

Encuesta coronavirus BID/Cornell:

Nota metodológica

Nicolas Bottan
Bridget Hoffmann
Diego A. Vera-Cossio

Departamento de Investigación
y Economista Jefe

NOTA TÉCNICA N°
IDB-TN-1936

Encuesta coronavirus BID/Cornell:

Nota metodológica

Nicolas Bottan*

Bridget Hoffmann**

Diego A. Vera-Cossio**

* Cornell University

** Banco Interamericano de Desarrollo

Junio 2020

Catalogación en la fuente proporcionada por la
Biblioteca Felipe Herrera del
Banco Interamericano de Desarrollo

Bottan, Nicolas.

Encuesta coronavirus BID/Cornell: nota metodológica / Nicolas Bottan, Bridget Hoffmann, Diego A. Vera-Cossio.

p. cm. — (Nota técnica del BID ; 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. Banco Interamericano de Desarrollo. Departamento de Investigación y Economista Jefe. IV. Título. V. Serie.
IDB-TN-1936

<http://www.iadb.org>

Copyright © 2020 Banco Interamericano de Desarrollo. Esta obra se encuentra sujeta a una licencia Creative Commons IGO 3.0 Reconocimiento-NoComercial-SinObrasDerivadas (CC-IGO 3.0 BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) y puede ser reproducida para cualquier uso no-comercial otorgando el reconocimiento respectivo al BID. No se permiten obras derivadas.

Cualquier disputa relacionada con el uso de las obras del BID que no pueda resolverse amistosamente se someterá a arbitraje de conformidad con las reglas de la CNUDMI (UNCITRAL). El uso del nombre del BID para cualquier fin distinto al reconocimiento respectivo y el uso del logotipo del BID, no están autorizados por esta licencia CC-IGO y requieren de un acuerdo de licencia adicional.

Note que el enlace URL incluye términos y condiciones adicionales de esta licencia.

Las opiniones expresadas en esta publicación son de los autores y no necesariamente reflejan el punto de vista del Banco Interamericano de Desarrollo, de su Directorio Ejecutivo ni de los países que representa.



Resumen

La encuesta del BID/Cornell sobre el coronavirus recolecta información relacionada a los impactos en el mercado laboral, la seguridad alimentaria, el distanciamiento social y el conocimiento de la pandemia Covid-19 a través de una encuesta en línea. Los módulos principales del cuestionario se estandarizaron en todos los países para poder agrupar los datos y realizar comparaciones entre países. La encuesta se lanzó en el primer país el 27 de marzo de 2020, y la mayoría de las respuestas se recolectaron durante abril de 2020. Los participantes fueron reclutados mediante redes sociales usando campañas publicitarias pagadas. Los datos se componen en su totalidad por las respuestas a una encuesta en línea y, por lo tanto, no recogen información sobre las personas más vulnerables, quienes probablemente no tienen acceso a internet o a redes sociales. Sin embargo, la encuesta contiene información correspondiente a personas de distintos estratos sociales y regiones sub-nacionales. En este sentido, los datos son útiles para analizar patrones distintos por grupos socioeconómicos dentro de un país y entre países. Los datos incluyen ponderadores para corregir posibles sesgos y lograr mayor representatividad a nivel nacional. Adicionalmente, se incluyen ponderadores que corrigen diferencias en el tamaño de la muestra entre países y proveen una mayor ponderación a las observaciones de países de mayor población.

Palabras clave: Encuesta en Línea, Pandemia del Coronavirus, Salud Pública, Mercado laboral, Desigualdad

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

Agradecimientos

Nos gustaría agradecer a Sebastián Espinoza y María Paula Medina por su magnífico trabajo como asistentes de investigación. Nos gustaría agradecer a Julián Cristia por su apoyo y asesoramiento en el proyecto. También queremos agradecer a Tom Sarrazin, Sebastián Oliva y Pablo Bachelet por su tremendo apoyo en la diseminación de la encuesta. También nos gustaría agradecer el apoyo de las oficinas de países del BID, representantes de países y economistas de países por su invaluable trabajo en la personalización de los cuestionarios para cada país. En particular, queremos agradecer a Marta Ruiz Aranz y Diether Beuermann por su apoyo en el lanzamiento de las encuestas en América Central y el Caribe. Quisieramos agradecer a Ana María Ibañez, Eric Parrado, Michael Lovenheim, y el equipo de investigadores en el departamento de investigación del BID por su valioso apoyo y retroalimentación en diferentes partes del proyecto.

Financiamiento: Este proyecto fue financiado por los fondos de investigación del coronavirus del BID (RG-E1700- encuesta coronavirus). La licencia de Qualtrics y la aprobación del IRB son de la Universidad de Cornell.

Datos

Recolección de Datos

Llevamos a cabo encuestas en línea a hogares de 17 países de América Latina y el Caribe. La encuesta se realizó en línea utilizando Qualtrics. La muestra consta de 230,540 observaciones en 8 países Sudamericanos (Chile, Colombia, Bolivia, Ecuador, Guyana, Perú, Suriname y Uruguay), 4 Norte y Centroamericanos (Costa Rica, El Salvador, México y Panamá), y 5 Caribeños (República Dominicana, Bahamas, Barbados, Jamaica y Trinidad & Tobago). La mayoría de los datos fueron recolectados durante las últimas dos semanas de Abril 2020. La encuesta fue lanzada por primera vez el 27 de Marzo, 2020 y se extendió a todos los 17 países el 17 de Abril, 2020. La columna (1) de la Tabla 1 muestra la fecha de lanzamiento de la encuesta en cada país. Con la excepción de Costa Rica, la recolección de datos continuó hasta el 30 de Abril, 2020. El número de observaciones varía según el país, desde 565 en Surinam hasta 35,556 en Chile (ver Tabla 1 para detalles específicos de cada país).

Construimos la base de datos siguiendo los siguientes pasos. Primero, restringimos la muestra a las encuestas completadas. Aproximadamente el 59% de las encuestas que se iniciaron fueron completadas. Segundo, eliminamos las encuestas asociadas a direcciones IP fuera de las fronteras del país en donde la persona respondió la encuesta. Entre todos los países de la muestra, menos del 1% de las encuestas completadas fueron eliminadas siguiendo este criterio. Tercero, excluimos las encuestas que Qualtrics marcó como posibles encuestas repetidas. Qualtrics marca las encuestas que son realizadas en el mismo dispositivo, que son probablemente respuestas repetidas hechas por el mismo individuo u hogar, basándose en las cookies almacenadas. Este es un filtro imperfecto. Por ejemplo, no reconocería encuestas repetidas por el mismo individuo u hogar que son completadas en diferentes dispositivos. Menos del 2.3% de las encuestas completadas fueron marcadas por Qualtrics, y eliminamos estas encuestas de la

muestra. En la columna (2) de la Tabla 1 se muestra el número de observaciones de la muestra por país.

Table 1: Fecha de Lanzamiento y Número de Observaciones por País

País	(1) Fecha de Lanzamiento	(2) No. de observaciones	(3) % de Localidades con observaciones
Chile	3/27/2020	35,556	97%
Bolivia	4/1/2020	25,970	83%
Panamá	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%

Fecha de lanzamiento es la fecha en la que las publicaciones en redes sociales empiezan. La fecha en la que las encuestas fueron lanzadas en cada país fue largamente determinada por procesos y aprobaciones burocráticas. Con la excepción de Costa Rica, la recolección de datos en cada país continuó hasta el 30 de Abril, 2020. El número de observaciones para cada país refleja el número en la muestra luego de la limpieza de datos. El porcentaje de localidades es el porcentaje de localidades para cada país para los cuales tenemos una o más observaciones en la muestra.

En general, nuestra muestra logra una amplia cobertura geográfica. Nuestros datos contienen observaciones en el 92% de las regiones sub-nacionales y en el 61% de las localidades en los países presentes en la muestra (ver Tabla 1). La Figura 1 detalla la amplia cobertura geográfica de nuestra muestra, mostrando el número de observaciones como proporción de la población (en %) por región sub-nacional para cada país en la muestra.

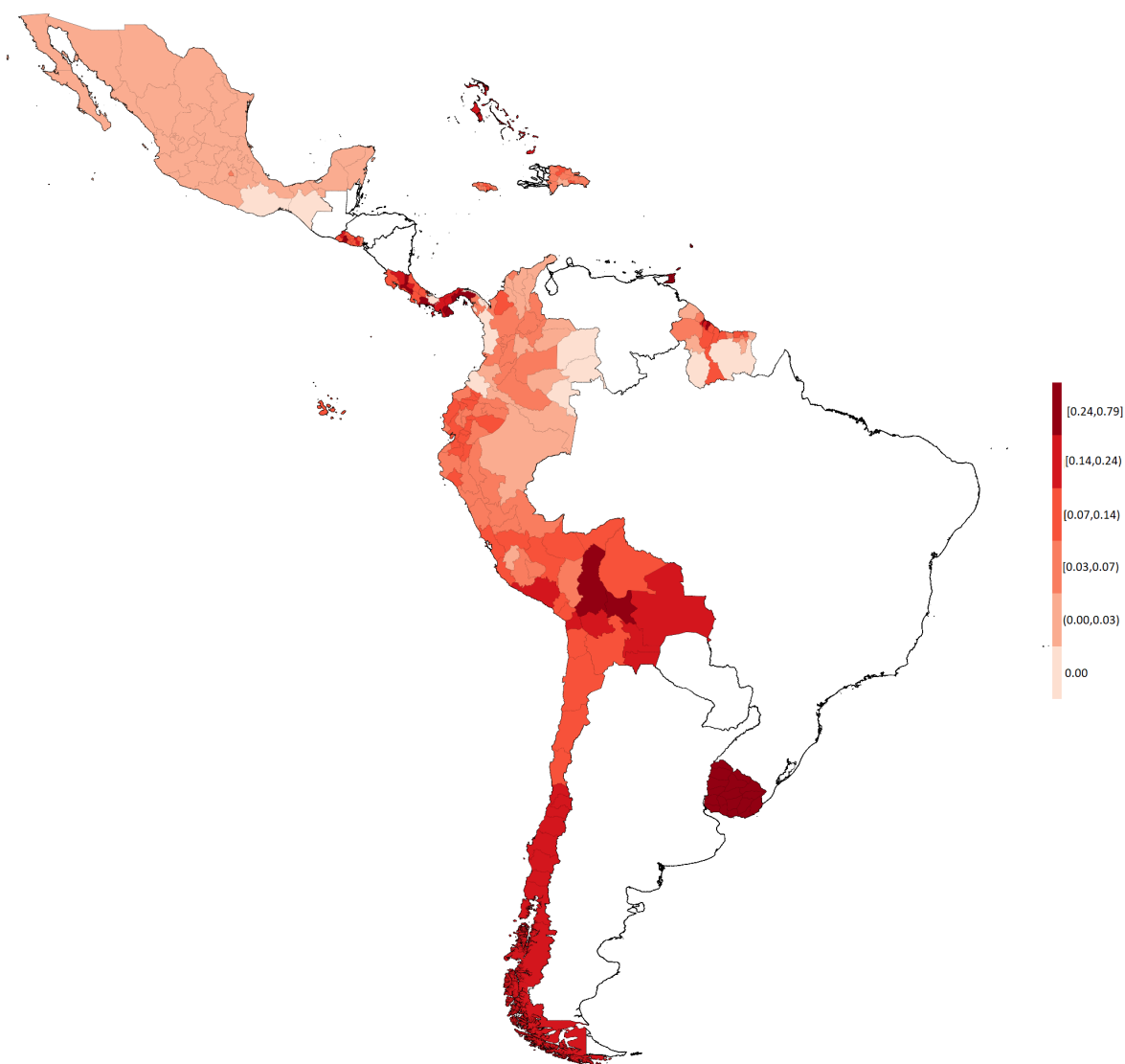


Figure 1: **Cobertura geográfica de la muestra a nivel sub-nacional.** Las regiones sub-nacionales en cada país están coloreadas de acuerdo al número de respuestas a la encuesta como porcentaje de la población (en %). Las fuentes de los datos de población para cada país se encuentran en la Sección de Material Suplementario.

Reclutamos participantes para la encuesta usando anuncios publicitarios mediante redes sociales. En cada país, usamos palabras clave de interés general que no estuvieran relacionadas con el Covid-19, como el nombre de celebridades del fútbol, para evitar reclutar a los participantes con base en su conocimiento o experiencia relacionada con la actual pandemia. Dirigimos los anuncios a personas con 19 años en adelante. Utilizamos una segunda campaña simultánea en las redes sociales con las mismas imágenes y palabras clave, pero para dirigirnos específicamente a personas entre los 24-35 años de edad, en los países en donde el promedio de edad de nuestra muestra fue alta o se esperaba fuese alta.

Las campañas publicitarias consistieron en tres publicaciones para cada país: una imagen de un hombre usando una mascarilla, una mujer usando una mascarilla y una imagen de un paisaje icónico del país. Las publicaciones fueron adaptadas a cada país de tres formas diferentes. Primero, la imagen del paisaje icónico fue específica a cada país. Segundo, las imágenes del hombre y la mujer se cambiaron según el contexto. Tercero, se cambió el idioma del texto de las publicaciones (español, inglés u holandés). Los algoritmos de Facebook seleccionaron qué publicación mostrar a cada usuario. Las publicaciones usadas en Uruguay se muestran a continuación a modo de ejemplo.

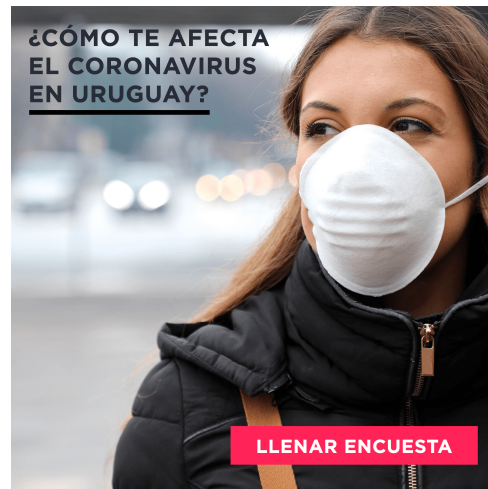


Figure 2: **Publicaciones en Facebook.** Creamos tres publicaciones de Facebook personalizadas para cada país, un hombre con una mascarilla, una mujer con una mascarilla y un paisaje icónico específico a cada país.

Diseño de la Encuesta

El objetivo principal de las encuestas fue medir el impacto de la pandemia en la economía y el bienestar de los hogares en América Latina y el Caribe. Por esta razón, el cuestionario se enfocó en recolectar datos en las siguientes categorías: resultados del mercado laboral, situación financiera, inscripción en programas sociales, hambre, escasez de bienes claves, y la aprobación de las diferentes políticas para ralentizar la propagación del Covid-19.

El cuestionario se estandarizó entre los países para la creación de una única base de datos que contuviera información de todos los países de la muestra.

Ejemplos de los cuestionarios en inglés, español y holandés pueden ser descargados aquí:
<https://www.dropbox.com/sh/uuv17cfaz94kw4h/AAAWvgJUNYEBHhNxZPcQVv3Za?dl=0>.

Validación

Estimación de los pesos

Para cada país, modelamos la probabilidad de ser representados en la encuesta representativa nacionalmente como una función de la demografía, características que no varían en el tiempo como género, edad y educación del que responde, indicadores para la presencia en el hogar de niños de 5 años o menos, y ancianos (60 años o más), así como el número de miembros en el hogar y el número de niños matriculados en la escuela. Finalmente, también incluimos indicadores específicos por regiones en cada país.

Realizamos este proceso país por país. Para cada uno, usamos la encuesta nacional representativa más reciente disponible en el repositorio de bases armonizadas del Banco Interamericano de Desarrollo. Estimamos el modelo ajustando a una función logística (logit) y computando las probabilidades predichas de estar en la encuesta nacionalmente representativa ($\hat{p}_{i,c}$) para cada encuestado i en el país c . Luego, usamos el inverso de probabilidad de los pe-

pesos ($ipw_{i,c} = 1/(1 - \hat{p}_{i,c})$) para que, al menos en términos de características observables, sean similares a las encuestas de hogares representativas.

Usamos estos propensity scores como semillas para la calibración de pesos usando método proporcional de ajuste iterativo (raking). Para computar los nuevos pesos calibramos los pesos para que calcen con la proporción de población de los encuestados basadas en diferentes variables. Usamos la distribución marginal de los encuestados por categoría de edad (< 40 , $40-60$, > 60 años), la proporción de mujeres, la proporción con educación superior, y la distribución marginal de ingresos (basada en 8 categorías) como los objetivos de momentos de población. Luego de aplicar los pesos del proceso iterativo $\tilde{w}_{i,c}$, el número de observaciones en la encuesta en línea en cada país coincide con la población total del país según los datos de la encuesta de hogares.

Para prevenir que diferencias en las tasas de respuesta determinen los resultados, re-escalamos los pesos dentro del país $\tilde{w}_{i,c}$ por el inverso de la proporción del número reponderado de respuestas por país, relativo a la población del país ($Population_c / (\sum_i^{N_c} \tilde{w}_{i,c})$) según el censo.

Validación

Validamos los pesos en la encuesta online de dos formas. Primero, comparamos las medias de las características demográficas con y sin pesos ponderados computados usando los datos online y las medias computadas usando los datos de las encuestas nacionalmente representativas que se realizan cara a cara.

La Tabla 2 muestra que los individuos que responden la encuesta online son más educados y más probables de ser mujeres que los que responden las encuestas de hogares (ver columna 1 y 3, respectivamente). Columna 4 y 5 reporta datos a partir de una sub-muestra de encuestas de hogares correspondiente a áreas urbanas y con hogares con acceso a internet en casa, y muestra que las características de los que encuestados online son similares a los encuestados por las

encuestas nacionalmente representativas. Columna 2 muestra que somos capaces de reducir las diferencias en las características del encuestado entre las encuestas online y de hogares al reponderar las observaciones de las encuestas online usando pesos recalibrados. En particular, las diferencias en genero y educación del encuestado son significativamente reducidas y nos acercamos bastante a la distribución de ingresos.

Table 2: Diferencias entre Datos Encuesta Online y Encuestas de Hogares

	Online		Encuesta de Hogar (en persona)		
	No ponderado (1)	Ponderado (2)	Total (3)	Urbano (4)	Internet (5)
Tamaño del hogar	4.38	4.22	3.86	3.87	3.96
Adultos mayores en el hogar (proporción)	0.35	0.3	0.36	0.35	0.34
Mujeres (proporción)	0.72	0.54	0.52	0.53	0.53
Educación: Nada	0	0.11	0.11	0.08	0.05
Educación: Primaria	0.03	0.18	0.31	0.29	0.22
Educación: Secundaria	0.23	0.38	0.41	0.4	0.41
Educación: Universitaria o Superior	0.74	0.33	0.16	0.2	0.3
Edad (en años)	39	40.47	43.49	42.88	42.09
0-0.5 SM	0.1	0.08	0.07	0.05	0.01
0.5-1 SM	0.16	0.13	0.11	0.09	0.04
1-2 SM	0.21	0.23	0.24	0.23	0.15
2-3 SM	0.14	0.16	0.18	0.19	0.18
3-4 SM	0.11	0.09	0.12	0.13	0.15
4-6 SM	0.1	0.12	0.14	0.16	0.21
6-8 SM	0.06	0.06	0.06	0.07	0.11
8-11 SM	0.05	0.05	0.04	0.05	0.08
11+ SM	0.07	0.06	0.04	0.04	0.07

La tabla presenta promedios de características demográficas usando datos de la encuesta online y las encuestas de hogares representativas, y las observaciones agrupadas de todos los países estudiados (ponderando por el tamaño del país). La columna (1) reporta las medias brutas usando todas las observaciones de las encuestas online. La columna (2) reporta las medias luego de reponderar los datos de las encuestas online. La columna (3) reporta medias para todas las observaciones disponibles en la encuesta de hogares usando los pesos de muestreo. Columna (4) y (5) reporta medias usando encuestas de hogares para la sub-muestra de hogares en áreas urbanas y hogares con acceso a internet en casa. SM se refiere a salario mínimo.

Población y Encuestas de Hogares

Año y Fuente de Encuestas de Hogares por País

País	Nombre encuesta	Año	Link al recurso
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
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
México	ENIGH	2018	https://www.inegi.org.mx/programas/enh
Panamá	EHPM	2018	https://www.inec.gob.pa/publicaciones/Default2.aspx?ID_CATEGORIA=5&ID_SUBCATEGORIA=38
Perú	ENAHO	2018	https://webinei.inei.gob.pe/anda_inei/index.php/catalog/672
República Dominicana	ENCFT	2018	https://www.bancentral.gov.do/a/d/2539
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

Las encuestas nacionales de hogares utilizadas fueron las más recientes disponibles en el repositorio de bases armonizadas del BID. Es posible acceder a cada encuesta en las páginas web de los organismos correspondientes a cada país. El link provisto es referencial para el acceso de los datos. Por favor considerar que en algunas ocasiones los microdatos deben ser solicitados directamente al organismo competente debido a restricciones de privacidad de los datos.

Año y Fuente de los Datos de Población por País

País	(1) Año	(2) Fuente
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
Panamá	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
Perú	2020	http://proyectos.inei.gob.pe/web/biblioineipub/bancopub/Est/Lib0846/libro.pdf
México	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
República Dominicana	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/

Las fechas de lanzamiento representan la fecha en la cual lanzamos la campaña en redes sociales para reclutar participantes en cada país. En todos los países excepto Costa Rica se continuó la recolección de datos hasta el 30 de Abril, 2020. La población proyectada para Panamá excluye Comarca Guna de Madugandi y Comarca Guna de Wargandi. Los datos de población entre los años 2018 y 2020 son proyectados.