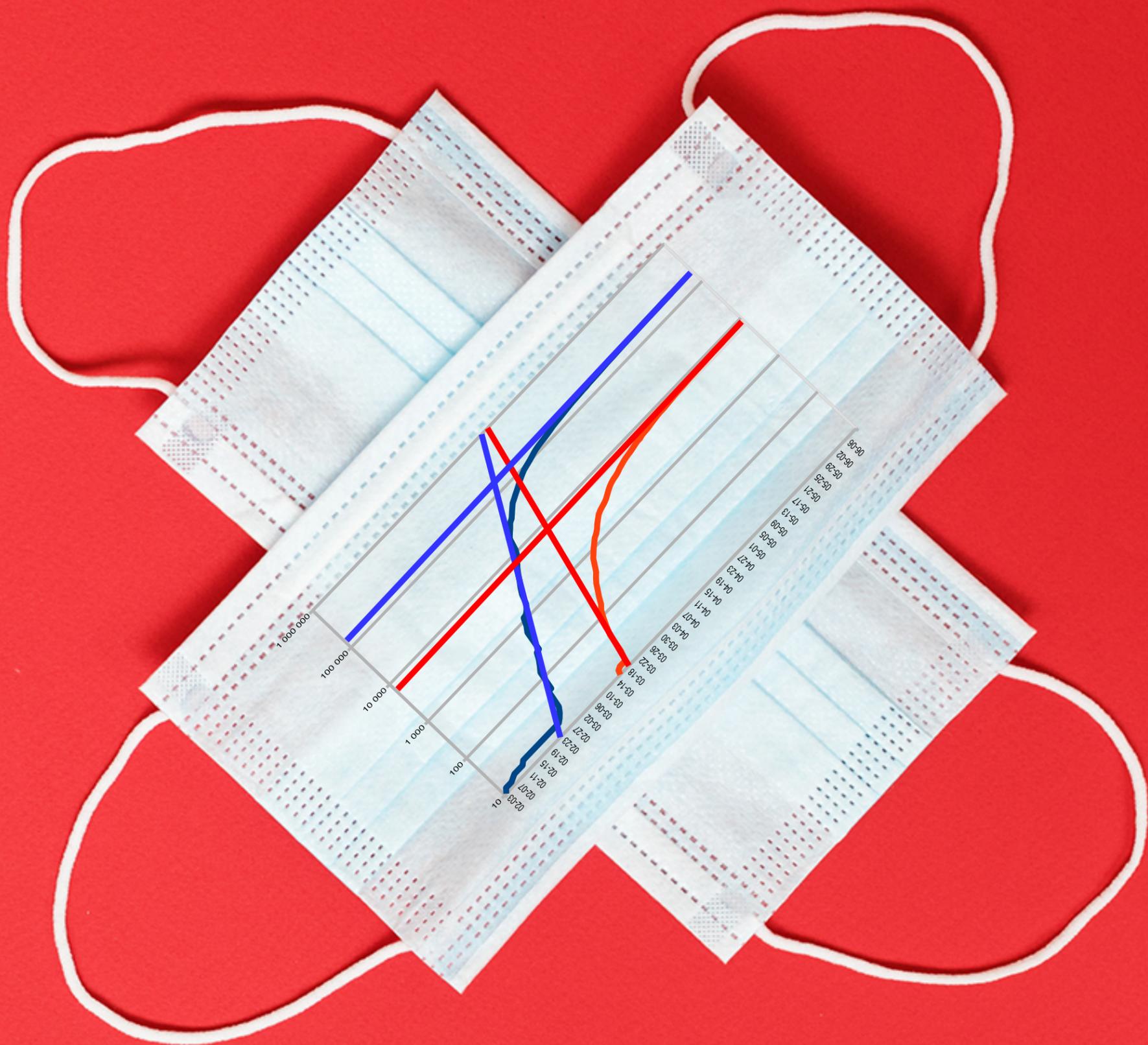


ep» el periodista

www.elperiodista.cl



SIETE LECCIONES PARA LIDIAR CON LOS DATOS DE UNA PANDEMIA

por Ricardo Baeza-Yates y Karma Peiró

Al inicio del 2020 prácticamente nadie había oído hablar del coronavirus, ni de cómo la enfermedad podría girar el mundo en cuestión de semanas.

Cinco meses más tarde, la economía mundial se ha desplomado, el teletrabajo se ha impuesto como solución a una subsistencia profesional e interpretamos datos a diario de nuevos infectados, muertos o recuperados.

Los datos están siendo el termómetro para entender la gravedad de la enfermedad, acatar con más exigencia el confinamiento, y valorar las estrategias de los políticos que nos gobiernan.

No obstante, si algo ha demostrado esta crisis es cuán complejo resulta aplicar los mismos criterios a los datos oficiales. Por ejemplo, cuando nos informan sobre una muerte por coronavirus, ¿se refieren a la persona que ha fallecido por el COVID19- o al enfermo terminal que ha muerto en período de pandemia? ¿Y los casos confirmados son todos los que existen? ¿Y los 'recuperados' están realmente recuperados?

La pandemia ha desbordado a los gobiernos de todo el mundo. La Organización Mundial de la Salud (OMS) es la primera que no ha sabido dar respuestas claras. La institución está muy preparada para afrontar al ébola en una zona de África, pero no para atender las demandas de todo el planeta a la vez. Y el caos de datos ofrecidos



por los gobiernos alimenta la incertidumbre, y las actuaciones públicas no siempre son las más acertadas.

En este ensayo repasamos siete lecciones que la crisis del coronavirus ha dejado hasta ahora. Las cuatro primeras son inherentes a la pandemia y las tres finales son válidas para los datos de cualquier crisis global.

LECCIÓN 1: ERRORES EN LA RECOGIDA DE DATOS



Sin datos no se puede entender cómo está progresando la pandemia, pero **sin saber cómo se han obtenido tampoco**. Según *Our World in Data* – científicos de diferentes grupos de investigación– “algunos países informan de las pruebas y otros de las personas evaluadas que pueden haberse hecho el test varias veces”. Y si tampoco se sabe





Organización Mundial de la Salud

cuándo se hicieron realmente el test, cada día se suman manzanas con naranjas.

Sobre los tests hay que mencionar también que no todos tienen la misma calidad. Los países que no hacen PCR –el recomendado por la OMS, capaz de detectar el virus con un %95 de acierto–, y usan sólo tests baratos y rápidos, quedan mal parados. Éste es el caso de Venezuela que muestra **distintas estadísticas generando confusión en la ciudadanía.**

Por otro lado, los laboratorios privados de algunos países no reportan a una autoridad central, por lo tanto sus datos tampoco son recogidos. Y muchas veces se desconocen con qué criterios se toman los tests: si se tienen que tener síntomas o no, si son reactivos (la gente va a los hospitales) o proactivos (muestras aleatorias en lugares de alta densidad), etc.

A todo esto hay que agregar todos los errores de la transferencia y procesamiento de los datos, desde su origen al destino donde sirven para decidir la políticas sanitarias. Por ejemplo, al transcribir un certificado de defunción hecho en papel por un doctor (con letra de doctor) a un soporte digital.



LECCIÓN 2: LA IMPRECISIÓN DE LOS DATOS



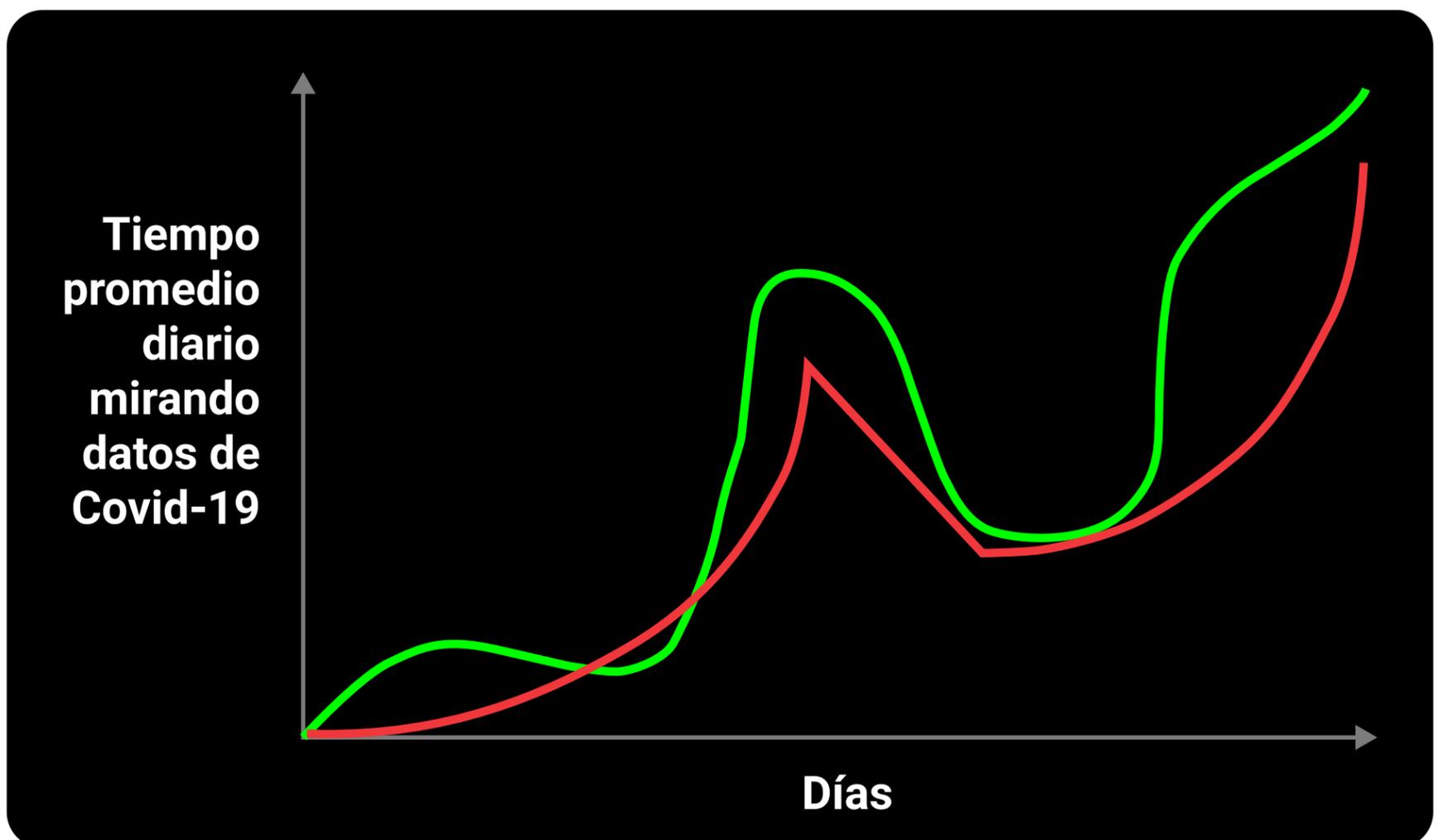
El virus se aprovecha de la cercanía física entre personas para expandirse y sobrevivir. Por esta razón, la pandemia es un proceso muy dinámico que depende de muchos factores y basta con que cada enfermo infecte a más de una persona para que el contagio crezca de forma exponencial. Y como además todo pasa muy rápido, hay imprecisión pues no lo sabemos todo.

Los casos informados que conocemos están lejos de ser los reales. Una gran proporción de los enfermos son asintomáticos (**estimados en un %40**), los que no saben que están enfermos y siguen contagiando. También este número depende de la cantidad de tests realizados: así que en la mayoría de los países hay el doble o más enfermos de los confirmados. **En Chile se estima**



que el factor es entre 3 y 4 veces, mientras que en Cataluña se estima en 10 veces.

Algo similar ocurre con los recuperados. Es muy difícil hacer seguimiento a cada paciente, sobre todo a los que nunca se han hecho un test. Sólo los hospitalizados son controlados mejor. Por eso hay países que tienen **muy pocos recuperados, como Reino Unido**. Y otros como **Chile usan fórmulas que los sobreestiman, generando problemas éticos**. Si los recuperados no se saben con exactitud, no sabemos tampoco cuántos enfermos activos – de los que conocemos – hay.



Incluso los fallecimientos están subestimados, siendo esto tan grave que merece una lección aparte. Y cuando llega el momento de calcular la letalidad de casos, **la mayoría la calcula mal**: hay que contar los enfermos que habían cuando los fallecidos se enfermaron y no los de hoy, **usando**



el número promedio de días que pasa entre el informe del test y el fallecimiento, por lo que siempre queda más baja de lo que realmente es. Pero por otro lado, la tasa de letalidad final será menor, pues como ya dijimos no sabemos el número real total de enfermos.

Debido a todas estas aproximaciones, cualquier análisis es impreciso y cualquier conclusión tiene que ser considerada con mucho cuidado. Es como ir en un coche con el parabrisas con barro: puede ser peligroso conducir en esas condiciones. Y en este caso no conducir no es una opción válida.

LECCIÓN 3: CAOS PARA CONTABILIZAR LOS MUERTOS



Algo que parece difícil de creer es que el número de fallecidos no sea exacto. Porque... ¿cuál es el motivo de fallecimiento si una persona tenía



una enfermedad previa y muere de COVID19-? La respuesta es distinta en cada país: en muchos **ya existe otra enfermedad**. En Estados Unidos hay un incentivo económico para decir que fue por COVID19-, ya que en ese caso el **hospital recibe más dinero de Medicare**, el programa federal de salud.

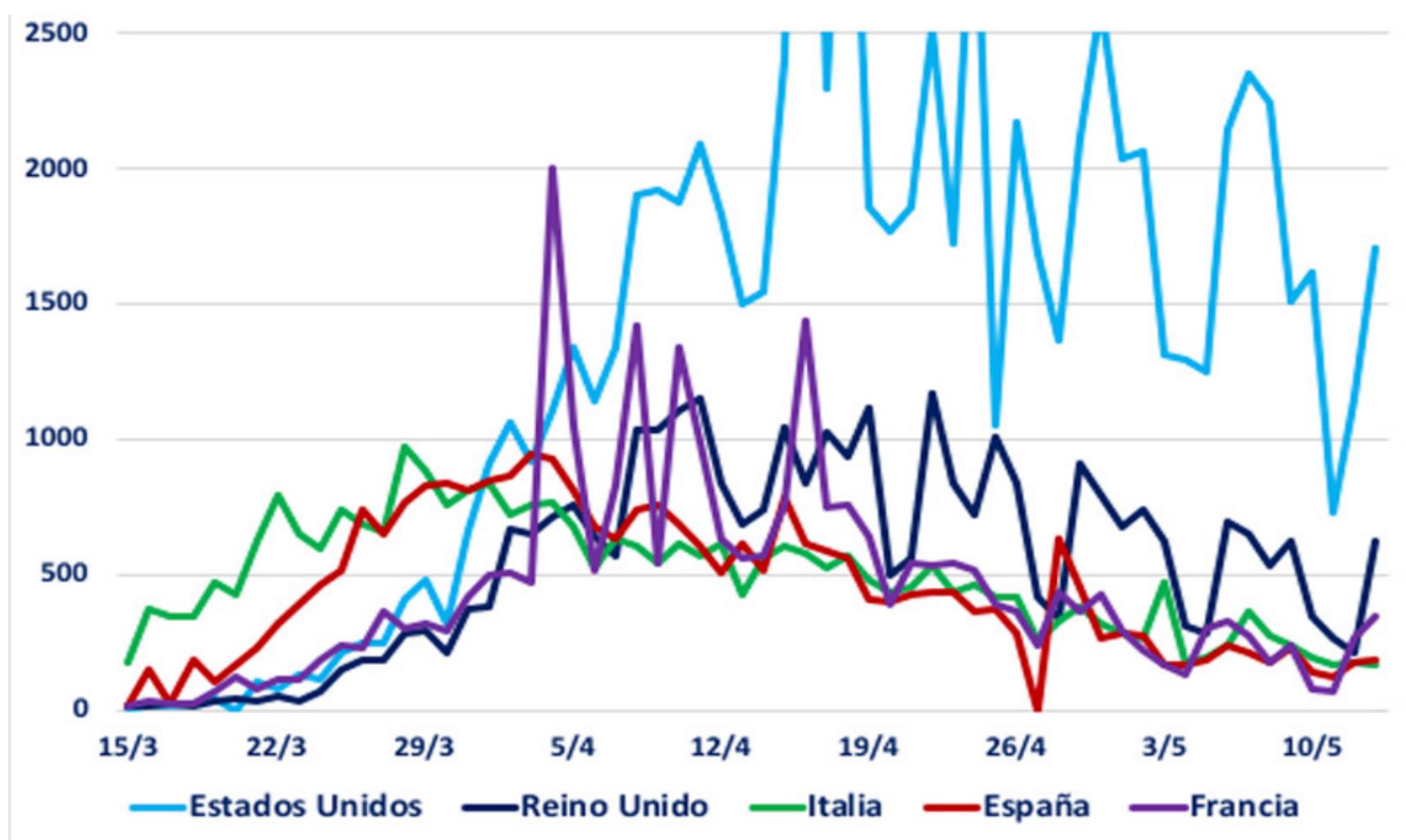
¿Y qué pasa si murió por COVID19- y no lo sabemos? En algunos países se hace un test, pero en la mayoría no. En otros, como **Bélgica, basta con la sospecha para contabilizarlo como parte de la pandemia**. Esta honestidad hace que sea uno de los países con más muertos per cápita.

Otro motivo de confusión son las fuentes utilizadas. En la mayoría de los países se contabilizan las muertes en hospitales y centros de salud, pero no en residencias de ancianos, ni se recogen los datos de las **funerarias que tienen información de fallecimientos en las casas sin causa confirmada**. Más aún, en el caso de España, las funerarias no están obligadas a dar sus datos; muchas residencias son de gestión privada y tampoco tienen por qué dar sus defunciones.

Cuando un país ha considerado estas fuentes han habido cambios brutales en los datos: aumentando en más de **4 mil los muertos en el Reino Unido** o en **más de 3 mil en Cataluña**. Esto produce discontinuidades en las estadísticas que



Muertes Diarias por Covid-19



generan un problema adicional a la imprecisión que ya teníamos.

Para estimar el número real de fallecidos se están usando las diferencias de un año a otro en los registros civiles de cada país, ajustando el aumento poblacional, para ver si han habido cambios adicionales a los fallecidos informados y verificar si estos son estadísticamente significativos. Por ejemplo en **España se estima que hay un %80 de fallecidos más.**

Y si esta estadística –vital para calcular tasas de mortalidad y letalidad– no es igual en cada país, implica un caos adicional para la OMS. Una solución sería tener estándares mundiales de cómo decidir la causa de la muerte y que fuentes usar para contabilizarlas.



LECCIÓN 4: LAS PARADOJAS TEMPORALES



Las lecciones anteriores enseñan que la pandemia es un proceso muy dinámico que depende de muchos factores, empezando por la educación cívica: basta con que cada enfermo infecte a más de una persona para que el contagio crezca de forma exponencial. Todo esto hace **muy complejo modelar cómo progresa la epidemia en el tiempo.**

Peor aún, **cada día estamos viendo los efectos del virus en el pasado.** Los fallecidos son personas que supimos que **estaban enfermas 10 o más días atrás** y los casos informados son de los **test que se hicieron varias horas o días antes.** El retardo depende del sistema de salud (público o privado), la disponibilidad del test y la ubicación geográfica (ya que, a veces, el test tiene que llevarse a donde hay laboratorios).



Sin embargo, **también cada gobierno contempla posibles futuros**. La experiencia de un país en un estadio más avanzado de la pandemia es útil para lidiar con nuestros datos ya obsoletos. Mientras más tarde ha llegado el virus, mejor es la bola de cristal. Así lo entendió muy bien Jacinta Ardern, primera ministra de Nueva Zelanda, que ni siquiera esperó a tener parte de los datos: decretó el cierre total de fronteras antes de la primera muerte. Por ello es uno de **los pocos países que puede decir que tiene el virus casi erradicado**.

LECCIÓN 5: LA IMPORTANCIA DE LA TRANSPARENCIA



La transparencia sobre las pruebas diagnósticas del coronavirus avanza a varias velocidades

Durante crisis que afectan a todos, en particular las de salud, la transparencia de los datos es un reflejo del nivel de democracia de cada gobierno y de la confianza de los ciudadanos en ese mismo

gobierno. Ocultar datos sólo genera desconfianza y problemas políticos que nadie quiere tener durante una crisis. En Guatemala, **el gobierno casi no entrega datos.**

En Chile hay bastantes datos agregados, pero pocos microdatos (es decir, información a nivel de paciente) lo que ha generado una **discusión pública prolongada.** En España, la organización ciudadana Civio **ha analizado qué comunidades autónomas son más transparentes** en dar datos sobre los tests realizados.

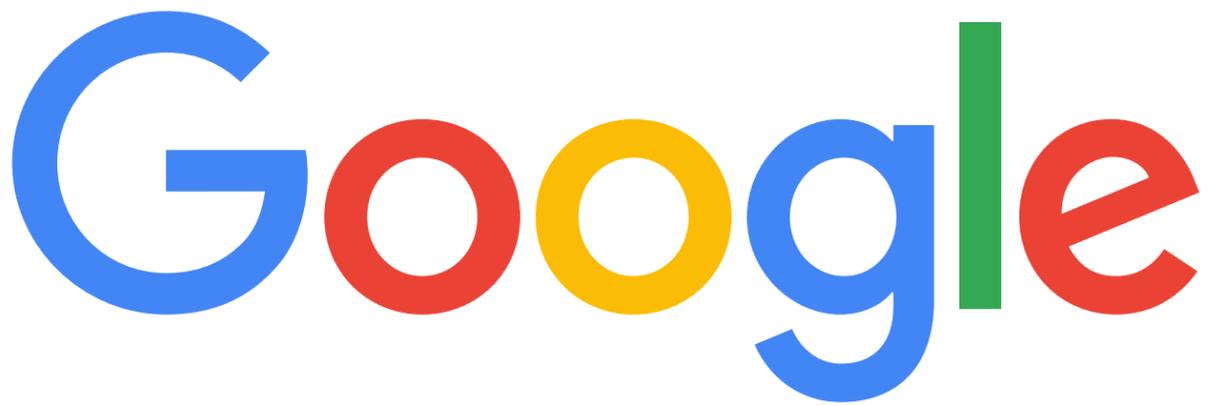
En la mayoría de los países más democráticos, los microdatos –debidamente anonimizados– (ver siguiente lección), se comparten. Un ejemplo es Nueva Zelanda, que entrega **información de sus casi 1.500 enfermos**, incluyendo casos potenciales.

Por otro lado, la fundación chilena Ciudadanía Inteligente hizo un **análisis en Latinoamérica** y concluyó que sólo 4 países entregaban microdatos: Colombia, Cuba, Ecuador y México. Y Nicaragua, ningún tipo de dato, ni siquiera agregado.

LECCIÓN 6: LA PRIVACIDAD EN TIEMPOS DE PANDEMIA

Otra cuestión fundamental es la privacidad de los datos, empezando por los microdatos. Para utilizarlos hay que eliminar cualquier característica





que identifique a una persona, asegurando que ningún individuo quede en un grupo de menos de 50 personas indistinguibles (en inglés, esto se conoce como *k-anonymity*). Ello implica usar rangos de edad, distritos geográficos, etc. Si en una región geográfica hay menos de 50 casos, debemos unirla a otra. Por supuesto 50 es un parámetro de la solución, podrían ser 100 si queremos incluso más privacidad.

La privacidad se ha cuestionado mucho a partir las apps de móvil presentadas por diferentes gobiernos, para tener localizados los posibles contagios que se produzcan en el desconfinamiento. El pasado **17 de abril, el Parlamento Europeo aprobó una resolución** donde hace una llamada a los operadores de telecomunicación para que entreguen los datos anonimizados y agregados de sus usuarios, para “monitorear movimientos, contactos y datos de salud”.

Unos días más tarde, **científicos, informáticos y matemáticos de todo el mundo** firmaron una carta conjunta para alertar del peligro del uso



de aplicaciones móviles que no respeten la privacidad. Por ello, **apoyan la propuesta de Google y Apple** que asegura la protección de los usuarios de una manera descentralizada. Esta medida estaría en sintonía con el enfoque descentralizado del Parlamento Europeo. Francia y **Reino Unido apuestan por un sistema alternativo.** España todavía no ha decidido su posición, aunque podría inclinarse por una **solución facilitada por el Barcelona SuperComputing Center**, donde no se identifica a individuos ni lugares, sino claves aleatorias de teléfonos que han estado a dos metros de la posible persona contagiada. Como siempre, la app será más o menos útil en función de la cantidad de gente que la tenga descargada. Y con tanta variedad de apps se puede dar el caso que una persona pase por diferentes países europeos, y que el rastreo no sea efectivo por incompatibilidad. Para poder evaluar el impacto de las apps recientemente se **ha propuesto un marco socio-tecnológico.**

Dicho esto, ¿es la privacidad de los datos el precio que debemos pagar para sobrevivir a una pandemia?, tal como se pregunta en el **documento colaborativo elaborado por diferentes expertos** (entre ellos, Ricardo Baeza- Yates, uno de los firmantes de este artículo). El estudio concluye – entre otras cosas- que esta es una falsa dicotomía y que es posible lograr ambos objetivos. También indica que **tanto la falta de privacidad como la**



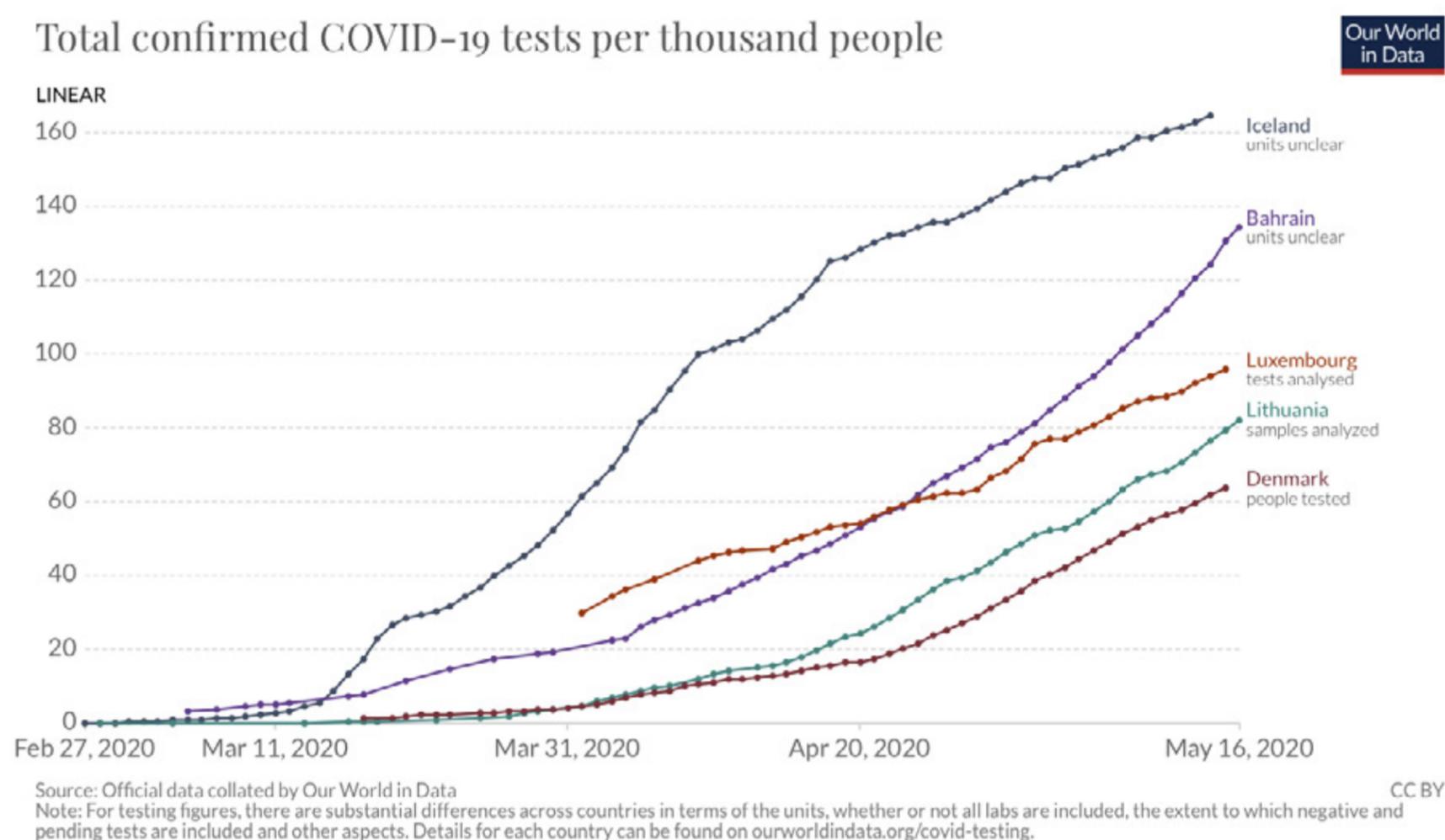
falta de transparencia de los gobiernos en el uso de los datos personales disminuye la confianza de la ciudadanía en el Estado.

LECCIÓN 7: LA OBSESIÓN POR COMPARARSE



Dado todo lo anterior, si los criterios son diferentes de un país a otro, es muy difícil compararlos, aunque midan lo mismo. ¿Tiene sentido comparar dos conjuntos de números si todos los demás factores que rodean la propagación de la enfermedad son diferentes? Pues eso es lo que más se ha hecho, aunque sea muy difícil hacer comparaciones justas entre países. Más aún, se distrae a la opinión pública midiendo los países que están peor, sesgando la comparación, ya sea por diseño o por ignorancia. Y pareciera que esta obsesión tiene un sesgo masculino, tanto en los líderes como en los analistas que lo hacen.

De la segunda lección aprendemos que comparar casos no tiene sentido, pues ellos dependen de la estrategia y número de test hechos. Comparar tests por millón de habitantes tiene sentido, ya que indica cuán bien conocemos la situación. Pero esto favorece a países con poca población, como son Islandia y Emiratos Árabes Unidos que **han testeado a más del %15 de sus habitantes** – si consideramos todos los tipos de tests (al 15 de mayo)–. Pero si sólo consideramos PCR –el test más exacto– el orden cambia, como vemos en el gráfico a continuación.



De la tercera lección podemos rescatar que comparar fallecimientos es lo único que podría tener sentido. Pero para ser justos debemos usar un punto de partida común para eliminar las diferencias temporales, como usar para cada país el **día en el cual ya fallecieron al menos 3**



personas. Por supuesto, los países más poblados prefieren usar muertes por millón de habitantes para salir favorecidos, pero esto es **incorrecto cuando el porcentaje de infectados es una minoría**, ya que el contagio avanza gracias a la densidad de gente en los medios de transporte y zonas más transitadas de las grandes ciudades, lo que no siempre implica más habitantes. Y tampoco hay que usar la tasa de letalidad, pues al final sólo depende del virus y de la demografía y salud de la población. Si lo hacemos, sólo estamos midiendo la prevalencia del contagio en los tests que hacemos.

¿Y por qué tenemos que compararnos? Mil fallecidos más o menos no significa que lo estamos haciendo peor o mejor, en ambos casos la situación es grave. Más aún, **para aprender de otros, no siempre necesitamos compararnos.**

EN FUTURAS PANDEMIAS...

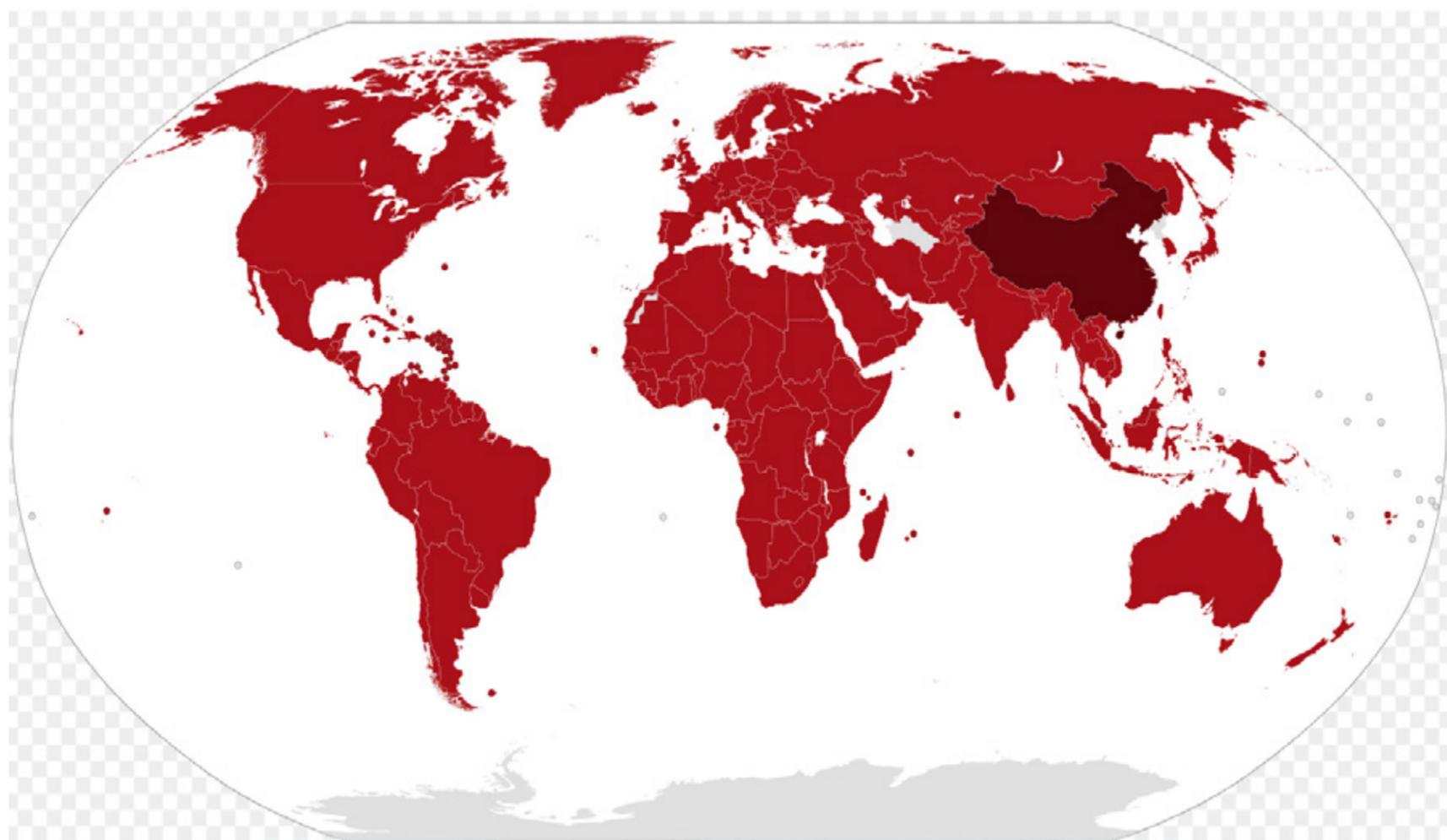
Esta crisis nos ha enseñado que los datos son prioritarios para avanzar y resolver. La OMS tendría que haber dado a los gobiernos un protocolo de recogida de datos confiables. No lo hizo para el COVID19-, pero confiamos que lo haga para futuras pandemias. Por otro lado, si en todas las reuniones de urgencia gubernamentales se hubiera incluido a una persona Responsable de los Datos, muy probablemente se hubiera comunicado



la evolución de la pandemia de una manera más clara y real. En una crisis es difícil tener datos de calidad, pero siempre se pueden tener protocolos claros para hacer lo mejor posible en una emergencia.

La primera lección nos indica la importancia de la recogida de datos que en este caso depende principalmente de las políticas de testeo. Por otro lado, los datos son imprecisos por la naturaleza de la pandemia y de la diversidad de criterios para contabilizarlos (lecciones 2 y 3), así que en este tipo de crisis, calidad al final significa veracidad, no exactitud. La lección 4 nos agrega la dimensión temporal que complica aún más la interpretación de los datos. Vemos además que es deseable tener datos abiertos resguardando la privacidad

Mapamundi del brote del Covid-19



(lecciones 5 y 6), que sin embargo **no sirven de nada si no son datos de calidad**. Esto ha pasado en México, **donde la transparencia de los datos ha reflejado la poca calidad de ellos**. La última lección nos enseña que no es necesario hacer comparaciones, más aún cuando es difícil que sean justas.

Quizás, antes de acabar, debiéramos plantearnos algunas preguntas éticas, surgidas durante la pandemia. **¿Es el beneficio social más importante que el individual?** ¿A cuántas personas podemos afectar si no hacemos un esfuerzo colectivo? Son conocidas las **declaraciones de los presidentes Trump y Bolsonaro** poniendo por delante la economía a la protección de las personas mayores. En una crisis socio-económica como la que estamos viviendo, ¿el gobierno tiene derecho a saltarse los derechos adquiridos con la promesa del bien común? ¿Hasta cuándo?

No tenemos todas las respuestas, pero las lecciones aprendidas deberían servirnos para mejorar las estrategias de los próximos meses y años, porque **después del coronavirus ya nada será igual**. Una primera medida sería tener planes de emergencia preparados para nuevas crisis. Otra, que se considerase la transparencia de los datos como un derecho ciudadano prioritario. Y una última, **que el acceso a Internet sea un nuevo derecho humano**, para que la brecha digital no ahonde más en las desigualdades. **ep**



Ricardo Baeza-Yates es Ph.D. en Ciencia de la Computación y Director de Ciencia de Datos de Northeastern University en el Silicon Valley, además de investigador a tiempo parcial en universidades de Cataluña y Chile. Es Fellow de ACM e IEEE.

Karma Peiró es periodista especializada en Tecnologías de la Información y la Comunicación desde 1995. Sus intereses son la ética de la inteligencia artificial y la transparencia algorítmica.

Este ensayo se publicó en El Periodista en 4 partes:

<https://www.elperiodista.cl/siete-lecciones-para-lidiar-con-los-datos-de-una-pandemia/> (2020/5/18)

<https://www.elperiodista.cl/siete-lecciones-para-lidiar-con-los-datos-de-una-pandemia-cap2/> (2020/5/20)

<https://www.elperiodista.cl/siete-lecciones-para-lidiar-con-los-datos-de-una-pandemia-parte-iii/> (2020/5/22)

<https://www.elperiodista.cl/siete-lecciones-para-lidiar-con-los-datos-de-una-pandemia-parte-iv/> (2020/5/24)



ep » **el periodista**

www.elperiodista.cl