

# Description of Data and Codes

Hiroki Kato

## 1 Folder Structure

Figures and tables and the manuscript were prepared in Rmarkdown. All codes to output figures and tables are in `Main Document-LaTeX.rmd` (main manuscript) and `Supplementary Material.rmd` (supplementary material). The analysis and other internal codes are in the `R` folder. The data used for the analysis are in the `data` folder. You can try all the codes in `Main Document-LaTeX.rmd` by placing them in the folder structure as shown below.

```
C:.\n|  Main Document-LaTeX.rmd\n|  Supplementary Material.rmd\n|\n|-data\n|    niid_prevalence.csv\n|    nudge_descript.csv\n|    online_survey.csv\n|    vars_descript.csv\n|\n|-R\n    misc.r\n    R6_EstimateEffect.r\n    R6_Mechanism.r\n    R6_MonetaryValue.r\n    R6_Regression.r\n    R6_StartAnalysis.r
```

## 2 Data

- `niid_prevalence.csv`: Rubella antibody prevalence data provided by 2018 National Epidemiological Surveillance of Vaccine-Preventable Diseases, NIID. We used to create Figure 1 in the manuscript. See below for definitions of each variable.
- `nudge_descript.csv`: List of message contents for each experimental arm (Japanese and English). We used to create Table 1 in the manuscript.

- `online_survey.csv`: Nationwide online survey data. See below for definitions of each variable.
- `vars_descript.csv`: List of definitions of variables used in analysis (Japanese and English). We used to create Table D1 in the Supplementary Material D.

## 2.1 Variable definitions: `niid_prevalence.csv`

- `age`: age
- `gender`: gender
- `total`: sample size
- `HI_less8`: Number of people with rubella HI antibody titer less than 8
- `HI8`: Number of people with rubella HI antibody titer of 8
- `HI16`: Number of people with rubella HI antibody titer of 16
- `HI32`: Number of people with rubella HI antibody titer of 32
- `HI64`: Number of people with rubella HI antibody titer of 64
- `HI128`: Number of people with rubella HI antibody titer of 128
- `HI256`: Number of people with rubella HI antibody titer of 256
- `HI512`: Number of people with rubella HI antibody titer of 512
- `HI_more1024`: Number of people with rubella HI antibody titer more than 1024
- `prevalence`: Antibody prevalence (%) calculated by  $100 * (HI8 + HI16 + HI32 + HI64 + HI128 + HI256 + HI512 + HI\_more1024) / total$

## 2.2 Variable definitions: `online_survey.csv`

- `id`: ID of respondent
- `nudge`: Assignment of text message reminders. A = MHLW (Control); B = MHLW (Age); C = Altruistic; D = Selfish; E = Social Comparison; F = Deadline; G = Convenient
- `test_int`: (Wave 1) Dummy variable of intention to undertake antibody testing. See the manuscript for how to create it.
- `vaccine_int`: (Wave 1) Dummy variable of intention to be vaccinated. See the manuscript for how to create it.
- `age`: (Wave 1) Age calculated by birth year and birth month.
- `coupont2019`: (Wave 1) Dummy variable of men aged 40–46 who automatically received the free voucher in FY2019.
- `know_coupon`: (Wave 1) Dummy variable of awareness of additional free routine immunization
- `exp_antibody`: (Wave 1) Dummy variable indicating previous rubella antibody testing
- `exp_vaccine`: (Wave 1) Dummy variable indicating previous rubella vaccination
- `education`: (Wave1) Years of education.
- `married`: (Wave1) Dummy variable taking one if a respondent is married.
- `exercise_w1`: (Wave1) Dummy variable taking one if a respondent exercises or plays sports more than once a week.

- **health\_check**: (Wave1) Dummy variable taking one if a respondent has had a medical examination in their city or place of employment in the past year from the time of Wave 1.
- **flushot**: (Wave1) Dummy variable taking one if a respondent is vaccinated against influenza every year.
- **handicap**: (Wave1) Dummy variable taking one if a respondent believes that if a woman in early pregnancy is infected with rubella, her child may be born with a disability.
- **generosity**: (Wave1) Five-point Likert scale for the question “I am happy when I pick up trash in the park.”
- **norm\_waiting\_line**: (Wave1) Five-point Likert scale for the question “I never interrupt someone in line.”
- **selfish\_anonymous**: (Wave1) Five-point Likert scale for the question “If I can never find it, I can do bad things (littering, parking tickets, etc.).”
- **income**: (Wave1) Household income. For those who did not respond with household income, the overall average was used.
- **noinfo\_income**: (Wave1) Dummy variable taking one if a respondent did not answer household income.
- **wtp\_vaccine**: (Wave 1) WTP for the rubella vaccination assuming the vaccination costs of JPY 5,000.
- **act\_test**: (Wave 2) Response of the antibody testing question presented in Supplementary Material B. 1 = (a); 2 = (c); 3 = (b).
- **act\_vaccine**: (Wave 2) Response of the vaccination question presented in Supplementary Material B. 1 = (a); 2 = (e); 3 = (c) or (d); 4 = (b).
- **handwash**: (Wave2) Five-point Likert scale for the question “I washed my hands and gargled frequently from the end of the previous survey to today.”
- **temp\_check**: (Wave2) Five-point Likert scale for the question “I took my temperature frequently from the end of the previous survey to today.”
- **avoid\_out**: (Wave2) Five-point Likert scale for the question “I refrained from going out from the end of the previous survey to today.”
- **avoid\_crowd**: (Wave2) Five-point Likert scale for the question “I avoided crowded places when I went out from the end of the previous survey to today.”
- **wear\_mask**: (Wave2) Five-point Likert scale for the question “I always wore a medical mask when I went out or met people from the end of the previous survey to today.”
- **aw1\_test**: (Wave 2) Dummy variable taking one if **act\_test** = 1 (Respondents had undertaken antibody testing between Wave 1 and Wave 2).
- **aw1\_testhave**: (Wave 2) Dummy variable taking one if **act\_test** = 1 and **act\_vaccine** = 4 (Respondents had undertaken antibody testing and confirmed a positive test between Wave 1 and Wave 2).
- **abw1\_test**: (Wave 2) Dummy variable taking one if **act\_test** = 1 or **act\_test** = 2 (Respondents had undertaken antibody testing before Wave 2).
- **aw1\_testvaccine**: (Wave 2) Dummy variable taking one if **act\_test** = 1 and **act\_vaccine** = 1 (Respondents had undertaken antibody testing and been vaccinated Wave 1 and Wave 2).
- **abw1\_testvaccine**: (Wave 2) Dummy variable taking one if **act\_test** = 1 or 2 and

`act_vaccine = 1 or 2` (Respondents had undertaken antibody testing and been vaccination before Wave 2).

- `follow`: Dummy variable taking one if a respondent participated in Wave 2.

### 3 Codes

We use object-oriented programming with the `R6` package. All R files in the `R` folder define classes (`R6` classes). `misc.r` defines custom functions such as graph templates. Analytical code such as regression analysis is defined as methods of objects. To reproduce, you need to read all the source files in the `R` folder, using following code.

```
library(here)
source(here("R/R6_StartAnalysis.r"))
```