

Datos abiertos en México durante la pandemia por COVID-19[#]

Erick Osorio López (@ErickOsorio33), Oswaldo Tostado Islas (@Islas_oswaldo)

Instructores del Departamento de Salud Pública de la Facultad de Medicina, UNAM

Resumen: En México y otros países del mundo, se han elaborado bases de datos de libre acceso que compilan características clínicas y sociodemográficas de los casos sospechosos y casos confirmados de COVID-19. Son múltiples los usos que se le han dado a las bases, de los cuales destaca su uso en artículos originales y la difusión de información actualizada, tanto al personal de salud como a la ciudadanía.

Palabras clave: COVID-19, datos abiertos, vigilancia epidemiológica, México.

1. Introducción

En México se han realizado diversos esfuerzos para mejorar la calidad de la información que almacenan las bases de datos de salud. En las últimas décadas, la Secretaría de Salud Federal ha desarrollado instrumentos que permiten difundir la información que contiene este tipo de bases de datos, por ejemplo: el *Boletín Epidemiológico Semanal*¹ y los *Anuarios de Morbilidad*² de la Dirección General de Epidemiología (DGE); o bien, los *Cubos Dinámicos* de la Dirección General de Información en Salud (DGIS).³

La incorporación de nuevas tecnologías ha posibilitado a los usuarios acceder de manera remota y libre a algunas bases de datos y sus publicaciones. Si bien son herramientas perfectibles en diferentes aspectos, no hay duda en que han logrado reducir las barreras del acceso a la información en salud.

2. Características de la base de datos abierta COVID-19

Durante la pandemia por COVID-19, y en concordancia con el Decreto que establece la regulación en materia de Datos Abiertos,⁴ la Secretaría de Salud ha publicado diariamente desde el 14 de abril del 2020 la base de datos abierta de COVID-19 de México.⁵ Esta base de datos contiene la información de todos los casos documentados de COVID-19, incluyendo a los casos sospechosos, casos confirmados y casos con prueba negativa, entre otros subgrupos. Es importante mencionar que la base no contiene datos personales que comprometan la privacidad o revelen la identidad de las personas.

Cada base de datos diaria sobre casos de COVID-19, contiene un total de 38 variables que abordan aspectos sociodemográficos, clínicos, y de las instituciones del Sistema Nacional de Salud. Dentro de dichas variables, destacan: características generales del paciente, como su edad, sexo, entidad de nacimiento y de residencia, así como comorbilidades previas. Otros datos clínicos respecto al cuadro de COVID-19: las fechas de inicio de síntomas, de ingreso y de defunción si el caso lo amerita. Por otro lado, se especifica si la persona requirió ingresar a alguna Unidad de Cuidados Intensivos. Algunas otras variables relevantes son la entidad federativa que reportó el caso de COVID-19 y la institución en que fue atendida la persona.

Las bases de datos diarias publicadas en el portal de la DGE se encuentran disponibles en formato *.CSV, cuyas siglas provienen del inglés “comma-separated values”, el cual almacena el valor de las variables separadas por comas, tal y como su nombre lo sugiere. Este formato es de gran utilidad para poder compilar grandes volúmenes de información, con la ventaja adicional de ser compatible con diversos programas que permiten procesar dichos datos.

Hasta el 20 de mayo del 2021, la base de datos estaba integrada por 6,905,098 registros.⁶ Por lo tanto, si se intenta abrir la base de datos utilizando *Microsoft Excel*, sólo se podría recuperar una parte de la base y no la totalidad de los casos, ya que este programa solamente es capaz de mostrar hasta

[#] El contenido de los artículos es responsabilidad de sus autores y no necesariamente refleja la postura de la Facultad de Medicina.

1,048,576 filas (registros). Como alternativa se puede utilizar alguno de los programas estadísticos más utilizados en salud pública para procesar grandes bases de datos, tales como el *SPSS*, *Stata* o *R*. Por ejemplo, la versión completa del programa *Stata*, permite procesar bases de datos de hasta 20 billones de registros.

La información que se presenta en esta base de datos ha sido utilizada para elaborar trabajos científicos, presentados en congresos (literatura gris) o publicados en revistas con *revisión de pares*.

Un ejemplo del tipo de información que se puede obtener a partir de la base de datos abierta, se muestra en la tabla I, que contiene las frecuencias de cada una de las 7 categorías que existen dentro de la variable “CLASIFICACION_FINAL”. Dichas categorías están codificadas con números del 1 al 7. Para identificar el significado de cada uno de esos códigos, es necesario consultar los documentos contenidos en el archivo denominado “Diccionario de Datos”, también disponible en el sitio web de la DGE.

Otro ejemplo de información obtenida a partir de las bases de datos, son algunos de los artículos de la sec-

ción *Actualización Epidemiológica*, que han sido publicados en diversos números de este Boletín. Estas contribuciones han analizado distintas variables que posee la base de datos y que contienen información epidemiológica relevante. Algunos de los análisis realizados abordan temas de incidencia acumulada de COVID-19, así como la mortalidad específica según grupos de edad, sexo y entidad federativa, entre otras variables.

3. ¿Cuál es el uso que se le puede dar a la base de datos abierta sobre COVID-19?

Dado que es una base de datos nacional —pero que contiene un desglose a nivel estatal y municipal— es posible hacer una gran cantidad de cruces de información, incluso con otras bases de datos.

Uno de los usos más representativos es la elaboración de trabajos de investigación. En el artículo *¿Qué se ha publicado sobre COVID-19 en México a un año de la pandemia?*,⁷ se referenciaron 186 artículos publicados en revistas indexadas hasta el 12 de abril del 2021, de los cuales el 23% utilizaron este tipo de bases de datos abiertas como fuente de información. Aunque los datos pueden ser utilizada por los toma-

Tabla I. Clasificación de casos de COVID-19 en la base de datos abierta de la Secretaría de Salud

Código	Descripción	Número de registros	%
1	Caso de COVID-19 confirmado por asociación clínica epidemiológica	147,228	2.1
2	Caso de COVID-19 confirmado por comité de dictaminación	8,819	0.1
3	Caso de SARS-CoV-2 confirmado	2,234,093	32.4
4	Inválido por laboratorio	10,987	0.2
5	No realizado por laboratorio	73,194	1.1
6	Caso sospechoso	353,692	5.1
7	Negativo a SARS-CoV-2	4,077,085	59
Total		6,905,098	100.0

Fuente: Referencia 6.

dores de decisiones para el manejo de la pandemia, también se utilizan para realizar investigación en salud pública. Sin duda, este porcentaje de publicaciones no refleja al total de personas que han utilizado las bases de datos abiertas para distintas finalidades.

Por otro lado, es importante mencionar que México no ha sido el único país en publicar bases de datos abiertas con información sobre los casos de COVID-19. En la tabla II se muestran algunos ejemplos de bases de datos abiertas en distintos países. Aunque

son repositorios heterogéneos en cuanto a sus variables, tienen como denominador común datos básicos: casos COVID-19, defunciones, pacientes recuperados, edad, sexo, entre otras. Al igual que en el caso de México, la mayoría están disponibles en formato *.CSV⁸

En el mundo existen distintos mapas interactivos que hacen uso de estas y otras bases de datos que, aunque no tienen un desglose a nivel individual, brindan un panorama, casi en tiempo real, sobre la cantidad de nuevos casos diagnosticados de COVID-19.

Tabla II. Algunos ejemplos de bases de datos nacionales sobre COVID-19

País	Institución	Repositorio	Vínculo en la Web
Argentina	Ministerio de salud	Covid19arData	https://bit.ly/2Spi4IM
Australia	Departamento de Salud Australiano	covid-19-au	https://bit.ly/3umKD6Q
China	Comisión Nacional de Salud de China	JHU Midas-China	https://bit.ly/3ukhSb1 https://bit.ly/3uIPHSi
Francia	Sistema de Salud Público Francés	opencovid19-fr FRANCE-COVID-19	https://bit.ly/3wx73Uz https://bit.ly/2QPwNMC
Alemania	Instituto Robert Koch	covid-19-germany-gae	https://bit.ly/34jyc12
Islandia	Gobierno de Islandia	gauri-covid19	https://bit.ly/3i6peg3
Italia	Departamento de Protección Civil de Italia	pcm-dpc	https://bit.ly/3fh3Z9f
Paraguay	Ministerio de Salud Pública y Bienestar Social	covidpy-rest	https://bit.ly/3vqjXn9
Sudáfrica	Instituto Nacional de Enfermedades Contagiosas	covid19za	https://bit.ly/3hWNWiC
Corea del Sur	Centro para el control y prevención de enfermedades	COVID19-Korea	https://bit.ly/3hPk3j
España	Ministerio de Salud	datadista-Covid.19	https://bit.ly/3yDQhVO
Reino Unido	Salud Pública Inglesa	covid-19-uk-data	https://bit.ly/3fj7w6U
Estados Unidos	Centro de Control y Prevención de Enfermedades	JHU Nytimes	https://bit.ly/3fkG4pF https://bit.ly/3oRq5IK
México	Dirección General de Epidemiología	Sitio DGE, mapa interactivo CONACyT	https://bit.ly/3vrE0BG

Fuente: Adaptación propia tomando como base la referencia 8.

Algunos gobiernos emiten actualizaciones sobre los datos epidemiológicos de su territorio; sin embargo, la información solamente se encuentra en el idioma de dicho país.

Las bases de datos se hacen tangibles a través de novedosas herramientas digitales que acercan esta información a las y los ciudadanos. Destaca el caso de Brasil, país que cuenta con el portal *Localiza SUS*, el cual contiene los registros de morbilidad y mortalidad, casos recuperados y casos nuevos por municipio. También contiene una sección de vacunas aplicadas, un boletín informativo y una base de datos abierta sobre COVID-19.⁹ En Chile, por su parte, también se utilizan los portales digitales: en el sitio del Ministerio de Salud alberga todo tipo de información sobre COVID-19, además de ser el sitio de publicación del reporte diario sobre COVID-19 con información sobre la morbimortalidad, incidencia y capacidad hospitalaria.¹⁰

A manera de conclusión

Las herramientas tecnológicas disponibles en la actualidad permiten construir y acceder a bases de datos, con una rapidez nunca antes vista en otras pandemias. Los datos tienen una gran aplicabilidad en distintos ámbitos: en la toma de decisiones por parte de los gobiernos; para la elaboración de trabajos de investigación; y no menos importante, para brindar información oportuna y de calidad a las y los ciudadanos. Es importante continuar sumando esfuerzos para mejorar los sistemas de información en salud, teniendo el objetivo claro de que, tanto gobierno como profesionales de la salud, puedan tomar acciones basadas en información oportuna y confiable.

Referencias

1. Dirección General de Epidemiología. Boletín Epidemiológico. Sistema Nacional de Vigilancia Epidemiológica. Sistema Único de Información; 2021. Disponible en: <https://buff.ly/3D4AuRg>.
2. Dirección General de Epidemiología. Anuarios de morbilidad 1984-2019. Ciudad de México: Secretaría de Salud. Sistema Nacional de Vigilancia Epidemiológica. Sistema Único de Información; 2020. Disponible en: <https://buff.ly/3CAXtUn>.
3. Dirección General de Información en Salud. Cubos dinámicos - Servicios SIS 2021. Disponible en: <https://buff.ly/3FaEz8h>.
4. Presidencia de la República. DECRETO por el que se establece la regulación en materia de Datos Abiertos. Diario Oficial de la Federación. 2015 20 de febrero. Disponible en: <https://buff.ly/3CZeqau>.
5. Secretaría de Salud. Datos Abiertos - Dirección General de Epidemiología 2020-2021. Disponible en: <https://buff.ly/38rFyRu>.
6. Secretaría de Salud. Datos Abiertos - Dirección General de Epidemiología [sitio de internet]. 2021. [actualizado al 20 de mayo]. Disponible en: <https://buff.ly/38rFyRu>.
7. Osorio López E, Tostado Islas O. ¿Qué se ha publicado sobre COVID-19 en México a un año de la pandemia? Boletín sobre COVID-19: Salud Pública y Epidemiología. 2021;2(17):14-16. Disponible en: <https://buff.ly/3t3IYUx>.
8. Alamo T, Reina DG, Mammarella M, Abella A. Covid-19: Open-Data Resources for Monitoring, Modeling, and Forecasting the Epidemic. 2020;9(5):827. Disponible en: <https://buff.ly/3D9H4WF>.
9. Localiza SUS. [Internet] [consultado: 22 de mayo de 2021] Disponible en: <https://buff.ly/2Ww7fqI>.
10. Ministerio de Salud – Gobierno de Chile. Casos confirmados en Chile COVID-19 [Internet] [consultado: 22 de mayo de 2021]. Disponible: <https://buff.ly/3djYun5>.