



Big Data, reproducibility, and Open Science: towards a new scientific culture

Lázaro Pablo Linares-Cánovas¹ 

¹University of Medical Sciences of Pinar del Río. "Luis Augusto Turcios Lima" University Teaching Polyclinic. Pinar del Río, Cuba.

Dear readers:

Biomedical science is currently undergoing a profound transformation, driven by the convergence of big data, real-world evidence, and open science principles. It is undeniable that enormous volumes of information are generated every day; while this opens up unprecedented opportunities to better understand different processes and complex diseases, and to make more precise health decisions, it also necessitates a serious examination of data quality, governance, and responsible use.⁽¹⁾ More than just a technological innovation, a new scientific culture is taking shape, in which integrating data platforms and artificial intelligence has become an international priority.⁽²⁾

Real-world evidence (RWE) has become one of the cornerstones of this transformation. Unlike traditional clinical trials, which are usually conducted under highly controlled conditions, RWE allows us to observe how interventions work in everyday practice, in diverse patients, and in real healthcare settings. This is why it is gaining importance in drug regulation, health technology assessment, and post-marketing surveillance.⁽³⁾ This is especially true in increasingly demanding healthcare systems, where having useful real-world evidence is no longer a complement, but a strategic necessity.

At the same time, massive databases have extraordinarily expanded the analytical possibilities of modern medicine. Electronic health records, population registries, as well as biobanks and omics platforms, among others, allow for the generation of continuous flows of information, guaranteeing the study of diseases on previously unimaginable scales and enabling scientific progress in this area to occur at an almost immeasurable pace. When this is combined with the use of advanced analytical methods, new perspectives emerge for the identification of risk factors, prediction of clinical outcomes, and the consequent design of more effective interventions.⁽⁴⁾ However, the enthusiasm for data volume should not overshadow a fundamental principle: methodological quality remains the primary determinant of scientific validity. Therefore, large datasets do not replace rigor; on the contrary, they demand higher standards of transparency, traceability, and bias control.

Citación: Linares Cánovas LP. Big Data, reproducibilidad y ciencia abierta: hacia una nueva cultura científica. Rev Ciencias Médicas [Internet]. 2026 [citado: fecha de acceso]; 30(2026): e7272. Disponible en: <http://revcmpinar.sld.cu/index.php/publicaciones/article/view/7272>

Recibido: 02 de octubre de 2026
Aceptado: 02 de octubre de 2026
Publicado: 03 de octubre de 2026

Over the past decade, one of the major concerns arising from the widespread expansion of big data has been the crisis of scientific reproducibility. Multiple disciplines have documented difficulties in replicating previously published results, a phenomenon that has generated debates about the reliability of scientific literature. In response to this reality, high-impact journals and funding agencies are increasingly promoting the pre-registration of protocols, the availability of analytical codes, the publication of underlying data, and the adoption of stricter reporting standards.⁽⁵⁾ Reproducibility should not be understood solely as a technical requirement, but as an expression of scientific integrity and an ethical commitment to the society that funds and uses research results. Furthermore, strengthening replication studies is an emerging priority on the international biomedical agenda.

In this new scenario, data sharing is no longer a secondary option but a key condition for accelerating knowledge. Therefore, open science proposes that results, databases, and code can be found and reused responsibly. This approach facilitates collaboration and improves the efficiency of scientific investment, something especially valuable in the face of global health challenges. It also necessitates the development of governance frameworks that protect privacy and promote common standards, enabling the reliable exchange of information.

From the Cuban perspective, these challenges take on particular nuances. Cuba possesses recognized strengths in epidemiological surveillance, primary healthcare, and the integration of healthcare networks, elements that generate a valuable source of data for public health research. However, the limited availability of advanced technological infrastructure, economic constraints (resulting from the economic embargo imposed on the island), and the need to strengthen national ecosystems for managing scientific data continue to pose significant challenges. Given this situation, it is essential to promote institutional policies that foster progressive digitization, system interoperability, data science literacy, and active participation in regional and international open science initiatives. These actions would enhance the visibility and global impact of national scientific output.

In light of these reflections, the Pinar del Río Journal of Medical Sciences, as a forum for scientific dialogue and knowledge dissemination, recognizes the transition to a culture based on open data, reproducible research, and real-world evidence as a strategic opportunity to strengthen the social impact of science and its contribution to health decision-making. Editorial excellence no longer depends solely on publishing novel results, but also on ensuring transparency, accessibility, and verifiability.

Therefore, we promote research committed to the principles of open science, data ethics, and methodological rigor, understanding that these elements are essential for building public trust and moving toward more equitable and sustainable healthcare systems. The future of medical research will not be defined by who accumulates the most information, but by who is able to transform it into useful, reproducible, and shared knowledge for the collective benefit.

BIBLIOGRAPHIC REFERENCES

1. Cremin CJ, Dash S, Huang X. Big Data: Historic Advances and Emerging Trends in Biomedical Research. Curr Res Biotechnol [Internet]. 2022 [citado 28/06/2026]; 4:138-151. Disponible en: <https://doi.org/10.1016/j.crbiot.2022.02.004>
2. Raghupathi W, Raghupathi V. Big data in healthcare and medicine revisited: design and managerial challenges in the age of artificial intelligence. Health Inf Sci Syst [Internet]. 2026 [citado 28/06/2026]; 14:12. Disponible en: <https://doi.org/10.1007/s13755-026-00433-2>
3. Burns L, Le Roux N, Kalesnik-Orszulak R, Christian J, Dudinak J, Rockhold F, et al. Real-world evidence for regulatory decision-making: updated guidance from around the world. Front Med (Lausanne) [Internet]. 2023 [citado 28/06/2026]; 10:1236462. Disponible en: <https://doi.org/10.3389/fmed.2023.1236462>
4. Beesley LJ, Salvatore M, Fritsche LG, Pandit A, Rao A, Brummett C, Willer CJ, Lisabeth LD, Mukherjee B. The emerging landscape of health research based on biobanks linked to electronic health records: Existing resources, statistical challenges, and potential opportunities. Stat Med [Internet]. 2020 [citado 28/06/2026]; 39(6):773-800. Disponible en: <https://doi.org/10.1002/sim.8445>
5. Ellis RJ. Questionable Research Practices, Low Statistical Power, and Other Obstacles to Replicability: Why Preclinical Neuroscience Research Would Benefit from Registered Reports. eNeuro [Internet]. 2022 [citado 28/06/2026]; 9(4):ENEURO.0017-22.2022. Disponible en: <https://doi.org/10.1523/eneuro.0017-22.2022>