IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Volume-6, Special Issue-2, March 2018 E-ISSN: 2347-2693

A Survey on Electronic Health Records and Big Data Analytics for Healthcare

S. Thirumurugam[1*], P. Thambidurai[2]

[1]Dept. of Computer Science,Bharathidasan University, Trichy, Tamilnadu, India
[2]Professor of CSE and Principal, PKIET, Karaikal,Tamilnadu, India

Available online at: www.ijcseonline.org

Abstract- Increase in storing the Electronic Health Records (EHR) of patients has developed a large scale. Healthcare data analytics is rapidly emerging with huge potential for organizations to provide healthcare by reducing the costs and improving healthcare decisions. Analytics help in gaining the information to improve decision making by using advanced data mining tools. A healthcare information and management system uses big data analytics for operational excellence. As Electronic Healthcare records are unstructured in nature, big data adoption is gaining importance in processing and visualizing the data. Big data utilizes Hadoop framework to process the large data sets in distributing computing environment. This paper discusses the survey on the analyzation of EHR by using Big Data Analytics.

Keywords: Health care, Electronic Health Record, Big Data Analytics

1. Introduction

The concept of "big data" [1] is not new; however the way it is defined is constantly changing. Various attempts at defining big data essentially characterize it as a collection of data elements whose size, speed, type, and/or complexity require one to seek, adopt, and invent new hardware and software mechanisms in order to successfully store, analyze, and visualize the data. Healthcare is a prime example of how the three Vs of data, velocity (speed of generation of data), variety, and volume, are an innate aspect of the data it produces. This data is spread among multiple healthcare systems, health insurers, researchers, government entities, and so forth. Furthermore, each of these data repositories is siloed and inherently incapable of providing a platform for global data transparency. To add to the three Vs, the veracity of healthcare data is also critical for its meaningful use towards developing translational research.

A collection of large and complex data sets which are difficult to process using common database management tools or traditional data processing applications. "Big data refers to the tools, processes and procedures allowing an organization to create, manipulate, and manage very large data sets and storage facilities". Standard medical practice is moving from relatively ad-hoc and subjective decision making to evidence-based healthcare.

2. Electronic Health Records

MJ Ball, N Carla Smith, RS Bakalar (2007) wrote that the purpose of a patient record is "to recall observations, to inform others, to instruct students, to gain knowledge, to monitor performance, and to justify interventions [2]." The many uses described in this statement, although diverse, have a single goal—to further the application of health sciences in ways that improve the well-being of patients, including the conduct of research and public health activities that address population health. Yet, observational studies of physicians' use of the paper-based record find that logistical, organizational, and other practical limitations reduce the effectiveness of traditional records for storing and organizing an ever increasing number of diverse data. An electronic health record (EHR) [3] is designed to overcome many of these limitations, as well as to provide additional benefits that cannot be attained by a static view of events.

An electronic health record (EHR) is a repository of electronically maintained information about an individual's lifetime health status and health care, stored such that it can serve the multiple legitimate users of the record [4]. Traditionally, the patient record was a record of care provided when a patient is ill. Managed care encourages health care providers to focus on the continuum of health and health care from wellness to illness and recovery. Consequently, the record must integrate elements regarding a patient's health and illness acquired by multiple providers across diverse settings.

An electronic health record (EHR) is a repository of electronically maintained information about an individual's lifetime health status and health care, stored such that it can serve the multiple legitimate users of the record [5]. Traditionally, the patient record was a record of care provided when a patient is ill. Managed care encourages health care providers to focus on the continuum of health and health care from wellness to illness and recovery. Consequently, the record must integrate elements regarding a patient's health

and illness acquired by multiple providers across diverse settings [6].

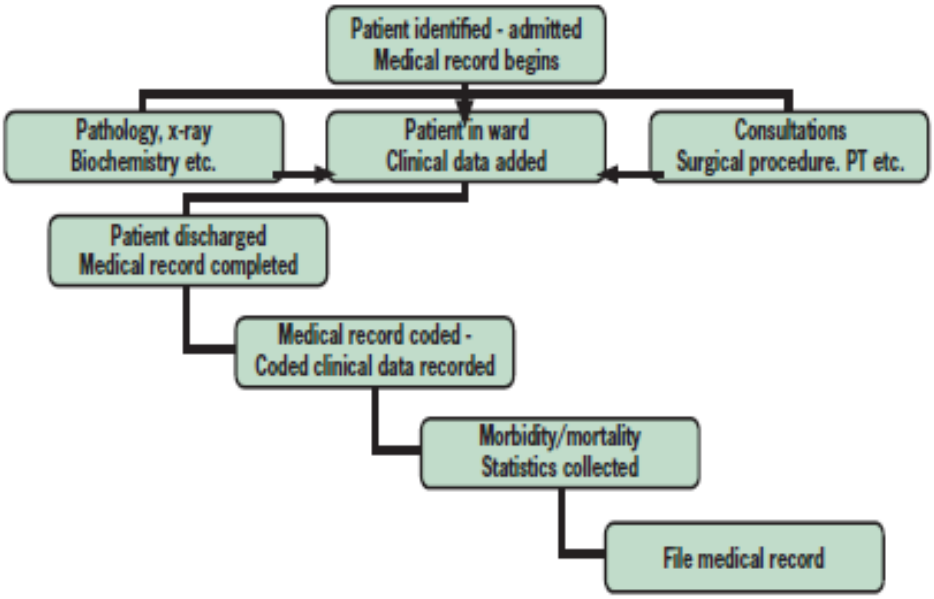

Figure 1: Manual Medical Record Information Flow

3. Big Data Analytics

Big Data technologies have already made some impact in fields related to healthcare: medical diagnosis from imaging data in medicine, quantifying lifestyle data in the fitness industry, to mention a few [7] [8]. Nevertheless, for several reasons that will be discussed in the report, the healthcare has been lagging in taking up Big Data approaches, which is a paradoxical situation, since it was already estimated by the Poneman Institute in 2012 that 30% of all the electronic data storage in the world was occupied by the healthcare industry. It is evident that within existing mounds of big data there is hidden knowledge that could change the life of a patient or, at a very large extent, change the world itself. Extracting this knowledge is the fastest, least costly and most effective path to improving peoples' health.

Big Data technologies will definitely open new opportunities and enable breakthroughs related to, among the others healthcare data analytics addressing different perspectives: (i) descriptive to answer what happened, (ii) diagnostic to answer the reason why it happened [9], (iii) predictive to understand what will happen and (iv) prescriptive to detect how we can make it happen.

4. Related Works on EHR and Big Data Analytics for Healthcare

The following table 1 depicts the related works done on the Electronic Health Record and

Big Data Analytics for Health care.

Author Name	Paper Title	Description	Findings
Jaeyong Bae, Jason M. Hockenberry, Kimberly J. Rask, Edmund R. Becker [10]	Evidence that electronic health records can promote physician counseling for healthy behaviors	This study estimates the impact of select EHR functionalities on the rate of health behavior counseling provided during primary care visits	The use of an EHR system with health information and data, order entry and management, result management, decision support, and a notification system for abnormal test results was associated with an approximately 25% increase in the probability of health behavior

				counseling delivered
Claudia Leonardi, Neal R. Simonsen, Qingzhao Yu, Chi Park, Richard A. Scribner [11]	Street Connectivity and Obesity Risk: Evidence From Electronic Health Records	This study aimed to determine the feasibility of using electronic health record (EHR) data from a federally qualified health center (FQHC) to assess the association between street connectivity, a measure of walkability for the local environment, and BMI obtained from EHRs	EHRs were successfully used to assess the relationship between street connectivity and BMI in a multilevel framework. Increasing street connectivity levels measured as intersection density were inversely associated with directly measured BMI obtained from EHRs, demonstrating the feasibility of the approach.	
Benjamin A Goldstein, Ann Marie Navar, Michael J Pencina, John PA Ioannidis [12]	Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review	Electronic health records (EHRs) are an increasingly common data source for clinical risk prediction, presenting both unique analytic opportunities and challenges. We sought to evaluate the current state of EHR based risk prediction modeling through a systematic review of clinical prediction	EHR data present both opportunities and challenges for clinical risk prediction. There is room for improvement in designing such studies.	
			studies using EHR data	
Martin C. Gulliford, Judith Charlton, Toby Prevost, Helen Booth, Alison Fildes, Mark Ashworth, Peter Littlejohns, Marcus Reddy, Omar Khan, Caroline Rudisill [13]	Costs and Outcomes of Increasing Access to Bariatric Surgery: Cohort Study and Cost-Effectiveness Analysis Using Electronic Health Records	To estimate costs and outcomes of increasing access to bariatric surgery in obese adults and in population subgroups of age, sex, deprivation, comorbidity, and obesity category	Diverse obese individuals may benefit from bariatric surgery at acceptable cost. Bariatric surgery is not cost-saving, but increased health care costs are exceeded by health benefits to obese individuals.
Yih-Ing Hser, Larissa J. Mooney, Andrew J. Saxon, Karen Miotto, Douglas S. Bell, David Huang [14]	Chronic pain among patients with opioid use disorder: Results from electronic health records data	To examine the prevalence of comorbid chronic pain among patients with opioid use disorder (OUD) and to compare other comorbidities (substance use disorder (SUD), mental health disorders, health/disease conditions) among patients in four categories: no chronic pain (No Pain), OUD prior to pain (OUD First), OUD and pain at the same time (Same Time), or pain condition prior to OUD (Pain First).	The alarming high rates of chronic pain conditions occurring before OUD and the associated severe mental health and physical health conditions require better models of assessment and coordinated care plans to address these complex medical conditions.
Juan D Chaparro, David C Classen, Melissa Danforth, David C Stockwell, Christopher A Longhurst [15]	National trends in safety performance of electronic health record systems in children's hospitals	To evaluate the safety of computerized physician order entry (CPOE) and associated clinical decision support (CDS) systems in electronic health record (EHR)	Pediatric computerized physician order entry (CPOE) systems on average are able to intercept a majority of potential

		systems at pediatric inpatient facilities in the US using the Leapfrog Group's pediatric CPOE evaluation tool.	medication errors, but vary widely among implementations. Prospective and repeated testing using the Leapfrog Group's evaluation tool is associated with improved ability to intercept potential medication errors
Julian Wolfson, David M. Vock, Sunayan Bandyopadhyay, Thomas Kottke, Gabriela Vazquez-Benitez, Paul Johnson, Gediminas Adomaviciu s, Patrick J. O'Connor [16]	Use and Customization of Risk Scores for Predicting Cardiovascular Events Using Electronic Health Record Data	Clinicians who are using the Framingham Risk Score (FRS) or the American College of Cardiology/ American Heart Association Pooled Cohort Equations (PCE) to estimate risk for their patients based on electronic health data (EHD) face 4 questions. (1) Do published risk scores applied to EHD yield accurate estimates of cardiovascular risk? (2) Are FRS risk estimates, which are based on data that are up to 45 years old, valid for a contemporary patient population seeking routine care? (3) Do the PCE make the FRS obsolete? (4) Does refitting the risk score using EHD	They conclude that published cardiovascular risk models can be successfully applied to EHD to estimate cardiovascular risk; the FRS remains valid and is not obsolete; and model refitting does not meaningfully improve the accuracy of risk estimates

		improve the accuracy of risk estimates?	
Matthew F. Daley, Douglas A. Newton, Lynn DeBar, Sophia R. Newcomer, Lisa Pieper, Joseph A. Boscarino, Sengwee Toh, Pamala Pawloski, James D. Nordin, Cynthia Nakasato, Lisa J. Herrinton, Regina Bussing [17]	Accuracy of Electronic Health Record–Derived Data for the Identification of Incident ADHD	To assess the accuracy of electronic health record (EHR)–derived diagnoses in identifying children with incident (i.e., newly diagnosed) Attention-Deficit/Hyperactivity Disorder (ADHD)	Studies predicated on the identification of incident ADHD cases will need to carefully consider study designs that minimize the likelihood of case misclassification
Martin R. Cowie, Juuso I. Blomster, Lesley H. Curtis, Sylvie Duclaux, Ian Ford, Fleur Fritz, Samantha Goldman, Salim Janmohamed, Jo¨rg Kreuzer, Mark Leenay, Alexander Michel, Seleen Ong, Jill P. Pell, Mary Ross Southworth, Wendy Gattis Stough, Martin Thoenes, Faiez Zannad, Andrew Zalewski [18]	Electronic health records to facilitate clinical research	Electronic health records (EHRs) provide opportunities to enhance patient care, embed performance measures in clinical practice, and facilitate clinical research.	Electronic health records are a promising resource to improve the efficiency of clinical trials and to capitalize on novel research approaches. EHRs are useful data sources to support comparative effectiveness research and new trial designs that may answer relevant clinical questions as well as improve efficiency and reduce the cost of cardiovascular clinical research. Initial experience

			with EHRs has been encouraging, and accruing knowledge will continue to transform the application of EHRs for clinical research
Stefan Koudstaa, Mar Pujades-Rodriguez, Spiros Denaxas, Johannes M.I.H. Gho, Anoop D. Shah, Ning Yu, Riyaz S. Patel, Chris P. Gale, Arno W. Hoes, John G. Cleland, Folkert W. Asselbergs, and Harry Hemingway [19]	Prognostic burden of heart failure recorded in primary care, acute hospital admissions, or both: a population-based linked electronic health record cohort study in 2.1 million people	The prognosis of patients hospitalized for worsening heart failure (HF) is well described, but not that of patients managed solely in non-acute settings such as primary care or secondary outpatient care. We assessed the distribution of HF across levels of healthcare, and assessed the prognostic differences for patients with HF either recorded in primary care (including secondary outpatient care) (PC), hospital admissions alone, or known in both contexts	In the general population, one in four patients with HF will not be hospitalized for worsening HF within a median follow-up of 1.7 years, yet they still have a poor 5-year prognosis. Patients admitted to hospital with worsening HF but not known with HF in primary care have the worst prognosis and management. Mitigating the prognostic burden of HF requires greater consistency across primary and secondary care in the identification, profiling, and treatment of patients.
Yichuan Wang, LeeAnn Kung, William Yu Chung Wang, Casey G. Cegielski [20]	An Integrated Big Data Analytics-Enabled Transformation Model:Application to Health Care	A big data analytics enabled transformation model based on practice-based view is developed which reveals the causal relationships among big data analytics capabilities, IT-enabled transformation practices, benefit dimensions and business value. This model was then tested in healthcare setting	One challenge in the health care industry is that their IT adoption usually lags behind other industries. Case organizations studied in this paper are "leaders" in their own rights. They are either top-ranked research hospitals or associated with top medical schools with resources, or highly profitable entities.
Hemingway H, Feder G S, Fitzpatrick N K, Denaxas S, Shah A D & Timmis A D. [21]	Using nationwide 'big data' from linked electronic health records to help improve outcomes in cardiovascular diseases: 33 studies using methods from epidemiology, informatics, economics and social science in the ClinicAl disease research using LInked Bespoke studies and Electronic health Records (CALIBER) programme	Electronic health records (EHRs), when linked across primary and secondary care and curated for research use, have the potential to improve our understanding of care quality and outcome. To evaluate new opportunities arising from linked EHRs for improving quality of care and outcomes for patients at risk of or with coronary disease across the patient journey.	Emerging 'big data' opportunities arising from the linkage of records at different stages of a patient's journey are vital to the generation of actionable insights into the diagnosis, risk stratification and cost-effective treatment of people at risk of, or with, CVD.
Christian R. Macedonia, Clark T. Johnson, Indika Rajapakse [22]	Advanced Research and Data Methods in Women's Health- Big Data Analytics, Adaptive Studies, and the Road Ahead	Technical advances in science have had broad implications in reproductive and women's health care. Recent innovations in population-level data collection and storage have	Examination of big data research examples contained in this article provides insight into the potential and the limitations of this data

		made available an unprecedented amount of data for analysis while computational technology has evolved to permit processing of data previously thought too dense to study. "Big data" is a term used to describe data that are a combination of dramatically greater volume, complexity, and scale. The number of variables in typical big data research can readily be in the thousands, challenging the limits of traditional research methodologies	science revolution and practical pathways for its useful implementation.
Scott M. Sutherland a Stuart L. Goldstein b Sean M. Bagshaw [23]	Leveraging Big Data and Electronic Health Records to Enhance Novel Approaches to Acute Kidney Injury Research and Care	While acute kidney injury (AKI) has been poorly defined historically, a decade of effort has culminated in a standardized, consensus definition. In parallel, electronic health records (EHRs) have been adopted with greater regularity, clinical informatics approaches have been refined, and the field of EHR-enabled care improvement and research has burgeoned.	The combination of a standard AKI definition, the pervasive growth of EHR adoption, and the development of novel informatics tools has created a unique set of circumstances capable of transforming AKI-related care and research. Application of the KDIGO criteria to the EHR dataset allows
		Although both fields have matured in isolation, uniting the 2 has the capacity to redefine AKI-related care and research. This article describes how the application of a consistent AKI definition to the EHR dataset can accurately and rapidly diagnose and identify AKI events. Furthermore, this electronic, automated diagnostic strategy creates the opportunity to develop predictive approaches, optimize AKI alerts, and trace AKI events across institutions, care platforms, and administrative datasets	accurate and timely diagnosis of AKI events
Vera Ehrenstein, Henrik Nielsen, Alma B Pedersen, Søren P Johnsen, and Lars Pedersen [24]	Clinical epidemiology in the era of big data: new opportunities, familiar challenges	Routinely recorded health data have evolved from mere by-products of health care delivery or billing into a powerful research tool for studying and improving patient care through clinical epidemiologic research. Big data in the context of epidemiologic research means large interlinkable data sets within a single country or networks of	Big data will also provide new possibilities for research by enabling access to linked information from biobanks, electronic medical records, patient-reported outcome measures, automatic and semiautomatic electronic monitoring devices, and social media. The sheer

		multinational databases	amount of data, however, does not eliminate and may even amplify systematic error. Therefore, methodologies addressing systematic error, clinical knowledge, and underlying hypotheses are more important than ever to ensure that the signal is discernable behind the noise.

V Conclusion

This review paper carried out the related works done on the Electronic Health Record (EHR) and Big Data Analytics in Healthcare. Accurate patient identification is the backbone of an effective and efficient health record system, whether manual or electronic. From the works, the following research issues and challenges are in HER: Inferring knowledge from complex heterogeneous patient sources. Leveraging the patient/data correlations in longitudinal records, Understanding unstructured clinical notes in the right context, Efficiently handling large volumes of medical data and extracting potentially useful information and biomarkers. • Analyzing data is a computationally intensive task and combining with standard clinical data adds additional layers of complexity. Capturing the patient's behavioral data through several sensors; their various social interactions and communications. Access control to ensure health records are available when needed for patient care and other official purposes but may not be accessed by unauthorized persons. Audit controls where access may be monitored to ensure only authorized persons use the system and to identify when changes are occurred.

VI FUTURE DIRECTION

The following works are the future direction of the research work. To improve the accuracy, analyzation of records can be done by classifying the records into district wise or zonal wise or state wise. The records are taken according to the age criteria by using Map Reduce operation. Considering the Lab/ Haemotology reports of the patients to classify the patients into three categories by means of Minimum, Maximum and Average. Finally, the decision support system is developed to take the decision based on the reports and to alert the doctor as well as patient according to their physical state. Finally, the decision support system is developed to take the decision based on the reports and to alert the doctor as well as patient according to their physical state.

References

[1] Xiaofeng, Meng, and Chi Xiang. "Big data management: concepts, techniques and challenges." *Journal of computer research and development* 1.98 (2013): 146-169.

[2] Ball, Marion J., N. C. M. N. Carla Smith, and Richard S. Bakalar. "Personal health records: empowering consumers." *J Healthc Inf Manag* 21.1 (2007): 77.

[3] Pizziferri, Lisa, et al. "Primary care physician time utilization before and after implementation of an electronic health record: a time-motion study." *Journal of biomedical informatics* 38.3 (2005): 176-188.

[4] Tang, Paul C., and Clement J. McDonald. "Electronic health record systems." *Biomedical informatics* (2006): 447-475.

[5] Tang, Paul C., and Clement J. McDonald. "Electronic health record systems." *Biomedical informatics* (2006): 447-475.

[6] Hoffman, Sharona, and Andy Podgurski. "E-Health hazards: provider liability and electronic health record systems." *Berkeley Technology Law Journal* (2009): 1523-1581.

[7] Zhang, Yin, et al. "Health-CPS: Healthcare cyber-physical system assisted by cloud and big data." *IEEE Systems Journal* 11.1 (2017): 88-95.

[8] M. Durairaj, T.S. Poornappriya, "A Review on Big Data Analytics Tools for Telecommunication Industry,"*International Journal of Control Theory and Applications*, Volume 9, Issue 27, pp.185-193, 2016.

[9] Gale, Nicola K., et al. "Using the framework method for the analysis of qualitative data in multi-disciplinary health research." *BMC medical research methodology* 13.1 (2013): 117.

[10] Jaeyong Bae, Jason M. Hockenberry, Kimberly J. Rask, Edmund R. Becker, "Evidence that electronic health records can promote physician counseling for healthy behaviors", *Health Care Manage Rev*, 42(3), pp. 258-268, 2017.

[11] Claudia Leonardi, Neal R. Simonsen, Qingzhao Yu, Chi Park, Richard A. Scribner, "Street Connectivity and Obesity Risk: Evidence From Electronic Health Records", *American Journal of Preventive Medicine*, 52(1S1):pp. S40–S47, 2017.

[12] Benjamin A Goldstein, Ann Marie Navar, Michael J Pencina, John PA Ioannidis, "Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review", *Journal of the American Medical Informatics Association*, pp.1-11, 2016.

[13] Martin C. Gulliford, Judith Charlton, Toby Prevost, Helen Booth, Alison Fildes, Mark Ashworth, Peter Little johns, Marcus Reddy, Omar Khan, Caroline Rudisill, "Costs and Outcomes of Increasing Access to Bariatric Surgery: Cohort Study and Cost-Effectiveness Analysis Using Electronic Health Records", *Value in Health*, pp. 85-92, 2017.

[14] Yih-Ing Hser, Larissa J. Mooney, Andrew J. Saxon, Karen Miotto, Douglas S. Bell, David Huang, "Chronic pain among patients with opioid use disorder: Results from electronic health records data", *Journal of Substance Abuse Treatment*, pp. 26–30, 2017.

[15] Juan D Chaparro, David C Classen, Melissa Danforth, David C Stockwell, Christopher A Longhurst, "National trends in safety

performance of electronic health record systems in children's hospitals", *Journal of American Medical Informatics Association*, pp. 1–8, 2016.

[16] Julian Wolfson, David M. Vock, Sunayan Bandyopadhyay, Thomas Kottke, Gabriela Vazquez-Benitez, Paul Johnson, Gediminas Adomavicius, Patrick J. O'Connor, "Use and Customization of Risk Scores for Predicting Cardiovascular Events Using Electronic Health Record Data", *Journal of American Heart Association*, pp. 1-18, 2017.

[17] Matthew F. Daley, Douglas A. Newton, Lynn DeBar, Sophia R. Newcomer, Lisa Pieper, Joseph A. Boscarino, Sengwee Toh, Pamala Pawloski, James D. Nordin, Cynthia Nakasato, Lisa J. Herrinton, Regina Bussing, "Matthew F. Daley, Douglas A. Newton, Lynn DeBar, Sophia R. Newcomer, Lisa Pieper, Joseph A. Boscarino, Sengwee Toh, Pamala Pawloski, James D. Nordin, Cynthia Nakasato, Lisa J. Herrinton, Regina Bussing, "Accuracy of Electronic Health Record–Derived Data for the Identification of Incident ADHD", *Journal of Attention Disorders*, pp. 1-10, 2014.

[18] Martin R. Cowie, Juuso I. Blomster, Lesley H. Curtis, Sylvie Duclaux, Ian Ford, Fleur Fritz, Samantha Goldman, Salim Janmohamed, Jo¨rg Kreuzer, Mark Leenay, Alexander Michel, Seleen Ong, Jill P. Pell, Mary Ross Southworth, Wendy Gattis Stough, Martin Thoenes, Faiez Zannad, Andrew Zalewski, "Electronic health records to facilitate clinical research", *Clin Res Cardiol*, pp. 1-9, 2016.

[19] Stefan Koudstaa, Mar Pujades-Rodriguez, Spiros Denaxas, Johannes M.I.H. Gho, Anoop D. Shah, Ning Yu, Riyaz S. Patel, Chris P. Gale, Arno W. Hoes, John G. Cleland, Folkert W. Asselbergs, and Harry Hemingway, "Prognostic burden of heart failure recorded in primary care, acute hospital admissions, or both: a population-based linked electronic health record cohort study in 2.1 million people", Euro*pean Journal of Heart Failure*, pp. 1-9, 2016.

[20] Yichuan Wang, LeeAnn Kung, William Yu Chung Wang, Casey G. Cegielski, "An Integrated Big Data Analytics-Enabled Transformation Model: Application to Health Care", *Information & Management*, pp. 1-50,2018.

[21] Hemingway H, Feder G S, Fitzpatrick N K, Denaxas S, Shah A D & Timmis A D, "Using nationwide 'big data' from linked electronic health records to help improve outcomes in cardiovascular diseases: 33 studies using methods from epidemiology, informatics, economics and social science in the ClinicAl disease research using LInked Bespoke studies and Electronic health Records (CALIBER) programme", *Programme Grants for Applied Research*, Volume: 5, Issue 4, 2017.

[22] Christian R. Macedonia, Clark T. Johnson, Indika Rajapakse, "Advanced Research and Data Methods in Women's Health- Big Data Analytics, Adaptive Studies, and the Road Ahead", *Clinical Expert Series*, Obstetrics & Gynecology, pp. 249-264, Volume 129, Number 2, 2017.

[23] Scott M. Sutherland, Stuart L. Goldstein, Sean M. Bagshaw, "Leveraging Big Data and Electronic Health Records to Enhance Novel Approaches to Acute Kidney Injury Research and Care", *Blood Purification*, Issue 44, pp. 68-76, 2017.

[24] Vera Ehrenstein, Henrik Nielsen, Alma B Pedersen, Søren P Johnsen, and Lars Pedersen, "Clinical epidemiology in the era of big data: new opportunities, familiar challenges", *Clinical Epidemiology*, 9: pp. 245–250, 2017.

