

# Global survey on COVID-19 beliefs, behaviors, and norms

This version: 30 November 2020

Avinash Collis<sup>a,b,1</sup>, Kiran Garimella<sup>c,1</sup>, Alex Moehring<sup>d,1</sup>, M. Amin Rahimian<sup>a,c,e,1</sup>, Stella Babalola<sup>f,h,2</sup>, Nina Gobat<sup>i,j,2</sup>, Dominick Shattuck<sup>h,k,2</sup>, Jeni Stolow<sup>l,j,2</sup>, Dean Eckles<sup>a,c,d,3</sup>, and Sinan Aral<sup>a,c,d,3</sup>

<sup>a</sup> Initiative on the Digital Economy, Massachusetts Institute of Technology

<sup>b</sup> McCombs School of Business, The University of Texas at Austin

<sup>c</sup> Institute for Data, Systems, and Society, Massachusetts Institute of Technology

<sup>d</sup> Sloan School of Management, Massachusetts Institute of Technology

<sup>e</sup> Department of Industrial Engineering, University of Pittsburgh

<sup>f</sup> Bloomberg School of Public Health, John Hopkins University

<sup>h</sup> John Hopkins Center for Communications Programming

<sup>i</sup> University of Oxford

<sup>j</sup> Global Outbreak Alert and Response Network

<sup>k</sup> Breakthrough ACTION

<sup>l</sup> College of Public Health, Temple University

<sup>1</sup> Authors listed in alphabetical order.

<sup>2</sup> Authors listed in alphabetical order.

<sup>3</sup> To whom correspondence should be addressed: [mit\\_covid\\_survey\\_researchers@mit.edu](mailto:mit_covid_survey_researchers@mit.edu), [eckles@mit.edu](mailto:eckles@mit.edu), [sinan@mit.edu](mailto:sinan@mit.edu).

## Abstract

Policy and communication responses to COVID-19 can benefit from better understanding of people's baseline and resulting beliefs, behaviors, and norms. We fielded a global survey on these topics. This report provides an overview of the motivation behind the survey design, outlines the survey content and its representation in the respondent-level data, and details the sampling and weighting designed to make the results representative of populations of interest.

## Data Access

Updated aggregate data can be found [here](#), and researchers can request access to individual level responses by requesting access [here](#).

## Citation

Collis, A., Garimella, K., Moehring, A., Rahimian, M.A., Babalola, S., Gobat, N., Shattuck, D., Stolow, J., Eckles, D., & Aral, S. (2020). Global survey on COVID-19 beliefs, behaviors, and norms. Technical report, MIT Sloan School of Management.

## **Table of Contents**

[Introduction](#)

[Sampling & Weighting](#)

[Completion types](#)

[Non-response modeling and post-stratification](#)

[Using the weights](#)

[Internal consistency of the non-response models](#)

[Survey Description & Data Dictionary](#)

[Additional Information](#)

[Acknowledgements](#)

[References](#)

[Appendix A: Data Dictionary](#)

[Missing Values](#)

[Sample Code](#)

[Python \(pandas\)](#)

[R](#)

[Meta Data](#)

[Information Exposure](#)

[Knowledge and Positive Cases](#)

[Vaccine and Healthcare](#)

[Demographics](#)

[Information Sources, Trust & Needs](#)

[Basic Knowledge](#)

[Distancing Familiarity, Importance & Norms](#)

[Risk Perceptions and Locus of Control](#)

[Prevention Behaviors in Practice](#)

[Behavioral Measures Taken, Efficacy & Norms](#)

[Employment and Work](#)

[Intentions to Visit Locations](#)

[Travel](#)

[Future Actions](#)

[Appendix B: Survey Change Log](#)

[Appendix C: Survey Instrument](#)

## Introduction

We are conducting a large-scale, international survey related to COVID-19 in over 60 countries to help policymakers and researchers better monitor and understand people's knowledge, beliefs, behaviors, and risk perceptions across the world. While similar in sampling design to the symptoms survey arising from the collaboration between Facebook, Carnegie Mellon University's Delphi Research Group, and the University of Maryland's Joint Program in Survey Methodology, the survey's purpose is to guide policy and research around individual responses to COVID-19 beyond symptoms and the most closely associated behaviors.

The survey went live on July 7, 2020 and data is continuing to be collected until at least December 2020. The survey was translated to 51 languages. As of December 4, 2020, we had received over 1,000,000 completes from 67 countries. The full survey instrument is provided in the Appendix.

## Sampling & Weighting

Two versions of the survey were fielded. First, in countries with a sufficient pool of users to sample, we fielded a multi-wave survey that runs continuously for multiple two-week waves. In each wave, Facebook aimed to deliver 3,000 respondents to our survey. In countries with a limited survey pool, we fielded a snapshot survey where Facebook aimed to deliver 3,000 respondents over a two-week period.

The Facebook team uses non-response modeling and poststratification techniques from survey statistics to design the two following components (Barkay et al. 2020):

1. Sampling - deciding who to present with the invitation to participate in the survey.
2. Weighting - providing a weight per user so that respondents better represent the target population as a whole.

Using the total survey error framework, the goal of the sampling and weighting steps are to minimize the representation error due to the coverage, sampling variability, and nonresponse biases.

## Completion types

The MIT team supplied binary survey completion flags (binary indicators of whether or not each respondent has completed the survey) along with a respondent identifier (a random number associated with each survey respondent) back to the Facebook team. No other data about respondents was sent by MIT to Facebook. We provide the completion flags for the following two analytical samples:

1. Respondents who have completed the basic knowledge and demographics parts of the survey. This part consists of a briefing followed by questions about information exposure, availability of treatments and vaccines, and contact with healthcare workers, as well as gender, age, education, overall health, country, and in the case of the US and India, state as well. We call this the *demographic completion type*.
2. Respondents who have reached the end of the entire survey, viewing (and typically answering) additional questions about information sources; information needs; their knowledge about high risk populations, methods of transmission, and disease symptoms; norms and beliefs about distancing, mask wearing, and other preventative measures; risk perception and locus of control; work, travel and intentions to visit various locations, followed by a debrief. We call this the *full survey completion type*, though note that there can still be missing data due to nonresponse to individual questions.

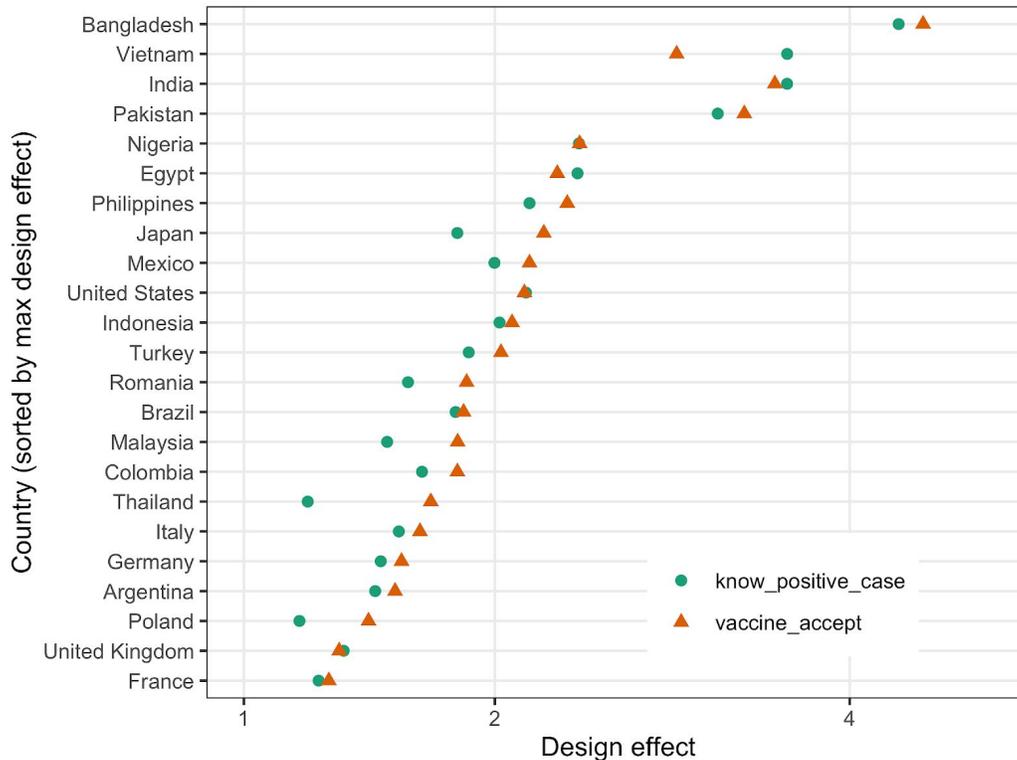
Subsequently, the Facebook team computed and returned sets of survey weights to the MIT team, one set for each analytical sample. No other data about respondents was sent by Facebook to MIT besides a respondent identifier (a random number associated with each survey respondent), their language preference, these survey weights, and an indicator of whether these survey weights were clipped (see below).

## Non-response modeling and post-stratification

Based on the various attributes of each client and their patterns of engagement with the Facebook platform, the Facebook team designed a regularized regression and post-stratification model to issue a set of weights for the survey respondents in each wave (Wang et al. 2015). Subsequently, the survey respondents are modeled as differing not only in their demographics (through post-stratification), but also in their response or nonresponse behaviors. The regression model outputs the probability of responding to the survey as a function of the sample attributes. After training the regression model, the resultant non-response weights are calculated for each survey respondent as the inverse of their response probability (i.e., the inverse probability of sampling weights, IPSW). The Facebook team tested multiple models of survey nonresponse behavior with varying degrees of complexity (in terms of the number of included attributes and their interactions) and chose the most parsimonious model with low design effects (equation 1). The non-response weights provide us with a representative sample of the Facebook active user base, irrespective of their non-response behavior or factors that might have prevented them from being exposed to Facebook's invitation to the survey.

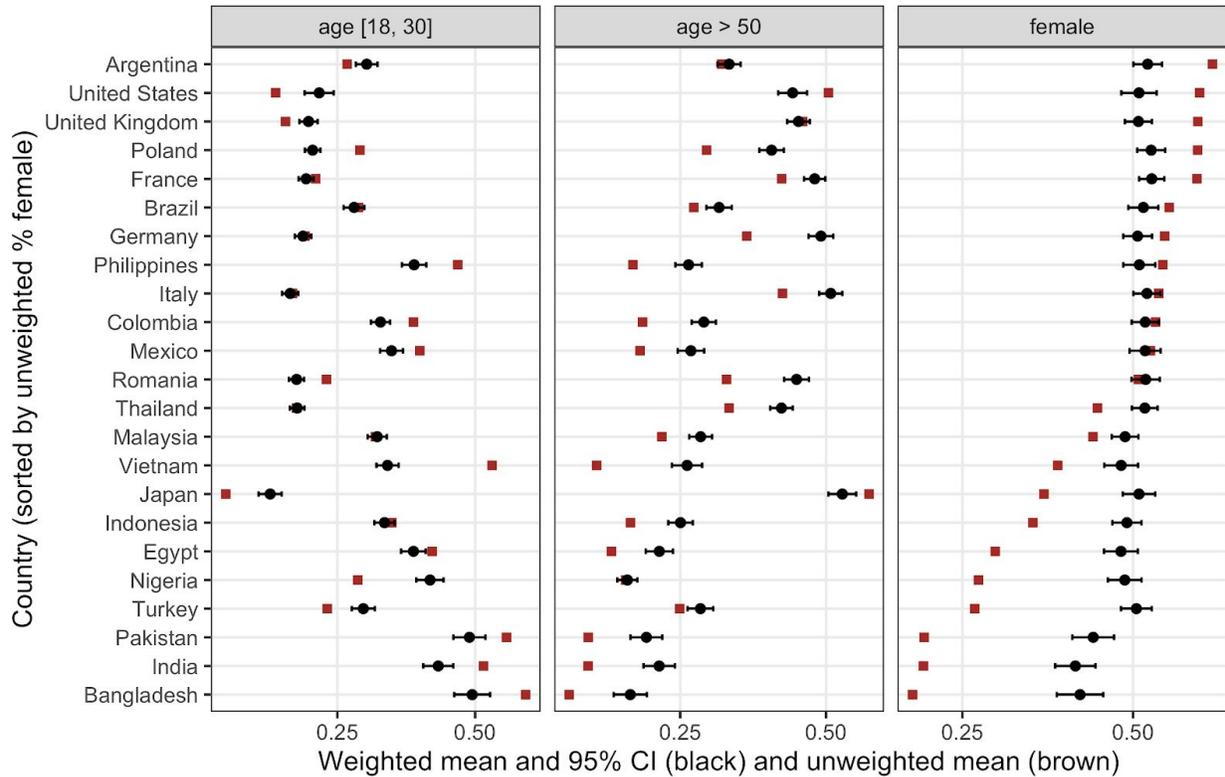
After weighting for nonresponse, by comparing the demographic data (age bracket and gender) of the respondents with the census data in each country, the weights were readjusted to match the target population which we aim to represent. This uses post-stratification, a common method in survey analysis to correct for known differences between the sample and target populations (Little 1993). For each country, the Facebook team considered a two-way cross-classification of the age brackets and gender, and provided weights for the country's adult population. Although post-stratification reduces the bias and increases the representativeness of the survey sample,

the added adjustment to the weights induces an increased design effect. We aimed to keep the overall design effects due to both the non-response and post-stratification weighting below four, and it remained below two for most counties (Figure 1).



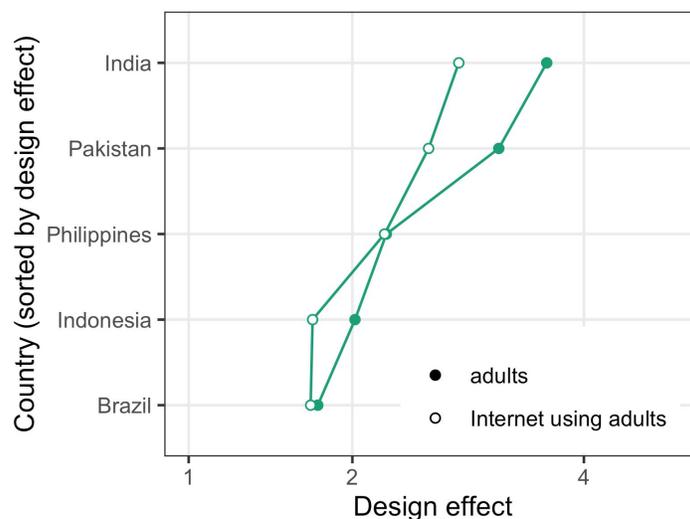
**Figure 1.** Design effect from weighting wave 1, using the demographic completion weights, for two of the questions that occur prior to the demographics block. Design effects for most countries are near or below two.

Using the survey data, we can plot the self-reported demographics of each country with and without the weights to see the effect of post stratification (Figure 2). In some cases, Facebook’s inferred demographics may not match what the users report in the survey. This causes a mismatch between the census data and weighted self-reported demographics (e.g., female population in Pakistan, India and Bangladesh in Figure 2 remain below their census levels). Since our survey results come with self-reported demographic information, the researchers who use this data can perform their own adjustments to the weight in trying to match their target population’s demographics.



**Figure 2.** Illustration of change in self-reported demographics from weighting of wave 1. Unweighted proportions in brown show country-specific biases. These are substantially reduced when using weights — here the weights for the demographic completion type. The self-reported fraction female remains statistically significantly below 0.5 in Pakistan, India, and Bangladesh even after weighting.

For some countries, weights for another target population are provided (in addition to the adult population weights). In particular, for some countries where data on their Internet-using population was available, there was (a) substantial divergence between their adult Internet users and adult population and (b) substantial divergence between the respondent population and the adult population, such that post-stratification by age and gender results in a large reduction in effective sample size and may also leave much bias remaining. For some of these countries, and where data on demographics of Internet users is available, we include weights where the target population is adult Internet users. Targeting the population of internet-using adults will reduce the design effect (Figure 3) as presumably the demographics of the Facebook active user base is more similar to that of adult Internet users than all adults.



**Figure 3.** Design effect from weighting when post-stratifying to match adult population or Internet-using adult population, for those wave countries where the latter are available. Both design effects are computed for wave 1 with the demographic completion type for knowing a positive case.

Finally, the weights were clipped (i.e., trimmed, Winsorized) from below and above to control their variance and especially to reduce sensitivity to a few samples with very large weights. In particular, Facebook trimmed the weights within each country at 10 times above and 30 times below their (untrimmed) means for that country; then, because of possible mismatch between the country associated with a respondent in Facebook data and that from the survey data, MIT made further adjustments to the weights within each country. Specifically, MIT has two ways of identifying each participant's country: (i) the respondent's self reported country in the demographics part of the survey and (ii) the GeoIP location data that MIT obtains from Qualtrics for each respondent. The Facebook team has their own way of determining each participant's country using their internal data which per our legal agreements is not shared with MIT. In addition to providing a weight for each survey respondent, the Facebook team also provides an indicator variable of whether the weight for that respondent has been clipped or not. Subsequently, the MIT team compares each respondent's weight with other respondents that have the same reported or GeoIP-identified country. If a participant's weight is larger than the level at which Facebook has clipped the weights for that country from above (inferred from the clipping indicators), then the MIT team determines that the participant's country is misidentified and such participants' weights are subsequently reset to zero (removed from the sample). Although such misidentified cases happen very rarely (a few such cases have been observed thus far), this correction is important because such large weights could substantially bias the survey results.

Given the final weight  $w_i$  for each respondent  $i$ , we can compute the design effect and effective sample size. First, one can compute these without reference to a particular outcome:

$$(Kish) \text{ design effect} = \frac{n \sum_{i=1}^n (w_i)^2}{(\sum_{i=1}^n w_i)^2}, \quad (1)$$

$$(Kish) \text{ effective sample size} = \frac{(\sum_{i=1}^n w_i)^2}{\sum_{i=1}^n (w_i)^2}, \quad (2)$$

where  $n$  is the total number of completed survey responses. These formulas were used by the Facebook team to benchmark the design effects as they did not have access to the outcome data (recall that only random respondent IDs and completion indicators were sent from MIT to Facebook). On the other hand, MIT can additionally compute design effects and effective sample sizes for each outcome, accounting for the correlation between the outcome and the weights. Then the design effect is estimated using the ratio between the variance of the weighted estimator and an unweighted estimator using an equality probability sample; this can also account for further changes in missingness for that particular item, which can be assumed to be missing unconditionally (completely) at random. This is what is reported in Figures 1 and 3.

## Using the weights

The weights are meant to be used in Hajek estimators (i.e., normalized importance sampling estimators) for measuring population means. Specifically, let  $Y_i$  be an outcome variable of interest measured for the respondent  $i$  whose weight is  $w_i$ . The Hajek estimator,  $\hat{Y}$ , for the population mean of the outcome,  $\bar{Y}$ , is given by:

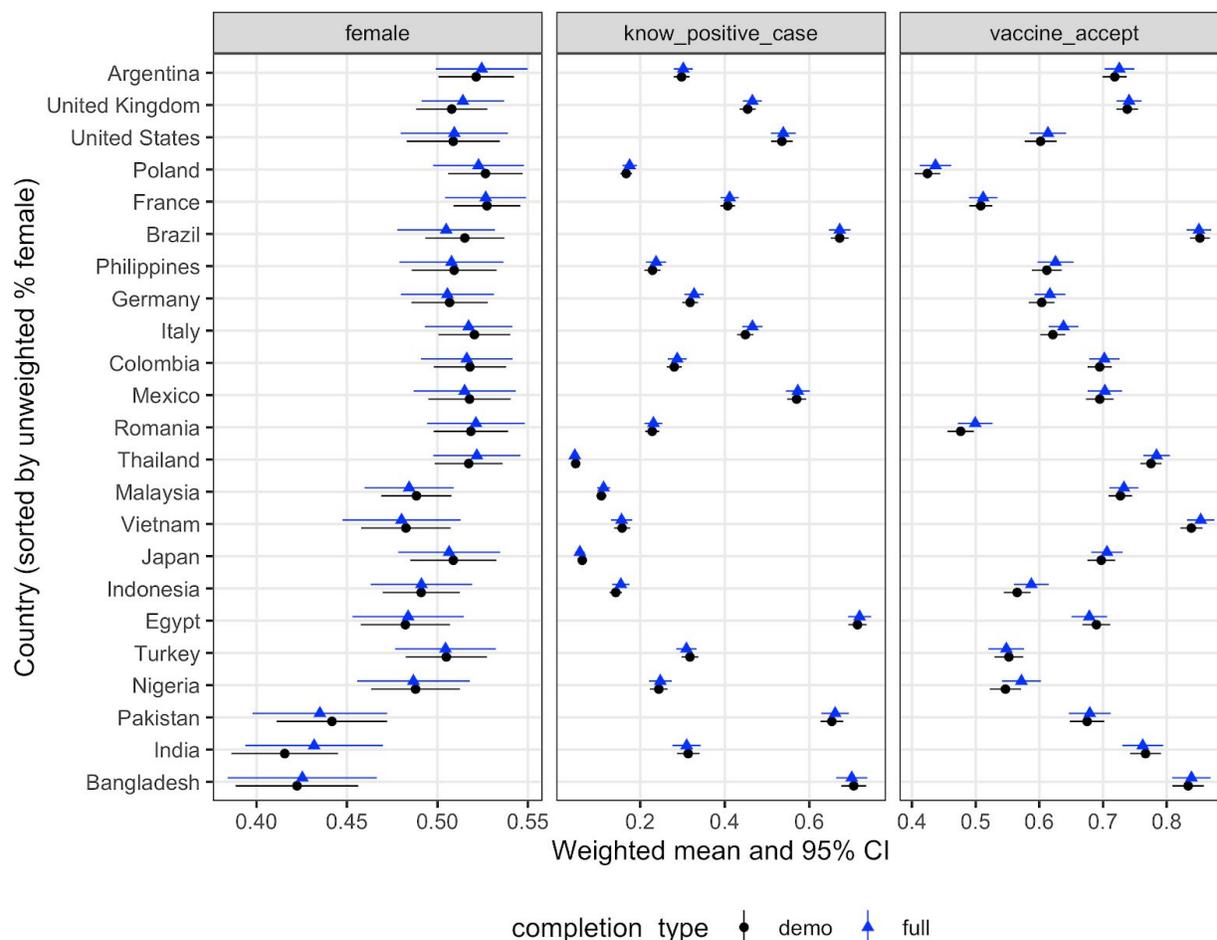
$$\hat{Y} = \frac{\sum_{i=1}^n w_i Y_i}{\sum_{i=1}^n w_i}, \quad (3)$$

This is the default in most statistical software for computing a weighted mean. Subsequently, if interested in population totals, analysts should use  $N\hat{Y}$  as an estimator of the total outcome level where  $N$  is the population size. That is, analysts should not use the weights in an unnormalized way, as in a Horvitz–Thompson estimator (i.e., an unnormalized importance sampling estimator), as, while the weights are approximately on the level of each country's adult population, the clipping and other adjustments to the weights make them unsuitable for direct estimation of total outcome levels without normalization. More generally, users can use these weights in other related estimators that appropriately normalize the weights.

## Internal consistency of the non-response models

Because of the creation of weights for two different completion types, it is possible to test some observable implications of the assumptions of the non-response model (e.g., whether someone responds is as-good-as-random conditional on the observables used in the nonresponse model above). In particular, for the questions in the initial blocks of the survey,

estimation using either the demographic completion type or full survey completion type should yield very similar answers. Analysis of two substantive questions suggests that these observable implications of this assumption are not egregiously violated (Figure 4).



**Figure 4.** Comparison of weighted estimates of proportions using the demographic completion and full survey completion weights for wave 1. Under conditional ignorability of nonresponse, these should be the same in large samples.

## Survey Description & Data Dictionary

The survey is organized into blocks based on the question topics. Every survey begins with questions from the same five blocks: information exposure, knowledge, vaccine and healthcare, and demographics. In snapshot countries, all respondents are shown an information block and then three additional blocks that are randomly selected from the remaining blocks. In multi-wave countries, respondents are shown four randomly selected blocks. Precise questions and the codebook for the data can be found in the appendix.

In constructing the survey instrument, we leveraged a wide set of domain experts as well as existing surveys in the field, including Fetzer et al. (2020) and Allcott et al. (2020). This survey is complementary to the existing surveys as we leverage Facebook’s weights that use the rich information Facebook has about users to balance covariates.

## Additional Information

- MIT portal for survey information ([link](#))
- KAP COVID Dashboard by Johns Hopkins University ([link](#))
- Facebook Data for Good ([link](#))
- Data Request Site ([link](#))

## Acknowledgements

This work was funded in part by a grant from Facebook to the MIT Initiative on the Digital Economy.

This survey was a collaborative effort involving contributions from individuals at multiple institutions, especially the Massachusetts Institute of Technology, Johns Hopkins University, the Global Outbreak Alert and Response Network (GOARN), the World Health Organization, and Facebook. At Facebook, key contributors were made by Esther Kim, Kelsey Mulchahy, Praveen Raja, Stephanie Sasser, Carlos Velasco, and Thomas Wynter.

## References

- Allcott, Hunt, Levi Boxell, Jacob Conway, Matthew Gentzkow, Michael Thaler, and David Yang. “Polarization and Public Health : Partisan Differences in Social Distancing during COVID-19.” *Journal of Public Economics (forthcoming)*, 2020.
- Barkay, Neta, Curtiss Cobb, Roe Eilat, Tal Galili, Daniel Haimovich, Sarah LaRocca, Katherine Morris, and Tal Sarig. 2020. “Weights and Methodology Brief for the COVID-19 Symptom Survey by University of Maryland and Carnegie Mellon University, in Partnership with Facebook.” arXiv [cs.SI]. arXiv. <http://arxiv.org/abs/2009.14675>.
- Fetzer, Thiemo, Marc Witte, Lukas Hensel, Jon Jachimowicz, Johannes Haushofer, Andriy Ivchenko, Stefano Caria, et al. “Global Behaviors and Perceptions in the COVID-19 Pandemic.” <https://doi.org/10.31234/osf.io/3kfmh>, n.d.
- Little, Roderick JA. "Post-stratification: A modeler's perspective." *Journal of the American Statistical Association* 88, no. 423 (1993): 1001-1012.

Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. "Forecasting elections with non-representative polls." *International Journal of Forecasting* 31, no. 3 (2015): 980-991.

## Appendix A: Data Dictionary

### Missing Values

Data can be missing for a number of reasons, some missing is by design through randomization and other is due to selective answering and dropout. For all questions, missing values are coded as follows.

- NULL: Null values indicate the question was not shown either due to randomization or because the respondent left the survey before this point.
- -1: The question was viewed but not answered and the respondent completed the block containing this question.
- -2: The question was viewed but not answered and the respondent did not complete the block containing this question.

### Sample Code

Below are commands that can load the data in commonly used data analysis software. The data is a compressed tab-delimited text file.

#### Python (pandas)

```
import pandas as pd
df = pd.read_table('covid_survey_responses.txt.gz', sep='\t',
low_memory=False)
```

#### R

```
library(readr)
df = read_tsv(
  "covid_survey_responses.txt.gz"),
  col_types = cols(
    us_state = col_character(),
    india_state = col_character(),
    geoip_state = col_character(),
    wave = col_double(),
    weight_demo = col_double(),
    weight_full_survey = col_double(),
    trim_indicator_demo = col_integer(),
    trim_indicator_full_survey = col_integer(),
    weight_internet_demo = col_double(),
    weight_internet_full_survey = col_double(),
    trim_indicator_internet_demo = col_integer(),
    trim_indicator_internet_full_survey = col_integer())
```

## Meta Data

Variable	Type	Description
id	string	Unique id assigned to a survey response
start_date	YYYY-MM-DD	Date survey was started
duration	integer	Duration of survey in seconds
language	string	Language code survey was taken in
country	string	Inferred country of residence (See notes below)
reported_country	string	Self-reported country of residence (See notes below)
geoip_country	string	Country of residence inferred from Qualtrics GeoIP (See notes below)
progress	integer	Integer representing share of survey complete, between 0 and 100.
survey_type	string	Survey version, either waves or snapshot
survey_version	string	One of v1, v2, v3, representing the iteration of the survey taken.
wave	int	Wave number if survey_type is waves, null otherwise
device_type	string	Device used to complete survey (desktop, mobile, other)
browser	string	Name of browser used to complete survey
demo_complete	bool	Indicator if the respondent answered all demographic questions not including location questions and had a non-null country
finished	bool	Indicator if the respondent reached the end of the survey

weight_[[completion type]]	float	Weight using respondents who complete [[completion type]] as an analytical sample and adult population as a target population
weight_internet_[[completion type]]	float	Weight using respondents who complete [[completion type]] as an analytical sample and adult internet users as a target population
trim_indicator	bool	Indicator if the respondent's weight was clipped by Facebook.
display_order	string	<p>JSON string containing the order questions, answers, and blocks were displayed.</p> <pre>{   "block_order": {     block name: block order   }   "question_order": {     question id: question order     within block. Only populated for     questions that have a randomized     order.   }   "block_completions": {     block: Dictionary indicating if     block was started and/or completed.   }   "last_finished_block": Last block   completed in the survey   "last_started_block": Last block   started in the survey   "answer_order": {     question: Dictionary mapping     question answers to display     positions. Only populated for     questions with randomization in     answers.   } }</pre> <p>Note for roughly 600 responses the display order was improperly logged and will be null. For these responses, we are also unable to provide the full missingness codes. For these 600 responses, null implies the question was not seen and -1 implies the question was shown but not answered.</p>

If either `geoip_country` or `reported_country` were not in the set of sampled countries, we report null for country. If self-reported country was present and in the sampled set, we infer country equal to the self-reported country. If a self-reported country was missing, or the self-reported country was not in the sampled set, we infer the country equal to the country from the GeolIP field. If the GeolIP country is missing or not in the sampled countries as well, then we do not infer a country and this field will be null.

## Information Exposure

- `info_exposure_past_week` (string): In the past week, how much, if anything, have you heard or read about coronavirus (COVID-19)? [Nothing, A little, A moderate amount, A lot]
- `info_exposure_more_less_wanted` (string): In the past week, did you see more or less news than you wanted to see about coronavirus (COVID-19)? [Much more, More, About the right amount, Less, Much less]

## Knowledge and Positive Cases

- `knowledge_existing_treatments` (string): Which of the following is correct? [There is a drug to treat COVID-19, There is a vaccine for COVID-19, There is both a drug for treatment and a vaccine for COVID-19, There is currently no drug treatment or vaccine for COVID-19, I am unsure which is correct]
- `know_positive_case` (string): Do you personally know someone who has tested positive for COVID-19? [Yes, No, Prefer not to say]

## Vaccine and Healthcare

- `vaccine_accept` (string): If a vaccine for COVID-19 becomes available, would you choose to get vaccinated? [Yes, No, Don't know]
- `healthcare_avoid_contact` (string): In the past week, have you avoided contact with health care facilities or health care workers due to fear of exposure to COVID-19? [Yes, No]
- `healthcare_availability` (string): In the past week, have you been able to get an appointment with a healthcare worker when you needed one, whether related to COVID-19 or not? [Yes, No, I needed one and was not able, No, I did not need an appointment]
- `Flu_vaccine`: (string) Have you taken a flu vaccine this fall or do you plan to take one in the coming weeks? [Yes, No, Don't Know]. This question was only fielded in North American countries.

## Demographics

- `gender` (string): [Male, Female, Other]

- `age` (string): [Under 20, 20-30, 31-40, 41-50, 51-60, 61-70, 71-80, Over 80]
- `education` (string): What is the highest level of education you have completed? [Less than primary school, Primary school, Secondary school, College / university, Graduate school]
- `own_health` (string): In general, how would you rate your overall health? [Excellent, Very good, Good, Fair, Poor]
- `country` (string): In which country do you currently reside?
- `us_state` (string): In which state do you currently reside?
  - Only asked if `country == United States of America`
- `india_state` (string): In which state or union territory do you currently reside?
  - Only asked if `country == India`
- `density` (string): Which of these best describes the area where you are currently staying? [City, Town, Village or rural area]

## Information Sources, Trust & Needs

- **Information Medium:** In the past week, from which of the following, if any, have you received news and information about COVID-19? Select all that apply. [1=option selected, 0=option shown but not selected, and another option was selected. Note: if all values have -1 or -2, this question was viewed and the options were presented, but no options were selected.]
  - `news_medium_online_sources`: Online sources (websites, apps, social media)
  - `news_medium_messaging_apps`: Messaging apps / SMS/ text messaging
  - `news_medium_newspapers`: Newspapers
  - `news_medium_television`: Television
  - `news_medium_radio`: Radio
- **Information Sources:** In the past week, from which of the following, if any, have you received news and information about COVID-19? Select all that apply. [1=option selected, 0=option shown but not selected, and another option was selected. Note: if all values have -1 or -2, this question was viewed and the options were presented, but no options were selected.]
  - `news_sources_local_health_workers`: Local health workers, clinics, and community organizations
  - `news_sources_scientists`: Scientists, doctors, and health experts
  - `news_sources_world_health_organization`: World Health Organization (WHO)
  - `news_sources_government_health_authorities`: Government health authorities or other officials
  - `news_sources_politicians`: Politicians
  - `news_sources_journalists`: Journalists
  - `news_sources_ordinary_people_i_know_personally`: Ordinary people I know personally

- news\_sources\_ordinary\_people\_i\_dont\_know\_personally: Ordinary people I don't know personally
- Information Medium Trust: How much do you trust each of the following as a source of COVID-19 news and information? [Do not trust, Somewhat trust, Trust]
  - news\_medium\_trust\_online\_sources: Online sources (websites, apps, social media)
  - news\_medium\_trust\_messaging\_apps: Messaging apps / SMS/ text messaging
  - news\_medium\_trust\_newspapers: Newspapers
  - news\_medium\_trust\_television: Television
  - news\_medium\_trust\_radio: Radio
- Information Sources Trust: How much do you trust each of the following as a source of COVID-19 news and information? [Do not trust, Somewhat trust, Trust]
  - news\_sources\_trust\_local\_health\_workers: Local health workers, clinics, and community organizations
  - news\_sources\_trust\_scientists: Scientists, doctors, and health experts
  - news\_sources\_trust\_world\_health\_organization: World Health Organization (WHO)
  - news\_sources\_trust\_government\_health\_authorities: Government health authorities or other officials
  - news\_sources\_trust\_politicians: Politicians
  - news\_sources\_trust\_journalists: Journalists
  - news\_sources\_trust\_ordinary\_people\_i\_know\_personally: Ordinary people I know personally
  - news\_sources\_trust\_ordinary\_people\_i\_do\_not\_know\_personally: Ordinary people I don't know personally
- Information Demand: Which of the following aspects of COVID-19 do you have the most questions about? [1=option selected, 0=option shown but not selected, and another option was selected. Note: if all values have -1 or -2, this question was viewed and the options were presented, but no options were selected.]
  - information\_demand\_the\_cause\_of\_the\_disease: The cause of the disease
  - Information\_demand\_symptoms\_and\_risk\_factors: Symptoms and risk factors
  - information\_demand\_treatment\_of\_the\_disease: Treatment of the disease
  - information\_demand\_how\_i\_can\_protect\_myself: How I can protect myself
  - information\_demand\_immunity: Immunity
  - information\_demand\_vaccine: Scientific progress in development of a vaccine or treatment
  - information\_demand\_how\_other\_people\_are\_coping: How other people are coping

- information\_demand\_caring\_for\_those\_most\_at\_risk\_of\_covid: **Caring for those most at risk of COVID-19**
- information\_demand\_education: **How I can best take care of my children's school education**
- information\_demand\_differences\_between\_covid\_and\_other\_diseases: **Differences between COVID-19 and other diseases (e.g. flu)**
- information\_demand\_the\_evolution\_of\_the\_pandemic\_in: **The evolution of the pandemic in [country selected]**
- information\_demand\_the\_evolution\_of\_the\_pandemic\_globally: **The evolution of the pandemic globally**
- information\_demand\_the\_economic\_impact\_of\_covid\_to\_me\_personally: **The economic impact of COVID-19 to me personally**
- information\_demand\_the\_economic\_impact\_of\_covid\_in: **The economic impact of COVID-19 in [country selected]**
- information\_demand\_how\_to\_maintain\_my\_mental\_health\_during\_the\_isolation: **How to maintain my mental health during the isolation**
- information\_demand\_how\_to\_maintain\_my\_social\_contact\_despite\_the\_physical\_distancing: **How to maintain my social contact despite the physical distancing**
- information\_demand\_other\_protection\_measures\_by\_the\_government\_and\_communities: **Other protection measures by the government and communities**

## Basic Knowledge

- Knowledge Risk Groups: Which of the following types of people are at the highest risk of severe illness from COVID-19? Select all that apply. [1=option selected, 0=option shown but not selected, and another option was selected. Note: if all values have -1 or -2, this question was viewed and the options were presented, but no options were selected.]
  - knowledge\_high\_risk\_groups\_people\_of\_certain\_religions: **People of certain religions**
  - knowledge\_high\_risk\_groups\_people\_of\_certain\_ages: **People of certain ages**
  - knowledge\_high\_risk\_groups\_people\_with\_certain\_medical\_conditions: **People with certain medical conditions**
  - knowledge\_high\_risk\_groups\_people\_with\_certain\_ethnic\_backgrounds: **People with certain ethnic backgrounds**
  - knowledge\_high\_risk\_groups\_none\_of\_the\_above: **None of the above**
- knowledge\_spread: Which of the following best describes how COVID-19 spreads? (string) [Human contact, coughing or sneezing, Other, Dirt or pollution, Exposure to animals, Mosquito bites]
- Knowledge Symptoms: Which of the following can be symptoms of COVID-19? Please select as many as apply. [1=option selected, 0=option shown but not selected, and

another option was selected. Note: if all values have -1 or -2, this question was viewed and the options were presented, but no options were selected.]

- knowledge\_symptoms\_fever: **Fever**
- knowledge\_symptoms\_cough: **Cough**
- knowledge\_symptoms\_shortness\_of\_breath: **Shortness of breath**
- knowledge\_symptoms\_sore\_throat: **Sore throat**
- knowledge\_symptoms\_runny\_or\_stuffy\_nose: **Runny or stuffy nose**
- knowledge\_symptoms\_muscle\_or\_body\_aches: **Muscle or body aches**
- knowledge\_symptoms\_headaches: **Headaches**
- knowledge\_symptoms\_fatigue: **Fatigue (tiredness)**
- knowledge\_symptoms\_diarrhea: **Diarrhea**
- knowledge\_symptoms\_loss\_of\_taste\_and\_smell: **Loss of taste and smell**
- knowledge\_symptoms\_none\_of\_these: **None of these**

## Distancing Familiarity, Importance & Norms

- distancing\_familiarity (string): Which of the following best describes your familiarity with the term “physical distancing” during the COVID-19 pandemic? [I have not heard of it, I have heard of it but do not know what it means, I have heard of it and have some idea of what it means, I have heard of it and know what it means]
- distancing\_importance (string): How important do you think physical distancing is for slowing the spread of COVID-19? [Not at all important, Slightly important, Moderately important, Very important, Extremely Important]
  - This question was only shown to respondents with  
distancing\_familiarity=I have heard of it and have some idea of what it means or distancing\_familiarity=I have heard of it and know what it means
- Distancing Norms: Out of 100 people in your community, how many do you think do the following when they go out in public?
  - distancing\_norms\_maintain\_a\_distance\_of\_at\_least\_1\_meter\_from\_others: **Maintain a distance of at least 1 meter from others**
  - distancing\_norms\_wear\_a\_face\_mask\_or\_covering: **Wear a face mask or covering**
- norms\_vaccine: Out of 100 people in your community, how many do you think would take a COVID-19 vaccine if it were made available?

## Risk Perceptions and Locus of Control

- risk\_community (string): How dangerous do you think the COVID-19 risk is to your community? [Not at all dangerous, Slightly dangerous, Moderately dangerous, Very dangerous, Extremely dangerous]

- `risk_infection` (string): How likely is it that someone of the same age as you in your community becomes sick from COVID-19? [Not at all likely, Slightly likely, Moderately likely, Very likely, Extremely likely]
- `control_infection` (string): Do you agree with this statement? "I have control over whether I will get COVID-19." [Strongly disagree, Somewhat disagree, Neither agree or disagree, Somewhat agree, Strongly agree]
- `infection_severity` (string): How serious would it be if you became infected with COVID-19? [Not at all serious, Somewhat serious, Very serious]

## Prevention Behaviors in Practice

- `prevention_distancing` (string): How often are you able to stay at least 1 meter away from people not in your household? [Never, Rarely, Sometimes, Often, Always]
- `prevention_hand_washing` (string): When you clean your hands, how often are you able to clean your hands with soap or alcohol-based handrub? [Never, Rarely, Sometimes, Often, Always]
- `prevention_mask` (string): How often are you able to wear a mask or face covering when you are in public? [Never, Rarely, Sometimes, Often, Always] This question was only added starting in wave 3 on 8/4/2020 at 2:15PM ET.

## Behavioral Measures Taken, Efficacy & Norms

- **Preventative Measures Taken:** What measures have you taken to prevent infection from COVID-19 in the past week? [1=option selected, 0=option shown but not selected, and another option was selected. Note: if all values have -1 or -2, this question was viewed and the options were presented, but no options were selected.]
  - `measures_taken_washing_hands`: Washing hands regularly using disinfectants or soap and water
  - `measures_taken_cover_coughs`: Covering mouth and nose when coughing or sneezing
  - `measures_taken_avoid_sick`: Avoiding close contact with anyone who has a fever and cough
  - `measures_taken_wearing_a_face_mask_or_covering`: Wearing a face mask or covering
  - `measures_taken_meter_distance`: Staying at least 1 meter away from other people when out in public
  - `measures_taken_avoid_touching_face`: Avoiding touching your eyes, nose, and mouth with unwashed hands
  - `measures_taken_taking_herbal_supplements`: Taking herbal supplements
  - `measures_taken_using_homeopathic_remedies`: Using homeopathic remedies
  - `measures_taken_caution_opening_mail`: Using caution when opening letters and packages

- `measures_taken_getting_the_flu_vaccine`: **Getting the flu vaccine**
- `measures_taken_eating_garlic`: **Eating garlic, ginger, or lemon**
- `measures_taken_cleaning_or_disinfecting_surfaces`: **Cleaning or disinfecting surfaces**
- `measures_taken_using_antibiotics`: **Using antibiotics**
- `measures_taken_cleaning_or_disinfecting_a_mobile_phone`: **Cleaning or disinfecting a mobile phone**
- `measures_taken_isolation`: **Self-isolation**
- `measures_taken_none_of_these`: **None of these**
- `effect_hand_washing` (string): **How effective is handwashing for preventing the spread of COVID-19? [Extremely effective, Very effective, Moderately effective, Slightly effective, Not effective at all]**
- `effect_mask` (string): **How effective is wearing a face mask for preventing the spread of COVID-19? [Extremely effective, Very effective, Moderately effective, Slightly effective, Not effective at all]**
- `country_management` (string): **How well is COVID-19 being handled in [country selected]? [Extremely well, Very well, Moderately well, Slightly well, Not well at all]**
- `community_management` (string): **How well is your community handling COVID-19? [Extremely well, Very well, Moderately well, Slightly well, Not well at all]**
- `community_action_importance` (string): **How important is it for you to take actions to prevent the spread of COVID-19 in your community? [Extremely important, Very important, Moderately important, Slightly important, Not important at all]**
- `community_action_norms` (string): **How important do other people in your community think it is to take actions to prevent the spread of COVID-19? [Extremely important, Very important, Moderately important, Slightly important, Not important at all]**
- **Community Norms: Out of 100 people in your community, how many do you think believe the following because of COVID-19? (Default slider at 50)**
  - `community_norms_social_gatherings_should_be_cancelled`: **Social gatherings should be cancelled**
  - `community_norms_maintain_1_meter`: **People should stay at least 1 meter away from other people when out in public**
  - `community_norms_mask`: **People should wear a face mask or covering when out in public**
  - `community_norms_close_retail`: **Non-essential retail shops should be closed**
  - `community_norms_a_general_curfew_should_be_enforced`: **A general curfew should be enforced**
  - `community_norms_travel_between_countries_should_be_restricted`: **Travel between countries should be restricted**
  - `community_norms_travel_within_the_country_should_be_restricted`: **Travel within the country should be restricted**

## Employment and Work

- `employed_2020` (string): Have you been working at some point in 2020? [Yes, No]
- `work_changes` (string): How has your work changed since January 31, 2020? [No longer employed, Newly employed, Employed in a different business, Role substantially changed with same business, Little change]
- `work_type` (string): Which best way to describe the work you do most of the time to make money? [I work for my own business, I work in a business that is run by my household or family member, I work in a business that is run by someone else, I work for the government, Other]
- `work_industry` (string): What is the main activity of the business or organization in which you were working before February 2020? [Agriculture, Buying and selling, Construction, Education, Electricity/water/gas/waste, Financial/insurance/real estate services, Health, Manufacturing, Mining, Personal services, Professional/scientific/technical activities, Public administration, Tourism, Transportation, Other]

## Intentions to Visit Locations

- **Open Locations:** Which of the following businesses, locations, or events are currently open and operating near you? [1=option selected, 0=option shown but not selected, and another option was selected. Note: if all values have -1 or -2, this question was viewed and the options were presented, but no options were selected.]
  - `locations_open_restaurants`: Restaurants
  - `locations_open_parks_and_beaches`: Parks and beaches
  - `locations_open_retail_shops`: Retail shops / shopping malls
  - `locations_open_schools`: Schools
  - `locations_open_performances_and_sporting_events`: Performances and sporting events
  - `locations_open_places_of_employment`: Places of employment
  - `locations_open_places_of_worship`: Places of worship
  - `locations_open_health_care_facilities`: Health care facilities
- **Locations Would Attend:** Which of the following businesses, locations, or events would you visit or attend in the coming two weeks if they were operating at full capacity? [1=option selected, 0=option shown but not selected, and another option was selected. Note: if all values have -1 or -2, this question was viewed and the options were presented, but no options were selected.]
  - `locations_would_attend_restaurants`: Restaurants
  - `locations_would_attend_parks_and_beaches`: Parks and beaches
  - `locations_would_attend_retail_shops`: Retail shops / shopping malls
  - `locations_would_attend_schools`: Schools
  - `locations_would_attend_performances_and_sporting_events`: Performances and sporting events

- `locations_would_attend_places_of_employment`: Places of employment
- `locations_would_attend_places_of_worship`: Places of worship
- `locations_would_attend_health_care_facilities`: Health care facilities
- Preventative Measures: Compared to a [location] with no precautions at all, are you more or less likely to visit a restaurant with the following precautions? [More likely to visit, Doesn't affect my actions, Less likely to visit]
  - Asked for the following locations: restaurant, healthcare facility, place of worship, and retail shop (location could be restaurants, worship, health or retail)
  - `prevent_measures_[location]_operating_at_limited_capacity`: Operating at limited capacity
  - `prevent_measures_[location]_everyone_has_to_wear_a_mask`: Everyone has to wear a mask
  - `prevent_measures_[location]_everyone_has_to_pass_a_temperature_check`: Everyone has to pass a temperature check
  - `prevent_measures_[location]_has_additional_hand_washing_stations`: Has additional hand washing stations

## Travel

- `travel_restrictions` (string): Have you refrained from traveling between cities because of restrictions on movement due to COVID-19? [Yes, No — I planned to travel and kept plans, No — I did not plan to travel]
- `travel_concerns` (string): Have you refrained from traveling between cities due to fears regarding COVID-19? [Yes, No — I planned to travel and kept plans, No — I did not plan to travel]

## Future Actions

- `future_masks` (string): Over the next two weeks, how likely are you to wear a mask when in public? [Always, Almost always, When convenient, Rarely, Never]
- `future_distancing` (string): Over the next two weeks, how likely are you to maintain a distance of at least 1 meter from others when in public? [Always, Almost always, When convenient, Rarely, Never]
- `future_vaccine` (string): If a vaccine against COVID-19 infection is available in the market, would you take it? [Yes, definitely; Probably; Unsure; Probably not; No, definitely not]
- Would you be more or less likely to take a vaccine against COVID-19 infection if it were made available and recommended to you by each of the following [More likely, Less likely, No impact]
  - `future_vaccine_friends_and_family` (string): Friends and family
  - `future_vaccine_local_health_workers` (string): Local health workers

- future\_vaccine\_world\_health\_organization (string): WHO
- future\_vaccine\_government\_health\_officials (string): Government health officials
- future\_vaccine\_politicians (string): Politicians

## Appendix B: Survey Change Log

- 2020-07-12: Changed the intentions to visit with precautions randomization to randomly sample two locations rather than one
- 2020-07-25: Added Taiwan as a country option
- 2020-07-27: Added Republic of Northern Macedonia as a country option
- 2020-07-28: Shortened allowed survey response time to record partial completes 4 hours after latest activity from 48 hours
- 2020-08-04: Added an additional mask question (How often are you able to wear a mask or face covering when you are in public?)
- 2020-09-14: Added race question for the US (Not yet released to partners given as we work on preserving privacy)
- 2020-09-14: Changed randomization so that the BEN and Prevention Behaviors in Practice blocks are sampled for all users. Added one more block to be sampled so other blocks are sampled at the same rate.
- 2020-10-28: Survey V2 launched
  - Added future behavior block
  - Added vaccine norm question
  - Added Flu question
  - Removed healthcare access question (replaced w/ flu)
  - Removed past travel block
  - Removed distancing familiarity question
  - Removed intentions to visit if open question
  - Removed locations open questions

## Appendix C: Survey Instrument