Stefan Anker MD PhD FESC FHFA
Professor Innovative Clinical Trials
Department of Cardiology & Pneumology
University Medical Center Göttingen (UMG)
Göttingen, Germany
Corresponding author: s.anker@cachexia.de

Folkert Wouter Asselbergs,
Professor of Cardiovascular Genetics
Department of Cardiology,
Division Heart & Lungs
University Medical Centre Utrecht, Netherlands
Institute of Cardiovascular Science
Faculty of Population Health Sciences,
University College London

Gunnar Brobert PhD
Associate Professor
Director of Epidemiology
TA Head Thrombosis
Bayer AB, Solna, Sweden

Panos Vardas MD PhD
Professor of Cardiology,
University of Heraklion, Crete
ESC Past President
Chief Strategy Officer,
ESC Brussels office

Diederick E Grobbee MD PhD FESC
Professor of Clinical Epidemiology
Distinguished Professor of International
Health Sciences & Global Health
Chief Scientific Officer,
Julius Clinical Academic Research Organization
Chair, Center for Circulatory Health,
University Medical Center Utrecht
Corresponding author: grobbee@xs4all.nl

Dr Maureen Cronin
Consultant at Vifor Pharma

Big Data in Cardiology

Will big data lead to big improvements in cardiovascular care?

A 36-year-old woman presents with dyspnoea and dizziness. The clinician orders blood tests, an echocardiogram, an electrocardiogram, and documents the history in the electronic health record. A computer algorithm finds the patient's prior test results, her genetic profile and demographic characteristics, and links in her wearable biosensor data. The algorithm creates a unique phenotype by processing all of the data sources, compares it with one million other patients, and suggests to the provider that the patient has hypertrophic cardiomyopathy, with an 83% predicted probability of sudden cardiac death in the next 10 years. The information supports an accurate and efficient diagnosis and provides individualized risk to inform shared decision making for a potential implantable cardioverter defibrillator (ICD).

This scenario illustrates the promise of big data for cardiovascular clinicians—an automated system combining seemingly disparate data from various sources to provide decision support for diagnosis and treatment, as well as individualized, high-accuracy predictive analytics. How realistic is this future?

Big data analytics—and more generally the field of data science—are not new. Other industries have already capitalized on the explosion in data availability, computing power, and analytic methods like machine learning. Yet the 'big data era' in healthcare is just beginning.[1,2] Big data analytics support the concept of artificial intelligence and lie at the heart of many new digital health platforms and precision health tools. Ideally, utilization of big data analytic tools in cardiovascular care will translate into better care and outcomes at a lower cost. It is not yet clear, however, to what degree the promise will be fulfilled. In this brief article, we highlight three promising applications for big data in cardiovascular care, followed by 'proof of concept' challenges to be met if the promise of big data is to be realized.

The promise of predictive analytics, phenomapping, and precision health

The potential for more powerful predictive models is an appealing application of big data analytics.[1,2] Historically, prediction models have relied on a limited number of specified variables, manually entered to estimate a 'risk score'. Such models generally lack precision: they perform 'reasonably well' at the population level, but not at the individual patient level.[3] And despite the existence of dozens of risk models related to cardiovascular conditions, few are utilized to make therapeutic decisions.

Big data analytics may yield more powerful prediction of outcomes ranging from mortality to patient-reported outcomes to resource utilization, and thus could be more clinically actionable. Machine learning, for example, evaluates patterns associated with an outcome directly from the data, rather than from a pre-specified set of variables. A full range of associations and interactions among the data are assessed. Whereas traditional statistical models are 'one and done', machine learning uses a training process whereby the model is iteratively given varied data sets to explore many combinations of predictive features to optimize prediction.

A hallmark of big data is combining disparate data sources and types. Current primary sources are electronic health records and administrative (claims) data. But wider ranges of data inputs are increasingly available to develop more robust 'exposomes' for each patient.

For example, data from mobile health technology, biosensors, imaging, environmental data (e.g. air pollution), and information from social media networks, to name a few. In addition, '-omic' data (genomic, proteomic, metabolomic) will be increasingly available, potentially fuelling more accurate outcome predictions as well as more robust disease classification and individualized treatment recommendations.

Phenomapping, or deep phenotyping, is another promising big data application.[2] Current disease classifications, or phenotypes, are imprecise and heterogeneous. Take, for example, non-ischemic cardiomyopathy: treatment guidelines lump treatment interventions, despite substantial within-group heterogeneity. Some patients have peripartum cardiomyopathy, whereas other have alcohol-related or non-compaction cardiomyopathy; each experience a different disease trajectory. Clinicians are keenly aware that patients with the 'same' disease respond differently to treatment—in other words, substantial heterogeneity is present. Big data analytics can identify similar patient clusters, creating multiple phenotypes within each disease entity. In theory, more refined phenomapping of disease states and trajectories should help inform more tailored-health decisions.

Precision health is an important corollary of phenomapping. Patients and clinicians want to know if a specific patient is going to benefit (or be harmed) by an intervention. For example, guidelines for using ICD's rely on a crude measurement, left ventricular ejection fraction. The majority of patients who receive ICDs never receive a life-saving shock and some are harmed by inappropriate shocks. Big data methods can support the combination of multiple data sources from large patient populations to better estimate the potential benefits of therapies such as ICD's for individual patients. Indeed, big data analytic methods are central to the success of precision health, given the growing interest in incorporating '-omic' data, which vastly increases the size and complexity of datasets. Such datasets require advanced analytic platforms and methods that are the hallmarks of big data analytics.

Proof of concept challenges

The development, validation, and integration of big data predictive models, phenomapping, and precision health tools into cardiovascular care are at a nascent stage.[1,2] Despite proliferation of companies claiming to have big data 'solutions' that improve outcomes, it is hard to find published evidence of their impact or examples of successful integration into routine care. To that end, we propose the following 'proof of concept' challengers:

- Establish that big data models can have superior predictive power. Initial studies comparing big data methods to more traditional statistical methods and existing predictive models or risk scores suggest minimal or no significant incremental predictive benefit.[2]

- Show that big data tools can provide 'actionable' insights. Limited studies of big data predictive tools largely reinforce that older, sicker, more complex patients have worse outcomes and have

higher utilization of resources. Also, big data methods emphasize associations without consideration of causation, yet causal associations are often critical to inform medical decisions. Big data predictive models might, therefore, increase predictive power but provide no actionable insights to guide care decisions. Phenomapping studies do not yet support that novel phenotypes should be treated differently. And initial studies of precision medicine genetic markers have raised concern about accuracy and reproducibility; this does not support their readiness for clinical deployment.[4]

- Demonstrate that big data 'solutions' are valid and stable over time when deployed. Most existing publications are initial validation studies using retrospective data. Before big data tools are used in routine care or health management, prospective evaluation of their validity and stability over time—even if they are constantly updated based on new data—is crucial. Outside of healthcare, less accurate or stable models may be acceptable, such as to guide consumer spending or entertainment choices. The stakes are much higher for health decisions. In addition, big data methods are generally tolerant of poor underlying data, especially where 'all of the data' is available. However, 'all of the data' are not available in healthcare. Missing healthcare data are often informative, and there may be treatment selection bias in existing data. Underlying data quality may be essential for big data analytics in healthcare compared to other sectors.[2] Finally, many big data analytic companies use proprietary, 'black box', modelling approaches, which raise scepticism about validity and stability if they are to be used to inform care: proof of clinical utility will be essential.

- Prove that big data solutions improve care efficiency and outcomes. The initial development of big data tools isn't sufficient to claim their effectiveness or cost effectiveness (since there is cost associated with the big data solutions). Evidence is needed that they can be successfully integrated into care, leading to more efficient and/or superior care outcomes, while avoiding unintended consequences. This evidence is lacking to date.

In an era of exponential growth in technology, healthcare will change. But even as care moves away from episodic (e.g. clinic-based) to longitudinal, remote care supported by technology, human connection will remain at the centre of health care. The clinician–patient interface will still drive most health decisions. Big data carries the promise that these decisions may be informed by more powerful predictive analytics, better phenomapping of patients, and precision health tools that guide individualized care.

However, hype without evidence is a threat to fulfilment of the promise of big data for cardiovascular care. Proof of concept must be evident. And while big data analytics are novel for cardiovascular care, successful integration into care harkens the challenges—and many examples of failure—of clinical decision support. Big data solutions will need to be coupled with successful strategies for clinical workflow change in order to succeed. At this point in the 'era of big data' in healthcare, there needs to be a shift in focus from what big data *might* do for cardiovascular care to proving what it *can* do, including careful planning of how to integrate these tools into evolving care models and demonstrating impact.

Rashmee U. Shah MD MS
Assistant Professor of Medicine
University of Utah School of Medicine
30 N. 1900 E. Room 4A100
Salt Lake City, UT 84132, USA
801-585-7676
rashmee.shah@utah.edu

Conflict of interest: none declared.

John S. Rumsfeld MD PhD
Chief Innovation Officer
American College of Cardiology
Washington D.C., USA
Professor of Medicine,
University of Colorado School of Medicine
Aurora, CO, USA
jrumsfeld@acc.org

References

Fractional Flow Reserve_CT (FFR_CT)

FFR_CT for the diagnosis of suspected ischaemic heart disease, reported by Dr Tim Fairbairn a principal investigator in the ADVANCE Registry

Introduction

The UK National Institute for Health and Clinical Excellence (NICE) has recently made two significant decisions that may influence the role of fractional flow reserve_CT (FFR_CT) in the diagnosis and management of suspected coronary artery disease.

The first was a clinical guidance (CG95) recommending coronary computed tomography angiography (CCTA) as the first line diagnostic strategy in the majority of suspected cardiac chest pains. Risk stratification was removed with other non-invasive tests (stress magnetic resonance imaging, stress echocardiography, and myocardial perfusion imaging) remaining the preferred option in individuals with previously known coronary artery disease. Invasive coronary angiography (ICA) was reserved as a third line test.

The second decision was a medical technologies approval (MTG32) for the use of HeartFlow CT Fractional Flow Reserve (FFR_CT, HeartFlow Inc, Redwood city, CA, USA) in a chest pain diagnostic pathway.

Coronary computed tomography angiography for diagnosing coronary artery disease

Coronary computed tomography angiography over the past decade has become an increasingly utilized tool for diagnosing coronary artery disease. Its high sensitivity (93%) and negative predictive value (>95%) have made it particularly useful in low-intermediate risk populations.[1] Limitations to CCTA include technical factors that affect image quality (heart rate, arrhythmias, partial volume averaging) and radiation dose. However, the major limitation is the lack of a functional ischaemic component to this anatomical test. The ability to identify coronary plaque and degree of stenoses by CCTA and ICA whilst excellent, does not correlate well to myocardial ischaemia.[2]

Functional or anatomical imaging

The importance of ischaemia to guide revascularization in stable ischaemic heart disease is contentious. COURAGE (Clinical Outcomes Utilizing Revascularization and Aggressive Drug Evaluation), STITCH (surgical Treatment for Ischaemic Heart Failure), and BARI 2D (Bypass Angioplasty Revascularization 2 Diabetes) indicated no prognostic benefit of revascularisation vs. medical therapy.

However, FAME (fractional flow reserve vs. angiography in multivessel evaluation)[3] showed ischaemia-driven therapy to be superior to anatomical guided therapy and several non-invasive studies have reported the importance of a high burden of ischaemia.[4,5] International guidelines still recommend ischaemia driven therapy, so revascularizing an angiographically significant stenosis of uncertain functional significance remains controversial.[6,7]

Fractional flow reserve_CT functional assessment

Computed tomography assessment of plaque burden and characteristics (spotty calcification, positive arterial remodelling, necrotic core) is excellent and can help predict the likely functional significance of a lesion.[8] However, the specificity of CT to determine functional significance is disappointingly low, as a significant proportion of cases remain inconclusive.[9]

The ability to combine a functional and anatomic non-invasive test is therefore highly desirable. Computed tomography perfusion can potentially address this, but exposes the patient to additional radiation and requires complex acquisition techniques. HeartFlow have used fluid dynamics computational models to calculate coronary blood flow and derive a FFR_CT from a 'standard' CTA image. This has