



Datos y pandemia en Chile





Foto: IMFD

JORGE PÉREZ

Profesor Asociado del Departamento de Ciencias de la Computación de la Universidad de Chile e Investigador del Instituto Milenio Fundamentos de los Datos. Sus intereses incluyen: datos Web, teoría de redes neuronales profundas, y el análisis de texto en medicina y política. Ha publicado en eventos como SIGIR, SIGMOD, VLDB, WWW, ICLR, IJCAI y NeurIPS. Su trabajo ha recibido el premio al mejor artículo en cinco conferencias internacionales. En Twitter lo encuentras como @perez.



Desde la llegada de la pandemia del COVID-19 a nuestro país, la relación entre datos, información oficial, y decisiones técnicas y políticas fue tensa. Llegamos al punto que, al momento de escribir este artículo, las autoridades a cargo de las decisiones iniciales respecto de la pandemia se encuentran enfrentando una investigación de la fiscalía nacional. La razón de la investigación es su presunta responsabilidad en muertes por COVID que podrían haberse evitado. La arista principal de esta investigación es un posible ocultamiento de información clave que habría impedido la toma de decisiones correctas y oportunas. Independiente del desenlace de esta investigación, el hecho es que tanto desde el punto de vista técnico como político, el manejo de la información durante la pandemia ha dejado mucho que desear.

En este artículo me centraré principalmente en cómo la información en torno al COVID ha moldeado las acciones de la autoridad y del mundo científico. Éste es un relato bastante personal y subjetivo a partir de mi experiencia “metido en los datos” de COVID desde marzo de 2020. El objetivo del artículo es principalmente entender qué aprendizajes podemos sacar de las situaciones desde el punto de vista técnico de cara a lo que todavía nos

Como es la autoridad la que decide qué cuenta, es muy tentador dejar datos fuera de estos conteos si no se alinean al discurso político oficial.

espera en esta pandemia. Pero no es sólo la pandemia actual y sus repercusiones, de este episodio también podemos obtener aprendizajes importantes para situaciones similares que se podrían dar en el futuro. Dado lo personal del relato, invito al lector a contrastar lo que acá expongo con hechos y opiniones de otras fuentes.

Cifras versus datos desagregados: desechando la oportunidad de cooperar

Los estamentos técnicos del Ministerio de Salud de Chile tienen una larga tradición de uso y publicación de datos clínicos estadísticos. En marzo de 2020, el Departamento de Epidemiología y el Departamento de Estadísticas e Información de Salud (DEIS) tomaron el control de la comunicación de la información relacionada con la pandemia. Crearon un sitio web en donde comenzaron a publi-

car información detallada de cada caso de COVID en Chile, incluyendo la edad, sexo, lugar de atención, y comuna de residencia de cada persona contagiada, además de una indicación de la severidad en la que estaba afectada su salud a causa del COVID. Si bien la publicación no era óptima (los datos se entregaban en un documento en formato pdf), los datos estaban disponibles y varios comenzamos a recopilarlos pues entendíamos que serían valiosos en el corto plazo. Este tipo de publicación es lo que la gente de computación e informática llama “datos desagregados” (también llamados “microdatos” en el área de estadística y economía).

Lamentablemente esta apertura duró sólo unos pocos días. Desde el 18 de marzo de 2020, los datos pasaron a publicarse en forma “agregada” en una tabla que indicaba sólo la cantidad de contagiados por región. Pasamos de contar con datos detallados para tomar decisiones y hacer proyecciones, a contar con cifras macro que esencialmente permiten hacerse una idea global de la situación. Aún no está claro por qué se decidió cambiar la forma de publicación, pero el hecho es que hasta el momento de escribir este artículo no hay datos desagregados de público acceso con el nivel de detalle que contábamos a inicios de marzo.

Con esta decisión, la autoridad política y sanitaria estaba esencialmente dejando fuera la posibilidad de que terceras personas pudieran apoyar, cooperar, evaluar el desempeño de servicios públicos y auditar las decisiones que nos afectarían a todos y todas. Si bien ha habido esfuerzos para lograr poco a poco más apertura en cuanto a datos disponibles, todos estos



1 | <https://www.minsal.cl/wp-content/uploads/2020/03/2020-03-13-Casos-Confirmados-Covid-19.pdf>.

esfuerzos han sido más dolorosos y lentos de lo que se podría esperar. Esto contrasta con varios países de la región (ver el caso de Colombia, Argentina y México) los cuales utilizaron sus plataformas oficiales de datos abiertos para publicar información detallada y desagregada de los casos contagiados (y a veces también sospechosos) de COVID-19.

La publicación de datos de manera agregada en la forma de cifras macro tiene varios efectos perjudiciales. Por un lado, aumenta la presión sobre los profesionales que calculan y publican las cifras pues son ellos quienes deben atender todos los requerimientos por parte de otros estamentos que necesitan datos para su gestión. Con datos desagregados públicos, cada estamento sería responsable de sus propios cálculos adaptados a su realidad y necesidad. Contar con cálculos adaptables a cada realidad es imprescindible en un país como Chile en el que hay gigantes desigualdades territoriales en acceso a servicios de salud y personal capacitado. Esta presión nos hizo rápidamente caer en un círculo vicioso en donde estamentos de la sociedad civil exigían más cifras lo que ponía más peso en las personas encargadas de generarlas con la subsecuente baja en la calidad y periodicidad de publicación de reportes. Fue así como algunos reportes que se producían diariamente, pasaron a producirse un par de veces por semana hasta que simplemente dejaron de publicarse. El caso más emblemático de esto fue el "Reporte de Situación Epidemiológica²", un informe con varias cifras agregadas que se publicó diariamente hasta el 27 de abril, luego bajó su periodicidad hasta que a mediados de mayo simplemente dejó de publicarse. Algunas cifras publicadas en este reporte, notoriamente las que tenían que ver con trazabilidad de casos reportados, no volverían a aparecer en ningún otro reporte hasta varios meses después.

Algunos reportes que se producían diariamente, pasaron a producirse un par de veces por semana hasta que simplemente dejaron de publicarse.

El segundo efecto de no compartir datos detallados desagregados es que la autoridad pasa a ser la única encargada de mantener la consistencia en las cifras que calcula y entrega. Errores, retrasos e inconsistencias causan confusión y minan la confianza de los consumidores de esas cifras. Finalmente, como es la autoridad la que decide qué cuenta, qué suma o qué promedia, es muy tentador dejar conjuntos de datos fuera de estos contextos si es que los datos de alguna forma no se alinean al discurso político oficial. De esta forma, el manejo y comunicación de datos se convierte en una herramienta principalmente discursiva y se corre el riesgo de traicionar el fin último para el que se debieran usar los datos: en este caso lidiar con el COVID usando todas las herramientas técnicas posibles y basándose en información oportuna y de calidad. Esta tentación por decidir qué contar y qué no, fue increíblemente problemática desde el inicio de la pandemia hasta uno de los momentos políticos más álgidos: la salida de quien hasta ese momento fuera el ministro de Salud de Chile.

En lo que sigue me centraré en dos situaciones técnicas que se desprenden de este último problema.

La semántica de los datos agregados jugando una mala pasada

Uno de los primeros problemas que ocurren cuando lo que se entregan son cifras calculadas sin dar a conocer los datos base, es que quien calcula y publica la cifra es quien le da la *semántica*. Es decir,

quien calcula decide el significado de lo que está publicando. Y jugar con *semántica* puede causar confusiones gigantes. Una cifra crítica en los primeros días fue la de "pacientes recuperados". En este caso la autoridad de salud entregaba un número cerrado que indicaba cuántos pacientes se encontraban recuperados del COVID-19. Pero ¿qué quería decir la autoridad cuando decía "recuperado"?, ¿cuál era la semántica de esa cifra? Cualquiera persona que haya estado enferma y se haya recuperado de la enfermedad, le daría una interpretación sumamente natural al número. Pero la autoridad de salud calculaba el número que reportaba simplemente restando de todas las personas contagiadas, las que habían empezado su enfermedad dos o más semanas antes de la fecha actual. Es decir, sin importar que una persona con un cuadro grave de COVID-19 estuviera internada en un hospital y conectada a un ventilador mecánico, si su enfermedad había empezado hace más de dos semanas, era contada como un recuperado.

Razones técnicas para dar esta cifra pueden haber. Lo que se argumentó en su momento fue que lo que se contaba era a los "no contagiados" porque se creía que una persona con COVID sólo podría contagiar a otras durante los primeros 14 días de su enfermedad. Independiente de si hay consenso o no de este período, lo cierto es que la palabra "recuperado" le entregaba una semántica poco clara a la cifra que se estaba publicando e hizo por mucho tiempo creer a la opinión pública en algo que no era exactamente lo que se estaba reportando.

Un caso similar ocurría con la cantidad de "exámenes PCR realizados" que se

2 | <http://epi.minsal.cl/informes-covid-19/>.



Uno de los problemas cuando se entregan sólo cifras calculadas, es que quien calcula y publica la cifra es quien le da la semántica.

publicaba diariamente. ¿Qué significa “realizar un examen”? Para una persona común que sólo está habituada a la palabra examen cuando se toma una muestra de sangre u orina, la realización del examen coincide con la toma de la muestra. Sin embargo la cifra entregada por la autoridad era la cantidad de muestras procesadas en el laboratorio en un día. Otra vez un problema de semántica. En este caso la semántica podría esconder varias otras cifras relevantes. Por ejemplo, si lo que se informa son las muestras procesadas, sería interesante saber cuánto tiempo tardó en ser procesada esa muestra desde la fecha en que la muestra fue obtenida. Esta información era virtualmente imposible de obtener desde las publicaciones del Ministerio de Salud, pero las estimaciones decían que una cantidad considerable de muestras podrían tardar más de dos semanas en recibir un veredicto.³

Combinando las dos cifras anteriores podemos obtener conclusiones que hasta parecen contradictorias. Por ejemplo, si un paciente presuntamente contagiado de coronavirus recibe el resultado de su examen como positivo dos semanas después de que se tomó la muestra, al momento de recibir el resultado ese paciente ya está “recuperado” según las cifras oficiales (porque su enfermedad comenzó más de dos semanas antes). O sea, para efectos de las cifras, ese paciente nunca estuvo “activamente contagiado” de COVID-19. Lo que parece irrisorio es un artefacto de la semántica que la autoridad le puso a cada una de las cifras que entregaba.

Fallecidos por COVID-19 y las confusas metodologías de cálculo

Saber cuántas personas mueren producto de la pandemia es una forma de cuantificar el impacto que realmente está teniendo en la sociedad. Varios países a los que la pandemia llegó antes que a Chile habían mostrado una alza en el número de fallecidos totales (independiente de la causa), alza que en algunos casos era extrema comparada con promedios históricos de fallecimientos. Si bien determinar el exceso real de muertes en un período de tiempo es un problema extremadamente complejo, cuando el número de muertes en un período de tiempo es muy alto es difícil desconocer el exceso, independiente de que puede haber discusión sobre la magnitud de éste.

La periodista Alejandra Matus fue la primera en denunciar un presunto exceso de muertes que no estaba siendo razonablemente reportado por las autoridades de salud. En el momento de estas declaraciones, las muertes COVID reportadas por la autoridad eran relativamente pocas y no explicarían la supuesta alza de muertes. La conclusión sería entonces que habría muertes producidas directa o indirectamente por COVID que no estaban siendo reportadas por la autoridad. En ese momento el ministro de Ciencia, Andrés Couve, salió a desmentir públicamente a Matus, criticando su metodología para calcular el exceso. Pero la discusión parecía totalmente anacrónica: tanto Matus como Couve argumentaban sin mostrar ninguna cifra base, sólo con cifras agregadas. Es decir, ninguna

persona (que no fueran ellos) podría corroborar o refutar las aseveraciones de ambos. Unos días después Matus hizo públicos sus datos y expuso sus cálculos y metodologías al escrutinio público. Los datos que supuestamente usó Couve, no los pudimos ver públicamente sino hasta el escándalo que terminó con la salida de quien hasta ese momento era el ministro de Salud.

Las discrepancias y polémicas en los fallecimientos por COVID continuaron por varias semanas. Otra vez el problema era la falta de datos base que permitieran entender cuándo una persona fallecida tenía (o no) COVID como causa principal de muerte. Uno de los momentos más complejos fue cuando a inicios de junio de 2020 se agregaron alrededor de 700 fallecidos por COVID al conteo oficial. Esta alza fue explicada por un “cambio de metodología” y durante esa semana se explicaron tres métodos distintos para contar fallecidos por COVID. Uno de estos métodos, y que fue el que supuestamente se usó para los reportes diarios de fallecidos, estaba basado en búsqueda de palabras clave dentro de los certificados de defunción ingresados al Registro Civil. Pero la búsqueda de palabras clave es un proceso muy poco preciso y demasiado propenso a errores, sobre todo cuando se busca en texto libre escrito por personas.

Estas metodologías estaban lejos de seguir estándares internacionales y contrastaban radicalmente con el hecho de que en Chile el DEIS es el organismo encargado de contar fallecidos y establecer causas de muerte. Este organismo dependiente del mismo Ministerio de Salud está compuesto por un grupo de profesionales entrenados y especializados en el tema. Por qué no simplemente confiar en estos profesionales y usar su conteo. A mediados del mes de junio de 2020, se dio a conocer a la opinión pública que los fallecidos con COVID como causa princi-

3 | <https://docs.google.com/document/d/1ZGexVNzrNPHRzVPjIRM1t3FLmHrJE1m28Fid1T3sXEM/edit>.



pal de muerte calculado por el DEIS era casi el doble de las muertes reportadas oficialmente (3.841 versus 7.144 al día 18 de junio de 2020). Si bien la diferencia se explicaba por las metodologías distintas, la polémica fue suficiente para implicar la renuncia del ministro de Salud de ese momento.

Al momento de escribir este artículo todavía existen discrepancias entre el valor de fallecidos con COVID como causa principal reportados por el DEIS y la cifra oficial que diariamente se entrega a la ciudadanía. Esto porque la autoridad ha insistido en contabilizar como fallecidos por COVID sólo a fallecidos que cuentan con una muestra de test PCR positiva. Es decir, aunque el diagnóstico clínico indique que la causa principal de muerte es COVID, si la persona fallecida no tiene una muestra PCR positiva, no se informa oficialmente en la cuenta diaria de fallecidos.

Cabe destacar que desde junio el DEIS es el único organismo del Ministerio de Salud que entrega pública y periódicamente

datos desagregados y anónimos, en su caso, datos anónimos de personas fallecidas y sus causas de muerte.

Qué podemos aprender para el futuro

Lo que describo en esta nota son sólo algunos ejemplos de los problemas que surgen cuando no hay público acceso a datos base desagregados que nos permitan auditar y así tener confianza en las cifras reportadas. No contar con estos insumos esenciales desde el inicio de la pandemia dejó fuera a muchos actores que podrían haber cooperado en la toma de decisiones basadas en evidencia. Por otro lado los problemas semánticos y metodológicos de las cifras agregadas causaron más de una confusión en la opinión pública. Estas confusiones inciden en el comportamiento de cada persona, y en una pandemia el comportamiento tanto individual como colectivo

afecta directamente en cómo evoluciona la enfermedad.

El COVID ha impactado fuertemente a nuestro país y mientras no exista una vacuna de amplio acceso en la población, las decisiones basadas en evidencia se vuelven críticas. Debemos aprender todos y todas que más transparencia en la entrega de información técnica es siempre mejor que menos transparencia. Es verdad que entregar información puede abrir espacio a críticas, pero las autoridades no deberían temerle a las críticas de sus ciudadanos, menos a las de sus científicos. Las críticas y las discrepancias nos permiten dudar, y el dudar y repetir un cálculo, nos permite asegurarnos de que no cometimos algún error, lo que nos lleva al final a tomar una mejor decisión. Más transparencia, aceptar críticas, estar dispuestos a dudar, y por sobre todo querer cooperar, son los aprendizajes que debemos sacar de cara a lo que se nos puede venir en el futuro. ■