

# Metodología para la evaluación de la calidad de apertura datos oficiales publicados por los gobiernos de América Latina respecto al COVID-19

Versión al 14 de mayo 2020

*Autores<sup>1</sup>: Rodrigo Balbontín y Juan Ignacio Vila<sup>2</sup>*

## **1. Importancia de evaluar los datos que publican los gobiernos**

Comunicar los datos relativos al impacto de la pandemia de COVID-19 de manera transparente es una potente herramienta para implementar de mejor manera las políticas para su combate. Por un lado, es útil para que la sociedad civil, las comunidades científicas y de expertos en políticas públicas puedan evaluar las medidas tomadas y hacer recomendaciones basadas en evidencia. Por otra parte -y tal vez más importante que lo anterior- es una poderosa herramienta para que los gobiernos puedan explicar de manera fundamentada las medidas que toman.

A raíz de la actual pandemia, se han creado innumerables iniciativas desde la sociedad civil y desde los principales usuarios de los datos relacionados al COVID-19. En este contexto, proveer datos más robustos y más abiertos a estos grupos significa una oportunidad para que los gobiernos puedan compartir sus decisiones a la sociedad civil con el potencial respaldo de la comunidad científica y de expertos en políticas públicas.

---

<sup>1</sup> Agradecimientos especiales a Cristóbal Cuadrado y Catterina Ferreccio por sus comentarios técnicos respecto a las variables a considerar en la presente metodología. También agradecemos a Auska Ovando, Catalina Balla, Vanessa González, Lissett García, Agustín Vázquez, Valentina Briones y Nicole Ríos por sus ideas, tiempo y poner a disposición las capacidades de la Fundación Ciudadanía Inteligente para esta iniciativa.

<sup>2</sup> Cualquier comentario o sugerencia respecto al presente documento enviar a [covid19@ciudadaniai.org](mailto:covid19@ciudadaniai.org)

Se ha observado también un proceso de aprendizaje evidente desde los tomadores de decisiones: desde mejorar la capacidad de testeo hasta anticiparse y preparar infraestructura para recibir un mayor flujo de infectados. Esperamos también ver una mejora en la calidad de apertura de datos, tanto para tener una ciudadanía mejor informada y preparada para combatir esta pandemia, como para tomar decisiones basadas en evidencia.

Así, el presente documento es un esfuerzo por establecer una metodología transparente y replicable para poder evaluar la calidad de la apertura de información oficial que entregan los distintos Gobiernos de América Latina respecto a la pandemia COVID-19.

## 2. Selección de las variables a evaluar

En primer lugar, es necesario especificar que este ejercicio de evaluación de la calidad de la apertura de los datos se limita estrictamente a evaluar la información oficial que entregan los Gobiernos Centrales de cada país. Quedan así fuera las potenciales asimetrías que puedan existir dentro de un mismo país respecto a la información que entregan las autoridades regionales o federales, según sea el caso. También quedan fuera de este análisis las iniciativas no-gubernamentales de recolección de datos. Es importante recalcar que este trabajo no constituye una evaluación respecto a la veracidad de las cifras entregadas por los gobiernos de cada país.

En segundo lugar, el tipo de información que entregan los gobiernos se divide en dos potenciales fuentes de datos:

- **Estadísticas Generales.** Corresponde a la información agregada que entregan los gobiernos respecto a datos relativos a la evolución de la enfermedad, ya sea a nivel nacional, sub-nacional (regiones, Estados federados o provincias, según sea el caso) o local (municipios o departamentos).
- **Microdatos anonimizados.** Se refieren a los datos individualizados, pero no identificables. En una base de datos se refleja en que una observación equivale a una persona testeada y/o infectada.

Una vez establecidos estos dos tipos de fuentes de datos, se seleccionaron 22 variables relativas a las Estadísticas Generales y 25 variables de microdatos anonimizados. Para la selección de estas variables fueron consultados expertos en epidemiología y salud pública. Así, este set de variables se entiende como la mínima información idónea sugerida para que un país publique de manera abierta, a disposición de la comunidad científica y expertos en política pública. En la publicación de los resultados del presente trabajo se incluirán además variables contextuales que algún país pueda reportar, pero no se incluye en la presente métrica (por ejemplo, trabajadores de la salud infectados, o un desglose del tipo de síntomas).

La forma de recolección de datos se realizó revisando los distintos informes periódicos, infografías y datos publicados por los respectivos gobiernos a partir de la fecha de esta publicación. Este proceso consta de dos etapas: en una primera se registraron las distintas fuentes de información por cada país, para luego chequear la presencia de las distintas variables de Estadísticas Generales y microdatos anonimizados en sus respectivos sitios.

Las variables de Estadísticas Generales se agruparon en cuatro dimensiones o tipos de variables:

#### **1. CARACTERÍSTICA DE BASE DE DATOS**

Variable:

1. Calidad de los datos del reporte

#### **2. TEST**

Variables:

1. Numero de test realizados
2. Numero de test informados con resultados
3. Test PCR
4. Test rápido
5. Resultado test

#### **3. INFECTADOS, RECUPERADO Y FALLECIDOS**

Variables:

1. Número de fallecidos
2. Sexo fallecidos
3. Edad fallecidos
4. Número de infectados
5. Número de recuperados
6. Fecha (o días previos) en que el infectado presenta síntomas
7. Sexo infectados
8. Edad infectados
9. Número de pacientes hospitalizados

10. Número de pacientes en camas UCI
11. Número de pacientes en camas UTI
12. Número de pacientes en ventilación mecánica invasiva

#### **4. INFRAESTRUCTURA**

Variables:

1. Número total de respiradores en el país
2. Número total de camas UCI en el país
3. Número de contactos en cuarentena
4. Capacidad máxima de los laboratorios

Análogamente, los indicadores relativos a los microdatos anonimizados se agruparon en cinco dimensiones o tipos de variables:

#### **1. GEOGRÁFICA**

Variables:

1. Nivel geográfico que se reporta
2. Distritos censales o unidades geográficas menor a un gobierno local

#### **2. DEMOGRÁFICA**

Variables:

1. Sexo
2. Edad o tramo de edad
3. Nacionalidad o estatus migratorio

#### **3. CONDICIÓN MÉDICA**

Variables:

1. Embarazo
2. Tabaquismo
3. Obesidad
4. Enfermedades asociadas

#### **4. TEST**

Variables:

1. Fecha test
2. Test PCR
3. Test rápido
4. Resultado test

#### **5. ESTADO DEL INFECTADO**

Variables:

1. Fecha de inicio de síntomas
2. Origen del caso (contacto local, desde el extranjero, etc.)
3. Número de contactos que ha tenido el infectado
4. Está recuperado
5. Está fallecido
6. Realiza cuarentena en algún establecimiento público (ej. albergue, hogar, hotel)
7. Infectado hospitalizado

8. Infectado en UCI
9. Infectado en UTI
10. Infectado en ventilación mecánica invasiva
11. Indica el o los medicamentos usados en el tratamiento
12. Persona testeada, con anticuerpos positivos

Cada variable se contabiliza con 0 en caso que no se reporte de manera idónea para efectos de la base de datos, o con 1 en caso que sí se reporte. Para las variables de testeo, infectados y fallecidos en el caso de las Estadísticas Generales se genera una variable categórica para capturar el nivel geográfico que cubre la información. El detalle de los valores asignados por cada variable se encuentra en los Anexo I y Anexo II.

Con las variables ya determinadas y agrupadas, se procede al levantamiento de información por cada país. Para minimizar el riesgo de omisión de información o error en el reporte, dos personas de manera independiente levantan la información de cada país, y luego se cruzan los datos. En caso de existir disensos en alguna variable para un determinado país, un tercer individuo evalúa qué valor asignar.

### **3. Construcción de puntaje de cada país**

Es necesario establecer una métrica para determinar si existen brechas en los datos que publica un país y el mínimo idóneo para efectos de investigación científica y medición de políticas públicas. Para ello, se construye un indicador de 0 a 100 tanto para la información de Estadísticas Generales como para los microdatos anonimizados, donde 100 representa que un país reporta la totalidad de las variables, en el formato y nivel geográfico adecuado.

En primer lugar, se establecieron distintos pesos relativos para las familias de variables, tanto para las Estadísticas Generales como para los microdatos anonimizados. Con ello, se busca indicar que los distintos grupos de variables tienen una prioridad disímil para efectos de usar los datos. Esto es especialmente relevante para, por ejemplo, realzar la relevancia de que las Estadísticas Generales sean publicados en un formato .csv o similar; o la relevancia del nivel geográfico que abarcan los microdatos anonimizados.

Sólo se consideran países cuyo(s) reporte(s) de testeados, infectados y/o fallecidos hayan sido elaborados al menos una semana antes del momento de levantar los datos.

Así, las tablas 01 y 02 reflejan los ponderadores para los distintos tipos de variables, para el caso de las Estadísticas Generales y microdatos anonimizados.

**TABLA 01.** Ponderadores para los tipos de variables cubiertos en las Estadísticas Generales

TIPO DE VARIABLE	PONDERACIÓN
Características de Base de Datos	20%
Test	20%
Infectados y fallecidos	50%
Infraestructura	10%

**TABLA 02.** Ponderadores para los tipos de variables cubiertos en los microdatos anonimizados

TIPO DE VARIABLE	PONDERACIÓN
Geográfica	10%
Demográfica	20%
Condición médica	20%
Test	25%
Estado del infectado	25%

Entonces, el indicador por país se construye determinando la suma de cada variable sobre su máximo posible de la variable, ponderado por el tipo de variable. Es decir:

$$PUNTAJE_i = 100 * \left( \sum_k \sum_j \alpha_k * \frac{x_{j,i}}{X_j} \right)$$

Donde:

$i$ : país

$\alpha_k$ : ponderador  $\alpha$  del tipo de variable  $k$

$x_{j,i}$ : valor de la variable  $j$ , para el país  $i$

$X_j$ : máximo valor posible de la variable  $j$

Finalmente, es importante reconocer que el total máximo de datos posibles relativos a personas testeadas, infectadas, recuperadas y fallecidas producto de esta pandemia es directamente proporcional al número de test que realizan los países. Es por ello que, para efectos de visualización de los resultados de los indicadores construidos, se publicará como referencia el número de tests por 1.000 habitantes, de acuerdo al proyecto Our World In Data de Global Change Data Lab (<https://ourworldindata.org/>).

## Anexo 01

Descripción de los valores asignados para cada variable de las Estadísticas Generales.

DESCRIPCIÓN VARIABLE	VALORES
País	Nombre país
Calidad de los datos del reporte	0=si es pdf, imagenes u otro formato no analizable, 1 = si es bulk/csv
Número de tests realizados	0=No se reporta, 1=Se reporta sólo a nivel nacional, 2=Se reporta hasta nivel sub-nacional, 3=Se reporta hasta nivel local
Número de tests informados con resultados positivos	0=No se reporta, 1=Se reporta sólo a nivel nacional, 2=Se reporta hasta nivel sub-nacional, 3=Se reporta hasta nivel local
Test PCR	0=No se reporta, 1=Se reporta sólo a nivel nacional, 2=Se reporta hasta nivel sub-nacional, 3=Se reporta hasta nivel local
Test rápido	0=No se reporta, 1=Se reporta sólo a nivel nacional, 2=Se reporta hasta nivel sub-nacional, 3=Se reporta hasta nivel local
Resultado test (positivo, negativo y –si reporta- muestras no procesadas)	0=No se reporta, 1=Se reporta sólo a nivel nacional, 2=Se reporta hasta nivel sub-nacional, 3=Se reporta hasta nivel local
Número de fallecidos	0=No se reporta, 1=Se reporta sólo a nivel nacional, 2=Se reporta hasta nivel sub-nacional, 3=Se reporta hasta nivel local
Sexo fallecidos	0=No se reporta, 1=Se reporta sólo a nivel nacional, 2=Se reporta hasta nivel sub-nacional, 3=Se reporta hasta nivel local
Edad fallecidos	0=No se reporta, 1=Se reporta sólo a nivel nacional, 2=Se reporta hasta nivel sub-nacional, 3=Se reporta hasta nivel local
Número de infectados	0=No se reporta, 1=Se reporta sólo a nivel nacional, 2=Se reporta hasta nivel sub-nacional, 3=Se reporta hasta nivel local

Número de recuperados	0=No se reporta, 1=Se reporta sólo a nivel nacional, 2=Se reporta hasta nivel sub-nacional, 3=Se reporta hasta nivel local
Fecha (o días previos) en que el infectado presenta síntomas	0=No se reporta, 1=Se reporta sólo a nivel nacional, 2=Se reporta hasta nivel sub-nacional, 3=Se reporta hasta nivel local
Sexo infectados	0=No se reporta, 1=Se reporta sólo a nivel nacional, 2=Se reporta hasta nivel sub-nacional, 3=Se reporta hasta nivel local
Edad infectados	0=No se reporta, 1=Se reporta sólo a nivel nacional, 2=Se reporta hasta nivel sub-nacional, 3=Se reporta hasta nivel local
Número de pacientes hospitalizados	0=No se reporta, 1=Se reporta sólo a nivel nacional, 2=Se reporta hasta nivel sub-nacional, 3=Se reporta hasta nivel local
Número de pacientes en camas UCI	0=No se reporta, 1=Se reporta sólo a nivel nacional, 2=Se reporta hasta nivel sub-nacional, 3=Se reporta hasta nivel local
Número de pacientes en camas UTI	0=No se reporta, 1=Se reporta sólo a nivel nacional, 2=Se reporta hasta nivel sub-nacional, 3=Se reporta hasta nivel local
Número de pacientes en ventilación mecánica invasiva	0=No se reporta, 1=Se reporta sólo a nivel nacional, 2=Se reporta hasta nivel sub-nacional, 3=Se reporta hasta nivel local
Número total de respiradores en el país	0=no reporta, 1=se reporta
Número total de camas UCI en el país	0=no reporta, 1=se reporta
Numero de contactos en cuarentena (personas que han tenido contacto con infectados confirmados y se encuentran en cuarentena)	0=no reporta, 1=se reporta
Capacidad máxima de los laboratorios	0=no reporta, 1=se reporta

## Anexo 02

Descripción de los valores asignados para cada variable de los microdatos anonimizados.

DESCRIPCIÓN VARIABLE	VALORES
País	Nombre país
Nivel geográfico que se reporta	0=No hay dato, 1=Nivel nacional, 2=Nivel sub-nacional, 3=Nivel local
Distrito censales o unidades geográficas menor a un gobierno local	0=no reporta, 1=se reporta
Sexo	0=no reporta, 1=se reporta
Edad o tramo de edad	0=no reporta, 1=se reporta
Nacionalidad o estatus migratorio	0=no reporta, 1=se reporta
Embarazo	0=no reporta, 1=se reporta
Tabaquismo	0=no reporta, 1=se reporta
Obesidad	0=no reporta, 1=se reporta
Enfermedades asociadas	0=no reporta, 1=se reporta
Fecha test	0=no reporta, 1=se reporta
Test PCR	0=no reporta, 1=se reporta
Test rápido	0=no reporta, 1=se reporta
Resultado test	0=no reporta, 1=se reporta
Fecha de inicio de síntomas	0=no reporta, 1=se reporta
Origen del caso (contacto local, desde el extranjero, etc)	0=no reporta, 1=se reporta

Número de contactos que ha tenido el infectado	0=no reporta, 1=se reporta
Está recuperado	0=no reporta, 1=se reporta
Está fallecido	0=no reporta, 1=se reporta
Realiza cuarentena en algún establecimiento público (ej. albergue, hogar, hotel)	0=no reporta, 1=se reporta
Infectado hospitalizado	0=no reporta, 1=se reporta
Infectado en UCI	0=no reporta, 1=se reporta
Infectado en UTI	0=no reporta, 1=se reporta
Infectado en ventilación mecánica invasiva	0=no reporta, 1=se reporta
Indica el o los medicamentos usados en el tratamiento	0=no reporta, 1=se reporta
Persona testeada, con anticuerpos positivos	0=no reporta, 1=se reporta