

Close

## THE CONVERSATION

Rigor académico, oficio periodístico



Shutterstock / Cryptographer

# Covid-19: la malinterpretación de los datos de la pandemia daña la confianza del público

4 noviembre 2020 20:50 CET

Dos cartas publicadas en *The Lancet* y *The Lancet Public Health* en los últimos meses defienden la necesidad de evaluar de forma independiente la respuesta española ante la covid-19. Estamos de acuerdo, pero nos gustaría complementarlas con tres puntos que nos acercan a la ciencia abierta. Nos referimos a la confusión terminológica, la calidad de los datos y su disponibilidad.

Según las cartas, una de las razones por las que España se ha visto más afectada por la pandemia es la poca confianza que hay en el asesoramiento científico. Creemos que gran parte de esta desconfianza está ocasionada por una importante confusión terminológica. Esta se produce en prácticamente todos los medios de comunicación (con notables excepciones), y conduce a una malinterpretación de los datos y, como consecuencia, a una pérdida de confianza en los sistemas de información sanitaria, en la investigación y la epidemiología.

### ¿Casos nuevos o positivos? ¿Sospechosos o confirmados? La importancia del contexto y la terminología

Cuando hablamos de confusión terminológica no solo nos dirigimos a investigadores, sino a comunicadores y divulgadores. Su responsabilidad es que la sociedad no se

### Autores



**Marc Saez Zafra**

Catedrático de Estadística y Econometría,  
Universitat de Girona



**Adolfo Figueiras Guzmán**

Full Professor of Public Health,  
Universidade de Santiago de Compostela



**Ainhoa Alustiza Galarza**

Demographer, Max Planck Institute for  
Demographic Research

confunda y pueda también entender y combatir la pandemia.

Oímos recientemente hablar de “casos positivos” (o, a menudo, simplemente “positivos”), que muchas veces se confunden con los “casos nuevos”.

## No consuma noticias, entiéndalas.

### Suscribirme al boletín

Los casos positivos son aquellos que han dado positivo en alguna prueba diagnóstica (PCR o test rápido). Es decir, personas infectadas por SARS-CoV-2, hayan desarrollado covid-19 o no.

Los casos nuevos (“incidentes”) son los que han dado positivo por primera vez. De hecho, los casos positivos no solo incluyen los casos nuevos, sino que también incluyen aquellos que ya habían dado positivo previamente (a menudo más de una vez) y lo han vuelto a dar en otra prueba diagnóstica.

Por tanto, **los casos positivos contabilizan más de una vez los mismos casos.**

Esto hace que sea frecuente que, de un día para otro, aumente el número de casos positivos (que dependen del número de pruebas realizadas) y disminuya el número de casos nuevos. De los dos conceptos, el relevante para la investigación epidemiológica es el de “casos nuevos” o “incidencia” de la enfermedad. Es decir, el número de casos nuevos en un período de tiempo determinado.

Cuando se informa de los casos positivos siempre se debería acompañar de la información sobre casos nuevos, ya que es el concepto que realmente informa del contagio. Además, se debería utilizar siempre el mismo período temporal. Por ejemplo, casos nuevos en las últimas 24 horas o incidencia acumulada en la última semana. En su lugar, a menudo se utilizan indistintamente y sin diferenciar.

También sería importante acompañar esta información con el número de pruebas positivas entre las realizadas (“tasa de positividad”), útil para conocer indirectamente la calidad del rastreo. Valores bajos indican que este es mejor y que se están realizando pruebas no solo a personas con síntomas.

Por otra parte, suelen confundirse “casos confirmados” con “casos sospechosos”. Los casos confirmados son los casos positivos. Los sospechosos son las personas que en algún momento han presentado síntomas y un profesional sanitario los ha clasificado como posible caso, pero no tienen una prueba diagnóstica con resultado positivo.



**Alberto Ruiz Cantero**

Jefe de Servicio de Medicina Interna,  
Escuela Andaluza de Salud Pública



**Alexandre Medeiros de Figueiredo**

Professor of Public Health, Universidade  
Federal da Paraíba



**Alex Sánchez-Pla**

Profesor Titular de Universidad. Estadística  
y Bioinformática. Director de la Unidad de  
Estadística y Bioinformática del Vall  
d'Hebron Research Institute, Universitat de  
Barcelona



**Anabel Forte Deltell**

Doctora en Matemáticas y profesora en la  
Universidad de Valencia, Departamento de  
Estadística e Investigación Operativa,  
Universitat de València



**Andrés Cabrera León**

Doctor en Estadística y Matemáticas.  
Profesor, Escuela Andaluza de Salud  
Pública



**Antonio Daponte Codina**

Médico, Escuela Andaluza de Salud  
Pública



**Aurelio Tobias**

Investigador Científico CSIC, Instituto de  
Diagnóstico Ambiental y Estudios del Agua  
(IDAEA - CSIC)

Otro término que habría que considerar en todas las estadísticas es el de “casos activos”. Se define como personas infectadas (PCR positivos, hayan desarrollado o no la enfermedad), menos las curadas (tanto infectadas asintomáticas como curadas de la covid-19) y las fallecidas.

En general, no se dispone de estos datos (el Ministerio de Sanidad no facilita datos de personas curadas). Además, se debería discriminar entre personas infectadas y enfermas.

## ¿Es posible mejorar la calidad de los datos? ¿Definimos todos igual la mortalidad?

Parte de la desconfianza de la sociedad y de la baja calidad de algunos trabajos de investigación están ocasionadas por problemas en la calidad de los datos.

Este es un problema estructural causado por la diversidad de criterios usados en la recogida sistemática de datos clínicos. En otras palabras, es un error asociado a la recogida de los datos y no un fallo causado por la pandemia. Como decía uno de los más prestigiosos estadísticos de la medicina, Douglas Altman, “para maximizar el beneficio para la sociedad, no solo se necesita investigar, sino hacerlo bien”.

Por ese motivo, preguntarse por la calidad de los datos (también por su transparencia, validez, integridad, relevancia, exhaustividad, utilidad, generalización, objetividad, y fiabilidad) y sus diseños es algo fundamental.

Es conocida la discrepancia sobre las cifras de mortalidad a consecuencia de la covid-19. A partir del 11 de mayo, las comunidades autónomas notifican diariamente los casos confirmados de forma individualizada al Ministerio de Sanidad, quien los publica en internet. Según este informe, hasta el 30 de octubre se habrían producido 35 878 muertes en toda España y, por poner un ejemplo, 5 991 en Cataluña. Sin embargo, según la Generalitat, en esta Comunidad Autónoma se habrían producido 13 955 muertes hasta ese día.

La diferencia se debe a la definición de la mortalidad como consecuencia de la covid-19: “fallecidos confirmados”, según el Ministerio de Sanidad; “fallecidos confirmados y sospechosos”, según la Generalitat de Cataluña. Es cierto que esta última definición refleja mejor la situación, pero tampoco proporciona el número real de fallecidos.

Según las estimaciones del exceso de mortalidad realizadas por el Centro Nacional de Epidemiología del Instituto de Salud Carlos III, el 25 de octubre se habría producido un exceso de 58 389 defunciones en el conjunto de España (44 585 entre el 11 de marzo y el 9 de mayo y 9 074 entre el 1 de septiembre y el 25 de octubre) y 13 245 en Cataluña (11 922 entre el 13 de marzo y el 8 de mayo y 778 entre el 9 de septiembre y el 25 de octubre). Esto implica unos 700 fallecidos menos que la cifra proporcionada por la Generalitat de Catalunya. Debe señalarse, además, que el MoMo contabiliza la



**Carmen Sánchez-Cantalejo Garrido**

Técnico de proyectos en el Área de Salud Pública. Experta en Epidemiología y Estadística. Miembro del Instituto de Investigación Biosanitaria de Granada, Escuela Andaluza de Salud Pública



**Cristian Tebé Cordomí**

Investigador en la Unitat de Bioestadística, Institut d'Investigació Biomèdica de Bellvitge (IDIBELL)



**Daniel Fernández Martínez**

Lector Serra-Hünter en Estadística, Universitat Politècnica de Catalunya - BarcelonaTech



**David Blanco de Tena-Dávila**

Asistente metodológico y estadístico de la facultad de Fisioterapia, Universitat Internacional de Catalunya



**Dominic Royé**

Investigador de Geografía Física, Universidade de Santiago de Compostela



**Erik Cobo Valeri**

Catedrático de Estadística e Investigación Operativa, Universitat Politècnica de Catalunya - BarcelonaTech

**Gemma Molist**

Profesora en la Facultad de Medicina. Responsable del Servicio de Estadística en el Área de Investigación e Innovación del

mortalidad bruta. Es decir no todo el exceso de muertes es debido a la COVID-19.

Otro concepto que debe considerarse es el de “letalidad”, es decir, aquellas muertes producidas por la propia infección viral. En este caso la causa inicial o fundamental del fallecimiento es la covid-19. No es lo mismo morir *por* esta enfermedad que morir *con* una infección por SARS-CoV-2. En ambos casos la PCR sería positiva, pero las cifras de muertes no distinguen ambas posibilidades.

También es conocido el problema sobre el número de fallecidos como consecuencia de la covid-19 en centros residenciales de personas mayores. Un documento interno del Ministerio de Sanidad estima que, hasta agosto, 19 672 personas fallecieron en esas residencias con covid-19 o síntomas similares (aproximadamente la mitad confirmados y la otra mitad, sospechosos).

Sin embargo, el propio Ministerio de Sanidad reconoció la baja calidad de los datos, que las Comunidades Autónomas no aportan de forma regular ni homogénea. Esto provoca errores en las fechas de fallecimiento, en las causas de muerte, en la suma de residentes, o incoherencias en la evolución temporal (por ejemplo, disminuciones en el número de fallecidos). Además, algunas comunidades no cuentan como fallecidos a residentes que murieron en hospitales, mientras que otras distinguen las personas fallecidas en distintos tipos de residencias.

Hospital General de Granollers., Universitat de Vic – Universitat Central de Catalunya



**Guadalupe Gómez Melis**

Professor Statistics. President Consell Català d'Estadística. Coordinator GRBIO, Universitat Politècnica de Catalunya - BarcelonaTech



**Inmaculada Mateo Rodriguez**

Doctora en psicología, Escuela Andaluza de Salud Pública



**Jorge Casillas**

Catedrático de Inteligencia Artificial, Universidad de Granada



**Jorge Mateu**

Profesor de Estadística, Universitat Jaume I



**José A. Martínez**

Profesor titular del Departamento de Economía de la Empresa, Universidad Politécnica de Cartagena



**José Antonio González Alastrué**

Profesor Titular de Universidad, Dept. Estadística e Investigación Operativa, Universitat Politècnica de Catalunya - BarcelonaTech



**Juan A. Cayla Buqueras**

Presidente Fundación de la Unidad de Investigación en Tuberculosis de

Barcelona, Unidad de Investigación en  
Tuberculosis de Barcelona (UITB)



**Juan Guàrdia Olmos**

Facultat de Psicologia. Secció de  
Psicologia Quantitativa, Universitat de  
Barcelona



**María A. Barceló Rado**

Catedrática acreditada de Estadística y  
Econometría, Universitat de Girona



**María del Mar Rueda García**

Catedrática de Estadística e Investigación  
Operativa, Universidad de Granada



**Martí Casals**

Profesor de Estadística en la Facultad de  
Medicina - Fundación de Estudios  
Superiores en Ciencias de la Salud  
(FESS), Grupo Investigación del Centro de  
Estudios en el Deporte y Actividad Física  
(CEEAF) en la UVic, Universitat de Vic –  
Universitat Central de Catalunya



**Miguel De la Fuente Sánchez**

Profesor de Métodos de Investigación,  
Universidad Alfonso X el Sabio



**Pau Fonseca i Casas**

Profesor en el área de Estadística e  
Investigación Operativa, Universitat  
Politécnica de Catalunya - BarcelonaTech



**Salvador Macip**

Professor and Researcher, UOC -  
Universitat Oberta de Catalunya



**Sergi Trias-Llimós**

Researcher at Centre for Demographic  
Studies (CED-UAB), Universitat Autònoma  
de Barcelona



**Timothy L M Riffe**

Research Scientist, Max Planck Institute for  
Demographic Research



**Ulises Cortés**

Catedrático de Inteligencia Artificial,  
Universitat Politècnica de Catalunya -  
BarcelonaTech



Shutterstock / Cryptographer

**¿Medimos todos igual la incidencia de la covid-19?**

Las discrepancias también se encuentran en la incidencia de la enfermedad. Los datos mostrados por el Centro Europeo para el Control y la Prevención de Enfermedades (ECDC, por sus siglas en inglés), agencia europea que recoge datos oficiales proporcionados por los Gobiernos, coinciden con los mostrados por el Ministerio de Sanidad, proveedor de esos mismos datos, solo en términos acumulados (337 334 casos confirmados en España hasta el 13 de agosto de 2020). Sin embargo, difieren mucho en los datos de incidencia diaria y semanal.

Así, según el ECDC, la incidencia en la semana del 10 al 16 de agosto fue de 28 451 casos. Según el Ministerio de Sanidad, fue de 16 269 casos (calculado como la diferencia entre los casos acumulados hasta el lunes 17 de agosto y los acumulados hasta el viernes 14 de agosto, puesto que, desde el 4 de julio, el Ministerio no publica datos los fines de semana) o 10 516 casos (calculado como la suma de la incidencia diaria desde el lunes 10 al viernes 14 de agosto, ambos incluidos).

Todos los días de esa semana, la incidencia diaria proporcionada por el Ministerio fue mucho menor que la proporcionada por el ECDC (desde un 45 % inferior el viernes 14, hasta un 83 % el lunes 10). Estas diferencias, que desafortunadamente no son esporádicas, no pueden explicarse exclusivamente por una falta de datos (algunos días algunas comunidades no actualizan los datos por problemas técnicos) sino por una falta de calidad de los mismos.

De hecho, el informe advierte todos los días que “se está realizando una validación individualizada de los casos, por lo que puede haber discrepancias respecto a la notificación de días previos”. No son pocos los días en los que algunas comunidades depuran sus datos y eliminan duplicados. Sin embargo, el Ministerio no vuelve a publicar, por lo menos en línea, la serie revisada con los datos que faltan ni con los depurados. Tampoco cuando se producen cambios metodológicos (por ejemplo, hasta el 10 de mayo las comunidades notificaban al Ministerio los datos diariamente de forma agregada), de frecuencia (desde el 4 de julio no se proporcionan datos los fines de semana) y de temporalidad (los datos de defunciones hasta el 23 de mayo hacen referencia a defunciones diarias y, a partir del 24, a defunciones de los últimos 7 días).

Esta práctica de no publicar toda la serie revisada no es exclusiva del Ministerio, sino que se repite en términos muy parecidos en prácticamente todas las instituciones públicas, incluidas las de las comunidades autónomas.

Tampoco se dispone de información alguna acerca de la aplicación práctica del protocolo de identificación de casos recogido en la *Estrategia de detección precoz, vigilancia y control de covid-19*. En ella se define como “caso sospechoso” aquel que presenta síntomas, y para el cual está indicada la realización de una PCR u otro tipo de prueba similar.

Pues bien, presentar “fiebre” (por ejemplo), en ausencia de otros síntomas, no conlleva la definición de “caso sospechoso”, y por tanto no implica la realización de la prueba PCR, a pesar de que la fiebre es el síntoma más frecuente de la infección. Queda a criterio de las CCAA, los servicios de salud, o la interpretación del personal médico.

Este hecho es relevante por dos razones:

- Además de los casos sintomáticos y asintomáticos, tendríamos también casos “no diagnosticados”, que son aquellos que, presentando síntomas, no se les ha realizado la prueba diagnóstica. Esto facilita la transmisión comunitaria que la estrategia antes mencionada pretende combatir.
- El número de casos positivos de covid-19 entre Comunidades Autónomas es poco comparable, ya que dependerá de las indicaciones concretas para la transformación de los “casos no diagnosticados” en positivos.



Shutterstock / Cryptographer

## **Necesitamos reproducibilidad, repetibilidad y fiabilidad**

La “repetibilidad” significa que los investigadores deben esperar producir resultados idénticos al aplicar los mismos métodos de análisis a los mismos datos para cualquier pregunta. Pero esto solo se consigue con unos datos de calidad suficiente, procedentes de un sistema de información, parafraseando al Ministerio de Sanidad, “fiable, comparable e interoperable”.

En cuanto al principio de la “reproducibilidad”, un investigador diferente debería poder realizar la misma tarea en la ejecución de un análisis en una base de datos y esperar producir un resultado idéntico al del primer investigador. En esta ocasión, la reproducibilidad depende de que el sistema de información, parafraseando de nuevo al Ministerio de Sanidad, sea “accesible y transparente”.

Podemos decir que la accesibilidad a los datos tiene dos componentes interrelacionados: facilidad de acceso y contenido. La accesibilidad a los datos estará muy limitada si, a pesar de que el acceso sea muy fácil, el contenido al que se accede es muy reducido o limitado.



En lo referido a la información sobre la covid-19, existe una gran facilidad de acceso. Tanto el Ministerio de Sanidad, como casi todas las comunidades y ciudades autónomas (con la excepción de Extremadura, Galicia, y Ceuta) e incluso algunos ayuntamientos (como el de Barcelona) tienen webs dedicadas a la información sobre este tema. Sin embargo, en la mayoría de ellas el acceso con finalidades de investigación es muy limitado, puesto que la descarga de datos solo está disponible en las webs de Andalucía, Canarias, Cataluña y Madrid, así como en la web del Ministerio de Sanidad, aunque ésta solo parcialmente (casos que han precisado hospitalización, ingreso en UCI y fallecidos por comunidades autónomas).

El resto de las webs, o bien muestran los datos en formato de tablas y mapas, o solo permite descargar los documentos en formato pdf. Incluso en alguna web (como la del Ministerio) se está restringiendo últimamente el acceso a datos de ingresos en UCI o en hospital, o modificando el formato y definición de las variables. Esto dificulta su descarga, integración y comparabilidad entre las fuentes de información.

En cuanto al contenido de los datos, el panorama es bastante más desfavorable. La mayor cantidad de información la proporciona la web del Ministerio de Sanidad. Sin embargo, una parte importante de la información que dispone el Ministerio no es accesible. Por ejemplo, como dijimos, la información sobre el número de fallecidos en centros residenciales de personas mayores no se ha hecho pública, circulando solo como un documento interno ya que el propio Ministerio reconoció que la calidad de los datos dificulta su difusión. Además, las webs autonómicas o estatales no disponen de un sistema de descarga automática de las actualizaciones de los datos (sí la Organización Mundial de la Salud), por lo que esta debe hacerse de forma manual, lo que dificulta aún más los análisis.

Por otro lado, el nivel máximo de desagregación al que se tiene acceso desde la web del Ministerio es el de comunidad autónoma. La existencia de grupos vulnerables de población que experimentan desigualdades de salud y socioeconómicas es otra de las razones para que en España haya afectado más la pandemia. Esas desigualdades solo pueden ser investigadas a un nivel de desagregación mucho menor que el de comunidad autónoma.

Los datos a nivel de lo que se conoce con el nombre de “área pequeña” solo los proporcionan algunas de las webs de las comunidades autónomas. En este sentido, a nivel de provincia: Aragón, Castilla–La Mancha, Castilla-León; a nivel de comarca, departamento, distrito o área sanitaria, zona o área básica de salud: Andalucía, Asturias, Castilla-León, Cataluña, Madrid, Navarra, La Rioja, Valencia; a nivel de municipio: Andalucía, Baleares, Canarias, Cataluña, Madrid, Murcia, País Vasco, La Rioja, Valencia. Cabe señalar que ciudades grandes como Madrid o Barcelona (aunque también otras) tienen decenas de áreas o zonas básicas de salud.

Así pues, recurriendo a las webs de las comunidades autónomas, sería posible estudiar las desigualdades socioeconómicas en salud y en la mortalidad por covid-19 a nivel de áreas pequeñas de distinto tamaño. Sin embargo, no es posible estudiar las desigualdades en la incidencia, ya que ninguna de las webs de las comunidades (ni tampoco de los ayuntamientos) proporciona información de la incidencia, aunque sí que lo hacen de los casos positivos. Hay que destacar que existe **alguna** iniciativa sin ánimo de lucro que sí que proporciona esa **información** a nivel de provincia.

## la información desagregada por grupos de edad es muy escasa

Por último, si el contagio y la muerte como consecuencia de la COVID-19 se produjese de forma más o menos homogénea para todas las edades bastaría con considerar lo que se denomina razones de incidencia o de mortalidad crudas. Es decir, incidencia o mortalidad por 100 000 habitantes, por ejemplo. Pero a diferencia del contagio, cuya tasa es muy parecida en todas las edades, las muertes se producen sobre todo en edades avanzadas.

En este caso deben utilizarse las razones de incidencia o de mortalidad estandarizadas. Es decir, diferenciar la incidencia o mortalidad por grupos de edad. Desafortunadamente, la información desagregada por grupos de edad es muy escasa a nivel nacional y se proporciona en muy pocas webs de las comunidades autónomas.

La incidencia acumulada en las últimas dos semanas se proporciona en la web del Ministerio (y, en el caso de la incidencia acumulada desde el 10 de mayo, desagregada además de por edad también por sexo), aunque para toda España, sin desagregar por comunidades autónomas. Andalucía (por distrito sanitario), Castilla-León (por provincia) y Cataluña (por comarca) proporcionan en sus webs información desagregada por edad y sexo.

Por otra parte, también desagregan la información por edad y sexo, la Comunidad Valenciana, aunque no de forma desagregada a nivel de áreas pequeñas, y el País Vasco, aunque sólo para el día de la consulta, no permitiendo recuperar la información retrospectiva.

En 2019, la Real Sociedad de Estadística británica publicó un informe en el cual ofreció diez recomendaciones sobre cómo el gobierno del Reino Unido podía mejorar sus sistemas de información. Las recomendaciones las divide en tres apartados: datos para una mejor formulación de políticas, los datos como motor de la productividad y prosperidad, y datos para fortalecer la democracia y confiabilidad.

En este sentido creemos que el conjunto de medidas emprendidas y que se puedan emprender para combatir la pandemia podría ser calificado como formulación de políticas.

Así, en el informe se lee que existe una gran oportunidad para una adopción de medidas más eficaz si se utilizan los datos para informar, lo que se está haciendo, y señala diversos aspectos en los que las instituciones públicas deberían enfocarse. Entre ellos: que se debe seguir avanzando en la apertura de los datos procedentes de las instituciones públicas (datos abiertos) y que tanto los investigadores como la opinión pública deben tener acceso a esos datos siempre que sea con fines de investigación y de educación científica para ayudar a la mejora de toma de decisiones para la Salud Pública y la prevención de enfermedades como la covid-19.

---

*También han participado en la elaboración de este artículo Fernando Rosell Ortiz, médico de emergencias y coordinador del Registro Español de Parada Cardíaca Extrahospitalaria; Pilar Rueda de la Puerta, directora técnica del Observatorio de Salud y Medio Ambiente de Andalucía (OSMAN) y Ray G. Butler, CEO de Butler Scientifics.*



datos

pandemia

COVID-19