

## Artículo especial

Evaluación de los aspectos metodológicos, éticos, legales y sociales de proyectos de investigación en salud con datos masivos (*big data*)

Itziar de Lecuona

Departamento de Medicina, Facultad de Medicina; Observatorio de Bioética y Derecho de la Universidad de Barcelona;  
Comité de Ética de la Investigación del Hospital Clínic de Barcelona, Barcelona, España

## INFORMACIÓN DEL ARTÍCULO

## Historia del artículo:

Recibido el 18 de enero de 2018

Aceptado el 13 de febrero de 2018

On-line el 31 de mayo de 2018

## Palabras clave:

Investigación en salud

Comités de ética de la investigación

Datos masivos

Intimidad y confidencialidad

Reglamento General de Protección de Datos

## RESUMEN

El actual modelo evaluador de la investigación con seres humanos depende básicamente de los procesos de toma de decisiones de los comités de ética de la investigación. Resulta prioritario que estos comités tomen conciencia de la relevancia del nuevo paradigma digital asentado en la explotación a gran escala de datos personales, incluidos los de salud. El artículo ofrece pautas para una adecuada evaluación de los proyectos que apliquen analítica de datos masivos en salud ante los cambios introducidos por el Reglamento General de Protección de Datos. Los procesos de recolección y explotación de datos constituyen el nicho donde desarrollar la investigación. En este contexto, los protocolos de obtención del consentimiento de los participantes han quedado claramente desfasados debido a que se presuponía no solo que los datos eran anónimos, sino que siempre lo seguirían siendo en el futuro. Por ese motivo, resulta imprescindible que los citados comités asimilen nuevas capacidades y procedan a una relectura de valores como la intimidad y la libertad, actualizando para ello protocolos, metodologías y procedimientos de trabajo. Este cambio de cultura de trabajo dotará de seguridad jurídica al personal implicado en las investigaciones, posibilitará que se garantice la protección de la intimidad de los titulares de los datos, y permitirá orientar la explotación de estos de tal forma que se evite la comercialización de conjuntos de datos personales en la era de la reidentificación, para que la investigación responda a las necesidades sociales y no a intereses espurios u oportunistas disfrazados de investigación.

© 2018 SESPAS. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Review of the methodological, ethical, legal and social issues of research projects in healthcare with big data

## ABSTRACT

The current model for reviewing research with human beings basically depends on decision-making processes within research ethics committees. These committees must be aware of the importance of the new digital paradigm based on the large-scale exploitation of datasets, including personal data on health. This article offers guidelines, with the application of the EU's General Data Protection Regulation, for the appropriate evaluation of projects that are based on the use of big data analytics in healthcare. The processes for gathering and using this data constitute a niche where current research is developed. In this context, the existing protocols for obtaining informed consent from participants are outdated, as they are based not only on the assumption that personal data are anonymized, but that they will continue to be so in the future. As a result, it is essential that research ethics committees take on new capabilities and revisit values such as privacy and freedom, updating protocols, methodologies and working procedures. This change in the work culture will provide legal security to the personnel involved in research, will make it possible to guarantee the protection of the privacy of the subjects of the data, and will permit orienting the exploitation of data to avoid the commodification of personal data in this era of deidentification, so that research meets actual social needs and not spurious or opportunistic interests disguised as research.

© 2018 SESPAS. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## La oportunidad de la investigación con datos masivos en salud

Las iniciativas tanto públicas como privadas priorizan hoy organizarse en torno al dato. El concepto *big data* se refiere al

tratamiento de grandes volúmenes de datos mediante el desarrollo de algoritmos matemáticos, para establecer correlaciones entre datos (incluidos los personales) y determinar patrones de comportamiento que permitan predecir tendencias para mejorar la toma de decisiones<sup>1</sup>. El *big data* forma parte de los planes de investigación y de la política de la Unión Europea para crear un mercado interior competitivo<sup>2</sup>. Esta tecnología contribuye, por ejemplo, a la medicina personalizada, a determinar los efectos adversos de los medicamentos en menos tiempo y a mejorar la eficacia de los

Correo electrónico: [itziardelecuona@ub.edu](mailto:itziardelecuona@ub.edu)

<https://doi.org/10.1016/j.gaceta.2018.02.007>

0213-9111/© 2018 SESPAS. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

sistemas de salud<sup>3</sup>. La explotación masiva de conjuntos de datos almacenados en distintas fuentes para establecer correlaciones y utilizar los resultados con el objetivo de desarrollar nuevos tratamientos e intervenciones en salud parece la consecuencia lógica de la informatización de las historias clínicas.

Ha llegado el momento de aplicar el Reglamento General de Protección de Datos, que trata de dar respuesta a la investigación con *big data*<sup>4</sup>. Sin embargo, los principios de protección de datos de carácter personal, como los de proporcionalidad, necesidad, limitación del propósito, seguridad, minimización y conservación de los datos, no discriminación y consentimiento informado, se tambalean ante la explotación a escala masiva, por la propia naturaleza de los conjuntos de datos, por lo que revelan y por lo que potencialmente pueden desentrañar. Voces desde la salud pública y la epidemiología alertan sobre los sesgos del *big data* y la necesaria revisión humana<sup>5</sup>. Expertos en ciencia de datos e inteligencia artificial reclaman un análisis interdisciplinar y un debate social para establecer los límites de actuación y las pautas éticas para la toma de decisiones automatizada<sup>6</sup>.

### **El reto que la investigación con datos masivos en salud plantea a los comités de ética de la investigación**

Los aspectos metodológicos, éticos y legales de cualquier proyecto de investigación en el que participen personas, o se utilicen sus muestras biológicas o datos personales, deben ser evaluados por los comités de ética de la investigación (CEI), que se articulan como mecanismos de protección de los derechos de las personas<sup>7</sup>. La defensa de la dignidad a través de la salvaguarda de los datos personales en investigación se ha convertido en una tarea prioritaria para estas instancias colegiadas e interdisciplinarias<sup>8</sup>.

En investigación con *big data* en salud conviven distintos agentes con diversos intereses en los mismos conjuntos de datos. Aquellos datos que para el investigador principal podrían ser irrelevantes adquieren pleno sentido para el promotor, que puede ser la banca, las industrias del móvil, biotecnológica, farmacéutica o alimentaria, organizaciones filantrópicas, etc. Por ello, es preciso concretar los objetivos que se persiguen e identificar a todos los actores que van a intervenir en el tratamiento de los datos, incluidos terceros que presten servicios, pues muchas actividades de tratamiento de los datos se van a externalizar y pueden ser objeto de transferencias internacionales<sup>9</sup>. El bien común, que debería guiar la investigación, puede colisionar con intereses particulares y disfrazar de investigación nuevos nichos de mercado basado en la monetización de datos personales de salud; datos que son sensibles y que requieren especial protección. No solo se trata de evitar el mal uso, sino el uso no deseado y desproporcionado, y frenar la tendencia a la acumulación y a la explotación de datos sin claros objetivos. El *big data* genera un entusiasmo tecnológico exacerbado que no permite la necesaria reflexión interdisciplinar para anticipar escenarios, identificar conflictos y proponer marcos de actuación y evaluación.

### **Pautas y requisitos para la evaluación de proyectos de investigación con datos masivos en salud**

A continuación se refieren propuestas para orientar un cambio que permita la adecuada evaluación de este tipo de proyectos e identificar aquellas investigaciones que lo parecen, pero no lo son, y para tomar decisiones correctas sobre qué investigaciones autorizar en *big data*; decisiones ponderadas, que incluyan el impacto social de los proyectos y que no dependan exclusivamente del consentimiento informado de los interesados y de la anonimización como garantía. Por ello, los principios de transparencia y rendición de cuentas, establecidos legalmente pero poco implementados en la práctica, resultan necesarios para

equilibrar la balanza en investigación. Se trata de decisiones que pueden impactar de manera negativa en las generaciones futuras y potenciar la estigmatización y la discriminación encubierta desde las mejores intenciones, de modo que el daño sea irreparable. De los CEI depende decidir qué intereses promover y proteger en una sociedad de mercado que cuenta con las bazas de la investigación y la innovación<sup>10</sup>. Se debe exigir a los CEI la necesaria actualización y formación en *big data* y en protección de datos personales. No todos deben saber de todo, pero sí deben desarrollar un lenguaje común y unos procedimientos de trabajo que permitan identificar ágilmente las verdaderas cuestiones no resueltas por las que apostar, los datos que utilizar y cómo hacerlo. Continuar aplicando antiguas soluciones a problemas nuevos puede perjudicar seriamente la confianza en el sistema investigador y no solo al ciudadano.

Desde la responsabilidad activa que incorpora el Reglamento, conviene comprobar qué tipo de tratamientos de datos plantea el proyecto y sustituir toda referencia a la anonimización en los protocolos, por cuanto estamos ante una situación de posible reidentificación. Hoy sabemos que con el código postal, la fecha de nacimiento y el sexo es posible identificar en un porcentaje elevadísimo a las personas<sup>11</sup>. Los CEI deben abandonar viejos conceptos completamente inútiles que generan falsas seguridades, para exigir las explicaciones oportunas desde la reidentificación como punto de partida. En este sentido, deben solicitar a los investigadores que expliciten las políticas de privacidad por diseño y por defecto, y centrarse en la evaluación del impacto del tratamiento de datos en los derechos y las libertades de las personas participantes e implicadas que los proyectos deben acompañar<sup>12</sup>.

Los CEI deben analizar cómo el proyecto va a dar cumplimiento a los principios de protección de datos y en especial al de minimización del dato. Concretar qué datos se van a utilizar, cómo y dónde se almacenan y por cuánto tiempo, quién es el responsable de coordinar el procesamiento y cuál es el objetivo que se persigue, ponderando riesgos y beneficios; también los planes para mitigar riesgos y evitar brechas de seguridad. Se hacen necesarios mecanismos de observación y seguimiento de los usos de los datos personales por parte de los CEI hacia investigadores, promotores y otros agentes habilitados para operar en el sistema de *i+d+i*, cuyo interés prioritario puede no ser la investigación. Por ejemplo, comités *ad hoc* formados también por legos a quienes reportar la evolución del proyecto<sup>13</sup>. Los CEI deben determinar las comunicaciones con el Delegado de Protección de Datos de la institución, asesor obligatorio en proyectos *big data*, y desarrollar aquellas medidas que permitan el seguimiento del uso y la explotación del dato.

Los CEI tienen que identificar claramente quién es el responsable y el encargado del tratamiento de los datos, y quién tiene acceso a ellos o va a tenerlo durante todo el proceso, en particular las transferencias y la externalización de actividades relacionadas con la explotación de los datos. Es aquí donde los CEI deben innovar para asegurar la trazabilidad del dato y su seguridad, y evitar posibles daños. También en cuanto a la gobernanza es preciso analizar pormenorizadamente el protocolo de investigación desde la perspectiva de la información que se ofrece a los participantes y titulares de los datos a explotar y correlacionar. Los CEI deben comprobar que el proyecto explica con claridad cómo ejercer los derechos de control sobre los datos por parte de los titulares, incluido el derecho al olvido y a la toma de decisiones automatizada. La política de portabilidad del dato que introduce el Reglamento no va a encontrar encajes fáciles en el ámbito de la salud, por lo que los CEI tendrán que idear soluciones creativas que permitan ponderar los derechos e intereses potencialmente en conflicto, para que no se menoscaben la libertad y la intimidad de las personas ni el tratamiento confidencial de sus datos, pero tampoco se limite de manera injustificada la libertad de investigación. Cobra mayor relevancia la transparencia de la información que la obsesión por el consentimiento informado

en todo momento. Es preciso comprobar que se informa a los participantes de los proyectos sobre el riesgo de reidentificación, que siempre debería ser residual.

Los CEI deben incorporar expertos en ciencia de los datos, pues solo estos perfiles van a poder identificar los problemas técnicos y ayudar a concretar si lo que parece un problema ético es en realidad una cuestión técnica. Es necesario que analicen los métodos para la disociación y la seudonimización de los datos, y las fortalezas y debilidades metodológicas, para contribuir a la protección de los datos personales<sup>14</sup>. Son miembros clave para un adecuado análisis del impacto en los derechos de los afectados en cada proyecto.

La investigación que aplique analítica de datos masivos requiere sistemas de supercomputación y nubes que no deberían externalizarse. Los CEI deben exigir que el cruce entre bases de datos de distinta índole, entre las que se encuentren historias clínicas, se efectúe en el perímetro de la institución sanitaria y no fuera. Las autoridades competentes deberían habilitar los correspondientes sistemas informáticos para que esta explotación sea viable y segura. Es necesario que todo personal que trate datos cumpla con el deber de confidencialidad.

Conviene revisar la situación de los miembros de los CEI en cuanto a su dedicación. Entre el altruismo y la profesionalización es preciso desarrollar una fórmula que permita una evaluación pausada y exhaustiva para tomar decisiones racionales y argumentadas<sup>15</sup>. La inversión de tiempo que implica el estudio de la evaluación del impacto en los derechos de los interesados que exige la normativa no es trivial. Las inercias generadas por la prisa y la complejidad favorecen que los proyectos se aprueben tras un análisis superfluo.

La evaluación de la investigación en salud basada en analítica de datos masivos aplicada a la salud obliga a estrechar las colaboraciones entre los directores de sistemas de información de los centros de investigación, los servicios jurídicos, el área de gestión de la investigación y el área de innovación. Esta última es de especial relevancia porque a menudo se someten a consideración de los CEI propuestas de dudosa finalidad, como puede suceder con el desarrollo de aplicaciones de salud que involucren big data. Se trata de clarificar si son proyectos de investigación con una sólida base científica o no. Los CEI deben contribuir al debate social sobre estas cuestiones promoviendo una cultura de la intimidad, y desarrollar su capacidad formativa y sensibilizadora en cuestiones tan relevantes como la protección de los datos de las personas en investigación en salud.

#### Editor responsable del artículo

Carlos Álvarez-Dardet.

#### Contribuciones de autoría

I. de Lecuona es la única autora del manuscrito.

#### Agradecimientos

A mis compañeros del Comité de Ética de la Investigación del Hospital Clínic de Barcelona.

#### Financiación

Ninguna.

#### Conflicto de intereses

Ninguno.

#### Bibliografía

1. Llàcer MR, Casado M, Buisán L, editores. Documento sobre bioética y Big Data en salud: explotación y comercialización de los datos de los usuarios de la sanidad pública. Barcelona: Publicacions i Edicions Universitat de Barcelona; 2015. p. 72.
2. Hacia una economía de los datos próspera. Comisión Europea; 2014. 13 p.
3. Parra CL. Big data en sanidad en España: la oportunidad de una estrategia nacional. Gac Sanit. 2016;30:63-5.
4. 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). R. (UE) 2016/679 (27 abril 2016).
5. Pérez G. Peligros del uso de los big data en la investigación en salud pública y en epidemiología. Gac Sanit. 2016;30:66-8.
6. Barcelona Declaration for the proper development and usage of artificial intelligence in Europe. Barcelona: B-debat Biocat and "La Caixa" Foundation; 2017.
7. Ley de Investigación Biomédica. L. N.º (14/2007). (3 julio 2007).
8. European Data Protection Supervisor. Dictamen 4/2015. Hacia una nueva ética digital. Datos, dignidad y tecnología. Unión Europea. 2015. 24 p.
9. Méndez M. Investigación y mercados emergentes: webs, apps, big data, muestras biológicas e información genética. En: Casado M, coordinador. De la solidaridad al mercado: el cuerpo humano y el comercio biotecnológico. Barcelona: Edicions de la Universitat de Barcelona; 2017. p. 257-66.
10. Mantelero A. Regulating big data. The guidelines of the Council of Europe in the context of the European data protection framework. Computer Law & Security Review. 2017;33:584-602.
11. Sweeney L. Simple demographics often identify people uniquely. Pittsburgh: Carnegie Mellon University, Data Privacy Working Paper 3; 2000. p. 34.
12. Sáiz CA, coordinador. Código de buenas prácticas en protección de datos para proyectos big data. Agencia Española de Protección de Datos - ISMS Forum Spain; 2018. 34 p.
13. Richards M, chairman. The collection, linking and use of data in biomedical research and health care: ethical issues. Nuffield Council on Bioethics. 2015:198.
14. D'Acquisto G, Domingo-Ferrer J, Kikiras P, et al. Privacy by design in big data. An overview of privacy enhancing technologies in the era of big data analytics. European Union Agency for Network and Information Security; 2015: 79.
15. Comité Director de la Bioética. Guía para los miembros de los comités de ética de investigación. Consejo de Europa; 2011. 60 p.