054) | (0.064) | (0.055) | (0.062) |
| R:1st int. | -0.156*** | 0.359*** | -0.029 | 0.475*** |
| | (0.050) | (0.102) | (0.053) | (0.177) |
| R:2nd int. | -0.127** | 0.494*** | -0.033 | 0.784*** |
| | (0.056) | (0.097) | (0.051) | (0.156) |
| R:3rd int. | -0.030 | 0.684*** | 0.124** | 0.792*** |
| | (0.063) | (0.115) | (0.061) | (0.185) |
| RIF:1st int. | -0.108 | 0.558*** | 0.007 | 0.632*** |
| | (0.098) | (0.210) | (0.099) | (0.243) |
| RIF:2nd int. | -0.008 | 0.755*** | 0.075 | 1.103*** |
| | (0.105) | (0.168) | (0.103) | (0.219) |
| RIF: 3rd int. | -0.002 | 0.833*** | 0.152 | 1.094*** |
| | (0.110) | (0.151) | (0.110) | (0.217) |
| RIFB:1st int. | -0.020 | 0.399*** | 0.051 | 0.524*** |
| | (0.074) | (0.090) | (0.067) | (0.115) |
| RIFB:2nd int. | 0.073 | 0.672*** | 0.158** | 1.013*** |
| | (0.076) | (0.112) | (0.070) | (0.190) |
| RIFB:3rd int. | 0.103 | 0.719*** | 0.217*** | 0.928*** |
| | (0.069) | (0.118) | (0.068) | (0.183) |
| Constant | 0.596*** | 0.084 | 0.518*** | 0.092 |
| | (0.082) | (0.120) | (0.101) | (0.133) |
| Observations | 168,568 | 147,115 | 212,071 | 103,612 |
| R-squared | 0.114 | 0.196 | 0.126 | 0.181 |

Notes: The table reports estimates from OLS regressions where the unit of observation is a CSR on a given day in which that CSR is active. The regressors are binary indicators for experimental conditions, and interactions between experimental conditions and interventions. R indicates the reminder e-mail condition, RIF the reminder + individual feedback condition, and RIFB the reminder + individual feedback + regional benchmark condition. The estimated parameter on a given interaction term (e.g., RIFB: 2nd int.) represents the estimated difference in daily signups between the period that the interaction term identifies, and the period before the first intervention for that same condition. The regressions include variables that measure the number of days a CSR was present in the sample at any given date, its square, the total number of interactions of a CSR in a given day, and year-month fixed effects. Standard errors are clustered both at the CSR level and at the level of office-intervention period. * p<0.1, ** p<0.05, *** p<0.01.

## 3.7 Findings from the post-intervention survey

Figure 8 shows the percentage of respondents, by experimental condition, who answered "agree" or "strongly" agree to a 5-point Likert scale (strongly disagree, disagree, neither agree nor disagree, agree, strongly agree) to eight statements. These answers represent the extent to which CSRs perceived signing up customers to the organ donor registry as a meaningful part of their job and whether the intervention messages were a source of motivation or pressure.[14] Although we found only little differences in signup rates between the two types of performance feedback, the results from the survey indicate that the CSRs perceived the two types of e-mails differently. Panels A through F show that the presence of both individual and regional benchmark performance feedback (RIFB) had a stronger motivating effect on CSRs than providing only individual performance feedback (RIF) or just an encouragement reminder (R). The differences between the support rates for these statements between the RIF and RIFB conditions are statistically significant, whereas the difference between conditions R and RIF are not. Panel G shows that providing performance feedback increased CSRs' sense of "competition" with their peers, especially when CSRs also received information about the regional benchmark.[15] However, providing only individual performance feedback without any reference also resulted in a statistically significant increase in CSRs' feelings of pressure to ask customers to join the donor registry. Thus, despite similar effects on the primary outcome of interest, the two types of performance feedback may have generated differential additional benefits and costs to the CSRs.[16]

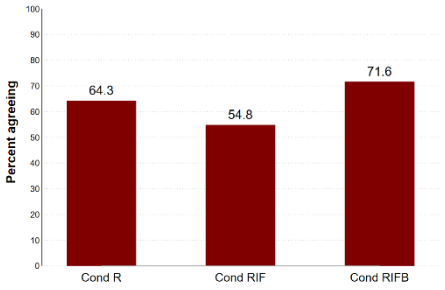
[14] Almost 79% of the respondents stated that they remembered receiving the intervention email, with only about 4% stating that they did not remember (17% were "not sure"). Moreover, we know from the administrative data that almost all CSRs (96.5%) did open the intervention emails.

[15] We also find that CSRs in conditions RIF and RIFB were more likely to report that they discussed the emails they received with their colleagues.

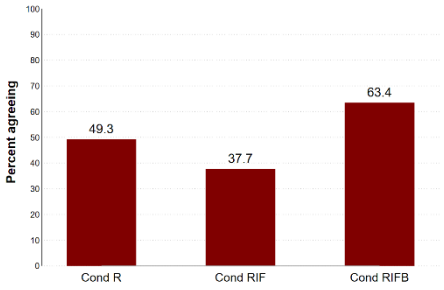
[16] In addition, the survey asked CSRs to rate statements that did not concern the organ donor registry signup activities, such as "This job is a good for me" and "I am contributing to the public good by doing this job". We did not find differences in these answers between the experimental conditions.

**Figure 8: Percent of survey respondents who reported they "agree" or "strongly agree" to various statements in the post-intervention survey**

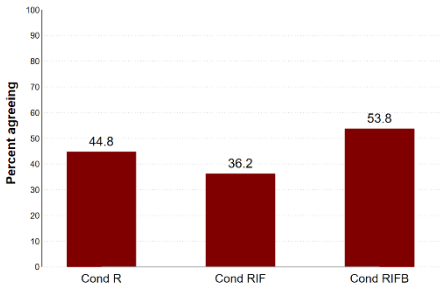
Getting customers to join the organ donor registry makes my job more meaningful



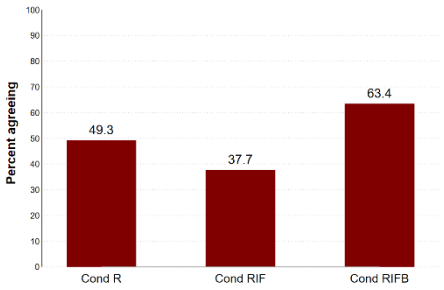

By remembering to ask customers consistently, I can increase registrations



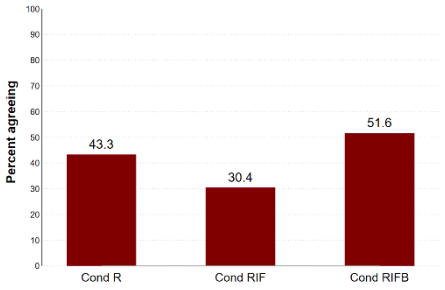

The e-mail (s) made me feel motivated to register more organ donors



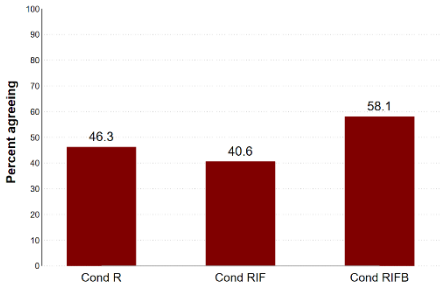

The e-mail (s) made me think it was possible to increase the number of organ donors



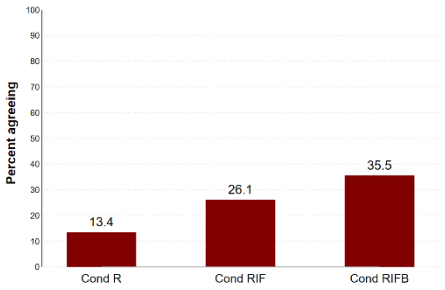

The e-mail (s) increased how important I believe organ donor registration is



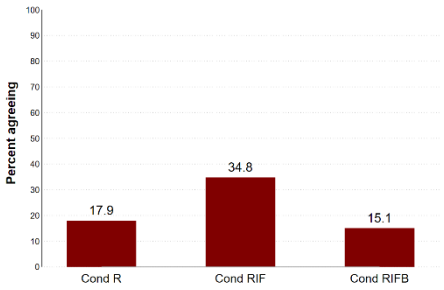

The e-mail (s) made me think senior management highly values registering more organ donors




The e-mail (s) made me feel like I was in competition with my colleagues




The e-mail (s) made me feel pressured to ask customers to join the organ donor registry



# 4. Discussion and conclusions

Theories in economics and behavioral science predict that performance feedback can affect employee performance even for activities that are not directly rewarded with explicit incentives. However, the theory is ambiguous about the direction of the effects, and, moreover, the available evidence is somewhat context-specific. Our study evaluated the effect of providing performance feedback to public-sector customer service representatives (CSRs) whose tasks include enrolling residents in the organ donor registry. Our randomized controlled trial results indicate that providing performance feedback, with or without a reference benchmark (i.e. regional average and 80$^{th}$ percentile performance), positively affected performance. Specifically, daily signup performance for CSRs who received performance feedback in addition to the standard reminder increased by about a quarter over the signups of those who only received the standard reminder (i.e. information and encouragement but no performance feedback). One concern with performance feedback is that it might backfire. Specifically, high-performing individuals might "relax" and reduce their effort, whereas low-performing ones might get discouraged and further reduce their effort (Bandiera et al. 2013; Allcott and Kessler 2019). We find no such adverse effects in our context. Instead, we find a strong, positive effect of performance feedback on high-performing CSRs and no effect on low-performing CSRs.

Another concern with providing employees with performance feedback is that it might create new psychological pressures to perform (Reiff et al., 2022). The evidence from our survey suggests that some of these considerations may be founded in our context. In particular, the fact that respondents in the RIF condition reported experiencing greater pressure from receiving the intervention e-mails suggests that there might be psychological costs to receiving behavioral feedback without any frame of reference. These findings should be taken into account when evaluating performance feedback policies.

Despite the success of our light-touch feedback interventions to achieve a statistically significant and relatively large 25% increase in daily organ donor registrations, it is important to acknowledge that the absolute take-up effect of the interventions is very low: about 0.13-0.15 signups per day, with a pre-intervention overall average of 0.6 daily signups as the reference.[17] Thus, implementing these types of feedback nudges alone cannot be expected to substantially

[17] The average intervention implemented by nudge units in the field generates a 8.1% increase over control, or 1.4 percentage points according to DellaVigna and Linos (2022).

reduce the intention-action gap when it comes to organ donor registrations. According to the Global Observatory on Donation and Transplantation (2021), just over 153,000 transplants were performed across all eighty-two member states in 2019, meeting less than 10% of the estimated need. In Ontario, only about 35% of the population is currently registered, and with significant medical and practical limitations restricting under what circumstances donations can happen after death, it seems unlikely that incremental improvements alone will meet the annual need of some 1,500 people waiting for a transplant in the Province.[18] Redirecting efforts to focus on system-level (rather than individual-level) policy frameworks (Chater and Loewenstein, in press), such as switching to a market approach with a regulated floor price in an attempt to encourage a massive increase in living donation (and possibly donation by the deceased), may be more adequate (see for example Taylor, 2005; Becker and Elías, 2007; Lacetera et al., 2013; and Sönmez & Ünver, 2017). Nevertheless, system-level changes can be hard and slow to get approved and implemented, while our study shows that there are improvements a government organization like ServiceOntario can make that are within its immediate realm of possibilities and do not require large resources and time to save or at least enhance the quality of life of a few more residents.

[18] Latest statistics may be retrieved from https://www.giftoflife.on.ca/en/publicreporting.htm.

# References


Allcott, H. (2011), Social norms and energy conservation, Journal of Public Economics, 95(9–10), 1082-1095. https://www.sciencedirect.com/science/article/abs/pii/S0047272711000478

Ball, S., Eckel, C., Grossman, P. J., & Zame, W. (2001). Status in markets. *The Quarterly Journal of Economics*, *116*(1), 161-188.

Bandiera, O., Barankay, I. and Rasul, I., 2013. Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association*, 11(5), 1079-1114.

Bandiera, O., Larcinese, V., & Rasul, I. (2015). Blissful ignorance? A natural experiment on the effect of feedback on students' performance. Labour Economics, 34, 13-25.

Becker, G. S. & Elías, J.J. (2007). Introducing incentives in the market for live and cadaveric organ donations. *Journal of Economic Perspectives*, 21(3), 3–24.

Charness, G., & Grosskopf, B. (2001). Relative payoffs and happiness: an experimental study. *Journal of Economic Behavior & Organization*, *45*(3), 301-328.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The quarterly journal of economics*, *117*(3), 817-869.

Charness, G., Masclet, D. and Villeval, M.C., 2011. Competitive preferences and status as an incentive: Experimental evidence. *CIRANO-Scientific Publications* 2011s-07.

Charness, G., Masclet, D. and Villeval, M.C., 2014. The dark side of competition for status. *Management Science*, 60(1), pp.38-55.Chater, N. and Loewenstein, G. (2022). The i-Frame and the s-Frame: How Focusing on the Individual-Level Solutions Has Led Behavioral Public Policy Astray. *Working Paper*, Available at SSRN 4046264.

Chater N. & Loewenstein, G. (in press). The i-frame and the s-frame: How focusing on individual-level solutions has led behavioral public policy astray. *Behavioral and Brain Sciences*, 2022 Sep 5:1-60. doi: 10.1017/S0140525X22002023.

Croson, R., Handy, F., & Shang, J. (2009). Keeping up with the Joneses: The relationship of perceived descriptive social norms, social information, and charitable giving. *Nonprofit Management and Leadership*, *19*(4), 467-489.

DellaVigna, S. and Linos, E. (2022). RCTs to Scale: Comprehensive Evidence from Two Nudge Units. *Econometrica*, 90 (1), 81–116.

Duflo, E., & Saez, E. (2002). Participation and investment decisions in a retirement plan: The influence of colleagues' choices. *Journal of public Economics*, *85*(1), 121-148.

Edelman, B. and Larkin, I. (2015). Social comparisons and deception across workplace hierarchies: Field and experimental evidence. *Organization Science*, *26*(1): 78-98.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives don't work to modify behavior. *Journal of Economic Perspectives*, 25(4), 191–210.

Global Observatory on Donation and Transplantation (GODT), (2021). Organ Donation and Transplantation Activities 2019 Report. See http://www.transplant-observatory.org/wp-content/uploads/2021/08/ImpactCOVID19-2.pdf, accessed: October 13, 2021.

Lacetera, N., Macis, M., & Slonim, R. (2013). Economic rewards to motivate blood donations. *Science*, 340(6135), 927-928.

Larkin, I., Pierce, L. and Gino, F. (2012). The psychological costs of pay-for-performance: Implications for the strategic compensation of employees. *Strategic Management Journal*, 33(10): 1194-1214.

Munshi, K. and Myaux, J., 2006. Social norms and the fertility transition. *Journal of Development Economics*, *80*(1), pp.1-38.

Reiff, J.S., Zhang, J.C., Gallus, J., Dai, H., Pedley, N.M., Vangala, S., Leuchter, R.K., Goshgarian, G., Fox, C.R., Han, M. and Croymans, D.M. (2022). When peer comparison information harms physician well-being. *Proceedings of the National Academy of Sciences*, *119*(29), p.e2121730119.

Robitaille, N., Mazar, N. Tsai, C.I., Haviv, A.M. and Hardy, E. (2021). Increasing Organ Donor Registrations with Behavioral Interventions: A Field Experiment, *Journal of Marketing*. https://journals.sagepub.com/doi/full/10.1177/0022242921990070.

Sönmez, T., & Ünver, M. U. (2017). Market design for living-donor organ exchanges: An economic policy perspective. *Oxford Review of Economic Policy*, *33*(4), 676-704.

Taylor, J. S. (2005). Stakes and Kidneys: Why Markets in Human Body Parts are Morally Imperative. *Live Questions in Ethics and Moral Philosophy*. Routledge.

Tran, A., & Zeckhauser, R. (2012). Rank as an inherent incentive: Evidence from a field experiment. *Journal of Public Economics*, *96*(9-10), 645-650.

Trillium Gift of Life Network (2017). Trillium Gift of Life Network Annual Report 2016/17. Available at https://www.giftoflife.on.ca/resources/pdf/Trillium_AR_16-17_ENG_access_10_12_2017.pdf. Accessed on 10/5/2021.

Trillium Gift of Life (2021). Public reporting. Available at: https://www.giftoflife.on.ca/en/publicreporting.htm#registration3yrhistory-cal. Accessed on 10/5/2021.

# APPENDIX

**Table A1: Effects of Mail, Individual feedback and Regional performance benchmark communications on the number of daily signups, by sub-periods within each intervention: differences from the Mail condition**

| | | Outcome variable: Daily signups |
|---|---|---|
| Pre-intervention | Subperiod -3 | -0.141** |
| | | (0.058) |
| | Subperiod -2 | -0.049 |
| | | (0.041) |
| | Subperiod -1 | -0.027 |
| | | (0.027) |
| | Subperiod -3 X RIF | 0.019 |
| | | (0.046) |
| | Subperiod -3 X RIFB | 0.003 |
| | | (0.047) |
| | Subperiod -2 X RIF | -0.023 |
| | | (0.029) |
| | Subperiod -2 X RIFB | -0.010 |
| | | (0.037) |
| | Subperiod -1 X RIF | -0.002 |
| | | (0.028) |
| | Subperiod -1 X RIFB | 0.009 |
| | | (0.028) |
| Post 1st intervention | Subperiod 1 | 0.078** |
| | | (0.037) |
| | Subperiod 2 | 0.146*** |
| | | (0.035) |
| | Subperiod 3 | 0.099** |
| | | (0.048) |
| | Subperiod 4 | 0.121* |
| | | (0.067) |
| | Subperiod 1 X RIF | 0.082 |
| | | (0.081) |
| | Subperiod 1 X RIFB | 0.042 |
| | | (0.047) |
| | Subperiod 2 X RIF | 0.015 |
| | | (0.062) |
| | Subperiod 2 X RIFB | 0.002 |
| | | (0.046) |
| | Subperiod 3 X RIF | 0.069 |
| | | (0.052) |
| | Subperiod 3 X RIFB | 0.048 |
| | | (0.045) |
| | Subperiod 4 X RIF | -0.043 |
| | | (0.059) |
| | Subperiod 4 X RIFB | 0.034 |
| | | (0.051) |
| Post 2nd intervention | Subperiod 5 | 0.107 |
| | | (0.075) |
| | Subperiod 6 | 0.179** |
| | | (0.080) |
| | Subperiod 5 X RIF | 0.152*** |
| | | (0.056) |
| | Subperiod 5 X RIFB | 0.151*** |
| | | (0.052) |
| | Subperiod 6 X RIF | 0.080 |
| | | (0.049) |
| | Subperiod 6 X RIFB | 0.129*** |
| | | (0.043) |

(continues at next page)

| | | |
|---|---|---|
| | Subperiod 7 | 0.245*** |
| | | (0.089) |
| | Subperiod 8 | 0.311*** |
| | | (0.093) |
| | Subperiod 9 | 0.363*** |
| | | (0.095) |
| | Subperiod 10 | 0.346*** |
| | | (0.099) |
| | Subperiod 11 | 0.325*** |
| | | (0.106) |
| | Subperiod 7 X RIF | 0.094 |
| | | (0.066) |
| | Subperiod 7 X RIFB | 0.141*** |
| | | (0.048) |
| Post 3rd intervention | Subperiod 8 X RIF | 0.066 |
| | | (0.065) |
| | Subperiod 8 X RIFB | 0.103** |
| | | (0.046) |
| | Subperiod 9 X RIF | 0.061 |
| | | (0.054) |
| | Subperiod 9 X RIFB | 0.052 |
| | | (0.044) |
| | Subperiod 10 X RIF | 0.050 |
| | | (0.052) |
| | Subperiod 10 X RIFB | 0.041 |
| | | (0.047) |
| | Subperiod 11 X RIF | 0.084 |
| | | (0.054) |
| | Subperiod 11 X RIFB | 0.070 |
| | | (0.053) |
| | Constant | -0.093 |
| | | (0.116) |
| | Observations | 265,475 |
| | R-squared | 0.297 |

Notes: The table reports the difference-indifferences estimates of the effect of the Reminder+Feedback (RIF) and the Reminder+Feedback+Benchmark (RIFB) conditions on daily signups per CSR with respect to the Mail condition, in sub-periods of sixty to seventy days within each post-intervention period, where the reference is the subperiod immediately preceding the first wave of treatment (subperiod 0, whose indicator variable is omitted from the regression), for each single condition. Each intervention consisted in sending an e-mail whose content differed according to the experimental conditions to which a CSR was randomly assigned. The e-mails were sent on June 20 2017 (1st intervention), January 29 2018, (2nd intervention) and June 15 2018 (3rd intervention). The regression includes variables that measures the number of days a CSR was present in the sample at any given date, its square, the total number of interactions of a CSR in a given day, and year-month and CSR-level fixed effects. Standard errors are clustered both at the CSR level and at the level of office-intervention period. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.