Data Task Report

Zerui Wu

1 Data Simulation and Merge

Answer: (**Simulation**) Here is the overview of the variables in simulated datasets and how they are produced:

- **Baseline Survey Data:** The dataset has 5,000 observations and 6 variables:
  - *id*: Identifier. The IDs are unique and range from 1 to 5000 based on the row order.
  - *vac_takeup0*: COVID-19 vaccination indicator **at baseline**, where 1 indicates vaccinated and 0 otherwise.
    - ∗ It is assumed to follow a Bernoulli distribution with a vaccine take-up probability of 0.3 for all individuals.
    - ∗ So, random values drawn from the uniform distribution in the interval $(0, 0.3)$ are coded as 1, and those in $(0.3, 1)$ as 0.
  - *age*: Baseline age. It is randomly drawn from a uniform distribution ranging from 18 to 80.
  - *female*: Female indicator. 1 is female, and 0 is male.
    - ∗ It follows a Bernoulli distribution such that the $Pr(female) = 0.5$. Then, random values drawn from the uniform distribution in the interval $(0.5, 1)$ are coded as 1, and those in $(0, 0.5)$ as 0.
  - *edu*: Baseline educational attainment.
    - ∗ It is randomly generated from a uniform distribution, with values falling in the intervals $(0, 0.1)$ assigned as Less than High School, $(0.1, 0.4)$ as High School, $(0.4, 0.6)$ as Some College, and $(0.6, 1)$ as College or higher.
    - ∗ This distribution approximates the educational attainment structure in the U.S.
  - *college*: Indicator for college or higher education. It equals 1 if *edu* = College or higher, and 0 otherwise.

- **Baseline Survey Data:** The dataset has 5000 observations and 2 variables:
  - *id*: Identifier. It is generated in the same way as in the Baseline Survey Data.
  - *treatment*: Treatment assignment variable.
    - ∗ It is generated based on the row order. The first third of observations are assigned to the Reason campaign, the second third to the Emotions campaign, and the remaining third to the Control group.
    - ∗ As the other two datasets are randomized, treatment status will be randomized in the final merged data.

- **Endline Survey Data:** The dataset has 4,500 [1] observations and 2 variables:

[1] The initial data has 5,000 observations. After generating all variables, a sample of 4,500 individuals is randomly selected to simulate the attrition.

- *id*: Identifier. It is generated in the same way as in the Baseline Survey Data.
- *vac_takeup1*: COVID-19 vaccination indicator **at endline** such that 1 is vaccinated, and 0 otherwises. It is created as follows:
  - ∗ *vac_takeup0* and *treatment* are temporary variables generated using the same procedures as described above. They serve as baseline conditions for constructing the endline vaccination indicator such that:
    - · If someone was already vaccinated at baseline, we assume they stay vaccinated at endline.
    - · For those not vaccinated at baseline, the probability of vaccination at endline is determined using a uniform random draw. Based on their treatment group, we assume that each person is assigned a chance of being vaccinated:

      (1). Those who are in the Reason campaign have a 55% chance of getting vaccinated.

      (2). Those who see the Emotions Ad have a 70% chance of vaccination.

      (3). Those in the Control group have a 30% chance, which matches the baseline rate.
    - · Observation that doesn't meet any of the conditions above is marked as not vaccinated.

(**Merge**) The final merged data has 4,500 observations and 8 variables. The data is well-balanced in terms of the baseline variables [2]. Below is an overview of how the three raw datasets are merged to create this final clean data:

- First merge: The baseline survey data, as the master dataset, is merged with the randomization data using a one-to-one merge on the unique *id* variable. All 5,000 rows were successfully matched one-to-one across both datasets.
- Second merge: Building on the first-stage merged dataset, we further merge it with the endline survey data using the same *id* variable as the key. A total of 4,500 observations are successfully matched one-to-one, and the remaining 500 unmatched rows are dropped to account for sample attrition.

2 Evaluate the Effectiveness of Campaigns

Answer: Both the Emotions and Reason campaigns are effective and significantly increase the **vaccination rate** and the **likelihood of COVID-19 vaccine uptake** for the sampled individuals, with the Emotions treatment having a stronger impact.

- **Vaccination Rate:** From Figure 1, both treatment arms show substantially **higher** vaccination rates compared to the control group after the intervention.

[2] The balance check is conducted in *Stata* in lines 204 to 213. No statistically significant differences can be found across baseline vaccination status, age, gender (female), and college education at the 10% level.

  - After the intervention, the Emotions treatment leads to the **largest** gain. At baseline, vaccination rates were similar across all three groups. By endline, **these rates diverged significantly**, rising to **0.675** in the Reason group by 139.36%, **0.794** in the Emotions group by 172.85%, and **0.516** in the Control group by 65.38%.

  - At endline, the observed differences in vaccination rates across the three groups are highly **statistically significant** and are **unlikely** to be due to chance. Using a χ^2-test conducted in $Stata$, we have $p < 0.01$, and this **rejects** the null hypothesis that the proportions of vaccinated individuals are equal for all three groups at endline even at the 1% significance level. This result provides strong evidence that **both** treatments had a positive effect on vaccination rate, compared to the control group.

- **Probability of COVID-19 Vaccine Take-Up**: Both treatment campaigns have a statistically significant and positive effect on the likelihood of vaccination, with the Emotions intervention having a stronger mariginal impact than the Reason campaign.

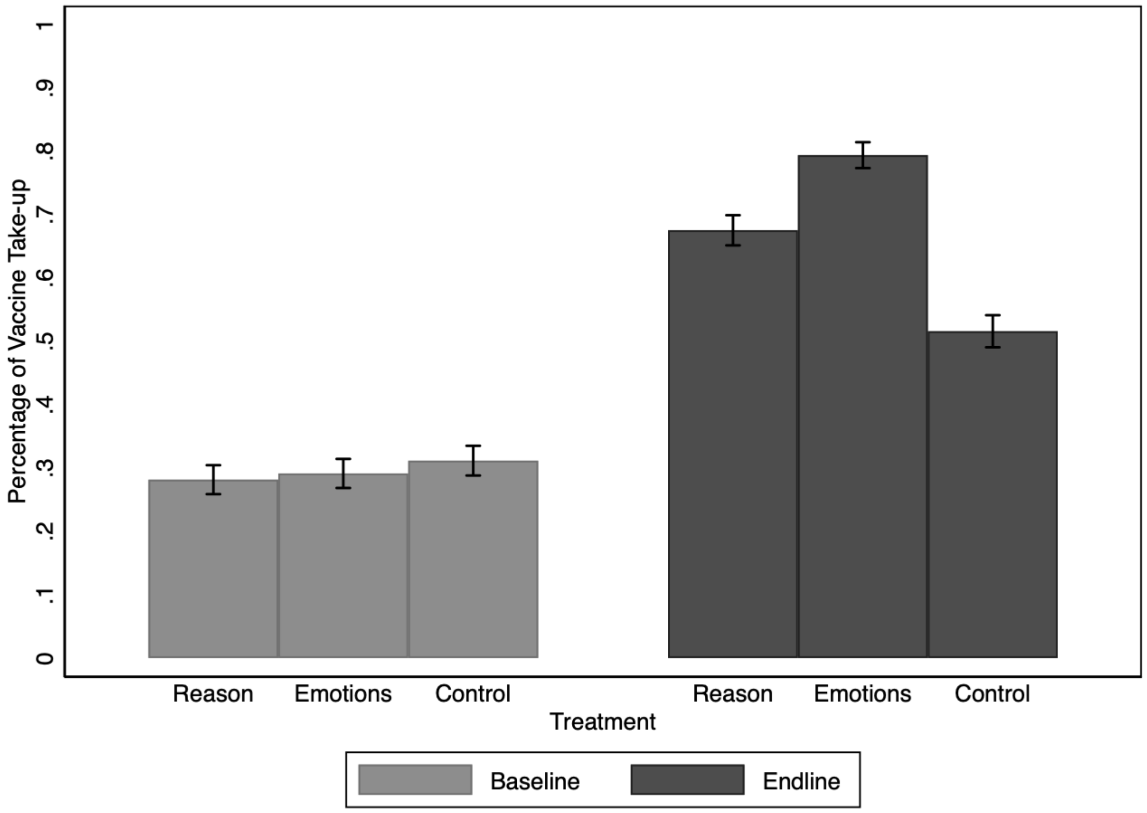
  - From Table 1, both coefficients on Reason Ad and Emotions Ad are **positive** and **statistically significant** at the 1% level. Controlling for variables such as college attainment, gender, age, and the baseline vaccination status, we regress the probability of being vaccinated at endline on the treatment assignments.

    - The estimated average treatment effect for the Reason Ad is approximately 0.178, indicating that on average individuals exposed to the Reason campaign are **17.8%** more likely to get vaccinated than those in the control group, holding other variables constant.
    - **The treatment effect for Emotions campaign is much stronger:** individuals in the Emotions campaign are expected to be **28.2%** more likely to get vaccinated compared to the control group, controlling for the same set of variables.

  - Figure 2 illustrates that the Emotions campaign is **more effective** than the Reason intervention in increasing vaccine take-up in our sample. This coefficient plot displays the comparative treatment effects of the Emotions and Reason campaigns. The point estimate and the 95% confidence interval for the Emotions Ad are visibly higher than those for the Reason Ad.

3 Appendix

Figure 1: Vaccine Take-Up Rates across Baseline and Endline



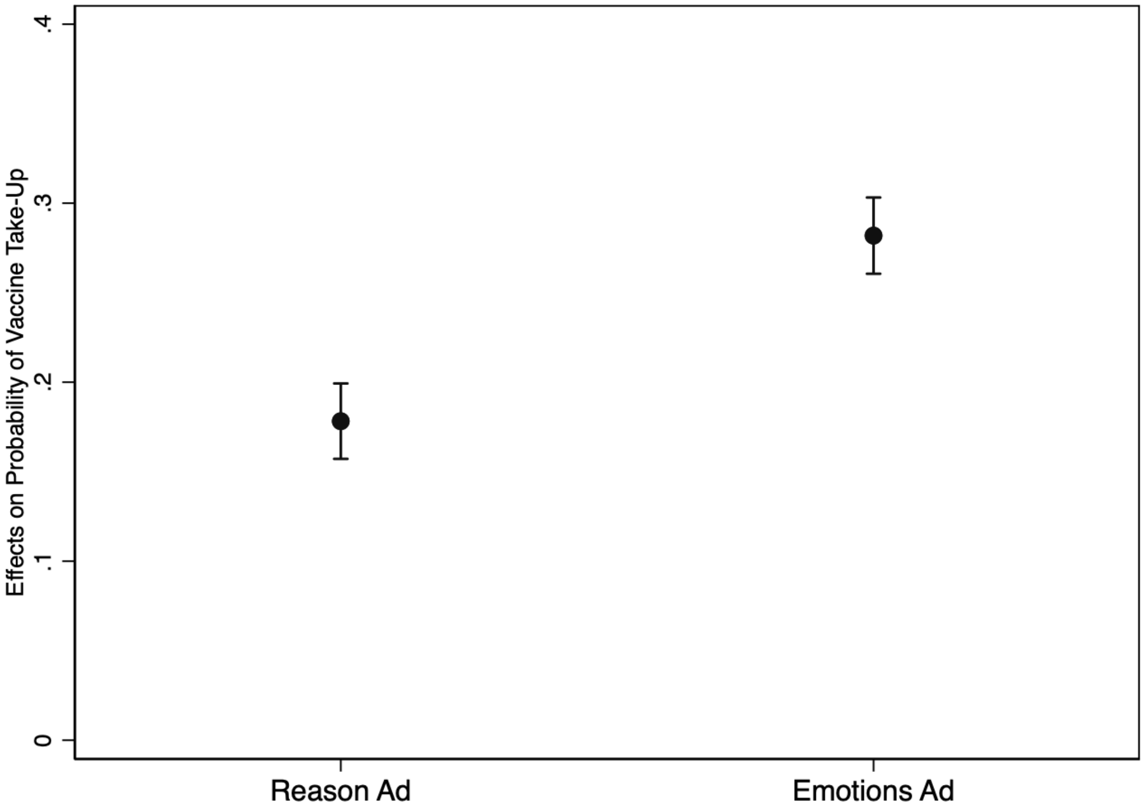
Note: This figure shows vaccination rates at baseline and endline across the Reason, Emotions, and Control groups. At baseline, vaccination rate is around 0.282 in the Reason group, 0.291 in the Emotions group, and 0.312 in the Control group. χ^2-test for the baseline proportion yields $p = 0.186 > 0.10$, and it suggests that we don't have enough evidence to reject that the vaccination rates at baseline are equal across these three groups at the 10% significance level. However, χ^2-test for the endline proportion yields $p < 0.001 < 0.01$, and this rejects the null hypothesis and concludes that there's a statistically significant difference after intervention across all three groups at the 1% significance level.

Table 1: Effects of Reason and Emotions Advertisement on Vaccination

	(1) Vaccine Takeup (Endline)
Reason Ad	0.1782***
	(0.0107)
Emotions Ad	0.2819***
	(0.0109)
College+	-0.0121
	(0.0089)
Female	-0.0040
	(0.0089)
Age	-0.0154***
	(0.0002)
Vaccine Takeup (Baseline)	0.4890***
	(0.0096)
Constant	1.1258***
	(0.0148)
R^2	0.61
Observations	4500

Note: This table reports the estimated average treatment effects (ATE) of the Reason and Emotions advertisements on COVID-19 vaccine uptake. The analysis uses a sample of 4,500 individuals. The linear probability model (LPM) regresses a binary indicator for endline vaccination status (1 = vaccinated, 0 = not vaccinated) on treatment assignment, controlling for baseline age, education, gender, and vaccination status. The reference group is the control arm. The model explains 61% of the variation in the dependent variable (R^2 = 0.61). Statistical significance is denoted as follows: p < 0.1 (*), p < 0.05 (**), p < 0.01 (***). Treatment effects are statistically significant at the 1% significance level for both Reason and Emotions campaigns.

Figure 2: Comparing Treatment Effects: Reason v.s. Emotions Campaigns



Note: This figure visualizes the point estimates of treatment effects for both Reason and Emotions campaigns, respectively, with 95% confidence intervals.