

Chapter

A Systematic Review of Knowledge Visualization Approaches Using Big Data Methodology for Clinical Decision Support

Mehrdad Roham, Anait R. Gabrielyan and Norm Archer

Abstract

This chapter reports on results from a systematic review of peer-reviewed studies related to big data knowledge visualization for clinical decision support (CDS). The aims were to identify and synthesize sources of big data in knowledge visualization, identify visualization interactivity approaches for CDS, and summarize outcomes. Searches were conducted via PubMed, Embase, Ebscohost, CINAHL, Medline, Web of Science, and IEEE Xplore in April 2019, using search terms representing concepts of: big data, knowledge visualization, and clinical decision support. A Google Scholar gray literature search was also conducted. All references were screened for eligibility. Our review returned 3252 references, with 17 studies remaining after screening. Data were extracted and coded from these studies and analyzed using a PICOS framework. The most common audience intended for the studies was healthcare providers ($n = 16$); the most common source of big data was electronic health records (EHRs) ($n = 12$), followed by microbiology/pathology laboratory data ($n = 8$). The most common intervention type was some form of analysis platform/tool ($n = 7$). We identified and classified studies by visualization type, user intent, big data platforms and tools used, big data analytics methods, and outcomes from big data knowledge visualization of CDS applications.

Keywords: systemic review, knowledge visualization, big data, clinical decision support, visual analytics, health care

1. Introduction

A clinical decision support system (CDSS) involves the use of digital information and communication technologies to bring relevant knowledge to bear on the healthcare and well-being of the patient ([1], p. 6). A CDSS has the following components ([1], p. 11):

Purpose: Task or process of clinical care which uses the CDSS.

Structure:

- a. Decision Model (Organization of data and/or knowledge to generate recommendations)
- b. Knowledge Base (Knowledge content used by the CDSS)
- c. Information Model (Representation of clinical and decision support parameters)
- d. Result Specification (Decision model output for user support)
- e. Application Environment (CDSS interaction with a host application (e.g. an electronic health record system))

The growing use of medical/healthcare big data and data analytics is having a profound impact on decision models and knowledge bases in CDSS development and applications of today. There is a particular need to address user requirements when attempting to recognize patterns in the massive volumes of data being presented to decision makers (as in the Result Specification component of a CDSS structure, as outlined above). In psychology and cognitive neuroscience, pattern recognition in humans is the result of a cognitive process whereby the brain attempts to match information from a stimulus that is received and entered into short term memory with certain content retrieved from long-term memory. While the stimulus may arise from any of our senses, in this chapter we will focus on visual sensations of information that can inform decision making.

Why is knowledge visualization important to clinical decision support? Knowledge visualization is essential when there is a need to augment human capabilities rather than to attempt to automate decision-making computationally. Data mining is widely used for discovering latent knowledge from databases. This can contribute to clinical decision making. Electronic health record (EHR) and other healthcare systems contain large volumes of patient data, often with different formats and structures, so these data are typically accessible only in a form that is not conducive to rapid synthesis and interpretation for potential therapeutic outcomes.

Medical data inherently contain information that provides support for patient-centric and personalized healthcare. Machine learning approaches, combined with high-performance distributed computing technologies such as Hadoop can assist in the exploitation of big healthcare databases. In a high pressure clinical environment, well-designed knowledge-based visualization can play a key role for clinicians and managers who must deal with complex issues and decisions. These may arise from algorithms derived through computational results from healthcare databases containing data from many patients, with each record having hundreds of attributes. As other chapters in this book discuss in detail, there are machine learning/deep learning and other approaches that can convert these massive databases into decision support resources. These resources are virtually never deterministic and are difficult to validate to a high degree of accuracy, but they can help to support decision makers needing to make individualized healthcare decisions.

The value in knowledge visualization lies in presenting knowledge to the user in a way that gives the user factual information that is useful for current decisions. Since there are multiple dimensions to every specific decision, the visualization designer must provide flexibility in potential displays, while taking into account

limitations on computer resources, human capabilities, and the display environment. Visualization system usage can be analyzed in terms of why the user needs it, what data is shown, and how the display idiom is designed ([2], p. 1).

The digitization of patient health records has profoundly impacted medicine and healthcare. The compilation of medical history provides a holistic account of patient condition, family history, social determinants, procedures, and medications. In addition to regular treatment benefits, the availability of this information in digital form has created opportunities for population-level monitoring and studies that can help to guide initiatives for improving quality of care. Each hospital client population base varies in geography, size, and organizational structure, and extracting meaning from the data to support a hospital's strategic mission requires a combination of clinical, statistical, and technical literacy. Effective visualization of data resources is essential to the efficient use of these resources by the healthcare facility [3].

The objectives of this study are:

- To explore the sources of big data in knowledge visualization for clinical decision support;
- To identify the visualization and interactivity approaches that have been used in the light of recent CDSS advances in the use of big data;
- To identify big data analytical techniques and technologies uses in CDSS with knowledge visualization;
- To determine the purposes of CDSS, and to illustrate the potential benefits of big data and knowledge visualization for these purposes;
- To recognize what appears to be the future of visualization in the support of big data and CDSS

By systematically investigating these objectives in detail, this systematic review will make a significant contribution towards understanding how big data knowledge visualization can be used in clinical decision support. This will benefit the development of future applications in this field.

2. Methodology

2.1 Design

This systematic review was carried out according to the recommendations and reporting specified in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [4, 5].

2.2 Inclusion and exclusion criteria

Any peer-reviewed studies published from January 2008 to April 2019 that explored the applications of big data knowledge visualization in a clinical decision support setting were retained for full-text review. Excluded studies included reports on results of: abstracts, surveys, focus groups, feasibility studies, monitoring devices, "what-if" analysis, big data collection techniques, commentaries, letters, editorials or short reports, mere usage of big data technologies or knowledge

visualization, and studies not published in English. Publications describing big data and visualization applications that were not developed for clinical decision support were also excluded.

2.3 Search strategy and screening process

We developed the search strategy according to the Cochrane Handbook of Systematic Reviews of Interventions [6]. Search keywords (terms) were grouped into three categories: big data, knowledge visualization, and clinical decision support (see **Table 1**). Searches were conducted in electronic databases: EBSCOHOST, EMBASE, MEDLINE VIA OVID, PUBMED, IEEE XPLORE, WEB OF SCIENCE, and GOOGLE SCHOLAR (the latter for a gray literature search) on April 5–6, 2019. We also used the related article function in PubMed, Science Direct, and Springer Link databases on initially included studies to identify additional studies.

Studies were selected in two steps. First, the title and abstract for each study were screened for inclusion and exclusion criteria and classified by two reviewers. If the title and abstract did not provide enough information to assess its relevance or if a final decision could not be made, we assessed the full article. Second, the full texts of the studies without enough information and/or deemed to be potentially relevant were reviewed randomly by two reviewers. Any disagreements between the two reviewers were resolved via discussion or by consultation with a third reviewer. Studies cited in eligible articles were also reviewed following a similar screening process. The studies identified for systematic review were examined by two reviewers qualitatively, as described below.

2.4 Data: collection, management, analysis and classification

All reviewers used a shared spreadsheet template to classify and summarize eligible studies and keep track of review progress. Counts and percentages were based on the databases from which each study was identified in the following order: WEB OF SCIENCE, IEEE XPLORE, PUBMED, EMBASE, EBSCOHOST and GOOGLE SCHOLAR. For example, if a study was identified in both WEB OF SCIENCE and EMBASE, it counted as being found within the WEB OF SCIENCE.

All eligible studies were classified using the Population/Problem–Intervention–Comparison–Outcomes–Study (PICOS) design framework [6, 7], which provides a common approach for detailed specification of the review questions and criteria during

KEYWORDS: (P) AND (I) AND (O):	
within each group the keywords are combined using the “OR” operator	
P (BIG DATA)	BIG DATA
I (KNOWLEDGE VISUALIZATION)	KNOWLEDGE VISUALIZATION OR VISUAL ANALYTICS OR INFORMATION VISUALIZATION OR VISUALIZATION OR BUSINESS INTELLIGENCE OR BI OR DASHBOARD
O (CLINICAL DECISION SUPPORT)	CLINICAL DECISION SUPPORT OR CLINICAL DECISION MAKING OR DECISION SUPPORT SYSTEMS OR HEALTHCARE OR HEALTH CARE OR HEALTH OR CARE OR CLINICAL INTELLIGENCE OR HOSPITAL OR CLINIC OR CLINICAL OR MEDICAL OR DIAGNOSIS OR TREATMENT

Table 1. Search terms used to identify publications related to knowledge visualization of big data for clinical decision support.

study design. In this systematic review, the same PICOS categories are used to synthesize the data extracted from eligible studies. All PICOS subcategories were identified and classified through the review process. Similarly [7] where applicable, pre-existing taxonomies such as interaction taxonomy [8, 9], major data types of big data in health care [10] and target audience [7] were integrated with the PICOS framework.

The synthesis of the findings contains results for each PICOS category (Population/Problem–Intervention–Comparison–Outcomes–Study Design) which covers all related subcategories.

3. Results

A total of 2338 references were retrieved from our first search of electronic databases. A second search of the grey literature via GOOGLE SCHOLAR (n = 780) and a third manual search (n = 134) yielded a total of 3252 studies. We then removed 866 (26.6%) duplicate studies. We screened all titles and abstracts for 2386 (73.4%) papers and excluded 1843 studies because the visualizations discussed were not for healthcare, studies were only for big data without visualization or were not for CDS (clinical decision support), papers that used genomics/DNA or image data, or presented only an abstract. The full text of each of the remaining 543 (16.7%) studies was then read. A search of citations yielded an additional 34 eligible studies. A total of 577 studies were classified and 560 of these studies were excluded. The PRISMA flow diagram (**Figure 1**) summarizes the screening process. The 17 studies remaining were included in the qualitative synthesis [11–27] that follows.

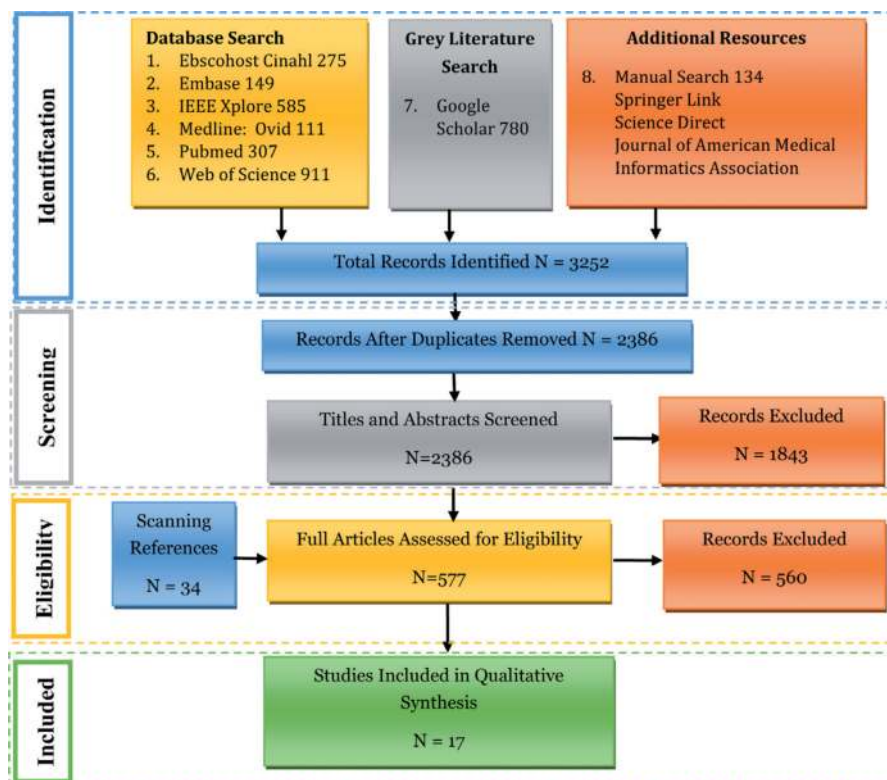


Figure 1.
PRISMA flow diagram.

3.1 Characteristics of included studies

The majority of the 17 included publications were indexed in the Web of Science (n = 12, 70.6%). The series Lecture Notes in Computer Science (LNCS) and the Journal of the American Medical Informatics Association (JAMIA) were the other sources of the studies (n = 3 (17.6%) and n = 2 (11.8%) respectively). More than half of the studies (n = 9, 52.9%) were journal publications and nearly 94% of the studies were published in the past 4.5 years (2015–2019). Studies originated from 10 countries: the USA tops the list with 9 (near 53%) publications, followed by Canada with 3 (17.6%), France and UAE with 2 each (11.8%), and Greece, Italy, South Korea, Portugal, Taiwan and Spain with 1 each (5.9%). 5 (29.4%) studies were conducted in two countries and more than half of the studies were conducted at a university. More than 82% of the studies discussed one specific type of visualization tool or system.

3.2 PICOS classification: part 1: population, patient and problem

3.2.1 Population (target users) and patients

We identified four main groups of users of big data visualization applications for clinical decision support: (1) academicians (clinical researchers and clinical epidemiologists, nurse educators) [11, 12, 17, 18, 26], (2) administrators (hospital administrators and managers) [12, 17, 25], ancillary staff (caretakers, lab managers, pharmacists) [11, 13, 25]; (3) health care providers (clinicians, nurses, physicians) [11–16, 18–27] and (4) patients [11, 14, 22, 23]. More than 94% of the studies (n = 16) were developed to support clinical healthcare provider decision-making.

A majority of the studies were in non-intensive hospital settings (n = 9, 53%), 2 in intensive surgical and trauma settings, 2 for outpatients, and 6 studies that did not specify hospital settings. Studies were designed to investigate different diseases or health problems, such as acute kidney injury [20], appendicitis [16], cardio, respiratory, and adverse events [19, 26], chronic diseases and diabetes [11, 13], hospital infections and sepsis [20, 25]. Two studies were developed to help predict various hospital complications after treatment [17, 20], and seven studies (41%) did not provide any information about specific patient diseases or problems.

3.2.2 Big data major types and sources

Big data in health care can be classified into four major types based on data sources: (1) big data in medicine, (or medical/clinical big data); (2) big data in public health and behavior; (3) big data in medical experiments; and (4) big data in medical literature [10]. We identified three major types of big data in reviewed studies (**Figure 2**).

Big data in medicine and clinics includes big data generated in hospitals, such as electronic health records (EHRs)/medical records (EMRs), personal health records (PHRs), and medical images (visual information of the internal human body). This group was the most frequent source of data in the 17 studies. Sixteen studies (94%) used big data from medicine and clinics. EHRs, EMRs (n = 12) and PHRs (n = 7) were the main sources of big data. These records consisted of some combination of clinical notes, laboratory and image reporting results, medical histories, hospital stay information, medication and allergies, patient demographics, diagnoses, sensor data, etc., all of which are the basis for personalized medicine.

Big data sourced from public health and patient activities are the second major sources of big data in our studies (n = 13, 76%). These focus on the physiological

BIG DATA TYPES	GROUP	BIG DATA AQUISION GROUP	BIG DATA AQUISION SOURCE
BIG DATA IN MEDICINE AND CLINICS	1	Electronic health record (EHR)/ electronic medical record (EMR)	Clinic or inpatient notes Clinic or inpatient notes (not for vis, only for evaluation) Clinical data from outpatient visit Clinical diagnosis Drug administration Electronic reports Hospital information resource (information about ward, bed, edc) Hospital stay Information(LOS, mortality, Length of ICU, septic shock, etc.) not for vis, only for evaluation) Identification information of patient Imaging report Information of patient Notes from physician Patient demographics (not for vis only for evaluation) Patient transfers to different hospital units (including emergency, ICU, surgical suite or ambulatory) Physical examination Physical examination (not for vis, only for evaluation) Sensor data (Glucose, Weight, Hemoglobin A1c, Medications dosage, Food/Activity Log, etc.) Sensor data (blood pressure) Surgery's work/protocols Treatment
	2	Personal health record (PHR)	Illnesses and hospitalizations Laboratory test results Laboratory test results (not for vis, only for evaluation) Medical history Medication data Medication history Medications and dosing Nursing assessments Patient demographics Provided by Government, Hospital or other organization
	3	Medical Images (visual information of interior human body)	X-ray
BIG DATA IN PUBLIC HEALTH AND BEHAVIOR	4	Measures and records of the electrical and electromagnetic signals from the body	A live data feed from patients' continuous monitoring systems Continuous activity data from a wearable device (Fitbit) Continuous physiologic data from the bedside monitors Electrocardiograph (ECG) signal Electroencephalography (EEG) signal Sensor data from The Internet of Things (IOT): the approach of connecting real-world objects "to the Internet via sensors, and microprocessor chips that record and transmit data" such as brain activity, heart rate, sound waves, and body temperature
	5	Vitals	Sensor data(Glucose, Weight, A1c, Medications dosage, Food/Activity Log, etc.) Acute Physiology and Chronic Health Evaluation(APACHE) score (not for vis, only for evaluation) Temperature, pulse, respiratory rate, and blood pressure
	8	Microbiology/Pathology Laboratory	Blood glucose, carbohydrates, medications, insulin, physical activity, blood pressure, body weight,Hemoglobin A1c levels Hemoglobin A1c levels
BIG DATA IN MEDICAL LITERATURE AND SOCIAL MEDIA	12	Structured knowledge	MeSH and International Classification of Diseases 10th revision (ICD-10), etc
	13	Social Media Data	Community Forums, Social Connectivity, Meal Plans, Food and Fitness Guides,Disease Guides, Health Information

Figure 2.
 Major types of big data in the 17 chosen studies.

data of users that are often collected by portable equipment [10] such as electrical and electromagnetic signals from body vitals collected by wearable devices, daily health records, and from sports and personal diets.

The main sources of the big data in public health and behavior include measures and records of electrical and electromagnetic signals from the body (n = 5, 29%) which comprise live data feeds from patient monitoring systems, electrocardiograph (ECG) and electroencephalography (EEG) signals, and sensor data from the Internet of Things (IOT). These approaches connect humans “to the Internet via sensors, and microprocessor chips that record and transmit data” such as brain activity, heart rate, sound waves, and body temperature” [22, 23].

Structured knowledge, such as big data from the medical literature under Medical Subject Headings (MESH) codes, the International Classification of Diseases 10th revision (ICD-10), laboratory test codes, etc. and social media data are also significant source of big data in a number of the 17 studies we selected for review (n = 3, 18%) (Figure 2).

3.2.3 Big data size

The studies reviewed included more than 5.4 million patients, mostly admitted to hospital. The sample size of patients in the studies ranged from 1757 to 2.9 million. Authors reported the number of records, tests and patient stay days that ranged from 15,700 to 17.8 million. Three studies [11, 18, 21] used structured and/or social

media big data ranging from 7000 to 300 million records. About 65% of the studies (n = 11) used real time data such as ECG streaming, producing 500,000 messages per minute and 80 different types of vitals totaling 100,000 messages [26].

3.2.4 Big data types

Big data visualization studies varied significantly among the data types they could handle [8].

The data information types could be categorical (nominal (diagnosis, treatment)) or ordinal (“high”, “low” blood pressure level), numerical (for example, cholesterol measure, temperature), texts, maps or networks. All 17 studies used categorical and numerical data information types, while more than 75% (n = 13) used text data types (clinician notes, laboratory results, notes, etc.). Applications in 13 studies used time series data, including real time signals in three studies (18%) [21, 22, 26]. Three studies [13, 15, 16] used maps in their applications to provide clinical decision support. All applications reviewed could deal with several types of data information, with the most common being combinations of categorical, numerical, text and time series.

We further categorized data by medical information type. Most of the studies dealt with physical examination of patients, patient outcomes and diseases, symptoms, and treatment problems which require support for clinical decision making (see **Table 3**).

A variety of data collection methods were used, with more than half of the studies using data from EHR/EMR systems. The other most common method involved receiving data from continuous monitoring via EEG, ECG, bedside monitoring, or IOT [11, 14, 19, 22, 23, 26]. Two studies used big databases [11, 20], one study used data from a provincial pathology laboratory [13] and one study used user input data in combination with IOT data [14].

3.3 PICOS classification: part 2: intervention

3.3.1 Intervention type and big data visualization techniques

We identified 11 subcategories of the PICOS intervention classification to help classify all interventions in the studies such as intervention type, big data visualization techniques, visualization types and others.

Seven studies had developed data analysis platform/tools to help the target audience (healthcare providers, academicians, etc.) [12, 16, 17, 21, 24–26]. Three studies developed and used mobile care coordination [14, 15, 20], four studies designed web portals for healthcare providers [11, 13, 22, 23] and patients [11]. Two studies used a multi-patient surveillance system [12, 18] and one study used a multi patient dashboard to support clinical decisions.

For big data visualization techniques 47% (n = 8) of all studies developed web application and 41% (n = 7) of all studies used dashboards for supporting clinical decisions.

3.3.2 Visualization types

We classified visualizations (36 different types) into 7 visualization categories (**Table 2**). On average, studies used three different visualization types. One study [21] used only one visualization type (tabular), five studies (29%) presented two different visualization types (multidimensional and tabular, tabular and temporal, etc.) [15, 17, 18, 20, 24].

PICOS subcategory: visualization types	All studies	All studies, %
3D/volumetric	3	17.6
Icons	5	29.4
Maps	3	17.6
Multidimensional	12	70.6
Tabular	14	82.4
Temporal	10	58.8
Textual	4	23.5

Visualization types subcategory.

Table 2.
 PICOS classification of all publications: part 2: intervention.

In 6 studies (35.3%) three different types of visualizations were used [11, 12, 14, 25–27] with the most common combination being multidimensional, tabular and temporal. Four studies (23.5%) presented four different types of visualization [13, 16, 19, 22], but only one study used six of the seven visualization categories we identified [23].

The most common visualization category (n = 14.82%) for our studies was tabular (table) visualization for seven studies [11, 14, 16, 17, 20, 26, 27]. Three studies used tabular visualization with color (9 different indicators) and tabular visualization with color indicating risks [12, 22, 23]. One study used tabular visualization with color coding indicating change from previous state [25] or tabular visualization with color coding indicating change in value and out of normal ranges or trend of changes [21]. One study used a very innovative tabular visualization with color comets [19], where the comet head representing risks at the current time, and a tail that is 3 h long.

More than 70% of the studies (n = 12) used multi-dimensional visualization techniques such as area charts (color grading/without color grading) [14, 19, 26], bar graphs [11, 13, 14, 16, 26], bipartite graphs [25], box plots [17], bubble charts [12], causal network visualizations and heatmaps [13], key performance indicators (KPIs) [12, 16], line graphs [11, 19, 24], pie charts [11, 14, 20] and scatter plots [23].

Temporal or timeline visualization is another major category of knowledge visualization for clinical decision support (n = 10, 59%). Two studies reported visualization of just simple time series graphs without color coding [23, 24], although six studies used time series line graphs with color coding indicating: different indicators [25, 27], different status [18, 19], changes from previous encounter (Emergency, Hospital Unit, Surgery) [18] or time changes (breakfast, lunch, dinner, etc.) [11]. Four studies (24%) presented visualizations of the time series trend line [14, 16, 18, 26].

Five studies used icons [12] or icons with color grading zones [15, 19, 22, 23], three studies used 3D visualizations [13] or 3D brain maps [22, 23]. The spatial context of the big data was presented via maps [16] and interactive geo-spatial maps [13, 15]. Four studies presented textual information using simple text [18, 22, 23] or word clouds for problem identification [27]. More information about specific techniques, including background, explanations, and concepts can be found in [28].

3.3.3 Support for user intent

Information visualization can be compared and classified by interaction features. There are a great variety of interaction visualization techniques [29], but we will use an intent model proposed by [9] and extended by [8]. The concept focuses

PICOS subcategory: user intent	All studies	All studies, %
Select: mark something as interesting	15	88.2
Explore: show me something else	13	76.5
Reconfigure: show me a different arrangement	13	76.5
Encode: show me a different representation	7	41.2
Abstract/elaborate: show me more or less detail	12	70.6
Filter: show me something conditionally	14	82.4
Connect: show me related items	10	58.8
Print, submit feedback, record/download: show me something to save	8	47.1

User intent subcategory.

Table 3.
PICOS classification of all publications: part 2: intervention.

on ‘What a user wants to achieve’ and described as “user intent,” quite effective technique to classify the low-level interaction techniques into seven descriptive high-level categories [9]. We also added an additional category that includes printing, submitting feedback or saving features (**Table 3**).

Most of the studies support user interactivity (n = 15, 89%). *Select: mark something as interesting* interaction techniques such as keeping track or managing a group are available in 15 and 13 studies respectively (77%). The *Filter: (show me something conditionally)* interaction technique was very useful for multiple patient applications/systems and enables users to change the set of data items conditionally (in some specific conditions). This second intent played a leading part in the reviewed applications and was used in 14 studies (82%).

The *Explore: show me something else* includes items such as reposition, sorting, editing, adjusting axis and *Reconfigure: show me a different arrangement* (switch representation technique, vary visual encoding) intents were available in 13 studies (77%), while *Abstract/Elaborate: show me more or less detail* intent gives users the ability to adjust the level of abstraction of a data representation (for example via time or time constraints) is available in 12 studies (71%).

In more than half of the studies (n = 10, 59%) the *Connect: show me related items* intent is supported, where the most common interaction technique was to show patient/group relationship (n = 9, 53%). *Print, Submit Feedback, Record/Download: show me something to save* (printing, submitting feedback and recommendations, recording or downloading) intent is enabled in 8 studies (47%) with the most common being printing. *Encoding: show me a different representation* (temporal data binning, altering fundamental representation of visual appearance (e.g. color, size, and shape) intent is supported in 7 studies (41%) with the most common being the switch representation technique (n = 5, 29%).

3.3.4 Cognitive presentation and units visualized

In most of the 17 studies, color and similarity (grouping) are the most supported cognitive presentations. Eight studies (47%) support all cognitive presentation types, such as color, size and similarity (grouping), seven studies (41%) used color and similarity (grouping) and only one study [20] supported color and size. In just one study [15], only color presentation is supported.

More than 76% of the studies (n = 13) supported visualization for single patients, nine studies (53%) enabled visualization for multiple patients. Nearly

30% of the studies (n = 5) visualized aggregated data, such as groups of patients. Application or visualization tools can be used for single users in one study [23]. In addition, only one study supported data for multi-products [25].

3.3.5 Big data analytics methods and problems

Three main groups of analytical methods were identified in the 17 studies. The most commonly used type was Descriptive (n = 16, 94%) which summarizes data to provide useful information and sometimes prepares it for further analysis. Predictive analytics (conditional logistic regression, deep learning, machine learning, data mining, AI) was the second most popular analytical method and was used in 11 studies (65%). Only two studies (12%) supported natural language processing (text mining, sentiment analysis).

These analytical methods were used to solve a variety of analytical problems such as analyzing all note types from different providers [27], classification [23] and pattern recognition [11], clustering [13, 23], electronic risk visualization of respiratory and cardiovascular events [19], identification of sequences of inappropriate drug administration cases [21], calculation of aggregated score and progression of the patient [26], perioperative risk prediction and visualization [20], and risk estimation [17, 18, 26]. One study [24] supported only summary statistics for contours (as in box-whisker, i.e. mean, median, percentiles 25 and 75, and outliers). One other study [25] used statistical measures such as prevalence and incidence of microorganisms, antibiotics, and microbiologicals.

3.3.6 Big data platforms and tools

The use of big data requires the support of new analytical and other types of tools for handling such massive and different types of data as well as technological innovations for data management, integration and interoperability, storage, distributed processing and analysis, and visualization. The studies we reviewed used a great variety of big data platforms and tools. We classified them into 13 categories which are presented in **Table 3**.

Cloud computing services (such as Amazon Web Services, Azure and Google clouds or other clouds not specified) were used in most of the studies (n = 11, 65%).

More than half of the studies (n = 9, 53%) supported big data maintenance/storage technologies based on NoSQL databases [13] and the Hadoop Distributed File System (HDFS) [13, 20, 25, 26]. This group also included studies using Apache HBase [13, 14, 27], Cassandra [20] and MongoDB [14, 15, 17, 24].

Big data maintenance/storage technologies based on SQL and SQL on Hadoop solutions were supported in six studies. The most frequent solutions used were MySQL [17, 22], PostgreSQL, Oracle or Microsoft SQL [21, 25], RSQL [18], Spark SQL [20], Apache Hive and Apache Impala [25].

Big data scalable distributed processing and analysis was supported in 10 studies (59%) and included Hadoop MapReduce [13, 15, 22, 23, 25, 27], Apache Spark [13, 14, 20, 25, 26], Apache Mahout [23, 25], Apache Kafka [26], Elasticsearch Engine [24], Pentaho [25] and TensoronSpark [20].

We also identified studies that supported a microservices architecture for big data applications [20], big data cluster resource management [13], big data integration and interoperability and data maintenance/storage technologies (Rest for Traditional Storage (RAID)) [26], messaging cloud based services such as Google Cloud Messaging (GCM) [13] and Java Message Service (JMS) [25], mobile application development framework [22].

PICOS subcategory: comparators	All studies	All studies, %
Concurrent Controls with No Visual	1	5.9
Different Users	1	5.9
Other System/Visualization	7	41.2
Other Web Portals/Social media websites	1	5.9
None	7	41.2

Table 4.
PICOS classification of all publications: Part 3. Comparison.

Twelve studies (71%) also used non-big data analytic tools and technologies, such as i2b2 platform [21], JavaScript Object Notation (JSON) [22–26], Java Database Connectivity (JDBC) [17], Matlab [23], NetBeans [22] and Python [20]. Six studies (35%) used R/R Server, and/or Shiny [13, 16–20] for their analyses and two studies reported using technologies to implement client-side solutions (Java, WebView, HTML, etc.) [22, 23].

3.4 PICOS classification: part 3: comparison

Table 4 presents the results of a PICOS Category Comparison. Seven studies (41%) provided or mentioned some description that compared them with another system and/or visualization tool [14, 18, 21, 23, 25–27].

Other studies provided comparison for concurrent controls with no visualization [19], different users [13], other web portals [11]. Seven studies (41%) did not provide or mention any comparison or comparators.

3.5 PICOS classification: part 4: outcomes

We identified six subcategories of the PICOS classification of outcomes (clinical decision support) which helped us to classify all outcomes in our studies such as system/tool type, clinical decision support purpose, clinical area support, measured outcome, clinical or patient outcomes divided into clinical and usability, outcomes effects or potential effects.

3.5.1 System/tool type

A majority of studies ($n = 16$, 94%) used decision support systems, six studies (35%) were designed for monitoring intelligent systems (sensors, devices), two (12%) used expert systems together with decision support systems [12, 25], one used a system for optimizing operations (lab resources depending on order volumes and treatments) and a decision support system [13], while only one study was developed as a visualization system [17].

3.5.2 Clinical decision support purpose

Each of the systems was designed to address a different purpose, and some were used for several purposes. We classified clinical decision support purposes into three categories of use: (1) early detection of diseases; (2) improvement of decision making; and (3) patient-centric care.

Improvement of decision making. This was the largest common purpose group ($n = 14$, 82%). The intent of these systems included utilizing diverse data to

provide automated and augmented insight, discovery, and evidence-based health and wellness decision support (n = 10, 59%) [11, 12, 14, 16, 18, 20, 22–25], followed by monitoring and assisting individuals with intelligent systems: sensors, devices and robotics, to maintain function and independence (n = 2, 12%) [14, 15], improving accuracy [15, 24], supporting clinical workflow [27] and monitoring clinical pathways [24]. Other purposes included recalling all similar patients from an institution's electronic medical record (EMR) repository, exploring “what-if” scenarios, and collecting these evidence-based cohorts for future statistical validation and pattern mining [24]. Also included were detecting inadequate treatments [21], tracking treatments [13], supporting and formulating guidelines (standards of care, expert standards, best practices) adherence to and formulating guidelines for interpreting and adjusting severity thresholds regarding the measurement of test results [11, 13].

Early detection of diseases. This was the second most frequently used common purposes group (n = 12, 71%). The major purpose was disease monitoring, such as supervising (the patients), monitoring (the clinical condition) and notifying (healthcare personnel) [12, 17, 25, 26], reporting on continuous updates on seizures [23], visualizing signal data on physician and clinician smartphone devices to analyze, inspect, and recommend the appropriate medical decisions [22, 23], assisting users in the prescription of appropriate empiric or targeted antibiotic treatments [25]. Other purposes in this group included adopting and tracking healthier behaviors [11, 21], monitoring patient safety for adverse events following medical procedures [17], identifying VAE (ventilator adverse events) risk for more rigorously identification of adverse events at an earlier stage [26], surveillance and management of antibiotics [24], health tracking [25, 27], and surgery risk assessment [20].

Patient-centric health care was supported in two studies (12%). This included planning post-discharge care coordination (which could include the follow-up visit with the primary care physician or with a subspecialist, or care on subsequent days at home by a visiting nurse) [16] and cyber-based empowering of patients and healthy individuals to play a substantial role in their own health and treatment [11].

3.5.3 Clinical area support

We further categorized our studies by clinical area they supported. We found that seven studies (41%) were used for several clinical areas [12, 16, 18, 19, 25–27].

The major supported areas were management of chronic medical conditions or preventive care (n = 7, 41%) [11, 14–16, 22, 23, 27] and management of acute medical conditions (n = 7, 41%) [12, 16–19, 26, 27]. This was followed in frequency by surgical procedures [18, 20, 24], monitoring surgical trauma intensive care units (ICU) [19, 26], pharmacotherapy (example of drug safety, infection control, rational use of antibiotics) [12, 21, 25], diagnosis [18], laboratory test ordering [13], surveillance and operational control of hospital-acquired infections or multidrug resistant infections [12, 26].

3.5.4 Were outcomes measured?

In most studies outcomes were not measured (n = 10, 59%) or reported (n = 2, 12%). Two studies reported more than one type of patient outcome measured (algorithm evaluation, time spent, etc.) [21, 27], specific tests such as oxygenation of the blood, heartbeats, daily steps, etc. [19], wound healing and/or length of stay, and time spent [12, 14].

3.5.5 Clinical and usability outcomes

Clinical outcomes. Clinical care, clinical research and patient outcomes are components of decision making and decision support processes that affect clinical outcomes. All of the studies [11–27] reported at least one clinical outcome, and most of them reported several. Fourteen studies (82%) [11–13, 16, 18–27] reported potential clinical care outcomes, such as quality of healthcare delivery and patient outcomes [18, 22–26], compliance [11, 25], adequate antibiotic use [12], identifying patients with decompensating physiology via a visual aid [18], clinician performance on clinically relevant tasks [27], accuracy (i.e. the system facilitates doctors developing preventive strategies depending on the timely and accurate identification of the greatest perioperative complication risks for patients, while the system provides accurate information such as improved calculated risk score) [20, 26]. Others included determining healthcare provider confidence (i.e. making physicians more confident that they were not missing crucial information) [27], efficiency (how to quickly detect inappropriate treatment, or help provide better care and faster response to adverse events) [21, 26]. Two studies (12%) in addition to clinical care outcomes reported that their system could be used in a research setting [11, 13]. Nearly half of the studies ($n = 8$, 47%) reported patient outcomes [14, 15, 17, 19, 20, 22, 23, 26] such as patient safety [17, 20], reduced hospital stay [19, 22, 23], earlier detection of disease, adverse events [19, 26], and time spent for each patient state [14, 15].

Usability outcomes. This refers to whether the system that was developed is easy to use, cost efficient and satisfactory for the users. Only two studies did not report usability outcomes [16, 26]. Ten studies (59%) reported effectiveness of their systems, such as gaining more knowledge [12, 18, 21–25] and decreasing errors [13], effectiveness usage to estimate changes over time or to compare the estimated risks from one hospital to another [17], and optimizing and planning resources [13]. Six studies (35%) reported efficiency outcomes, such as low costs [17, 18] and time needed on clinically relevant tasks [11, 19, 20, 26]. Almost 30% of the studies ($n = 5$) included satisfaction outcomes, such as satisfaction with system usage [23, 27], visual validity (i.e. real time and interactive visualization on physician smartphone applications) [14, 22, 23], improvement and user friendliness [16].

3.5.6 Outcomes effect

Only six studies (35%) included quantitative measures that could validate the outcome changes in their studies [12, 14, 19, 21, 26, 27].

One study [12] reported that their system increased predictive power for patient admissions (ICU) and 30% success in predicting an event 7 days before it occurs. This study reported that when the system was used, they found decreasing negative patient outcomes (death, disease, adverse reactions, infections)—50% of patient deaths were predicted within 7 days before the event occurred. Ninety-five percent of all decisions (an increase from 30%) were based on information coming from the system [12]. They concluded that monitoring antibiotic use in real-time, either from an institutional or individual perspective, immediately generated targeted interventions that led to more adequate antibiotic use.

Another study [14] reported that both doctors and normal users are getting more and more familiar with the system as a function of “disease numbers” (comorbidity) increases. A person with four diseases understands visualization outputs three times more quickly than a person with only one disease (30 s vs. 90 s). User satisfaction increases as well since they can now understand disease diagnosis

results diagrams faster (from 6 (for patient with 1 disease) to near 10 (for patient with 4 diseases) on a scale of 0–10.

In another study, the rate of septic shock in the Surgical and Trauma Intensive Care Unit (STICU) decreased by more than half after the display of the system monitor was made available to the STICU ($p < 0.05$). These results remained statistically significant even after adjusting for other control variables. The rates of respiratory failure, hemorrhage and mortality did not change significantly in either unit when comparing the periods before and after monitor display [19].

Time to detection of an inappropriate treatment decreased when using an algorithm (sorted and ordered sequences) developed for one system [21]. The study also reported that the algorithm is a very good classifier when compared with a pharmacovigilance expert (gold standard review). Another study [27] reported a non-statistically significant difference in time on clinically relevant tasks, but many others reported improved potential outcomes without any quantitative measures.

Fourteen studies (82.4%) mentioned the potential gain in knowledge for clinical decision support [11, 13, 14, 17–27], eight studies (47%) reported a potential to decrease negative patient outcomes (death, disease, adverse reactions, infections) [12, 16, 18–20, 24–26], seven studies (41%) reported improved clinical decisions based on information [12–15, 21, 23, 25]. Six studies (35%) mentioned improved/increased patient outcomes (e.g. blood sugar, decreased length of stay, saved time) [11, 16, 18, 23, 24, 26]. Three studies (18%) reported that system implementation will decrease cost of care [22, 23, 26], and one study (6%) reported increased predictive power due to system use [12].

3.6 PICOS classification: part 5: study design

The subcategories of the PICOS classification of the study design are presented in **Table 5**. Three subcategories were identified that could classify the study design, including: (1) analytics and descriptive study design, (2) study design score and (3) whether the system was a prototype or actually used in practice. Thirteen

PICOS subcategory		Description	All studies	All studies, %
STUDY DESIGN	Analytics	Analytic; Case-Crossover	1	5.9
		Analytic; Experimental;	3	17.6
		Analytic; Observational	1	5.9
	Descriptive	Qualitative	13	76.5
		Qualitative; Survey	1	5.9
		User-centered, iterative design	1	5.9
STUDY DESIGN SCORE	1 (qualitative design)	10	58.8	
	2 (quantitative descriptive design)	1	5.9	
	3 (mixed qualitative and quantitative descriptive)	6	35.3	
PRACTICE OR PROTOTYPE	Practice (used in practice)	4	23.5	
	Prototype (real data)	13	76.5	

Table 5.
 PICOS classification of all publications: Part 5: study design.

studies (77%) were identified as descriptive (qualitative) studies, although 10 (59%) of them were identified only as qualitative designs [11, 13, 16, 18, 20, 22–26], three were identified as mixed qualitative and quantitative descriptive [12, 14, 19], one was identified as descriptive qualitative with survey [15] and one was descriptive with user-centered, iterative design [27]. Five studies (29%) used analytics designs [12, 14, 17, 19, 21].

Most of the studies reviewed ($n = 13$, 77%) were prototypes but with results based on real data. Only four of the studies reported systems (23%) actually used in practice.

4. Conclusion

To provide effective patient-centered healthcare, it is essential to manage and analyze huge amounts of data. In the past decade, the variety and volume of health data sources have both increased dramatically, making traditional data management and analysis tools insufficient. Big data has emerged as a response to the growing need for health organizations to have new tools capable of processing massive amounts and varieties of healthcare data [30]. A major advantage of big data techniques is the use of advanced analysis techniques such as predictive analytics to improve clinical care, quality of care and patient outcomes.

This systematic review identified 17 studies with different visualization types and user intents, with a wide variety of data collection methods, big data platforms and tools, clinical decision support purposes to understand and synthesize existing approaches of big data knowledge visualization for clinical decision support. The results of this review emphasize the use of common types of big data knowledge visualization, patterns of big data analytics methods and classification of outcomes for clinical decision support.

The study demonstrated big differences in terms of visualization techniques used, user intents, big data tool implementations and outcome effects. Most studies reported only potential effectiveness from using knowledge visualization in clinical setting. This included gaining more knowledge for clinical decision support and improved clinical decision making based on better and more timely information, decreasing negative patient outcomes (i.e. death, disease, adverse reactions, infections) and improving patient outcomes (i.e. blood sugar, decreased length of stay and time saving), increasing predictive power of adverse events, and decreasing cost of care. Much more research is needed on implementing different techniques for big data knowledge visualization and evaluating the resulting outcomes from clinical decision support. Additional study is also needed to provide solid evidence that clinical outcomes can be improved through clinical decision support through big data knowledge visualization.

5. Limitations

Our study has three limitations. First, we only searched publications in English, and thus did not capture studies published in other languages. We also did not include commercial applications/systems which are used for clinical decision support and which may be using advanced techniques for knowledge visualization support. Lastly, papers on big data knowledge visualization but not implemented on big data platforms/tools were excluded from our review. Additional knowledge to understand big data knowledge visualization for clinical decision support could be possible through a review of such systems.

Author details

Mehrdad Roham, Anait R. Gabrielyan and Norm Archer*
DeGroote Business School, McMaster University, Hamilton, ON, Canada

*Address all correspondence to: rohamm@mcmaster.ca

IntechOpen

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Greenes RA. Definition, scope, and challenges. In: *Clinical Decision Support*. Academic Press; 2007. pp. 3-29
- [2] Munzner T. *Visualization Design and Analysis: Abstractions, Principles, and Methods* [Internet]. 2012. Available from: <http://web.cse.ohio-state.edu/~machiraju.1/teaching/CSE5544/ClassLectures/PDF-old/book.120803.pdf>
- [3] Stadler JG, Donlon K, Siewert JD, Franken T, Lewis NE. Improving the efficiency and ease of healthcare analysis through use of data visualization dashboards. *Big Data*. 2016;**4**(2):129-135
- [4] Moher D, Liberati A, Tetzlaff J, Altman DG, Altman D, Antes G, et al. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement (Chinese edition). *PLoS Medicine*. 2009;**6**(7):e1000097
- [5] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine*. 2009;**6**(7):e1000100
- [6] Higgins J, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. Cochrane Collaboration; 2011
- [7] Wu DTY, Chen AT, Manning JD, Levy-fix G, Backonja U, Borland D, et al. Review. Evaluating visual analytics for health informatics applications: A systematic review from the American medical informatics association visual analytics working group task force on evaluation. *Journal of the American Medical Informatics Association*. 2019;**26**(February):314-323
- [8] Rind BA, Wang TD, Aigner W, Miksch S, Wongsuphasawat K, Plaisant C, et al. Interactive information visualization to explore and query electronic health records and ben Shneiderman. *Foundations and Trends® in Human-Computer Interaction*. 2013;**5**(3):207-298
- [9] Yi JS, Kang Y, Stasko JT, Jacko JA. Interaction taxonomy. *IEEE Transactions on Visualization and Computer Graphics*. 2007;**13**(6):1224-1231
- [10] Hong L, Luo M, Wang R, Lu P, Wei Lu LL. Big data in health care: Applications and challenges. *Data and Information Management*. 2018;**2**(3):175-197
- [11] Chen H, Compton S, Hsiao O. DiabeticLink: A health big data system for patient empowerment and personalized healthcare. *Lecture Notes in Computer Science (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. 2013;**8040 LNCS**:71-83
- [12] Feng Z, Bhat RR, Yuan X, Freeman D, Baslanti T, Bihorac A, et al. Intelligent perioperative system: Towards real-time big data analytics in surgery risk assessment. In: *Proc - 2017 IEEE 15th Int Conf Dependable, Auton Secur Comput 2017 IEEE 15th Int Conf Pervasive Intell Comput 2017 IEEE 3rd Int Conf Big Data Intell Compu*. Vol. 2018-Jan. 2018. pp. 1254-1259
- [13] Ledieu T, Bouzille G, Plaisant C, Thiessard F, Polard E, Cuggia M. Mining clinical big data for drug safety: Detecting inadequate treatment with a DNA sequence alignment algorithm. *AMIA Annual Symposium Proceedings*. 2018;**2018**:1368-1376
- [14] Serhani MA, El Menshawy M, Benharref A, Harous S, Navaz AN. New

- algorithms for processing time-series big EEG data within mobile health monitoring systems. *Computer Methods and Programs in Biomedicine*. 2017;**149**(2017):79-94
- [15] Serhani MA, El Menshawy M, Benharref A. SME2EM: Smart mobile end-to-end monitoring architecture for life-long diseases. *Computers in Biology and Medicine*. 2016;**68**:137-154
- [16] Li P, Yates SN, Lovely JK, Larson DW. Patient-like-mine: A real time, visual analytics tool for clinical decision support. In: *Proc - 2015 IEEE Int Conf Big Data*. 2015. pp. 2865-2867
- [17] Morales A, Cánovas-Segura B, Campos M, Juárez JM, Palacios F. Proposal of a big data platform for intelligent antibiotic surveillance in a hospital. In: Oscar L, Gámez JA, Barrenechea E, Troncoso A, Galar M, Quintián H, et al., editors. *Advances in Artificial Intelligence*. Cham: Springer International Publishing; 2016. pp. 261-270
- [18] Baljak V, Ljubovic A, Michel J, Montgomery M, Salaway R. A scalable realtime analytics pipeline and storage architecture for physiological monitoring big data. *Smart Healing*. 2018;**9-10**:275-286
- [19] Hirsch JS, Tanenbaum JS, Gorman SL, Liu C, Schmitz E, Hashorva D, et al. HARVEST, a longitudinal patient record summarizer. *Journal of the American Medical Informatics Association*. 2015;**22**(2):263-274
- [20] Almeida JP. A disruptive big data approach to leverage the efficiency in management and clinical decision support in a hospital. *Porto Biomedical Journal*. 2016;**1**(1):40-42
- [21] Li D, Park HW, Batbaatar E. In: Nguyen N, Hoang D, Hong TP, Pham H, Trawiński B, editors. *Intelligent Information and Database Systems*. Vol. 10751. 2018. pp. 684-693
- [22] Galletta A, Carnevale L, Bramanti A, Fazio M. An innovative methodology for big data visualization for telemedicine. *IEEE Trans Ind Informatics*. 2019;**15**(1):490-497
- [23] Suresh S. Big data and predictive analytics: Applications in the care of children. *Pediatric Clinics of North America*. 2016;**63**(2):357-366
- [24] Caron A, Chazard E, Muller J, Perichon R, Ferret L, Koutkias V, et al. IT-CARES: An interactive tool for case-crossover analyses of electronic medical records for patient safety. *Journal of the American Medical Informatics Association*. 2017;**24**(2):323-330
- [25] Badgeley MA, Shameer K, Glicksberg BS, Tomlinson MS, Levin MA, McCormick PJ, et al. EHDViz: Clinical dashboard development using open-source technologies. *BMJ Open*. 2016;**6**(3):1-11
- [26] Abusharekh A, Stewart SA, Hashemian N, Abidi SSR. H-DRIVE: A big health data analytics platform for evidence-informed decision making. *Proc - 2015 IEEE Int Congr Big Data*. 2015; (c):416-423.
- [27] Ruminski CM, Clark MT, Lake DE, Kitzmiller RR, Keim-Malpass J, Robertson MP, et al. Impact of predictive analytics based on continuous cardiorespiratory monitoring in a surgical and trauma intensive care unit. *Journal of Clinical Monitoring and Computing*. 2019;**33**:703
- [28] Aigner W, Miksch S, Schumann H, Tominski C. *Visualization of Time-Oriented Data*. Human-Computer Interaction. Springer-Verlag London Limited; 2011

[29] von Landesberger T, Fiebig S, Bremm S, Kuijper A, Fellner DW. Interaction taxonomy for tracking of user actions in visual analytics applications. In: Huang W, editor. Handbook of Human Centric Visualization. New York, NY: Springer New York; 2014. pp. 653-670

[30] Alonso SG, de la Torre Díez I, Rodrigues JJPC, Hamrioui S, López-Coronado M. A systematic review of techniques and sources of big data in the healthcare sector. *Journal of Medical Systems*. 2017;**41**(11)