

Los Big Data: conceptos relacionados y algunas aplicaciones en pediatría

The Big Data: related concepts and some applications in pediatrics

Ana J. Zepeda Ortega^a

^aTecnóloga Médica. Escuela de Tecnología Médica, Facultad de Medicina, Universidad de Valparaíso y Centro Interdisciplinario de Investigación en Salud Territorial (CIISTE)

Recibido: 20 de junio de 2019; Aceptado: 24 de junio de 2019

Resumen

El avance de la tecnología médica, el registro de salud electrónico (EHR, por sus siglas en inglés) y la producción compleja de datos biomoleculares están generando grandes volúmenes de información, en varios formatos y de múltiples fuentes, que se conocen como “Big Data”. El análisis integrado de estos datos ha abierto una amplia posibilidad para explorar respuestas a problemas de salud. En pediatría, se han incrementado los estudios se analizan Big Data o se utilizan las herramientas informáticas que se han desarrollado para analizar estos datos. Los propósitos de esos estudios han sido variados, por ejemplo: en la detección y prevención temprana de una amplia gama de afecciones médicas, mejoramiento de los diagnósticos, para especificar tratamientos o anticipar el resultado de alguna patología, etc. El presente documento tiene como objetivo revisar los conceptos principales involucrados en el análisis de Big Data o en las tecnologías informáticas asociadas, así como también examinar sus aplicaciones, potencialidades y limitaciones actuales. Este estudio se realizó sobre la base de una revisión bibliográfica no sistemática, centrada en el campo de la pediatría. En la selección de los ejemplos de aplicación, se consideró que eran fuentes primarias, publicadas en los últimos cinco años y con poblaciones infanto-juveniles.

Palabras clave:

Big Data;
datos masivos;
pediatría;
macrodatos

Abstract

Medical technology advances, the Electronic Health Record (EHR), and the complex production of biomolecular data are generating large volumes of information, in various formats and from multiple sources, which are known as “Big Data”. The integrated analysis of these data has opened up a wide possibility to explore answers related to health problems. In pediatrics, there has been an increase in the studies on Big Data, or the computer tools use that have been developed to analyze these data. The purposes of those studies have been diverse, for example, for earlier detection and prevention of a wide range of medical conditions, improvement of diagnoses, to specify treatments or anticipate

Keywords:

Big Data;
massive datasets;
pediatrics;
data science

Correspondencia:
Ana J. Zepeda Ortega
ana.zepeda@uv.cl

Cómo citar este artículo: Rev Chil Pediatr. 2019;90(4):376-383 DOI: 10.32641/rchped.v90i4.1306

the outcome of some pathology and so on. For this reason, this contribution aims to review the main concepts involved in the analysis of Big Data or its related computer technologies, as well as to examine their current applications, potentialities, and limitations. This study was carried out based on a non-systematic bibliographical review, focused on the field of pediatrics. In the selection of application examples, it was considered that they were primary sources, published in the last five years and with child and youth populations.

Introducción

En la actualidad, cualquiera de nuestras acciones, ya sea de manera consiente o inadvertidamente, son susceptibles de producir información registrable y analizable. El programa de televisión o la película que seleccionamos para ver, los lugares que frecuentamos, el vestuario, medicamento o cualquier otro artículo que consumimos son algunos ejemplos de nuestros actuar cotidiano que va dejando una huella digital de nuestro comportamiento. Esto ha hecho que cada día se produzca un gran volumen de información, donde se ha estimado que en el mundo se producen diariamente 2.5×10^{30} bytes, es decir 2,5 quintillones de datos¹.

El avance tecnológico, tanto en lo que se refiere al almacenamiento como al análisis de esta producción masiva de datos, promete convertirse en una poderosa herramienta, que si se utiliza adecuadamente podría ser muy beneficiosa para las personas y, en el mejor escenario, redundar en una mejor calidad de vida. Sin embargo, también genera muchas dudas en cuanto a un uso malicioso y controlador por los intereses de grandes compañías de la industria², así como también se advierte sobre los excesos de la cuantificación³, lo que podría llevar de vuelta al reduccionismo cuantitativo, implicando un retroceso en cuanto al valor que ha cobrado la comprensión de los fenómenos en estudio.

Por otra parte, la datificación y la digitalización son dos conceptos que en ocasiones se mencionan como sinónimo, pero no lo son. Tal como lo ilustra Mayer y Cukier, la datificación ocurrió mucho antes de que aparecieran las Tecnologías de Información (TI) y, por ende, es anterior a la digitalización⁴. Esto porque la datificación hace alusión al dato, el cual es un registro de información, en cualquier medio. La transformación de estos datos, por medio del análisis y la reorganización, en información susceptible de ser utilizada en cualquier área del conocimiento (medicina, aeronáutica, física, astronomía, etc.), es lo que se podría llamar “datificación”. Por tanto, antes de que pudiéramos contar con máquinas de procesamiento de datos, en forma manual o con herramientas rudimentarias, ya se efectuaba la datificación, lo cual fue la llave maestra del avance del conocimiento científico. En cambio, la digi-

talización hace referencia al proceso mediante el cual una información que se encuentra en formato analógico se convierte a códigos binarios de ceros y unos, para que así sea legible y puedan ser procesados por un computador.

De esta forma se puede afirmar que la “datificación” en el ámbito de la salud es una tradición muy antigua en cambio, la digitalización ha llegado de la mano con la incorporación de las TI en la atención de salud. Éstas han permitido, entre otros, la generación del Registro Electrónico de datos de Salud (RES o en inglés EHR) o Ficha Clínica Electrónica (FCE), los cuales compilan las historias clínicas de una persona⁵ y que pueden ser consultados por personal de salud⁶. Esta innovación tecnológica ha puesto a disposición una cantidad ilimitada de información susceptible de ser analizada, la cual abre múltiples oportunidades para explorar soluciones clínicas o administrativas que vayan en beneficio de la salud de un individuo, pero también, se abren flancos de precaución en cuanto al tratamiento confidencial de los datos clínicos y sus aspectos legales, entre otros.

La pediatría no está ajena a esta realidad, pues cada vez se emplean tecnologías más avanzadas que están generando grandes volúmenes de datos, los cuales impactarán en la forma en que se conduzcan las investigaciones en el futuro inmediato. De hecho, la digitalización creciente es mencionada como uno de los desafíos y oportunidades que tendrán que enfrentar los pediatras de las próximas décadas⁷.

Este artículo tiene como objetivo describir los conceptos y terminología relacionada con la producción masiva de datos, que son conocidos por el tecnicismo “Big Data”. Así también, se pretende brindar algunos ejemplos en los cuales se están usando este tipo de datos en el ámbito de la pediatría. Finalmente, se hace una reflexión sobre la ventana de oportunidad que generan los Big Data y los retos que conlleva como para que se conviertan en un verdadero beneficio para proveedores de salud pediátricos. Este trabajo fue realizado sobre la base de una revisión bibliográfica no sistemática, centrada en ámbito de la pediatría y para la elección de los ejemplos de aplicación se consideró que fueran de fuentes primarias y reportados en los últimos cinco años.

Conceptos y terminología

1. Macro, Micro y Metadatos

Los Macrodatos o Big Data, hasta ahora no tiene una definición muy precisa, pero hay cierta coincidencia en definirlos como aquella información recopilada electrónicamente que, por su volumen, formato y diversidad, no puede ser procesada con las herramientas informáticas comunes, por lo que requiere de soluciones computacionales de alta complejidad⁸⁻¹¹.

Por otra parte, los Microdatos corresponden a los registros de las características de las unidades de análisis de una población en estudio (individuos, hogares, establecimientos, entre otros), los cuales son recogidos por medio estadísticos formales. Por ejemplo, los registros de atención de salud permanente, las encuestas nacionales de salud y censos de población podrían dar origen a Microdatos.

Por último, los Metadatos se refieren a aquella información que describe a otros datos, es decir, corresponden al contenido informativo de algún recurso informático. Por ejemplo, en una fotografía el metadato sería la información adicional que registra la cámara, como sería la fecha y hora.

2. Características, estructura y procesamiento de los Big Data

2.1. Las características "V"

La mínima cantidad de información que puede ser procesada por un aparato tecnológico es el bit, el cual sólo puede ser expresado en ceros o unos, mientras que un byte es un conjunto de 8 bits. Así entonces, un Gigabyte (Gb) corresponde a 10^9 byte, un Terabyte (Tb) a 10^{12} , un Petabyte (Pb) a 10^{15} y un Exabyte (Eb) a 10^{18} bytes. Por tanto, cuando se habla de datos masivos, se está refiriendo a una generación de datos del orden de los Pb y Eb o superior. Por lo tanto, uno de sus primeros distintivos es el volumen⁸.

El otro aspecto que define a los Big Data es la velocidad, lo cual hace alusión al ritmo en que los datos se producen, pero también a la capacidad de respuesta en el análisis como para hacer de la información un producto que dé soluciones innovadoras y eficientes⁹.

La variedad es otra característica de los datos masivos, lo cual hace referencia a las diferentes fuentes y tipos de datos que lo conforman. Por ejemplo, pueden ser archivos de texto o de datos, imágenes, videos, datos de sistemas de posicionamiento global (GPS), de sensores digitales de equipos (médico, industriales, medidores de electricidad, etc.). Estos pueden registrar información de posición geográfica, variaciones atmosféricas, cambio de temperatura, velocidad de movimiento, entre muchos otros¹².

Las tres características mencionadas anteriormente

—volumen, velocidad y variedad— son conocidas como “las tres V” de los datos masivos. Luego se han descrito otras, como son: Veracidad, que sería el grado de fiabilidad de los datos; Validez, que sean datos correctos y precisos para el análisis deseado; Volatilidad, se refiere al tiempo en que permanecen de los datos para ser analizados; Viabilidad, capacidad de uso eficaz de los datos; Visualización, que sería la capacidad de poder identificar patrones y claves útiles; y por último el Valor, lo cual se refiere a la posibilidad que dichos datos puedan convertirse en conocimiento^{9,13}.

2.2. La estructura

La distribución y orden de los Big Data pueden ser estructurada, no estructurada o una mezcla de ambos (semiestructurados). Los datos estructurados suelen tener una longitud, formato y tamaño claramente definidos, tales como bases de datos en planilla de cálculo o formatos de paquetes computacionales de análisis de bases de datos. Así entonces, a una base de Microdatos se podría decir que en general son datos estructurados. En cambio, los datos no estructurados no tienen un formato específico, es decir, pueden haber documentos con múltiples formatos, tales como: documentos portables o de procesadores de texto, correos electrónicos, carpetas de imágenes, audio o video, entre otros (Figura 1). Finalmente, los datos semi estructurados son una mezcla de los dos anteriores, donde no hay una estructura, pero si hay una cierta organización en sus metadatos, como es el caso del Hipertexto⁹.

2.3. El procesamiento informático y el análisis estadístico

Las fuentes de producción de datos masivos pueden tener múltiples orígenes, como son: las redes sociales (Facebook, Twitter, Instagram, blogs, etc.); comunicación entre máquinas (equipo móvil a computador fijo, sensores o medidores de evento a otros dispositivos; cámaras de seguridad a su central, etc.); registro de transacciones (bancarias, compras, registro de llamadas, entre otras); biometría (huellas digitales, escaneo de la retina, reconocimiento facial, etc.) y generación humana (mensajería de texto o voz, correos electrónicos, etc.)¹⁴.

Así entonces, el mayor reto que plantean los Big Data es convertir los datos en información utilizable. Para ello, se han desarrollado plataformas y herramientas computacionales que apuntan básicamente a: la captura, el procesamiento, el almacenamiento y el análisis (Figura 2). A modo de ejemplo, en el ámbito de la captura se han creado interfaces de programación de aplicaciones (Application Programming Interfaces, API) especiales para permitir la comunicación entre diferentes programas computacionales que realizan la captación de datos. Para el procesamiento, se ha ge-

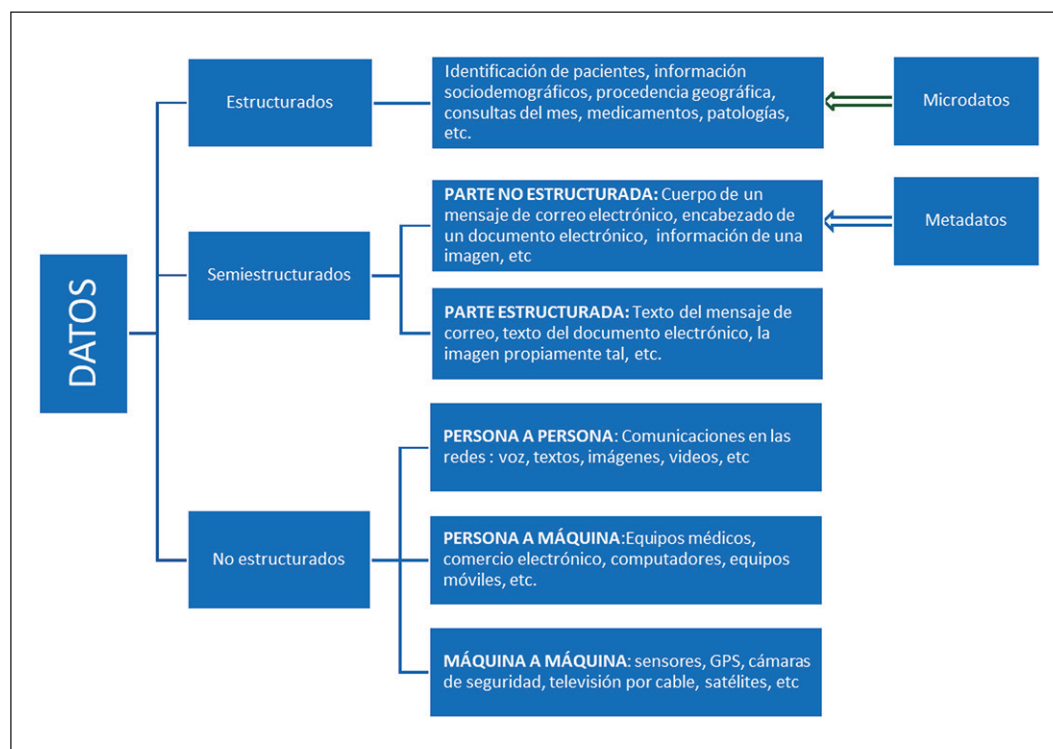


Figura 1. Fuentes y tipos de datos que pueden conformar los Big Data

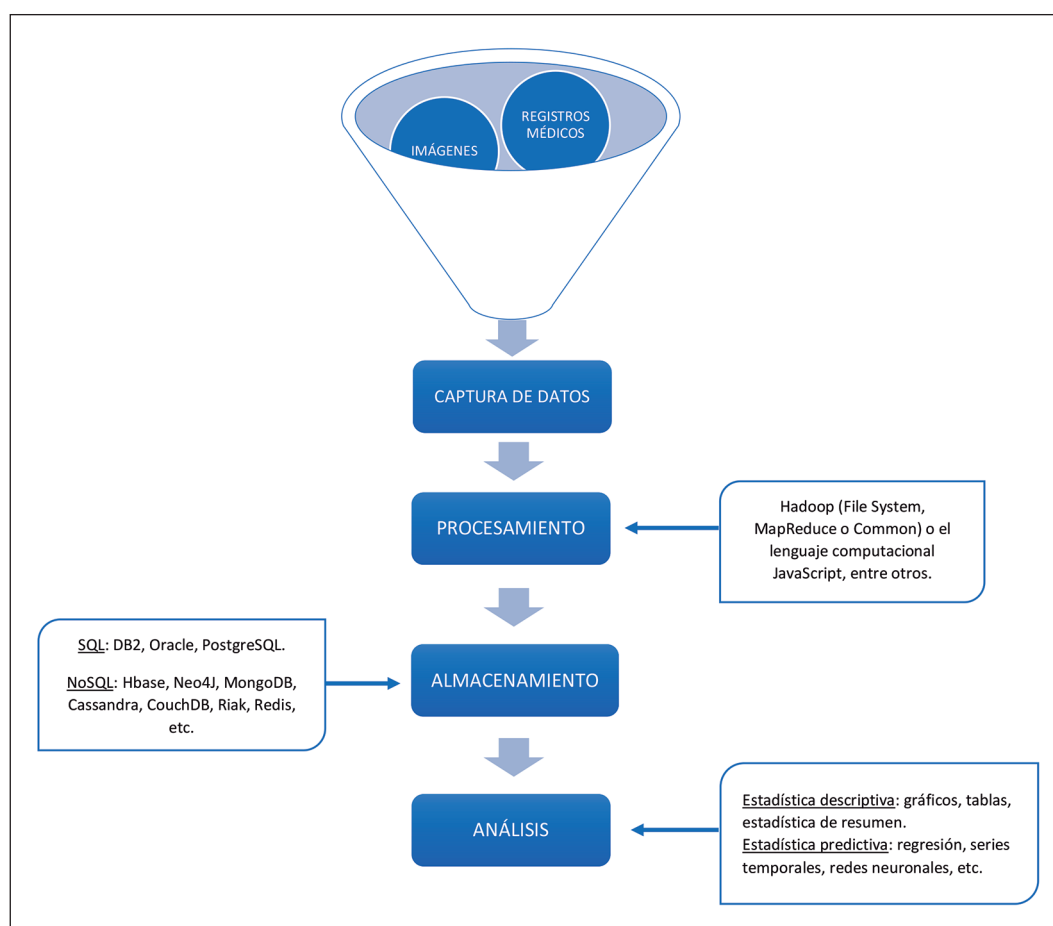


Figura 2. Etapas de la gestión de los Big Data.

nerado la plataforma de código abierto Hadoop, tales como el Distributed File System (HDFS), el MapReduce y el Common, así como también el lenguaje de programación JavaScript¹². En cuanto al almacenamiento, las bases de datos se han tenido que trabajar según su estructura y el lenguaje informático de consulta a la base de datos o sistemas de información (Query Language, QL). Por tanto, si la base de datos es de lenguaje de consulta estructurada, se usan herramientas SQL (Structured Query Language), tales como: DB2, Oracle, PostgreSQL, etc. Por contraste, el lenguaje no estructurado se llama NoSQL y dentro de éste, se puede mencionar a Hbase, Neo4J, MongoDB, Cassandra, CouchDB, Riak, Redis, entre otras^{12,14,15}. Finalmente, para el análisis y dependiendo del propósito, esto van desde la estadística básica y univariada hasta la multivariada. Ejemplos de estas últimas son los análisis de regresión, las series temporales y las redes neuronales, entre otros^{8,16}.

3. La ciencia de los datos y su relación con los Big Data

La ciencia de datos es un concepto que, al igual que el de Big Data, aún no está enteramente definido. Sin embargo, se acepta que se trata del “estudio científico de la creación, validación y transformación de datos para crear significado”, es decir, la ciencia que permite extraer valor y conocimiento de los datos. Por tanto, los Big Data están intrínsecamente relacionados con la “ciencia de datos” debido a que son su materia prima^{17,18}.

4. Aprendizaje automático y sus modalidades

El aprendizaje automático o Machine Learning se refiere a la capacidad de las computadoras de poder aprender de los datos. Esto está cimentado en el desarrollo de algoritmos que permiten que las máquinas puedan cambiar su comportamiento, basándose en el análisis de los datos nuevos que está recibiendo^{17,19}. Actualmente, existen varias modalidades de este tipo de aprendizajes, entre los que se cuentan los aprendizajes supervisados, no supervisados y los profundos (*deep-learning*). De manera sencilla, se puede decir que un aprendizaje supervisado es cuando el equipo recibe información sobre las características de lo que tiene que identificar y las respuestas correctas (etiqueta). Esto se puede hacer con un propósito de predicción o de clasificación. Un ejemplo de predicción sería pronosticar la muerte de un paciente, hecha en base al análisis de un conjunto de características clínicas, tales como: signos vitales, imagenología, determinaciones de laboratorio, entre otros²⁰. En cambio, para un fin de clasificación, al equipo se le entregan patrones de sujetos y con estos debe ubicar a nuevos individuos dentro de alguno de estos perfiles²¹.

Por otro lado, en el aprendizaje no supervisado, el algoritmo del equipo no posee la respuesta correcta, por lo que debe generar los perfiles o predicciones solo con la información histórica que se le entrega. Por ejemplo, se le podría solicitar identificar a grupos de pacientes con ciertas características clínicas y luego, si esta clasificación tiene algún sentido clínico, se podría utilizar para tomar decisiones, por ejemplo, sobre el beneficio de algún determinado tratamiento que algún grupo podría tener²².

En otro orden, el aprendizaje profundo se refiere a la forma en que la máquina va obteniendo la “respuesta correcta”. El procedimiento se va realizando por capas o nodos, de menor a mayor complejidad, el cual se puede realizar con sistemas de aprendizaje supervisados o no supervisados. Este tipo de aprendizaje se ha usado en la clasificación de imágenes, por ejemplo, en el estudio de cálculos renales usando un conjunto de datos con imágenes de cálculos en color e infrarrojo²³.

Aplicaciones en pediatría

El horizonte de uso de la producción masiva de datos en salud es inconmensurable y, por ende, cada año se incrementa exponencialmente el número de artículos científicos que reportan estudios donde se usan Big Data en diferentes disciplinas relacionadas con la salud^{16,24}. A continuación, se describirán algunas experiencias de aplicación en pediatría de datos masivos y las estrategias de procesamiento y análisis descritas anteriormente en el acápite anterior.

En los últimos diez años han proliferado los registros de pacientes gracias a la implementación de FCE, los que inicialmente se almacenaban como Microdatos. Pero la incorporación de toda la información disponible, en grandes volúmenes y de variados formatos, hizo que, en algunos casos, estos datos se convirtieran en macrodatos. Estos micro o macrodatos se convirtieron en la fuente principal que alimentan los llamados “Sistema de apoyo a la Decisión Clínica”, SDC o DSS en inglés²⁵. Los SDC consisten en brindar información filtrada de manera inteligente y presentada para ayudar a los clínicos. La alerta electrónica (e-alerta) es un subproducto de estos sistemas, las cuales son emisiones de recordatorios con recomendaciones, basadas en guías clínicas y en la situación clínica del paciente, con el fin de apoyar el diagnóstico y las decisiones clínicas del personal de salud.

Las e-alertas se han usado para la definición y estandarización de la insuficiencia renal aguda (IRA), sobre la base de los RES y los sistemas integrados de cuidados intensivos²⁶. En pacientes pediátricos hospitalizados, sin enfermedades críticas, las e-alertas han sido usadas en sistemas de vigilancia de nefrotoxina relacionadas

con IRA, lo cual ha permitido reducir la intensidad de IRA en 42% de los casos²⁷, aunque hay autores que advierten que la evidencia científica es aún limitada como para establecer con certeza que las e-alarmas sean de utilidad clínica²⁸.

Otra aplicación de los RES ha sido en la investigación fármaco-epidemiológica, donde se ha podido estudiar el uso de medicamentos, la eficacia y la seguridad, particularmente para población pediátrica. Esto debido a la baja exposición a los medicamentos que suelen tener los niños en comparación con los adultos²⁹.

En el ámbito de cuidados intensivos pediátricos, se han reportado varios estudios donde se han utilizado herramientas de análisis, tales como redes neuronales y aprendizajes automatizados con el fin de darle uso a los volumen y variabilidad de datos que se producen en las Unidades de Cuidado Intensivo Pediátricas. Así es como se han identificado patrones predictores de riesgo de mortalidad, como también información relevante como para el pronóstico, diagnóstico y la terapia de niños en cuidados críticos, entre otros^{30,31}.

En imagenología pediátrica, el aprendizaje automático se ha utilizado para desarrollar modelos de detección y segmentación de imágenes³². Así, por ejemplo, se estudiaron 14.036 radiografías infantiles de mano y sus respectivos informes para elaborar un modelo, basado en análisis de redes neuronales, para hacer estimaciones de edad ósea. En dicho estudio, se observó que el modelo puede estimar la madurez esquelética con una precisión similar a la de un radiólogo experto³³.

A la par con el incremento de registros de datos de pacientes, se está generando una gran cantidad de banco de datos, mayoritariamente de acceso público, los cuales están facilitando la investigación a nivel molecular^{34,35}. De ahí que los datos masivos y las complejas estrategias de análisis antes descritas han permitido buscar patrones de personas, ya sea para clasificarlas con fines diagnósticos o terapéuticos, así como también para la elaboración de fármacos dirigido a estos patrones únicos. Así, por ejemplo, en el ámbito de la genómica, se realizó el estudio de caracterización de los genes conductores de cáncer y mutaciones, donde se identificaron 299 genes conductores con sus respectivos sitios anatómicos y tipos de cánceres involucrados³⁶. En cáncer infantil, y al alero del proyecto Pancer, se han analizado 961 tumores de niños, adolescentes y adultos jóvenes que comprenden 24 tipos moleculares distintos de cáncer. En este estudio, se ha podido determinar que casi el 50% de las neoplasias pediátricas tienen un evento potencialmente farmacológico, lo que plantea una oportunidad para avanzar en el conocimiento a través de ensayos clínicos³⁷.

En esta misma línea, se han usado múltiples fuen-

tes de datos integrados (genómicos, transcriptómicos y clínicos) con el fin de identificar los factores genéticos y moleculares que pueden ser cruciales en el comportamiento del neuroblastoma, el cual es un tumor sólido extracraneal muy común en niños³⁸.

Discusión

Los temas tratados en esta contribución son la punta de iceberg, puesto que son una ínfima parte de los desafíos y oportunidades que involucra la puesta en valor de la producción de datos en el ámbito de la salud. Sin embargo, da luces sobre el avance tecnológico de tratamiento y análisis de datos a gran escala, así como también de sus aplicaciones en la pediatría. Pero es importante tener en cuenta que estos promisorios avances también conllevan situaciones de conflicto o de riesgos potenciales, algunas de las cuales se discutirán a continuación.

En primer término, es importante tener en cuenta que el registro médico surge de la práctica clínica, por tanto, el dato que se registra corresponde a un conjunto de acciones realizadas por profesionales de salud, en pos de atender o dar respuesta a una necesidad de salud de quien la requiera. En consecuencia, los datos que se generan no necesariamente son con fines de investigación científica, por lo cual pueden contener errores, estar incompletos o pueden haberse registrado bajo diferentes reglas de clasificación o unidades de medida, entre otros. Por lo cual, un desafío importante cuando se emprende el análisis de estos datos es la limpieza y validación de éstos²⁰.

Otro aspecto importante para tener en cuenta es que estos datos corresponden a una muestra de sujetos u objetos de estudios que no han sido seleccionados aleatoriamente, por lo cual es susceptible que tengan sesgos de selección³⁹.

La seguridad y privacidad de los registros médicos son algunas de las inquietudes que surgen en relación al uso de datos personales con otros fines que no sean para lo cual fueron registrados. Es así como el consentimiento informado de los propietarios de los datos surge como un elemento importante, pero también se advierte sobre la relevancia de definir qué se entenderá por datos personales⁴⁰. Por esto y otras razones, la Unión Europea (UE) ya ha tomado precauciones, mediante un reglamento de protección de las personas con relación al tratamiento de datos personales y a la libre circulación⁴¹.

En Chile, existen varias leyes y normativas que regulan la investigación clínica. En lo relativo a los datos, se ha dictaminado que la información que surja desde la ficha clínica, los estudios y demás documentos de registro de procedimientos y tratamientos, serán con-

siderados como “dato sensible”, por lo cual deben ser resguardados y no pueden ser utilizados sin el consentimiento informado de los pacientes⁴². Aun así, existe la posibilidad de usar datos recolectados para otros fines y, por ende, sin consentimiento para estudios de investigación, previa autorización a un Comité de Ética.

Por otra parte, en las normas dictadas por el Consejo de Organizaciones Internacionales de las Ciencias Médicas (CIOMS) de 2016, se da la posibilidad de recolectar y almacenar datos con fines de investigación, haciendo uso de un consentimiento informado amplio⁴³. En consecuencia, ya se han emprendido iniciativas como para generar registros de datos médicos que puedan servir para investigaciones futuras, donde en el presente no se tiene claridad de la pregunta de investigación a la que pueden dar respuesta. Esto se está haciendo mediante un consentimiento informado amplio, protocolos y documentos regulatorios⁴⁴. Esto, por cierto, abre la puerta a la utilización de datos masivos

con fines de investigación, pero no resuelve las aprehensiones que existen con relación al resguardo de la privacidad, discriminación, intencionado o no, de los algoritmos, la manipulación de los datos como para obtener respuestas que favorezcan interés corporativas, entre otras^{39,40,45,46}.

En suma, aunque exista la percepción de que estos datos masivos están siendo subutilizados y no se está obteniendo el beneficio que podrían ofrecer⁴⁷, resulta relevante también que se reflexione sobre sus riesgos y limitaciones, para así estimular una investigación científica con una cultura ética del uso de la información personal de sujetos que pueden ser titulares de derechos⁴⁸.

Conflicto de intereses

La autora declara no tener conflicto de intereses.

Referencias

- Cotino L. Big data e inteligencia artificial. Una aproximación a su tratamiento jurídico desde los derechos fundamentales. *Dilemata*. 2017;24:131-50.
- Monleón-Getino A. El impacto del Big-data en la Sociedad de la Información. Significado y utilidad. *Historia y Comunicación Social*. 2015;20(2):427-45.
- Marí-Sáez VM. Comunicar para transformar, transformar para comunicar. *Tecnologías de la información desde una perspectiva de cambio social* Madrid: POPULAR; 2011.
- Mayer-Schönberger, Cukier K. Datafication. In Mayer-Schönberger V, Cukier K. *Big Data. A Revolution That Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt; 2013. p. 49-51.
- Ramos, Arenas A. Chile. Acceso a la Ficha Clínica para Investigación Científica. *Revista chilena de derecho*. 2013; 40(3).
- FALMED. La Ficha Clínica Electrónica. [En línea]; 2019. Disponible en: <http://www.falmed.cl/prevencion-y-educacion/ficha-clinica-electronica.aspx>.
- Moraga F. La digitalización de la pediatría: Una realidad que nos alcanzó. *El estetoscopio*. 2019;102:18-21.
- Barranco R. ¿Qué es Big Data? Todos formamos parte de ese gran crecimiento de datos. *DeveloperWorks*. 2012;2-11.
- Maté C. Big data. Un nuevo paradigma de análisis de datos. *Anales de mecánica y electricidad*. 2014;10-6.
- Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*. 2014;2(3).
- Barea J, Rovira J, Quecedo L, Gol J, Del Llano J. Oportunidades y Retos de los Macrodatos (Big Data) en la toma de decisiones sanitarias Madrid: Fundación Gaspar Casal; 2018.
- Hurwitz JS, Nugent A, Halper F, Kaufman M. *Big data for dummies*: John Wiley & Sons. 2013.
- IIC. Infografía Big Data: las 7 V. [En línea]; 2016 [citado 7-6-2019. Disponible en: <http://www.iic.uam.es/innovacion/big-data-infografia-7-v/>].
- Universidad Complutense de Madrid FdEE. ¿Qué es big data? [En línea]; 2019 [citado 7-6-2019. Disponible en: <https://www.masterbigdataucm.com/que-es-big-data/>].
- IIC. Herramientas Big Data para conseguir mejores resultados. [En línea]; 2017 [citado 29-5-2019. Disponible en: <http://www.iic.uam.es/innovacion/herramientas-big-data-conseguir-mejores-resultados/>].
- Andreu-Perez J, Poon CCY, Merrifield RD, Wong STC, Yang GZ. *Big Data for Health*. *IEEE Journal of Biomedical and Health Informatics*. 2015;19(4).
- Association DS. Code of Conduct. *Data Science Code of Professional Conduct*. [En línea]. [citado 8-6-2019. Disponible en: <http://www.datascienceassn.org/code-of-conduct.html>].
- Provost F, Fawcett T. Data science and its relationship to big data and data-driven decision making. *Mary Ann Liebert, Inc*. 2013;1(1).
- Sánchez-Pinto LN, Luo Y, M. CM. Big Data and Data Science in Critical Care. *Chest*. 2018.
- Andía M, Arrieta C, Sing Long CA. Una guía conceptual para usar y entender Big Data en la investigación clínica. *Revista Médica de Clínica Las Condes*. 2019;30(1):83-94.
- Zambrano J. ¿Aprendizaje supervisado o no supervisado? Conoce sus diferencias dentro del machine learning y la automatización inteligente. [En línea]; 2018 [citado 9-6-2019. Disponible en: <https://medium.com/@juanzambrano/aprendizaje-supervisado-o-no-supervisado-39ccf1fd6e7b>].
- Santa María CR, Santa María V, Ávila L, Otaegui J, Soria M. Clasificación por Enterotipos y Grupos Ortólogos del Microbioma Humano con Métodos No Supervisados. In XVIII Workshop de Investigadores en Ciencias de la Computación (WICC); 2016; Entre Ríos, Argentina. p. 218-22.
- Asensio-Casas V. Clasificación de cálculos renales con técnicas de Deep Learning. 2016.
- Sutherland S, Kaelber D, Downing N, Goel V, Longhurst C. Electronic health record-enabled research in children using the electronic health record for clinical discovery. *Pediatr Clin North Am*. 2016; 63(2): p. 251-268.
- Walsh S, De Jong EE, Van Timmeren JE, et al. Decision support systems in oncology. *JCO clinical cancer informatics*. 2019;3:1-3.
- Kashani KB. Automated acute kidney injury alerts. *Kidney International*. 2018;1-7.

27. Goldstein SL, Kirkendall E, Nguyen H, et al. Electronic health record identification of nephrotoxin exposure and associated acute kidney injury. *Pediatrics*. 2013;132(3):e756-e67.
28. Hoste EAJ, Kashani K, Gibney N, et al. Impact of electronic-alerting of acute kidney injury: workgroup statements from the 15th ADQI Consensus Conference. *Canadian Journal of Kidney Health and Disease*. 2016;3(10):1-9.
29. Christensen ML, Davis RL. Identifying the "blip on the radar screen": leveraging big data in defining drug safety and efficacy in pediatric practice. *The Journal of Clinical Pharmacology*. 2018;58:S86-S93.
30. Williams JB, Ghosh D, Wetzel RC. Applying Machine Learning to Pediatric Critical Care Data. *Pediatric Critical Care Medicine*. 2018;19(7).
31. Aczon M, Ledbetter D, Ho L, et al. Dynamic mortality risk predictions in pediatric critical care using recurrent neural networks. *arXiv*. 2017;1701.06675.
32. Moore MM, Slonimsky E, Long AD, Sze RW, Iyer RS. Machine learning concepts, concerns and opportunities for a pediatric radiologist. *Pediatric Radiology*. 2019;49:509-16.
33. Larson DB, Chen MC, Lungren MP, Halabi SS, Stence NV, Langlotz CP. Performance of a Deep-Learning Neural Network Model in Assessing Skeletal Maturity on Pediatric Hand Radiographs. *Radiology*. 2018; 287(1):313-22.
34. Yang Y, Dong X, Xie B, et al. Databases and web tools for cancer genomics study. *Genomics, proteomics & bioinformatics*. 2015;13(1):46-50.
35. Frigolet ME, Gutiérrez-Aguilar R. Ciencias "ómicas", ¿cómo ayudan a las ciencias de la salud? *Revista Digital Universitaria*. 2017; 18(7):1-7.
36. Bailey MH, Tokheim C, Porta-Pardo E, et al. Comprehensive characterization of cancer driver genes and mutations. *Cell*. 2018;173(2):371-85.
37. Gröbner SN, Worst BC, Weischenfeldt J, et al. The landscape of genomic alterations across childhood cancers. *Nature*. 2018;555(7696):321-56.
38. Salazar BM, Balczewski EA, Ung CY, Zhu S. Neuroblastoma, a Paradigm for Big Data Science in Pediatric Oncology. *International Journal of Molecular Sciences*. 2017;18(37):1-27.
39. Pérez G. Peligros del uso de los big data en la investigación en salud pública y en epidemiología. *Gaceta Sanitaria*. 2016;30(1):66-8.
40. Buenadicha C, Galdon G, Hermosilla MP, Loewe D, Pombo C. La Gestión Ética de los Datos. Por qué importa y cómo hacer un uso justo de los datos en un mundo digital BID, editor; 2019.
41. Unión Europea. Reglamento relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos). 2016.
42. Ley 20584. Regula los derechos y deberes que tienen las personas en relación con acciones vinculadas a su atención en salud. *Diario Oficial de la República de Chile*, Santiago, Chile, 24 de abril de 2012.
43. Consejo de Organizaciones Internacionales de las Ciencias Médicas (CIOMS). Pautas éticas internacionales para la investigación relacionada con la salud con seres humanos: CIOMS-OPS-OMS; 2016.
44. Simian D, Martínez-Jalilíe M. Experiencia en la creación de registros con fines de Investigación Clínica. *Revista Médica Clínica Las Condes*. 2019;30(1):76-82.
45. De Lecuona I. Evaluación de los aspectos metodológicos, éticos, legales y sociales de proyectos de investigación en salud con datos masivos (big data). *Gaceta Sanitaria*. 2018;32(6):576-8.
46. Lefranc Weegan FC, Montes Bracchini LM, Margalli Luna G. E-health Reporter. [En línea].; 2018 [citado 17-6-2019]. Disponible en: <https://ehealthreporter.com/es/noticia/etica-big-data-y-salud/>.
47. Belle A, Thiagarajan R, Soroushmehr SM, Navidi F, BDA, Najarian K. Big data analytics in healthcare. *BioMed research international*. 2015;2015:1-16.
48. Tractenberg RE. Creating a culture of ethics in biomedical big data: Adapting 'guidelines for professional practice' to promote ethical use and research practice. In Mittelstadt BD, Floridi L. *The ethics of biomedical big data.*: Springer; 2016:367-93.