BIG DATA AND CARDIOLOGY: TIME FOR MASS ANALYTICS?

***Rahul Potluri,[1] Ignat Drozdov,[2] Paul Carter,[3] Jaydeep Sarma[4]**

1. ACALM Study Unit in collaboration with Aston Medical School, Aston University, Birmingham, UK
2. Bering Limited, London, UK
3. Royal Free London NHS Trust, London, UK
4. North West Heart Centre, University Hospital of South Manchester, Manchester, UK
**Correspondence to rahulpotluri@outlook.com*

Disclosure: Dr Rahul Potluri is the Founder of the ACALM Study Unit which undertakes clinical epidemiology research utilising anonymous routinely available data. Dr Ignat Drozdov is an employee of Bering Limited. The views expressed in the publication are his own and do not necessarily reflect those of the company.

The dawn of the millennium coupled with developments in technology has led to the computerisation of information previously collected with paper and filed in large warehouses. If the first decade of this century brought about the era of data collection, the current decade belongs to big data analysis. Big data covering topics such as how we shop, spend our money, go on holiday, seek healthcare, and finally the range of medical conditions we have, are all recorded, analysed, and scrutinised to the minutest detail for information, profit, and general and scientific knowledge; we have all benefitted from this.

In the commercial world, multinational giants such as Tesco have shown the way by introducing reward schemes, collecting information about consumer spending patterns and ingeniously encouraging customers to spend more. In healthcare, hospital and primary care records, demographics, medical conditions, and comorbidity and mortality information have all been captured in the majority of the Western world, primarily for financial reasons, but to a lesser extent for audit and monitoring purposes. At the touch of a button, analyses can inform about the financial status of a department, hospital, region, or even a country. However, the healthcare sector has been slow to realise the true potential of this information. There are an exponential number of methods in which this data can be used to plan clinical, social, and auxiliary services, offer patient-tailored care, develop the arguments for health economics, and advance medical science.

As cardiologists, it is the enhancement of medical science and the opportunity to develop tailored patient services that interests us the most with big data. Large databases in the field of cardiology are not new; the Framingham study,[1] which followed generations of patients from Framingham, Massachusetts, USA, has led the way since the 1950s and a number of clinical risk models are based on this study.[2-4] Scandinavian countries have large registry data from the 1960s[5,6] which have led to numerous large-scale studies.[7] Work from Birmingham, UK, has led to vast improvements in the study of common conditions such as atrial fibrillation and development of clinically useful risk scores, such as CHA_2DS_2-VASc,[8] to determine the stroke risk of these patients and the use of anticoagulation in their treatment. The development of the National Institute for Cardiovascular Outcome Research (NICOR) and the British Cardiovascular Intervention Society (BCIS) datasets in the UK have led to vast improvements in our understanding of acute coronary syndrome and interventional cardiology.[9,10] In the USA, Medicare datasets have aided in large analyses related to cost-benefit ratios of cardiovascular procedures and related conditions.[11] These datasets have also informed us about the growth of revascularisation procedures such as percutaneous coronary intervention and coronary artery bypass grafting.[12]

Cardiology is an evidence-based speciality and clinical trials make up the backbone of clinical decision making and management.[13,14] However, in the face of increased scrutiny and regulation, the costs of these trials are rising steeply and have already run into billions of dollars. It is not uncommon to require a 20-year period for the development of a new treatment from its inception before it becomes mainstream, and the drawbacks and risks are large and very expensive. Therefore, with increasing technology and knowledge, it is disappointing to see the lack of significant major developments in the treatment of routine cardiovascular conditions, particularly when looking at medical therapy. Less expensive, registry-based clinical trials are being implemented.[15] In other cases, trials are designed with less significant endpoints and shorter follow-up periods to lower the cost. Moreover, none of these strategies hide the fact that patients in clinical trials are highly selected and in the face of an ageing population in the Western world, are less representative of real-world populations.

All the large datasets we have discussed so far in this article are collected data and/or registries ranging from hundreds of patients to many hundreds of thousands of patients. So we already have big data; why bother with more? Well, what about all that routinely collected data in the healthcare sector? The population of the UK is in excess of 60 million; the Western world, where such routine information has been collected over the last decade, has a population of over a billion. The developing world, which is not far behind in terms of data collection (namely Southeast and East Asia), has a population numbering multiple billions. Such enormous numbers pale the existing cardiology datasets, provided such datasets can be utilised, developed, and applied. The ACALM (Algorithm for Co-morbidities, Associations, Length of stay, and Mortality) study unit has been developed with such a dataset in mind. This algorithm differs from existing datasets because it utilises completely anonymous, routinely available healthcare information and transforms the data into fully functional, cross-sectional, and/or longitudinal research databases with real-life outcomes. The ACALM study unit is only 2 years old but has datasets of millions of patients and has already undertaken a number of studies addressing poorly researched areas, such as the interplay between mental and physical health in cardiology and factors influencing outcomes such as mortality and length of hospital stay.[16-18]

This is just the beginning. Datasets such as this one have the potential to not only answer questions that cannot be answered, but to revolutionise the way research is performed. It is the norm to perform literature reviews and to develop and test hypotheses with a carefully planned study. What if the research question is not known until the data is analysed? What if the data suggests information which we could not think of as plausible with our existing knowledge?

Complex modelling and further algorithms adapted from the world of computing, mathematics, and statistics can be used to enhance our knowledge and generate hypotheses for further research. Recent advances in both algorithm development and hardware infrastructure have paved the way for rapid adaptation of artificial intelligence and machine learning in medicine. Indeed, large-scale datasets generated in routine clinical settings can now be mined using sophisticated algorithms to accurately predict the onset of septic shock,[19] optimise patient-specific anticoagulation regimes,[20] and assess the prospective-risk of myocardial infarction.[21] Machine learning offers an opportunity to identify concepts rather than correlations in clinical data, thus promising to become an invaluable tool for data-aided decision making.

As with anything, there are limitations ranging from the quality of data collection, to the practicality of resourcing and running such large datasets, and other logistical and bureaucratic factors. However, the significant cost advantages of utilising routinely collected data, the sheer size of the datasets, and the fact that such data cannot otherwise be collected weigh heavily in favour of such research. We believe big data analytics will delineate a paradigm shift in cardiovascular medicine.

REFERENCES

1. Dawber TR et al. Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health Nations Health. 1951;41(3):279-81.

2. Freund KM et al. The health risks of smoking. The Framingham Study: 34 years of follow-up. Ann Epidemiol. 1993;3(4):417-24.

3. Benjamin EJ et al. Impact of atrial fibrillation on the risk of death: the Framingham Heart Study. Circulation. 1998;98(10):946-52.

4. Mahmood SS et al. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet. 2014;383(9921):999-1008.

5. Brandsaeter B et al.; Norwegian Heart Failure Registry. Gender differences among Norwegian patients with heart failure. Int J Cardiol. 2011;146(3):354-8.

6. Floderus B et al. Smoking and mortality: a 21-year follow-up based on the Swedish Twin Registry. Int J Epidemiol. 1988;17(2):332-40.

7. Nyboe J et al. Risk factors for acute myocardial infarction in Copenhagen. I: Hereditary, educational and socioeconomic factors. Copenhagen City Heart Study. Eur Heart J. 1989;10(10): 910-6.

8. Lip GYH et al. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. Chest. 2010;137(2):263-72.

9. Gale CP et al.; NICOR Executive. Engaging with the clinical data transparency initiative: a view from the National Institute for Cardiovascular Outcomes Research (NICOR). Heart. 2012;98(14):1040-3.

10. McAllister KS et al.; British Cardiovascular Intervention Society and the National Institute for Cardiovascular Outcomes Research. A contemporary risk model for predicting 30-day mortality following percutaneous coronary intervention in England and Wales. Int J Cardiol. 2016;210:125-32.

11. Johnston SS et al. Coronary artery bypass graft surgery in acute coronary syndrome: incidence, cost impact, and acute clopidogrel interruption. Hosp Pract (1995). 2012;40(1):15-23.

12. Fosbøl EL et al. Repeat coronary revascularization after coronary artery bypass surgery in older adults: the Society of Thoracic Surgeons' national experience, 1991-2007. Circulation. 2013;127(16):1656-63.

13. Scandinavian Simvastatin Survival Study Group. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). Lancet. 1994;344(8934):1383-9.

14. Yusuf S et al. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. N Engl J Med. 2000;342(3):145-53.

15. Fröbert O et al.; TASTE Trial. Thrombus aspiration during ST-segment elevation myocardial infarction. N Engl J Med. 2013;369(17):1587-97.

16. Carter P et al. The impact of psychiatric comorbidities on the length of hospital stay in patients with heart failure. Int J Cardiol. 2016;207:292-6.

17. Potluri R et al. The role of angioplasty in octogenarian patients with acute coronary syndrome. Int J Cardiol. 2016;202:430-2.

18. Potluri R et al. The role of angioplasty in patients with acute coronary syndrome and previous coronary artery bypass grafting. Int J Cardiol. 2014;176(3):760-3.

19. Henry KE et al. A targeted real-time early warning score (TREWScore) for septic shock. Sci Transl Med. 2015;7(299):299ra122.

20. Ghassemi MM et al. A data-driven approach to optimized medication dosing: a focus on heparin. Intensive Care Med. 2014;40(9):1332-9.

21. Zampetaki A et al. Prospective study on circulating MicroRNAs and risk of myocardial infarction. J Am Coll Cardiol. 2012;60(4):290-9.