Big data: increasing productivity while reducing costs in health and social care

Guest Editorial

Pavanraj Jessal[1]

What is big data?

In recent years there has been widespread interest in the business and technology press surrounding a data-driven revolution, which is being brought about through the exponential growth of data that are being collected, stored and transferred. This vast amount of data is currently referred to as 'big data' and includes data that are continually collected through devices and technologies, such as credit cards and customer loyalty cards, the internet and social media.

In healthcare, big data include clinical data (clinicians' notes and prescriptions, medical imaging results, and laboratory, pharmacy and other administrative data); patient data in Hospital Episodes Statistics (HES), machine generated or sensor data, such as from monitoring vital signs; and articles in medical journals. At the same time, new advanced analytical techniques are allowing practitioners to connect and interrogate datasets that were once separate. By finding links and understanding patterns and trends in the data, big data analytics has the potential to improve care, increase efficiency and lower costs.

Potentially big data analytics can lead to better outcomes across many different areas of healthcare. Examples of this include:

- analysing patient characteristics and the cost and outcomes of care, to identify the most clinical- and cost-effective treatments
- applying advanced analytics to patient profiles (for example, segmentation and predictive modelling) to proactively identify people who would benefit from preventative care or lifestyle changes
- undertaking large-scale disease profiling to identify predictive events and support prevention initiatives
- identifying, predicting and minimising fraud by implementing advanced analytic systems for fraud detection (Raghupathi et al., 2014).

Volume, velocity and variety

Gartner, the US information technology research specialist and consultancy, first developed a model for big data. Its '3V' model encompassed 'volume, velocity and variety'. Gartner states that large amounts of data become big data when they meet three criteria: volume, variety and velocity. Gartner formalised its definition in 2012: 'big data are high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization' (Beyer, 2012).

Volume: Already huge quantities of data exist in the healthcare system, such as HES data, patient records and prescribing information. Over time, even more data will be created and accumulated, leading to an enormous volume of data. This is perhaps the most immediate challenge of big data, as it requires advances in data management. Virtualisation and cloud computing are facilitating the development of platforms for more effective capture, storage and manipulation of these large volumes of data.

Variety: Variety refers to a collection of many types of data, both structured and unstructured, including multimedia, social media and financial transactions, GPS tracking information, audio and video streams, and web content. Although standard techniques and technologies exist to deal with large volumes of structured data, it becomes a significant challenge to analyse and process a large amount of highly variable data and turn it into actionable information. The ability to perform real-time analytics against high-volume data across all specialties would improve healthcare.

Velocity: Data are being collected in real time and at a rapid pace or velocity. Traditionally in health and social care, data tended to be collected periodically; big data are processed and analysed in real or near-real time. This is advantageous in health and social care for areas such as clinical decision support, where access to up-to-date information is essential for correct and timely decision-making and elimination of errors.

[1] For further information, contact Pavanraj Jessal at the National Institute for Health and Care Excellence (NICE): Pavanraj.Jessal@nice.org.uk.

Some practitioners and researchers have introduced a fourth characteristic of big data: '*veracity*' (or data assurance). This means the big data, analytics and outcomes are error-free and credible. This is the aim but not yet the reality. If decisions are going to be made using this data, then it is imperative that the data are high quality and accurate.

Traditionally, most data collected appeared in a structured and semi-structured format but, increasingly, unstructured data are being collected. Structured data describes data that are grouped into a relational scheme (for example, rows and columns in a standard database). Because of the data's configuration and consistency, it can be queried simply to arrive at usable information, based on an organisation's parameters and need. Semi-structured data refer to data that may have some relational structure but which are incomplete or irregular. Unstructured data describe data of all formats that cannot easily be indexed into relational tables for analysis or querying. Examples include text and web pages, social network content and blog posts, images, audio and video (CEBR, 2012).

Of all of the sectors in the economy, healthcare is one of the leading generators of administrative data. During each episode of a patient's care, huge amounts of information can be generated, ranging from demographics to symptoms, observations and investigations, through to diagnoses, treatments, procedures and outcomes. Increasingly, this information is being recorded and stored electronically rather than on paper; however, only a tiny fraction of this information is currently fed into central flows of data where the information is collated and used for wider patient benefits.

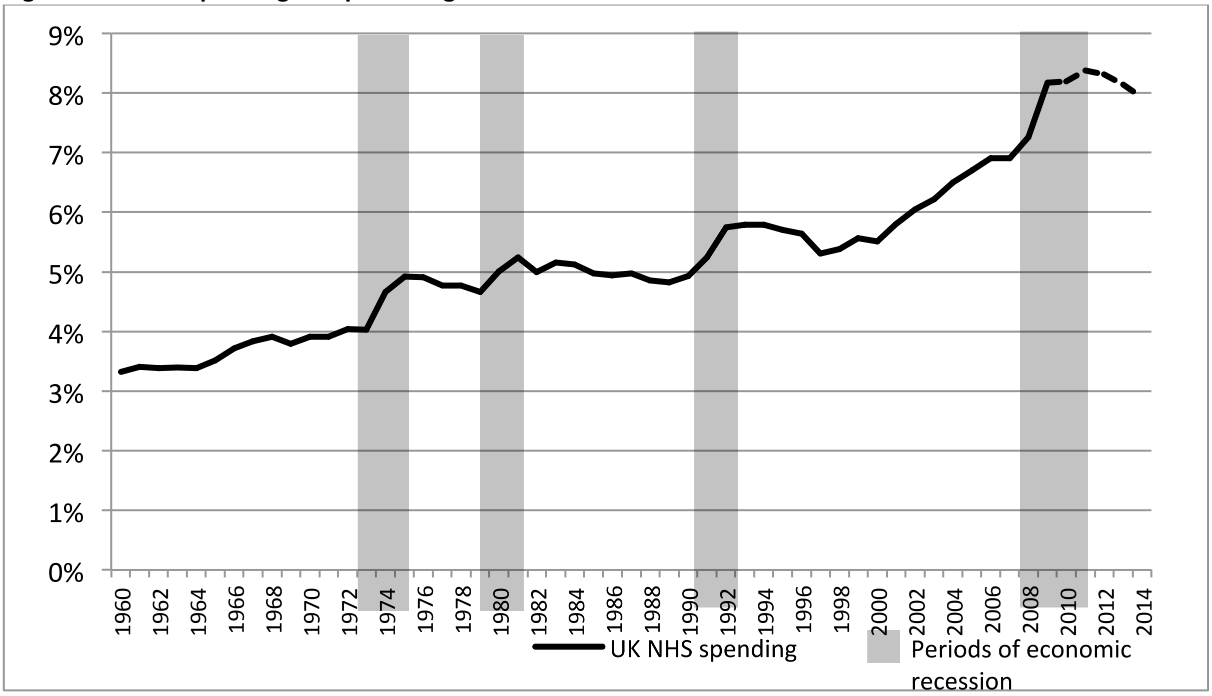
The health and social care sector is in an excellent position to try and harness the potential of big data. Already the NHS has HES, which contains information about all admissions, outpatient appointments and A&E attendances at NHS hospitals in England. Across these three care settings, a number of common sections are recorded, such as patient identity, activity levels and clinical diagnoses. HES contains information about every hospital admission that has occurred since 1989, covering the whole population of England and spanning every hospital provider in the country. Over the coming years, HES will transform into the Care Episodes Service (CES), which will not only include a far richer hospital dataset but will also be expanded to include all other care settings including primary care, mental health, clinical audit and social care data. Examples of some of the new datasets which will be included are patient ward and theatre details, pharmacy data, pathology data and patient feedback information. This will provide a much more detailed picture of the patient and any variation in healthcare provided.

By expanding the amount of data collected from different care settings, it is hoped that these data can then be used to improve integration between separate services. One of the biggest perceived benefits of more integration is improved pathways of care, for example through reduced duplication, or provision of services in different locations. The next article in this volume, '*Cost of Integrated* Care' (Bardsley & Street) gives some examples of how this could be achieved, explaining the importance of population data linkage methods so that costs can be looked at for individuals over time, not just for particular interventions or treatments.

Rationale for analysing big data

Fifty years ago the UK's NHS consumed around 3.4 per cent of gross domestic product (GDP) (Organisation for Economic Co-operation and Development, 2012). Now, public spending on the NHS is nearly two-and-a-half times greater – amounting to 8.2 per cent of GDP and equivalent to seven times more in real terms (Appleby, 2013).

Figure 1 UK NHS spending as a percentage of GDP



Source: Organisation for Economic Co-operation and Development (2012): author estimates

The historic tendency has been for healthcare spending to grow, and the income elasticity of demand for healthcare tends to be above one, with increases in national income (GDP) leading to proportionately higher increases in healthcare spending.

Expenditure on healthcare is increased to obtain more in terms of volume and quality. However, increased and improved outputs and outcomes do not necessarily need more inputs. Increasing the productivity of each pound spent on health and social care would also produce better outcomes. The Office for Budget Responsibility's (OBR) projection taking health spending to 16.6 per cent of GDP by 2061/2, for example, assumes annual productivity gains in the NHS of just 0.8 per year (Office for Budget Responsibility, 2012). Higher productivity would reduce the need to spend more, while maintaining improvements in volume and quality. Experts are commenting that these productivity gains could be informed through the analysis of big data and big data analytics.

Accurate costing can contribute to the efficient allocation of resources in the health and social care system, and help identify where cost reduction is feasible and justifiable. Conversely, misleading or absent cost data can lead to unfair comparisons and flawed policy choices. Developing unit costs of health and social care can take considerable time and effort, and the precision of the cost will vary depending on the data and information used to calculate the cost. 'The least precise estimates are likely to be based on average per diems (or daily costs); the most precise estimates are likely to be based on micro-costing' (Drummond et al., 2005). With more data available across the health and social care sector, the data available to produce accurate and timely micro-costing will increase and, as long as the veracity of this data is strong, the unit costs developed will become more precise.

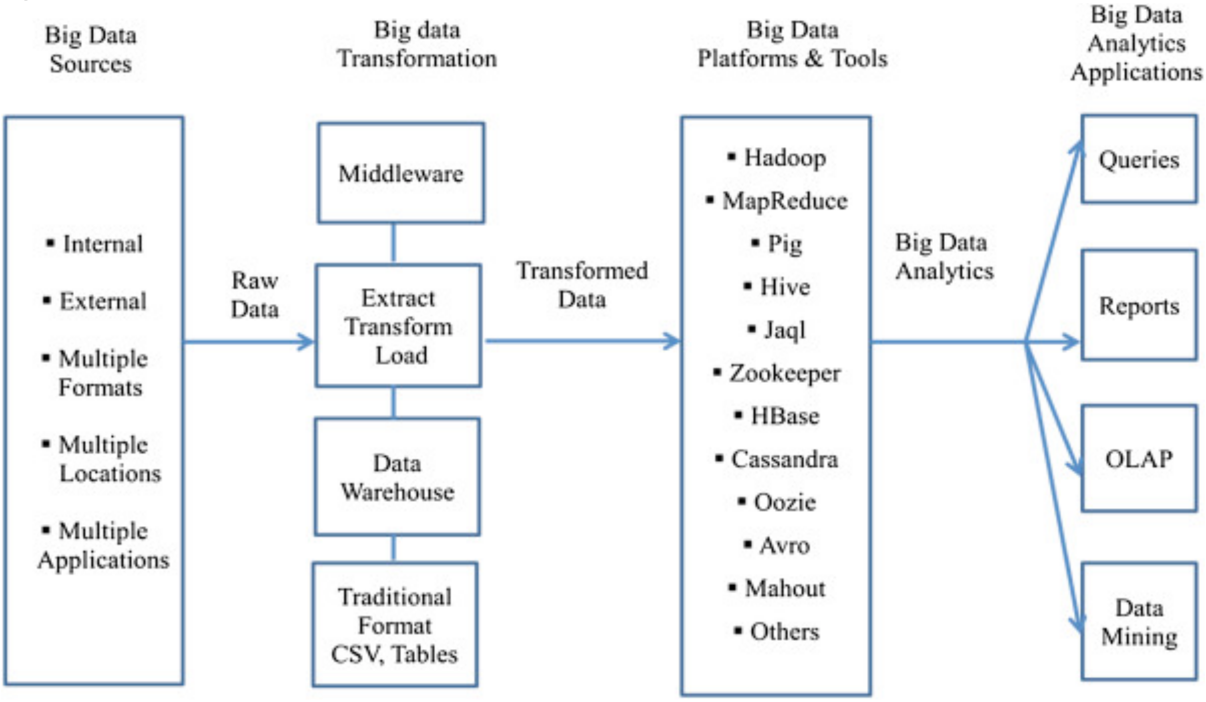
Architectural framework

The conceptual framework for a big data analytics project in healthcare is similar to that of a traditional health informatics or analytics project. The main difference lies in how big data need to be broken down and processed across multiple nodes. This can make big data analytics tools extremely complex, programming intensive and require the application of specialist skills.

Big data can come from internal or external sources in multiple formats and from multiple locations. For the purpose of big data analytics, these data need to be brought together (see Figure 2, Big data sources). In the second component (Figure 2, Big data transformation), the raw data need to be processed or transformed, and several options are available to do this. A service-oriented architectural approach combined with web services (middleware) is one possibility (Raghupathi et al., 2007). The data stay raw, and services are used to call, retrieve and process the data. Another approach is data warehousing, in which data from numerous sources are collected and prepared for processing, although the data are not available in real time. Through the steps of extraction, transformation and loading (ETL), data from various sources are cleaned and made ready. Several data formats can be input to the big data analytics platform.

In this next component in the conceptual framework (Figure 2, Big data platforms and tools), several decisions need to be made about the data input approach, distributed design, tool selection and analytics models. Finally, these three components lead to the four typical applications of big data analytics in healthcare (Figure 2, Big data analytics and applications). These include queries, reports, online analytical processing (OLAP) and data mining. Visualisation is an overarching theme across the four applications. Drawing from such fields as statistics, computer science, applied mathematics and economics, a wide variety of techniques and technologies has been developed and adapted to collect, manipulate, analyse and visualise big data in healthcare (Raghupathi et al., 2014).

Figure 2



Source: Raghupathi & Raghupathi (2014) – *Health Information Science and Systems*

Examples of how big data have been used to increase efficiency

Staff performance information can be dynamically monitored and forecast through predictive analytic tools, allowing departments to link strategic objectives with service-user outcomes. The use of predictive key performance indicators (KPIs), balance scorecards and dashboards in health and social care can bring operational benefits, provided that the required data are fully accessible to operations managers (CEBR, 2012). Also, a detailed understanding of how much time staff are spending on tasks and how well they are performing these tasks will help to produce more robust unit costs of labour time. For example, it could improve the accuracy of estimates made about the ratio of face-to-face contact a nurse has with a patient.

A wider set of patient data to analyse allows healthcare providers to accurately apply the latest findings of medical research; thus being able to efficiently prevent complications and new disease developments. Clinical decision support systems can compare patient information with research literature and medical guidelines, highlighting potential errors such as adverse drug reactions and enhancing the efficiency and quality of care. For example, chest pain can result in approximately 100 different diagnoses; decision support systems that can narrow down the alternatives can still leave the final decision to the physician, but will greatly speed up the process. The dependability and the comprehensiveness of support offered to healthcare providers will be further improved as these solutions develop and include other capabilities, such as image analysis (Piai & Claps, 2013).

Personalised medicine and evidence-based practices will also integrate more cohesively with chronic disease management programmes. Big data platforms enable healthcare providers to better control information coming from remote patient monitoring (RPM) systems checking patient adherence to prescriptions and to improve future treatment options and reduce complications. By effectively using information from RPM systems, healthcare providers will be able to reduce inpatient stays, limit emergency department visits, and improve the effectiveness of homecare and outpatient appointments. By applying advanced analytics such as segmentation and predictive modelling to patient profiles, healthcare providers can identify anomalies and find patients who are at high risk of developing a specific disease or complication and would benefit from a preventive care programme or a disease management programme (Piai & Claps, 2013).

Analysing datasets from the patient pathway as well as other administrative processes allows health and social care provider managers to identify inconsistencies, bottlenecks and misuse of resources. By mapping processes, healthcare providers will have greater visibility of areas where operations need to be streamlined. A leaner process will reduce costs, release unused resources and deliver a better service. For example, Nice University Hospital in France uses a radiofrequency identification (RFID)-based system for managing approximately 57,000 biological samples in the hospital's biobank. Previously, the hospital relied on a paper-based traceability process that was time-consuming and error-prone, and could result in lost samples and compromised security. The analysis of datasets and the integration of the workflows of pathology and the biobank centre has led to a saving of more than 50 per cent in time, increased traceability and timely delivery of biospecimen samples (Piai & Claps, 2013).

Conclusion

Big data and big data analytics are expected to transform health and social care. Organisations need to collect and store data, start to gather and share information, and implement strategies developed from the information. Health and social care organisations will need to invest in new technologies and new ways of working, and it will be important to recruit the right talent and develop the right culture. There will be challenges such as issues around privacy and confidentiality, data veracity and contextual meaningfulness, information assurance and organisational change. However, in order to make the improvements in the quality demanded from the health and social care services, while providing efficiency savings, the potential of innovative ideas such as big data and big data analytics will need to be fully explored.

References

Appleby, J. (2013) Spending on health and social care over the next 50 years: Why think long term? http://www.kingsfund.org.uk/sites/files/kf/field/field_publication_file/Spending%20on%20health%20...%2050%20years%20low%20res%20for%20web.pdf [accessed 4 June, 2014].

Beyer, M. & Laney, D. (2012) *The importance of 'big data': a definition*, http://www.gartner.com/id=2057415 [accessed 10 June 2014].

Centre for Economics and Business Research (2012). *Data equity unlocking the value of big data*, http://www.sas.com/offices/europe/uk/downloads/data-equity-cebr.pdf [accessed 10 June 2014].

Curtis, L. & Burns, A. (2014) *Unit costs of health and social care*, University of Kent, Canterbury.

Drummond, M., Schulper, M. & Torrence, G. (2005) *Methods for the economic evaluation of health care programmes*, Oxford University Press, Oxford.

Office for Budget Responsibility (2012) Office for budget responsibility: economic and fiscal outlook. Report. Cm 8303, The Stationery Office, Norwich.
http://budgetresponsibility.independent.gov.uk/wordpress/docs/March-2012-EFO1.pdf [accessed 10 June 2014].

Organisation for Economic Co-operation and Development (2012*) OECD Health Data 2012 – frequently requested data*, http:/www.oecd.org/els/healthpoliciesanddata/oecdhealthdata2012-frequentlyrequesteddata.htm [accessed 10 June 2014].

Piai, S. & Claps, M. (2013) *Bigger data for better healthcare*, http://www.intel.com/content/dam/www/public/us/en/documents/white-papers/bigger-data-better-healthcare-idc-insights-white-paper.pdf [accessed 10 June 2014].

Raghupathi, W. & Kesh, S. (2007) *Interoperable electronic health records design: towards a service-oriented architecture*, https://muse.jhu.edu/login?auth=0&type=summary&url=/journals/eservice_journal/v005/5.3raghupathi.html [accessed 10 June 2014].

Raghupathi, W. & Raghupathi, V. (2014) *Big data analytics in healthcare: promise and potential*, http://www.hissjournal.com/content/2/1/3 [accessed 10 June 2014].