EDITORIAL

Big data: Value creation in clinical nutrition[☆]

Big data, creación de valor en nutrición clínica

Julia Alvarez Hernández

Sección de Endocrinología y Nutrición, Hospital Universitario Príncipe de Asturias, Universidad de Alcala, Alcalá de Henares, Madrid, Spain

Artificial intelligence (AI) has made an unstoppable entry into our lives. Its origins date back to the mathematical logic and computational works of Alan Turing[1] and McCulloch and Pitts, published over 80 years ago.[2] Artificial intelligence is a term coined by Minsky and McCarthy in 1956 at the Dartmouth conference.[3] It is defined as the scientific field of computing that focuses on the creation of programs and mechanisms that can display behavior considered intelligent. In other words, ''machines that think like human beings.''

But in order to develop the concept, AI draws on a large amount of data (big data [BD]), and uses it to develop algorithms and generate its own logic. Ultimately, AI uses data to gain information from and interact with the environment accordingly. Thus, AI and BD are closely linked terms that offer opportunities for improvement in all disciplines if we are able to take advantage of the information systems available to us.

Big data is the term used in the information and communication technology sector in reference to a body of data which, because of its volume and variety, and the speed at which the data need to be processed, exceeds the capabilities of standard computer systems. Regarding Spanish, it has been proposed that BD be translated as ''macrodata'', this being presented as a valid alternative, since like the term ''big'', it refers to the concept of largeness. As a solution it is to the point, and unlike the term ''megadata'', it produces no confusion with ''mega'' – which is also often used in the same scenarios.[4]

Big data implicates all aspects of human life, including biology and medicine.[5] The advances in recent decades in the world of ''omics'' (genomics, proteomics, metabolomics) and other types of technologies, as well as the implementation of electronic health records (EHRs), have led to an exponential growth in data volume, contributing to the reality of BD in the healthcare setting. Today it constitutes a genuine augmentation of knowledge allowing us to innovate and improve the quality and efficiency of care. According to Margolis, ''big data is not only a new reality for biomedical scientists, but also an imperative that must be understood and used effectively in the search for new knowledge''.[6]

Some authors suggest that the term BD does not have an adequate definition in the Medline (MeSH) thesaurus. An in-depth review reveals that the term which best defines BD in healthcare publications is ''volume''.[7]

It is clear that the field of medicine is dominated by a vast volume of healthcare data. Such data include personal medical records, medical images, genetic data, genomic sequences of population-based data, clinical research data (observational studies, clinical trials, etc.) and much more. More recently, this exponential growth has also been fueled by three-dimensional (3D) images, as well as readings from biometric sensors or wearable devices, i.e., devices incorporated into clothing or used bodily as implants or accessories which may act as an extension of the user's body or mind, some them being widely used in endocrinology and clinical nutrition.

While the concept of data volume in medicine, as in other disciplines, is relevant to BD, the concept of BD is becoming

increasingly associated with the five Vs: volume, velocity, variety, veracity and value. We could say that we are talking about a data set or combinations of large bodies of data (volume), of a diverse and complex nature (variety), which grow rapidly and need to be processed (velocity) and to be real, authentic and of quality (veracity), and which when appropriately analyzed afford value.

Considering the above, we cannot stay on the conceptual surface of the term, and the great challenge, in my opinion, lies in the continuous development and real, extensive and collaborative implementation of data analyzing systems that are more powerful than the traditional methods. We must focus on developing platforms for the more effective capture, storage and handling of these large volumes of data, to allow us to add value to our healthcare activity.

A good example of this is the Savana Manager tool, used by Ballesteros et al. in their study.[8] This innovative system, using EHRead technology, is able to automatically analyze and extract the relevant clinical information contained in the free text of EHRs using natural language and BD processing techniques, and to transform it into ordered information for research purposes.[9]

In their study, Ballesteros et al. tested one of the BD utilities in clinical research and healthcare management. The analysis of a large body of EHR data has shown that the underdiagnosis of disease-related malnutrition (DRM) remains a reality. This circumstance is widely highlighted by different initiatives in the fight against malnutrition, and is the constant concern of endocrinologists dedicated to clinical nutrition.[10]

The BD of the study in question assessed more than 180,000 hospitalization records and, with less effort than that required in classical prevalence studies, allowed for an assessment of the characteristics of the study population. The patients identified as malnourished were mainly individuals with heart failure (35%), respiratory infection (23%), urinary infection (20%) and chronic kidney disease (15%). The study also established that these patients were older (75 vs. 59 years), with greater mortality (7.08% vs. 2.98%) and longer stay (8 vs. 5 days; $p < 0.0001$) as compared to patients with no diagnosis of malnutrition. The study established that 2.47% of the episodes included the diagnosis of DRM, a figure far from the over 23% reported in the PREDyCES study.[11]

The true value of this exercise in BD analysis lies in the fact that, with relatively less effort, the authors have relevant information for managing the approach to DRM at their center. With data showing under-diagnosis, they are better able to adopt measures of improvement in care units. This facilitates the early detection of malnourished patients or individuals at risk of malnutrition who can benefit from a specific intervention to both improve the quality of care and to lower the costs.

We have sufficient evidence regarding how advances in BD and AI, along with human intelligence, lead to the practice of high-performance medicine. The current applications of AI range from embryo selection in in vitro fertilization processes or medical monitoring using Alexa-type verbal language devices, to mental health controls, the monitoring of parameters of interest (blood pressure, heart rate, electrocardiogram, blood glucose, etc.) or therapeutic adherence. Other applications refer to paramedic interventions in the cardiological or neurological setting (myocardial infarctions,

stroke, etc.), aids for reading and interpreting radiological images, the prevention of blindness (retinography readings), the identification of mutations causing cancer, the promotion of patient safety, and even the prediction of mortality in the hospital setting.[12] It is to be expected that these advances will allow for further improvements in disease prevention, diagnosis and treatment, as well as better and more efficient healthcare management, with a positive impact upon quality and efficiency.

However, in order to reach these goals capable of transporting the healthcare system to a new era, we need to work together to implicate those involved in the development of the digitalization process (public administration, private companies, hospitals, physicians, research centers, universities, etc.). One of the most important challenges in this development process in all countries is to integrate the technology with privacy and confidentiality policies, infrastructures, and a data sharing culture.[13] Although we are aware that digital transformation in health in Spain has already begun, authoritative voices have proposed for some years the development of a National Digital Health Strategy, taking into account all of these factors, incorporating a clear framework of cooperation, prioritizing the implementation of shared value use cases, and making it feasible to measure the resulting impact.[14]

References

1. Turing AM. Computing machinery and intelligence. Mind. 1950;59:433–60.
2. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys. 1943;5:115–33.
3. Conferencia de Darthmounth; 1956. Available from: https://darthmouthconference.wordpress.com/
4. Barea Mendoza J, Rovira Forns 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. Pinto, Madrid: Fundación Gaspar Casal ed.; 2018.
5. Zhang Z. Big data and clinical research: focusing on the area of critical care medicine in mainland China. Quant Imaging Med Surg. 2014;4:426–9.
6. Margolis R, Derr L, Dunn M, Huerta M, Larkin J, Sheehan J, et al. The National Institutes of Health's Big Data to Knowledge (BD2K) initiative: capitalizing on biomedical big data. J Am Med Inform Assoc. 2014;21:957–8.
7. Baro E, Degoul S, Beuscart R, Chazard E. Toward a literature-driven definition of big data in healthcare. BioMed Res Int. 2015;9, http://dx.doi.org/10.1155/2015/639021 [article ID 639021].
8. Ballesteros MD, Pintor de la Maza B, Barajas Galindo D, Cano Rodríguez I. Persiguiendo la desnutrición relacionada con la enfermedad mediante herramientas de big data. End Nutr y Diab. 2020.
9. Espinosa-Anke L, Tello J, Pardo A, Medrano I, Ureña A, Salcedo I, et al. Savana: a global information extraction and terminology expansion framework in the medical domain. Proc Leng Natural. 2016;57:23–30. Available from: http://hdl.handle.net/10230/33531
10. Fight against malnutrition. Available from: https://www.espen.org/fight-against-malnutrition
11. Álvarez-Hernández J, Planas Vila M, León-Sanz M, García de Lorenzo A, Celaya Pérez S, García Lorda P, et al. Prevalence and costs of malnutrition in hospitalized

patients; the PREDyCES Study. Nutr Hosp. 2012;27:1049–59, http://dx.doi.org/10.3305/nh.2015.32.4.9700.Predyces.

12. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25:44–56.

13. Menasalvas E, Gonzálo C, Rodriguez-González A. Big data en salud: retos y Oportunidades. Economía Industrial. Available from: https://www.mincotur.gob.es/Publicaciones/Publicacionesperiodicas/EconomiaIndustrial/RevistaEconomia Industrial/405/MENASALVAS,%20GONZALO%20Y%20RODR%C3% 8DGUEZ.pdf

14. Parra Calderón CL. Big data en sanidad en España: la oportunidad de una estrategia nacional. Gac Sanit. 2016;30:63–5.