Impact of COVID Misinformation on Vaccination Intent - A Reanalysis Identifying and Addressing Covariate Imbalances

Zhongren Chen
Department of Statistics
Stanford University
Stanford, CA 94305
zhongren.chen@stanford.edu

Matheus Dias
Department of Statistics
Stanford University
Stanford, CA 94305
mattdias@stanford.edu

Quentin Hsu
Department of Statistics
Stanford University
Stanford, CA 94305
qhsu@stanford.edu

Abstract

This paper reanalyzes a randomized controlled trial which exposed participants to online misinformation and measured its impact in vaccination intent. We evaluate their randomization and show that it is significantly imbalanced (p-value ¡ 0.0001) using a Monte Carlo simulation of the Mahalanobis distance between Treatment and Control. Moreover, we reduce the bias of the estimates by applying matching estimators and performing regression adjustment. Our findings reveal that the original study overestimated the impact of the treatment and that the adjusted treatment effect is a 7.9% decrease in vaccine intent after exposure to misinformation. We also explored heterogeneous treatment effects and provide some intuitive insights.

1 Introduction

We perform a reanalysis of a study performed in 2020 about the effect of COVID misinformation and the intent to get the COVID vaccine [6]. The study asked participants many pre-treatment questions that served as interesting covariates to explore. We explored the imbalance of these covariates and demographics in the treatment/control groups assigned in the study. We found significant imbalance of the covariates in the treatment/control groups, bringing question into the exact randomization procedure used in the original study. We adjusted for the imbalances to report more externally valid treatment effects. Finally, we looked into heterogenous treatment effects using personas created from these additional covariates.

All of our code can be found in this google colab.

2 Related Work

In late 2020, a study was ran that tested whether exposure to misinformation about COVID vaccines would reduce the intent to be vaccinated. This study recruited 4000 participants in each of US and UK, randomized the participants to form 3000 people in treatment and 1000 people in control. All participants were asked several questions prior to treatment, including whether they intended to take

the COVID vaccine if it was available. Participants were then exposed to 5 images that were either misinformation about COVID vaccines if they were in treatment or facts about COVID if they were in control. The subjects were then asked several post-treatment questions including whether they intended to take the COVID vaccine.

The original study found that exposure to vaccines caused a statistically significant reduction in the intent to receive the COVID vaccine.

There are also other studies that examine the effect of misinformation related to the pandemic. For instance, effect of misinformation on public trust of vaccine [8], effect of information channel on vaccine acceptance [9] and relationship between vaccine and prevalence conspiracy theories [10]. However, there are very limited studies that are focus directly on the effect of misinformation on the vaccine acceptance other than [6]. Hence a careful re-analysis is needed to derive a more accurate treatment effect that accounts for the imbalances of covariates.

3 Data

3.1 Data Sourcing

The original study's github page provided us with a CSV and other supplementary information for the experiment data. However, we noticed that the published CSV was a small subset of all the questions that were asked in the questionnaire. Many of the additional questions were important for our analysis as they would serve as pre-treatment covariates.

We dug through the version history of the github and were able to recover the original SPSS save file with all of the data. We leveraged some of the provided python code to extract the full dataset to use for our re-analysis. We include the full dataset as well as the code used to generate the CSV in our code samples.

3.2 Defining an Outcome

The study asked a 4 level question (yes, maybe yes, maybe no, no) about vaccine intent before and after the treatment. This is an ordinal outcome that is likely not linear (eg. going from yes to maybe yes is easier than going from maybe yes to maybe no). Many of the methods we learned in class do not directly support ordinal outcomes, so we needed to define a potential outcome that is usable for analysis.

We define the outcome of study as whether there is a strict decrease in vaccine intent after exposure to misinformation. This allows us to answer the same question as what the original study did, but lose some power by converting to a binary outcome. Also, we had to remove around 15% of the data since people who said they were not going to take the vaccine before the treatment are not able to decrease their vaccine intent. This 15% was distributed evenly between treatment and control, so there was no imbalance from removing this data.

3.3 Creating Personas from Covariates

The study asked several pre-treatment questions that could help us form personas of people to explore for imbalance/heterogenous treatment effects. We created the following personas as additional pre-treatment covariates to analyze along with demographics.

- know.anyone.covid - subject personally knows someone (eg themselves, family, friends) who has caught COVID
- Shielding - subject is shielding because they are in a vulnerable group for COVID
- bad_covid_knowledge - subject disagreed with commonly known facts based on several questions asked about COVID
- bad_covid_vax_trust - subject does not trust the safety/efficacy/importance of COVID vaccines based on several questions asked
- bad_general_vax_trust - subject doest not trust the safety/efficacy/important of vaccines in general basd on several questions asked

- covid_disrupt - subject's health/financial stability was negatively impacted by COVID

For heterogenous treatment effect analysis, we also looked at the individual questions that were asked to form some of the personas above. We collapsed the questions from a 5 level likert scale to a 3 level scale so that each stratum had more samples.

4 Methods

We explore 3 main questions for our project:

1. **Imbalance** - Is there imbalance in different pre-treatment covariates in the dataset?
2. **Adjustment** - How do we adjust for the imbalance in pre-treatment covariates and what do the results look like afterwards?
3. **Heterogenous Treatment Effects** - Using those pre-treatment covariates, can we identify heterogenous treatment effects that provide useful intuitive insights?

4.1 Imbalance

4.1.1 Overall Mahalanobis Distance

We tested the original experiment's randomization by looking for imbalance in the personas and demographics we created using pre-treatment covariates. To get a formal sense of whether the original randomization was imbalanced, we first calculated the mahalanobis distance between the treatment and control groups based on our established personas/demographics. We then ran monte carlo simulations to permute the treatment assignment 10,000 times and recalculated the mahalanobis distance for each permutation. This gave us a distribution of distances, which we could then use to calculate the probability of obtaining the original randomization's mahalanobis distance due to chance.

4.1.2 Imbalanced Covariates

The previous test only lets us identify whether the randomization was imbalanced or not and does not tell us specifically which covariates were imbalanced. We search for the imbalanced covariates by running individual t-tests for each covariate by themselves to see how likely it is to have the proportion of that covariate in treatment and control due to chance.

Since we are running multiple t-tests, we do face the issue of multiple testing. We adjust for this using Bonferroni's correction [7] and change our significant threshold from 0.05 to 0.05/14 = 0.00357 since we have 14 covariates.

4.1.3 Impact of Imbalanced Covariates

The pre-treatment intent to take the vaccine has a big impact on whether the user would have a decrease in intent to take the vaccine. For example, a user that did not want to take the vaccine before the treatment likely won't change their behavior after exposure.

One way to measure how impactful a covariate would be is to look at the distribution of pre-treatment vaccine intent amongst the strata of the covariates. If there is a large difference in vaccine intent per strata, imbalance of that covariate could also significantly impact the treatment effect we are trying to estimate (decrease in vaccine intent).

One example of this is distrust in vaccines. Distrust in vaccines is correlated with the same user not wanting to take the vaccine prior to the treatment. If we had a lot more people that distrust vaccines in the treatment group, our treatment effect could be underestimated since we have too many users that started with not wanting to take the vaccine.

4.2 Adjustment

We explore 2 methods of adjustment for bias.

4.2.1 Regression Adjustment - Lin's Estimator

We perform regression adjustment using Lin's estimator [5], which is a technique used in regression adjustment to minimize variance. The estimator is given by the coefficient of treatment in the model: $lm(Y \sim Z+W+Z*W)$ where Y stands for outcome, W for centered covariates and Z for treatment.

4.2.2 Matching

We use a matching procedure with robust mahalanobis distance (smahal). We flipped our matching process to use mahalanobis (smahal) to match each control unit with 2 treatment units since we had more treatment units than control units.

After matching, we evaluate on the subgroups as a whole after the match instead of doing matched pairs difference in means. This is since we care about understanding the treatment effects for the broader population and not of the individual units we happen to match (external validity). Pairs also wouldn't make sense since these are fundamentally different people that were formed to balance the covariate distribution of the group. This also assumes that our control is representative of the general population.

4.2.3 Comparing Bias Adjustment Methods with Placebo Test

To evaluate if the adjustment method is effective and which one performs better, we created a placebo test as a benchmark. We generated a placebo outcome by fitting a logistic regression model on the covariates we were adjusting for to predict our original outcome (decrease in vaccine intent). This generates an outcome that is completely independent of the treatment effect but still related to our covariates. The true average treatment effect of this placebo outcome is 0.

We then run a basic Fisher Randomization Test [1] (using a t-test) to calculate the average treatment effect of the placebo outcome. This gives us a baseline for how biased the dataset is based on the imbalanced covariates.

We run each of our adjustment procedures again to calculate the average treatment effect of the placebo outcome. The closer the method is to predicting an average treatment effect of 0 (or having a larger p-value), the better the method was at adjusting for covariates [4].

4.3 Heterogeneous Treatment Effects

From the personas we created using pre-treatment covariates, we evaluate for heterogeneous treatment effects to help gain more intuitive insights into how different types of people would react to the covid misinformation treatment.

5 Experiments

5.1 Imbalance

5.1.1 Measuring Overall Imbalance with Mahalanobis Distance

The original study has a mahalanobis distance of 0.41. We ran a monte carlo simulation on the original dataset to randomly permute the assignment into 3000 treatment and 1000 control. From the simulation of 10,000 permutations, none of the permutations were remotely close to the distance of the original study. This indicates that probability that we obtained a mahalanobis distance as big as the original study is practically 0.

See the results of the monte carlo simulation in Figure 1. Note that the distance of the original study was too far out to be seen on the graph.

5.1.2 Visualizing the Imbalance

We can view the overall imbalance visually by plotting the propensity scores of Treatment and Control (Figure **??**). Notice how our propensity score is highly skewed, when we were expecting it to be the same between Treatment and Control.

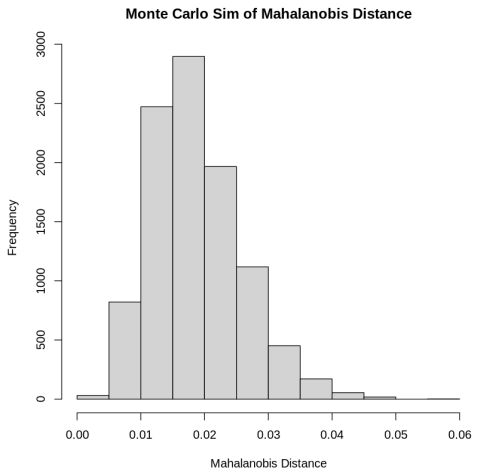

Figure 1: Monte Carlo simulation of Mahalanobis distances

Figure 2: Imbalance Seen in Propensity Score Distributions

We can also look at matching plots to see which covariates are imbalanced between the two groups. See Figure 6. We can see how Age, Employment Ethnicity, Gender, and covid_disrupt seem to have large imbalances. We will formalize the identification of these covariates in the next section 5.1.3.

5.1.3 Identifying Specific Imbalanced Covariates

We ran 14 individual t-tests comparing each individual covariate to treatment assignment ($X \sim Z$). The results of the t-tests can be found in Figure 3. After bonferroni's correction, we see that Age, Employment, Ethnicity, covid_disrupt, and Gender show significant imbalance.

variables	statistic	p.adj	p.adj.signif
<chr>	<dbl>	<dbl>	<chr>
Age	-17.2654778	1.3370e-60	****
Employment	-8.7647881	5.5300e-17	****
Ethnicity	4.9574744	1.1074e-05	****
covid_disrupt	4.4650867	1.1900e-04	***
Gender	-3.7735170	2.3240e-03	**
bad_covid_knowledge	-2.7140540	9.3940e-02	ns
Education	2.0158926	6.1600e-01	ns
Religion	1.9627419	6.9720e-01	ns
bad_covid_vax_trust	-1.8518380	8.9880e-01	ns
bad_general_vax_trust	-1.4297878	1.0000e+00	ns
Income	0.3340919	1.0000e+00	ns
know.anyone.covid	1.5623138	1.0000e+00	ns
Political	1.3468168	1.0000e+00	ns
Shielding	-1.3596203	1.0000e+00	ns

Figure 3: T-Tests for Individual Covariate Imbalance

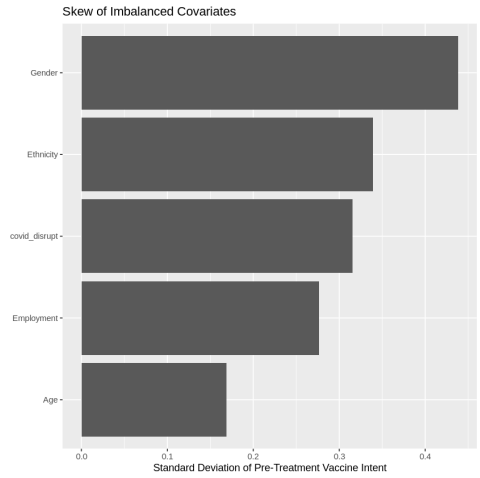

Figure 4: Skew (Standard Deviation) of Imbalanced Covariates

5.1.4 Impact of Imbalanced Covariates

We stratified the data by each individual persona/demographic and calculated the average vaccine intent prior to the treatment. The results of the pre-treatment vaccine intent per stratum for the 5 imbalanced covariates we found in section 5.1.3 can be seen in Figure 5.

To get a sense of how skewed the stratums are, we then calculated the standard deviations of the stratums compared to the average vaccine intent of the full dataset. The results for this can be seen in Figure 4. It seems like Gender, Ethnicity, and Employment have > 0.3, indicating there is > 0.6 spread between the stratums. It is thus, important for those covariates to be adjusted.

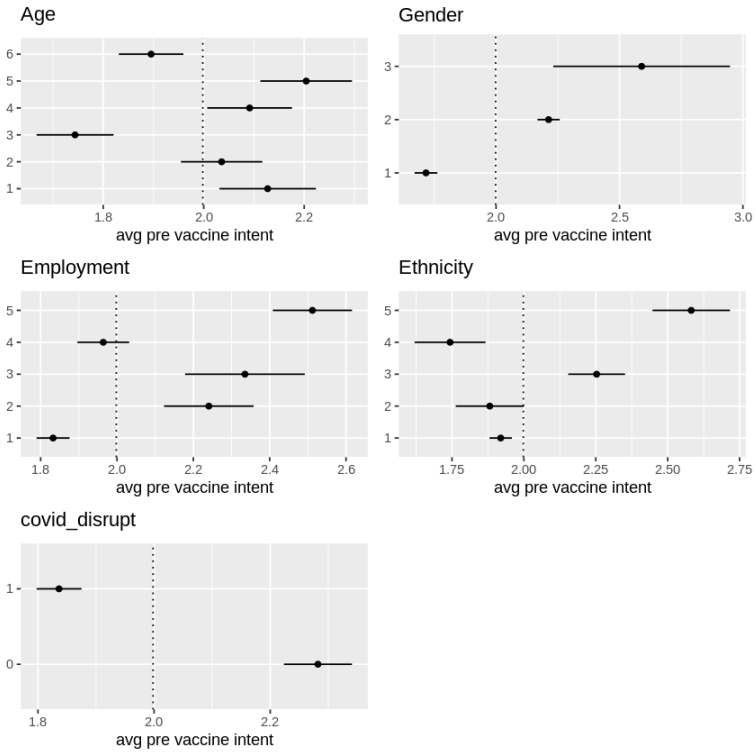

Figure 5: Avg Pre-Treatment Vaccine Intent of Imbalanced Covariates

Note that even though Age is the most imbalanced, it has a relatively small variation in pre-treatment vaccine intent, so adjusting for Age likely does not have a huge impact on our final treatment effect.

Charts for the full set of covariates we looked at can be seen in the Appendix in Figures 8 9 and Figure 10.

5.2 Adjustment

5.2.1 Summary of Adjustment Methods

Table 1 shows the overall treatment effects under the different adjustment methods and outcomes we performed.

Method	Outcome	Estimate	Std. Error	p
Baseline (FRT)	Decrease in Vaccine Intent	0.0866	0.0130	1.92e-11
Lin's Estimator	Decrease in Vaccine Intent	0.0738	0.0327	2.39e-02
Matching	Decrease in Vaccine Intent	0.0794	0.0119	1.53e-11
Baseline (FRT)	Placebo Test	0.0025	0.00138	6.87e-02
Lin's Estimator	Placebo Test	0.0093	0.00565	9.94e-02
Matching	Placebo Test	-0.000498	0.0012	6.73e-01

Table 1: Performance Summary

5.2.2 Regression Adjustment with Lin's Estimator

We first tried Lin's estimator to adjust for covariates. The result can be seen in 1. The estimate is now 13.9, which is smaller than the 14.1 given by the baseline FRT test. Also note that the standard error 9.72 is larger than the 6.00 of FRT.

5.2.3 Matching Adjustment

We used robust mahalanobis distance (smahal) to match each control unit with 2 treatment units. The covariate distribution before and after our matching exercise can be seen in Figure 6.

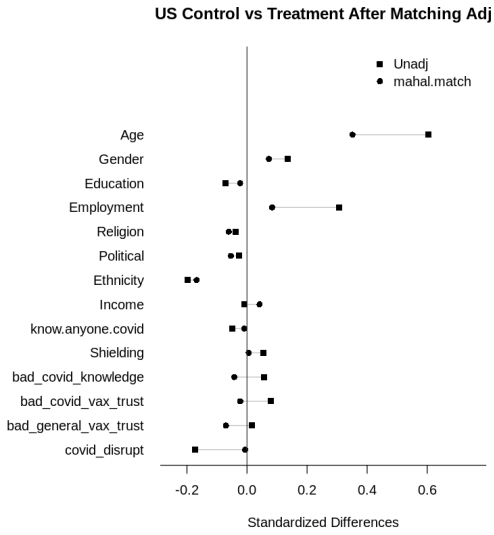

Figure 6: Covariates Before and After Matching Adjustment

Post-matching, we performed a Fisher Randomization Test by running a t-test to calculate the average treatment effect. We obtain a treatment effect of x with standard error of y. This is a smaller effect than the baseline of the original study.

Our covariate balance after matching can be seen in Figure, with an overall mahalanobis distance of 0.14 (which is a 65% increase in similarity). Note that we are throwing out 1000 treatment units, so our matching process trades off balanced covariates with sample size. In the end the average treatment effect calculated under this method is 0.0794392523 with a standard error of 0.011915672, which is slightly smaller than the original study.

5.2.4 Comparing Bias Adjustment Methods with Placebo Test

We generated a placebo outcome and calculated the average treatment effect for the placebo outcome using our baseline Fisher Randomization Test, Lin's Estimator [5], and Matching [11] [3]. The performance of these methods can be seen in table.

We see that the best performing method was matching since it was the closest to 0 and had the largest p-value. Lin's Estimator performed better than the baseline Fisher Randomization Test, so it was adjusting for the covariates.

5.3 Heterogeneous Treatment Effects

We stratified the samples by many covariates and looked at the treatment effect per stratum to identify differences between stratrums.

We'll highlight a few general observations, referencing Figure 7 for heterogenous treatment effects. The full list of plots can be found in the appendix in Figure 11 and 12.

Having good COVID knowledge/trust in vaccines seems to result in a larger decrease in vaccination intent after exposure to COVID misinformation. This seems counter-intuitive at first, but we notice that these 2 personas are directly correlated with a difference in pre-treatment vaccination intent (see Figure 13 in appendix). Those that have bad COVID knowledge/distrust in vaccines also did not intend to take the vaccine prior to the treatment exposure, so there was not much more of a decrease that can be seen from exposure to misinformation.

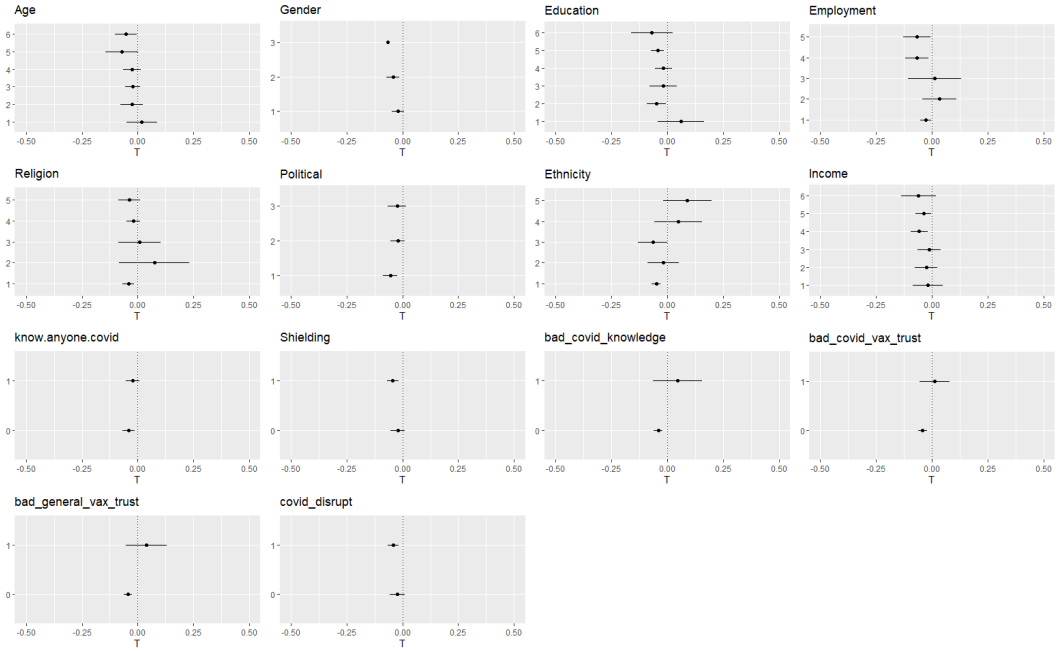

Figure 7: Heterogeneous Treatment Effects

A similar result can be seen by people who mentioned that COVID disrupted their mental/financial well-being. Those that were impacted by COVID also tended to have higher pre-treatment vaccination intent than those that were not impacted and thus, have more room to have decreased vaccination intent after exposure to misinformation.

We also see this for people that personally know someone that had COVID and people that are shielding because they are in a vulnerable group for COVID.

6 Conclusion

Overall, we found that the randomization of the original study in the US is suspicious and that there was a significant imbalance in Age, Employment, Ethnicity, Gender, and people who felt that COVID disrupted their mental/financial well-being. Different Gender, Ethnicity, and Employment tended to have different pre-treatment vaccine intentions, which would affect the final treatment effect.

We adjusted for the imbalances using Lin's Estimator and Matching. We used a placebo outcome to test the effectiveness of each method on the dataset and found that Matching was the better method at adjusting for the covariates for this dataset. Under this method, we found that exposure to COVID misinformation images in this experiment caused 7.9% more participants to have decreased intent to get vaccinated compared to being exposed to COVID facts.

We also found that different people would respond differently to the treatment.

6.1 Future Extensions

For the average treatment effect, one can also use Horvitz–Thompson [2] or Hajek estimator to accounts for imbalance of covariates. Moreover, we can also investigate into the effect of the amount of misinformation on the vaccine acceptance rate. The understanding of difference in effects caused by various types misinformation (e.g. videos, twitters) is also beneficial to policy making.

References

[1] Ronald Fisher. *Design of Experiments*. Oliver Boyd, 1935.

[2] D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1952.

[3] Guido W. Imbens. Matching methods in practice: Three examples. *Journal of Human Resources*, 50(2):373–419, 2015.

[4] Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press, 2015.

[5] Winston Lin. Agnostic notes on regression adjustments to experimental data: Reexamining freedman's critique. *The Annals of Applied Statistics*, 7(1):295–318, 2013.

[6] Sahil Loomba, Alexandre de Figueiredo, Simon J. Piatek, Kristen de Graaf, and Heidi J. Larson. : Measuring the impact of covid-19 vaccine misinformation on vaccination intent in the uk and usa. *Nature Human Behavior*, 5:337–348, 2021.

[7] J. H. McDonald. *Handbook of Biological Statistics*. Sparky House Publishing, 2014.

[8] Sachiko Ozawa and Meghan L Stack. Public trust and vaccine acceptance-international perspectives. *Human Vaccines & Immunotherapeutics*, 9(8):1774–1778, 2013. PMID: 23733039.

[9] Rachael Piltch-Loeb, Elena Savoia, Beth Goldberg, Brian Hughes, Tanner Verhey, Juliette Kayyem, Cynthia Miller-Idriss, and Marcia Testa. Examining the effect of information channel on covid-19 vaccine acceptance. *PLOS ONE*, 16(5):1–14, 05 2021.

[10] Daniel Romer and Kathleen Hall Jamieson. Conspiracy theories as barriers to controlling the spread of covid-19 in the u.s. *Social Science Medicine*, 263:113356, 2020.

[11] Elizabeth A. Stuart. Matching methods for causal inference: A review and a look forward. *Statist. Sci.*, 25(1):1–21, 2010.

7 Appendix

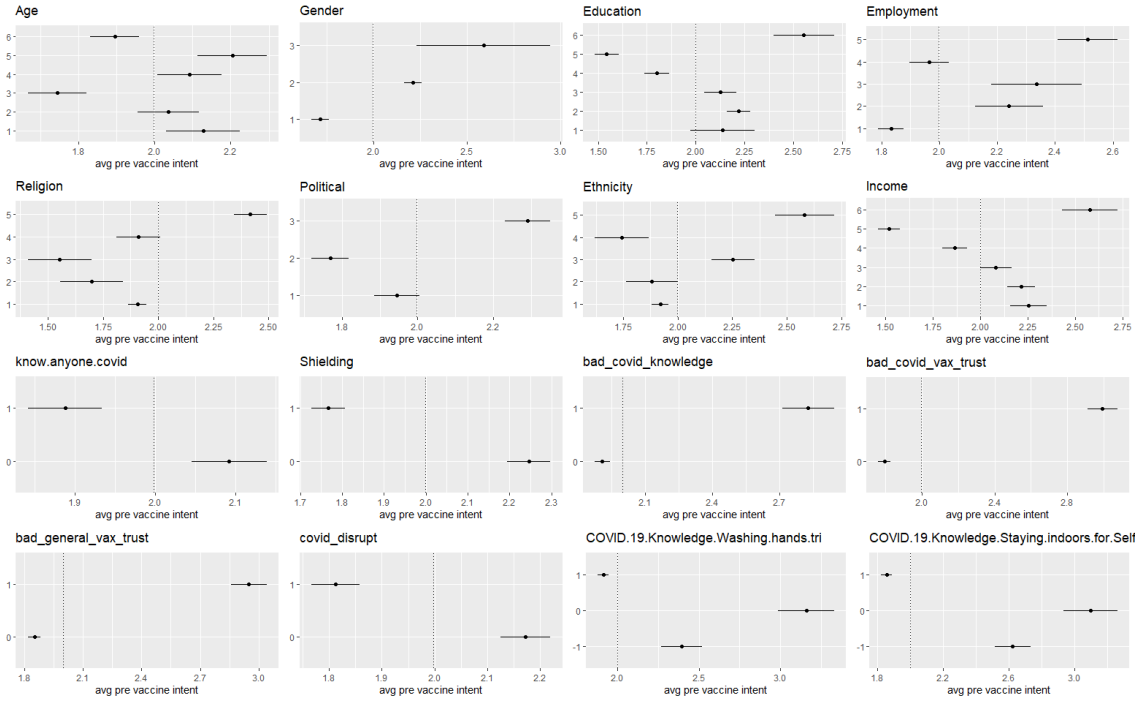

Figure 8: Avg Pre-Treatment Vaccine Intent of All Covariates

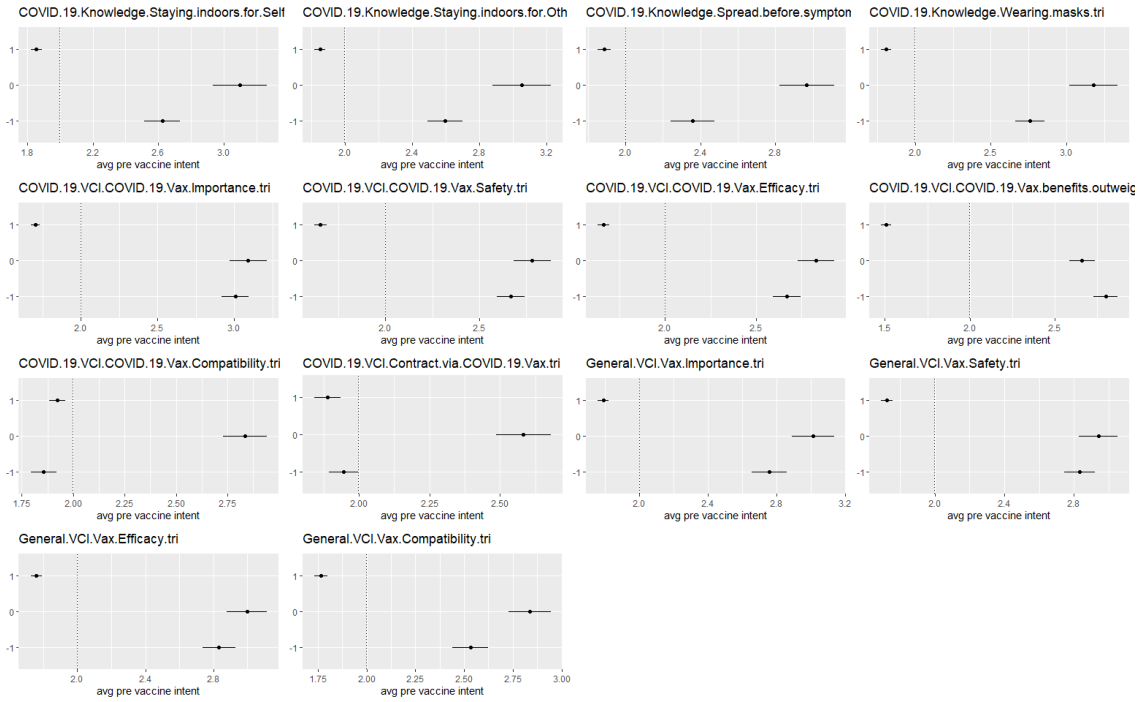

Figure 9: Avg Pre-Treatment Vaccine Intent of All Covariates

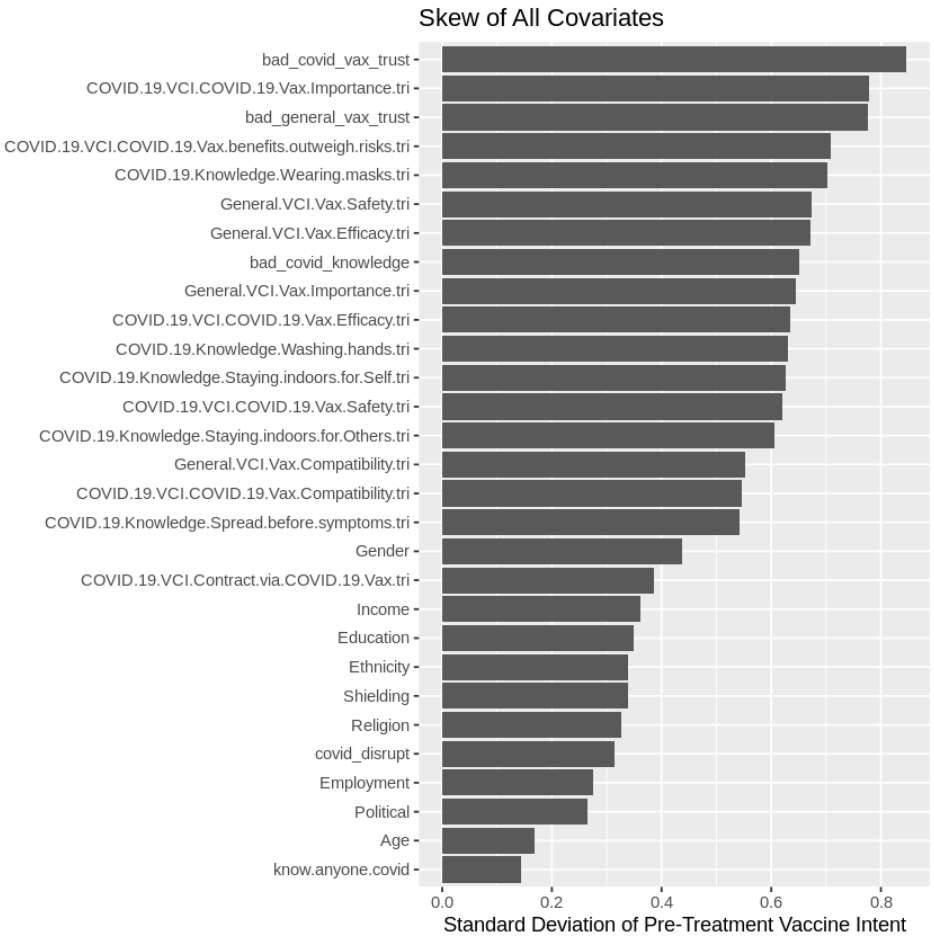

Figure 10: Skew (Standard Deviation) of All Covariates

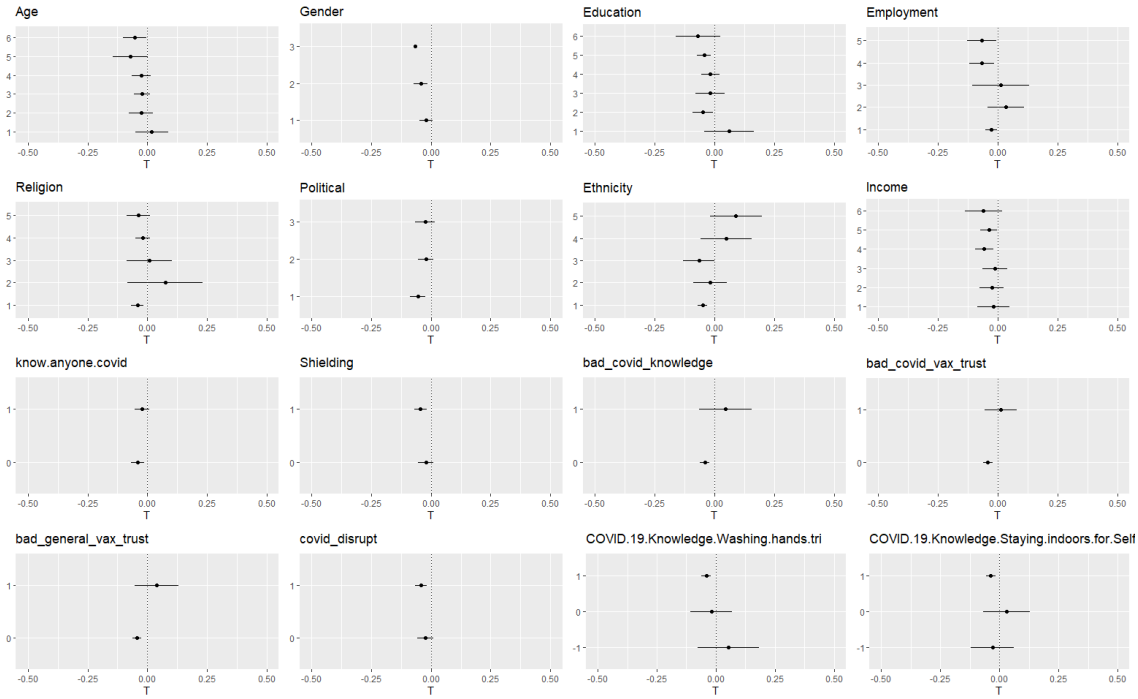

Figure 11: Heterogeneous Treatment Effects

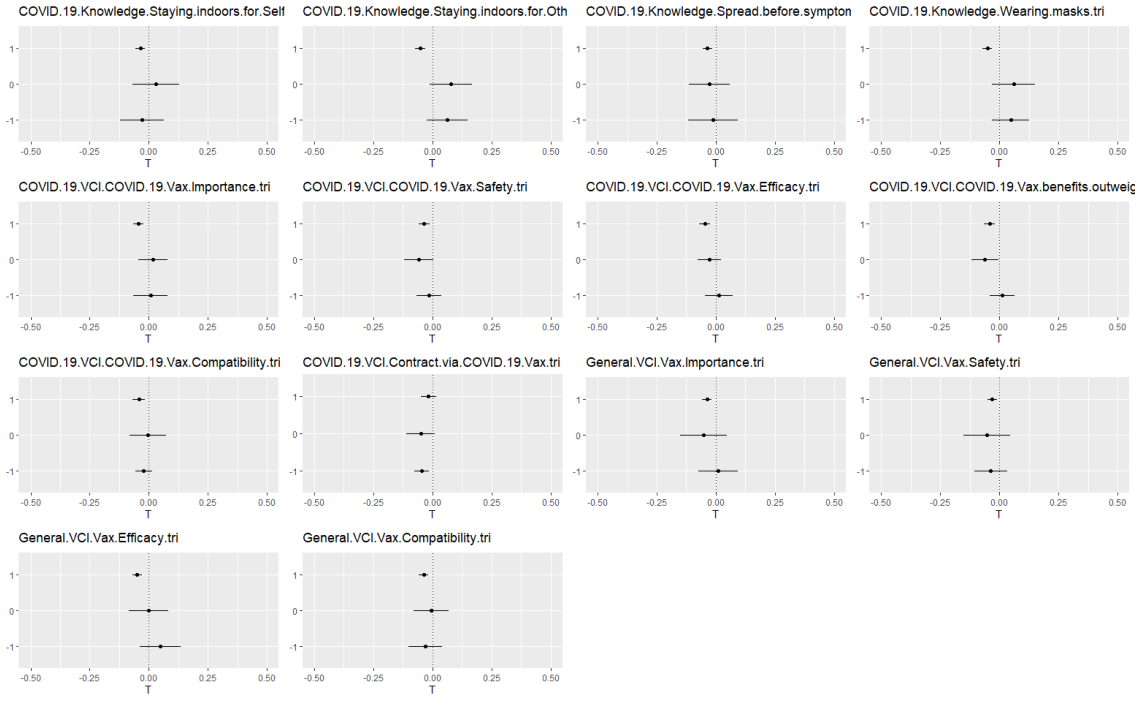

Figure 12: Heterogeneous Treatment Effects

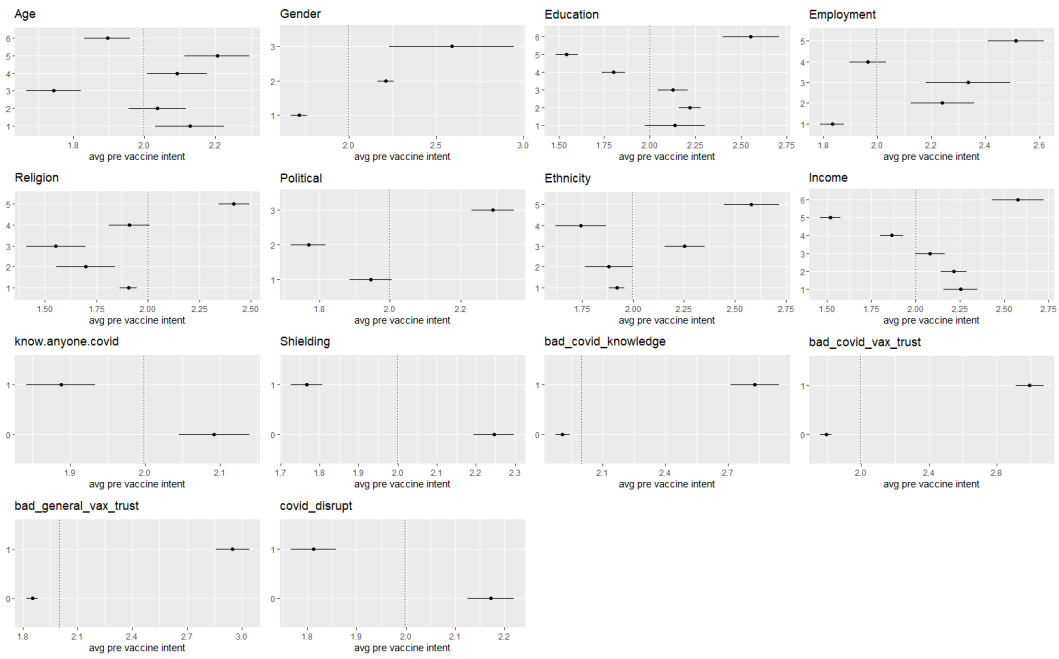

Figure 13: Avg Pre-Treatment Vaccine Intent by Covariates