

REC: CardioClinics

www.reccardioclinics.org

Editorial

El escenario y la relevancia de la investigación clínica en la época de los macrodatos



Scenario and relevance of clinical research in the era of big data

Daniel Fernández-Bergés^{a,b,*}, Luis Lozano Mera^{a,b,c} y Francisco Javier Félix-Redondo^{a,b,d}^a Unidad de Investigación, Área de Salud Don Benito-Villanueva, Sistema Extremeño de Salud, Don Benito, Badajoz, España^b Instituto Universitario de Investigación Biomédica de Extremadura, INUBE, Badajoz, España^c Centro Salud Mérida Urbano I, Servicio Extremeño de Salud, Mérida, Badajoz, España^d Centro de Salud Villanueva Norte, Servicio Extremeño de Salud, Villanueva de la Serena, Badajoz, España

«Duda siempre de ti mismo, hasta que los datos no dejen lugar a dudas».

Louis Pasteur.

En el comienzo de los tiempos, el principio de autoridad rigió el conocimiento. No es sino hasta el siglo XVII en el que René Descartes, en su célebre *Discurso del método*, establece un punto de inflexión entre la escolástica y la investigación moderna¹. Este filósofo propuso la duda como un camino en la búsqueda de la verdad y los datos obtenidos como el eje fundamental que nos acerque a tal encuentro.

La herramienta utilizada para describir y analizar datos ha sido la estadística, que surgió hace siglos de las necesidades del estado (de ahí su nombre) de elaborar censos que permitieran una adecuada gestión de recursos y necesidades². Con el tiempo, y la ayuda de las matemáticas, se pasó progresivamente de lo puramente descriptivo a lo inferencial. Este desarrollo matemático también ayudó a que los modelos causales de la epidemiología evolucionaran desde la multicausalidad imperante hasta el último cuarto del siglo XX, cuando irrumpe el modelo probabilístico, que define la moderna estadística con el concepto de riesgo. Si hubiera que nombrar un solo estudio que estimuló conceptos y nuevas herramientas para analizar la relación entre variables y con ello el desarrollo

de la bioestadística, este ha sido sin duda el del *Framingham Heart Study*³. Además de definir el concepto de factor de riesgo, con él aprendimos que era asociación estadística, magnitud de la asociación e independencia. Estas nociones además sirvieron para identificar qué eran «pruebas» sobre las que deberíamos tomar las decisiones clínicas y que con el nombre de «medicina basada en la evidencia», para el mundo anglosajón, se ha convertido en el nuevo paradigma de la propia medicina, que ha propiciado su desarrollo como nunca antes, ayudado por las tecnologías de la información y la comunicación (TIC) que han favorecido que el conocimiento y las posibilidades de colaboración lleguen a todo el mundo.

El diseño de una investigación requiere establecer, en primer lugar, qué hipótesis se quiere verificar o rechazar. Para ello es necesario la redacción de un protocolo, que permita su reproducibilidad, una descripción de los datos a recoger, definir las variables explicativas, el cálculo de un tamaño muestral necesario para contrastar la hipótesis, la aleatorización de la muestra para disminuir los sesgos que generen confusiones, establecer el análisis estadístico que se seguirá y el envío del protocolo al comité de ética de la institución involucrada en la investigación.

Realizada la investigación, hay que efectuar un análisis profundo de los resultados con sus pruebas específicas para establecer significación y asociación, una discusión exhaustiva de los hallazgos obtenidos para finalizar con unas

* Autor para correspondencia.

Correo electrónico: polonibo@gmail.com (D. Fernández-Bergés).
<https://doi.org/10.1016/j.rccl.2019.10.002>

2605-1532/© 2020 Sociedad Española de Cardiología. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

conclusiones que deben ajustarse al objetivo establecido al comienzo de la misma.

Este escenario está sufriendo un cambio en el cual debemos participar sin demora. El paso del dato al macrodato o *big data*⁴, la integración de la inteligencia artificial⁵, el replanteamiento del análisis o la forma de enunciar nuestros resultados con el empleo de la estadística⁶, requiere una participación y una amplitud de criterio como nunca antes se nos había planteado. Todo ello sin mencionar el impacto que tendrá la introducción del análisis de macrodatos con ordenadores cuánticos, el paso del bit al cúbit⁷.

Del dato al macrodato

Dato (del latín *datum*) es «la información que se da de algo concreto que permite su conocimiento exacto o sirve para deducir las consecuencias derivadas de un hecho»⁸.

Existen numerosas definiciones de macrodato: Barea Mendoza et al.⁹ escogen la propuesta de la International Data Corporation, que define macrodato como el «conjunto de nuevas tecnologías y arquitecturas diseñadas para la obtención de valor de grandes volúmenes y variedad de datos de una forma rápida, facilitando su captura, procesamiento y análisis». Estos autores hacen notar los beneficios de esta definición dado que evidencia la necesidad de involucrar tecnologías y datos para la obtención de valor, lo que caracteriza a dichos datos por su volumen, variedad y velocidad de generación a los que se han agregado últimamente veracidad y valor (5 V). La adición de estas 2 últimas «V» no es baladí, dado que explicita la necesidad de que esa información sea verdadera y que aporte valor de cara a la ciudadanía⁴.

La utilidad de los macrodatos dependerá del uso que se haga de los mismos cuando se trate de un detalle descriptivo, predictivo o prescriptivo. Cada intención llevará un tratamiento específico por parte de personal cualificado. Los análisis y sus resultados permitirán continuar con la sana costumbre de establecer hipótesis de trabajo que se deberán validar en estudios prospectivos o, en su defecto, el análisis de datos de la vida real.

Merece una consideración específica el análisis de macrodatos de imágenes y el aumento de la precisión diagnóstica con el empleo de técnicas de inteligencia artificial; el futuro que se abre a esta modalidad de exploración y explotación de datos de imagen es sencillamente fantástica.

El uso de los datos generados por la ciudadanía con respecto al tratamiento de datos personales, su libre circulación y la protección de las personas físicas, ya ha recibido respuesta por parte de la Unión Europea con la publicación del reglamento general de protección de datos¹⁰.

Inteligencia artificial: del machine learning al deep learning

La identificación de patrones recurrentes en estos conjuntos de datos es la base que permite el aprendizaje en modelos de inteligencia artificial. Los algoritmos de *machine learning* o aprendizaje automático aprenden de los datos con los que se los alimentan, proceso que se puede, o no, supervisar por

el hombre¹¹. Son capaces de hacer predicciones y orientar la toma de decisiones. Cuando se habla de *deep learning* o aprendizaje profundo se hace referencia a algoritmos de altísima complejidad, que imitan la red neuronal de un ser humano. Este salto solo ha sido posible gracias a ordenadores de nueva generación más potentes. Y tan solo estamos en la antesala de un cambio realmente revolucionario: la generalización de los ordenadores cuánticos. Los bits se sustituirán por cúbits y la velocidad de cálculo se disparará convirtiendo en solucionables problemas que hoy son irresolubles. Aunque empresas como Google ya son poseedoras de esta tecnología, IBM ha anunciado que desde octubre del 2019 pone a disposición de investigadores y empresas su ordenador cuántico más potente hasta la fecha: Big Blue de 53 cúbits¹².

De lo multidisciplinar a lo interdisciplinar

Durante años se ha insistido en la necesidad de constituir grupos integrados por diferentes especialistas, a fin de ampliar el ángulo de la mirada, a fin de no segmentar sino integrar el conocimiento y compartir espacios de investigación e incertidumbre. En los últimos años, y con la irrupción de las nuevas TIC, se ha hecho evidente la necesidad de cruzar lo multidisciplinar para pasar a lo interdisciplinar. La integración de ingenieros, informáticos y también educadores que transfieran el conocimiento a la población será, en nuestro criterio, un requisito a la hora de validar grupos de investigación por los organismos oficiales pertinentes. La propia sociedad está abriendo camino en Europa con la corriente de «ciencia ciudadana» buscando formas de integrarse a la investigación en alguna de sus facetas y no solamente como objeto de estudio¹³.

La evolución de la estadística como herramienta

En este contexto es necesario plantearse cómo utilizaremos la estadística o qué pruebas serán necesarias para analizar los macrodatos cuando desde los ámbitos académicos se nos recuerda la distorsión que introduce en la investigación aquello de «estadísticamente significativo»^{14,15} o el cuestionamiento del valor del paradigma de la prueba de significación de hipótesis nulas⁶.

El desafío es apasionante y las propuestas múltiples.

El ensayo clínico ha respondido en gran medida a los intereses de los financiadores, que no es necesariamente el interés de la sociedad. La posesión de macrodatos por parte de los sectores públicos de salud abre un campo de investigación predictivo que se deberá explorar con rigor y tiempo, pero sin duda cambiará la forma de interpretar los resultados en la vida real cuando los criterios de exclusión no dejen fuera del análisis importantes porcentajes de la población que luego es la que teóricamente más se beneficiaría de aquellos resultados. ¿Se deberá reformular el conocido discurso en cuanto a la validez interna y la validez externa?

La investigación en términos generales ha carecido de valor social y esto se traduce en una apuesta todavía deficitaria por parte de los sectores públicos. Hoy más que nunca

es necesario que los gestores apoyen la investigación como un aspecto esencial de la vida sanitaria. La mirada asistencial del sistema de salud es su fin último y no se debe abandonar, pero la transformación de un servicio de asistencia a uno de conocimiento es esencial para dar una atención más personalizada y eficiente.

La llegada del concepto de utilización del macrodato es de vital importancia para la sanidad y sus usuarios. No vislumbrar la importancia que tiene para la población un análisis no sesgado de sus propios datos es sencillamente un suicidio.

Los macrodatos y la inteligencia artificial están entre nosotros, se quedarán y aumentarán el conocimiento con herramientas de análisis conocidas y otras que se irán desarrollando. Se debe integrar lo antes posible, a nivel público, equipos interdisciplinarios que trabajen generando información y conocimiento en beneficio de la población.

Conflicto de intereses

Ninguno que declarar.

BIBLIOGRAFÍA

1. Descartes R. *Discurso del método. Biblioteca Universal del Círculo de Lectores: Discurso del método. La dióptrica. Los meteoros. La geometría.* Barcelona: Círculo de Lectores.; 1996:73-139.
2. Arribas Macho JM, Gómez García B, Almazán Llorente A, Villarejo Ramírez C. *Historia de la Estadística.* Madrid: UNED; 2007:8-34.
3. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: A historical perspective. *Lancet.* 2014;p.383:999-1008.
4. Pérez Huerta P, Edo Solsona MD, Poveda Andrés JL. *Cuadernos de gestión sanitaria.* N° 4. Datos de vida real y macrodatos: Su incorporación a la evaluación. Madrid: Fundación Gaspar Casal;2018.p.13-36.
5. Russell S, Norvig P. *Artificial intelligence. A modern approach.* 3rd ed. New Jersey: Pearson; 2010:1044-1052.
6. Mcshane BB, Gal C, Gelman A, Robert C, Tackett JL. Abandon statistical significance. *Am Stat.* 2019;73:235-245.
7. Katz J. How quantum mechanics can change computing. En: Phus.org. [consultado 16 Sept 2019]. Disponible en: <https://phys.org/news/2017-08-quantum-mechanics.html>.
8. Real Academia de la Lengua. *Diccionario de la Real Academia Española.* [consultado 14 Sept 2019]. Disponible en: <https://dle.rae.es/?id=Bskzsq5BsnXzV1>.
9. Barea Mendoza J, Rovira Fornis J, Quecedo Gutiérrez L, Gol Montserrat J, del Llano Señarís J. *Oportunidades y retos de los macrodatos (Big Data) en la toma de decisiones sanitarias.* Madrid: Fundación Gaspar Casal; 2019. p.25-34.
10. Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo (27 de abril de 2016) relativo a la Protección de las Personas Físicas en lo que respecta al Tratamiento de Datos Personales. [consultado 15 Sep 2019]. Disponible en: <https://www.boe.es/doue/2016/119/L00001-00088.pdf>.
11. Dorado-Díaz PI, Sampedro-Gómez J, Vicente-Palacios V, Sánchez PL. Applications of artificial intelligence in cardiology. The future is already here. *Rev Esp Cardiol.* 2019, <http://dx.doi.org/10.1016/j.recesp.2019.05.016>.
12. Macías E. IBM anuncia el ordenador cuántico más potente disponible para uso público, En: Computerworld.es. [consultado 15 Sept 2019]. Disponible en: <https://www.computerworld.es/tecnologia/ibm-anuncia-el-ordenador-cuantico-mas-potente-disponible-para-uso-publico>.
13. Citizen Science Association. [consultado 24 Ago 2019]. Disponible en: <https://www.citizenscience.org/>.
14. Ioannidis JPA. Publishing research with p-values: Prescribe more stringent statistical significance or proscribe statistical significance? *Eur Heart J.* 2019;40:2553-2554.
15. Ioannidis JPA. Options for publishing research without any p-values. *Eur Heart J.* 2019;40:2555-2556.