

Big Data and Precision Medicine: an Overview into the Future

Walter Masson 

Chief of Cardiovascular Prevention, Hospital Italiano.
Ciudad Autónoma de Buenos Aires, Argentina.

Acta Gastroenterol Latinoam 2023;53(3):207-210

Received: 21/08/2023 / Accepted: 21/09/2023 / Published online: 30/09/2023 / <https://doi.org/10.52787/agl.v53i3.345>

Unlike the "one size fits all" approach, precision medicine uses a person's medical (including genetic), behavioral and environmental information to further individualize their care. This could lead to better prediction of disease in an at-risk individual and more effective diagnosis and treatment when the disease is present. Big data allows for much greater precision and tailoring than ever before, linking disparate data sets to reveal previously unknown correlations and causal pathways. But there are

also ethical issues related to balancing interests, the feasibility of anonymity, the impact on family and groups, and genetic discrimination.

Behavioral and lifestyle factors (such as diet, daily activities, even use of social networks) interact with genetic factors to cause disease. This means that lifestyle data also contain information on the key risk determinants for many of the common chronic diseases that afflict modern societies. These behavioral factors can interact with biological factors to cause disease. Access to data on these behavioral factors not only allows for a better understanding of biological effects, but also for the identification of behavioral changes that can mitigate the effects of biological variants on disease susceptibility.

The generic term "big data" or "Big Data" refers to the set of structured and unstructured data that is so large and complex that it cannot be collected, processed, stored and analysed using traditional methods.¹ In other words, this large amount of information can only be processed and analysed using new technological and digital tools.

Some of the main characteristics of the "Big Data" concept are volume (enormous amount of data), speed (continuous and exponential growth of information), variety (data from different information sources, high

Correspondence: Walter Masson
Email: walter.masson@hospitalitaliano.org.ar

complexity), veracity (reliable sources of information) and added value (generation of new development opportunities).²⁻³ Advances in technology and information technology have made it possible to process this big data, searching for trends or patterns in the information analysed, with the aim of answering questions from different disciplines.

The advanced technologies developed for Big Data have promoted its application in many fields, such as the fight against crime and issues related to insecurity, business management, finance, the Global Positioning System (GPS), commerce, tourism, meteorology, biology or the environment, among others.⁴ On the other hand, the applications of Big Data in health sciences in general, and in medicine in particular, are diverse and cover practically all areas: genomics, epidemiology, clinical trials, diagnostic and prognostic algorithms, telemedicine, administrative management, etc. In other words, the applicability is infinite and includes issues related to research, teaching and healthcare tasks. The sources of medical data on which we could apply Big Data analysis techniques are very diverse, ranging from personal and socio-economic data, clinical characteristics, administrative information to data related to complementary laboratory or imaging studies. The information obtained after analysing the data using Big Data techniques makes it possible to optimize medical practice, making it more "personalized" or "patient-centered".⁵ In this sense, big data analytics makes it possible to "model" the patient and offer each person what best suits his or her personal characteristics. In addition, Big Data applications could make medicine more participatory.⁶ With the information generated, the patient can actively participate in decisions related to his or her health, facilitating, among other things, better adherence to treatment.

Two other scenarios in which Big Data analytics techniques are applied are preventive medicine and predictive medicine, which are much more tangible for the general practitioner. In the first case, we can cite as examples the possibility of carrying out more effective surveillance of epidemic outbreaks, health emergencies, pharmacovigilance⁷⁻⁸ or the development of better public health campaigns, from vaccination to the prevention of obesity or suicidal behavior.⁹⁻¹¹ In the second case, starting from the analysis based on Big Data techniques, more accurate predictive models can be created, for example to estimate the risk of presenting a cardiovascular emergency, hospitalization, or having a good or bad evolution after an oncological diagnosis.¹²⁻¹⁴

Finally, some concerns related to Big Data techniques should be considered. In this context, it is important to consider certain methodological issues when interpreting the information, such as registration and association biases. Some ethical issues, such as patient privacy, should also be considered.¹⁵

In the example shown below, we will evaluate the evidence from the information provided by Big Data.

Example

A very important issue in routine clinical practice is the relationship between gastroesophageal reflux disease (GERD) and so-called extraesophageal symptoms. Much has been said about the relationship between GERD and exacerbations of chronic obstructive pulmonary disease (COPD), although it remains controversial. Previous studies have shown that patients with coexisting GERD and COPD have worse quality of life and more severe shortness of breath than those with COPD but no GERD. It has also been suggested that GERD is a risk factor for acute exacerbations. More controversial is the use of proton pump inhibitors (PPIs) to prevent acute exacerbations in patients with COPD.

The following example analyses the effects of treating GERD with PPIs on the risk of acute exacerbations and pneumonia in COPD patients based on the following scientific study: *Respir Res.* 2023 Mar 11;24(1):75. This is a study using a large population-based database with information on medical diagnoses and treatments.

The aim of the study was to assess the risk of both exacerbation and pneumonia following PPI treatment for GERD in patients with COPD.

This study used a reimbursement database from the Republic of Korea. Patients aged ≥ 40 years with a primary diagnosis of COPD who received PPI treatment for GERD for at least 14 consecutive days between January 2013 and December 2018 were included.

A self-controlled case-series analysis was performed to estimate the risk of moderate and severe exacerbation and pneumonia, with each study subject serving as his or her own control, to minimize the influence of individual risk factors for exacerbation or pneumonia.

Results

A total of 104,439 COPD were treated with PPIs for GERD. The risk of moderate exacerbation was signifi-

cantly lower during PPI treatment than at baseline. The risk of severe exacerbation increased during PPI treatment but decreased significantly in the post-treatment period. The risk of pneumonia was not significantly increased during PPI treatment.

Conclusions

The risk of exacerbation was significantly reduced after treatment with PPIs compared to the period when they were not treated. Severe exacerbations may increase due to uncontrolled GERD, but then decrease after PPI treatment. There was no evidence of an increased risk of pneumonia.

Strengths and limitations of the study

The strength of this work is that it uses a series of patients with a large sample size, which makes it possible to describe the population and the size of the problem (representativeness).

Large population databases allow the construction of hypotheses based on their findings and the impact of multiple factors simultaneously. The construction of explanatory or predictive models from these data is another strength of these large databases.

At the same time, it is a self-controlled series of cases, which favors the minimization of the confounder problem.

Weaknesses include biases related to the selected population, data quality, and external validity, as the sample was taken from a Korean population and may not be applicable to other populations. Some issues should also be considered in relation to Big Data, such as heterogeneity, data processing, associated data techniques that are different from traditional ones, high expert cost, cost, and handling of data privacy manipulation.

In short, Big Data analysis techniques are here to stay. Their proper application will lead to precision medicine, which focuses on disease prevention and treatment, while taking into account each person's individual genetic variability, environment and lifestyle.

Intellectual Property. *The author declares that the data and table presented in the manuscript are original and were carried out at his belonging institution.*

Funding. *The author declares that there were no external sources of funding.*

Conflict of interest. *The author declares that he has no conflicts of interest related to this article.*

Copyright



© 2023 Acta Gastroenterológica latinoamericana. This is an open-access article released under the terms of the Creative Commons Attribution (CC BY-NC-SA 4.0) license, which allows non-commercial use, distribution, and reproduction, provided the original author and source are acknowledged.

Cite this article as: Masson W. Big Data and Precision Medicine: an Overview into the Future. Acta Gastroenterol Latinoam. 2023;53(3):207-210. <https://doi.org/10.52787/agl.v53i3.345>

References

- Gomes MAS, Kovalski JL, Pagani RN, da Silva VL, Pasquini TCS. Transforming healthcare with big data analytics: technologies, techniques and prospects. J Med Eng Technol. 2023 Jan;47(1):1-11. <https://doi.org/10.1080/03091902.2022.2096133>
- Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2(1). <https://doi.org/10.1186/2047-2501-2-3>
- Pramanik PK, Pal S, Mukhopadhyay M. Healthcare big data: a comprehensive overview. In research anthology on big data analytics, architectures, and applications. IGI Glob 2022;19-47. <https://doi.org/10.4018/978-1-6684-3662-2.ch006>
- Hong L, Luo M, Wang R, Lu P, Lu W, Lu L. Big Data in Health Care: Applications and Challenges. Data and Information Management. 2018;2(3):175-197. <https://www.sciencedirect.com/science/article/pii/S2543925122000791?via%3Dihub>
- Schulte T, Bohnet-Joschko S. How can Big Data Analytics Support People-Centred and Integrated Health Services: A Scoping Review. Int J Integr Care. 2022 Jun 16;22(2):23. DOI: 10.5334/ijic.5543
- Konstantinidis M, Lalla EA. Clinical anisotropy: A case for shared decision making in the age of too much data and patient disintegration. J Eval Clin Pract. 2020 Apr;26(2):604-609. DOI: 10.1111/jep.13312
- Bouzillé G, Poirier C, Campillo-Gimenez B, Aubert M-L, Chabot M, Chazard E. Leveraging hospital big data to monitor flu epidemics. Comput Methods Programs Biomed. 2018 Feb;154:153-160. DOI: 10.1016/j.cmpb.2017.11.012
- Trifiro G, Sultana J, Bate A. From big data to smart data for pharmacovigilance: the role of healthcare databases and other emerging sources. Drug Saf. 2018;41:143-149.
- Mills G. Big data drive efficient rabies vaccination. Vet Rec. 2021 Feb;188(3):88-89. DOI: 10.1002/vetr.150.

10. Detecting Suicide and Self-Harm Discussions Among Opioid Substance Users on Instagram Using Machine Learning. *Front Psychiatry*. 2021 May 31;12:551296. DOI: [10.3389/fpsyt.2021.551296](https://doi.org/10.3389/fpsyt.2021.551296)
11. Tu B, Patel R, Pitalua M, Khan H, Gittner LS. Building effective intervention models utilizing big data to prevent the obesity epidemic. *Obes Res Clin Pract*. 2023 Mar-Apr;17(2):108-115. DOI: [10.1016/j.orcp.2023.02.005](https://doi.org/10.1016/j.orcp.2023.02.005)
12. Scali ST, Stone DH. The role of big data, risk prediction, simulation, and centralization for emergency vascular problems: Lessons learned and future directions. *Semin Vasc Surg*. 2023 Jun;36(2):380-391. DOI: [10.1053/j.semvasc Surg.2023.03.003](https://doi.org/10.1053/j.semvasc Surg.2023.03.003)
13. Schulte T, Wurz T, Groene O, Bohnet-Joschko S. Big Data Analytics to Reduce Preventable Hospitalizations-Using Real-World Data to Predict Ambulatory Care-Sensitive Conditions. *Int J Environ Res Public Health*. 2023 Mar 7;20(6):4693. DOI: [10.3390/ijerph20064693](https://doi.org/10.3390/ijerph20064693)
14. Choi JW, Kang S, Lee J, Choi Y, Kim HC, Chung JW. Prognostication and risk factor stratification for survival of patients with hepatocellular carcinoma: a nationwide big data analysis. *Sci Rep*. 2023 Jun 27;13(1):10388. DOI: [10.1038/s41598-023-37277-9](https://doi.org/10.1038/s41598-023-37277-9)
15. Kayaalp M. Patient Privacy in the Era of Big Data. *Balkan Med J*. 2018 Jan 20;35(1):8-17. DOI: [10.4274/balkanmedj.2017.0966](https://doi.org/10.4274/balkanmedj.2017.0966)