

Ciencia de Datos para la Salud Pública: Una conversación necesaria

RENÉ LAGOS BARRIOS^{1,2,3,*}, SANDRA FLORES-ALVARADO^{1,2,4,**},
ANDRÉS GONZÁLEZ-SANTA CRUZ^{1,2,***}, FELIPE A. MEDINA MARÍN^{1,4,****}

¹GRUPO DE CIENCIA DE DATOS Y SALUD PÚBLICA, ESCUELA DE SALUD PÚBLICA, UNIVERSIDAD DE CHILE

²PROGRAMA DE DOCTORADO EN SALUD PÚBLICA, ESCUELA DE SALUD PÚBLICA, UNIVERSIDAD DE CHILE

³DEPARTAMENTO DE SALUD DIGITAL, SERVICIO DE SALUD METROPOLITANO SUR ORIENTE

⁴PROGRAMA DE BIOESTADÍSTICA, ESCUELA DE SALUD PÚBLICA, UNIVERSIDAD DE CHILE

La digitalización está generando una transformación profunda en la sociedad y la pandemia aceleró y profundizó la digitalización en muchas áreas, incluida la salud pública. Numerosas iniciativas surgieron o desplegaron su potencial en este periodo tales como ICOVID Chile, Covid Analytics o el Hospital Digital [1, 2, 3]. ¿Hacia dónde está llevando la digitalización a la práctica e investigación de la Salud Pública? y ¿qué rol está jugando la comunidad salubrista en este nuevo campo de desarrollo? El aniversario número 80 de la Escuela de Salud Pública y 40 del Magíster en Bioestadística de la Universidad de Chile brindan una oportunidad para discutir sobre estos cambios, sus potencialidades y curso. En este artículo destacamos algunos conceptos para delinear los alcances de la discusión.

¿Qué es la digitalización en salud pública?

La digitalización es un concepto amplio que se ha aplicado en salud pública con distintos alcances [4]. Por ejemplo, la política Sistema de Información de la Red Asistencial, SIDRA, informatizó los registros clínicos de los centros de salud que se hacían en papel a través de la implementación de soluciones comerciales y desarrollos informáticos propios, sin producir cambios sustanciales en los procesos asistenciales [5]; la implementación del sistema de trazabilidad de enfermedades de notificación obligatoria, Epivigila, optimizó la vigilancia epidemiológica mediante un manejo dinámico de los datos [6]; y la implementación del sistema de gestión de la demanda como Telesalud ha establecido nuevos canales de comunicación y relación de los centros de salud con los usuarios, pudiendo cambiar más profundamente la forma en que tradicionalmente se han vinculado con la red de salud [7]. Todos son ejemplos de políticas de digitalización en salud pública con distintos niveles de impacto en diferentes instancias de la institucionalidad sanitaria. Así, es posible desagregar este concepto en acepciones más específicas, distinguiendo la digitalización de la digitización y de la transformación digital [4]. En efecto, la

digitización (o informatización) comprende el proceso técnico de transformar registros análogos a digitales; la digitalización, constituye el proceso organizacional para mejorar la eficiencia y los resultados con tecnología; y la transformación digital, consiste en un cambio disruptivo en las relaciones entre los participantes del sistema con el uso de tecnología. Cada uno de ellos constituye un nivel de profundización que engloba a otro.

¿Qué es la ciencia de datos para la salud pública?

Producto de la digitalización, de la disponibilidad de datos y del desarrollo de métodos para explotarlos y analizarlos, ha emergido el campo de Ciencia de Datos para la Salud Pública [8]. La contribución de la ciencia de datos a la salud pública no es nueva. Su vínculo se ha ido estrechando gracias al importante volumen de datos secundarios, generado tanto por los sistemas de salud como por instituciones afines, y se ha implementado para la formulación de modelos predictivos, selección de predictores asociados a una variable respuesta y para la formulación de nuevas hipótesis derivadas del descubrimiento de patrones ocultos en los datos. Estas metodologías se han perfeccionado para el trabajo con volúmenes de datos cada vez mayores en que el ruido, la escala y su complejidad dificultan los abordajes convencionales, aportando técnicas para realizar el procesamiento de datos de una manera sistemática, coherente y consistente entre proyectos [8]. Estos métodos de análisis pueden presentar barreras para la interpretación causal de sus resultados, sin embargo, una predicción precisa es igualmente valiosa para la práctica de la prevención en salud pública. Junto con eso, recientemente existen desarrollos en metodologías computacionalmente intensivas que, simultáneamente, permiten realizar inferencia causal y de ese modo contribuir no sólo a la predicción sino también a la comprensión de los fenómenos en estudio [9].

El desarrollo de estas nuevas tecnologías, frente al aumento del volumen y complejidad de los datos de

* rlagos@ing.uchile.cl

** sflores@uchile.cl

*** gonzalez.santacruz.andres@gmail.com

**** f.medina@uchile.cl

salud pública, requiere que los salubristas desarrollen habilidades propias de la ciencia de datos, esto con el fin de orientar la investigación y práctica de la salud pública basada en la generación de conocimiento a partir de esta información. Este nuevo conocimiento debe aspirar a ser producido de una forma ética, centrada en la interpretabilidad y reproducibilidad de los resultados, y enfocada en la superación de las desigualdades que el trabajo con fuentes de datos de recolección rutinaria podría perpetuar [8]. A partir de esto, la ciencia de datos para la salud pública se define como “el estudio de la formulación y respuesta rigurosa de preguntas para mejorar la salud y bienestar usando un proceso data céntrico que enfatiza claridad, reproducibilidad, comunicación efectiva y prácticas éticas” [8]. Esta definición mantiene el foco propio de la salud pública en la protección y promoción de la salud de las poblaciones, e incorpora la forma en que la emergencia de grandes volúmenes de datos, computadores con gran capacidad de análisis y el trabajo interdisciplinario entre las ciencias de la computación, la estadística y la salud pública se complementan en la búsqueda de este objetivo. En este sentido, trasciende las áreas tradicionales de la salud pública, como la epidemiología y bioestadística, así como también delimita un campo de aplicación y reflexión para la ciencia de datos. De este modo, se configura como una interdisciplina que, a la vez que toma elementos de las disciplinas que le dieron origen, es capaz de delinear un camino de desarrollo propio.

Una propuesta para delimitar la discusión

Sostenemos que para comprender la magnitud y dirección de los cambios en curso se requiere considerar tanto la digitalización en salud pública como la ciencia de datos para la salud pública. La digitalización en salud pública (1 y 3 en el diagrama de Venn) abarca tanto los procesos de informatización, digitalización y transformación digital, así como aborda los procesos en que la tecnología es incorporada en salud pública y produce cambios en los actores y relaciones al interior del sistema de salud. Políticas como el desarrollo de Epivigila, el escalamiento del sistema Telesalud a toda la red pública, la implementación del Hospital Digital y la integración de registros clínicos electrónicos pueden ser comprendidos bajo la perspectiva de digitalización en salud pública. Por otra parte, la ciencia de datos para la salud pública (2 y 3 en el diagrama de Venn) abarca los métodos y las aplicaciones en que los datos pueden ser explotados para que sirvan al propósito de la salud pública. Iniciativas como ICOVID y Covid Analytics pueden ser entendidas desde esta mirada.

A la intersección de estos dos conceptos en un nuevo campo de desarrollo interdisciplinario y de aplicación le denominamos **digitalización y ciencia de datos para la salud pública** (3 en el diagrama de Venn). En es-

ta interdisciplina, ambos conceptos se entrelazan para considerar las necesidades y oportunidades existentes dentro de sus propios procesos, así como en los procesos del otro. El objetivo de esta “simbiosis” apunta a garantizar que los datos y los análisis utilizados para mejorar la salud pública sean claros, reproducibles, eficientes y éticos. Si bien este nuevo campo tiene mucho en común con los ya instalados de Salud Digital y los Sistemas de Información para la Salud [10], éste se diferencia en que su horizonte de aplicación es la salud pública, entendida como la actividad encaminada a mejorar la salud de la población y la respuesta social organizada a sus demandas [11]. Es decir, su horizonte es lo poblacional que, si bien dialoga con el nivel biológico y clínico, lo trasciende.



Diagrama de Venn que representa la relación entre la ciencia de datos, digitalización y salud pública, y delimita las intersecciones de interés para la discusión.

Los límites de este nuevo campo aún son difusos y se confunden con aquellos de las disciplinas que le dieron origen. A esta dificultad se suma la difuminación que se produce entre la información clínica a nivel de paciente y la información poblacional, dada por la gran disponibilidad de datos digitales levantados a nivel individual, que a su vez permiten responder a interrogantes a nivel poblacional. Vemos ahí una oportunidad de diálogo y aprendizaje para los que nos abocamos a estos temas: científicos de datos, investigadores y profesionales de la salud pública y profesionales de la salud digital.

Nuevos desafíos éticos e institucionales

Si se necesita formular y responder preguntas de manera rigurosa con el objetivo de promover la salud y el bienestar, empleando un enfoque basado en datos

que prioriza la reproducibilidad, la comunicación eficaz y la ética, es esencial no solo adaptarse a las nuevas tecnologías, sino también liderar su evolución [8]. Con todo, dicha evolución debe velar por conservar el respeto a los principios de la salud pública, que apuntan a proteger el bien común e intervenir en su favor en contextos de incertidumbre [12]. Parafraseando a Keyes y Galea [13], la tecnología y su vertiginoso desarrollo no deben orientar nuestras preguntas, sino mejorar nuestra capacidad para responderlas. Lo anterior también implica preguntarse por los dilemas éticos e institucionales que ésta genera [8].

Uno de los dilemas éticos más relevantes en este contexto es la pregunta por los intereses a los que responden los modelos de digitalización en salud pública, así como las amenazas que pueden representar para la autonomía y/o la privacidad de las personas. Si bien estos modelos apuntan a optimizar la atención médica, prevenir enfermedades y mejorar la calidad de vida, también pueden estar influenciados por intereses económicos y de control social. Así, pueden representar amenazas para la autonomía y privacidad de las personas, ya que la recolección masiva de datos [14] puede también implicar la descontextualización y abstracción de las diferencias individuales, invisibilizando y/o exacerbando inequidades existentes [15]. Además, la vigilancia y obtención de datos en línea plantean preocupaciones sobre consentimiento, privacidad y responsabilidades de los equipos de investigación [8], por ejemplo, en el hecho de que las personas ingresan de manera activa (ej., instalando una aplicación de monitoreo del sueño, o ingresando su peso), naturalizada y cotidiana al proceso de generación de datos, a diferencia como lo harían en una encuesta [14]. De igual manera, la omnipresencia de la información [4] y la tecnologización de eventos vitales han desencadenado prácticas de autogobernanza y autogestión en materia biomédica [16], lo que puede afectar la relación entre poder y sujeto desde una perspectiva biopolítica [17]. Ejemplos como el fenotipado digital y el monitoreo a través de marcadores de eventos ingeribles [18] evidencian el alcance de la intervención biomédica en relación con las TICs, lo que puede resultar en consecuencias no intencionadas y perjudiciales para la autonomía y privacidad de las personas.

Otro dilema de importancia refiere a cómo evitar el desarrollo de estudios y políticas que ignoren sesgos, que confundan asociaciones y causas, y/o que ignoren el bagaje de conocimiento cualitativo en salud pública. La salud pública se caracteriza por su enfoque crítico y el uso del método científico, reconociendo que algunos métodos novedosos pero opacos, como “cajas negras”, podrían propagar sesgos en los datos [8]. Para evitar el desarrollo de estudios y políticas con lógicas dataístas (archi-empiristas) que ignoren sesgos, confundan asociaciones y causas, se limiten solamente a describir

resultados paradójicos o contraintuitivos, o desestimen el conocimiento cualitativo, es esencial reconocer los factores que limitan o potencian las elecciones individuales [19]. Además, es importante cuestionar la estructura de los datos administrativos [20] y no limitarse a análisis superficiales que reafirmen o simplemente constaten un mundo predefinido [14]. Lo anterior se traduce en la necesidad de equilibrar el enfoque cuantitativo con el cualitativo, evitando caer en un positivismo tecnocrático que sólo reafirme las fronteras de conocimiento sobre la base de lo meramente observable [21] y volviendo a preguntarnos por “las causas de las causas”.

Por último, cabe preguntarnos respecto a la manera en que la explosión y el aprovechamiento de los datos puede contribuir a generar y reproducir inequidades. La introducción de la ciencia de datos en salud puede ser vista como parte de la creciente influencia del bio-capitalismo en la medicina, lo que puede llevar a socavar el bien común y la solidaridad en los sistemas de salud [15]. La globalización capitalista y la visión anti-estadista generan ciudadanos autónomos pero con limitada capacidad para ejercer dicha autonomía, debido a incertidumbres y desprotección [22, 23]. La salud pública debe priorizar la justicia social [24] en un contexto donde los datos personales, considerados “el petróleo del siglo XXI” en palabras de Stefan Groß-Selbeck (CEO 2009-2012 de la red social para profesionales Xing) [21], por lo que están sujetos a una distribución desigual. La brecha digital actual dificulta mantener objetivos de salud pública, y una transformación tecnológica podría agravar las inequidades. Foucault argumenta que la biopolítica responsabiliza a los individuos para gobernarlos eficazmente, lo que lleva a la subyugación y aumento de vulnerabilidad para quienes no acceden a nuevas tecnologías [13, 15]. Por ello, es necesario promover una ciencia de la salud poblacional [13] e impulsar cooperativas y repositorios de datos abiertos en salud pública desde una perspectiva poblacional y humana, pero conservando los ideales modernos ilustrados que fundamentan la práctica científica y que promueven la democratización del conocimiento a personas marginadas [17].

Una conversación necesaria

Para motivar esta conversación, a lo largo de 2023 realizaremos las 1as Jornadas de Ciencia de Datos y Salud Pública, organizadas desde la Universidad de Chile y abiertas a toda la comunidad interesada en ser parte de la discusión. En este evento queremos reunir a las y los investigadores y profesionales que están participando o tienen interés en participar de estos cambios: 1) académicas y académicos que investiguen en salud pública o ciencia de datos, 2) tomadores de decisiones e impulsores de políticas públicas, 3) profesionales

y empresas que impulsan el desarrollo tecnológico en salud pública, y 4) estudiantes y/o profesionales con interés de aprender sobre estas temáticas.

Queremos comprender los cambios en el campo de la salud pública y ser conscientes del rol que desempeñamos como salubristas, académicos y profesionales. Para encaminarnos en la comprensión de este fenómeno, proponemos un marco amplio para la conversación que abarca los principales cambios en el campo de la Salud Pública producto de la digitalización y de su abordaje en investigación a través de la ciencia de datos. Con esto en consideración, invitamos a la comunidad salubrista, científicos de datos y a profesionales de la digitalización a participar de esta discusión sumándose a las 1as Jornadas de Ciencias de Datos en Salud Pública, a realizarse en una serie de eventos a lo largo del año 2023 y que comenzó por una reunión de lanzamiento¹ el 23 de mayo en la Escuela de Salud Pública de la Universidad de Chile, y que será seguida por actividades en agosto y octubre.

Financiamiento

La Agencia Nacional de Investigación y Desarrollo de Chile (ANID) financia los estudios de doctorado de Sandra Flores Alvarado, René Lagos y Andrés González a través de la Beca de Doctorado Nacional (CONICYT-PFCHA/ Doctorado Nacional/ 2020-21200398, Subdirección de Capital Humano/Doctorado Nacional/2022-21220971 y Subdirección de Capital Humano/Doctorado Nacional/2023-21230172). Las 1as Jornadas de Ciencia de Datos y Salud Pública cuentan con el financiamiento del Departamento de Postgrado y Postítulo de la Vicerrectoría de Asuntos Académicos de la Universidad de Chile a través del concurso “Estímulo para proyectos académicos de estudiantes de postgrado de la Universidad de Chile”, convocatoria 2022 – 2023.

Referencias

- [1] ICOVID Chile [internet] [consultado el 3 de mayo de 2023], sitio web <https://www.icovidchile.cl/>.
- [2] Basso LJ, Goic M, Olivares M, Sauré D, Thraves C, Carranza A, et al. Analytics Saves Lives during the Covid Crisis in Chile.
- [3] ¿Qué es Hospital Digital? - Hospital Digital [Internet] [consultado el 18 de noviembre de 2022]. Disponible en <https://www.hospitaldigital.gob.cl/>.
- [4] Iyamu I, Xu AXT, Gómez-Ramírez O, Ablona A, Chang HJ, Mckee G, et al. Defining Digital Public Health and the Role of Digitization,

Digitalization, and Digital Transformation: Scoping Review. *JMIR Public Health Surveill*; 2021; 7(11):e30399.

- [5] Borgoño V. Fin de un Ciclo: Status de Avances en la Implementación de SIDRA. *Revista de Informática Médica*, 2014; 15:42–9.
- [6] Taramasco C, Rimassa C, Acevedo Romo J, Céspedes Zavando A, Fuentes Bravo R. Epidemiological surveillance in COVID-19 pandemic: EPIVIGILA system. *Medwave*; 2022; 22(05):e2560–e2560.
- [7] Telesalud [internet] [consultado el 3 de mayo de 2023]. Disponible en: <https://telesalud.gob.cl/>.
- [8] Goldsmith J, Sun Y, Fried LP, Wing J, Miller GW, Berhane K. The Emergence and Future of Public Health Data Science. *Public Health Rev*; 2021; 42:1604023.
- [9] Deng Z, Zheng X, Tian H, Zeng DD. Deep Causal Learning: Representation, Discovery and Inference. ArXiv Prepr ArXiv221103374; 2022.
- [10] Pan American Health Organization (PAHO/WHO). Information Systems for Health (IS4H) [internet] [consultado el 3 de abril de 2023]. Disponible en: <https://www.paho.org/ish/>.
- [11] Martínez F, Castell PL, Navarro V. Concepto actual de la salud pública. *Salud Pública*; 1998; 49(54):49–54.
- [12] Bayer R, Fairchild AL. The Genesis of Public Health Ethics. *Bioethics*; 2004; 18(6):473–92.
- [13] Keyes KM, Galea S. Setting the Agenda for a New Discipline: Population Health Science. *Am J Public Health*; 2016; 106(4):633–4.
- [14] Savage M. Identities and social change in Britain since 1940: The politics of method. OUP Oxford; 2010.
- [15] Giroux É. Is personalized medicine humanist? *Arch Philos*; 2020; 83(4):59–82.
- [16] Rose N. The human sciences in a biological age. *Theory Cult Soc*; 2013; 30(1):3–34.
- [17] Villarroel R. Consideraciones bioéticas y biopolíticas acerca del Transhumanismo: El debate en torno a una posible experiencia posthumana. *Rev Filo*; 2015; 71:177–90.
- [18] Cosgrove L, Karter JM, McGinley M, Morrill Z. Digital phenotyping and digital psychotropic drugs: Mental health surveillance tools that threaten human rights. *Health Hum Rights*; 2020; 22(2):33.

¹<https://uchile.cl/sp204605>

- [19] Linares-Pérez N, Arellano OL. La equidad en salud: propuestas conceptuales, aspectos críticos y perspectivas desde el campo de la salud colectiva. *Med Social*; 2008; 3(3):247-59.
- [20] Hand DJ. Statistical Challenges of Administrative and Transaction Data. *J R Stat Soc Ser A Stat Soc*; 2018; 181(3):555-605.
- [21] Hurtado A, García J. Smarts cities: Un reto antropológico. *Methaodos Rev Cienc Soc* 2022; 10(2):366-78.
- [22] Kottow M. Bioética: una disciplina en riesgo. *Rev Red bioética Unesco*; 2010; 1(1):158-72.
- [23] Kottow M. La urgente tarea de estructurar una bioética latinoamericana. *Rev Latinoam Bioét* 2018; 19(36-1):5-7.
- [24] Gostin LO, Powers M. What Does Social Justice Require For The Public's Health? Public Health Ethics And Policy Imperatives. *Health Aff (Millwood)*; 2006; 25(4):1053-60.

Fotografías del lanzamiento de las 1ras Jornadas de Ciencia de Datos y Salud Pública.
Créditos: Verónica Zúñiga, Encargada de Comunicaciones Escuela de Salud Pública.



Presentación del proyecto y actividades planificadas para el año 2023.



Discusión en grupos sobre el concepto de Ciencia de Datos para la Salud Pública y sus desafíos.