

Revista del Centro de Investigación de la Universidad La Salle
Vol. 15, No. 57, Enero-Junio, 2022: 227-266
DOI: <http://doi.org/10.26457/recein.v15i57.2841>

**Datos mexicanos de COVID-19 en sus etapas iniciales
¿estratégicamente distorsionados?**

**Mexican COVID-19 data in its early stages, strategically
distorted?**

Julio César Arteaga García
Universidad Autónoma de Nuevo León (México)

Edgar Mauricio Luna Domínguez¹
Universidad Autónoma de Nuevo León (México)

Recibido: 11 de diciembre de 2020
Aceptado: 12 de noviembre de 2021
Publicado: 09 de mayo de 2022

Resumen

El contar con mediciones insesgadas de indicadores relacionados con la pandemia de COVID-19 es primordial para establecer estrategias de políticas públicas que permitan mitigar la dispersión de la enfermedad y evaluar los resultados de dichas políticas. El objetivo de este trabajo es identificar entidades que realizan más cambios en mediciones en fechas muy distantes a los días en los cuales se publican los reportes del COVID-19 en México – denominados distorsiones– y los cofactores que contribuyen a explicar las magnitudes de estas distorsiones, durante la primera ola de esta pandemia. Con la finalidad de identificar las

¹ Email: edgar.lunadmz@uanl.edu.mx



entidades que más distorsionan la información, se utilizan análisis de *clúster* jerárquico y pruebas de cointegración. Además, a través del análisis de regresión se relaciona la magnitud de las distorsiones con información de indicadores de competitividad estatal. De acuerdo con los resultados, se encuentra que las entidades distorsionan información de contagios para no incidir negativamente en determinados sectores económicos. Estos resultados sugieren evidencia de posibles acciones estratégicas realizadas por las entidades con respecto al flujo de información del COVID-19. Dentro de las limitaciones del estudio destacan que el número de observaciones está limitado a la cantidad de entidades federativas y los cofactores son invariantes en el tiempo. Por otra parte, la magnitud real de las distorsiones en mediciones del COVID-19 podría ser mayor ya que se ha reconocido el sub-registro en la información de la Secretaría de Salud.

Palabras clave: COVID-19, Distorsión, Entidades.

Abstract

Having unbiased measurements of indicators related to the COVID-19 pandemic is essential to establish public policy strategies to mitigate the spread of the disease and to evaluate the results of such policies. The objective of this work is to identify entities that make more changes in measurements on dates far distant from the days on which COVID-19 reports are published in Mexico –called distortions– and the co-factors that contribute to explain the magnitudes of these distortions. We focus on the first wave of this pandemic. In order to identify the most distorting entities, we use hierarchical cluster analysis and cointegration tests. In addition, we use regression analysis to relate the magnitude of the distortions with information from state competitiveness indicators. According to the results, we found that entities distort contagion information in order to avoid a negative impact on certain economic sectors. These results suggest evidence of possible strategic actions taken by the states with respect to the flow of COVID-19 information. Among the limitations of the study are that the number of observations is limited to the number of states and the cofactors are time invariant. On the other hand, the real magnitude of the distortions in COVID-19 measurements could be greater since the underreporting of information from the Ministry of Health has been recognized.

Keywords: COVID-19, Distortion, States.

JEL: I18, I19, H75

Introducción

Dentro de la literatura relacionada con el COVID-19, mucho se ha escrito acerca de los posibles factores que influyen en la propagación del virus (Auler *et al.*, 2020; Bolaño-Ortiz *et al.*, 2020; Borjas, 2020; Valero y Valero, 2021); además se han realizado pronósticos de la evolución de la tasa de fatalidad del COVID-19 ya sea a nivel regional, nacional o mundial (Atkeson, 2020). Similarmente, se han elaborado estudios que pronostican las consecuencias económicas de esta pandemia (CONEVAL, 2020; Laborde *et al.*, 2020; Pujol, 2020; Stock, 2020), así como de la manera en que se podrían mitigar los efectos de la pandemia sobre la economía a través de políticas públicas (Vázquez, Muller y Zavaleta, 2021; Gómez-Ramírez y Villarraga-Orjuela, 2021; Desilus, 2021; Andrade *et al.* 2021). Sin embargo, existen pocos estudios que abordan la distorsión de la medición de los datos relacionados con esta enfermedad, es decir aquellos cambios en la información de fechas muy distantes al día del reporte, y parecería que no hay acerca de la posible decisión estratégica por parte de las autoridades locales para diferir el flujo de información acerca de este virus.

De los estudios que se han realizado acerca de la medición de indicadores relacionados del COVID-19, unos reconocen que los contagios confirmados y las muertes pueden estar subestimados. Por ejemplo, Favero (2020) demuestra que el número de casos confirmados en Estados Unidos se puede ajustar con base en el porcentaje de pruebas por COVID-19 que resulten positivas. Por su parte, Burki (2020) reporta 20,000 muertes adicionales en los asilos de Inglaterra y Gales durante marzo y abril, en lugar de las 12,526 muertes por COVID-19 que indican las cifras oficiales. Para el caso de nuestro país, Zavala y Despeghel (2020) estiman que hasta el 20 de mayo se habían emitido 8,072 actas de defunción más que el promedio de actas emitidas entre enero y mayo de los últimos cuatro años. Ellos sugieren que esa comparación de actas de defunción puede contribuir para tener una mejor idea del ritmo de expansión de COVID-19 en el país. En un trabajo posterior, estiman un exceso de mortalidad de 83,235 personas entre marzo del 2020 y febrero del 2021, lo cual equivale a más de 2.5 veces las defunciones reportadas por la Secretaría de Salud (Zavala y Despeghel, 2021). De igual manera, investigadores como Ederly (2021)² y Ximénez-Fyvie (2021) han

² <https://sites.google.com/site/arturoerdely/mi-opinion-sobre/covid-19-en-méxico/art%C3%ADculos>

cuestionado las cifras y el manejo de la pandemia. En particular, Ederly (2021) sugiere una subestimación del factor de expansión calculado por la Secretaría de Salud con base en el método centinela a principios de la pandemia.

Seoane (2020) señala que un problema con el hecho de que las estadísticas oficiales del COVID-19 tengan anomalías es que el diseño de las estrategias públicas a seguir en la nueva normalidad se basa en la evolución de las mismas. También menciona que es altamente probable que el nivel real de subestimación sea específico para cada país.

En nuestro país, la Secretaría de Salud a nivel federal concentra y reporta la información que proveen las secretarías de salud estatales. Aunque la actualización de la información es diaria, se permiten cambios en los registros ya anunciados de días previos. Asimismo, se ha modificado la manera de medir la evolución de los indicadores relacionados con la pandemia. En particular, a partir de octubre cambiaron la metodología y los criterios que definen a una persona como positivo al COVID-19. Desde entonces, se incluye la definición de casos confirmados por asociación epidemiológica; es decir, individuos que muestran síntomas relacionados con la enfermedad y que tuvieron contacto con personas confirmadas por COVID pero que no han tenido acceso a una prueba (Secretaría de Salud, 2020a). Este cambio en la metodología limita el análisis de reportes para un periodo más amplio.

En este sentido, el propósito de esta investigación se centra en analizar una parte de los cambios que hacen las entidades federativas en la información que reportan del COVID-19, particularmente las modificaciones hechas en fechas previas muy distantes del día del reporte. Estos cambios los denominamos distorsiones y difieren de las actualizaciones explicadas por el tiempo que transcurre entre el inicio de los síntomas, la consulta, la toma de muestra, el traslado de la muestra al laboratorio, el análisis de la muestra, la obtención de los resultados y su reporte a la Secretaría de Salud. El objetivo de este trabajo es el de identificar los cofactores que contribuyen a explicar los cambios en mediciones hechos por las entidades de nuestro país en fechas muy distantes a los días en los cuales se publican los reportes del COVID-19 en México, durante la primera ola de esta pandemia.

Para cumplir con el propósito de identificar a las entidades que realizan los mayores cambios en la información, se utilizan técnicas de análisis de *clúster* jerárquico y pruebas de cointegración. Por su parte, se asocian los subíndices que publica el Instituto Mexicano para la Competitividad (IMCO) en su análisis para la competitividad estatal del año 2020 como posibles cofactores de las distorsiones en las mediciones de las entidades del país. Para la realización de estos tres ejercicios, se consideraron los datos del número de confirmados y defunciones diarios asociados con COVID-19 para la primera etapa de la pandemia (hasta el 17 de agosto) publicadas por la Secretaría de Salud (2020). Sin embargo, para los análisis de *clúster* y regresión, se dividió la muestra en tres periodos que coinciden con las fechas en que estuvieron suspendidas las actividades no esenciales, la estimación del primer pico de la pandemia y el principio de la nueva normalidad. Se considera esta división con el fin de saber cuáles gobiernos estatales tuvieron incentivos a distorsionar la información durante diferentes etapas de la pandemia y si los cofactores que explican la distorsión de la información cambiaron durante esos tres periodos.

Los resultados que surgen del uso de técnicas de cointegración a la información diaria evidencian sub-registros de muertes, contagios o ambos en al menos seis entidades federativas. Por otra parte, se encuentra evidencia que indica que las distorsiones que hacen las entidades en la información previamente reportada son realizadas de manera estratégica puesto que mayores distorsiones en contagios están asociadas con un mejor posicionamiento en determinados sectores económicos, al tiempo que las distorsiones en muertes se relacionan negativamente con el estado de derecho funcional en las entidades.

El resto del trabajo se organiza como sigue: la sección dos presenta un análisis descriptivo de las distorsiones en las mediciones de casos confirmados y defunciones de las entidades y confirma que no son homogéneas entre las entidades del país. En la sección tres, se realizan ejercicios de análisis de *clúster* jerárquico y se identifican los factores que están asociados con las distorsiones en las mediciones. En la sección cuatro, se realiza un análisis de cointegración entre las series de casos confirmados y defunciones de las entidades. Por último, la sección cinco concluye.

1. Análisis descriptivo

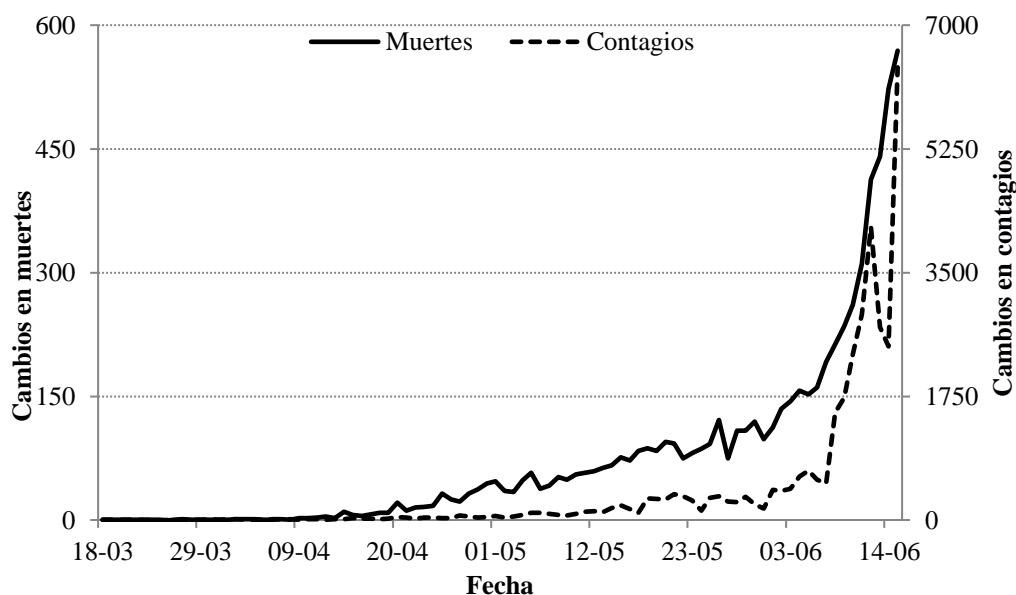
A nivel federal, la Secretaría de Salud concentra y reporta diariamente la información que proveen las Secretarías de Salud de cada entidad. Por ejemplo, el 15 de junio se reportó que a esa fecha había en el país un acumulado de 150,266 casos confirmados y 17,582 muertes por COVID-19. Sin embargo, el proceso diario de actualización permite a las entidades modificar la información reportada de días previos. De esta manera, en el reporte proporcionado por la Secretaría de Salud el 17 de agosto, se observa que de los 525,733 contagiados y 57,023 muertes anunciados, 184,364 contagiados y 24,192 muertes se registraron antes del 15 de junio; es decir, un cambio en los datos que implica 22.7% más casos y 37.6% más muertes que lo originalmente reportado.

La Gráfica 1 muestra los cambios en la información diaria presentada el 15 de junio con respecto a la del 17 de agosto por parte de la Secretaría de Salud para el periodo del 18 de marzo al 15 de junio. Es decir, esta gráfica presenta cómo están distribuidos los 34,098 contagios y las 6,610 muertes adicionales de la base de agosto³. Es fácil observar que la mayor parte de los cambios se concentran en fechas cercanas al 15 de junio; por ejemplo, para ese mismo día la base del 17 de agosto agrega 6,403 contagios y 569 muertes, siendo el día con mayor cambio entre ambas bases. Sin embargo, llama la atención que también se observan cambios a la información de días de abril y de mayo, sobre todo en el caso de las muertes.

³ La gráfica se presenta a partir del 18 de marzo, ya que es la fecha en que se reportan las primeras muertes por COVID-19 en el país. Tanto Durango como el Estado de México reportan una muerte esa fecha.

Gráfica 1.

Evolución de los cambios en las mediciones del COVID-19 en México



Fuente: Elaboración propia con datos de la Secretaría de Salud (2020).

Nota: Los cambios se calculan comparando la información diaria en las bases del 15 de junio y del 17 de agosto.

Una razón que podría contribuir a explicar los cambios en la medición de casos y muertes es el tiempo que pasa entre la realización de la prueba para detectar la presencia de COVID-19 y la obtención del resultado⁴. De hecho, en la gráfica se puede calcular que 76.3% de los cambios observados en los contagios, y 52.5% en las muertes, se concentran en los once días más cercanos al 15 de junio.

En el Cuadro 1, se presenta cómo se distribuyen las 32 entidades del país en términos de la diferencia acumulada al 15 de junio tanto en contagios como en muertes con base en la comparación de los datos presentados el 15 de junio y el 17 de agosto. Para controlar por el tamaño de población y poder realizar comparaciones más adecuadas entre estados, las diferencias acumuladas observadas se calculan por cada millón de habitantes. Se construyen cuatro rangos tomando como referencia el valor mediano de la diferencia acumulada en cada

⁴ En general, el cambio en la información se puede deber al tiempo que transcurre entre el inicio de los síntomas, la consulta, la toma de muestra, el traslado de la muestra al laboratorio, el análisis de la muestra, la obtención de los resultados y su reporte a la Secretaría de Salud.

indicador. En términos de contagios, son cinco entidades las que presentan una diferencia que es menor a la mitad del valor mediano, mientras que ocho estados presentan un cambio en muertes relativamente pequeño. Cabe señalar que Colima, Querétaro y Zacatecas reportan diferencias bajas en ambas mediciones.

Cuadro 1.

Distribución de entidades por tamaño de la diferencia acumulada (D) en mediciones del COVID-19 en México al 15 de junio

<i>Rango</i>	<i>Diferencia acumulada en contagios</i>	<i>Diferencia acumulada en muertes</i>
D < Mediana/2	5	8
Mediana/2 < D < Mediana	11	8
Mediana < D < Mediana*3/2	7	5
D > Mediana*3/2	9	11

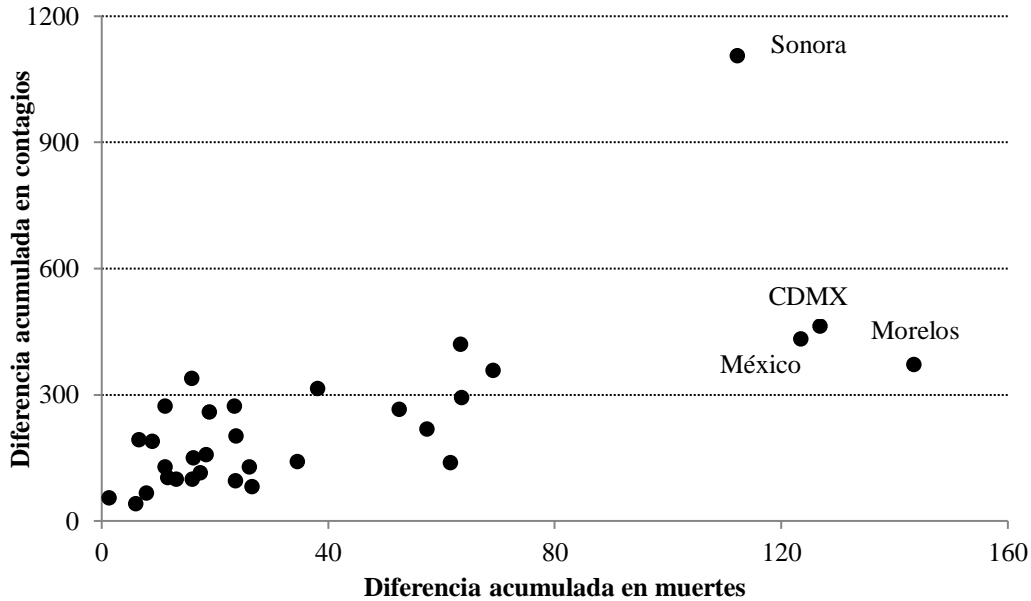
Fuente: Elaboración propia con información de la Secretaría de Salud (2020).

Nota: Las diferencias acumuladas por millón de habitantes se calculan comparando el total acumulado de cada medición al 15 de junio de las bases del 15 de junio y del 17 de agosto. La diferencia se divide entre la población estimada (en millones). La diferencia acumulada mediana de contagios es de 190.69 contagios por cada millón y la de muertes, 23.54 defunciones por millón de habitantes.

El Cuadro 1 muestra que hay más entidades que reportan cambios grandes (mayor a 1.5 veces el valor mediano) en ambas mediciones al 15 de junio, de los cuales ocho (Baja California, Chihuahua, Ciudad de México, Estado de México, Morelos, Sinaloa, Sonora, y Tabasco) coinciden en esta clasificación. Debido a que hay grupos de entidades que pertenecen a los mismos rangos en ambas mediciones, resulta interesante presentar en la Gráfica 2 la correlación que hay entre las diferencias acumuladas que tienen las entidades.

Gráfica 2.

Relación entre las diferencias acumuladas en las mediciones por entidad federativa al 15 de junio



Fuente: Elaboración propia con datos de la Secretaría de Salud (2020).

Nota: Las diferencias acumuladas por millón de habitantes se calculan comparando el total acumulado de cada medición al 15 de junio de las bases reportadas el 15 de junio y el 17 de agosto. La diferencia se divide entre la población estimada (en millones).

En la Gráfica 2, se observa que Sonora es el estado que presenta el mayor cambio en el acumulado de contagios, al tiempo que ocupa la cuarta posición más alta en términos de la diferencia acumulada en muertes, superado por Morelos, Ciudad de México y Estado de México. Cabe indicar que el coeficiente de correlación entre ambas diferencias al considerar todas las entidades se ubica en 0.692.

2. Las distorsiones en mediciones de COVID-19 y sus cofactores.

Seoane (2020) indica que la subestimación en el número oficial de casos de COVID-19 está en función de la capacidad de cada país para detectarlos y se debe a que típicamente se detectan aquellos casos que presentan una sintomatología que rebasa la indicada por las autoridades de cada país. Adicionalmente, menciona dos razones por las que las estadísticas relacionadas con muertes causadas por este virus también son defectuosas e incompletas: (1) solo se contabilizan si el fallecido es un contagio confirmado y (2) la mayoría de los países

considera solamente las muertes que ocurren en hospitales. Por otro lado, dado que en la sección 2 se presenta evidencia de que la mayor parte de los cambios se concentra en las fechas cercanas al reporte inicial, se podría apoyar la idea de que estos cambios son una actualización de la base de datos explicada por el tiempo que tardan en confirmarse los resultados de las pruebas de COVID-19. Sin embargo, dos hechos sugieren la posibilidad de que estos cambios sean una decisión estratégica por parte de algunas entidades: los cambios a información de días muy distantes del día del reporte y la alta correlación entre las diferencias acumuladas de ambas medidas (Gráfica 2). Otro argumento que fortalece la idea del movimiento estratégico en la información es que las estrategias gubernamentales a seguir en la nueva normalidad están basadas, entre otras cosas, en la evolución que presenten los contagios y muertes.

A pesar de la amplia literatura que se ha publicado con respecto al COVID-19, no hemos localizado trabajos que traten la posible decisión estratégica por parte de las autoridades locales a distorsionar el flujo de información acerca de este virus. En ese sentido, este trabajo es un primer intento por encontrar factores que se relacionen con los cambios que han reportado los gobiernos estatales durante los primeros meses de la pandemia. Por ello, en esta investigación denominamos distorsión en la medición a los cambios en la información reportada de casos confirmados o de muertes que se hacen en fechas muy anteriores. Con el propósito de ver si las decisiones estratégicas para distorsionar la información varían con el tiempo, se definen tres periodos en los que se estima la distorsión en las mediciones del COVID-19 de cada entidad. Para cada periodo, se realizan pruebas para ver cuáles estados muestran magnitudes similares en la distorsión de mediciones. Adicionalmente, se pone a prueba si las distorsiones que presentan las entidades en las mediciones del COVID-19 las realizan de manera estratégica; para ello se verifica si están asociadas con los subíndices utilizados por el IMCO para elaborar el índice de competitividad estatal 2020.

Para construir las medidas de distorsiones, tomamos las bases de datos de los lunes 11 de mayo, 15 de junio, 13 de julio y 17 de agosto publicadas por la Secretaría de Salud (2020). Dado que el interés es en las modificaciones hechas a la información de días previos muy distantes, no se consideran los cambios realizados en los 17 días más cercanos a la fecha de la base de datos. Así, comparamos el número de contagios y muertes acumulados a los

últimos viernes de los meses de abril, mayo y junio de diferentes bases de datos. El Cuadro 2 sintetiza la manera en cómo se generan las medidas de distorsión. Al estar interesados en ver si la distorsión puede ser explicada por factores diferentes a lo largo de los diferentes periodos de análisis, se calculan las distorsiones acumuladas al 24 de abril, las que prevalecen entre el 25 de abril y el 29 de mayo y las que hay entre el 30 de mayo y el 26 de junio. En este sentido, el primer periodo de análisis coincide con las fechas iniciales en que se suspendieron las actividades no esenciales en el país (del 30 de marzo al 30 de abril) y el segundo periodo abarca los días que dicha suspensión fue ampliada (del 1 al 30 de mayo). Por su parte, el tercer periodo coincide con los primeros días de la reapertura de las actividades económicas (la nueva normalidad).

Cuadro 2.

Construcción de las distorsiones en mediciones del COVID-19

Periodo considerado	Fecha de base inicial	Fecha de base final
Al 24 de abril	11 de mayo	15 de junio
Del 25 de abril al 29 de mayo	15 de junio	13 de julio
Del 30 de mayo al 26 de junio	13 de julio	17 de agosto

Fuente: Elaboración propia con información de la Secretaría de Salud (2020).

Para controlar por las diferencias poblacionales entre entidades, los cálculos de las distorsiones se normalizan por cada millón de habitantes. El Cuadro 3 presenta el promedio y la desviación estándar de las distorsiones en las mediciones para cada periodo. Se puede ver que, en el inicio de la pandemia, la distorsión de ambas medidas es pequeña. En términos de muertes, el mayor promedio se observa en el periodo entre el 25 de abril y el 29 de mayo y se ubica en 9.93 defunciones por cada millón de habitantes, mientras que, para los contagios, el promedio más alto se presenta en el tercer periodo, con 22.54 casos por cada millón de habitantes, aunque la cifra es muy similar a la del periodo previo.

Cuadro 3.

Promedio y desviación estándar de distorsiones en mediciones del COVID-19

Periodo	En contagios		En muertes	
	Promedio	Desv. Est.	Promedio	Desv. Est.
Al 24 de abril	3.28	7.62	1.29	2.49
Del 25 de abril al 29 de mayo	22.29	41.66	9.93	18.21
Del 30 de mayo al 26 de junio	22.54	38.56	7.11	8.62

Fuente: Elaboración propia con información de la Secretaría de Salud (2020).

Al analizar las distorsiones de los contagios, se observa que el primer periodo es el único en el cual hay entidades que no modifican su información de días muy anteriores; esto es, no hay distorsión en los reportes que emiten Baja California Sur, Colima, Durango y Tlaxcala. Por otro lado, es interesante observar que en los tres periodos hay algunas entidades que reportan distorsiones negativas en contagios (menos contagios acumulados en la segunda base de datos que en la primera), lo cual no ocurre en las distorsiones en muertes. Así mismo, en los tres periodos hay entidades que no distorsionan su información de defunciones: 14 en el primer periodo, 4 en el segundo y solo 2 no modifican su información en el tercero.

2.1. Análisis de clúster

La Secretaría de Salud ha argumentado que existen rezagos y distorsiones en los datos que los diversos estados reportan al gobierno federal. Ni el flujo de información ni las distorsiones en la misma son homogéneos entre entidades. Existen diferencias significativas relacionadas con la magnitud de la distorsión de medición. Por ejemplo, mientras que al 24 de abril Morelos muestra una diferencia negativa en casos confirmados, -5.8 casos por millón de habitantes, la Ciudad de México reporta una distorsión de 27.7 contagios por cada millón. Similarmente, Chiapas y Nuevo León presentan valores bajos de distorsión en defunciones, al tiempo que Baja California y Ciudad de México tienen distorsiones significativas a finales del mes de abril.

Con la finalidad de identificar cuáles estados muestran magnitudes similares en la distorsión de ambas mediciones para cada uno de los tres periodos incluidos en este documento, se realiza un análisis de *clúster* jerárquico. Esta metodología permite agrupar los datos basándose en la distancia que hay entre ellos, buscando que los datos dentro de un mismo grupo o *clúster* sean lo más similares posibles entre sí. Este análisis se realiza utilizando la matriz de distancia euclidiana clásica tanto para las distorsiones en casos confirmados como para las defunciones de las 32 entidades federativas. El índice de agregación de Ward es usado para estimar la similitud entre grupos de estados a clasificar. El análisis de *clúster* jerárquico se realiza para cada uno de los tres periodos considerados.

La Gráfica 3 presenta el dendograma que sintetiza los resultados del ejercicio para el periodo al 24 de abril⁵. En particular, el eje vertical muestra la distancia que existe entre cada una de las observaciones o grupos de acuerdo con las distorsiones, mientras que el eje horizontal indica las entidades que tienden a ser similares. Cabe señalar que generalmente el análisis de *clúster* requiere que se defina el número de grupos de manera exógena y un poco arbitraria. Asimismo, existen diversos indicadores que permiten obtener el número de *clústeres* óptimos de manera endógena. Sin embargo, cada uno de estos indicadores puede arrojar diferentes números de agrupamientos. Con la finalidad de seleccionar el número óptimo de *clústeres*, se estimaron diversos indicadores. El Cuadro A1 del Anexo 1 resume los resultados del número de grupos seleccionado de acuerdo con el indicador utilizado para cada uno de los periodos analizados en este estudio. En particular, la última fila de este cuadro indica el número óptimo de *clústeres*. Este número se selecciona de acuerdo con la regla de la mayoría; es decir, del conjunto de indicadores de *clúster* óptimo mostrados en ese cuadro⁶, se elige como óptimo el consenso de la mayoría. Para este periodo (24 de abril), el número óptimo de *clústeres* es 2, de manera que, como se observa en la gráfica, Chihuahua, Baja California y Ciudad de México se agrupan en un *clúster*, mientras que el resto de las entidades conforman el grupo 1.

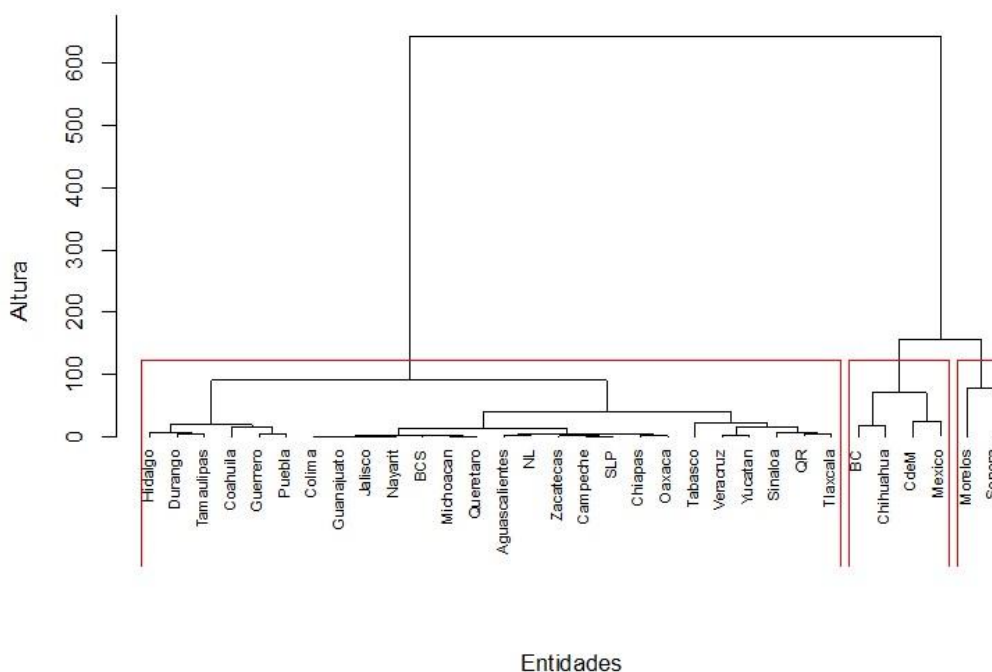
⁵ Un dendograma es una representación gráfica forma de árbol que organiza los datos en subcategorías que se van dividiendo en otros hasta llegar al nivel de detalle deseado.

⁶ El conjunto de indicadores son los utilizados por Charrad *et al.* (2014).

conjunto, Baja California, Chihuahua, Ciudad de México y Estado de México conforma otro grupo, mientras que el resto se agrupan en otro *clúster*⁷. Morelos y Sonora tienen la mayor distorsión promedio en ambas mediciones, 152.4 casos confirmados y 49.6 defunciones, durante la etapa en la que se extendió la suspensión de actividades no esenciales. El segundo grupo con mayor promedio lo integran Baja California, Chihuahua, Ciudad de México y Estado de México, siendo de 69.8 en casos confirmados y de 30.6 en defunciones. El *clúster* conformado por el resto de las entidades muestra una distorsión promedio sustancialmente más baja en ambas mediciones, 4.9 para casos confirmados y 3.7 para defunciones.

Gráfica 4.

Dendograma para clúster jerárquico de las distorsiones en medición (entre el 25 de abril y el 29 de mayo)



Fuente: Elaboración propia con datos de la Secretaría de Salud (2020)

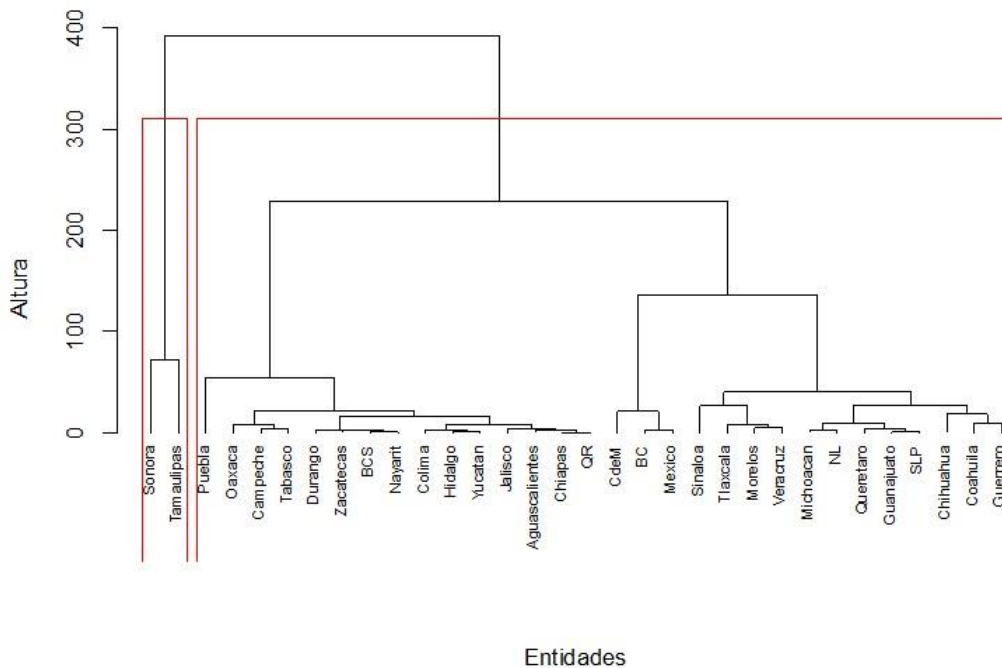
En relación con el análisis de *clúster* jerárquico para el tercer periodo de este estudio, el cual abarca la fecha en la que se dan los pasos para la nueva normalidad, el conjunto de indicadores señala que el número óptimo de *clústeres* es 2. La Gráfica 5 muestra el

⁷ En la Gráfica 4, se puede apreciar que en caso de que el número de *clústeres* se fijara en dos, uno de los grupos se conformaría por las seis entidades mencionadas previamente.

dendograma que resume los resultados de este análisis. Uno de los grupos se conforma por Sonora y Tamaulipas (*clúster 2*) mientras que el resto de las entidades se agrupan en el *clúster 1*. De acuerdo con el valor del promedio de las distorsiones de ambos grupos, el *clúster* conformado por Sonora y Tamaulipas presenta valores mayores ambas mediciones: una distorsión promedio de 145 casos confirmados y 28 defunciones por cada millón de habitantes. Por su parte, los promedios para el resto de las entidades se ubican, de manera respectiva, en 14 y 5. Resulta relevante observar que Baja California, Chihuahua y Ciudad de México dejan de distinguirse como entidades con mayores distorsiones una vez que comienza la nueva normalidad.

Gráfica 5.

Dendograma para clúster jerárquico de las distorsiones en medición (entre el 30 de mayo y el 26 de junio)



Fuente: Elaboración propia con datos de la Secretaría de Salud (2020)

En un escenario adicional, se agrupan las distorsiones de los dos primeros periodos en uno solo; esto es, las distorsiones acumuladas al 29 de mayo, justo antes de comenzar la nueva normalidad. En este contexto, la Ciudad de México pertenece a un *clúster* diferente del resto

de las entidades, caracterizándose por tener mayores distorsiones en muertes y contagios, a pesar de estar normalizado por cada millón de habitantes. Este resultado va en línea con lo encontrado por Zavala y Despeghe (2020) quienes, con base en los registros de defunciones de la Ciudad de México, argumentan que el número de muertes relacionado con la actual pandemia está sub-reportado en las cifras oficiales del gobierno capitalino.

Los resultados anteriores muestran que durante la primera ola de la pandemia algunos estados distorsionaron más su información relativa al COVID-19 que otros. Con el fin de saber si las entidades que reportaron más distorsiones en la medición de defunciones a principios de la pandemia son también los estados que reportan más subregistros de muertes, se estimó el exceso de muertes para cada entidad federativa. Para estimar este exceso de defunciones se utiliza el número de defunciones registradas entre enero y agosto del 2020 con base en los registros administrativos de defunciones. Esta información la publicó el INEGI a principios del año 2021⁸. Para estimar el porcentaje de exceso de defunciones y siguiendo la metodología de INEGI, se utiliza la siguiente definición:

$$\text{Exceso de Muertes} = \left(\frac{\text{Muertes observadas}_{2020}}{\text{Muertes esperadas}_{2020}} - 1 \right) * 100$$

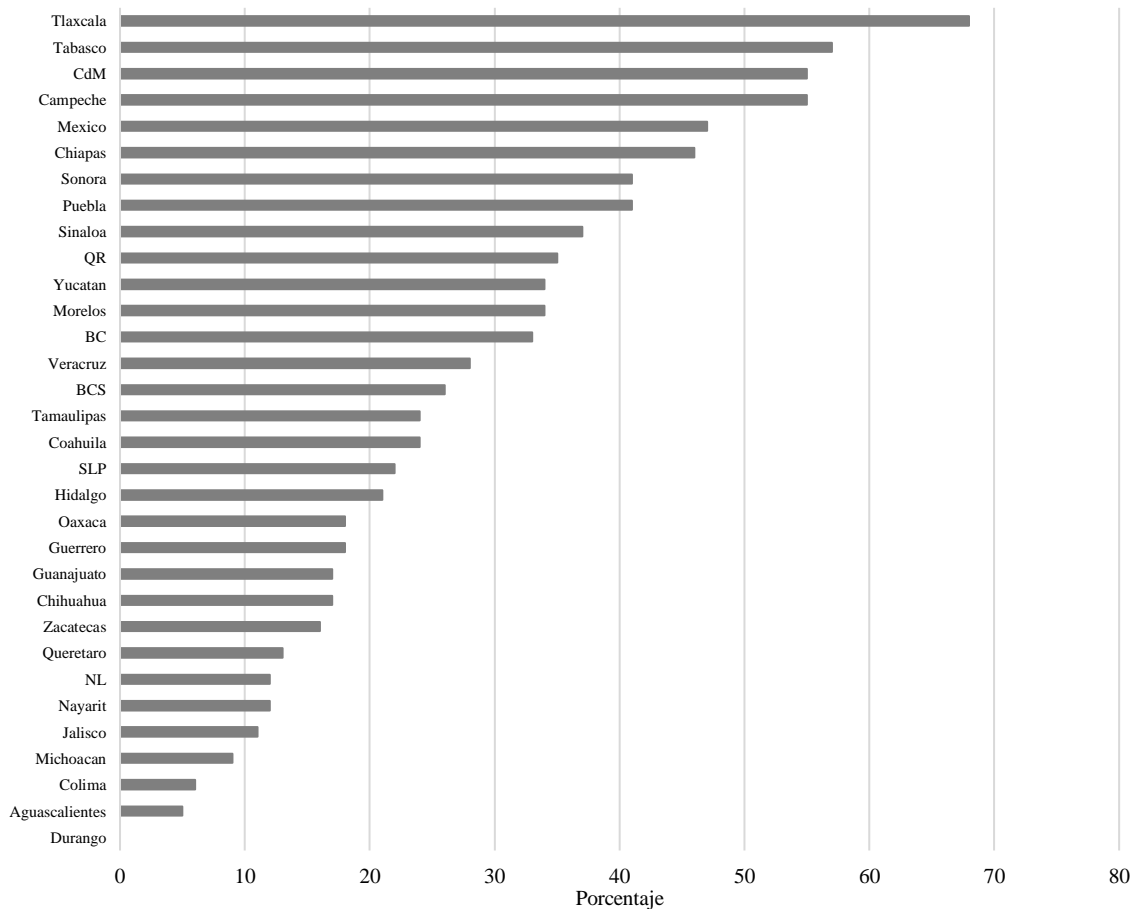
Donde las muertes observadas son las muertes reportadas con base en los registros administrativos y las muertes esperadas son las pronosticadas con base en los datos anuales desde 1990 hasta 2019 que publica también el INEGI. Los pronósticos se realizaron utilizando modelos ARIMA para cada una de las entidades federativas ya que esta metodología permite calcular el número de muertes para el año 2020 dada información hasta el 2019, año en el cual no había ningún caso registrado de COVID-19. Para hacer comparable el valor esperado con el observado, el pronóstico se multiplicó por 8/12 con el fin de obtener el número de muertes esperadas hasta el mes de agosto. Cabe señalar que nuestra definición de distorsión no es igual al sub-registro mostrado por el exceso de muertes del INEGI. Sub-registros se refieren a casos de defunciones no contabilizados mientras que en nuestro estudio se analiza el cambio de información de los casos contabilizados en fechas muy distantes al

⁸ INEGI (2021). “Características de las defunciones registradas en México durante enero a agosto de 2020”. <https://www.inegi.org.mx/app/saladeprensa/noticia.html?id=6244>

día del reporte, definimos como distorsión. Por lo tanto, las entidades que más distorsionan no son necesariamente las que muestran más sub-registro de defunciones. La Gráfica 6 muestra el exceso de muertes por entidad federativa de acuerdo con el indicador estimado. Como se observa en la gráfica, tres de los primeros seis lugares de las entidades con mayor exceso de muertes, Ciudad de México, Estado de México y Sonora, se encuentran también entre los que tienen mayores distorsiones en las defunciones en alguno de los tres periodos contemplados en este estudio. Contrariamente, Chihuahua y Baja California, que pertenecen al *clúster* con mayores distorsiones promedio en dos de los tres periodos, presentan porcentajes de excesos de muertes inferiores al 35%, mientras que los cuatro estados con mayor subregistro tienen un exceso de defunciones superior al 50%.

Gráfica 6.

Exceso de Defunciones por Entidad Federativa (Porcentaje)



Fuente: Elaboración propia con datos de defunciones registradas por entidad federativa publicados en el INEGI (2021).

2.2. El comportamiento estratégico de las distorsiones y los subíndices del IMCO

Para determinar si las distorsiones que presentan las entidades en las mediciones del COVID-19 las realizan de manera estratégica, se consideran los 10 subíndices utilizados por el IMCO para elaborar el índice de competitividad estatal 2020. Dado que el IMCO define competitividad como la capacidad de generar, atraer y retener talento e inversiones, se considera apropiado utilizar la información de IMCO (2020) para ver si las distorsiones en las mediciones del COVID-19 están asociadas con algunos de los subíndices considerados para determinar la competitividad de las entidades.

Los subíndices identifican 10 dimensiones que se asocian con la competitividad que tiene cada estado. De acuerdo con IMCO (2020), estas dimensiones surgen de la teoría económica, la experiencia internacional y el sentido común. Cada subíndice se genera de un conjunto de indicadores y, en total, se cuenta con 97 indicadores para cada una de las entidades que son publicados regularmente en fuentes neutrales de prestigio. Los subíndices que considera el IMCO son: (1) sistema de derecho confiable y objetivo, (2) manejo sustentable del medio ambiente, (3) sociedad incluyente, preparada y sana, (4) sistema político estable y funcional, (5) gobiernos eficientes y eficaces, (6) mercado de factores eficientes, (7) economía estable, (8) sectores precursores de clase mundial, (9) aprovechamiento de las relaciones internacionales y (10) innovación y sofisticación en los sectores económicos. De esta manera, se tiene la hipótesis de que las magnitudes de las distorsiones hechas por las entidades en los primeros meses de la pandemia están asociadas con estas 10 áreas con el fin de no limitar su competitividad.

Para analizar los posibles efectos de los 10 subíndices presentados por el IMCO sobre las distorsiones, y dado que se cuenta solo con 32 observaciones por periodo a nivel estatal, primero se analiza el efecto independiente que tiene cada subíndice en la distorsión de cada periodo. El Cuadro 4 presenta los signos de los subíndices cuyos coeficientes son significativos al realizar regresiones simples. Se observa que, salvo por el subíndice de sistema de derecho confiable y objetivo, los coeficientes son positivos, indicando que mejores posicionamientos en esas dimensiones por parte de los estados se asocian con mayores distorsiones en sus mediciones. En el caso del sistema de derecho, la relación es

inversa; es decir, entidades mejor posicionadas en esta dimensión presentan menores distorsiones en muertes para los dos primeros periodos.

Cuadro 4.
Subíndices del IMCO significativos para explicar distorsiones en mediciones

Subíndices	Periodo 1		Periodo 2		Periodo 3	
	Contagios	Muertes	Contagios	Muertes	Contagios	Muertes
Sistema de derecho confiable y objetivo		-		-		
Manejo sustentable del medio ambiente		+				
Sociedad incluyente, preparada y sana						+
Economía estable	+	+				+
Sectores precursores de clase mundial	+	+		+		+
Aprovechamiento de las relaciones internacionales	+					
Innovación y sofisticación en los sectores económicos				+		+

Fuente: Elaboración propia con información de la Secretaría de Salud (2020).

Resulta interesante observar que los subíndices que tratan las dimensiones de sistema político estable y funcional, gobiernos eficientes y eficaces, y mercado de factores eficientes nunca resultan ser estadísticamente significativos. Del Cuadro 4, resalta que cuatro subíndices están relacionados con las distorsiones en muertes del primer periodo y otros cuatro con las de contagios del tercero. Por otro lado, ninguno de los subíndices contribuye a explicar la distorsión en muertes para el periodo entre el 30 de mayo y el 26 de junio, el cual abarca el inicio de la nueva normalidad.

Una vez que se sabe cuáles de los subíndices estimados por el IMCO influyen sobre las distorsiones y dada la restricción en el número de observaciones, a cada tipo de distorsión y periodo se le aplica el método hacia atrás de eliminación de variables (*backward stepwise*), fijando como criterio de eliminación un nivel de significancia de 10%. El Cuadro 5 muestra los resultados de las estimaciones finales. Por ejemplo, para el caso de las distorsiones en contagios en el primer periodo, de los tres subíndices que, de manera independiente, son significativos, solo el relacionado con sectores precursores de clase mundial mantiene su significancia. Asimismo, se observa que solo tres de los siete subíndices que explicaban de manera individual alguna distorsión siguen siendo estadísticamente significativos. De esta manera, las distorsiones en contagios observadas en los periodos uno (al inicio de la pandemia) y tres (al inicio de la nueva normalidad) se asocian positivamente con el subíndice de sectores precursores de clase mundial, mientras el subíndice que explica la distorsión en el periodo en el que se extendió la suspensión de actividades no esenciales es el de innovación y sofisticación en los sectores económicos.

Cuadro 5.

Subíndices del IMCO determinantes de las distorsiones en mediciones

	Periodo 1		Periodo 2		Periodo 3	
	Contagios	Muertes	Contagios	Muertes	Contagios	Muertes
Constante	-8.723** (4.053)	0.517 (1.973)	-31.180 (21.150)	48.930** (18.530)	-16.740* (9.673)	7.107*** (1.523)
Derecho		-0.061** (0.025)		-0.671** (0.279)		
Precursores	0.324** (0.118)	0.116*** (0.0294)			1.060*** (0.336)	
Innovación			1.245** (0.591)			
Observaciones	32	32	32	32	32	32
R ²	0.268	0.522	0.218	0.262	0.112	0.000
F	7.485	12.08	4.443	5.770	9.945	0

Fuente: Elaboración propia con información de la Secretaría de Salud (2020).

Errores estándares robustos entre paréntesis. *** p<0.01, ** p<0.05, * p<0.10.

Las distorsiones en muertes en el inicio de la pandemia (primer periodo) son las únicas que presentan dos cofactores cuando se utiliza el método hacia atrás de eliminación de variables. La relación con el subíndice sistema de derecho confiable es negativa, mientras que el subíndice sectores precursores de clase mundial presenta una asociación positiva. Las distorsiones del segundo periodo, que coincide con las fechas en las que se extendió la suspensión de actividades no esenciales, también se relacionan negativamente con el subíndice del sistema de derecho, mientras que, como ya se había mencionado, ningún subíndice contribuye para explicar las distorsiones una vez que empieza la nueva normalidad (tercer periodo).

De acuerdo con IMCO (2020), el subíndice sectores precursores de clase mundial incluye indicadores de los sectores financiero, telecomunicaciones y transporte e indica que el buen funcionamiento de estos sectores es una condición necesaria para impulsar el crecimiento económico, la inversión y la generación de empleo. Por su parte, el subíndice innovación y sofisticación en los sectores económicos mide la capacidad de las entidades federativas para competir con éxito en la economía, pues al contar con sectores económicos más innovadores se es capaz de atraer y retener más talento e inversión. En este sentido, las mayores distorsiones en contagios ocurren en entidades con mejores indicadores en algunos sectores económicos, entre los que están el financiero, el de telecomunicaciones y el transporte. Así, estos estados en un primer reporte indicaron pocos contagios quizá con el propósito de no incidir negativamente en esos sectores económicos pensando en que la pandemia acabaría pronto. Para el caso de las distorsiones en muertes, el comportamiento estratégico ocurre en entidades cuyo estado de derecho no es funcional.

El Anexo 2 analiza si las magnitudes en la actualización (cambios a la información en fechas cercanas al reporte inicial) durante la primera ola de la pandemia están asociadas con los subíndices del IMCO. La comparación de los resultados para explicar las distorsiones y las actualizaciones muestra que los subíndices que explican los cambios en las mediciones en el primer periodo son los mismos para ambos tipos de modificación. Sin embargo, en los otros dos periodos, los subíndices que explican las distorsiones no contribuyen en explicar las actualizaciones. De igual manera, las actualizaciones están relacionadas con dos subíndices que no contribuyen a explicar las distorsiones.

Un ejercicio adicional consiste en analizar cuáles de los subíndices explican las distorsiones (normalizadas por millón de habitantes) acumuladas en todo el periodo; es decir, las diferencias acumuladas en contagios y muertes al 26 de junio de las bases de datos del 13 de julio y del 17 de agosto. Siguiendo una estrategia similar al análisis previo, los modelos de regresión simple indican que las distorsiones en contagios se pueden explicar por los subíndices de manejo sustentable del medio ambiente, sociedad incluyente, preparada y sana, economía estable, sectores precursores de clase mundial, o el de innovación y sofisticación en los sectores económicos. Por su parte, la dimensión de sistema de derecho confiable y objetivo o el subíndice de sectores precursores de clase mundial son estadísticamente significativos para explicar las distorsiones en muertes por cada millón de habitantes.

Cuadro 6.

Determinantes de las distorsiones en mediciones acumuladas al 26 de junio

	(A)	(B)	(C)	(D)
	Contagios	Muertes	Contagios	Muertes
Constante	-68.36** (25.18)	38.45 (25.14)	-67.798** (25.763)	24.997 (19.840)
Derecho		-0.672** (0.303)		-0.473* (0.249)
Precursores	1.200* (0.685)	0.511** (0.238)	1.121 (0.729)	0.463** (0.208)
Innovación	1.676* (0.917)		1.651* (0.902)	
Afín			15.838 (26.088)	16.774* (9.831)
Observaciones	32	32	32	32
R ²	0.233	0.288	0.241	0.366
F	7.432	10.40	8.28	12.47

Fuente: Elaboración propia con información de la Secretaría de Salud (2020).

Errores estándares robustos entre paréntesis. *** p<0.01, ** p<0.05, * p<0.10.

Las columnas (A) y (B) del Cuadro 6 presentan las estimaciones finales de aplicar el método hacia atrás de eliminación de variables que utiliza como criterio de eliminación un nivel de significancia de 10%. Resulta interesante observar que los subíndices sectores precursores de clase mundial e innovación y sofisticación en los sectores económicos se mantienen como los determinantes de las distorsiones acumuladas en contagios por cada millón de habitantes. Por su parte, las distorsiones acumuladas en muertes se relacionan negativamente con el subíndice sistema de derecho confiable y objetivo, pero positivamente con el de sectores precursores de clase mundial. Estos resultados confirman que las entidades del país estratégicamente difirieron el flujo de la información relacionada con el COVID-19 durante la primera ola de la pandemia para tratar de no incidir negativamente en determinados sectores económicos.

Debido a que los valores de la R^2 son bajos, se infiere que existen otros factores que pueden contribuir a explicar las distorsiones en las mediciones del COVID-19 en nuestro país. De los análisis de la subsección anterior se observa que algunas de las entidades que reportan mayores distorsiones están gobernadas por el mismo partido político que el ejecutivo federal. De esta manera, las columnas (C) y (D) del Cuadro 6 incluyen como posible determinante una variable dicotómica (Afin) que toma el valor de 1 si el partido político que gobierna la entidad es del mismo partido político que el ejecutivo federal. Se observa que, para el caso de la distorsión en contagios, la inclusión de esta variable, además de no ser significativa, causa que el subíndice de sectores precursores de clase mundial también deje de serlo. Sin embargo, esta variable sí contribuye a explicar la distorsión en muertes de las entidades del país en el periodo analizado. Con base en las estimaciones, entidades que son gobernadas por partidos políticos afines al del ejecutivo federal distorsionaron la información de muertes por COVID-19 en 16.8 unidades por cada millón de habitantes más que el resto de los estados.

3. Análisis de cointegración.

Con la finalidad de corroborar los resultados previos relacionados con la identificación de entidades que muestran una distorsión de casos o muertes causadas por este virus, se realizan pruebas de cointegración entre estas variables. Los datos de las variables utilizadas provienen

de la actualización presentada el 17 de agosto del 2020 por parte de la Secretaría de Salud, pero se utiliza la información que cubre del 17 de marzo al 2 de agosto del 2020 para cada una de las 32 entidades federativas.

Dentro de la literatura de la macroeconomía y finanzas, el concepto de cointegración ha sido ampliamente utilizado para caracterizar las relaciones de largo plazo entre variables⁹. Se dice que cuando dos variables están cointegradas existe una relación de largo plazo que hace que estas dos series no se puedan apartar mucho una de la otra ni de su equilibrio porque las fuerzas del mercado hacen que se restablezca dicho equilibrio. Generalmente, en macroeconomía estas relaciones de largo plazo implican un periodo prolongado de tiempo. Sin embargo, en el área de Finanzas estudios entre los que se encuentran los de Bentes (2015) y Clinet y Potiron (2021) utilizan datos de alta frecuencia ya que largo plazo puede implicar un periodo de días, horas e incluso algunos minutos ya que las fuerzas del mercado actúan rápidamente para restablecer el equilibrio.

En el contexto de la pandemia por COVID-19, es de esperarse que las series de casos confirmados y de defunciones no se aparten mucho en el largo plazo ya que la dinámica de la enfermedad tendría que encargarse de restablecer el equilibrio. De acuerdo con el Ministerio de Sanidad Español, el coronavirus tarda entre dos y ocho semanas en provocar la muerte¹⁰. Así, el largo plazo podría implicar un par de semanas ya que es el tiempo que se estima tardan los pacientes con diagnóstico positivo a COVID-19 para mostrar complicaciones relacionadas con dicha enfermedad. Una razón por la cual las dos series pueden diferir en el largo plazo, no estando cointegradas, es un problema de sub-registro de las defunciones, los contagios o de ambas mediciones ya que sería poco probable que el número de defunciones atribuidas al coronavirus se redujera en periodos de incrementos significativos de casos.

Una de las condiciones para que exista cointegración entre dos variables es que ambas series tengan el mismo orden de integración, el cual debe ser mayor o igual a uno. Para determinar el grado de integración de las variables se realizan pruebas de raíces unitarias para

⁹ Entre muchos, se pueden mencionar a Christopoulos y Tsionas (2004) y a Moosa y Vaz (2016).

¹⁰ https://pnsd.sanidad.gob.es/noticiasEventos/actualidad/2020_Coronavirus/pdf/20200421_INFORME_CIENTIFICO_SANIDAD_COVID-19.pdf

cada una de las variables ya que estas pruebas permiten saber si las variables son integradas de orden 0, 1 o 2. El Cuadro A6 del Anexo 3 muestra los resultados de la prueba de Dickey Fuller Aumentada (DFA) de raíces unitarias para el logaritmo del número de casos confirmados y defunciones por COVID-19, en niveles, primeras y segundas diferencias, para el agregado nacional y las 32 entidades federativas. La hipótesis nula de esta prueba establece que la serie tiene una raíz unitaria. Como se puede observar en ese cuadro, solo en el caso del número de muertes para el Estado de México y San Luis Potosí se rechaza la hipótesis nula de presencia de una raíz unitaria a un nivel de significancia del 1 por ciento. En el resto de las entidades, con excepción Sonora y Veracruz, las series de confirmados y defunciones muestran evidencia de ser integradas de orden 1 ya que las series no presentaron evidencia de una raíz unitaria en primeras diferencias pero sí en niveles¹¹. Los resultados para las series de interés de Sonora y Veracruz sugieren que están integradas de orden 2, ya que se requiere calcular segundas diferencias para que estas variables no muestren la presencia de una raíz unitaria (ver cuadro A6, columnas 8 y 9).

Engle y Granger (1987) proponen una prueba para comprobar estadísticamente la existencia de una relación de cointegración entre dos variables. En particular, esta prueba consiste en dos fases. En la primera se estima el vector de cointegración entre las variables integradas de orden uno o superior. Para estimar este vector de cointegración, Stock y Watson (1993) sugieren el método de mínimos cuadrados dinámicos (MCD), el cual, además de incluir las variables cointegradas y variables determinísticas, incluye rezagos y adelantos de las diferencias de las variables regresoras. El vector de cointegración está basado en la siguiente especificación:

$$d_t = \alpha + \delta t + \beta c_t + \sum_{j=-p}^p \theta_j \Delta c_{t-j} + \varepsilon_t$$

Donde d es el logaritmo del número de defunciones diarias, c es el número de casos confirmados diarios, t es la tendencia y ε , el término del error.

¹¹ En algunos de los casos se rechaza la hipótesis nula de raíz unitaria al 10 o al 5% pero no al 1% por lo que en estos casos se opta por considerar que la serie era no estacionaria.

En la segunda fase se verifica que los residuales calculados utilizando los coeficientes del vector de cointegración estimados en la primera fase no presenten una raíz unitaria. Para esta verificación se utiliza la prueba de DFA descrita anteriormente. Bajo la hipótesis nula no existe cointegración entre las variables por lo que el rechazo de esta hipótesis sugeriría la presencia de un vector de cointegración. La evidencia de no cointegración podría atribuirse a la distorsión de los datos, al sub-reporte de cifras, entre otros factores. Cabe indicar que esta prueba no es posible realizarla para el Estado de México y San Luís Potosí ya que sus series de defunciones tienen un grado de integración diferente a las de los casos confirmados. El Cuadro 7 presenta la clasificación de los estados de acuerdo con la evidencia mostrada por las pruebas de cointegración de Engle-Granger para las 30 entidades y el agregado nacional¹². La primera columna muestra el conjunto de entidades en las cuales se encuentra evidencia de cointegración al 1% de significancia. La segunda columna muestra las entidades en las cuales existe evidencia de relación de largo plazo al 5% y 10%, pero no al 1%. Finalmente, la columna 3 muestra las entidades en las cuales no existe evidencia de cointegración. Cabe destacar que este cuadro muestra el p-valor asociado con la prueba de cointegración en paréntesis.

Cuadro 7.

Clasificación de entidades de acuerdo con las pruebas de cointegración entre las defunciones y los casos de COVID-19 en México

Entidades Federativas		
Fuerte evidencia (al 1%)	Evidencia inconclusa (significancia 5% y 10%)	Sin evidencia
Chihuahua (0.00)	Aguascalientes (0.06)	Baja California Sur (0.18)
Colima (0.00)	Baja California (0.03)	Chiapas (0.11)
Ciudad de México (0.00)	Campeche (0.02)	Coahuila (0.15)
Durango (0.00)	Guanajuato (0.02)	Sonora (0.21)
Guerrero (0.00)	Hidalgo (0.02)	Tabasco (0.12)

¹² Para la estimación de MCD se seleccionan el número de adelantos y rezagos utilizando el criterio BIC. Para la segunda fase de la prueba se incluye un intercepto y/o una tendencia de acuerdo con su significancia. El número óptimo de rezagos para esta fase se selecciona de acuerdo con el criterio MIAC.

Michoacán (0.00)	Jalisco (0.06)	Veracruz (0.17)
Oaxaca (0.00)	Morelos (0.01)	
Puebla (0.00)	Nayarit (0.05)	
Querétaro (0.00)	Nuevo León (0.02)	
Sinaloa (0.00)	Quintana Roo (0.01)	
Tlaxcala (0.00)	Tamaulipas (0.02)	
	Yucatán (0.04)	
	Zacatecas (0.02)	
	Nacional (0.02)	

Fuente: Elaboración propia con base en los resultados del Cuadro A6.

Nota: Entre paréntesis se muestran el p-valor asociado con la prueba de raíces unitarias.

En el Cuadro 7 podemos observar que en once entidades se encuentra una fuerte evidencia de la existencia de una relación de largo plazo entre el número de casos y defunciones, la cual hace que estas dos series no se aparten mucho entre ellas. Este resultado sugeriría que en estas 11 entidades la distorsión en mediciones relacionadas con el COVID-19 es muy pequeña o nula. Sin embargo, el resultado del análisis de *clúster* jerárquico de la sección anterior agrupa a Chihuahua y Ciudad de México en *clústeres* con altas distorsiones en ambas mediciones en los primeros dos periodos. Así, no sorprende que las series del número de confirmados y de defunciones cointegren en estas dos entidades, pero debido a que presentan altas distorsiones en ambas mediciones.

Son seis entidades en las cuales existe evidencia de no cointegración: Baja California Sur, Chiapas, Coahuila, Sonora, Tabasco y Veracruz; es decir, que sus casos de contagios y muertes pueden diferir en el largo plazo. Sonora sobresale por el hecho ser el único estado que se ubica en los *clústeres* más altos de distorsión en mediciones por COVID-19 durante los últimos dos periodos. En el caso de la columna central del Cuadro 7, son 13 entidades, además del agregado nacional, en las cuales existe evidencia de cointegración al 5% y 10% de significancia, pero no del 1%. Esto implica que en estas entidades podría existir o no un problema de distorsión de defunciones y/o casos relacionados con COVID-19. De las 13 entidades, sobresalen Baja California, Morelos y Tamaulipas ya que en la sección 3.1 son

identificados por pertenecer, en alguno de los periodos, a un *clúster* con distorsiones altas en las mediciones.

4. Conclusiones

Tanto las autoridades gubernamentales como algunos investigadores reconocen que la información oficial de contagios confirmados y de muertes por COVID-19 está subestimada. Un problema asociado con la subestimación es que las estrategias públicas que enmarcan la nueva normalidad están basadas en la evolución de las mediciones. De esta manera, y dado que las entidades son las encargadas de proveer la información, además de subestimar es posible que tengan incentivos para distorsionar el flujo de las cifras locales. En este sentido, este trabajo es un intento para analizar la conducta estratégica de las entidades mexicanas con respecto a la difusión de la información de este virus.

Este estudio muestra que, aun normalizando por el tamaño poblacional, la distorsión en las mediciones relacionadas con el COVID-19 no es homogénea entre entidades del país. De hecho, se observa una relación directa entre el tamaño de la distorsión que se hace en una medición y el tamaño de la otra. El análisis de *clúster* jerárquico para cada uno de los tres periodos considerados durante la primera ola indica que ninguna entidad forma parte del *clúster* con mayores distorsiones en los tres periodos. Baja California, Chihuahua y Ciudad de México forman parte en los dos primeros periodos, esto es, antes de comenzar con la nueva normalidad, mientras que Sonora, en los últimos dos; es decir, a partir de que se extiende la suspensión de actividades.

En relación con el objetivo de esta investigación, de los factores que pueden asociarse con el tamaño de las distorsiones se infiere que estados mejor posicionados en ciertos sectores económicos (financiero, telecomunicaciones, transporte y otros más innovadores) inicialmente reportaron pocos contagios quizá con el propósito de no incidir negativamente en esos sectores, pensando en que la pandemia acabaría pronto. Por su parte, en la medida en que el estado de derecho de un estado sea más funcional, las distorsiones en muertes son menores. Adicionalmente, entidades con gobernadores de un partido político afín al del ejecutivo federal retrasan más el flujo de información de defunciones, pero no hay evidencia

de que hayan distorsionado en mayor cuantía la información de contagios relativo al resto de los estados.

El análisis de cointegración muestra que en seis entidades no hay evidencia de una relación de largo plazo entre el número de defunciones y el de casos confirmados, indicando la posibilidad de un problema en los datos de las defunciones, los contagios o ambas mediciones en esas entidades. Al inicio de la pandemia, las autoridades federales indicaron que el pico de la pandemia ocurriría en el mes de mayo. Aunque este pronóstico no fue acertado y extendieron la suspensión de actividades no esenciales, es posible que entidades como Baja California, Chihuahua y Ciudad de México hayan tenido incentivos a retrasar el flujo de información en los inicios de la pandemia, pero una vez que se implementó la nueva normalidad lo dejaron de hacer.

Este trabajo ha mostrado evidencia de posibles acciones estratégicas por parte de las entidades con respecto al flujo de información relacionada con el COVID-19. En la forma organizacional actual, las autoridades locales son las encargadas de proveer la información pero al mismo tiempo esa información es parte fundamental de las políticas que se siguen con el propósito de reactivar actividades. En este sentido, sería conveniente tener un sistema de contraloría independiente a las autoridades locales que permita mejorar los sistemas de mediciones estatales.

Finalmente, dentro de las limitantes del estudio está el hecho de no contar con indicadores de frecuencia mensual para las 32 entidades federativas que permitan incluir otros factores que podrían contribuir a explicar las distorsiones en las mediciones del COVID-19. Similarmente, el estudio está sujeto a tener una muestra de datos de 32 observaciones, correspondientes a las entidades del país. Por otro lado, una posible extensión al trabajo es la utilización de métodos de regresión no lineales con el propósito de incrementar el poder explicativo de los cofactores.

Referencias

- Andrade Rosas, L. A., S. León López, N. Escobar Loza, C. A. Jiménez Bandala (2021). La política monetaria mexicana ante la crisis petrolera y la pandemia de CoVid-19. *Revista del Centro de Investigación de la Universidad La Salle*. Vol. 14, No. 55, Enero-Junio, 2021: 25-64. <http://doi.org/10.26457/recein.v14i55.2646>
- Atkeson, A. (2020). How deadly is COVID-19? Understanding the difficulties with estimation of its fatality rate. *National Bureau of Economic Research*, Working Paper No. 26965.
- Auler, A.C., F.A.M. Cássaro, V.O. da Silva, L.F. Pires. (2020). Evidence that high temperatures and intermediate relative humidity might favor the spread of COVID-19 in tropical climate: A case study for the most affected Brazilian cities, *Science of The Total Environment*. Vol. 729, August, 139090. <https://doi.org/10.1016/j.scitotenv.2020.139090>
- Bentes, Sónia R. (2015). On the Integration of financial markets: How strong is the evidence from five international stock markets. *Physica A: Statistical Mechanics and its Applications*. Volume 429, 1 July 2015, Pages 205-214. <https://doi.org/10.1016/j.physa.2015.02.070>
- Bolaño-Ortiz, T.R., Camargo-Caicedo, Y., Puliafito, S.E., Ruggeri, M.F., Bolaño-Díaz, S., Pascual-Flores, R., Saturno, J., Ibarra-Espinosa, S., Mayol-Bracero, O.L., Torres-Delgado, E., Cereceda-Balic, F. (2020), Spread of SARS-CoV-2 through Latin America and the Caribbean region: a look from its economic conditions, climate and air pollution indicators, *Environmental Research*, <https://doi.org/10.1016/j.envres.2020.109938>
- Borjas, G. (2020). Demographic determinants of testing incidence and COVID-19 infections in New York City neighborhoods. *National Bureau of Economic Research*, Working Paper No. 26952.
- Burki, T. (2020). England and Wales see 20000 excess deaths in care homes. *Lancet* 395: 1602. [https://doi.org/10.1016/S0140-6736\(20\)31199-5](https://doi.org/10.1016/S0140-6736(20)31199-5)
- Charrad M., Ghazzali N., Boiteau V. & Niknafs A. (2014). NbClust: An R Package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, 61(6), 1-36.

- Christopoulos, D. y Tsionas, E. (2004). Financial development and economic growth: evidence from panel unit root and cointegration tests. *Journal of Development Economics*. Vol. 73, <https://doi.org/10.1016/j.jdeveco.2003.03.002>
- Clinet, Simon and Potiron, Y. (2021). Cointegration in high frequency data. *Electronic Journal of Statistics*. Vol. 15 1263–132. <https://doi.org/10.1214/21-EJS1805>
- CONEVAL. (2020). La crisis sanitaria generada por la covid-19 y sus consecuencias económicas ponen en riesgo avances en desarrollo social y puede afectar en mayor medida a grupos vulnerables. Consejo Nacional de Evaluación de la Política de Desarrollo Social. Recuperado: https://www.coneval.org.mx/Evaluacion/IEPSM/Paginas/Politica_Social_COVID-19.aspx (Agosto 17, 2020).
- Desilus, B. (2021). The Modern Money Theory in the Time of COVID-19. *Revista del Centro de Investigación de la Universidad La Salle*. Vol. 14, No. 55, Enero-Junio, 2021: 65-80. <http://doi.org/10.26457/recein.v14i55.2697>
- Engle, R. y Granger, C. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica*. Vol. 55, 251-276.
- Favero, N. (2020). Adjusting confirmed COVID-19 case counts for testing volume. *medRxiv*, <https://doi.org/10.1101/2020.06.26.20141135>
- Gómez- Ramírez, L. y A. Villarraga-Orjuela. (2021). Expansionary Policy in Pandemics, a Dynamic Model Examination. *Revista Mexicana de Economía y Finanzas Nueva Época REMEF*, [S.l.], v. 16, n. 3, p. e622, june 2021. <https://doi.org/10.21919/remef.v16i3.622>
- IMCO. (2020). *Índice de competitividad estatal 2020. Que no vuelva a pasar: estados prevenidos valen por dos*. Recuperado de <https://imco.org.mx/indices/>
- Laborde, David, Will Martin and Rob Vos (2020). Poverty and food insecurity could grow dramatically as COVID-19 spreads. International Food Policy Research Institute: Research Post, April 16, 2020. https://doi.org/10.2499/p15738coll2.133762_02
- Moosa, I. y Vaz, J. (2016). Cointegration, error correction and exchange rate forecasting. *Journal of International Financial Markets, Institutions and Money*. Vol. 44. <https://doi.org/10.1016/j.intfin.2016.04.007>
- Pujol, T. (2020). The long-term economic cost of Covid-19 in the Consensus Forecasts. *Covid Economics Vetted and Real-Time Papers*, Issue 44, 225-237.

- Secretaría de Salud. (2020). COVID-19 México. Información general. Recuperado de <https://coronavirus.gob.mx/datos/> (Bases reportadas los días 11 de mayo, 15 de junio, 13 de julio y 17 de agosto de 2020).
- Secretaría de Salud (2020a). Versión estenográfica. Conferencia de prensa. Informe diario sobre coronavirus COVID-19 en México. 6 de octubre. <https://www.gob.mx/presidencia/es/articulos/version-estenografica-conferencia-de-prensa-informe-diario-sobre-coronavirus-covid-19-en-mexico-254220?idiom=es>
- Seoane, B. (2020). A scaling approach to estimate the COVID-19 infection fatality ratio from incomplete data. *medRxiv*, <https://doi.org/10.1101/2020.06.05.20123646>
- Stock, J. (2020). Data gaps and the policy response to the novel coronavirus. *Covid Economics Vetted and Real-Time Papers*, Issue 3, 1-11.
- Stock, J. y Watson, M. (1993). A Simple Estimator of Cointegrating Vectors in Higher Order Integrated System. *Econometrica*, Vol. 61. <http://dx.doi.org/10.2307/2951763>
- Valero, M. and Valero-Gil, J.N. (2021). Determinants of the number of deaths from COVID-19: differences between low-income and high-income countries in the initial stages of the pandemic. *International Journal of Social Economics*. Vol. ahead-of-print No. ahead-of-print. <https://doi.org/10.1108/IJSE-11-2020-0752>
- Vázquez Muñoz J. A., N. I. Muller Durán y J. Zavaleta González. (2021). Public Deficits in USMCA Economies During the COVID-19 Economic Crisis. *Revista Mexicana de Economía y Finanzas Nueva Época REMEF*, [S.l.], v. 16, n. 3, p. e608, may 2021. <https://doi.org/10.21919/remef.v16i3.608>
- Ximénez-Fyvie L.A. (2021). Un daño irreparable: La criminal gestión de la pandemia en México. Editorial Planeta. Primera Edición.
- Zavala, M. & Despeghe, L. (2020, mayo). ¿Qué nos dicen las actas de defunción de la CDMX? *Nexos*. Recuperado de <https://datos.nexos.com.mx/?p=1388> (Agosto 24, 2020).
- Zavala, M. & Despeghe, L. (2021, febrero). ¿Qué nos dicen las actas de defunción de la CDMX? Actualización al 14 de febrero de 2021. *Nexos*. Recuperado de <https://datos.nexos.com.mx/que-nos-dicen-las-actas-de-defuncion-de-la-cdmx-actualizacion-al-14-de-febrero-de-2021/> (Julio 9, 2021).

Anexo 1

Cuadro A1.
Número de Clústeres Óptimo Sugerido por cada Indicador

Índice	Número de Grupos Óptimo		
	Al 24 de abril	25 de abril-29 de mayo	30 de mayo- 26 de junio
Krzanowsky y Lai (KL)	2	6	2
Calinski y Harabasz (CH)	16	16	16
Hartigan	3	3	5
Cubic Clustering Criterion (CCC)	16	16	15
Scott and Simon	4	3	5
Marriot	4	3	5
Milligan y Cooper (TrCovW)	5	3	3
Milligan y Cooper (TraceW)	4	3	3
Friedman	15	16	15
Rubin	14	13	12
Categorical index	3	2	4
Davies y Bouldin	16	16	14
Silhouette	2	2	2
Duda	2	2	2
PseudoT2	2	3	2
Beale	2	2	2
Rarkowsky	2	2	2
Ball	3	3	3
PtBiserial	2	3	2
Frey	2	1	2
McCalin	2	3	2
Dunn	2	13	2
Hubert	0	0	0
SDindex	6	6	5
Dindex	0	0	0
SDbw	16	16	16
Número óptimo de acuerdo con la regla de mayoría de votos	2	3	2

Anexo 2. Relación entre la actualización de datos y los subíndices del IMCO

En la sección 2 se menciona que los cambios a la información en las fechas cercanas al reporte inicial se pueden considerar actualización de la base de datos debido al tiempo que tardan en confirmarse los resultados de las pruebas de COVID-19, mientras que las modificaciones a fechas previas muy distantes pueden ser una decisión estratégica de las entidades. En este sentido, uno esperaría que la actualización a la base de datos no fuera estratégica; es decir, no se relacionara con los subíndices del IMCO.

Aunque se reconoce que no es sencillo determinar hasta qué día los cambios en la información se consideran actualizaciones y a partir de cuándo son distorsiones, para este ejercicio se considera como actualización al cambio realizado en la información de un día específico cuando se comparan reportes de dos fechas. Dado que la Gráfica 1 muestra que el mayor cambio al comparar los reportes del 15 de junio y del 17 de agosto es precisamente el día en que se presenta el primer reporte, el Cuadro A2 describe la manera en que se calcula la actualización en cada una de las tres fechas analizadas: 11 de mayo, 15 de junio y 13 de julio. Por ejemplo, para el 11 de mayo, se comparan los contagios y las muertes anunciados en el reporte de ese día con las cifras que el 15 de junio se reportan que ocurrieron el 11 de mayo.

Cuadro A2.

Construcción de las actualizaciones en mediciones del COVID-19

Fecha considerada	Fecha de base inicial	Fecha de base final
11 de mayo	11 de mayo	15 de junio
15 de junio	15 de junio	13 de julio
13 de julio	13 de julio	17 de agosto

Fuente: Elaboración propia con información de la Secretaría de Salud (2020).

Al igual que el análisis de las distorsiones de la sección 3, la actualización observada se normaliza por cada millón de habitantes. El Cuadro A3 presenta el promedio y la desviación estándar de las actualizaciones tanto de contagios como de muertes para cada una de las tres fechas. Se observa un crecimiento en el promedio de las actualizaciones de ambas mediciones con el paso del tiempo.

Cuadro A3.

Promedio y desviación estándar de actualizaciones del COVID-19

Fecha	En contagios		En muertes	
	Promedio	Desv. Est.	Promedio	Desv. Est.
11 de mayo	17.96	15.68	1.54	1.81
15 de junio	48.50	25.54	3.88	2.31
13 de julio	70.14	41.11	5.28	3.22

Fuente: Elaboración propia con información de la Secretaría de Salud (2020).

El Cuadro A4 sintetiza los resultados del efecto independiente que tiene cada subíndice del IMCO en la actualización de cada fecha. Se presentan los signos de los subíndices cuyos coeficientes son estadísticamente significativos al realizar regresiones simples. Se observa que el subíndice de sectores precursores de clase mundial se relaciona positivamente con la actualización de contagios en las tres fechas. Por su parte, la actualización de las muertes por COVID-19 solo tiene relación con algunos subíndices en la actualización del 11 de mayo.

Cuadro A4.

Subíndices del IMCO significativos para explicar la actualización en mediciones

Subíndices	11 de mayo		15 de junio		13 de julio	
	Contagios	Muertes	Contagios	Muertes	Contagios	Muertes
Sistema de derecho confiable y objetivo		-				
Sociedad incluyente, preparada y sana					+	
Mercado de factores eficientes			+		+	
Economía estable		+			+	
Sectores precursores de clase mundial	+	+	+		+	

Fuente: Elaboración propia con información de la Secretaría de Salud (2020).

El Cuadro A5 muestra los resultados de las estimaciones finales de aplicar el método hacia atrás de eliminación de variables (*backward stepwise*), fijando como criterio de eliminación un nivel de significancia de 10%. Destaca que de los cuatro subíndices que, de manera independiente, explican la actualización de los contagios del 13 de julio solo el de mercado de factores eficientes mantiene su significancia. Por su parte, el subíndice de sectores precursores de clase mundial se relaciona positivamente con la actualización de contagios y de muertes del 11 de mayo.

Cuadro A5.
Subíndices del IMCO determinantes de la actualización de mediciones

VARIABLES	11 de mayo		15 de junio		13 de julio	
	Contagios	Muertes	Contagios	Muertes	Contagios	Muertes
Constante	-14.42 (9.987)	0.919 (1.412)	16.45 (13.60)	3.882*** (0.408)	-31.85 (31.29)	5.281*** (0.569)
Derecho		-0.0423** (0.0176)				
Mercado de Factores					1.812*** (0.545)	
Precursores	0.874*** (0.287)	0.0831*** (0.0200)	0.865** (0.349)			
Observaciones	32	32	32	32	32	32
R ²	0.460	0.501	0.170	0.000	0.270	0.000
F	9.290	14.55	6.130	0	11.08	0

Fuente: Elaboración propia con información de la Secretaría de Salud (2020).
Errores estándares robustos entre paréntesis. *** p<0.01, ** p<0.05, * p<0.10.

La comparación de los resultados del método hacia atrás de eliminación de variables análisis para las distorsiones y para las actualizaciones muestra que los subíndices que explican los cambios en las mediciones en el primer periodo son los mismos (el subíndice de sectores precursores de clase mundial para ambas medidas y el subíndice de sistema de derecho confiable y objetivo para el cambio en muertes). Para los otros dos periodos, los subíndices que explican las distorsiones no contribuyen en explicar las actualizaciones, las cuales están relacionadas con dos subíndices que no contribuyen a explicar las distorsiones.

Anexo 3.

Cuadro A6.
P-valores de la Prueba DFA asociados con las series de confirmados (Conf.) y de defunciones (Def.)

Entidades	Niveles		Primeras Diferencias		Segundas Diferencias		Orden de Integración	
	Conf.	Def.	Conf.	Def.	Conf.	Def.	Conf.	Def.
Aguascalientes	0.90	0.98	0.00	0.00	0.00	0.00	1	1
Baja California							1	1
Norte	0.07	0.09	0.00	0.00	0.00	0.00		
Baja California							1	1
Sur	0.79	0.96	0.00	0.00	0.00	0.00		
Campeche	1.00	0.64	0.00	0.00	0.00	0.00	1	1
Chiapas	0.99	0.98	0.08	0.00	0.00	0.00	1	1
Chihuahua	0.15	0.12	0.00	0.00	0.00	0.00	1	1
Coahuila	0.95	0.85	0.00	0.00	0.00	0.00	1	1
Colima	0.28	0.09	0.00	0.00	0.00	0.00	1	1
Ciudad de México	0.52	0.03	0.00	0.00	0.00	0.00	1	1
Durango	0.99	0.82	0.00	0.00	0.00	0.00	1	1
Guanajuato	1.00	0.60	0.00	0.00	0.00	0.00	1	1
Guerrero	0.98	1.00	0.00	0.00	0.00	0.00	1	1
Hidalgo	0.95	0.92	0.00	0.00	0.00	0.00	1	1
Jalisco	0.98	0.91	0.00	0.00	0.00	0.00	1	1
Estado de México	0.53	0.00	0.00	0.05	0.00	0.00	1	0
Michoacán	0.80	0.85	0.00	0.00	0.00	0.00	1	1
Morelos	0.14	0.86	0.00	0.00	0.00	0.00	1	1
Nayarit	0.98	0.62	0.00	0.00	0.00	0.00	1	1
Nuevo León	0.99	0.40	0.00	0.00	0.00	0.00	1	1
Oaxaca	1.00	0.99	0.00	0.00	0.00	0.00	1	1
Puebla	1.00	0.25	0.00	0.00	0.00	0.00	1	1

Querétaro	0.97	0.89	0.00	0.00	0.00	0.00	1	1
Quintana Roo	0.66	0.12	0.00	0.00	0.00	0.00	1	1
San Luis Potosí	0.95	0.00	0.00	0.00	0.00	0.00	1	0
Sinaloa	0.89	0.57	0.00	0.00	0.00	0.00	1	1
Sonora	1.00	1.00	0.10	0.15	0.00	0.00	2	2
Tabasco	0.76	0.40	0.00	0.00	0.00	0.00	1	1
Tamaulipas	1.00	0.89	0.00	0.00	0.00	0.00	1	1
Tlaxcala	0.86	0.96	0.00	0.00	0.00	0.00	1	1
Veracruz	0.67	0.98	0.22	0.11	0.00	0.00	2	2
Yucatán	0.39	0.88	0.00	0.00	0.00	0.00	1	1
Zacatecas	0.50	0.42	0.00	0.00	0.00	0.00	1	1
Nacional	0.93	0.25	0.00	0.00	0.00	0.00	1	1

Notas: (1) Para cada una de las pruebas se incluye un intercepto y/o una tendencia de acuerdo con su significancia. (2) El número óptimo de rezagos se seleccionó de acuerdo con el criterio MIAC.