Commentary

Secondary Prevention of Cardiovascular Disease: Time to Rethink Stratification of Disease Severity?

Ralph K Akyea[1]; Joe Kai[1]; Nadeem Qureshi[1]; Hasidah Abdul Hamid[1]; Stephen F Weng[1]

Affiliations:

[1] Primary Care Stratified Medicine (PRISM), Division of Primary Care, University of Nottingham, Nottingham, UK

Correspondence to:

Dr Ralph Kwame Akyea

Primary Care Stratified Medicine (PRISM), Division of Primary Care, Floor 13-15, The Tower Building, University Park Campus, University of Nottingham, NG7 2RD UK

Tel: +44 (0) 115 748 6834

Email: Ralph.Akyea@nottingham.ac.uk

Word count: **1,532** (including references); **1,152** (without references)

Cardiovascular disease (CVD) remains a leading cause of mortality globally,[1] but with improved management, CVD it less fatal.[2] However, this means a large proportion of the people are living with this long term condition. As of 2015, one in seventeen (1 in 17) of the global population had CVD.[3]

There is lack of evidence across countries to truly quantify the significant economic consequences of this pressing global health issue.[4] Overall, CVD is estimated to cost the European Union (EU) economy €210 billion a year – around 53% (€111 billion) due to health care costs, 26% (€54 billion) to productivity losses and 21% (€45 billion) to the informal care of people with CVD.[5]

The primary prevention of CVD focuses on altering known modifiable risk factors such as improving one's diet, exercising more, losing weight or quitting smoking to prevent disease onset. However, there are many genetic and environmental factors that cannot be controlled by an individual most of the time. For instance, an individual with an undiagnosed inherited condition which greatly increases cholesterol will struggle to reduce their risk through lifestyle changes alone. In these instances, high-intensity lipid therapies will be needed. Where primary prevention of CVD fails due to unmodifiable risk factors, secondary prevention becomes important. For people with established CVD, the priority is to prevent a second CVD event and improve the quality of their lives.

Secondary prevention focuses on minimizing the impact of a disease condition, and its impact on patient lives. This may reduce the socioeconomic burden on individual households as well as the health care system in general. A crucial step for secondary prevention is the early identification of risk markers for subsequent events or complications which might be prevented. Crucially, this needs to embrace the identification of risk factors, their interactions, and how these variations in these risk factors may relate to CVD severity. Although this process might seem more complex than primary prevention, the benefits of secondary prevention can be comparatively substantial compared to tertiary prevention (such as cardiac rehabilitation programmes) for preserving quality of life for patients and reducing costs to individuals or the health system.

Advances in technology in medicine have enabled clinicians to identify and manage patients with various conditions more effectively and efficiently. For instance, recent innovations in artificial intelligence, are improving the earlier and timelier prediction of disease conditions or their complications, thus improving their prognosis.[6] Earlier prediction can ensure early treatment interventions before patients' disease conditions worsen or suffer complications such as heart attack or stroke, thus improving life expectancy and quality of life.[7]

Patients' electronic health records (EHRs), employed routinely for clinical care, provide large quantities of clinical data that could usefully drive further new research using these innovative methodologies to inform secondary prevention. EHRs provide low-cost means of accessing potentially rich longitudinal data on large populations at the granular level of patients, across different types of health care settings. Despite the challenges and inherent biases associated with using routinely collected data in EHRs, their potential has been recognised, with their utility and functionality increasing rapidly in the past decades for health research.[6,8] To improve the validity of findings from using routinely available clinical data in EHRs, robust approaches are constantly being developed to improve the quality of the data available. For instance, creating new data linkages from biobanks which contain genomic data to EHRs offers the potential for conducting casual epidemiological study designs.[9] This opportunity allows us to derive meaningful and valid conclusions from studies using EHRs.[10]

There is also increasing potential for research to improve the accuracy of risk stratification for secondary prevention. By interrogating the large volumes of clinical data, we can stratify or cluster all CVD patients into different risk groups according to risk factors or clinical characteristics. Traditionally, this has been explored using multivariable analysis, but with advances in data science we can use machine-learning techniques such as gradient boosting and neural networks. The advantage of this innovative methodology in developing robust disease prediction models to guide clinical decision-making is the ability to uncover 'hidden' interactions based on patient characteristics and applying these to predict future clinical events.[11]

Key benefits of using a data science approach for disease prediction are the accuracy and efficiency of the prediction process. Conventional risk prediction models usually estimate a risk score for an individual based on a weight assigned to a defined set of usually limited number of variables to reflect an average value in the population that the model was derived from. This process has been shown to often overestimate risk in low-risk individuals and underestimate risk in high-risk individuals.[12] Such inaccuracies and the clinical experience of this occurring in individual patients' trajectories', may cause a loss of trust in risk prediction models and may result in risk factors not being given the needed attention. By contrast machine learning approaches have the ability to efficiently signal correlations between risk factors that were previously unknown, something not easily achieved with conventional techniques for developing risk prediction models. These signals can then be verified using traditional epidemiology study designs to establish causality or association. This emerging approach has considerable potential to identify previously unknown correlations for risk prediction and thus improve the accuracy of stratification models that are developed. Figure 1 provides a conceptual representation of the role of data science (machine-learning) in risk stratification.

Moreover, such novel data science approaches may enable prediction of disease types (coronary heart disease, stroke, and peripheral heart disease), to enable more precise and focused management of individual patients. Rather than predict overall risk of CVD, we are now able to predict specific subtypes of disease,[8] such as CVD or even focused disease areas, example subtypes of coronary heart disease. For instance, Schiele et al,[13] recommend a careful selection of patients with stable coronary artery disease (with high residual risk and low therapeutic risk) for intense secondary prevention therapy. The intensified therapy, of anti-thrombotic and lipid-lowering medications, is efficacious in terms if ischaemic events and CVD mortality but could incur an excess haemorrhagic risk. EHRs and data science approaches offer the opportunity to profile the unique characteristics of this specific group of patients, ensuring intensified therapy is made available to those with greatest need and greatest benefit.

Secondary prevention is now one of the highest priorities across countries if we are to make a meaningful difference to the health and lives of the many individuals with established CVD. Alleviating pressure on already stretched health and social care services would be attendant major benefits. There is, however, a lack of consensus regarding the definition of severe CVD. There have been large advances in behavioural, therapeutic and interventional management of CVD in recent decades. However, given wide variations in disease severity in those with CVD, more precise stratification of differing severity risk is needed if the right patients are to benefit most. The potential to exploit data science to help in achieving this goal is considerable.

Author contributions

RKA and SFW contributed to the conception and drafted the manuscript. RKA, JK, NQ, HAH, SW critically revised the manuscript. All gave final approval and agree to be accountable for all aspects of work ensuring integrity and accuracy.

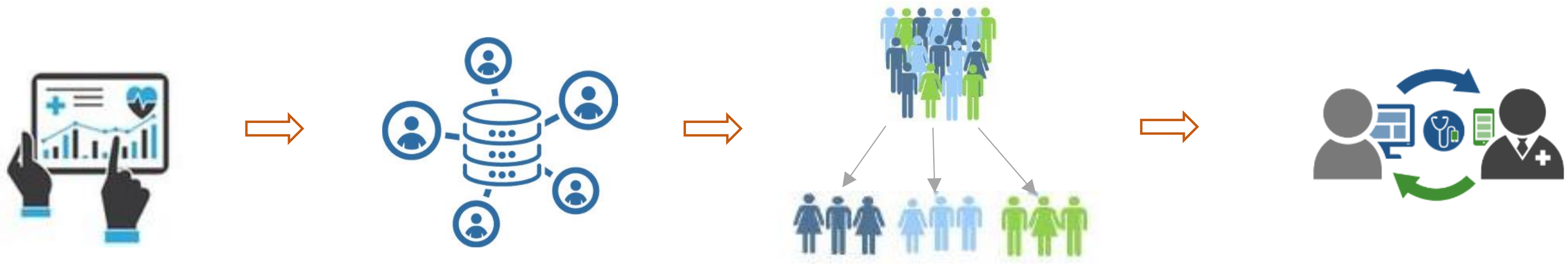
Conflict of Interest Disclosure

SFW is a member of the Clinical Practice Research Datalink (CPRD) Independent Scientific Advisory Committee (ISAC) and previously held an NIHR-SCPR Career Launching Fellowship Award (2015–2018). RKA currently holds a National Institute for Health Research (NIHR) School for Primary Care Research funded Studentship (2018–2021).

The views expressed are those of the authors and not necessarily those of the NIHR, the NHS, or the Department of Health and Social Care.

REFERENCES

1. Roth GA, Johnson C, Abajobir A, et al. Global, Regional, and National Burden of Cardiovascular Diseases for 10 Causes, 1990 to 2015. *J Am Coll Cardiol* 2017; 70: 1–25.
2. Rosamond WD, Chambless LE, Folsom AR, et al. Trends in the Incidence of Myocardial Infarction and in Mortality Due to Coronary Heart Disease, 1987 to 1994. *N Engl J Med* 1998; 339: 861–867.
3. Mozaffarian D. Global Scourge of Cardiovascular Disease: Time for Health Care Systems Reform and Precision Population Health. *J Am Coll Cardiol* 2017; 70: 26–28.
4. IOM (Institute of Medicine). *Promoting Cardiovascular Health in the Developing World: A Critical Challenge to Achieve Global Health*. Washington, DC: National Academies Press (US), 2010. Epub ahead of print 2010. DOI: 10.17226/12815.
5. Wilkins E, Wilson L, Wickramasinghe K, et al. *European Cardiovascular Disease Statistics 2017*. Brussels, 2017.
6. Henry KE, Hager DN, Pronovost PJ, et al. A targeted real-time early warning score (TREWScore) for septic shock. *Sci Transl Med* 2015; 7: 299ra122.
7. Adams A, Bojara W, Schunk K. Early Diagnosis and Treatment of Coronary Heart Disease in Asymptomatic Subjects With Advanced Vascular Atherosclerosis of the Carotid Artery (Type III and IV b Findings Using Ultrasound) and Risk Factors. *Cardiol Res* 2018; 22–27.
8. Ahlqvist E, Storm P, Käräjämäki A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *lancet Diabetes Endocrinol* 2018; 6: 361–369.
9. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014; 23: R89-98.
10. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009; 338: b2393.
11. Weng SF, Vaz L, Qureshi N, et al. Prediction of premature all-cause mortality: A prospective general population cohort study comparing machine-learning and standard epidemiological approaches. *PLoS One* 2019; 14: e0214365.
12. DeFilippis AP, Young R, Carrubba CJ, et al. An Analysis of Calibration and Discrimination Among Multiple Cardiovascular Risk Scores in a Modern Multiethnic Cohort. *Ann Intern Med* 2015; 162: 266.
13. Schiele F, Ecarnot F, Chopard R. Coronary artery disease: Risk stratification and patient selection for more aggressive secondary prevention. *Eur J Prev Cardiol* 2017; 24: 88–100.



Clinical records

Individual patient records from primary, secondary and tertiary care settings.

Linked dataset

Electronic health records linked to data from other national data registers. Provides information on clinical outcomes and wider social determinants of health.

Risk stratification

Using data science approach to group patients based on risk factors.

Clinical Management

Development of CVD risk stratification tool. Integration of tool into patient care pathways.

Figure 1. Conceptual representation of the role of data science (artificial intelligence) in risk stratification of patients