Cardio-oncology
The Role of Big Data

Anant Mandawat, MD[a], Andrew E. Williams, PhD[b], Sanjeev A. Francis, MD[c,*]

KEYWORDS

• Big data • Cardio-oncology • Pharmacovigilant

KEY POINTS

- Despite its challenges, a "big data" approach offers a unique opportunity within the field of cardio-oncology.
- A pharmacovigilant approach using large data sets can help characterize cardiovascular toxicities of the rapidly expanding armamentarium of targeted therapies.
- Creating a broad coalition of data sharing can provide insights into the incidence of cardiotoxicity and stimulate research into the underlying mechanisms.
- Population health necessitates the use of big data and can help inform public health interventions to prevent both cancer and cardiovascular disease.
- As a relatively new discipline, Cardio-Oncology is poised to take advantage of big data.

Big data is like teenage sex; everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it.[1]

—Dan Ariely, Duke University Professor of Behavior Economics

Patients with cancer are living longer due to improvements in treatment.[2] As cancer survivors live longer, they are at risk of complications of heart disease in part due to effects of cardiotoxicity from chemotherapy and radiotherapy. Cardio-oncology is the discipline that focuses on the intersection of cancer and cardiovascular disease.[3] A major facet of this intersection is chemotherapy-related or radiotherapy-related heart disease.[2,4] Cancer therapy may result in damage to the heart from a direct toxic effect of the therapy or indirectly by worsening cardiac risk factors like hypertension.[2] However, at present, there is variation in screening and treatment practices.[5] For a relatively young discipline like cardio-oncology, research and discovery are essential for growth and understanding. In particular, there is a call to study the direct and indirect effects of chemotherapy and radiation on heart function. This call includes strategies for primary and second prevention, identification of surrogate markers, guidelines for screening, and risk prediction models.[2,6] More information on the early and late effects of systemic treatments and mechanisms of injury is needed. The focus of this review is to describe the potential of big data to achieve these goals in cardio-oncology.

Cardio-oncology is an exciting new field primed for the application of big data. In 2015, President Barack Obama announced the Precision Medicine Initiative, and in 2016, Vice-President Joe Biden declared the Cancer Moonshot, which will allocate more than $1 billion in advancing patient level care and oncology. Importantly, these initiatives will direct significant resources that could be used toward the collection and application of big data in cardiology and oncology.[7,8] Research in

[a] Division of Cardiovascular Medicine, Duke University School of Medicine, 2301 Erwin Road, Durham, NC 27710, USA; [b] Division of Health Services Research, Center for Outcomes Research and Evaluation, Maine Medical Center Research Institute, 509 Forest Avenue, Suite 200, Portland, MW 04101, USA; [c] Division of Cardiovascular Medicine, Maine Medical Center, 22 Bramhall Street, Portland, ME 04102, USA
* Corresponding author.
E-mail address: SFrancis@mmc.org

Heart Failure Clin 13 (2017) 403–408
http://dx.doi.org/10.1016/j.hfc.2016.12.010

cardio-oncology is a unique opportunity for both clinical and research collaboration between the cardiology and oncology.[9]

Big data has been an elusive term to define, as characterized in Prof Dan Ariely's opening quote.

It is estimated that there is 2.3 trillion gigabytes of data created each day. As of 2011, the global size of data is health care is estimated to be 161 billion gigabytes (IBM).[10] A framework for defining big data has been developed using the 4 "V's": volume, variety, veracity, and velocity (IBM). Importantly, big data is not just about large amounts of data; rather, it is characterized by data heterogeneity, validity, and its high speed of new inputs. Big data in health care includes insurance claims data, public health data, electronic health records, genomic data, mobile health data, and social health data. The goal of big data is to leverage information from such sources to meaningfully inform patients, physicians, and researchers. Examples can be found in the priorities of the new Cancer Moonshot Task Force to support data sharing, analytical collaborations, and public-private partnerships between the Veterans Administration and IBM Watson.[8]

The European Organization for Research and Treatment of Cancer has advocated for a systematic collection of data in cardio-oncology.[2] Perhaps more than any other discipline within cardiology, cardio-oncology would benefit from leveraging multiple sources of data. Statistically, the absolute event rate of chemotherapy-induced cardiomyopathy may be relatively low depending on the specific drug and patient population being studied[11] (rate of cardiac events between 1.3 [control] and 4.0% [trastuzumab arm] at 7 years follow-up of 1830 patients). Big data presents an opportunity to detect relatively low-frequency events that nonetheless can have significant clinical impact. Aleman and colleagues[2] support an international registry for cardio-oncology. Unlike acute coronary syndrome, cardio-oncology is a relatively young discipline in which new fundamental pathophysiology and phenotypes continue to be understood, and whose identification and surveillance will require large amounts of data. In addition, chemotherapy-induced cardiomyopathy is a complex interaction between a patient's demographics, genetics, comorbidities, medications, laboratory tests, imaging, and functional status, requiring integration of multiple data sources, including electronic health records and genomics, for research on risk assessment and treatment.

The current infrastructure in cardio-oncology is limited. Recent initiatives are designed to streamline development of new cancer therapeutics at an accelerated rate.[8] However, cancer therapeutics receives approval based on randomized clinical trials with cardiac events prone to variable ascertainment and adjudication. In addition, patients enrolled in randomized trials may represent the "healthier" patients with cancer who are younger with a lower incidence of cardiovascular disease or risk. The incidence of cardiotoxicity observed in clinical trials may not be an accurate reflection of the risk in typical clinical practice.

There are several examples of using big data within cardiology. Patel and colleagues[12] used data from CathPCI, a registry part of the American College of Cardiology National Cardiovascular Data Registry, to analyze almost 400,000 patients from 663 hospitals undergoing elective coronary angiography. They showed that almost two-thirds of patients undergoing the procedure had nonobstructive disease, encouraging the need for better risk stratification before angiography. Speaking to the power of big data, the CathPCI registry from its inception in 1998 to 2013 contained data on more than 12 million records from 1577 participating US centers.[13] This year, the *International Journal of Medical Journal Editors* released a proposal advocating for release of deidentified data within 6 months of trial publication.[14] Recent high-profile editorials have supported this push for transparency and open data access of clinical trial data,[15–17] an important source of data for toxicities of new cancer therapeutics.

Chen and colleagues[18] and Ezaz and colleagues[19] are 2 examples of using big data in cardio-oncology. Both studies used the Surveillance, Epidemiology, and End Results–Medicare data. Chen and colleagues[18] calculated incidence rates of heart failure or cardiomyopathy over 3 years in 45,000 women aged 67 to 94 with breast cancer who received anthracycline, trastuzumab, a combination of both, or neither. It leveraged big data to elucidate incidence rates that otherwise would be difficult to assess due to low event rates. Ezaz and colleagues[19] analyzed data on 1664 women with breast cancer receiving adjuvant trastuzumab and created a risk score for the development of heart failure and cardiomyopathy. Unlike other disciplines that have to use existing tools, which at times are outdated, cardio-oncology has the luxury to design its own research infrastructure from the ground up.

Cardio-oncology would benefit from several initiatives that could contribute to the big data research mission. First, a clinically meaningful and consistent definition of cardio-toxicity is needed. The most adopted definition is the Cardiac Review and Evaluation Committee of trastuzumab-associated cardiotoxicity definition