

PROF. FEDERICO CARBONE (Orcid ID : 0000-0003-2957-4078)

Article type : Editorial

## **Big data and data sharing: opportunities for the urgent challenges in cardiovascular disease**

Federico Carbone<sup>1,2</sup> and Fabrizio Montecucco<sup>2,3</sup>

<sup>1</sup>Ospedale Policlinico San Martino Genoa – Italian Cardiovascular Network, 10 Largo Benzi, Genoa 16132, Italy. <sup>2</sup>First Clinic of Internal Medicine, Department of Internal Medicine, University of Genoa, 6 viale Benedetto XV, Genoa 16132, Italy. <sup>3</sup>First Clinic of Internal Medicine, Department of Internal Medicine and Centre of Excellence for Biomedical Research (CEBR), University of Genoa, 6 viale Benedetto XV, Genoa 16132, Italy.

**Correspondence to:** Federico Carbone, MD, PhD. First Clinic of Internal Medicine, Department of Internal Medicine, University of Genoa, 6 viale Benedetto XV, 16132 Genoa, Italy. Tel: +39 010 33 51054; Fax: +39 010 353 8686; E-mail: federico.carbone@unige.it

**Running title:** Big data in cardiovascular research

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/ECL.13188](https://doi.org/10.1111/ECL.13188)

This article is protected by copyright. All rights reserved

The past half-century has witnessed unbelievable progress in cardiovascular (CV) medicine, determining a swirling reduction in mortality. Despite this progress, the decline in incidence and mortality leveled off in recent years. Age-adjusted CV mortality remained flat independently of race and ethnicity, with only few exceptions. In other words, disparity in CV health is still observed across sex, race and ethnicity. The rise of obesity, diabetes and other risk factors has been widely used to explain this trend in CV mortality. However, inequality in CV health recognizes other determinants. Geographic location, wealth and education are associated with different access to basic primary care and approach to modifiable CV risk factors <sup>1</sup>. As extensively reviewed, they influence awareness, treatment and control of dyslipidemia, hypertension, stroke, myocardial infarction and childhood obesity among the others.

Noteworthy, this stagnating trend in CV outcome comes together with a progressive rise in CV spending, which includes both direct and indirect costs (e.g. loss of work productivity, and the need of household assistance). As similar or even better outcomes may be reproduced at substantially lower costs in certain context, this raises questions about the suitability of CV healthcare spending <sup>2</sup>. In light of this, the American Heart Association recently recognized six categories of missed opportunities to improve CV care and avoid unnecessary costs: i) risk factor modifications; ii) patient engagement and involvement; iii) correct diagnosis; iv) adherence and proper use of first-line treatments; v) proper use of advance treatments; vi) better use of supportive devices <sup>3</sup>. The current challenge is then to address the shortfalls of existing treatments, also facing the continuous innovations in CV health management. Similar to the European Heart Network, also the American Heart Association has recently established the Value in Health Care Initiative to increase awareness, access and affordability while eliminating barriers to necessary care <sup>3</sup>.

However, the planned goals of those initiatives require new forms of CV research, able to analyze large amounts of information. Large volumes of data are often available as the sample size of CV studies hugely increased along time. While the Framingham Heart study started in 1948 with 5,209 patients at first enrollment, the most recent studies are enrolling hundreds of thousands of patients and even more than one million patients are expected in future studies. The electronic health record (EHR) is now considered the

Accepted Article

main source of “big data” for biomedical research. They merge traditional data (e.g. clinical and imaging) with other relevant information from omics, data internet use, wearable devices and others <sup>4</sup>. As the information increases, their variety also increases. Much of the information obtained from HER are unstructured and therefore difficult to categorize and analyze. Also considering the time for information creation and storing, this remains one of the main challenge in big data analysis <sup>5</sup>. Machine/deep learning approach integrate and interpret complex biomedical and healthcare data also developing automated risk prediction algorithms. CV disease represents a groups of diseases that can benefit greatly those algorithms. Their application in prospective clinical trials may will have a role for identifying behavioral drivers and therapeutic pathways, essential for implementing precision diagnostics, risk stratification and personalized therapies.

Other times, data are part of datasets not linked each other. Here, the challenge is to remove barriers between data sources by developing specific platforms. Much data generated in clinical trials (someone claims up 50%) are kept out, but not at random. Positive and statistically significant trials are more likely to be published. Many data from clinical trials are often underutilized, whereas third-party scientists could provide an independent validation or even provide response to unaddressed question. Data sharing revolution has taken shape in early 2000’s and was endorsed by prominent politicians in US. To date, the international Committee of Medical Journal Editors argues for an “ethical obligation to responsibly share data”, whereas the National Institute of health requires data sharing for investigator founded with more than a half-million dollars. In parallel, industry-, university- and non-profit-based sharing platform are rising <sup>6</sup>. A very interesting example of sharing data potential has been observed with the SPRINT trial of which analysis was implemented and event corrected by external researchers <sup>7,8</sup>.

The road map established for an affordable and sustainable CV health is bound together with big data development and data sharing. As further step to promote CV health globally, patient and public attitude to share health data for research should be implemented. Media sources (internet, television, social media) will may have a role in this process <sup>9,10</sup>.

## References

1. Dean CA, Zhang D, Kulchyski KT, Ventline B, Jagirdar R, Milan RA. Social Determinants of Cardiovascular Disease among Michigan Residents: A Call to Action for Population Health Initiatives. *Journal of racial and ethnic health disparities*. 2019.
2. Alyesh DM, Seth M, Miller DC, et al. Exploring the Healthcare Value of Percutaneous Coronary Intervention: Appropriateness, Outcomes, and Costs in Michigan Hospitals. *Circulation Cardiovascular quality and outcomes*. 2018;11(6):e004328.
3. McClellan M, Brown N, Califf RM, Warner JJ. Call to Action: Urgent Challenges in Cardiovascular Disease: A Presidential Advisory From the American Heart Association. *Circulation*. 2019;139(9):e44-e54.
4. Hemingway H, Asselbergs FW, Danesh J, et al. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. *European heart journal*. 2018;39(16):1481-1495.
5. Silverio A, Cavallo P, De Rosa R, Galasso G. Big Health Data and Cardiovascular Diseases: A Challenge for Research, an Opportunity for Clinical Care. *Frontiers in medicine*. 2019;6:36.
6. Dey P, Ross JS, Ritchie JD, Desai NR, Bhavnani SP, Krumholz HM. Data Sharing and Cardiology: Platforms and Possibilities. *Journal of the American College of Cardiology*. 2017;70(24):3018-3025.
7. Burns NS, Miller PW. Learning What We Didn't Know - The SPRINT Data Analysis Challenge. *The New England journal of medicine*. 2017;376(23):2205-2207.
8. Warner F, Dhruva SS, Ross JS, Dey P, Murugiah K, Krumholz HM. Accurate estimation of cardiovascular risk in a non-diabetic adult: detecting and correcting the error in the reported Framingham Risk Score for the Systolic Blood Pressure Intervention Trial population. *BMJ open*. 2018;8(7):e021685.
9. Agarwala A, Kohli P, Virani SS. Popular Media and Cardiovascular Medicine: "with Great Power There Must Also Come Great Responsibility". *Current atherosclerosis reports*. 2019;21(11):43.

10. Kalkman S, van Delden J, Banerjee A, Tyl B, Mostert M, van Thiel G. Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. *Journal of medical ethics*. 2019.