

# Big Data and Visual Analytics in Health and Medicine: From Pipe Dream to Reality

Noreen Kamal<sup>1\*</sup>, Samuel Wiebe<sup>2</sup>, Jordan DT Engbers<sup>3</sup> and Michael D Hill<sup>4</sup>

<sup>1</sup>Department of Clinical Neurosciences, Faculty of Medicine, University of Calgary, Calgary, AB Canada

<sup>2</sup>Department Clinical Neurosciences, Associate Dean (Clinical Research), Faculty of Medicine, University of Calgary, Calgary, AB Canada

<sup>3</sup>Clinical Research Unit, Faculty of Medicine, University of Calgary, Calgary, AB Canada

<sup>4</sup>Department Clinical Neurosciences, Associate Dean (Clinical Trials), Faculty of Medicine, University of Calgary, Calgary, AB Canada

## Introduction

The increasing move to electronic data platforms for the management of health information is creating an untapped resource with the power to change health care. These Big Data open up possibilities for better quality health care, and better and faster clinical research. A key problem with Big Data is understanding the information quickly, so by applying a Visual Analytics approach, the initially overwhelming scale of Big Data becomes a valuable asset. Interactive visualization permits everyone to understand large, multi-source, variable-type, and time varying data.

Big Data have been defined as large datasets from a variety of sources and data types (such as numeric, text-based and imaging data) that cannot be managed or processed using standard software tools within a reasonable time [1]. Practically, it refers to datasets of a sufficient volume, variety, and velocity so as to allow discovery of new insights or new forms of value that could not be found in smaller datasets [2,3]. Visual Analytics supports Big Data by providing interactive visualizations that allow people to navigate these datasets. Visual Analytics has been defined as “the science of analytical reasoning facilitated by interactive visual interfaces [4].” Visual Analytics is more than just visualization of the data; it is an approach that combines visualization, human factors and data analysis [5]. Big Data are more powerful when combined with Visual Analytics, as the synthesized information that Big Data provides can be understood more quickly. This increases utility of Big Data for decision-making processes that need to be made with limited time in real-time. Although there is time for sober reflection after clinical decisions are made, quality cares provided through correct clinical decisions are influenced by information in real-time.

Both of these emerging fields will expand current discussion on Big Data in health, which have described the potential to generate new clinical knowledge by analyzing unstructured text-based data, assisting with knowledge dissemination by using data-driven decision support tools, translating personalized medicine (e.g. genomics) into clinical practice, and delivering information directly to patients [6].

Both Big Data and Visual Analytics can contribute significantly to improving the quality of health care. A key tenet in quality improvement (QI) is to measure the process where improvement is sought, which can often be a single measure such as time for a patient to receive a specific treatment [7]. However, each data point contains a large amount of supporting information (i.e. a wide dataset), which can include: patient demographic information; patient co-morbidities; time of day that the treatment happened; the specific diagnoses; text-based physician notes; the physician(s) and healthcare professionals that were involved in the care; the prescriptions that the patient was taking; imaging data etc. This supporting information can account for a proportion of the variability in the measure that is being improved. Through interactive visualization of the measure, a QI consultant or analyst can explore the factors associated with this variability and gain a deeper understanding of possible causality. For example, patients with two or more co-

morbidities may have delays in treatment due to extended times needed to ensure a proper diagnosis, or patients that arrive for treatment during off-hours may have delays in treatment because of availability of certain specialists. By applying Big Data and Visual Analytics research, various individuals can explore this wide dataset to understand factors influencing quality of care. Generally, these disciplines can allow for: 1) the development of better QI methods; 2) implementation of more meaningful system improvements by taking out the guess work in the Model for Improvement’s plan-do-study-act iterative cycle [7] (or six-sigma’s define-measure-analyze-improve-control process) [8]; and 3) the ability to engineer sustainable improvements achieving tightly controlled excellence in health care despite variation in patients and care settings.

Similarly, Big Data and Visual Analytics can significantly contribute to clinical research by using de-identified health data and pragmatic clinical trials [9]. Randomized Controlled Trials (RCTs) provide the highest level of evidence when implementing clinical research into clinical practice [10]. However, the cost of running RCTs is often prohibitive, and new, more pragmatic solutions to traditional RCTs are being discussed, which include randomized registry trials [11] and pragmatic RCTs [12]. By using a randomized registry approach, a more pragmatic, generalizable and cost effective approach to clinical research can be undertaken. The randomized registry trial as a backbone offers the potential to quickly recruit patients, test interventions, and translate successes into clinical best practice for acute stroke care. Novel software solutions can help to facilitate consent and randomization into clinical trials by working with researchers to better understand the time constraints, ethical concerns, and privacy issues. Furthermore, the electronic visualizations of the data can ensure appropriate anonymization required by the trial protocols. Visual Analytics can provide the ability to dynamically interact with the data and seamlessly provide ways to move between meaningful overviews and patient specific information as needed by the researchers.

Despite the potential, creation of the one-source-of-truth registry file in a real-time and dynamic manner holds many challenges. Additionally, there are challenges to using these large datasets to improve the quality of health care and providing cost effective ways

**\*Corresponding author:** Noreen Kamal, Department of Clinical Neurosciences, Faculty of Medicine, University of Calgary, AB Canada, Tel: +1-403-210-8895; E-mail: [nrkamal@ucalgary.ca](mailto:nrkamal@ucalgary.ca)

**Received** February 04, 2014; **Accepted** February 05, 2014; **Published** February 12, 2014

**Citation:** Kamal N, Wiebe S, Engbers JDT, Hill MD (2014) Big Data and Visual Analytics in Health and Medicine: From Pipe Dream to Reality. J Health Med Informat 5: e125. doi:10.4172/2157-7420.1000e125

**Copyright:** © 2014 Kamal N, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

of running clinical research. Principally, the data often resides across multiple data sources, which create difficulties in having fully linked big data. Furthermore, it is often difficult to understand data quality and fidelity, which may lead to misleading results. By creating secure linked health information from all necessary sources, we can start to learn about the causes of poor quality and high variation in our complex health system [13] and be able to create solutions for the entire system [14].

Despite these challenges, there are current examples of Big Data solutions that are being used for QI in healthcare and for clinical research, which provide promise of moving forward. The Get With the Guideline Registry (GWTG) is a very large American registry created by the American Heart Association/American Stroke Association involving over one thousand sites [15]. This registry has shown good data quality [16], and spawned important QI research in stroke [17], coronary heart disease [18], and heart failure [19]. This example shows how large datasets from diverse organizations can be developed when there is a will, leading to improved outcomes for patients. On the clinical research side, the randomized registry trial has been successfully applied in Scandinavia with 5000 patients [20]. They used an existing Swedish Coronary Angiography and Angioplasty registry to obtain all baseline data, where numerous data points were collected directly in the catheterization laboratory via a web interface. The extra work of randomization was done through a small randomization module within the registry. This type of registry trial can allow for reduction in the cost to run an RCT, while proving to be more efficient and providing better data quality by having a centralized repository. Similarly, the Resuscitation Outcomes Consortium has set-up a registry and data collection system in routine paramedic care for their pre-hospital ED trials.

The new world of Big Data is changing the way collaboration occurs in health research. Multi-institution, industry, multi-centre and international collaborations often come together to tap into the enormous potential of Big Data in health. Several groups are working toward this vision across Canada. "Open data" and "data sharing" models are being increasingly developed. In the province of Alberta, for example, initiatives are afoot to develop province-wide, consolidated health data repositories and high-performance, cloud and agile computing platforms; additionally expertise is being developed to explore specific health questions using Big Data and Visual Analytics. Similarly, groups are working to create a central clinical and imaging registry for stroke. Alberta is in a unique position in Canada, as it has a single health authority and single health data custodian for the entire province. A clinical registry that is able to securely fill data from the electronic medical records and link to the imaging registry to obtain a full record for each stroke patient will potentially provide a backbone for quality improvement and randomized registry trials that are running across the province. Examples in other Canadian provinces include: the Southern Ontario Smart Computing Innovation Platform, a multi-university partnership with a large focus on health; and the Atlantic University Partnership on Analytic Skills, an initiative focusing largely on skills development for the novel field of Big Data and Visual Analytics.

Big Data combined with Visual Analytics provide promise for improving the delivery of health services and changing how clinical research is performed. Despite the challenges of residing across multiple data sources and unknown data quality, there are examples of Big Data registries for improving health outcomes and clinical trials. Furthermore, Canadian examples provide a glimpse into the future of

Big Data and Visual Analytics to improve health outcomes and develop better and cheaper ways of running clinical trials.

## References

1. Diebold FX (2012) A Personal Perspective on the Origin(s) and Development of 'Big Data': The Phenomenon, the Term, and the Discipline, Second Version. PIER Working Paper No 13-003.
2. Mayer-Schönberger V, Cukier K (2013) Big data: A revolution that will transform how we live, work, and think. 2013: Houghton Mifflin Harcourt.
3. Laney D (2001) 3D Data Management: Controlling Data Volume, Velocity and Variety Gartner.
4. Thomas JJ, Cook KA (2006) A Visual Analytics Agenda. *IEEE Computer Graphics and Applications.* 10-13.
5. Keim DA, Mansmann F, Schneidewind J, Ziemer H (2006) Challenges in Visual Data Analysis. Tenth International Conference on Information Visualization, 9-16.
6. Murdoch TB, Detsky AS (2013) The inevitable application of big data in health care. *JAMA* 309: 1351-1352.
7. Berwick DM (1996) A primer on leading the improvement of systems. *BMJ* 312: 619-622.
8. De Feo JA, Barnard W (2005) JURAN Institute's Six Sigma Breakthrough and Beyond-Quality Performance Breakthrough Methods. Tata McGraw-Hill Publishing Company Limited.
9. Larson EB (2013) Building trust in the power of "big data" research to serve the public good. *JAMA* 309: 2443-2444.
10. Oxman AD (2004) GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 328: 1490-1494.
11. Lauer MS, D'Agostino RB (2013) The randomized registry trial—the next disruptive technology in clinical research? *N Engl J Med* 369: 1579-1581.
12. Chalkidou K, Tunis S, Whicher D, Fowler R, Zwarenstein M (2012) The role for pragmatic randomized controlled trials (pRCTs) in comparative effectiveness research. *Clinical Trials* 9: 436-446.
13. Stermann JD (2006) Learning from evidence in a complex world. *Am J Public Health* 96: 505-514.
14. Smith M, Halvorson G, Kaplan G (2012) What's needed is a healthcare system that learns. *JAMA* 308: 1637-1638.
15. American Heart Association (2014) Get With the Guidelines—HF/Stroke/Resus.
16. Xian Y, Fanarow GC, Reeves MJ, Webb LE, Blevins J, et al. (2012) Data quality in the American Heart Association Get With the Guidelines—Stroke (GWTG-Stroke): results from a national data validation audit. *Am Heart J* 163: 392-398.
17. Cumber E, Wald H, Bhatt DL, Cox M, Xian Y, et al. Quality of stroke and outcomes for in-hospital ischemic stroke: findings from the national get with the guidelines-stroke. *Stroke* 45: 231-238.
18. Thukkani AK, Fonarow GC, Cannon CP, Cox M, Hernandez AF, et al. (2014) Quality of Care for Patients With Acute Coronary Syndromes as a Function of Hospital Revascularization Capability: Insights From Get With The Guidelines-CAD. *Clin Cardiol.*
19. Allen LA, Fonarow GC, Grau-Sepulveda MV, Hernandez AF, Peterson PN, et al. (2014) Hospital Variation in Intravenous Inotrope Use for Patients Hospitalized with Heart Failure: Insights from Get With The Guidelines. *Circ Heart Fail.*
20. Fröbert O, Lagerqvist B, Gudnason T, Thuesen L, Svensson R, et al. (2010) Thrombus Aspiration in ST-Elevation myocardial infarction in Scandinavia (TASTE trial). A multicenter, prospective, randomized, controlled clinical registry trial based on the Swedish angiography and angioplasty registry (SCAAR) platform. Study design and rationale. *Am Heart J* 160: 1042-1048.