Pre-Analysis Plan: *"The Voice of the Emperor and Personalized Messages during the COVID-19 Pandemic"*

Shuhei Kitamura* Katsunori Yamada†

April 27, 2020

Abstract

This document describes the design and an analysis plan for an online RCT experiment (#N = 3,500) to be held between April the 17th, 2020 and May the 3rd, 2020 in Japan under the influence of COVID-19 pandemic.

Contents

*Assistant Professor. Osaka School of International Public Policy, Osaka University, 1-31 Machikaneyama, Toyonaka, Osaka, 560-0043 Japan. Email: kitamura@osipp.osaka-u.ac.jp. Phone: +81 6-6850-5645. Fax: +81 6-6850-5656.

†Associate Professor. Faculty of Economics, Kindai University. 228-3, Shin-Kamikosaka, Higashi-Osaka city, 577-0813 Japan. Email: kyamada@kindai.ac.jp.

1 Introduction

The COVID-19 pandemic has threatened many human lives and economic activities in the world. Governments worldwide took several measures to prevent the spread of coronavirus including law-, military-, and/or technology-enforced lock-downs. However, not all the governments immediately took such "hard" measures, but some, such as Sweden and Japan, took "soft" measures, e.g., providing political messages to citizens so that they take self-restraint actions on their own.

Despite the flood of administrative messages asking for self-restraints of staying home, and the emergency declaration by the government on April the 7th, since there are no practical fines against violations in the Japanese legal system, there are people who would not comply. This situation has become a social issue and revealed how difficult it is for a government to guide citizens by providing advisory messages even when the whole society obviously benefits in the long run by following them.

With this circumstance in Japan, this study conducts a Randomized Controlled Trial experiment to investigate causal effects of information provisions to individuals on their self-restraint behaviors with consideration of two novel features in our messages. The first feature we test is the power of the Royals over the political authority. The second one to be examined is the power of personalized messages over commonly shared slogans. We will test if those new features could enhance self-restraint behaviors among Japanese citizens.

2 Experimental Design

2.1 Timeline

The experiment takes place between April the 17th and May the 3rd, 2020. It is a three week panel survey for around 3,500 subjects in Japan and we collect data on hours of outings per day on every Saturday and Sunday, starting on April the 18th. The information of self-reported hours of outings will be our outcome variable. Before these main surveys, there is a baseline survey consisting of 20 questions to collect data on individual characteristics on April the 17th. Finally, Our randomized manipulations occur on April the 23rd (Thursday), right before the main surveys for the second week. We have four categories for our treatments and their details are explained in the treatment section below.

With this setting of the three week panel survey, we will use information of hours of outings from the second week as our outcome variable to investigate effects of our randomized manipulations.[1] The information of outing hours from the first week can be used to give

[1] It is important for our online experiment to make sure that our subjects actually are exposed to assigned treatments before the weekend of the second week. Hence, we close the window of information provision

feedback in randomly assigned information provisions in the second week, and to control the baseline conditions for respective participants. Information from the final week can be used to test lasting effects of the manipulations.[2]

2.2 Sampling

2.2.1 Sampling Frame

The target population for the study is registered individuals aged 20s-50s living in one of the seven prefectures in Japan (Chiba, Fukuoka, Hyogo, Kanagawa, Osaka, Saitama, and Tokyo). The emergency declaration by the government was announced for these seven prefectures on April the 7th, and was expanded to include the rest of prefectures on April the 16th (Figure 1). We limited the sampling place to these seven prefectures in order to control for the baseline conditions.[3] The sample is stratified based on gender and age group and is made proportional to the most recent Basic Resident Register in 2019.

2.2.2 Recruitment

To obtain a representative sample of the Japanese population in terms of gender and age, we hire MyVoice Inc. to administer the experiment. This local survey company has more than 1 million pre-registered participants.

In order to avoid experimenter demand effects, participants are not informed of the purpose of the survey except that it is being conducted for academic research by Osaka University and Kindai University. The anonymity of responses are secured and we guarantee the freedom of quitting the survey at any point.

In order to encourage our subjects to finish all the survey, we employ an inclination allocation of participation reward toward the end of the survey. By finishing all the survey, our participant can earn up to 180 Japanese Yen.

We do not expect attrition to be significant or systematic in our experiment because the experiment involves merely simple questions. In finishing our questionnaire, we make sure that there will be no complications or confusions upon answering the questions and that there will be no experimenter demand effect caused by the questions, either. Still, there is a possibility that systematic attrition happens after being exposed to different types of information. We will check if the attrition occurs in a systematic way.

2.2.3 Statistical Power

A usual procedure for deciding sample size requires standard deviations from pilot data or related studies. However, given the urgency of circumstances and unprecedentedness of the issue, we choose the sample size by minimizing minimum detectable effect (MDE) so that it gets smaller than the effect size of 0.2 standard deviations. Since, as explained below, there

sessions by the end of the 18th (Friday). Those who do not answer the questions in these sessions are dropped from our study. We will check if this attrition occurs in a systematic way.

[2] In addition to the main outcome, we also have the other outcome variable in intentions of self-restraint, which subjective information is provided right after our information manipulations.

[3] We also did not know that emergency declaration by the government would be extended for the rest of prefectures at the time of designing the experiment.

are four arms in the 2-by-2 factorial design experiment, the target sample size is set to be 3,500, implying 875 individuals per arm.

Figure 2 shows an MDE curve for pairwise comparisons when power is 0.8 and the false discovery rate is 0.05. The minimum detectable effect between a pair of groups is 0.1897 standard deviations when the group size is 875.

We do not expect attrition to be significant in our experiment because the experiment involves a simple task in which participants answer a small number of survey questions. People are allowed to leave the experiment early, but this would entail forgoing payments earned from completing the experiment.[4] We will check if the attrition occurs in a systematic way.

In addition, we make survey questions intuitive and do not include unnecessary questions to avoid any complication or confusion upon answering the questions. The baseline survey, which includes the largest number of questions in this experiment, has just 20 questions, and is supposed to be completed within 10 minutes.

2.3 Survey Questions

There are in total eight surveys for the study. All the questions asked in the surveys (originally in Japanese) will be provided as a supplementary material. Both the original versions and the English-translated versions will be provided.

Baseline Survey (April the 17th): the baseline survey includes demographic questions (gender, age, education, marital status, number of people living together, residential place, employment status, self-reported health status, household pre-tax annual income, and income shock) as well as questions to measure time preferences, risk preferences, social preferences, tendency to do herd behaviors, political orientations, policy evaluations, religious beliefs, national identities, and types of mobile/smart phones. The total number of questions is twenty.

Survey on Outing Behavior (April the 18th, 19th, 25th and 26th, May the 2nd and and the 3rd): the surveys on outing behaviors ask two questions: hours of outings on the day, and that of necessary outings such as ones to work place, schools, grocery shops, or hospitals, on the day.

Pre- and Post-Treatment Survey (April the 23rd): As a pre-treatment question, we ask predictions about relative length of outing hours for each participant, as compared to that of other participants. We have four post-treatment questions, which include a comprehension question regarding the information provided, an attention question, a question to check if the information provided was new to the subjects, and finally a question about self-reported intention on how much they would comply with self-restraint behavior thereafter.

[4] To prevent sample attrition, we use the incentive scheme in which the payment to the participants increases towards the end of the experiment.

2.4 Information treatments

2.4.1 Types of treatment

We conduct randomized information manipulations on April the 23rd, exploiting a 2-by-2 factorial design (Table 1). Information provisions are in Japanese and intervention screens will be provided as a supplementary material.

To the horizontal direction in Table 1 we have two conditions regarding the sender of the messages. Specifically, we consider the political authority (the current prime minister) and the current Emperor as the ones who send messages. Since there is already a flood of administrative messages on staying home at the time of the experiment, we consider the condition of the political authority to be our control group. In Japan, the Imperial Family is well-respected and is supposed to be the "symbol" of Japan. As such, even without any political power with him, the Emperor can influence the society with his messages. As such, we will treat this condition as our treatment, and test the symbolic power of the Royals.

In both cases, the senders of messages hope that the society will overcome the COVID-19 pandemic. For the sake of the salience of the experiment, we choose actual statements made by either sender before our experiment set off, rather than we make up fictional messages. Those messages come with drawing images of the senders in order to enhance the reality of the messages.

Next, to the vertical direction in Table 1, we have two conditions regarding personalized messages. When you are in "NO" condition along *Feedback* in Table 1, you will receive some common information for everybody, just as it is happening in the real world. When you are in "Yes" condition, you will receive personalized messages. Those personalized messages contain information of their own hours of outings from the previous week, together with the median value of outing hours for the rest of participants.[5] As such, from the messages it is easy for our participants to grasp where they stand in the distributions of outing hours. The aim of our personalized massages is to make the information in the society more salient to individuals, thereby promoting responsible behaviors on their own.

2.4.2 Random assignments to treatments

We randomly assign our subjects into four groups using STATA 15's `randomize` command. We make sure that the sample distributions regarding important factors such as age, gender, residence place, and so forth, are balanced across the groups. Then, the survey company sends the corresponding messages to respective groups at the time of treatments.

2.5 Others

2.5.1 Instruments

As it is already documented, our experiment is conducted online. All the participants remotely answer survey questions using their own computer, tablet, or smartphone.

[5] Because the weather can be different across prefectures, we set the median values of outing hours at the prefecture level.

2.5.2 Data Collection and Processing

Data are collected by the survey company and sent to us by email every week. Data are kept anonymous and protected by passwords.

Data processing occurs as results come in. We clean, manage, and analyze data. Data are stored on Dropbox shared between the authors.

3 Empirical Strategy

3.1 Balancing Checks

We will check the balance across conditions. The variables that we will check the balance are (a) gender, (b) age group, (c) education, (d) marital status, (e) per capita household income, (f) negative income shock, and (g) prefecture. Any variables that are found to be unbalanced at the false discovery rate of 0.05 will be included in the regression. Since there are four arms, we report the results unadjusted and adjusted for multiple comparisons when we compare means across arms.

We will also check the balance between attritors and non-attritors using the same set of variables and the significance level. If there is any difference, then Lee (2009)'s bounds will be performed.

3.2 Empirical Specification

3.2.1 Main Specification

We examine effects of various types of information provisions on outing behaviors and intentions. From our experimental setting, the empirical specification is straightforward and read as

$$Y_i^k = \beta_0^k + \beta_1^k \text{Emperor}_i + \beta_2^k \text{Feedback}_i + \beta_3^k \text{Emperor}_i \times \text{Feedback}_i + X_i' \Gamma^k + \phi_p + \varepsilon_i^k, \qquad (1)$$

where i denotes individuals, Y_i^k is outcome k, Emperor_i is the indicator for the Emperor's message treatment, Feedback_i is the indicator for the feedback treatment, X_i is a vector of variables that are unbalanced between treatment and control groups, and the pre-treatment outing behavior, ϕ_p is the prefecture fixed effects, and ε_i^k is the error term. Given the nature of our treatment, i.e., individuals are separately treated online, we use robust standard errors to account for heteroskedasticity but not inter-class correlation.

We use the following outcome variables ($k \in \{1,2,3\}$) in the main analyses:

1. hours of outings
2. hours of outings for unnecessary purposes
3. intentions of outings

Regarding 1 and 2, we will sum up hours of outings over the weekend. The data values from the second week will be used to examine immediate effects of manipulations while those from the third week will be used to test lasting effects of manipulations.

We hypothesize that individuals who receive personalized feedback (that is, for those with Feedback_i takes the value of 1) are more likely to stay at home after the treatment due to salience of the message, and that those who receive a message from the Emperor (that is, Emperor_i takes the value of 1) are less likely to go out after treatment. Therefore, in the model (1), we are particularly interested in the estimates of β_1^k, β_2^k, and β_3^k.

We run the model (1) using the ordinary least squares regression. For the third outcome, we also the run the models using the ordered logistic regression.

3.2.2 Heterogeneous effects of information treatments

Subjects may react to our personalized messages in different ways, depending on their traits. There are varieties of driving forces for the heterogeneity. We will consider effects through; (1) time preference, (2) risk preference, (3) social preference, (4) herd behavior, (5) political orientation, (6) policy evaluation, (7) religious belief, (8) national identity, and (9) if they are above/below the social median values in the distributions of outing hours.

Those potential heterogeneous effects of manipulations are detected by the following model:

$$\begin{aligned} Y_i^k &= \varphi_0^k + \varphi_1^k \text{Emperor}_i + \varphi_2^k \text{Feedback}_i + \varphi_3^k \text{Emperor}_i \times \text{Feedback}_i \\ &+ \varphi_4^k Z_i + \varphi_5^k \text{Emperor}_i \times Z_i + \varphi_6^k \text{Feedback}_i \times Z_i + \varphi_7^k \text{Emperor}_i \times \text{Feedback}_i \times Z_i \\ &+ X_i' \Gamma^k + \varepsilon_i^k. \end{aligned} \tag{2}$$

where Z_i captures one of the variables above mentioned. The definitions of the other variables are the same as those in the model (1).

Table 2 shows the list of the expected direction of the effect (if anything) for these variables. We expect that patient individuals would care more about the future, and therefore avoid outings to get infected. By contrast, risky individuals would tend to go out by taking the risk of being infected. Socially-minded individuals would avoid going out to further spread the coronavirus. Individuals who are worried about others' opinions would be less likely to go out. The right-wing individuals, individuals who are skeptical about the current government policies, or individuals who have stronger religious belief or national identity would be more likely to stay home if they receive a message from the Emperor. Finally, we expect that effect is larger (smaller) for those who are above (below) the social median. It is also interesting that we may detect a non-monotonic effect for those who receive messages that they stay longer at home. After knowing that, they would no longer stay home as in the previous week, which will be regarded as "moral licence effect". We are interested in the estimates of γ_5^k, γ_6^k, and γ_7^k.

We run the model (2) using the ordinary least squares regression. For the third outcome, we also the run the models using the ordered logistic regression.

References

Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *Review of Economic Studies 76*(3), 1071–1102.

Tables and Figures

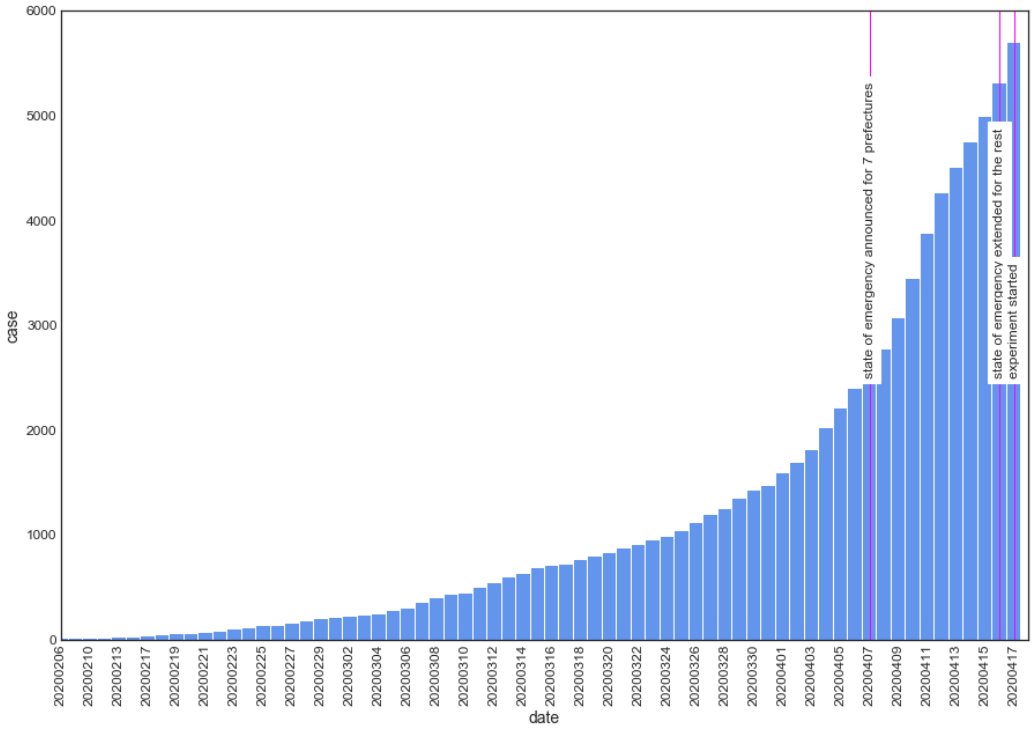
Table 1: Experiment Arms

		Message	
		Political Leader	**Emperor**
Feedback	**No**	Group A	Group C
	Yes	Group B	Group D

Table 2: Expected Direction of the Heterogeneous Effect

Variable	**Expected Direction**
Time preference	Effect is larger for more patient individuals
Risk preference	Effect is smaller for more risky individuals
Social preference	Effect is larger for more socially-motivated individuals
Herd behavior	Effect is larger for those who are more worried about others' opinions
Political orientation	Effect is larger for more rightist individuals
Policy evaluation	Effect is larger for those who are more skeptical about current policies
Religious belief	Effect is larger for those who have stronger religious views
National identity	Effect is larger for those who have stronger national identity
Position effect	Effect is larger (smaller) for those who are above median (below median)

Figure 1: Spread of Coronavirus Infections in Japan



Note. Data are taken from the Ministry of Health, Labour and Welfare.

Figure 2: MDE Curve