Measuring Worldwide COVID-19 Attitudes and Beliefs

Thiemo Fetzer, Marc Witte, Lukas Hensel, Jon M. Jachimowicz*, Johannes Haushofer, Andriy Ivchenko, Stefano Caria, Elena Reutskaja, Christopher Roth, Stefano Fiorin, Margarita Gomez, Gordon Kraft-Todd, Friedrich M. Goetz, Erez Yoeli

* To whom correspondence should be addressed; E-mail: jjachimowicz@hbs.edu

This paper introduces data from a large-scale survey covering respondents from more than 170 countries. The data was collected via snowball sampling starting March 20, 2020 through a survey instrument that was translated by volunteers into 69 languages. Data collection is ongoing and may be complemented with representative samples in the future. This document outlines the survey instrument, provides details over the variables and their descriptions, and illustrates how it can be used for analysis.

1 Introduction

We recruited volunteers via social media and personal contacts of the research team to translate the survey into 69 languages. The research team used online translation tools to conduct basic checks of the translations. The link to the survey went live at 8 PM CET / 5pm EST on March 20, 2020, with a call to participate being launched via social media (see Figure 1). That launch produced nearly 1.4 million impressions on Twitter, and was retweeted 4,892 times. The initial tweets account for 31,211 clicks on the landing page (`http://www.covid19-survey.org`). This is a minority of all clicks on the landing page, indicating that there was a substantial

multiplication due to the link to the survey being broadly shared on social media, via email and through other channels. Several media outlets featured links to the landing page and the call to participate. The landing page was accessed by 391,476 different users from March 20 - April 8, 2020. Out of those visitors, 157,264 started the survey, while 112,136 survey completes were registered.

2 Data Dictionary

The survey instrument is provided in the Appendix. This document provides a description of the variables that can be found in the data file on our OSF page (`https://osf.io/3sn2k/`). We also indicate below which variables were derived measures and how they have been computed. Derived variables are created in the accompanying do-file and not part of the main dataset.

2.1 Meta-information

The survey collection software Qualtrics provides the following meta information:

`StartDate` – for confidentiality reasons, we do not provide the exact start time but only the date on which a participant took the survey. Dates are coded based on the time of interview recorded in GMT time zone.

`ResponseID` – Unique ID for an individual survey attempt.

`Duration` – This variable measures the number of seconds the respondent took to fill out the survey.

`UserLanguage` – This variable captures the ISO language code for the language in which the survey was taken.

`CountryofLiving` – This variable indicates the country in which a respondent lives.

`iso2c` – ISO 3166 alpha-2 country code

`year` – year that survey was taken

`month` – month that survey was taken

`day` – day that survey was taken

`region` – region

`continent` – continent

2.2 Past Behaviors

To what extent do the following statements describe your behavior for the past week? [0 = Does not apply at all; 100 = Applies very much]

- `beh_stayhome` I stayed at home.
- `beh_socgathering` I did not attend social gatherings.
- `beh_distance` I kept a distance of at least two meters to other people.
- `beh_tellsymp` If I had exhibited symptoms of sickness, I would have immediately informed the people around me.
- `beh_handwash` I washed my hands more frequently than the month before.

As a derived variable we generate the `beh_index`, which is the sum of the above five `beh_*` dummies pertaining to protective behaviors. This variable is then standardized.

2.3 Future Behaviors

- `leavehome` Do you need to leave your home in the next 5 days? [0 = No; 1 = Yes]
- What are the reasons for you to leave your home (check all that apply)? Please try to be as honest as possible. Your answers will be kept confidential. [0 = did not check; 1 = checked]

- leavehome_reason_work Going to work
- leavehome_reason_pet Walking a pet
- leavehome_reason_physical Doing physical activity (e.g. exercising, jogging)
- leavehome_reason_food Procuring food for yourself or family
- leavehome_reason_pharmacy Going to the pharmacy
- leavehome_reason_hospital Going to the hospital / receiving medical treatments
- leavehome_reason_care Taking care of dependents
- leavehome_reason_friends Meeting friends or relatives
- leavehome_reason_tired Getting tired of being inside of the house
- leavehome_reason_bored Getting bored
- leavehome_reason_adrenaline Getting some adrenaline (from breaking the law)
- leavehome_reason_freedom Exercising my freedom
- leavehome_reason_other Other

As derived variables, we distinguish "good" and "bad" reasons for leaving the house:

- leavehome_bad This variable is the sum of all "bad" reasons to leave the home, which is leaving the house to: visit friends, because one is bored, for the adrenaline, and to exercise one's freedom.
- leavehome_good This variable is the sum of all "good" reasons to leave the home, which is leaving the house to: go to work, walk a pet, get exercise, procure food, go to the pharmacy, the hospital, or to care for somebody else.

2.4 Personal attitudes about coronavirus measures

- `fob_social` - What do you think: should people in your country cancel their participation at social gatherings because of the coronavirus right now? [No = 0; Yes = 1]

- `fob_handshake` - What do you think: should people in your country not shake other people's hands because of the coronavirus right now? [No = 0; Yes = 1]

- `fob_stores` - What do you think: should all shops in your country other than particularly important ones, such as supermarkets, pharmacies, post offices, and gas stations, be closed because of the coronavirus right now? [No = 0; Yes = 1]

- `fob_curfew` - What do you think: should there be a general curfew in your country (with the exception of grocery shopping, necessary family trips, and the commute to work) because of the coronavirus right now? [No = 0; Yes = 1]

We refer to these variables as "first-order beliefs". As a derived variable, we construct the weighted average country-level beliefs about coronavirus attitudes (first-order beliefs), meaning that we generate the weighted average of respondents per country agreeing with each of the beliefs above. We use within-country weights (`weight`) to account for selection on age, gender, education, and income (see Section 3 below for further details on weight construction).

- `m_fob_social` - the country-level average of `fob_social` [min 0; max 100]

- `m_fob_handshake` - the country-level average of `fob_handshake` [min 0; max 100]

- `m_fob_stores` - the country-level average of `fob_stores` [min 0; max 100]

- `m_fob_curfew` - the country-level average of `fob_curfew` [min 0; max 100]

2.5 Perception of others' beliefs about coronavirus measures

- `sob_social` - How many of 100 people in your country do you think believe that participation at social gatherings should be cancelled because of the coronavirus right now? [slider ranging from 0 to 100 — initiated at 0]

- `sob_handshake` - How many of 100 people in your country do you think believe that one should not shake other people's hands because of the coronavirus right now? [slider ranging from 0 to 100 — initiated at 0]

- `sob_stores` - How many of 100 people in your country do you think believe that all shops in your country other than particularly important ones, such as supermarkets, pharmacies, post offices, and gas stations, should be closed because of the coronavirus right now? [slider ranging from 0 to 100 — initiated at 0]

- `sob_curfew` - How many of 100 people in your country do you think believe there should be a general curfew in your country (with the exception of grocery shopping, necessary family trips, and the commute to work) because of the coronavirus right now? [slider ranging from 0 to 100 — initiated at 0]

We refer to these variables as "second-order beliefs".

Misperceptions: Differences between first- and second-order attitudes As derived variables, we construct individual-level misperceptions. We construct our misperception measures by taking the difference between people's beliefs about others' attitudes, minus the weighted average of actual attitudes of other respondents ($diff\_ind\_X$ = `m_fob_X` − `m_fob_X`). We do this for four dimensions: (i) cancellation of social gatherings, (ii) appropriateness of handshakes, (iii) store closures, (iv) a general curfew. The resulting misperception variables are:

- diff_ind_social - Difference between first- and second order attitudes regarding social gatherings. [min -100; max 100]
- diff_ind_handshake - Difference between first- and second order attitudes regarding hand shaking. [min -100; max 100]
- diff_ind_stores - Difference between first- and second order attitudes regarding store closures. [min -100; max 100]
- diff_ind_curfew - Difference between first- and second order attitudes regarding curfew. [min -100; max 100]

We use the z-score transformation of the sum of all four misperception items to construct our misperception index (misperception_index). Higher values indicate higher levels of misperceptions.

2.6 Financial sanctioning of risky behaviors

- financialpunishment - What do you think: should risky behaviors, which might enable further spread of the coronavirus, be financially punished? [0 = No; 1 = Yes]
- Which fines should be enforced for the following risky behaviors (amount in your country currency)?
  - financialpunish_1: Participation at social gatherings (amount in country currency) [value as entered by respondent, numerical values only]
  - financialpunish_2: Going out despite exhibiting symptoms of coronavirus [value as entered by respondent, , numerical values only]

2.7 Case predictions

- `infect_now` - Without looking it up, what is your estimate of the number of people in your country who are currently infected? [value as entered by respondent]

- `infect_onemonth` - How many people in your country do you think will be infected 1 month from now? [value as entered by respondent]

We also use the log of case growth as a derived variable:

```
log_case_growth = log((infect_onemonth-infect_now)/infect_now)
```

2.8 Perceptions of government/public response & efficacy

- `perceivedreaction` - Do you think the reaction of your country's government to the current coronavirus outbreak is appropriate, too extreme, or not sufficient? [5-point scale; 1 = The reaction is much too extreme; 2 = The reaction is somewhat too extreme; 3 = The reaction is appropriate; 4 = The reaction is somewhat insufficient; 5 = The reaction is not at all sufficient]

- `govtrust` - How much do you trust your country's government to take care of its citizens? [5-point scale; 1 = Strongly distrust; 2 = Somewhat distrust; 3 = Neither trust nor distrust; 4 = Somewhat trust; 5 = Strongly trust]

- `govfact` - How factually truthful do you think your country's government has been about the coronavirus outbreak? [5-point scale; 1 = Very untruthful; 2 = Somewhat untruthfu; 3 = Neither truthful nor untruthful; 4 = Somewhat truthful; 5 = Very truthful]

- `react_pub_appr` - Do you think the reaction of your country's public is appropriate, too extreme, or not sufficient? [5-point scale; 1 = The reaction is much too extreme; 2 =

The reaction is somewhat too extreme; 3 = The reaction is appropriate; 4 = The reaction is somewhat insufficient; 5 = The reaction is not at all sufficient]

- `perceivedeffectivnes` - What do you think: How effective are social distancing measures (e.g., through a general curfew) to slow down the spread of the coronavirus? [5-point scale; 1 = Not at all effective; 2 = Not effective; 3 = Neither effective nor ineffective; 4 = Effective; 5 = Very effective]

For the analysis, we generate dummies of all measures of the perceptions of government and public response and efficacy described above. Those equal 1 if the individual's response is above the midpoint of the Likert scale, and 0 if it is below the midpoint. When the data is collapsed at the country-level, it can then be interpreted as the share of respondents who gave a response above the midpoint.

2.9 Worries battery

- `mh_anxiety_1` - I am nervous when I think about current circumstances. [5 point scale; same for all items in this battery; 1 = Does not apply at all; 2 = Somewhat does not apply; 3 = Neither applies nor does not apply; 4 = Somewhat applies; 5 = Strongly applies]
- `mh_anxiety_2` - I am calm and relaxed. [reverse coded]
- `mh_anxiety_3` - I am worried about my health.
- `mh_anxiety_4` - I am worried about the health of my family members.
- `mh_anxiety_5` - I am stressed about leaving my house.

As a derived variable, we generate the "worries index" (`mh_index`), which is the z-scored sum of the 5 worries questions questions above. Higher values indicate higher levels of worries.

2.10 Depression questionnaire (PHQ9)

The data includes responses eight questions from the commonly used PHQ9 depression questionnaire (Kroenke et al., 2001) (without the suicide question). Higher values indicate higher levels of depression.

How often have you been bothered by the following over the past 2 weeks?

- `PHQ9_1` - Little interest or pleasure in doing things? [4 point scale; same for all items in this battery; 1 = Not at all; 2 = Several days; 3 = More than half the days; 4 = Nearly every day]
- `PHQ9_2` - Feeling down, depressed, or hopeless?
- `PHQ9_3` - Trouble falling or staying asleep, or sleeping too much?
- `PHQ9_4` - Feeling tired or having little energy
- `PHQ9_5` - Poor appetite or overeating?
- `PHQ9_6` - Feeling bad about yourself — or that you are a failure or have let yourself or your family down?
- `PHQ9_7` - Trouble concentrating on things, such as reading the newspaper or watching television?
- `PHQ9_8` - Moving or speaking so slowly that other people could have noticed? Or so fidgety or restless that you have been moving a lot more than usual?

Derived variable: The "depression index" (`phq9_index`) is the z-scored sum of the 8 PHQ9 questions above.

2.11 Personality Battery

The data also includes a ten item version of the Big-Five personality questionnaire (Gosling et al., 2003).

To which extent do the following questions apply to you? I see myself as . . .

- `personality_b5_1` - Extroverted, enthusiastic [7 point scale; same for all items in this battery; 1 = Disagree strongly; 2 = Disagree moderately; 3 = Disagree a little; 4 = Neither agree nor disagree; 5 = Agree a little; 6 = Agree moderately; 7 = Agree strongly]
- `personality_b5_2` - Critical, quarrelsome
- `personality_b5_3` - Dependable, self-disciplined
- `personality_b5_4` - Anxious, easily upset
- `personality_b5_5` - Open to new experiences, complex
- `personality_b5_6` - Reserved, quiet
- `personality_b5_7` - Sympathetic, warm
- `personality_b5_8` - Disorganized, careless
- `personality_b5_9` - Calm, emotionally stable
- `personality_b5_10` - Conventional, uncreative

2.12 Personal Information

- `age` - Age in years (2020 - year of birth).
- `educ` - How many years of education did you complete? [numerical value provided by participants]

- `income` - What is your monthly household income, before tax, in your country's currency? [numerical value provided by participants]
- `marital_status` - What is your marital status? [1 = married/co-habitating, 2 = single/divorced]
- `hhmember` - How many people live in your household? [numerical value provided by participants]
- `gender` - Which gender do you identify with? [1 = Male; 2 = Female; 3 = Other]
- `health` - How healthy are you? [1 = poor; 2 = fair; 3 = good; 4= excellent]
- `Comorbidities` - How many of the following conditions do you have: cardiovascular diseases, diabetes, hepatitis B, chronic obstructive pulmonary disease, chronic kidney diseases, and cancer? [values ranging from 0 to 5 or more]

Derived variables:

- `age_yr` - Age in years (2020- year of birth).
- `age_yr_bin` - categorical variables subdividing individuals into five year bin groups.

2.13 Merged external data

2.13.1 John's Hopkins Data

We merge the JHU COVID19 data introduced in Dong et al. (2020). The data are matched to each countries ISO2 letter country code and merged to individual respondents based on the date on which a respondent took the survey. This implies that the variables are varying within-country over time.

The variables are:

- `covid_confirmed` - confirmed cases in country on date respondent participated
- `covid_death` - confirmed COVID19 deaths in country on date respondent participated
- `covid_recovered` - patients recovered from COVID19 in a country on date respondent participated
- `l1covid_confirmed` - confirmed cases in country on one day prior to date that respondent participated
- `l1covid_death` - confirmed COVID19 deaths in country one day prior to date respondent participated
- `l1covid_recovered` - patients recovered from COVID19 in a country one day prior to date respondent participated
- `l2covid_confirmed` - confirmed cases in country two days prior to the day that respondent participated
- `l2covid_death` - confirmed COVID19 deaths in country two days prior to date on which respondent participated
- `l2covid_recovered` - patients recovered from COVID19 in a country two days prior to date on which respondent participated

2.13.2 Oxford COVID-19 Policy tracker

We merged the Oxford COVID-19 Policy tracker introduced in Hale et al. (2020). This data tracks government policy relating to COVID-19 responses. The data is described in detail in Hale et al. (2020). The data, being a country-by-day panel, is merged based on the date at which

a respondent took the survey and thus reflects the policy landscape at the date the person took the interview.

For the purposes of our analysis, we further constructed some derived statistics separately. We leverage the data from Hale et al. (2020) that is up to date as of April 6th. The data is a country-by-daily data set capturing the different measures countries adopted to constrain the spread of COVID-19. We focus on the main subcomponents S1-S6, capturing government actions grouped as: S1 School closures, S2 Workplace closures, S3 Cancellation of public events, S4 Closure of public transportation, S5 Public information, and S6 Restrictions on internal movement.

The data distinguishes between general country-wide restrictions versus targeted ones. In addition, the data also distinguishes between recommendations versus requirements. We construct restrictions indices that are specific to our sample-countries and time window, and do not use the stringency index that is provided by Hale et al. (2020). This is because for the sample period under consideration, from March 20 to April 5, 2020, most countries had already adopted quite stringent measures with regard to international travel and public information campaigns. As a result, these and a few other sub-components add little variation.

Instead, we use the Hale et al. (2020) data to create a set of indicator variables that capture whether a country applied measures in a specific domain S1-S6 that are general, i.e., apply to the country as a whole. Similarly, we also use a robustness measure that constructs an index based on transformed indicator variables capturing whether restrictions in the domain S1-S6 are general and are mandatory. Using these sets of dummy variables, we then construct the first principal component of the data for the set of countries in our estimations and for the time period under considerations.

Table 1 presents the factor loadings for the first three principal components for the measures S1-S6 that are coded as applying countrywide (but are not necessarily mandatory). We observe

that the first principal component is positively loaded with little weight being placed on the Public Information component. This is not surprising as most countries had ongoing public information efforts by March 20th, 2020 (the day our data collection began).

Table 1: Principal components to our construction of a COVID-19 country restriction index: general-country wide (but not necessarily mandatory) measures

	Comp1	Comp2	Comp3	Unexplained
General recommended/mandated School closures	0.407	-0.050	0.889	0.020
General recommended/mandated Workplace closures	0.451	-0.199	-0.388	0.329
General recommended/mandated Public event cancelation	0.425	0.207	-0.126	0.479
General recommended/mandated Public transport closure	0.400	-0.334	-0.159	0.458
General recommended/mandated Public information	0.222	0.892	-0.077	0.094
General recommended/mandated Restrictions internal movement	0.491	-0.086	-0.107	0.357

Table 2 presents the factor loadings for the first three principal components constructed on the dummy variables capturing general and mandated policy changes. Naturally, information campaigns do not have a "mandatory" dimension. As such, this feature, that already added little variation, is dropped. As before, we observe positive loading on all subcomponents.

Table 2: Principal components to construction COVID-19 country restriction index: general and mandatory restrictions

	Comp1	Comp2	Comp3	Unexplained
General mandated School closures	0.400	0.613	0.681	0.000
General mandated Workplace closures	0.476	-0.010	-0.252	0.393
General mandated Public event cancelation	0.436	0.347	-0.565	0.220
General mandated Public transport closure	0.412	-0.672	0.373	0.130
General mandated Restrictions internal movement	0.503	-0.228	-0.118	0.319

We use these two first principal components to study the impact of country-level policy changes on perceptions at the individual level.

3 Weight Construction

This section describes the construction and use of weights included with the survey data. The included weights correct for differences in income, education, age, and gender between survey respondents and the general population in each country. For countries that lack data on one of

the dimensions, the weights correct for the available dimensions. We use data on the population structure from the United Nations statistical agency to construct the weights.[1] To weight by income, we use data from the Gallup World Poll.[2]

Table 3 displays the age bins used for reweighting for each gender to account for the population structure. This definition means that we cannot construct weights for respondents who indicate 'other' as gender. To construct income weights, we use country-level income quintiles. Finally, we use three education categories to construct weights: less than 8 years of education, between nine and 14 years of education, and 15 and more years of education.

Table 3: Age bins used for reweighting

Age bin
18 - 19
20 - 24
25 - 29
30 - 34
35 - 39
40 - 44
45 - 49
50 - 54
55 - 59
60 - 64
65+

These age bins are then used to construct weights based on the frequency of observations in the survey data according to the following formula:

$$j\_weight_{ibc} = \frac{weight\_\mathtt{jb} \cdot N_c}{N_{jb}} \tag{1}$$

where j_weight_{ibc} is the weight for individual i in bin b, for category j (age-gender, income, education), and country c. $weight_{jb}$ is the fraction of the population in bin b of category j. N_{jb} is

[1] The data can be accessed here: *http://data.un.org/Data.aspx?d=POP&f=tableCode%3A22* .

[2] We use the latest available wave of data for each country.

the number of individuals in our survey in bin b and N_c is the number of observations in country c. Intuitively, this formula put more weight on individuals in in bins with few observations and individuals in larger bins.

To construct aggregate individual-level weights, we multiply the weights in different categories.[3]

$$weight_{ijbc} = \Pi_{j \in (educ,inc,ag)} j\_weight_{ibc} \tag{2}$$

We also construct weights that account for the differential sample size across countries (weighting all countries equally) by dividing the weights by the number of observations in our sample.

$$weight\_sample_{ijbc} = weight_{ijbc}/N_c \tag{3}$$

3.1 Included weight variables

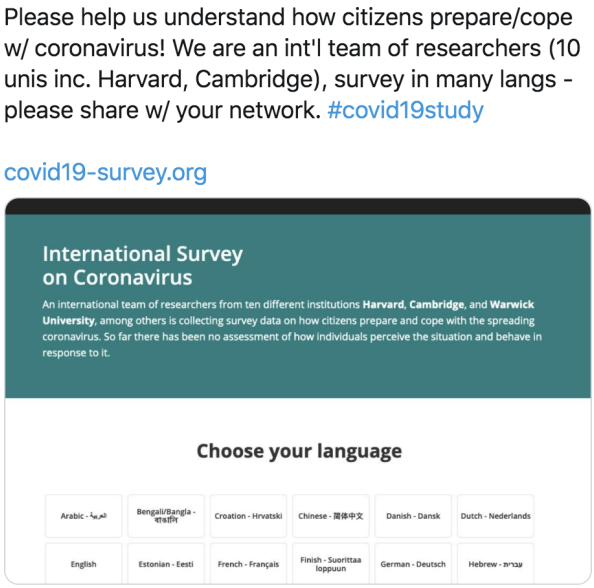
- `weight_sample` - weights to reweight individuals within country weighting all countries equally.
- `weight` - weight to reweight individuals within country.
- `ag_weight` - age-gender weights based on UN population data.
- `no_ag_weights` - dummy indicating missing age-gender weights.
- `educ_weight` - education weights based on GALLUP World Poll education data.
- `no_educ_weights` - dummy indicating missing education weights.
- `inc_weight` - income weights based on GALLUP World Poll income data.

[3]This implicitly assumes independence of distributions of the different categories.

- `no_inc_weights` - dummy indicating missing income weights.
- `N_country` - number of respondents in country of respondent.

Figures

Figure 1: Survey Launch Tweet



Notes: Survey launch announcement went live on March 20, 2020. The initial tweet generated around 1.5 million impressions.

References

Dong, E., H. Du, and L. Gardner (2020). An interactive web-based dashboard to track covid-19 in real time. *The Lancet Infectious Diseases*.

Gosling, S. D., P. J. Rentfrow, and W. B. Swann (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality 37*(6), 504–528.

Hale, T., A. Petherick, T. Phillips, and S. Webster (2020). Variation in government responses to COVID-19.

Kroenke, K., R. L. Spitzer, and J. B. Williams (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine 16*(9), 606–613.