Pre-Analysis Plan

Minnesota COVID-19 Testing

This Draft: September 17, 2020

1. Introduction

In the United States, recent statistics show that African American and Latinx communities bear a disproportionate burden from COVID-19. Reaching vulnerable and underserved populations is therefore crucial to combating the disease. However, most public messaging campaigns are not targeted toward underserved communities and don't address fears of social stigma, mistrust in the healthcare system, or concerns about immigration status.

The goal of this project is to help the state of Minnesota understand why individuals are not getting tested and potentially identify trusted individuals or organizations that could be used in follow-up work to send messages. To do so, we are deploying flyers through 11 Twin City area food shelves and potentially through public housing units with information on how to answer an online questionnaire.

This provides us with an opportunity to study who answers surveys and why - and what questions are particularly sensitive. This is of general interest to academicians and policymakers alike.

According to Meyer, Mok, and Sullivan (2015) the quality of household surveys is in decline, for three main reasons. First, households have become increasingly less likely to answer surveys at all (unit nonresponse). Second, those that respond are less likely to answer certain questions (item nonresponse). Third, when households do provide answers, they are less likely to be accurate (measurement error). This is important since household surveys help to estimate the employment rate, healthcare needs and of course the census determines resources/representation.

We focus on the first two issues of unit and item nonresponse, which is not random across the population and thus could lead to *nonresponse bias*.[1] Griffin (2002) found that census tracts with predominantly Hispanic or Black residents had significantly lower response rates to the American Community Survey as compared to the response rates in predominantly white tracts. Similarly, Maitland et al. (2017) found that response rates to the Health Information National Trends Survey (HINTS) were lower in areas with higher levels of Hispanic and minority residents.[2]

We hypothesize that financial incentives may encourage unit response; conversely, a close association with the government may discourage response. To test these hypotheses, we plan to cross-randomize the incentive amount offered and the emphasis placed on government involvement in the study on flyers advertising our survey. Individuals will see either a) a 10 dollar incentive, or b) a 20 dollar incentive; and either a) messaging that emphasizes government involvement in the study, or b) messaging that

[1] Nonresponse bias is often defined as the product of the difference in mean responses between those who answer vs. those that do not, and the nonresponse rate, see Maitland et al. (2017).

[2] see also https://www.nytimes.com/interactive/2015/04/20/upshot/missing-black-men.html

emphasizes the involvement of academic researchers. Flyers will be randomized at the foodshelf-date level.

To test what affects item non-response on potentially sensitive questions, such as questions which ask for health information, we hypothesize that ethical framing may encourage individuals to answer questions. This takes two forms — the deontological (or duty based) frame, and the consequential (or cost-benefit) frame. Moreover, knowing others feel the same way (regarding the obligation or benefits of providing health information) may amplify motivation. Finally, there is the possibility that emphasizing the importance of ethnic and racial disadvantage associated with COVID-19 outcomes may also be important.

Upon completion of the demographic module of the survey but prior to starting several potentially sensitive survey modules, individuals will see a message that either a) emphasizes the public health benefits of answering the survey questions (cost-benefit frame); b) emphasizes an individual's responsibility to their community (duty frame); c) emphasizes the disproportionate impact of COVID-19 on certain ethnic and racial groups; or d) provides no messaging. Messaging content will be randomized at the individual level.

2. Sampling and Experimental Protocols

2.1 Recruitment and Sampling

We are recruiting subjects through 11 food shelves in the Twin Cities area. Food shelves will place flyers (see Figure 1 and 2) advertising the baseline survey in bags of food and prepackaged meals that are distributed to food shelf users. We require all participants to be age 18 or older and to speak English or Spanish. We will use a touchless delivery systems to drop off and redeem flyers on a daily basis. Our target sample size is 1000 survey responses. We may stay in the field (i.e. oversample) at foodshelves that tend to serve minority individuals so that we can increase their representation in the survey.

2.2 Experimental Protocols

The structure of our experiment is as follows. Figure 3 provides additional information on study flow.

1. Randomization of Flyers (Unit Response)
    (a) Randomly assign survey advertising flyers to food shelves stratified by bins of dates. These include four bins over the two weeks of planned enrollment. Bins 1 and 3 will be comprised of Monday, Tuesday, and Wednesday on week 1 and 2, respectively. Bins 2 and 4 will similarly be comprised of Thursday, Friday, and Saturday. The unit of analysis will be the foodshelf-date as will the level of randomization. See Appendix A for flyer examples.
    (b) Food shelves distribute the randomized daily flyer via food bags and prepackaged meals.
2. Recruitment and Baseline Survey
    (a) For all potential study participants, elicit a preference for English or Spanish. Those with a preference for Spanish will be given a Spanish language consent form and survey questions.

(b) Individuals provide consent.

(c) Collect demographic information.

3. Randomization of messaging content before sensitive survey modules (Item Response)

   (a) After completing the demographics survey module, participants are randomly assigned to one of four ethical motivation messaging options; deontological, consequential, acknowledgement of racial inequities, or nothing.

   (b) Participants are shown the same randomly assigned message prior to each sensitive survey module.

   (c) Participants answer questions in survey modules 2-5 on media attitudes, health and COVID-19, and discrimination in healthcare.

4. Participants answer debrief questions and are prompted to get tested for COVID-19 if they have any officially recognized symptoms. Participants are also notified that they may share the survey with family and friends.

5. Electronic gift cards are texted to participants who complete the survey to compensate them for their time.

2.3 Randomization

Within each strata we will randomly assign foodshelf-dates to the four different treatments using Stata 14.2 for the unit nonresponse outcomes. Robust standard errors will be used as the level of outcome is the same as the unit of randomization. For item nonresponse outcomes, we will randomize at the individual level in Qualtrics.

2.4 Survey Messaging Content

One of the following message options will be randomly assigned to participants after completing the demographics section of the survey module.

1. **No framing** – No message displayed.

2. **Consequential framing** – "Answering the questions in this survey is an easy way that you can help improve the public health response to COVID-19 in your community."

3. **Deonteological framing** – "It is important for everyone to do their part to protect their community during the COVID-19 pandemic."

4. **Acknowledgement of racial inequities** – "COVID-19 is affecting everyone, but is hitting African American and Latinx communities particularly hard."

3. Hypotheses Tested

In this study, we seek to understand survey takeup and messaging surrounding COVID-19 in minority and low SES populations.

H1. Do higher monetary incentives increase unit response?

H2. Does a government frame reduce unit response?

H3. Are incentives and a de-emphasis on government complements or substitutes in increasing unit response?

H4. a. Which frame/incentive combinations diffuse most from the initial point of distribution throughout the community?

b. How do characteristics of individuals nudged that respond differ by treatment (i.e. frames and incentive structures)?

c. Can we extrapolate using statistical methods and the randomized estimates to obtain population level estimates and approximate missing mass in government surveys?

H5. Do various ethical and (racial/ethnic) acknowledgement frames improve item nonresponse? (including quality)

H6. How do item nonresponse frames interact with variation in the composition of respondents induced through the randomized incentives/frames?

4. Data Collection and Outcomes

We will run our experiment beginning on September 28, 2020.

Flyers advertising the baseline survey will be distributed to the 11 participating food shelves for two weeks. All of the survey responses will be downloaded in a .csv file for cleaning and analysis in Stata.

4.1 Oversampling by Race and Ethnicity

In order to achieve our target population numbers in African American and Latinx communities, we may oversample from food shelves that are located in areas with higher concentrations of these populations. In particular, we may extend the enrollment period in these specific food shelves.

4.2 Baseline

The baseline survey includes demographic characteristics, attitudes towards different media sources, and healthcare experience.

- Gender and age
- Current residence and country of origin
- Education

- Income
- Insurance
- Media attitudes
- COVID-19 knowledge
- Healthcare usage and attitudes on the COVID-19 pandemic
- Discrimination in healthcare

4.3 Data quality checks

Within the survey, we will include a question aimed at capturing respondent attention. We will indicate whether respondents are among the top 5% fastest in terms of total time spent on the survey questions. We will define different samples based on quality cutoffs as well as assess whether quality of responses is affected by our treatments. We will collect phone numbers in order to distribute electronic gift cards. Phone numbers will also be used to identify and drop duplicate survey responses.

4.4 Attrition from the Sample

Attrition will be defined as those who have been given the randomized messaging content but then dropout of the survey and do not complete it within four days of initiation. Attrition will be assessed in real time in order to assess any problematic survey questions. We will check that attrition is not differential across study arms.

5. Empirical Analysis

5.1 Econometric Specification

Our primary specification for evaluating **H1**, **H2**, and **H3** (unit nonresponse) is as follows:

$$Y_{jt} = \alpha + \beta_1 \mathbb{1}_{jt}^{high\_incentive} + \beta_2 \mathbb{1}_{jt}^{research} + \beta_3 \mathbb{1}_{jt}^{research*high\_incentive} + X_j'\Omega + Strata + \epsilon_{jt} \quad (1)$$

Where Y is a measure of unit response at foodshelf j on date t. Since we have no characteristics on individuals who did not visit the landing page, we run this specification at that foodshelf-date level. $\mathbb{1}^{research}$ is an indicator for the framing used on the advertising flyers where a value equal to 0 indicates government sponsor framing and a value equal to 1 indicates academic research sponsor framing.

$\mathbb{1}^{high\_Incentive}$ is an indicator for the \$20 incentive amount – the \$10 category will be omitted.

We will include strata fixed effects. We may also include a set of background characteristics on the foodshelf X in specifications chosen by LASSO (Chernozhukov et al., 2016). Inclusion of these LASSO

covariates may increase the precision of our treatment effect estimates.[3]

The primary outcome Y measure for evaluating unit nonresponse in Equation 1 (hypotheses **H1-H3**) will be visiting the survey link provided on the advertising flyer. This will be calculated as the fraction of distributed flyers at a given foodshelf that are redeemed. The denominator will come from the difference between the total number of flyers dropped off and picked up on a certain day by an RA. The numerator will be the number of individuals who have provided the given the passcode on the flyer (which is uniquely associated with a given foodshelf-date) and respond that they learned about the survey through the foodshelf (an early question on the survey).

We will also explore a secondary outcome of completion of the entire survey calculated using the same denominator and numerator, conditional on completion.

1. **H1**, providing higher incentives, predicts that $\beta_1 > 0$.
2. **H2**, providing research framing, predicts that $\beta_2 > 0$.
3. **H3**, the interaction effect between incentives and researcher framing, is indifferent to the sign on β_3. $\beta_3 > 0$ suggests de-emphasis of government involvement and higher incentives are complementary. $\beta_3 < 0$ suggests they are substitutes.
4. **H4a** replaces the outcome with the difference between all landing page visits from respondents with a given foodshelf-date with all those who were recruited from the foodshelf. The difference reflects the amplification of the flyers throughout social networks. The denominator will continue to be the number of flyers distributed at that foodshelf-date.

$$Y_{ijt} = \alpha + \beta_1 \mathbb{1}_{jt}^{high\_incentive} + \beta_2 \mathbb{1}_{jt}^{research} + \beta_3 \mathbb{1}_{jt}^{research*high\_incentive} + Strata + \epsilon_{jt} \tag{2}$$

To evaluate **H4b** the outcome of Equation 2 will be demographic characteristics amongst those who have responded. For those who respond we will compare the characteristics (e.g. income, ethnicity/race, age, gender) of those that respond to a low vs. high incentive and emphasis vs. de-emphasis government.[4].

See Section 5.3 on testing for heterogeneous treatment effects.

To test **H5** and **H6**, the effect of ethical messaging on item nonresponse will be assessed according to the following specification:

$$Y_i = \alpha + \lambda_1 \mathbb{1}_i^{community} + \lambda_2 \mathbb{1}_i^{duty} + \lambda_3 \mathbb{1}_i^{burden} + \beta_1 \mathbb{1}_{jt}^{research} + \beta_2 \mathbb{1}_{jt}^{high\_incentive} + X'_{ijt}\Omega + Strata + \epsilon_{ijt} \tag{3}$$

Where Y_i is a measure of responsiveness to sensitive questions conditional on takeup of the survey. We are precoding the sensitivity of each question asked in a "sensitive block" by three individuals and

[3]Other fixed effects might also be important for precision, e.g. day of week and food shelf.

[4]Note - we may also receive administrative data on some group characteristics allowing us to run a version like Equation 1

using the average score as a measure for sensitivity. The outcome will therefore be a) a binary outcome that indicates whether the respondent answered any sensitive questions, the b) fraction of sensitive questions answered, c) the mean sensitivity across sensitive questions answered, and the d) overall sum. Note - we will code answered as 1 if answer the question and do not chose "prefer not to answer". Symmetrically, non-response will be coded as skipping a question or choosing "prefer not to answer" and total and fraction non-response will be an additional outcome. *community* is an indicator equal to 1 if the message emphasizing the cost and benefits of answering the survey question is displayed. *duty* is an indicator equal to 1 if the message appealing to a participant's sense of of community is displayed. *burden* is an indicator equal to 1 if the message acknowledging racial and ethnic inequities is displayed. The dummy for no message is omitted. We will also test a specification for item nonresponse that includes the full set of interaction terms among the incentive, framing, and messaging treatments. We hypothesize the $\lambda_1 > 0$, $\lambda_2 > 0$ and $\lambda_3 > 0$. It will also be of interest to test for differences across the treatments.

5.2 Balance Checks

We will conduct a series of balance tests across treatment arms to ensure that there are no chance differences between subjects in the various arms. For the unit nonresponse, food shelf level-date, we will look at differences in characteristics of food shelves across treatment groups including strata fixed effects.

For the item nonresponse, balance tests will be conducted at the individual level on the following baseline characteristics.

- Survey language selected
- Sex
- Age
- Household size
- Race
- Education
- Income
- Insurance coverage

5.3 Heterogeneous Effects

Important secondary analyses will include investigating variation in the treatment response. In particular, we are interested in heterogeneity by:

- Age
- Sex
- Education

- Race
- Food pantry locations

We may also pursue a machine learning-driven approach to uncover heterogeneous treatment effects using the techniques from Chernozhukov et al. (2017) and Chernozhukov et al. (2018b).

5.4 Standard Error Adjustments

For testing the effect of framing and incentives on unit nonresponse, standard errors will be clustered at the foodshelf-date level. For the item nonresponse outcomes, robust standard errors will be used. For assessing the interaction between sponsor framing, incentives, and messaging on item nonresponse,we will employ two-way clustered standard errors.

6. Power

Power calculations for unit nonresponse and item response are shown in Appendix B.

Our power calculations for the effect of non-government framing on unit response compute minimum detectable effects (MDEs) conditional on given levels of control group unit response (i.e. take-up), which we vary from 10% to 30%. We also vary the amount of intra-cluster correlation (ICC) as our unit of randomization is the foodshelf-date. Our calculations assume that our data features 84 food shelf-dates, evenly allocated to two framing arms — government, and researcher — with an average of 100 flyers distributed per food shelf-date.

Fixing government framing as the control group and control take-up at 10%, our calculations imply a MDE of 1.93 percentage points (or 19.3% greater than the control group mean) in the absence of intra-cluster correlation. Holding control take-up constant at 10%, the required MDE rises to 3.43 percentage points (+34.3%) given 2% intra-cluster correlation and to 6.94 percentage points (+69.4%) given 10% ICC. If take-up in the control group is 30%, the required MDE falls to 2.87 percentage points (or 9.6% greater than the control mean) in the absence of ICC, 5.00 percentage points (+16.7%) given 2% ICC, and 9.68 percentage points (+32.3%) given 10% ICC.

In computing MDEs for item response, we assume our target sample size of 1000 survey respondents is evenly divided across four messaging arms. Given 70% item response in the control group of no messaging, we would require a 10.77 percentage point effect size (or 15.38% higher than the control mean) for treatment effects to be detectable. As item response in the control group rises, MDEs fall: given 90% item response in the no messaging group, the MDE is 6.31 percentage points (+7.01% from the control mean).

7. Funding and Human Subjects Review

Funding is provided by the J-PAL State and Local Innovation Initiative. Study approval has been granted by the Harvard Institutional Review Board.

8. References

[1] Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, and Whitney Newey. "Double/Debiased/Neyman Machine Learning of Treatment Effects." American Economic Review 107, no. 5 (2017): 261-65. https://10.1257/aer.p20171038.

[2] Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. "Double/Debiased Machine Learning for Treatment and Structural Parameters." The Econometrics Journal 21, no. 1 (2018a): C1–68. https://doi.org/10.1111/ectj.12097.

[3] Chernozhukov, Victor, Mert Demirer, Esther Duflo, and Iván Fernández-Val. "Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments." Working Paper. Working Paper Series. National Bureau of Economic Research, June 2018b. https://doi.org/10.3386/w24678.

[4] Griffin, Deborah H. "Measuring Survey Nonresponse by Race and Ethnicity." 4700 Silver Hill Road, Washington, DC 20233-8700 Bureau of the Census: United States Bureau of the Census, 2002.

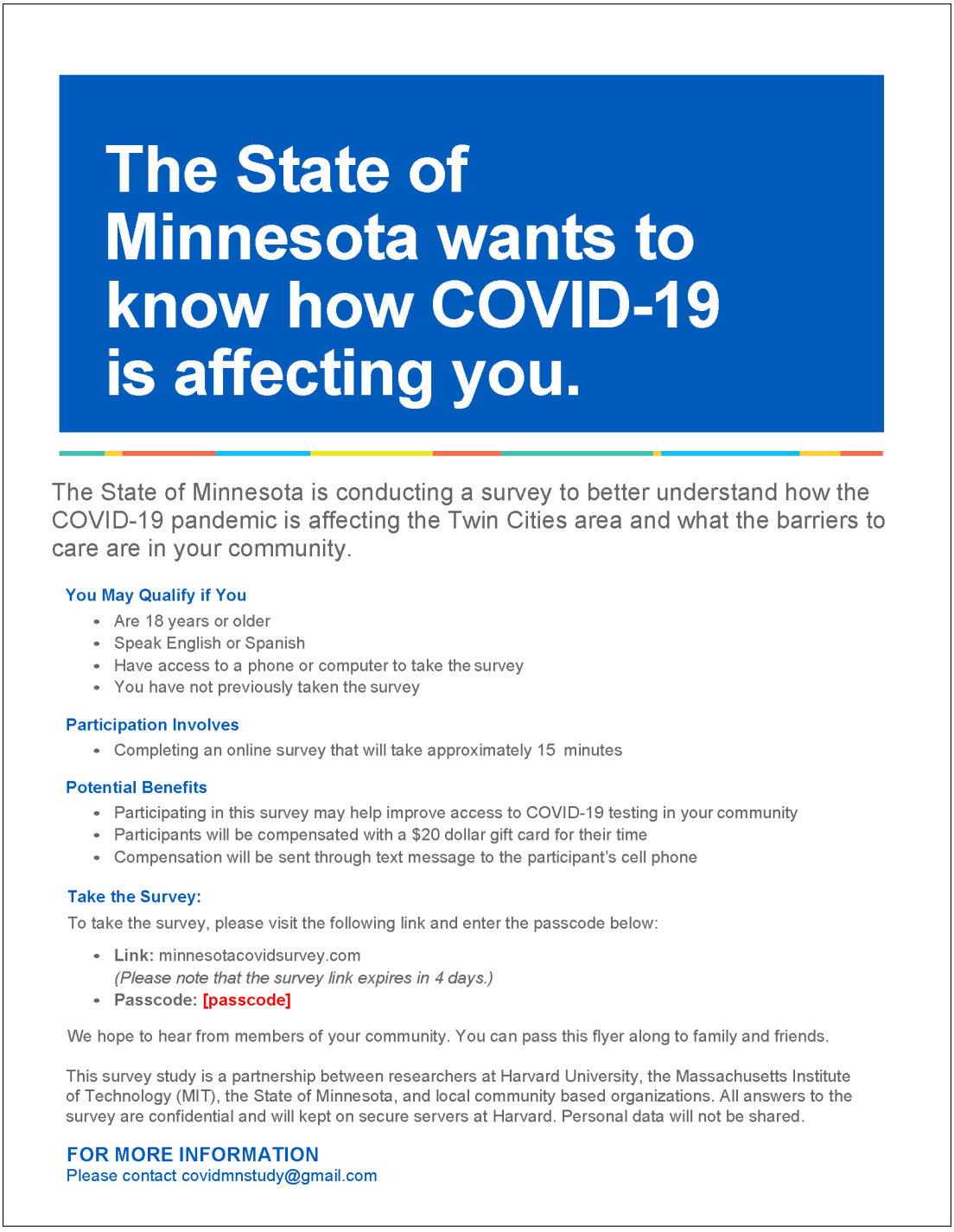
[5] Maitland, Aaron, Amy Lin, David Cantor, Mike Jones, Richard P. Moser, Bradford W. Hesse, Terisa Davis, and Kelly D. Blake. "A Nonresponse Bias Analysis of the Health Information National Trends Survey (HINTS)." Journal of Health Communication 22, no. 7 (July 2017): 545–53. https://doi.org/10.1080/10810730.2017.1324539.

[6] Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan. 2015. "Household Surveys in Crisis." Journal of Economic Perspectives, 29 (4): 199-226.

[7] Wolfers, Justin, David Leonhardt, and Kevin Quealy. "1.5 Million Missing Black Men." The New York Times, April 20, 2015, sec. The Upshot.

Appendix

A. Survey Advertising Fliers



The State of Minnesota wants to know how COVID-19 is affecting you.

The State of Minnesota is conducting a survey to better understand how the COVID-19 pandemic is affecting the Twin Cities area and what the barriers to care are in your community.

You May Qualify if You

- Are 18 years or older
- Speak English or Spanish
- Have access to a phone or computer to take the survey
- You have not previously taken the survey

Participation Involves

- Completing an online survey that will take approximately 15 minutes

Potential Benefits

- Participating in this survey may help improve access to COVID-19 testing in your community
- Participants will be compensated with a $20 dollar gift card for their time
- Compensation will be sent through text message to the participant's cell phone

Take the Survey:

To take the survey, please visit the following link and enter the passcode below:

- **Link:** minnesotacovidsurvey.com
  (Please note that the survey link expires in 4 days.)
- **Passcode:** **[passcode]**

We hope to hear from members of your community. You can pass this flyer along to family and friends.

This survey study is a partnership between researchers at Harvard University, the Massachusetts Institute of Technology (MIT), the State of Minnesota, and local community based organizations. All answers to the survey are confidential and will kept on secure servers at Harvard. Personal data will not be shared.

FOR MORE INFORMATION
Please contact covidmnstudy@gmail.com

Figure 1: Government Framing Flier

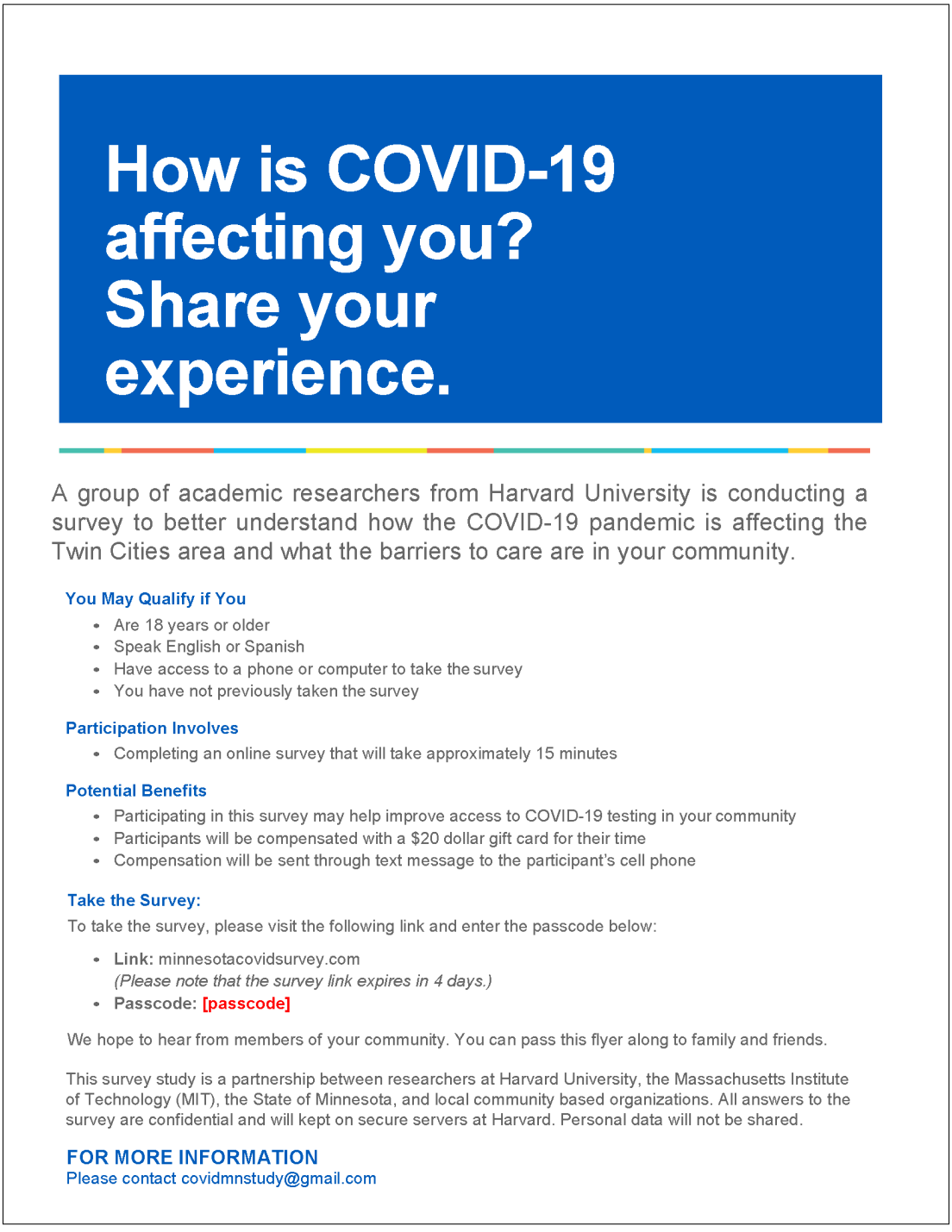


Figure 2: Academic Researcher Framing Flier

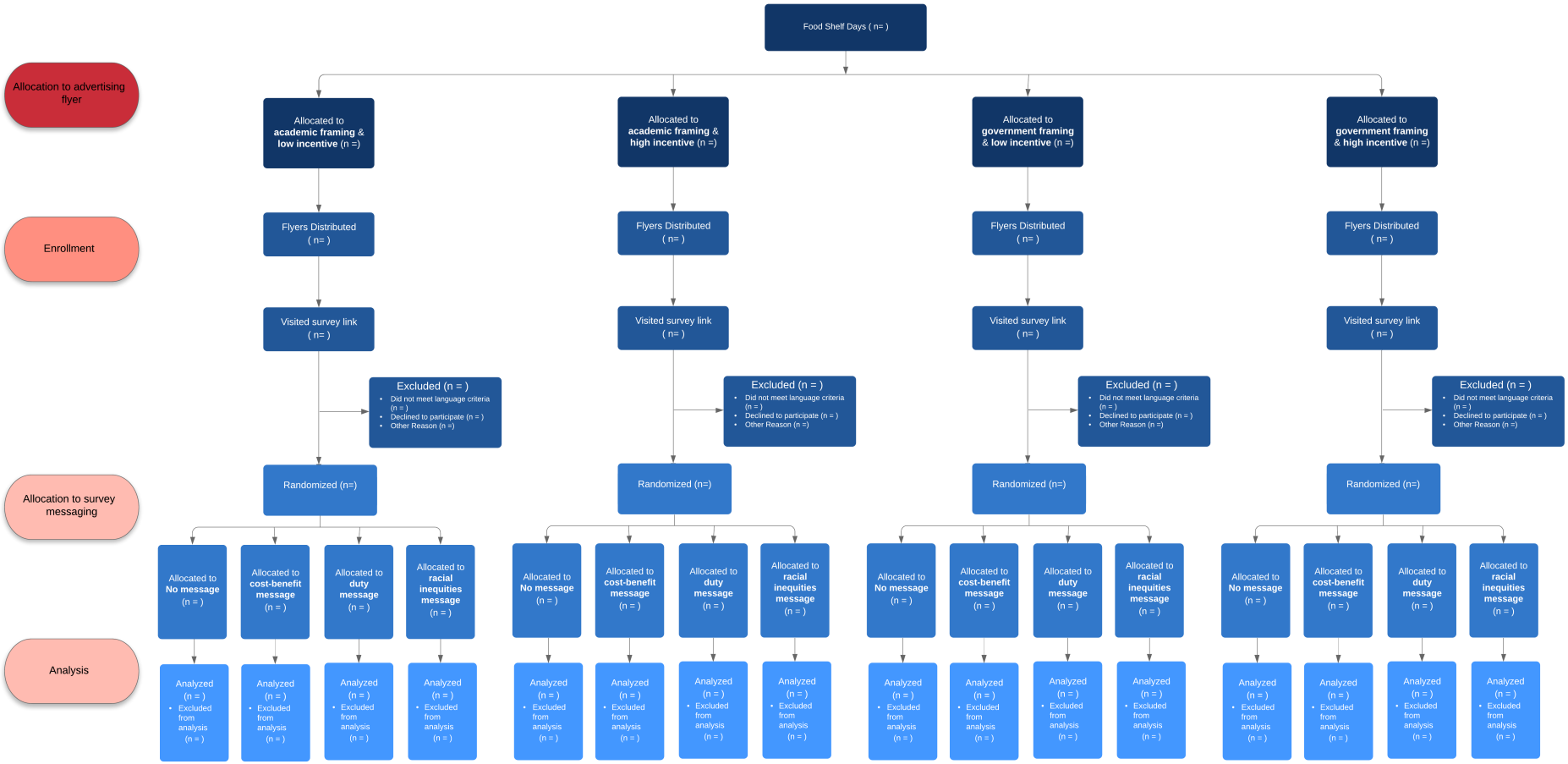


Figure 3: Consort Diagram

B. Power Calculations

All power calculations are shown in Table 1 and Table 2 are MDEs assuming α of 0.05 and power of 80%. MDEs are displayed both as percentage points and as percentage of the control group mean.

Table 1: Unit Response MDEs by Control Group Response Rate and Intra-Cluster Correlation

Control Group Response	ρ	MDE (Percentage Points)	MDE
10%	0%	1.93	19.3%
10%	2%	3.43	34.3%
10%	4%	4.51	45.1%
10%	6%	5.41	54.1%
10%	8%	6.21	62.1%
10%	10%	6.94	69.4%
20%	0%	2.53	12.7%
20%	2%	4.43	22.2%
20%	4%	5.78	28.9%
20%	6%	6.88	34.4%
20%	8%	7.85	39.3%
20%	10%	8.72	43.6%
30%	0%	2.87	9.6%
30%	2%	5.00	16.7%
30%	4%	6.48	21.6%
30%	6%	7.69	25.6%
30%	8%	8.74	29.1%
30%	10%	9.68	32.3%

Power calculations for unit response were conducted using the *clustersampsi* package in STATA 14.

Table 2: Item Response MDEs by Control Group Response Rate

Control Group Response	MDE (Percentage Points)	MDE
70%	10.77	15.38%
75%	10.00	13.33%
80%	9.04	11.30%
85%	7.85	9.23%
90%	6.31	7.01%