Information technology innovation: the power and perils of big data

The global health-care system is consistently under tremendous pressure to lower health-care costs, maintain high efficiency and quality of care, and remain up-to-date technologically in an era of instantaneous information exchange. In the UK, around 8.4% of the gross domestic product is spent on health care (approximately 0.19 trillion GBP). In the USA, this number is 17.9% of gross domestic product, or 2.7 trillion USD. With the introduction of health-care reform and a shift in payment structure to pay-for-performance, further pressure has been placed on the health-care system to reduce costs and increase health-care quality. Additionally, a shift in patient characteristics to an ageing population and improved access to care have increased the number of patients seeking care. Compounding the situation is a shortage of key practitioners, including nursing staff, in the medical workforce.

As a result of staff shortage and external pressure and regulations from government agencies to reduce costs, the health-care system must find a way to improve the quality of patient care for more patients with fewer resources. With these difficulties in mind, and the additional challenges that lie ahead, the health-care system, including anaesthetists, must continue to use innovative medical technologies and become more efficient in the collection and analysis of this information to drive cost-effective clinical practice.

As discussed in the article by Simpao and colleagues, technological advancements in health care have led to an explosion in data collection, increasing storage and analysis needs. In 2011, there were 1.8 zettabytes of data created globally. In the same year, it was estimated that data from the US health-care system reached 150 exabytes. This number will continue to grow to reach zettabyte ($10^{21}$ gigabytes) followed by yottabyte ($10^{24}$ gigabytes) levels over time. Data of this magnitude are known as ‘big data,’ defined as electronic data sets so large and complex that they are difficult or impossible to manage with traditional software, hardware, or both; nor can they be easily managed with traditional or common data-management tools and methods. There are three primary characteristics of big data: volume (the amount of data generated by organizations, individuals, or machines), variety (data in all forms; structured, unstructured, and semi-structured), and velocity (the speed of data generation, delivery, or processing). The creation of these massive data sets with varying formats is a result of the proliferation of electronic health records (EHRs). The EHRs have vastly improved the maintenance of health information and have promoted the collection and sharing of information among providers across all health-care disciplines, leading to a more collaborative approach to patient care. In the field of anaesthesia, the EHRs, also known as Anaesthesia Information Management Systems (AIMS) or Anaesthesia Information Systems (AIS), have decreased inaccuracies, incompleteness, biases, and inherent errors. However, implementation of these EHRs has created data sets that can be difficult to analyse for quality control or research purposes. In one study of AIMS, event recording dependent on user input can have a low sensitivity (38%), leading to under-reporting of key clinical events,

There are several systems that have been developed to overcome structural and analytical data issues. The most popular system at this time is the open-source distributed data processing platform, Hadoop (Apache platform). Initially, Hadoop was developed as a platform to aggregate Web search indexes. Using numerous servers, known as nodes, Hadoop has the potential to store and process extremely large amounts of data by allocating partitioned data sets to each node. An analysis request in Hadoop (a Map Reduce request) is allocated to each node and each data set, and executed at the data level in parallel (the Map process), and the results are integrated and aggregated for the final result (the Reduce process).

Hadoop has the ability to analyse unstructured, semi-structured, and structured data. As a data-protection method, Hadoop maintains redundant data sets in different nodes to protect the data and analyses from system crashes. If a node becomes unusable, an additional node will be used to continue the requested analyses. Therefore, Hadoop is structured to serve dual roles, namely the ability to store massive amounts of data
efficiently and to analyse these data. Being open source, Hadoop has great benefits to the medical community and health-care researchers. However, there are several drawbacks to this technology. Although Hadoop is open source and free of charge, the lack of technical support and concerns with security must be addressed. Furthermore, the expertise required to use this system is often greater than most health-care institutions can provide at this time.

Even though there are challenges ahead with the use of big data, significant gains have been reported. The global use of big data in health care has many benefits to be realized, including the detection of diseases at early stages, enhancements in disease management, and detection of health-care fraud. It has been estimated that the U.S. health-care system could save more than $300 billion per year by using big data analytics. These savings come from the reduction of waste and inefficiencies in clinical operations, research and development, public health, evidence-based medicine, genomic analytics, pre-adjudication fraud analysis, device or remote monitoring, and patient profile analytics. To realize these benefits, the health-care system must begin to use the massive amounts of collected data appropriately.

Providing pre-, intra-, and postoperative care to patients, anaesthetists have access to large amounts of patient data and should learn from previous efforts to use big data effectively. The ability to analyse these data efficiently can assist in the provision of high-quality care and have a significant impact on clinical practice by allowing early identification of adverse or clinically relevant events before their occurrence. Use of sophisticated vital sign monitoring equipment allows collection of a plethora of data points in second or even millisecond intervals, resulting in billions of data points to analyse. Additionally, with proliferation of EHR systems and the support of government agencies around the world, the creation of national and regional registries has become more prevalent. In the field of perioperative outcomes research, there are numerous databases available for use, including the National Surgical Quality Program (NSQIP), the Society of Thoracic Surgeons Database, the American Society of Anaesthesiology Closed Claims Registry of settled malpractice cases, the Multicenter Perioperative Outcomes Group (MPOG), the Anesthesia Quality Institute’s National Outcomes Registry (NACOR), and multiple international databases through the Healthcare Quality Improvement Partnership (HQIP) and the Health Services Research Centre and National Audit Projects (NAP), to name a few. Furthermore, there are numerous large billing databases and national surveys available to conduct health outcomes and resource utilization studies, including the Center for Medicare and Medicaid Services (CMS), the Healthcare Cost and Utilization Project (HCU) maintained by the Agency for Healthcare Research and Quality (AHRQ), and various national health surveys. Anaesthetists should take advantage of these national registries and local EHR data to help drive improvement in clinical care. Although retrospective studies will not replace or provide the level of evidence of the randomized control trial, researchers should look to these retrospective, big data studies as initial evidence to drive clinical investigations. Thus, the use of big data analytical tools enables hypothesis generation and exploration of massive data sets in a meaningful way.

Beyond clinical research, anaesthetists can use big data to enhance operational efficiencies, resource utilization, and patient safety. Big data analytical tools have been used to monitor and reduce drug-drug interactions, evaluate perioperative transfusion practices, and identify risk factors associated with laryngeal mask airway failure in children. Finally, big data can assist departments of anaesthesia in identifying billing anomalies. Currently, most departments find these errors through the time-consuming and error-based approach of complicated algorithms and manual audits on a sample of data. However, using advanced analytical approaches, one health-care system identified up to 2% of out-patient revenue that was previously unbilled and reduced audit expenses by as much as 75%.

Although big data offers significant gains for anaesthetists and the health-care system, there are still significant challenges to overcome. As previously mentioned, the skill set necessary to use big data analytics is specialized and is not widely available in the health-care system at the present time. In order for these tools to become more widely used, they must become more user friendly. In addition, analytical tools must provide real-time information to care providers in order to make a significant impact on clinical practice, quality, and operations. Finally, although analytical systems have the ability to work with and provide data cleaning of unstructured data, the present state of EHR data collection will require immense data cleaning, and the findings of these studies should be interpreted with caution.

Data security is a significant concern. It is the duty of health-care providers and institutions to ensure the security of patient information; however, as greater volumes of data are collected, concern about large-scale data breaches increases. The question remains as to whether patient informed consent is necessary to conduct research using data collected for an alternative purpose. As the use of big data intensifies, these issues will become more heavily debated, and solutions will need to be approached with caution.

In summary, big data is currently underused in health care despite the significant advantages it presents. Simpao and colleagues’ highlight the potential uses of big data, visual analytic techniques available, and challenges to its use in anaesthesia. Given the role of the anaesthetist in maintaining patient safety and the vast availability of data, departments of anaesthesia should be on the front-line of big data analytics. The ability to analyse data efficiently has serious implications for the care and treatments provided by anaesthetists, particularly in the operating room, where a patient’s condition can change from minute to minute and the ability to process data quickly facilitates identification of pathological states and early treatment. In this context, clinicians are provided with large amounts of data to evaluate a patient’s condition on several monitor screens in order to identify potential problems. The time for a physician to process these data leads to diagnostic delays, potentially causing harm to the patient. Using big data analytics, researchers can analyse billions of data points collected during the perioperative period to identify patients at risk for adverse intraoperative and postoperative events, and through development of early warning systems, provide the clinician with proactive, rather than reactive, support in care. Supported by emerging technologies, the health-care system, especially the data-intensive specialty of anaesthesia, must continue to evolve with the ever-changing patient population and economic pressures. Innovation has become and needs to remain a critical capability of all health-care organizations, and anaesthetists are poised to lead in the application of big data to acute care.

Declaration of interest
Hugh C. Hemmings, MD is an Editor of the British Journal of Anaesthesia.
References


G. P. Giambrole1, H. C. Hemmings1, M. Sturm2, and P. M. Fleischut1
1Department of Anesthesiology, Weill Cornell Medical College, 525 East 68th Street, Box 124, New York, NY 10065, USA
2Department of Information Services, NewYork-Presbyterian Hospital, 333 East 38th Street, New York, NY 10065, USA
*Corresponding author. E-mail: pmf9003@med.cornell.edu, pfleischut@gmail.com