
Uso de simulaciones para evaluar métodos estadísticos

Felipe Medina M.

Programa de Bioestadística, Escuela de Salud Pública
Facultad de Medicina, Universidad de Chile

1 de diciembre de 2020

“Revisamos cómo las simulaciones pueden asistir a la inferencia estadística. Primero revisamos el método *bootstrap* para calcular intervalos de confianza y vimos

cómo adaptarlo al ajuste de curvas de germinación para además guiar la selección de modelos. Luego revisamos un estudio basado en simulaciones para evaluar la influencia del diseño experimental sobre la calidad de las estimaciones hechas con diferentes métodos. Finalmente, compartimos un trabajo que propone cómo planificar y reportar estudios de simulación para evaluar metodologías estadísticas.”

→ [Enlace a la grabación](#)

Big data y open data en la pandemia de COVID-19: cuestionamientos éticos de la recolección y publicación de datos de salud pública

SANDRA FLORES ALVARADO

PROGRAMA DE BIOESTADÍSTICA, ESCUELA DE SALUD PÚBLICA, FACULTAD DE MEDICINA, UNIVERSIDAD DE CHILE

MARCELA FERRER LUES

DEPARTAMENTO DE SOCIOLOGÍA, FACULTAD DE CIENCIAS SOCIALES, UNIVERSIDAD DE CHILE

El *big data* es un campo de desarrollo relativamente reciente, que se caracteriza por comprender grandes volúmenes de datos que pueden abarcar poblaciones completas, producirse rápidamente y servir para una multiplicidad de propósitos [1, 2]. En paralelo al aumento del volumen de datos, se ha producido una tendencia hacia la apertura respecto a su acceso, uso o distribución, con miras a la democratización del conocimiento, lo que ha sido conocido como el movimiento de open data [1].

La salud pública ha sido impactada por la llegada del *big data* y *open data*, que han cobrado protagonismo gracias a los múltiples repositorios web y a la acelerada producción de conocimiento científico que ha acompañado la pandemia de COVID-19 [3]. El *big data* y *open data* en salud pública constituyen valiosas herramientas para la toma de decisiones. Sin embargo, introducen importantes cuestionamientos éticos que enfrentan la democratización del conocimiento y transparencia en las políticas de salud, por una parte, y la potencial vulneración a la privacidad de los individuos, por otra. Otro conjunto de cuestionamientos éticos dice relación con los procesos de consentimiento informado para la obtención de los datos y su utilización, tema crucial que no profundizaremos en este artículo.

En cuanto a la democratización del conocimiento y transparencia en las políticas de salud, los datos pueden ser entendidos como un bien de uso público [4, 5]. La OPS/OMS enfatiza en sus guías éticas de respuesta a la pandemia que se debe implementar el derecho a

la información [6], la cual debe ser proporcionada de manera regular, exacta y confiable.

Esto depende directamente de los datos a partir de los cuales se construye esa información, los cuales, no sólo deben ser obtenidos y publicados de manera rápida y de fácil acceso, sino también ser transparentes, confiables y veraces, además de encontrarse apropiadamente documentados y contextualizados [1, 5, 7, 8]. Los datos de buena calidad, tras ser recopilados y pasar por un buen proceso de curatoría, se almacenan a nivel individual y son agregables y vinculables a otras fuentes de datos [1, 9]. La falta a estos requisitos puede conducir a la generación de conocimiento cuestionable, que no resulta de utilidad para la formulación de políticas sanitarias. Los datos confiables y de buena calidad permiten estimar los riesgos de la pandemia y reducir la incertidumbre que la caracteriza. Su ausencia o ambigüedad, en tanto, puede aumentar la incertidumbre y provocar desconfianza en la ciudadanía no sólo con respecto a esta información, sino también respecto a las medidas que se toman para mitigar la situación [10]. La confiabilidad, validez y transparencia de la información ayudan a mantener la credibilidad de las autoridades, lo que facilitaría que la población respete las medidas de protección de la salud dictadas.

Junto a la publicación de datos institucionales de vigilancia epidemiológica, la OPS/OMS plantea la apertura de los datos generados en investigación como una estrategia para enfrentar de mejor manera la pandemia [11, 12]. En ciencias, la idea de abrir los datos

se entiende como algo intrínsecamente bueno, ya que permite su reutilización para el planteamiento de nuevas preguntas y la generación de nuevo conocimiento [13, 14]. Sin embargo, dado que la información de salud se considera un dato sensible, la apertura de esta información ha sido cuestionada éticamente [15]. Los principales cuestionamientos se refieren a que los volúmenes de datos actuales y la velocidad con la que se publican podrían conducir a que el resguardo de la información sensible no sea considerado rigurosamente y se produzcan vulneraciones a la seguridad; los datos no se encuentren apropiadamente anonimizados y resulten fácilmente asociables a otros datos; o no se definan claramente sus términos de uso. Todo esto implica riesgos importantes para el resguardo de la privacidad de las personas.

Tanto en vigilancia epidemiológica como en investigación en salud pública, los datos suelen ser recolectados a nivel individual y por sus potencialidades de análisis son más informativos cuando se comparten de esta manera. Sin embargo, al disponerse de manera individual se encuentran más vulnerables a la exposición de la información sensible y a la vinculación con otras fuentes, lo que puede conducir, entre otros efectos, a la estigmatización de los individuos o grupos a los que pertenecen. La posible re-identificación no implica exclusivamente un riesgo para el individuo que brinda su información, sino que, por su capacidad de vinculación, puede serlo también para otras personas de su familia o comunidad. Por ello, los datos personales deben ser entendidos como información que es personal para un número mayor de personas que aquellas de quienes se levanta el dato o quienes dan su consentimiento para su uso [15]. Esto hace que el problema de la privacidad se extienda más allá de la dimensión individual.

La protección de la privacidad es un valor fundamental en nuestra sociedad, que puede ser cautelado. La tecnología actual permite no sólo recolectar y analizar grandes volúmenes de datos, sino también permite que los registros que contienen datos sensibles sean adecuadamente anonimizados, sin que ello implique no acceder en absoluto al contenido, facilitando el tratamiento ético de la información. Sin embargo, la aplicación de estos tratamientos a los datos no debe depender sólo de la buena voluntad de quienes los producen, sino que requiere de una actualización de la gobernanza en la materia, como ha ocurrido en otros países como parte de la respuesta frente a la pandemia [16, 17]. En Chile, carecemos de directrices respecto al manejo de los datos que se adecúen al contexto de la crisis sanitaria actual.

Una respuesta ética a una emergencia de salud pública requiere datos confiables, de buena calidad y que permitan tanto la auditoría de las decisiones de la autoridad, como la investigación independiente, con miras a

generar políticas públicas oportunas, efectivas y justas. El *big data* y *open data* en investigación y vigilancia en salud pública quedarán instaladas en nuestras prácticas para abordar otros problemas de salud poblacional más allá de la pandemia [18]. Debido a esto, la reflexión respecto a los dilemas éticos en torno a las nuevas tecnologías es algo que no sólo es necesario como una respuesta reactiva al acontecer actual, sino que además dejará lecciones valiosas para el futuro de la ética en salud pública. Sobre todo, la importancia tanto de la privacidad como de la transparencia en el manejo de los datos de salud, con miras a mejorar la credibilidad de las autoridades e investigadoras(es) y, con ello, el apoyo de la ciudadanía.

Referencias

- [1] R. Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. SAGE Publications Inc., 2014.
- [2] F. X. Diebold, "On the Origin(s) and Development of the Term 'Big Data,'" *SSRN Electronic Journal*, 2012.
- [3] U. Gasser, M. Ienca, J. Scheibner, J. Sleight, and E. Vayena, "Digital tools against COVID-19: taxonomy, ethical challenges, and navigation aid," 2020.
- [4] Nuffield Council of Bioethics, "Introduction," in *Public health: ethical issues Introduction*, ch. 1, pp. 1–9, London: Nuffield Council of Bioethics, 2007.
- [5] World Health Organization, "WHO Data Principles," 2020.
- [6] Organización Panamericana de la Salud / Organización Mundial de la Salud, "Promoción de la equidad en la salud, la igualdad étnica y de género, y los derechos humanos en la respuesta a la COVID-19: consideraciones clave," tech. rep., Organización Panamericana de la Salud / Organización Mundial de la Salud, 2020.
- [7] K. Robinson, "A false promise of COVID-19 'big' health data? Health data integrity and the ethics and realities of Australia's health information management practice," *Health Information Management Journal*, p. 183335832094119, jul 2020.
- [8] S. Leonelli, "The disruptive potential of data publication," *Notes and Records: the Royal Society Journal of the History of Science*, vol. 70, no. 4, pp. 393–395, 2016.

- [9] D. Rosenberg, "Data before the fact," in *'Raw Data' is an Oxymoron* (L. Gitelman, ed.), pp. 15–40, Cambridge, MA: MIT Press, 2013.
- [10] B. Hofmann, "The first casualty of an epidemic is evidence," *Journal of Evaluation in Clinical Practice*, vol. 26, pp. 1344–1346, oct 2020.
- [11] Organización Panamericana de la Salud / Organización Mundial de la Salud, "Orientación ética sobre cuestiones planteadas por la pandemia del nuevo coronavirus (COVID-19)," tech. rep., Organización Panamericana de la Salud / Organización Mundial de la Salud, 2020.
- [12] World Health Organization, "Ethical standards for research during public health emergencies: Distilling existing guidance to support COVID-19 R&D," tech. rep., World Health Organization, 2020.
- [13] J. Lezaun, "The escalating politics of 'Big Biology'," *BioSocieties*, vol. 8, no. 4, pp. 480–485, 2013.
- [14] S. Leonelli, "Global data for local science: Assessing the scale of data infrastructures in biological and biomedical research," *BioSocieties*, vol. 8, no. 4, pp. 449–465, 2013.
- [15] C. Aicardi, L. Del Savio, E. S. Dove, F. Lucivero, N. Tempini, and B. Prainsack, "Emerging ethical issues regarding digital health data. On the world medical association draft declaration on ethical considerations regarding health databases and biobanks," *Croatian Medical Journal*, vol. 57, no. 2, pp. 207–213, 2016.
- [16] European Data Protection Board, "Statement by the EDPB Chair on the processing of personal data in the context of the COVID-19 outbreak," 2020.
- [17] M. Ienca and E. Vayena, "On the responsible use of digital data to tackle the COVID-19 pandemic," 2020.
- [18] D. S. W. Ting, L. Carin, V. Dzau, and T. Y. Wong, "Digital technology and COVID-19," 2020.

¿Porcentaje respecto a qué? Un ejercicio con datos de COVID-19

¿Qué enfermedades crónicas tienen los hospitalizados o cuánto se hospitalizan los enfermos crónicos?

MAURICIO FUENTES ALBURQUENQUE

PROGRAMA DE BIOESTADÍSTICA, ESCUELA DE SALUD PÚBLICA, FACULTAD DE MEDICINA, UNIVERSIDAD DE CHILE

Según los datos reportados en los informes epidemiológicos del Ministerio de Salud (MINSAL) sobre la enfermedad COVID-19, un 8 % de los casos confirmados requieren hospitalización, es decir, menos de uno de cada diez casos. Considerando que hay muchos casos asintomáticos que no consultan ni se hacen el examen PCR, se podría esperar que el porcentaje de hospitalización sea aún menor. Podríamos decir entonces que, de manera general para la población chilena, si una persona se contagia de COVID-19 tiene un riesgo o probabilidad menor al 10 % de requerir hospitalización.

Pero ¿es realmente éste el riesgo de hospitalización para todas las personas? Evidentemente que no, y se sabe que la edad avanzada y la presencia de enfermedades crónicas aumenta el riesgo de gravedad de la enfermedad y, en consecuencia, el de hospitalización. Más aún, en las últimas semanas se ha reportado que este perfil de hospitalización ha ido variando.

Si bien aún no se dispone públicamente de informa-

ción individualizada para realizar análisis y estimaciones más precisas sobre el comportamiento de esta enfermedad en Chile, lo que hay hasta ahora permite realizar algunos cálculos relativamente sencillos sobre este aspecto, es decir, la hospitalización.

En sus informes periódicos, el Departamento de Epidemiología del MINSAL publica el porcentaje de casos confirmados que presentan una cierta enfermedad crónica de base, esto separado en pacientes con hospitalización y sin hospitalización. En las figuras 1 y 2 se observan los gráficos y la respectiva tabla entregados en el Informe Epidemiológico N° 115 (el último a la fecha de elaboración de esta nota). Si tomamos el caso de la hipertensión arterial, se observa que cerca del 40 % de los pacientes hospitalizados presentan esta enfermedad, mientras que en los no hospitalizados un poco menos del 10 %. Es importante tener en cuenta que las personas pueden tener más de una enfermedad crónica simultáneamente, lo que no se muestra ni en los gráficos ni en la tabla y eso explica por qué los