

IDB/Cornell Coronavirus Survey:

Methodological Notes

Nicolas Bottan
Bridget Hoffmann
Diego A. Vera-Cossio

Department of Research and
Chief Economist

TECHNICAL
NOTE Nº
IDB-TN-1936

IDB/Cornell Coronavirus Survey:

Methodological Notes

Nicolas Bottan*

Bridget Hoffmann**

Diego A. Vera-Cossio**

* Cornell University

** Inter-American Development Bank

June 2020

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Bottan, Nicolas.

IDB/Cornell coronavirus survey: methodological notes / Nicolas Bottan, Bridget Hoffmann, Diego A. Vera-Cossio.

p. cm. — (IDB Technical Note ; 1936)

1. Coronavirus infections-Social aspects-Latin America. 2. Coronavirus infections-Social aspects-Caribbean Area. 3. Household surveys-Latin America. 4. Household surveys-Caribbean Area. I. Hoffmann, Bridget. II. Vera-Cossio, Diego A. III. Inter-American Development Bank. Department of Research and Chief Economist. IV. Title. V. Series.

IDB-TN-1936

<http://www.iadb.org>

Copyright © 2020 Inter-American Development Bank. This work is licensed under a Creative Commons IGO 3.0 Attribution-NonCommercial-NoDerivatives (CC-IGO BY-NC-ND 3.0 IGO) license (<http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) and may be reproduced with attribution to the IDB and for any non-commercial purpose. No derivative work is allowed.

Any dispute related to the use of the works of the IDB that cannot be settled amicably shall be submitted to arbitration pursuant to the UNCITRAL rules. The use of the IDB's name for any purpose other than for attribution, and the use of IDB's logo shall be subject to a separate written license agreement between the IDB and the user and is not authorized as part of this CC-IGO license.

Note that link provided above includes additional terms and conditions of the license.

The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.



Overview

The IDB/Cornell Coronavirus Survey collects information on labor market impacts, food security, social distancing, and knowledge regarding the Covid-19 pandemic through an online survey. The primary modules of the questionnaire were standardized across countries to allow data pooling and cross-country comparisons. The survey was launched in the first country on March 27, 2020, and the majority of responses were collected during April 2020. Respondents were recruited through social media using paid social-media advertisement campaigns. The dataset is entirely composed of responses to an online survey, and therefore it is unable to capture information regarding the most-vulnerable people, who may not have access to internet or social media. However, the survey was still able to collect information from different socioeconomic backgrounds and subnational regions. Thus, the data are useful for analyzing patterns across different socioeconomic groups and for making cross-country comparisons. The dataset includes weights to increase the within-country representativeness of the data, as well as weights that correct for differences in the response rate across countries and provide a larger weight to observations from countries with larger populations.

Keywords: Online Survey, Coronavirus Pandemic, Public Health, Labor Markets, Inequality

JEL codes: D31, H55, I14, I32, I38, J01

Acknowledgements

We would like to thank Sebastian Espinoza and Maria Paula Medina for superb research assistance. We would like to thank Julián Cristia for his encouragement and advice with the project. We also want to thank Tom Sarrazin, Sebastián Oliva and Pablo Bachelet for their tremendous support in the dissemination of the survey. We also want to thank the support of IDB's country

offices, country representatives, and country economists for their invaluable input in customizing the questionnaires for each country. In particular, we want to thank Marta Ruiz Aranz and Diether Beuermann for their support in the rollout of the surveys in Central America and the Caribbean. We would like to thank Ana María Ibañez, Eric Parrado, Michael Lovenheim, and the team of researchers in IDB's Research Department for valuable support and feedback in different parts of the project.

Funding: This project was funded by the IDB's Coronavirus research funds (RG-E1700 - Coronavirus survey). The Qualtrics license and IRB approval are from Cornell University.

Data

Data Collection

We conducted online surveys of households in 17 countries in Latin America and the Caribbean. The survey was implemented online using Qualtrics. Our sample is comprised of 230,540 observations across 8 South American countries (Chile, Colombia, Bolivia, Ecuador, Guyana, Peru, Suriname, and Uruguay), 4 North and Central American countries (Costa Rica, El Salvador, Mexico, and Panama), and 5 Caribbean countries (Dominican Republic, Bahamas, Barbados, Jamaica, and Trinidad and Tobago). The majority of the data collection occurred during the second half of April 2020. The survey was launched in the first country on March 27, 2020 and rolled out to all 17 countries by April 17, 2020. The date of launch for each country is displayed in Column (1) of Table 1. For all countries except Costa Rica, data collection continued until April 30, 2020. The number of observations varies by country, ranging from 565 in Suriname to 35,556 in Chile (see Table 1 for country-specific details).

We construct the sample according to the following steps. First, we only include completed surveys in the sample. Approximately 59% of all surveys that are started are completed. Second, we exclude surveys associated with IP addresses outside the borders of the country for which the respondent is completing the survey. Across all countries in the sample, less than 1% of completed surveys are excluded due to the location of the IP address. Third, we exclude surveys that Qualtrics flags as likely to be repeat surveys. Qualtrics flags surveys that are completed on the same device and likely to be repeat surveys completed by the same individual or household based on cookies. This is an imperfect indicator. For example, it will not recognize repeated surveys by the same individual or household that are completed on different devices. Fewer than 2.3% of completed surveys are flagged, and we exclude these surveys from the sample. Column (2) of Table 1 shows the numbers of observations in the sample by country.

Table 1: Date of Launch and Number of Observations by Country

Country	(1) Launch Date	(2) No. of Observations	(3) % of Localities with Observations
Chile	3/27/2020	35,556	97%
Bolivia	4/1/2020	25,970	83%
Panama	4/3/2020	15,521	100%
Uruguay	4/3/2020	21,191	64%
Peru	4/7/2020	25,452	47%
Mexico	4/13/2020	19,483	51%
Costa Rica	4/13/2020	9,151	90%
Colombia	4/15/2020	23,458	71%
El Salvador	4/16/2020	13,456	50%
Dominican Republic	4/16/2020	7,965	94%
Barbados	4/16/2020	2,072	100%
Jamaica	4/16/2020	2,547	91%
Guyana	4/16/2020	1,670	96%
Bahamas	4/16/2020	896	81%
Trinidad and Tobago	4/16/2020	4,683	100%
Ecuador	4/16/2020	18,688	68%
Suriname	4/17/2020	565	100%

Date of launch is the date on which the social media posts began. The date that the survey was rolled out in each country was largely determined by bureaucratic processes and approvals. With the except of Costa Rica, data collection in each country continued until April 30, 2020. The number of observations for each country reflected the number in the sample after data cleaning. The percent of localities is the percent of localities of each country for which we have one or more observations in the sample.

Overall, our sample has broad geographic coverage. Our sample contains observations in 92% of the sub-national regions and in 61% of the localities of countries in our sample (see Table 1). Figure 1 illustrates the broad geographic coverage of our sample, showing the number of observations as a share of population (in %) by sub-national region for each country in our sample.

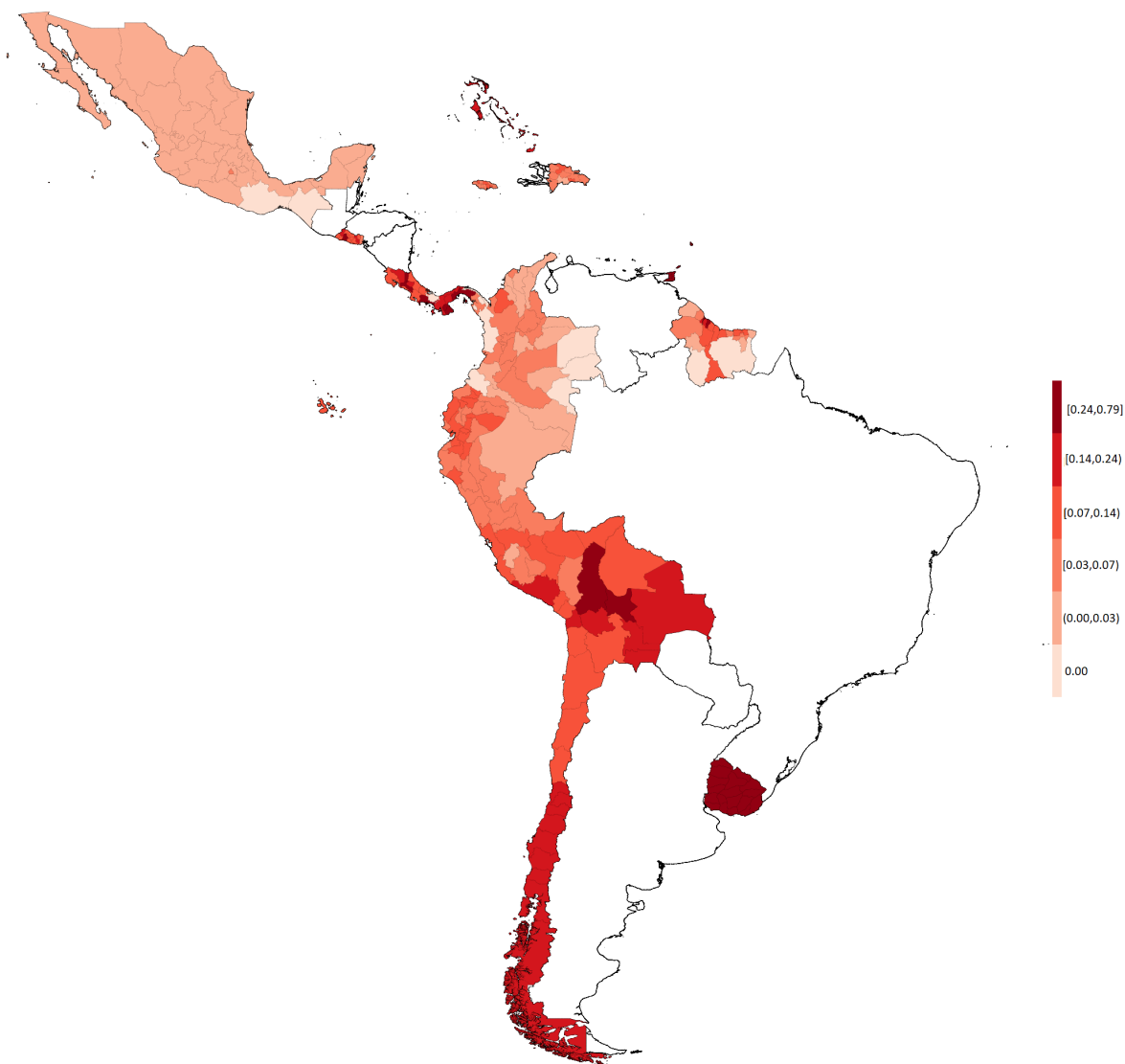


Figure 1: **Geographic Coverage of Sample at the Sub-National Level.** The sub-national regions of each country in the sample are shaded according to number of observations as a share of population (in %). Sources of population data for each country are shown in the Supplementary Material Section.

We recruited participants for the survey using paid advertisements. In each country, we used keywords with broad appeal that were unrelated to Covid-19, such as soccer celebrities, to avoid recruiting respondents based on their knowledge or experience during the pandemic. We targeted the advertisements to ages 19 years old and up. We utilized a second, simultaneous social media campaign with the same images and keywords to specifically target millennials aged 24-35 in countries in which the average age of our sample was high or expected to be high.

The advertisement campaigns consisted of three posts for each country: an image of a male wearing a face mask, an image of a female wearing a face mask, and an image of an iconic landmark in that country. The posts were adapted to each country in three ways. First, the image of the landmark was country-specific. Second, the images of the male and female were changed as appropriate for the context. Third, the language of the text of the posts was changed (Spanish, English, or Dutch). Facebook algorithms select which one of the three posts to show to each Facebook user. The posts that were used for Uruguay are shown below as an example.



Figure 2: **Facebook Posts.** We used three Facebook Posts customized to each country, a male wearing a mask, a female wearing a mask, and a landmark.

Survey Design

The primary objective of the survey was to measure the impacts of the current pandemic on households' economic situation and well-being in Latin America and the Caribbean. Therefore, the questionnaire focused on collecting data in the following categories: labor market outcomes, financial situation, social program enrollment, hunger, shortages of key goods, and agreement with different policies to slow the spread of Covid-19.

The questionnaire was standardized across countries for the creation of a single dataset containing data for all countries in the sample.

Sample questionnaires in English, Spanish, and Dutch can be downloaded here: <https://www.dropbox.com/sh/uuv17cfaz94kw4h/AAAWvgJUNYEBHhNxZPcQVv3Za?dl=0>.

Validation

Estimation of Weights

For each country, we modeled the probability of being in the nationally representative sample as a function of demographic, time-invariant characteristics such as gender, age, and education of the respondent, indicators for the presence of children of 5 years old or younger in the household and elderly (60 years old or older) in the households, as well as the number of household members and the number of children enrolled in school. Finally, we also include region-specific indicators, and indicators for income categories.

We conducted this process country by country. For each country, we used the most recent nationally representative survey available in the Inter-American Development Bank harmonized survey data repository. We estimated the model by fitting a logistic function (logit) and computed predicted probabilities of being in the nationally representative data set ($\hat{p}_{i,c}$) for each respondent i in country c . We then used inverse probability weights ($ipw_{i,c} = 1/(1 - \hat{p}_{i,c})$) to

resemble those from nationally representative surveys, at least in terms of observable characteristics.

We used these propensity scores as seeds for the calibration of weights using the iterative proportional fitting method (raking). To compute the raked weights we calibrated the weights to match the population proportion of respondents based on different variables. We used the marginal distribution of respondents' age categories (< 40 , $40-60$, > 60 years old), the share of female respondents, the share of respondents with college education, and the marginal distribution of income (based on eight categories) as target population moments. After applying the raked weights, $\tilde{w}_{i,c}$, the number of observations in the online survey in each country matched the total population of the country according to the household survey data.

To prevent differences in response rates across from driving the results, we re-scale the within-country weights $\tilde{w}_{i,c}$ by the inverse share of re-weighted number of responses per country, relative to the country's population size ($Population_c / (\sum_i^{N_c} w_{i,c})$) according to the census.

Validation

We validate the online survey data in two ways. First, we compare unweighted and weighted means of demographic characteristics using the online data to means computed using nationally representative household face-to-face surveys.

Table 2 shows that the respondents to the online data are more educated and more likely to be females than the respondents of representative household surveys (see Columns 1 and 3, respectively). Columns 4 and 5 report data from the subset of household-survey observations corresponding to urban areas and households with access to the internet at home, and they show that characteristics of online respondents are similar to those of respondents in nationally representative surveys. Column 2 shows that we were able to reduce the differences in respondent characteristics between the online and household surveys by reweighting the observation from

the online survey using our calibrated weights. In particular, the imbalances in respondents' gender and education are substantially reduced, and we quite closely match the income distribution.

Table 2: Differences between Online Survey Data and Household (field) Survey Data

	Online		Face-to-face Household Survey		
	Unweighted (1)	Weighted (2)	Total (3)	Urban (4)	Internet (5)
Household size	4.38	4.22	3.86	3.87	3.96
Elderly in household (share)	0.35	0.3	0.36	0.35	0.34
Female (share)	0.72	0.54	0.52	0.53	0.53
Education: None	0	0.11	0.11	0.08	0.05
Education: Primary	0.03	0.18	0.31	0.29	0.22
Education: High School	0.23	0.38	0.41	0.4	0.41
Education: University	0.74	0.33	0.16	0.2	0.3
Age (in years)	39	40.47	43.49	42.88	42.09
0-0.5 MW	0.1	0.08	0.07	0.05	0.01
0.5-1 MW	0.16	0.13	0.11	0.09	0.04
1-2 MW	0.21	0.23	0.24	0.23	0.15
2-3 MW	0.14	0.16	0.18	0.19	0.18
3-4 MW	0.11	0.09	0.12	0.13	0.15
4-6 MW	0.1	0.12	0.14	0.16	0.21
6-8 MW	0.06	0.06	0.06	0.07	0.11
8-11 MW	0.05	0.05	0.04	0.05	0.08
11+ MW	0.07	0.06	0.04	0.04	0.07

The table presents means of household and survey respondent demographic characteristics using data from the online survey and nationally representative surveys, and pooling observations from all study countries (each country with equal weight). Column (1) reports raw means using all the observations from the online surveys. Column (2) reports means after reweighting the data from the online survey. Column (3) reports means using all available observations in the household (field) surveys using sampling weights. Columns (4) and (5) report means using household-survey data for the subpopulations of households in urban areas and households with access to the internet at home. MW stands for minimum wage.

Population and Household Survey Data

Year and Source of Household Survey Data by Country

Country	Survey Name	Year	Link to source
Bahamas	LFS	2014	http://www.bahamas.gov.bs/wps/portal/public
Barbados	BSLC	2016	http://sistemasintegrales.cl/project/barbados-survey-of-living-conditions/
Bolivia	ECH	2018	https://www.ine.gob.bo/index.php/herramientas/bases-de-datos-catalogo-anda/bases-de-datos-encuestas-sociales/
Chile	Casen	2017	http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/basedatos.php
Colombia	GEIH	2018	http://microdatos.dane.gov.co/index.php/catalog/659/get_microdata
Costa Rica	ENAHO	2018	https://www.inec.cr/noticias/enaho
Dominican Republic	ENCFT	2018	https://www.bancentral.gov.do/a/d/2539
Ecuador	ENEMDU	2018	https://www.ecuadorencifras.gob.ec/estadisticas/
El Salvador	EHPM	2018	http://www.digestyc.gob.sv/index.php/temas/des/ehpm.html
Guyana	LFS	2018	https://statisticsguyana.gov.gy/data/databases/
Jamaica	SLC	2014	https://statinja.gov.jm/living_conditions_poverty.aspx
Mexico	ENIGH	2018	https://www.inegi.org.mx/programas/enh
Panama	EHPM	2018	https://www.inec.gob.pa/publicaciones/Default2.aspx?ID_CATEGORIA=5&ID_SUBCATEGORIA=38
Peru	ENAHO	2018	https://webinei.inei.gob.pe/anda_inei/index.php/catalog/672
Suriname	SLC	2017	https://statistics-suriname.org/en/
Trinidad & Tobago	CSSP	2015	https://cso.gov.tt/methods/classifications/
Uruguay	ECH	2018	http://www.ine.gub.uy/encuesta-continua-de-hogares1

We used the most recent household survey available at the IDB's harmonized database repository. Each survey can be accessed on the websites of the specific institutions in each country. The table provides links to original sources for data access. In some instances, the microdata must be requested directly from the competent body due to privacy restrictions.

Year and Source of Population Data by Country

Country	(1) Year	(2) Source
Chile	2020	https://www.ine.cl/estadisticas/sociales/demografia-y-vitales/proyecciones-de-poblacion
Bolivia	2020	https://www.ine.gob.bo/subtemas_cuadros/demografia_html/PC20106.htm
Panama	2020	https://www.inec.gob.pa/publicaciones/Default3.aspx?ID_PUBLICACION=499&ID_CATEGORIA=3&ID_SUBCATEGORIA=10
Uruguay	2020	http://www.ine.gub.uy/estimaciones-y-proyecciones
Peru	2020	http://proyectos.inei.gob.pe/web/biblioineipub/bancopub/Est/Lib0846/libro.pdf
Mexico	2020	https://www.gob.mx/cms/uploads/attachment/file/63977/Documento_Metodologico_Proyecciones_Mexico_2010_2050.pdf
Costa Rica	2020	https://www.inec.cr/poblacion/estimaciones-y-proyecciones-de-poblacion
Colombia	2020	https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion
El Salvador	2020	https://www.transparencia.gob.sv/search?utf8=%E2%9C%93&ft=Proyecciones+municipales
Dominican Republic	2020	https://www.one.gob.do/demograficas/proyecciones-de-poblacion
Barbados	2010	https://web.archive.org/web/20170118220332/http://www.barstats.gov.bb/files/documents/PHC_2010_Census_Volume_1.pdf
Jamaica	2018	https://statinja.gov.jm/Demo_SocialStats/PopulationStats.aspx
Guyana	2012	https://statisticsguyana.gov.gy/publications/#elementor-tab-content-1465%20%3E%20ul:nth-child(3)%20%3E%20li:nth-child(1)%20%3E%20span
Bahamas	2010	https://www.bahamas.gov.bs/wps/wcm/connect/22f9b2b0-68fa-4a26-8bd8-474952e42dc2/Population+Projection+Report+2010-2040.pdf?MOD=AJPERES
Trinidad and Tobago	2011	https://cso.gov.tt/census/2011-census-data/
Ecuador	2020	https://sni.gob.ec/proyecciones-y-estudios-demograficos
Suriname	2012	https://statistics-suriname.org/en/census-statistics-2012/

The launch date represents the date at which we launched the social media campaign to recruit participants in that country. All countries except Costa Rica continued data collection until April 30, 2020. The projected population data for Panama exclude Comarca Guna de Madugandi and Comarca Guna de Wargandi. Population data for years 2018 through 2020 are projected.