2014

Big data in healthcare: What is it used for?

Rebecca Hermon
Edith Cowan University, rhermon0@our.ecu.edu.au

Patricia A H Williams
Edith Cowan University, trish.williams@ecu.edu.au

DOI: 10.4225/75/57982b9431b48
Originally published in the Proceedings of the 3rd Australian eHealth Informatics and Security Conference. Held on the 1-3 December, 2014 at Edith Cowan University, Joondalup Campus, Perth, Western Australia.

BIG DATA IN HEALTHCARE: WHAT IS IT USED FOR?

Rebecca Hermon[1] Patricia A H Williams[2]
[1,2]eHealth Research Group, School of Computer and Security Science
Edith Cowan University, Perth, Western Australia
[1]Rhermon0@gmail.com.au, [2]trish.williams@ecu.edu.au

Abstract

Big data analytics is a growth area with the potential to provide useful insight in healthcare. Whilst many dimensions of big data still present issues in its use and adoption, such as managing the volume, variety, velocity, veracity, and value, the accuracy, integrity, and semantic interpretation are of greater concern in clinical application. However, such challenges have not deterred the use and exploration of big data as an evidence source in healthcare. This drives the need to investigate healthcare information to control and reduce the burgeoning cost of healthcare, as well as to seek evidence to improve patient outcomes. Whilst there are a number of well-publicised examples of the use of big data in health, such as Google Flu and HealthMap, there is no general classification of its uses to date. This study used a systemic review methodology to create a categorisation of big data use in healthcare. The results indicate that the natural classification is not clinical application based, rather it falls into four broad categories: administration and delivery, clinical decision support (with a sub category of clinical information), consumer behaviour, and support services. Further, the results demonstrate that the use of big data in all examples in the literature is not singular in its approach and each study covers multiple use and application areas. This study provides a baseline to assess the proliferation of the use of big data in healthcare and can assist in the understanding the breadth of big data applications.

Keywords
Big Data, Healthcare, Big Data Analytics, Literature Review, Systematic Review.

INTRODUCTION

According to a 2013 Commonwealth of Australia report, about 90% of data today was created in the last 2 years. It has been calculated that the production of data will be 44 times greater in 2020 than it was in 2009. Other calculations suggest that data is being created at 2.5 quintillion bytes a day. For the purpose of this study, the following definition has been adopted: "high-volume, high velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight, decision making, and process optimization" (Commonwealth of Australia, 2013, p. 8). This definition recognises three dimensions of big data. However, there are arguably two additional challenges: veracity and value.

Big data is recognised as a multidisciplinary information processing system. Areas of business, government, media, and in particular healthcare, are increasingly incorporating big data into information processing systems. To make effective use of the potential of big data in healthcare, an understanding of what the 2.5 quintillion bytes of data consists of, where they reside, are they raw, processed or derived artefacts, and what the delineation between public and private access is required. What is missing currently is any answers to these questions. A first step to answering such a question, is to construct a meaningful picture using categorisation of the areas of current use.

This initial review attempts to provide foundation for understanding the use of big data in healthcare, with a view to explore how big data can be applied to particular areas to gain the maximum benefit for the targeted research. Big data analytics relates to healthcare as an option for solving information system complexities within healthcare. Although the five dimensions of big data are categorised separately, in fact, they intertwine.

- Volume (scale of data): This is the management of the amount of data, usually referred to in terms of terabytes or petabytes of data. It involves management of data storage (Feldman, 2012).
- Variety (different forms of data): The format of data can be structured, semi structured and unstructured (Feldman, 2012)
- Velocity: The frequency of data that is produced, processed, and analysed (Feldman, 2012).
- Veracity (uncertainty of data): The quality, relevance, predictive value and meaning of data (Clifford, 2008).

- Value: The worth of information to various stakeholders/ decision makers (Clifford, 2008).

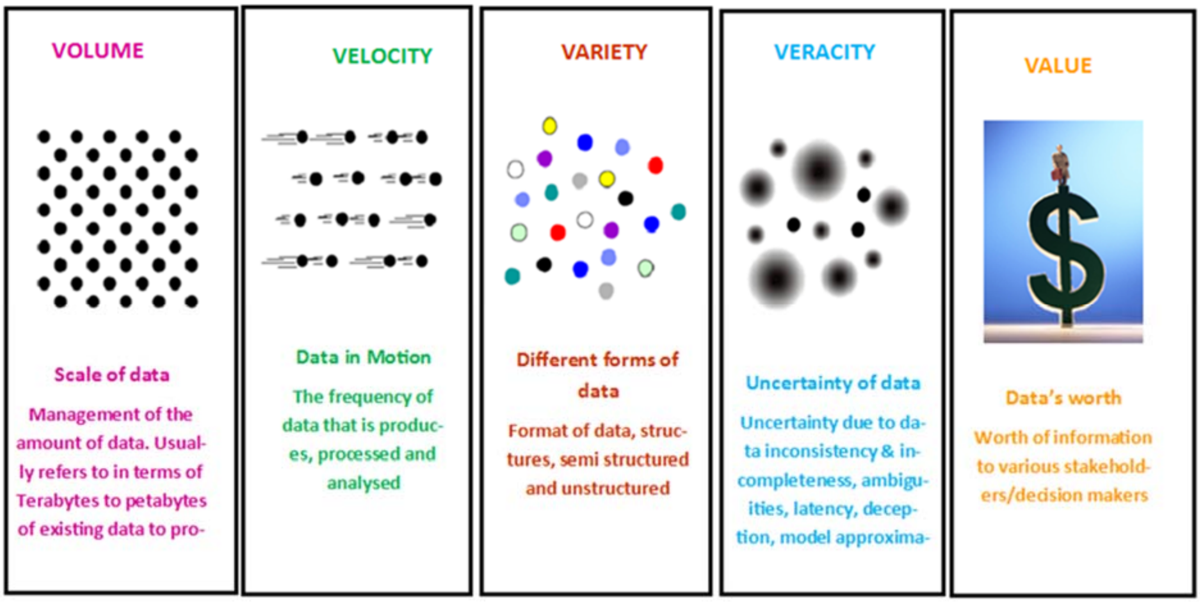

Figure 1: The five dimensions of big data (adapted from Haas, 2013)

These five dimensions play an important role in the issues facing big data analytics. Big data analytics involves the analysis of high volume data that stems from a variety of sources, including structured and unstructured data (Commonwealth of Australia, 2013). It also refers to the volume of the data sets within the data analysis and the velocity in which they are analysed.

In recent years, the concept of big data has been introduced to the healthcare system as a solution to a variety of healthcare related information system problems as health systems grown increasingly complex and expensive (Sun, 2013). Estimates suggest that in 2012, healthcare data reached about 500 petabytes. While future estimates suggest that by 2020 healthcare data will equal 25,000 petabytes. Effective integration of such data using data mining and medical informatics may result in lower costs and improved patient care via well informed decision making (Sun, 2013).

Big data is the intelligence for Electronic Health Records (EHRs), as it has the ability to connect financial, operational, and clinical analytic systems, and may support evidence-based healthcare. Evidence based healthcare involves the systematic reviewing of previous clinical data in order to provide decision makers with information. Evidence suggests that big data can be used to detect disease and in particular assist in clinical genomic analysis of HIV patients (Feldman, 2012). However, big data requires careful data management in order to fulfil the goals of big data analytics (Commonwealth of Australia, 2013). This includes data governance of the data sources, data content, data quality, data consistency, data access and security, user training, and data stewardship (Shaw, 2013). Data related issues can arise without effective management and governance, and these include unreliable, inaccessible, missing, or inaccurate data (Shaw, 2013). Therefore, data management aims to provide authentic, accurate, and reliable data essential for good decision making in healthcare. Consequently, big data initiatives must consider the various aspects of big data processes such capturing, storing, searching, sharing, and analysing of data (Commonwealth of Australia, 2013).

Whilst tackling these issues is important, it is also vital to examine the landscape of trial use and leading applications in big data in order to investigate the practical application and potential of this new phenomenon. To date, this type of examination is not evident in the literature. The first step in making effective big data use more prominent is to identify and categorise the current innovative uses. Hence, this paper describes a systematic review of big data to provide a categorisation of the uses of big data in healthcare. The research process and an explanation of the choice of articles selected for the review is presented. Subsequently, an analysis of the articles and the resultant mapping of the categories are described.

METHODOLOGY

The purpose of a systematic review is to summarize the available research on a specific subject, by sourcing evidence from other studies (Clarke, 2011). A rigorous systematic review must follow a predetermined plan as with any research endeavour. It is a structured method to analyse the literature on a specific subject that, like any primary research, is documented such that it could be replicated. It involves formulation of a review/research question, scoping the evidence base and devising appropriate keywords, conducting a comprehensive search of the literature, devising inclusion and exclusion criteria, synthesising the results from individual studies, analysis, and reporting the findings (JBIEBNM, 2000).

Research/Review Question Formulation

The first step in the methodology defined an appropriate question. The primary research question was "Can big data be categorised into distinct utilisation categories to improve semantic misinterpretation, leading to a better performance of big data analytics within healthcare". Sub-questions to support this question are:

- What is big data used for in healthcare?
- How is big data use categorised?

Scoping the Evidence Base, Devising Appropriate Keywords and Selection Criteria

An initial literature review was conducted to identify potential literature sources, terminology, and keywords related to big data use in healthcare. The scope of the research was initially limited to publications that within the last 14 years. This was due to the relative recent evolution of big data and health analytics. Therefore, publications that were less than 3 years old had higher value than those published 14 or more years ago. High quality research conducted by reliable sources was deemed valid for use. Literature from internet websites such as blogs was not classified as high research quality material. Appropriate literature was identified by searching in online academic databases, including such as JAMIA, Wiley Online and IEEE. Conference papers, journal articles, books, research papers, fact sheets, government reviews and standards, as well statistics were also included within the initial research phase. Peer reviewed literature was deemed reliable and literature that did not fit the appropriate criteria was excluded.

The second step was to further define the initial search terms and search appropriate literature that met the criteria. The initial search of the literature resulted in 30 articles selected as potentially suitable for inclusion in the study. From initial literature review search terms were then created to conduct a full literature search methodology. Each article was categorised into search terms (Table 1) and database source with number of articles found (Figure 2).

Table 1: Number of publications located using final search terms

Search Terms Used	Number of Publications
Big data and clinical decision support	7
Big data and health care delivery	6
Big data and health services	4
Big data and health care admin	0
Big data consumers	4
Big data patients	2
Big data behaviour	3
Big data clinical data management	4
Big data health care data management	1
Big data and Electronic Health Records	4
Big data and public health	5
Total number of publications	40

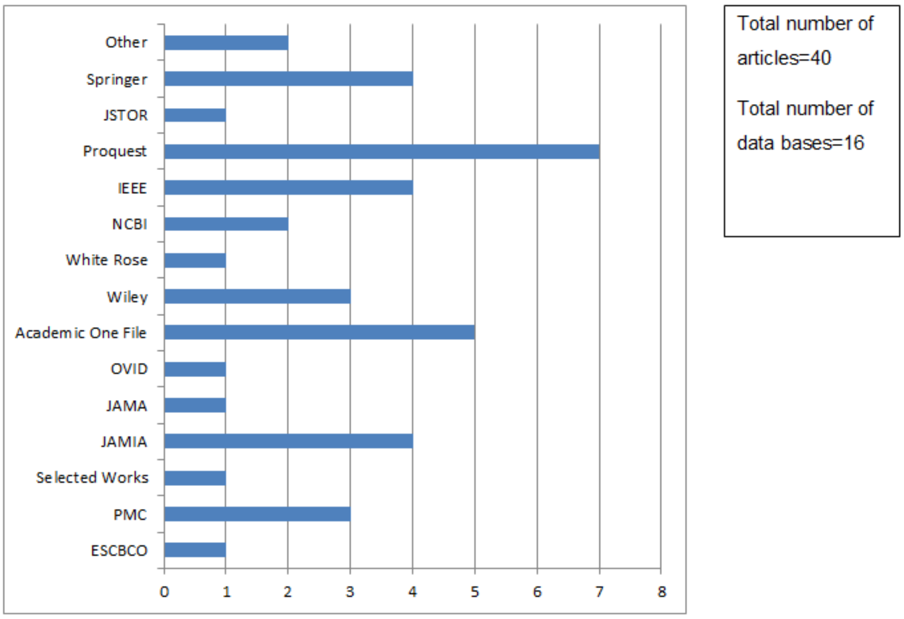

Figure 2 Databases and number of articles identified

The search of the literature using the above terms resulted in 40 articles. No relevant articles were found by using the search term big data and healthcare admin. This is not to say that there is no research concerning big data and healthcare admin, however, it does suggest that there are limited search results from using that particular research term. Five out of the 40 articles came from JAMIA while seven came from ProQuest. Big data within healthcare is a new and growing field, which encompasses articles that rarely categorise the use of big data within healthcare. Through inspection of where the articles were found, assessment of what is currently being published regarding big data use in healthcare can be identified.

Synthesising the Results from Individual Studies

Step 3 involved combing the results of the literature search. The literature collected was sorted from high quality useful material to low quality research literature. High quality material refers to articles found from peer reviewed journal articles that contained relevant information. By reading and analysing the literature it was possible to identify those articles relevant to the research question. Given the diversity in context, and in order to allow preliminary analysis, four broad categories below, as suggested by Groves (2013), were used to provide an initial categorisation:

1. Administration and delivery: managing healthcare delivery costs : Administration and delivery refers to the management of healthcare delivery. Big data analytics can be shown to utilise numerous activities that can be categorised into admin and delivery. Therefore, where the use influences the management of healthcare delivery, it is placed in this category.

2. Clinical decision support: Big data can be used to support clinical decision making. Various components that allow clinicians to gain additional information to aid in the decision making process are categorised within clinical decision support.

3. Clinical information: This category represents the information and data sets that are specifically available for big data analytics. The information systems that hold the clinical information as well as the various types of clinical information systems are evident within this category.

4. Behaviour/ consumer: Within the behaviour/consumer category, big data uses are recognised for the purpose of demographical analysis. Behaviour and lifestyle factors of both individuals and the public are also represented here.

As the analysis progressed, it was realised that a fifth category was required for articles and annotations of usage that did not fit into one of the primary broad Groves categories. This category was 'support information'.

5. Support Information: This category refers to big data uses that cannot be distinctly categorised into one of the other four categories. Either its use surpasses one category and fits into several or has been shown to fit into its own category of use.

ANALYSIS AND RESULTS

The articles revealed numerous potential usage applications. Table 2 presents the multiple utilisations based on the extended general categories, as extrapolated from Groves (2013).

Table 2: Analysis of articles using Groves (2013) extended categories

Groves Extended Categories	Article Reference	Type of Information Located
Administration and delivery: managing healthcare delivery costs	Murdoch & Detsky, 2013	*Electronic Health Records, decision making, observational based evidence, patient data, trails, case studies, quantitative data, real time patient data, genomics demographics, lifestyle information from other sources, healthcare deliver, economic values, medication lists, family history, practitioner data, knowledge dissemination, direct information to patients*
	Harper, 2013	*Predict public health, patient care, and population health, Clinical decision support: automated algorithms, electronic health record secondary data. business data, economic data, billing data, financial data, reduced healthcare costs, education*
	Rose & Burgin, 2014	*Monitoring of evidences based guideline for prevention screening , treatment plan, Electronic Health Records -population data, patient specific information, patient reported data, physician collected information, medication records, patient behaviour, reduce costs, predictive analysis-risk accuracy, medication adherence*
	Monheit & Berk, 2001	*Health care costs, health insurance*
	Sheldon, 1998	*Organizational healthcare, assurance healthcare, state medicine, quality indicators,*
	Ferlie & Shortell, 2001	*Track population, disease registry, clinical information -outcome data, evidence based, financial data, process data, legal data, education, academic, guidelines*
Clinical decision support	Adler-Milstein & Jha, 2013	*Early detection of treatment, lower costs, Clinical decision support, Electronic Health records-Bio surveillance, Accurate predictions of sick, health behaviours*
	Bates et al., 2003	*Health information systems, Evidence based medicine, clinical decision support, Electronic Health Records*
	Weber et al., 2009	*Ontology, health observations, demographics, Clinical data, health research information, patient information*
	Thadani, Weng, Bigger, Ennever, & Wajngurt, 2009	*Electronic Health Record, clinical trials, semantic dictionary, Predictive analysis, E-screening, demographics*
	Swinglehurst, Pierce, & Fuller, 2001	*Decision making, GP evidence based medicine, primary care delivery, clinical information, Local health, community health*
	Peleg & Tu, 2006	*Clinical data , standard terminology, healthcare delivery, patient care, reducing costs-prevention of medication errors, educational institutions*
	Cooper, Kennedy, & Springer, 2007	*Health care management-healthcare services, laboratory results, healthcare databases*
	Chen, Mao, & Liu, 2014	*Public health, biomedical data, R and D, genomics, clinical gene diagnosis fraud claims, error investigation, tax claims*
	Cushing, 2013	*Visual analytics, research*
	Howe et al., 2008	*Biological data, research, semantics*
Clinical information	Liu & Park, 2014	*External health data (social media), Electronic Health Record, mobilised Health Record, health monitors, genetic sequencing, biomedical sensors*
	Doug et al., 2008)	*Confidential data-machine learning-clinical trials, epidemiological database, research*
	Schouten, 2013	*Predictive modelling, anomaly detection, pattern detection*
	Sanat, 2013)	*Real time consumer sentiment, social interaction, financial data, predictive diagnostic, warrant y analysis, genomic research*
	Chawla & Davis, 2013	*Decision making, disease risk profiles, disease prevention, disease management, clinical data, population based evidence healthcare, individual based evidence healthcare, genomics, lifestyle and environmental behaviours to contribute to factors*
	Kuriyan & Cobb, 2013	*Reduce costs, insurance claims, health claims data, demographics: social determinants-health population, prescription data-forecasting chronic disease*
	Eramo, 2013	*Electronic health record clinical data, disease registry data*
	Skolnik & Bertman, 2013	*Electronic health records, population chronic disease management, physician behaviour, treatment management, clinical decision support, evidence based medicine, diagnosis and therapeutic interventions, Electronic health record data analytics*
	Kevin, 2008	*Electronic health records*
	Bonney, 2013	*Healthcare reform, electronic health record, population healthcare*
	Ackerman, 2012	*Genomics disease treatment, outcomes*
	Green, 2006	*Clinical care, disease epidemiology, social behaviour*
	Diehr & Lumley, 2002	*Predictive public health, public health*

Behaviour/ consumer	Ringel & Skiera, 2014	*Consumers, competition*
	Kaye, 2013	*Marketing, consumer behaviour*
	Wang & Strong, 1996	*Advertising, customers, behaviour*
	Timothy & Susan, 2002	*Health care consumer, demographics, external healthcare tracking*
	Kallus, 2014	*Predictive analysis within the population*
	Bentley, O'Brien, & Brock, 2014	*Decision making, health behaviour, social phenomenon, public health, research*
	Yoon, Park, Kim, ChaeTae & JunHyun, 2014	*Behaviour of individuals and public*
Support Information	Gross & Bates, 2007	*Clinical trials, clinical information, clinical laboratory, cost effective, detection of errors, medications, clinical knowledge for education-clinical decision support, evidence based medicine*
	Ola & Sedig, 2014	*Decision making, policy visual analytics, analytic reasoning, interpretation, problem solving, education*
	Hoffman & Podgurski, 2013	*Clinical care policy, immunisations, treatment effectiveness, electronic health records, clinical outcomes, public health research, surveillance of infectious disease, outbreaks, lab results, research, public health interventions*

As can be seen in Table 2, even from these 40 articles, there is a broad range of uses of big data in healthcare. A summary of these findings include:

- Seven out of the forty articles suggested that predictive analysis or some type of disease or predictive model can be used with big data analytics.
- Seven discussed a potential to reduce costs.
- Three articles contained information regarding semantic standards within big data analytics. In addition, primary care, health care delivery and a variety of other uses were found within the literature review.

A pattern was noted in regards to the sub categories; these either directly or indirectly influenced management of healthcare delivery. Therefore, it was deemed suitable to place each utilisation that had activities involving management into the category of administration and delivery: managing healthcare delivery costs.

- Twelve articles suggested that electronic health records would be integrated within big data analytics.
- Five suggested that evidence based medicine would be a significant outcome of big data analytics. Other big data activities within healthcare resulted in further support for clinical decision support.
- Ten articles suggest that big data analytics within healthcare will produce clinical decision support.

It can be inferred from the literature that electronic health records as well as other data can be process within big data analytics and result in extraction of valuable information that can be used as evidence based medicine and clinical decision-making support.

- Sixteen articles indicated a variety of data and information that would be utilised within big data healthcare analytics.
- Five articles expressed big data analytics role in genomics.
- Fifteen articles represented a big data role in public health.

The big data healthcare activities found within the literature displayed a pattern of databases and information that could be used by big data analytics to map out potential outbreaks of disease within the public. This would be possible as individual data is mapped out. These activities were deemed suitable for the category of clinical information, as they involve the various types of clinical information available. Information such as and not limited to genomics, patient data and clinical trials.

- Fifteen articles from the literature review demonstrated the utilisation category behaviour/consumer.

Literature regarding health behaviours, health observations, demographics, and information regarding the lifestyles of patients were noted. It is shown that big data analytics has the ability to track and map the lifestyle of individuals, and this can result in predictive analysis of individuals and the public. Therefore, a pattern involving the integration of all the categories is evident.

Support information and big data utilisation that did not fit the categories created were placed in other. It should be noted that activities performed by big data analytics within the category support information' also intertwine with the other categories. It is evident that the big data utilisation categories although distinct do interrelate.

Figure 3 was constructed from analysis of the articles in Table 2 and represents the currency of big data use within healthcare.

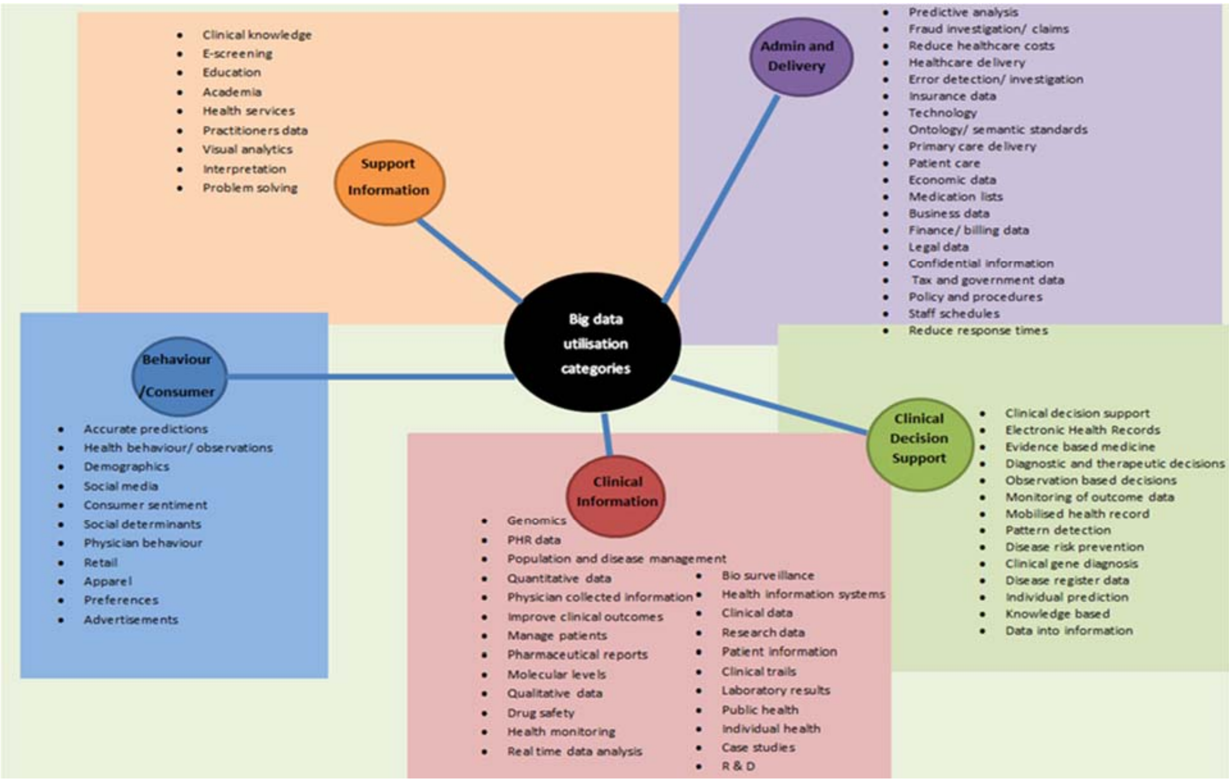

Figure 3 Big data uses in healthcare categorisation

CONCLUSION AND FUTURE RESEARCH

Effective integration of data mining and medical informatics and its subsequent analysis using big data techniques will no doubt impact healthcare delivery costing and improved healthcare results via well informed decision making (Sun, 2013). From a systematic review of the literature, a cross section of variety of articles was extracted for use within the study. Literature was located in multiple databases, which suggests that yet no standard or definitive natural place for publishing on big data in healthcare has been established. Further, the limited number of academic articles found suggest that as an aspect of healthcare, health informatics and data science, it has yet to establish a sound academic base and body of literature, despite this, there is a plethora of marketing cases and suggested uses evident through the popular press and websites. Future similar research may include analysis of such materials and examples of vendor studies, as well as identification of other material from reference citations from the initial articles selected.

From the systematic review of the literature, utilisation categories were created. The categories incorporate semantics. The creation of categories can potentially result in creation of standard terms and vocabulary used within big data analytics. Indeed, this could result in more refined searches and improved yield to support evidence based medicine, and form reliable and efficient decision support for diagnosis and treatment of patients. Identification of what literature is currently being published regarding big data analytics in healthcare may lead to more defined publishing that can assist in identifying the current uses of big data analytics. In addition, identification of the databases can also result in determining what databases are helpful for future research in big data analytics.

This research suggested big data utilisation categories, which the existing literature can be classified into: administration and deliver, clinical decision support, clinical information, behaviour/consumer and support information. Using an extended classification system to classify studies, a broad categorisation of big data uses in healthcare has been mapped. This provides a basis upon which to explore 'taming' big data use, so that a greater understanding can be promoted. Further, this will allow subsequent identification of the types of analytical tools that can be employed. The study provides a picture of the disparate and diverse uses of big data in healthcare. This indicates that today, its use is not systematic but opportunistic. The potential is being explored through these initial uses, and demonstrates a promise of rapid expansion and exploration in a proliferating data environment. This preliminary review provides a basis from which to investigate and target fundamental questions on the data content, storage, processing, delineation and access. Future research should draw attention to these factors in addition to questioning the errors in such data and sources, and the metrics with which the use of big data can be measured. This paper adds to the discourse, as it is the first systematic review of big data uses in the current literature.

REFERENCES

Ackerman, M. (2012). Big data. *The Journal of Medical Practice Management : MPM, 28*(2), 153.

Adler-Milstein, J., & Jha, A. K. (2013). Healthcare's "Big Data" challenge. *American Journal of Managed Care, 19*(7), 537-538.

Bates, D. W., Middleton, B., Kuperman, G. J., Wang, S., Gandhi, T., Kittler, A., . . . Tanasijevic, M. (2003). Ten commandments for effective clinical decision support: Making the practice of evidence-based medicine a reality. *Journal of the American Medical Informatics Association, 10*(6), 523-530. doi: 10.1197/jamia.M1370

Bentley, R. A., O'Brien, M. J., & Brock, W. A. (2014). Mapping collective behavior in the big-data era. *Behavioral and Brain Sciences, 37*(01), 63-76.

Bollier, D. (2010). The promise and peril of big data. Washington, DC: The ASPEN Institute.

Bonney, S. (2013). HIM's role in managing big data: Turning data collected by an EHR into information. *Journal of AHIMA / American Health Information Management Association, 84*(9), 62.

Botsis, T., Hartvigsen, G., Chen, F., & Weng, C. (2010). Secondary Use of EHR: Data Quality Issues and Informatics Opportunities. *Summit on Translational Bioinformatics, 2010, 1–5*. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041534/.

Chawla, N. V., & Davis, D. A. (2013). Bringing big data to personalized healthcare: a patient-centered framework. *Journal of General Internal Medicine, 28 Suppl 3*(3), S660-665. doi: 10.1007/s11606-013-2455-8

Chen, M., Mao, S., & Liu, Y. (2014). Big sata: A survey. *Mobile Networks and Applications, 19*(2), 171-209. doi: 10.1007/s11036-013-0489-0

Clarke, J. (2011). What is a systematic review? *Evidence Based Nursing, 14*(3), 64. doi: 10.1136/ebn.2011.0049

Clifford, L. (2008). How do your data grow? *Nature, 455*(7209), 28.

Commonwealth of Australia. (2013). *The Australian public service big data strategy: Improved understanding through enhanced data analytics capability*. Canberra: Commonwealth of Australia.

Cooper, R., Kennedy, J., & Springer, V. (2007). Data management: data, data everywhere, In Proceedings of the *24th British National Conference on Databases, BNCOD 24, Glasgow, UK, July 3-5, 2007 : proceedings* (Vol. 4587.). New York: Springer Berlin Heidelberg.

Cushing, J. B. (2013). Beyond big data? *Computing in Science & Engineering, 15*(5), 4-5. doi: 10.1109/mcse.2013.102

Diehr, P., & Lumley, T. (2002). The importance of the normality assumption in large public health data sets. *Annu Rev Public Health, 23*, 151-169.

Doug, H., Susan St, P., Simon, T., Owen, W., Seung Yon, R., Maria, C., . . . Mary, S. (2008). The future of biocuration. *Nature, 455*(7209), 47.

Eramo, E. (2013). Healthcare’s data revolution. *Journal of AHIMA, 84*(9), 27.

Feldman, B., Martin, E., & Skotnes, T. (2012). *Big data in healthcare: Hype and hope*. Retrieved from http://www.west-info.eu/files/big-data-in-healthcare.pdf.

Ferlie, E. B., & Shortell, S. M. (2001). Improving the quality of health care in the United Kingdom and the United States: a framework for change. *The Milbank Qquarterly, 79*(2), 281-315. doi: 10.1111/1468-0009.00206

Graeff, T., & Harmon, S. (2002). Collecting and using personal data: consumers’ awareness and concerns. *Journal of Consumer Marketing, 19*(4), 302-318. doi: 10.1108/07363760210433627

Green, L. W. (2006). Public health asks of systems science: to advance our evidence-based practice, can you help us get more practice-based evidence? *American Journal of Public Health, 96*(3), 406-409. doi: 10.2105/AJPH.2005.066035

Gross, P. A., & Bates, D. W. (2007). A pragmatic approach to implementing best practices for clinical decision support systems in computerized provider order entry systems. *Journal of the American Medical Informatics Association, 14*(1), 25-28. doi: 10.1197/jamia.M2173

Groves, P., Kayyali, B., Knott, D., & Van Kuiken, S. (2013). *The 'big data' revolution in healthcare: Accelerating value and innovation*: McKinsey & Company.

Haas, L. (2013). *IBM ALmaden Research Centre.* Paper presented at the HISA Big Data 2013 Melbourne Presentation, Melbourne

Harper, E. M. (2013). The economic value of health care data. *Nursing Administration Quarterly, 37*(2), 105.

Heubusch, K. (2008). Certified EHRs. *Journal of AHIMA, 79*(8), 34.

Hoffman, S., & Podgurski, A. (2013). Big bad data: Law, public health, and biomedical databases. *The Journal of Law, Medicine & Ethics, 41*(Journal Article), 56-60. doi: 10.1111/jlme.12040

Howe, D., Costanzo, M., Fey, P., Gojobori, T., Hannick, L., Hide, W., . . . St Pierre, S. (2008). Big data: The future of biocuration. *Nature, 455*(7209), 47-50.

JBIEBNM. (2000). *Appraising systematic reviews*. Retrieved from http://connect.jbiconnectplus.org/ViewSourceFile.aspx?0=4311

Jee, K., & Kim, G. H. (2013). Potentiality of big data in the medical sector: focus on how to reshape the healthcare system. *Healthc Inform Res, 19*(2), 79-85. doi: 10.4258/hir.2013.19.2.79

Kallus, N. (2014). *Predicting crowd behavior with big public data.* Paper presented at the Proceedings of the companion publication of the 23rd international conference on World wide web companion.

Kaye, K. (2013). Big data: Marketers are tracking consumers, but regulation is a moving target. *Advertising Age, 84*(29), S014.

Kuriyan, J., & Cobb, N. (2013). Forecasts of cancer and chronic patients: Big data metrics of population health. Retreived from http://arxiv.org/abs/1307.3434.

Lassila, O. (1998). Web metadata: a matter of semantics. *IEEE Internet Computing, 2*(4), 30-37. doi: 10.1109/4236.707688

Liu, W., & Park, E. K. (2014, 3-6 Feb. 2014). *Big data as an e-health service.* Paper presented at the Computing, Networking and Communications (ICNC), 2014 International Conference.

Monheit, A. C., & Berk, M. L. (2001). The concentration of health care expenditures, revisited. *Health affairs, 20*(2), 9-18. doi: 10.1377/hlthaff.20.2.9

Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *JAMA, 309*(13), 1351.

Ola, O., & Sedig, K. (2014). The challenge of big data in public health: an opportunity for visual analytics. *Online J Public Health Inform. 2014 Feb 5;5*(3):223. doi: 10.5210/ojphi.v5i3.4933. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24678376.

Peleg, M., & Tu, S. (2006). Decision support, knowledge representation and management in medicine. *Yearb Med Inform, 45*, 72-80.

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems 2014, 2*:3 doi:10.1186/2047-2501-2-3. Retreived from http://www.hissjournal.com/content/2/1/3

Ringel, D. M., & Skiera, B. (2014, 6-9 Jan. 2014). *Understanding Competition Using Big Consumer Search Data*. Paper presented at the System Sciences (HICSS), 2014 47th Hawaii International Conference

Rose, J., & Burgin, M. (2014). Disrupting health care through big data and predictive analytics. *Managed Care Outlook, 27*(1), 11.

Russom, P. (2011). *Big data analytics: TDWI best practices report.* Renton, WA: TDWI Research.

Sanat, J. (2013). Big data. *InTech, 60*(3), 40.

Sanders, D. (2013). *Before you buy: A checklist for evaluating your analytics vendor*. Utah: Health Catalyst.

Schouten, P. (2013). Big data in health care solving provider revenue leakage with advanced analytics. *Healthcare Financial Management, 67*(2), 40.

Shaw, G. (2013). *Data analytics: Actionable information to improve healthcare*. Retreived from http://www.fiercehealthit.com/offer/dataanalytics?source=library

Sheth, A. (1996). *Data semantics: what, where and how.* Paper presented at the 6th IFIP Working Conference on Data Semantics (DS-6), Atlanta, GA.

Skolnik, N., & Bertman, J. (2013). EHR report: Big data is coming--are you ready? *Internal Medicine News, 46*(6), 30.

Skolnik, N., & Bertman, J. (2013). EHR report: CME, EHR integration improves clinical outcomes. *Family Practice News, 43*(6), 19.

Steinbrook, R. (2008a). Personally controlled online health data--the next big thing in medical care? *The New England Journal of Medicine, 358*(16), 1653-1656. doi: 10.1056/NEJMp0801736

Steinbrook, R. (2008b). Personally controlled online health data — The next big thing in medical care? *The New England Journal of Medicine, 358*(16), 1653-1656. doi: 10.1056/NEJMp0801736

Sun, J., & Reddy, C. (2013). *Big data analytics for healthcare*. Paper presented at the Tutorial presentation at the SIAM International Conference on Data Mining. Austin, TX.

Swinglehurst, D. A., Pierce, M., & Fuller, J. C. A. (2001). A clinical informaticist to support primary care decision making. *Quality in Health Care, 10*(4), 245-249. doi: 10.1136/qhc.0100245.

Thadani, S. R., Weng, C., Bigger, J. T., Ennever, J. F., & Wajngurt, D. (2009). Electronic screening improves efficiency in clinical trial recruitment. *Journal of the American Medical Informatics Association, 16*(6), 869-873. doi: 10.1197/jamia.M3119

Transformation, I. f. H. T. (2013). *Transforming healthcare through big data: Strategies for leveraging big data in healthcare industry*. New York: Institute for Healthcare Technology Transformation.

Vemuganti, G. (2013). Metadata management in big data. *Big Data: Countering Tomorrow's Challenges*, 3.

Villars, R., Olofson, C., & Eastwood, M. (2011). Big data: What it is and why you should care. *White Paper, IDC*.

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *J. of Management Information Systems, 12*(4), 5-33.

Weber, G. M., Murphy, S. N., McMurry, A. J., MacFadden, D., Nigrin, D. J., Churchill, S., & Kohane, I. S. (2009). The Shared Health Research Information Network (SHRINE): A prototype federated query tool for clinical data repositories. *Journal of the American Medical Informatics Association, 16*(5), 624-630. doi: 10.1197/jamia.M3191

Yoon, S., Park, J., Kim, M., Lim, C., & Cho, J. (2014). *Behavior signature for big data traffic identification.* Paper presented at the Big Data and Smart Computing (BIGCOMP), 2014 International Conference.