

Editorial

The Big Data Era in Cardiology and Cardiovascular Medicine: Advanced Analytics for Truly Personalized Care

Pasquale Ambrosino ^{1,*}, Fabio Manzo ^{2,†}, Claudio Candia ³, Giorgio Alfredo Spedicato ⁴, Guido Grassi ^{5,*} and Mauro Maniscalco ^{6,7}

¹ Istituti Clinici Scientifici Maugeri IRCCS, Scientific Directorate of Telese Terme Institute, 82037 Telese Terme, Italy

² Fleming Clinical Laboratory, 81020 Casapulla, Italy; fabioce26@gmail.com

³ Department of Biomedicine, Neurosciences and Advanced Diagnostics, University of Palermo, 90100 Palermo, Italy; c.candia@outlook.it

⁴ Department of Statistics and Quantitative Methods, Milano-Bicocca University, 20126 Milan, Italy; spedicato_giorgio@yahoo.it

⁵ Department of Medicine and Surgery, Milano-Bicocca University, 20126 Milan, Italy

⁶ Istituti Clinici Scientifici Maugeri IRCCS, Pulmonary Rehabilitation Unit of Telese Terme Institute, 82037 Telese Terme, Italy; mauro.maniscalco@icsmaugeri.it

⁷ Department of Clinical Medicine and Surgery, Federico II University, 80131 Naples, Italy

* Correspondence: pasquale.ambrosino@icsmaugeri.it (P.A.); guido.grassi@unimib.it (G.G.)

† These authors contributed equally to this work.

In recent years, a growing body of alarming epidemiological evidence has suggested that cardiovascular diseases (CVDs) are the leading cause of disability and premature mortality worldwide. According to estimates from the Global Burden of Disease (GBD) study, CVDs were responsible for approximately 19 million deaths globally in 2021 and accounted for 438 million disability-adjusted life years (DALYs) lost [1]. This causes a profound impact on healthcare systems and national economies worldwide, posing an increasingly urgent challenge for both the scientific community and public health authorities [2]. The situation is further exacerbated by discouraging future projections, which anticipate a continued rise in prevalence, mortality, and DALYs between 2025 and 2050, a trend largely driven by the aging global population [3]. These alarming forecasts highlight the pressing need for two complementary approaches, the first involving the implementation of broad population-based prevention strategies, and the second focusing on the promotion of increasingly personalized pharmacological and rehabilitation interventions tailored to the specific needs of individual patients [4].

Although the ultimate goal of personalized medicine is to deliver the right treatment to the right patient [5], its implementation in cardiovascular care is still limited by the fact that most clinical guidelines are derived from evidence based on the “average patient” model [6]. Randomized controlled trials (RCTs), for instance, typically report aggregate outcomes across entire study populations and only subsequently attempt to assess the heterogeneity of treatment effects using traditional subgroup analyses, with forest plots stratifying results by single clinical or demographic variables. While this approach provides a valuable overview, it fails to fully account for the true clinical complexity of individual patients and the interplay of multiple variables coexisting across different patient clusters [4]. In this context, advanced statistical tools and top-tier evidence, including meta-analyses of RCTs, can offer more robust and consistent insights. However, while methodologically sound and effective at the population level, this approach remains fundamentally distant from the core principles of precision medicine [5]. Truly personalized care comes from



Received: 26 June 2025

Accepted: 1 July 2025

Published: 3 July 2025

Citation: Ambrosino, P.; Manzo, F.; Candia, C.; Spedicato, G.A.; Grassi, G.; Maniscalco, M. The Big Data Era in Cardiology and Cardiovascular Medicine: Advanced Analytics for Truly Personalized Care. *Diagnostics* **2025**, *15*, 1705. <https://doi.org/10.3390/diagnostics15131705>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

personalized diagnostics, which requires a deep understanding of the biological and molecular mechanisms underlying disease processes and their phenotypic expression in different individuals [6]. Precision medicine, therefore, is not merely about stratifying existing data, but about integrating molecular (genomic, transcriptomic, proteomic, metabolomic), phenotypic, and clinical data to generate patient-specific approaches, thus improving both prevention and treatment strategies while enhancing health outcomes [7]. In recent years, the field of omics has made significant progress, laying the foundation for future generations to benefit from truly personalized medicine. However, this approach is still often inaccessible due to its high costs, advanced technical requirements, and limited feasibility, especially in resource-limited contexts [8].

Concurrently, healthcare systems are being reshaped by several changes, many of which have been facilitated by the COVID-19 pandemic. Specifically, this global crisis induced governments around the world to step up their digital transformation efforts, promoting telemedicine and patient electronic monitoring, processes that would have otherwise taken much longer to implement [9]. Therefore, with massive volumes of health information collected from electronic health records (EHRs), the healthcare industry is now uniquely positioned to benefit from data-driven approaches to guiding innovation and improving care at both the individual and population levels [10]. Data from EHRs are a relatively recent development and should be integrated with complementary data sources, including administrative databases, imaging data, patient-reported outcomes, and environmental or socioeconomic data [11]. An example of such integration is the TriNetX platform, which was launched in 2014 and is now powered by real-world data from nearly 300 million patients across more than 30 countries, sourced from EHRs, claims, and other databases from hundreds of healthcare organizations [12]. This network represents a unique opportunity to enhance the optimization of RCTs and facilitate large-scale observational studies, particularly in the cardiovascular field [12]. Numerous articles based on thousands of patient records have been recently published utilizing the TriNetX network, with a particular focus on the clinical efficacy of novel cardiovascular therapies such as glucagon-like peptide-1 (GLP-1) receptor agonists [13,14]. These investigations aim to employ large-scale real-world data to rapidly address clinical questions on novel pharmacological and rehabilitation approaches that traditionally would have required epidemiological evidence collected over many years. Additionally, the platform has contributed to recent findings regarding a reduced incidence of CVD following herpes zoster vaccination [15], the underdiagnosis of cardiac sarcoidosis in young patients with unexplained heart block [16], and the role of routine cardiac biomarkers in predicting major adverse cardiac events in patients with glomerulonephritis [17].

In this evolving landscape, a growing number of clinical registries, promoted by scientific societies since the 1980s, further contributes to the expanding volume of large-scale health information now available in the digital format [11]. Some of the most representative CVD registries currently in use have been developed through the initiatives of the American College of Cardiology (ACC), the Society of Thoracic Surgeons (STS), the American Heart Association (AHA), and the American Stroke Association (ASA) [18]. Among these, the Transcatheter Valve Therapy (TVT) registry plays an important role in collecting and analyzing data on patients undergoing transcatheter valvular procedures [19]. With millions of patient records and hundreds of high-profile studies, the largest-scale initiative remains the National Cardiovascular Data Registry (NCDR) with its sub-registries, created by the ACC in 1987 to describe the clinical characteristics and outcomes of patients undergoing cardiac catheterization and coronary interventions [20]. Similarly, the Get With The Guidelines (GWTG) Database is another hospital-based initiative led by the AHA and the ASA,

collecting data from thousands of hospitals to improve diagnostics and quality of care in CVDs, such as stroke, atrial fibrillation, and heart failure [21].

Alongside this vast amount of data collected for professional and scientific purposes, the massive volume of information collected from outside healthcare systems through smartphones and wearable technologies is increasingly facilitating the continuous monitoring of lifestyle and cardiovascular parameters, including heart rate, blood pressure, oxygen saturation, sleep quality or even electrocardiographic (ECG) data [22]. Therefore, a number of large-scale studies have been initiated, including the Apple Heart Study, which investigated the ability of optical sensors embedded in a commercial smartwatch to detect atrial fibrillation in real-world settings [23]. This shift toward real-time, longitudinal health monitoring represents a potentially revolutionary step forward, especially in cardiovascular care, due to the nature of the parameters most easily captured by wearable devices [24].

Overall, while omics technologies are not yet widely available or fully integrated into clinical practice [8], “big data” derived from EHRs and other complementary sources now represent a solid base for generating predictive models and identifying patterns [25], supported by the large amount of observations and increasingly sophisticated analytical methods. Big data, in fact, usually describes extremely large and complex datasets that exceed the capacity of conventional tools and processing infrastructures to effectively manage and analyze information. Artificial intelligence (AI), including machine learning and deep learning techniques, as well as data science approaches have recently shown potential in transforming risk assessment, early diagnosis, and the personalization of therapy in cardiovascular medicine [26]. In fact, machine learning models can capture complex non-linear relationships and interactions among variables, perform effective feature selection, and offer interpretability through specific tools (e.g., Shapley values), which reduce the perception of opacity often associated with these approaches [27]. Additionally, given the multifactorial nature of CVDs, multimodal deep learning models even enable the integration of different data sources in different formats, such as images, ECG sequences, and tabular clinical data, thus facilitating a unified analysis of heterogenous cardiac information to improve diagnostic and prognostic assessments [27].

Traditionally, imaging and ECG signal interpretation have represented the earliest and most successful areas of integration between AI and cardiovascular medicine, supporting early detection while minimizing scan duration and exposure to ionizing radiation [28]. Since the first algorithm received regulatory clearance in 1995, the Food and Drug Administration has approved more than 60 clinical AI applications specifically for cardiovascular medicine, out of a total of over 600 approvals, with the majority granted in the past five years [26]. Thus, as coronary artery calcium scoring becomes a key tool in cardiovascular risk stratification, AI algorithms are now able to automatically quantify calcium scores not only from low-dose chest computed tomography (CT) scans, but also from nuclear imaging techniques such as positron emission tomography (PET)/CT [28]. AI-assisted applications have also been successfully implemented in the volumetric and structural analysis of cardiac chambers using cardiac magnetic resonance (CMR) [29], as well as in guiding procedures for the ablation of infarct-related ventricular tachycardia [30]. Similarly, AI has been used in electrocardiography to diagnose atrial fibrillation or predict imminent ventricular arrhythmias [31], in coronary angiography to identify stenosis from angiograms [29], and in echocardiography to automatically evaluate valvular structures and flow gradients, calculate ejection fraction, measure longitudinal strain, and detect abnormalities in cardiac wall motion [28]. The Digital Aortic Stenosis Severity Index (DASSi) is an example of an AI tool that can not only recognize severe aortic stenosis from a single echocardiographic image but also stratify individuals based on their future risk of disease onset or rapid progression [32]. Models based on more computationally intensive methods,

such as boosting algorithms (e.g., gradient boosting machines, LightGBM, XGBoost) and deep learning architectures, have been applied to EHRs in various clinical studies to predict the risk of incident hypertension, heart failure readmissions, and ischemic stroke [28,31]. Moreover, the opportunities introduced by virtual reality and wearable technologies are stimulating increasing interest in the use of AI to monitor and support physical exercise within both home-based and in-hospital rehabilitation programs [33–35].

Overall, as we gradually approach the era of clinical omics, we already possess an enormous volume of heterogeneous electronic information from a wide range of sources. The central challenge in this evolving landscape is achieving effective data integration and interoperability across systems and platforms. Although important concerns regarding privacy and equity persist, this is precisely where AI and advanced computational approaches are beginning to make a significant difference, helping to extract and unify complex datasets and drive the transition toward truly individualized, data-informed cardiovascular diagnostics and care.

Author Contributions: Conceptualization, P.A. and F.M.; writing—original draft preparation, P.A., F.M., G.G. and M.M.; writing—review and editing, C.C. and G.A.S.; supervision, M.M. and G.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by the “Ricerca Corrente” funding scheme of the Ministry of Health, Italy.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No datasets were generated or analyzed during the current study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. GBD 2021 Causes of Death Collaborators. Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990–2021: A systematic analysis for the Global Burden of Disease Study 2021. *Lancet* **2024**, *403*, 2100–2132. [[CrossRef](#)] [[PubMed](#)]
2. Wu, Y.; Benjamin, E.J.; MacMahon, S. Prevention and Control of Cardiovascular Disease in the Rapidly Changing Economy of China. *Circulation* **2016**, *133*, 2545–2560. [[CrossRef](#)]
3. Chong, B.; Jayabaskaran, J.; Jauhari, S.M.; Chan, S.P.; Goh, R.; Kueh, M.T.W.; Li, H.; Chin, Y.H.; Kong, G.; Anand, V.V.; et al. Global burden of cardiovascular diseases: Projections from 2025 to 2050. *Eur. J. Prev. Cardiol.* **2024**. [[CrossRef](#)] [[PubMed](#)]
4. Yeh, R.W.; Kramer, D.B. Decision Tools to Improve Personalized Care in Cardiovascular Disease: Moving the Art of Medicine Toward Science. *Circulation* **2017**, *135*, 1097–1100. [[CrossRef](#)]
5. Abrahams, E. Right drug-right patient-right time: Personalized medicine coalition. *Clin. Transl. Sci.* **2008**, *1*, 11–12. [[CrossRef](#)] [[PubMed](#)]
6. Leopold, J.A.; Loscalzo, J. Emerging Role of Precision Medicine in Cardiovascular Disease. *Circ. Res.* **2018**, *122*, 1302–1315. [[CrossRef](#)] [[PubMed](#)]
7. Giugni, F.R.; Berry, J.D.; Khera, A.; Shah, A.M.; de Lemos, J.A. Precision Medicine for Cardiovascular Prevention and Population Health: A Bridge Too Far? *Circulation* **2024**, *150*, 1720–1731. [[CrossRef](#)]
8. Tarazona, S.; Arzalluz-Luque, A.; Conesa, A. Undisclosed, unmet and neglected challenges in multi-omics studies. *Nat. Comput. Sci.* **2021**, *1*, 395–402. [[CrossRef](#)]
9. Horgan, D.; Hackett, J.; Westphalen, C.B.; Kalra, D.; Richer, E.; Romao, M.; Andreu, A.L.; Lal, J.A.; Bernini, C.; Tumiene, B.; et al. Digitalisation and COVID-19: The Perfect Storm. *Biomed. Hub.* **2020**, *5*, 1341–1363. [[CrossRef](#)]
10. Sauer, C.M.; Chen, L.C.; Hyland, S.L.; Girbes, A.; Elbers, P.; Celi, L.A. Leveraging electronic health records for data science: Common pitfalls and how to avoid them. *Lancet Digit. Health* **2022**, *4*, e893–e898. [[CrossRef](#)]
11. Weintraub, W.S. Role of Big Data in Cardiovascular Research. *J. Am. Heart Assoc.* **2019**, *8*, e012791. [[CrossRef](#)] [[PubMed](#)]
12. Topaloglu, U.; Palchuk, M.B. Using a Federated Network of Real-World Data to Optimize Clinical Trials Operations. *JCO Clin. Cancer Inform.* **2018**, *2*, 1–10. [[CrossRef](#)] [[PubMed](#)]

13. Kishimori, T.; Kato, T.; Wada, A.; Tani, A.; Yamaji, R.; Koike, J.; Iwasaki, Y.; Matsumoto, T.; Yagi, T.; Okada, M. Comparative Outcomes of Glucagon-Like Peptide-1 Receptor Agonists to Dipeptidyl Peptidase 4 Inhibitors in Patients with Heart Failure and Type 2 Diabetes. *J. Am. Heart Assoc.* **2025**, *14*, e037510. [[CrossRef](#)]
14. Wu, J.Y.; Tseng, K.J.; Kao, C.L.; Hung, K.C.; Yu, T.; Lin, Y.M. Clinical effectiveness of tirzepatide for patients with atrial fibrillation and type 2 diabetes: A retrospective cohort study. *Diabetes Res. Clin. Pract.* **2025**, *225*, 112279. [[CrossRef](#)]
15. Kornelius, E.; Lo, S.C.; Huang, C.N.; Wang, C.C.; Wang, Y.H.; Yang, Y.S. Association of herpes zoster vaccination and cardiovascular risk in patients with diabetes: Long-term insights from a retrospective cohort study. *BMJ Open* **2025**, *15*, e090428. [[CrossRef](#)] [[PubMed](#)]
16. Goldar, G.; Sifuentes, A.A.; Gokul, K.; ElBakkush, U.M.; Mhanna, M.; Farjo, P.; Dominic, P. Raising the Alarm: Underdiagnosis of Cardiac Sarcoidosis in Young Patients with Unexplained Heart Block. *JACC Adv.* **2025**, *4*, 101890. [[CrossRef](#)] [[PubMed](#)]
17. Davies, E.M.; Buckley, B.J.R.; Austin, P.; Lip, G.Y.H.; Oni, L.; McDowell, G.; Rao, A. Routine cardiac biomarkers for the prediction of incident major adverse cardiac events in patients with glomerulonephritis: A real-world analysis using a global federated database. *BMC Nephrol.* **2024**, *25*, 233. [[CrossRef](#)]
18. Meltzer, S.N.; Weintraub, W.S. The Role of National Registries in Improving Quality of Care and Outcomes for Cardiovascular Disease. *Methodist. Debakey Cardiovasc. J.* **2020**, *16*, 205–211. [[CrossRef](#)]
19. Carroll, J.D.; Edwards, F.H.; Marinac-Dabic, D.; Brindis, R.G.; Grover, F.L.; Peterson, E.D.; Tuzcu, E.M.; Shahian, D.M.; Rumsfeld, J.S.; Shewan, C.M.; et al. The STS-ACC transcatheter valve therapy national registry: A new partnership and infrastructure for the introduction and surveillance of medical devices and therapies. *J. Am. Coll. Cardiol.* **2013**, *62*, 1026–1034. [[CrossRef](#)]
20. Masoudi, F.A.; Ponirakis, A.; Yeh, R.W.; Maddox, T.M.; Beachy, J.; Casale, P.N.; Curtis, J.P.; De Lemos, J.; Fonarow, G.; Heidenreich, P.; et al. Cardiovascular care facts: A report from the national cardiovascular data registry: 2011. *J. Am. Coll. Cardiol.* **2013**, *62*, 1931–1947. [[CrossRef](#)]
21. Ellrodt, A.G.; Fonarow, G.C.; Schwamm, L.H.; Albert, N.; Bhatt, D.L.; Cannon, C.P.; Hernandez, A.F.; Hlatky, M.A.; Luepker, R.V.; Peterson, P.N.; et al. Synthesizing lessons learned from get with the guidelines: The value of disease-based registries in improving quality and outcomes. *Circulation* **2013**, *128*, 2447–2460. [[CrossRef](#)] [[PubMed](#)]
22. Williams, G.J.; Al-Baraikhan, A.; Rademakers, F.E.; Ciravegna, F.; van de Vosse, F.N.; Lawrie, A.; Rothman, A.; Ashley, E.A.; Wilkins, M.R.; Lawford, P.V.; et al. Wearable technology and the cardiovascular system: The future of patient assessment. *Lancet Digit. Health* **2023**, *5*, e467–e476. [[CrossRef](#)]
23. Perez, M.V.; Mahaffey, K.W.; Hedlin, H.; Rumsfeld, J.S.; Garcia, A.; Ferris, T.; Balasubramanian, V.; Russo, A.M.; Rajmane, A.; Cheung, L.; et al. Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation. *N. Engl. J. Med.* **2019**, *381*, 1909–1917. [[CrossRef](#)]
24. Hughes, A.; Shandhi, M.M.H.; Master, H.; Dunn, J.; Brittain, E. Wearable Devices in Cardiovascular Medicine. *Circ. Res.* **2023**, *132*, 652–670. [[CrossRef](#)] [[PubMed](#)]
25. Lau, E.; Wu, J.C. Omics, Big Data, and Precision Medicine in Cardiovascular Sciences. *Circ. Res.* **2018**, *122*, 1165–1168. [[CrossRef](#)] [[PubMed](#)]
26. Stamate, E.; Piraianu, A.I.; Ciobotaru, O.R.; Crassas, R.; Duca, O.; Fulga, A.; Grigore, I.; Vintila, V.; Fulga, I.; Ciobotaru, O.C. Revolutionizing Cardiology through Artificial Intelligence-Big Data from Proactive Prevention to Precise Diagnostics and Cutting-Edge Treatment-A Comprehensive Review of the Past 5 Years. *Diagnostics* **2024**, *14*, 1103. [[CrossRef](#)]
27. Kolk, M.Z.H.; Ruiperez-Campillo, S.; Allaart, C.P.; Wilde, A.A.M.; Knops, R.E.; Narayan, S.M.; Tjong, F.V.Y.; DEEP RISK investigators. Multimodal explainable artificial intelligence identifies patients with non-ischaemic cardiomyopathy at risk of lethal ventricular arrhythmias. *Sci. Rep.* **2024**, *14*, 14889. [[CrossRef](#)]
28. Armoundas, A.A.; Narayan, S.M.; Arnett, D.K.; Spector-Bagdady, K.; Bennett, D.A.; Celi, L.A.; Friedman, P.A.; Gollob, M.H.; Hall, J.L.; Kwitek, A.E.; et al. Use of Artificial Intelligence in Improving Outcomes in Heart Disease: A Scientific Statement from the American Heart Association. *Circulation* **2024**, *149*, e1028–e1050. [[CrossRef](#)]
29. Elias, P.; Jain, S.S.; Poterucha, T.; Randazzo, M.; Lopez Jimenez, F.; Khera, R.; Perez, M.; Ouyang, D.; Pirruccello, J.; Salerno, M.; et al. Artificial Intelligence for Cardiovascular Care-Part 1: Advances: JACC Review Topic of the Week. *J. Am. Coll. Cardiol.* **2024**, *83*, 2472–2486. [[CrossRef](#)]
30. Prakosa, A.; Arevalo, H.J.; Deng, D.; Boyle, P.M.; Nikolov, P.P.; Ashikaga, H.; Blauer, J.J.E.; Ghafoori, E.; Park, C.J.; Blake, R.C., 3rd; et al. Personalized virtual-heart technology for guiding the ablation of infarct-related ventricular tachycardia. *Nat. Biomed. Eng.* **2018**, *2*, 732–740. [[CrossRef](#)]
31. Khera, R.; Oikonomou, E.K.; Nadkarni, G.N.; Morley, J.R.; Wiens, J.; Butte, A.J.; Topol, E.J. Transforming Cardiovascular Care with Artificial Intelligence: From Discovery to Practice: JACC State-of-the-Art Review. *J. Am. Coll. Cardiol.* **2024**, *84*, 97–114. [[CrossRef](#)] [[PubMed](#)]
32. Holste, G.; Oikonomou, E.K.; Mortazavi, B.J.; Coppi, A.; Faridi, K.F.; Miller, E.J.; Forrest, J.K.; McNamara, R.L.; Ohno-Machado, L.; Yuan, N.; et al. Severe aortic stenosis detection by deep learning applied to echocardiography. *Eur. Heart J.* **2023**, *44*, 4592–4604. [[CrossRef](#)] [[PubMed](#)]

33. Witharana, P.; Chang, L.; Maier, R.; Ogundimu, E.; Wilkinson, C.; Athanasiou, T.; Akowuah, E. Feasibility study of rehabilitation for cardiac patients aided by an artificial intelligence web-based programme: A randomised controlled trial (RECAP trial)-a study protocol. *BMJ Open* **2024**, *14*, e079404. [[CrossRef](#)] [[PubMed](#)]
34. Sotirakos, S.; Fouda, B.; Mohamed Razif, N.A.; Cribben, N.; Mulhall, C.; O'Byrne, A.; Moran, B.; Connolly, R. Harnessing artificial intelligence in cardiac rehabilitation, a systematic review. *Future Cardiol.* **2022**, *18*, 154–164. [[CrossRef](#)]
35. Torres, R.; Zurita, C.; Mellado, D.; Nicolis, O.; Saavedra, C.; Tuesta, M.; Salinas, M.; Bertini, A.; Pedemonte, O.; Querales, M.; et al. Predicting Cardiovascular Rehabilitation of Patients with Coronary Artery Disease Using Transfer Feature Learning. *Diagnostics* **2023**, *13*, 508. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.