The Power of Big Data Mining to Improve the Health care System in the United Arab Emirates

Khawla Eissa Alhajaj
Mohammed Bin Rashid School of Government
Immanuel Azaad Moonesar (✉ immanuel.moonesar@mbrsg.ac.ae)
Mohammed Bin Rashid School of Government

Case Report

Keywords: Big Data, Healthcare System, UAE, Health Informatics

Posted Date: May 16th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1601389/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Journal of Big Data on February 1st, 2023.
See the published version at https://doi.org/10.1186/s40537-022-00681-5.

Abstract

Context

Collecting and analyzing data has become crucial for many sectors, including the health care sector, where a hefty amount of data is generated daily. Over time, the amount and complexity of this data increase substantially. Consequently, it is considered as *big data* that cannot be stored or analyzed conveniently unless advanced technologies are incorporated. Latest advances in technology have divulged new opportunities to use big data analysis to track a patient's record and their health. Still, it has also posed new challenges in maintaining data privacy and security in the healthcare sector.

Purpose

This study aims to determine the importance of big data and health informatics in the health care sector and to identify the opportunities related to big data mining in health care. Furthermore, the study focuses on various challenges posed by big data mining in the health care system and explores the tools and technologies for securely harvesting big data in this sector.

Methods

The research follows the methodology of PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) for reporting the reviews and evaluating the randomized trials. Furthermore, to evaluate each article in terms of quality, the Critical Appraisal Checklist for PRISMA 2009 was applied for the research.

Findings

The study concludes that the healthcare systems in the UAE can be improved through the use of big data; however, the authorities within the UAE must acknowledge that the development of efficient frameworks for the performance and quality assessment of the new health care system is significant. The said goal can be achieved via the integration of big data and health informatics with the help of IT specialists, health care managers and stakeholders.

Contribution to Knowledge

By discussing numerous issues and presenting solutions linked with big data, the current study makes a substantial contribution to the knowledge of big data and its integration into health care systems in the UAE.

1 Introduction

Technology has become a vital part of our daily lives, and the adoption of new technologies has drastically altered our daily lives. Technology has made our lives more comfortable and efficient,

regardless of our age, level of knowledge, or even the reason for utilizing it. We are experiencing technological wonders in the shape of smartphones, Internet of Things (IoT), robotic surgery, and applications that use Artificial Intelligence (AI). We are shifting from regular technology use to more complicated technologies which are connected via a powerful internet and frequently generate vast amounts of data. With the growth of data from numerous mobile networks, cloud computing systems, health applications, and electronic medical records, there is an increased need for a comprehensive approach to maintaining and updating information. The expanding data and information of the patients and the relevant health care activities are getting more challenging due to the speed, amount, and complexity of the data. Muni Kumar and Manjula (2014) reported that the health care facilities generate an abundant amount of data each day that is centered around the patients, medicines, treatments, diseases, research and other similar factors. To manage this data more efficiently, modern health care units choose to digitize the data related to patients. Worldwide, medical institutes shifted from the traditional paper-based medical file to the electronic medical record, providing help in managing patient information, lab tests, medications, and medical imaging.

Electronic Medical Records (EMR) are considered to be an essential, rich platform containing patient information. EMR captures all demographic data, lab results, radiology images and free-text notations. That collective information is extremely useful as a database for many longitudinal studies. Mining data from EMR can help understand disease signs and symptoms and the progression of a particular disease. It also offers improvement in clinical knowledge and understanding of a specific phenomenon and assists in clinical trials and disease management and therapeutic trials (Coorevits, 2013; Effoe, 2016). Further, it assists in predicting disease progression, comorbidities and mortalities (Paxton, 2013; Benjamin, 2017).

Data comes from various sources, including electronic medical files, home sensors, and wearable devices. As such, it will generate a massive amount of data known as *big data*. *Big data* refers to massive data, although the term has no universally accepted definition. The oldest definition is provided by Laney (2001), who observed that (big) data was growing in three different dimensions, namely: volume, velocity, and variety (known as the *three V's*). This definition has been expanded by Demchenko, Zhao, Grosso, Wibisono & De Laat (2012), who define big data by *five V's*: volume, velocity, variety, veracity and value. Volume refers to the amount of data, which is massively generated and requires a unique storage format. Data velocity means the high speed of data generated from different resources. Variety of data means the complexity of datathat varies from numerical data to text notation or in the form of, from numerical data to text notation or a (radiological) image. Finally, vacity refers to the accuracy of the data and value evaluates the quality of data (Srikanth Thudumu, 2020).

Wang (2016) expands on the idea of big data and defines it as a set of data that cannot be analyzed by a standard computerized method. Big data is segregated in the type of structured, unstructured and semi-structure forms of the data. The structured data can be stored, accessed and processed in a specific format. It is an already-segregated and dedicated form of easily retrievable and readable data. The unstructured data is not specific in its form, as it was discussed for the structured form of the data. As

stated by Wu and Lin (2018), this type of data possesses multiple challenges in terms of processing as well as retrieving valuable information from it. A typical example of this form of data is the data that comes from heterogeneous sources, meaning a combination of text files, images and videos. Data heterogeneity is due to the mixing of structured and unstructured data, having its roots in various platforms that are either quantitative or qualitative. The quantitative sources of data include laboratory tests, images, sensor data and gene array. The qualitative data sources include demographics and textual information (Shelton, Poorthuis, Graham & Zook, 2014). One of the critical challenges in this regard is related to the accuracy and trustworthiness of the data since the credibility of the data may be challenged as it is from unmanaged sources. To get the ultimate benefits out of big data technology, health care systems need to analyze the unstructured and semi-structured data coherently. The extraction and retrieval of big data may be subject to challenges related to social and legal technicalities. These social and legal issues might be generated due to problems associated with data ownership, privacy, identification and governance (Mittelstadt & Floridi, 2016).

1.1 Big Data and Health care System within the UAE

The UAE health care system is operated by government-funded health services and the rapidly growing private health sector. The standards of the health services provided by both sectors are acceptable. The healthcare industry of the UAE is realizing the potential of big data analysis and has the capability to transform the health care system. According to Bani-issa, Eldeirawi & Al Tawil (2014), such developments are the inactive lifestyle among the residents, leading to an increase in chronic diseases such as diabetes. Several regions in the Middle East, including the UAE, have undergone or are considering implementing health care insurance, which then needs to perform an analysis of the big volume of health data generated from claims. The UAE introduced a standardized insurance coding system to deal with the situation and improve process efficiency. The insurers in the UAE are pricing premiums based on little historical data due to the lack of big data analysis tools and the sophisticated nature of the big data. The availability of big data will enable insurers to paint a clear picture of health care in the region. It will allow them to accurately predict the validity of the claims (Hamidi et al., 2014).

The UAE vision is to provide world-class healthcare by 2021 and the government's direction is to foster innovation in the healthcare system to be able to achieve its vision. Many strategies have been explored to ensure that people are provided with a high-quality care system and to ensure the implementation of SDGs, particularly Goal 3 (ensuring healthy lives and promoting well-being for all ages) (UAE Government, 2020).

With advanced technology in the UAE, and smart government and public service, big data helps in providing a big database within the country, especially in the healthcare sector, which can assist in a better understanding of population health and provide the required service. To be able to achieve the government vision of providing world-class healthcare and ensure maintaining sustainability in delivering health and well-being to everyone living in the UAE, this can be facilitated by mining big data and

understanding the data to be able to provide better services and health plans to ensure a healthier, happier community. Despite the availability of big data in the UAE and the potential to utilize big data as a government looking for innovation and using big data, there are limitations and a lack of published research on big data mining in the UAE, in different sectors and especially the health care system. Although, there is no standardized government approach and policy regarding big data mining or storing. However, there is a big amount of data generated on a daily basis among different entities within the UAE. In terms of consensus, the UAE open data policy was launched in 2018 as per the UN eGovernment Survey (2018) to help access data without restriction. By 2020, not all data is accessible and there remains restrictions on available data from the entities (2021 UAE, 2021).

A lot of research in the literature review provides agreement about big data mining and its beneficial role toward enhancing the health care system, yet there is still no unified process or solutions for big data mining and how to make it possible. This research will help understand the importance of harnessing big data and utilizing it to enhance the health care system and identify the challenges and limitations of harnessing big data.

1.2 Big Data and Sustainable Development Goals

The UN developed a 2030 strategy to fight poverty, ensure equity among people and address the global challenges through 17 sustainable goals (United Nations, 2015). Policymakers, decision-makers, and investors, according to Wu (2018), need authentic, accurate, and real-time data to adopt the proper policies and decisions in order to accomplish the Sustainable Development Goals (SDGs). They then need to be able to check the impact of the policy, which can be achieved through the analysis of big data from different sources. Similarly, as stated in a report released by United Nations (2018), the big data revolution can contribute to SDG by providing accurate and reliable data and analyzing and analyzing the data to develop policy and plans to achieve SDGs 2030. The main concerns were about the inadequacy in technology adoption among all countries, and data privacy and transparency. Big data analysis can provide the ability to monitor the progress toward achieving SDGs by 2030. Big data analysis can be more cost-effective and faster in tracking SDGs than, for example, tracking poverty by traditional methods such as by questionnaires or interviews, which can be ineffective and time-consuming as well as require significant effort (Blumenstock, Cadamuro & On, 2015; Steele et al., 2017).

The focus on SDGs is stated in Goal 3, to “ensure healthy lives and promote well-being for all at all ages” (United Nations, 2015). Big data can help in providing precise and clear information about health. Barrett, Humblet, Hiatt & Adler, (2013) makes this point by better adopting big data analysis to understand population behavior and social and environmental factors. This will help in population health management and prevent the disease and target subpopulations by having accurate and real-time data. Big data analysis can help achieve SDGs by promoting well-being and chronic disease prevention through big data analysis.

1.3 Research Gap

As discussed in the literature, big data is emerging as a great source of improvement in different sectors of the world, especially for the countries that are adopting advanced health care systems. Most developed countries have recognized the importance of big data and have shown interest in improving the health care system through the collection and analysis of big data (Catalyst, 2018; Pastorino at al., 2019).The UAE is an example of a nation whose healthcare systems are up to date and equipped with modern health facilities. However, there is limited research in this context that have considered, that despite existing challenges of data security, data classification, data modeling, data storage, data accommodation and technology incorporation, whether the integration of big data and health care can emerge as a sustainable system. In countries similar to the UAE, with a high population and complex health care systems, the implementation of big data analysis will be a challenging task. The current study will focus on these challenges related to big data and the health care system, focusing on the context of the UAE.

This research study aims to gain further insight using a systematic approach to review the role and effectiveness of big data in the area of health care within the UAE. The objectives of the study are:

1. To investigate the role of big data in the health care system.
2. To identify opportunities to enhance quality-of-care services through integrating big data in the health care system.
3. To understand the challenges related to implementing and using big data technologies.

2 METHODOLOGY

The current research follows the methodology of PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) for reporting the reviews and evaluating the randomized trials. According to Tricco et al. (2018), to assess the quality of the selected research articles for current research, the critical appraisal PRISMA checklist 2009 was used (Moher, Liberati,Tetzlaff, Altman & PRISMA Group, 2009). Due to limitations in publication or access control, the limitation of access to a wide range of journals will limit the overview of available literature during the study time. Restriction to published papers in the English language will restrict the information to publish a valuable thesis in other languages.

2.1 Inclusion Criteria

The article searching methodology discussed above resulted in thousands of research articles. Instead of considering all of them, the most relevant research articles were segregated from those that were less relevant. The titles, abstracts and keywords of the research articles were screened and those articles which discussed the relation of big data to health care systems were separated for full-text review. The

screening process was made more efficient by removing duplicate research articles at the eligibility phase of PRISMA.

A systematic review of the studies was conducted and the articles were judged according to their objectives, methodology and study design, use of authentic data sources, validity and reliability of the study method, analysis and consideration of ethical issues. Further, comprehensiveness of description of the findings and outcomes, the appropriateness of the tools used for data mining, the suitability of the qualitative methodology, the use of valid research designs, and clearly stated research findings were considered in the review. Linkage to the current study questions and objective was taken in consideration. Only one study was conducted on big data within the context of the UAE. The study used quantitative data to better understand the topic, with the same context of the UAE as the current study.

2.2 Exclusion Criteria

The different research articles were excluded for consideration in the research based on the criteria of quantitative studies, surveys, focus groups from the research other than health care, feasibility studies, work environments other than health care, data collection techniques adopted, editorials and short reports, and articles which were not reputably published, such as in international journals. The assessment of the research articles based on the aforementioned criteria further narrowed the number of articles to be included in the current research and it was ensured that the remaining articles are the most relevant and high-quality manuscripts, which make the findings and outcomes of the research authentic and most reliable. Quantitative data were excluded as well. The research that made mention of computerized or digital tools to analyze big data and technical details about the transfer of big data, processing of big data, storing the data, cleaning the data and analyzing the data were excluded as well as these are not one included in the research objectives and technical details will not add value to current study. Artificial Intelligence algorithms and the role of big data in Artificial Intelligence studies were further excluded.

2.3 Quality Assessment and Processing Steps

To evaluate each article in terms of quality, the Critical Appraisal PRISMA Checklist 2009 was applied in the current research for qualitative studies. In addition, the quality of articles was judged according to their objectives, as was the approach of the research methods, objectives and abstract use of authentic data sources. The validity and reliability of the research design approaches adopted were evaluated. There was also a consideration of ethical issues. A thorough evaluation was done of the comprehensive description of findings and outcomes, and of the appropriateness of the qualitative methodology, use of valid research designs, and clearly stated findings of the research., The use of valid research designs and clearly stated findings of the research were analyzed. The flow diagram of the PRISMA methodology adopted for the research is illustrated in Figure 2. Extended results attached in appendix (I)

3 Results And Discussion

3.1 Big Data in Health Care

Big data is now considered the gold standard of the new technological era, especially for an institute that encounters great quantities of data daily, such as in the health care sector. Considering the importance of big data in healthcare, the content is analyzed to showcase how data is utilized in the health care sectors of the UAE and the certain challenges faced by the healthcare industry in this regard.

Big data can improve the efficiency and effectiveness of health strategy and policy. They can shift the policymaking from the patient visit to more advanced (value-based) policies with big data analysis, which can accurately reflect the population, as Gamache (2018) mentioned. It was also noted by Auffray et al. (2016) that big data can help in formulating a prevention and prediction strategy, in addition to improvement in the health of the population. Big data can help to improve the progress of individual health and shift toward personalized medicine. It can enable the predication of personal health and improve clinician decisions (Leyens, Reumann, Malats & Brand, 2017). In addition, big data can assist in a clinical trial to choose suitable participants and make the process less expensive. Big data can also give more insight into drug safety, early detection and trace adverse drug side effects. Big data can also help monitor trends of infectious disease and track the cases among the population, which can assist in making the right decision and act immediately to limit the spread of contagious disease.

It is mentioned by Zeng (2015) that big data integration with different behavioral and social factors can help in better understanding of health care and health disparities, to enhance the population health and reduce such disparities. Zeng (2015) discusses the vital two areas where big data can play an essential role in the health care system: it can help integrate and understand social factors impacting population health to improve the population's quality of care.

The USA HITECH Act helps in the adoption of the EMR, which contributes to the generation of data and reduces health disparities. The same applies to the context of the UAE where an EMR is adopted among various authorities in the public and private sectors. The EMR is considered as a new opportunity to create a vast amount of data to better understand other demographic factors and their effects on a large population. Still, there is an issue related to the standardization of medical notes worldwide, which play an essential role in a better understanding and analysis of data. Big data analysis can improve health care delivery and reduce unnecessary costs. Big data can help better understand disease progression, the side effect of medications and ensure the health care delivers equally among the population. In addition to EMR data, integration with other devices such as home monitoring devices and smartphone applications may better insight into the population data.

Big data can improve public health surveillance and address disparities in the health care system. In the USA, the implementation of the Affordable Care Act and widespread insurance play an essential role in expanding the health care system and enhance the accessibility of health care. The aforementioned factors help acquire extensive data from the population, across different socioeconomic backgrounds, to

better understand additional population requirements and health status and formulate policy and strategy based on population requirements. This approach will help in better resource allocation and enhance the health care system by utilizing big data analysis.

Geographic Information Systems (GISs) can help better understand the population at risk, health requirements and make the right intervention based on real-time data through big data analysis. Another example provided by Gamache (2018) uses the GIS to allocate vaccines targeting a particular population in response to an outbreak in a specific geographic area.

Social media data can play an essential role in understanding and monitoring population behavior and the infectious disease spread. For example, Young, Rivers & Lewis (2014) analyzed 553,186,061 tweets and found a correlation between the prevalence of HIV and the geographic location of HIV-related tweets. This can be expanded more by knowing and analyzing the current status, and utilizing the data for future prediction of cases by using big data modeling, analysis of population behavior, and linkage to social media data. Future projections can help in making the right intervention to reduce the spread of infections, which will reduce the cost of the disease burden on the community and health care system.

Zeng (2015) mentioned that big data modeling could be more accurate than traditional methods. Big data modeling can include predictive modeling to forecast a disease occurrence and forecast complications related to the particular disease, which can help in a better interventionnd enhance the health outcome and reduce the disease progression.

3.2 Challenges related to Big Data

There are many challenges identified in the literature concerning the application of big data in the health care system. The study performed by Auffray et al. (2016) and Zeng (2015) highlighted some of the key challenges concerning the development of utilizing big data. The most common areas of concern were data privacy, data storage, data structure, data ownership and governance.

3.2.1 Data Security

Health care data are considered sensitive and most of the literature agree that big data raise a security issue. It is considered a challenge to process big data, as several researchers in the USA mentioned. The Health Insurance Portability and Accountability Act of 1996 (HIPAA) law prevent handling the client's data without prior consent. The same applies within the context of the UAE Article 379 of the UAE Penal Code, which requires prior permission to handle client’s data. A simple de-identification process was not the final solution as mentioned by Adibuzzaman, DeLaurentis, Hill, Benneyworth (2017) The De-identification still cannot protect client data within the ID process. Further, it is easy to re-identify the person by other location or demographic information. Data related to health issues are considered highly sensitive and private. This is why specific regulation regarding accessibility and the availability of the data with consideration to the customer is required and prior consent mandatory, even if the data were to be coded and de-identified

3.2.2 Data Storage

Data storage has been a concern in many studies due to data security. The legal issue within different countries and accessibility of data is limited to research and governments and is without a clear policy related to data storage and accessibility. Big data cannot be stored by ordinary means, especially from different sources, including electronic medical records, monitoring devices, images, and lab tests.

Initiatives and solutions of data storage were discussed by Adibuzzaman et al. (2017), including the platform Informatics for Integrating Biology and the Bedside (i2b2). It is a platform of more than 100 hospitals where patient data is de-identified and stored for research purposes since the hospital should use another software to transfer data. The author argues that this initiative does not permit patient to access their data. Moreover, the hospital required much effort to de-identify the data and used a particular platform to transfer data. If the same system is to be applied, an additional budget is required and the system will be limited to structured data. It is also mentioned by Dash, Shakyawar, Sharma & Kaushik (2019) that data stored at the same time it is generated is less, compared to if the data is transferred using another system. Most of the evaluated authors agreed that structured data storage and analysis is more manageable than transferal to another platform. Cloud-based storage remains the ideal solution, yet the security issue remains a challenge to overcome.

3.2.3 Missing Data and Unstructured Data

One of the common challenges found in the literature review is missing data in the electronic health record, unstructured data and free-text data, which is very difficult to process and analyze. The study conducted by Zeng (2015) mentioned electronic health records lack social or behavioral data and that there was no standardization of the data format, leading to disparities in health information. This can prevent or make big data analysis impossible. In addition, lack of data standardization makes data transfer and data acquisition is impossible, as stated by Eysenbach (2016).

Most of the electronic health records are designed to make the diagnosis coding and, billing easier, and to make the information more explicit but these are not yet advanced enough to make the analysis and data linkage useful, as mentioned by Gamache (2018). One of the examples of the failure of EMR for data analysis is Medical Information Mart for Intensive Care (MIMIC III) which has collected data for more than 50,000 patients from Beth Israel Deaconess Hospital dating back to 2001. The researcher aimed to conduct studies to answer different questions, such as the drug-drug interaction between antihistamine and antidepression. When they applied the selection criteria and checked the files, they ended up with a minimal sample size, which statistically is not representative. This issue can also apply to EMRs in the UAE. There are multiple EMR systems in the UAE with different software, and there is no standardization of medical record notes and fragmented systems between local and federal authorities. This will make data transfer and analysis is complicated, which should be taken into consideration.

3.2.4 Data Ownership and Governance

Health care information is considered as sensitive data. The ownership and sharing of data is an unclear and negotiable challenge among different countries and has been mentioned by many studies (Auffray et al., 2016). The privacy policy across Europe varies, and there is no approach to big data sharing that can fit the existing policies in other countries. The EC for general data protection regulation (2012/0011COD) tried to synchronize the fragmented health care system to be able to make the data available and useable among different European countries. One suggestion was to share data across a block-chain where all the transactions would be recorded, and the data accessibility would be monitored, however, this is not yet adopted as security was a concern. This would not be the case in the UAE as the policy and federal law is unified, and data control and accessibility could be maintained at the federal level.

Auffray et al., (2016) mentioned that the USA is a “(p)atient-driven economy” where patients own their data. This is a step forward and a valid approach for a fragmented health care system where different healthcare systems are present, as in the case of Europe and even the UAE. This would help the patient own their digital data, but it requires a digital infrastructure and storage system to ensure the data is transferred into a cloud. It also required that it is in a structured format to access data, understand the data and easily analyze the data. This approach can help enhance health tourism and a health-driven economy.

Adibuzzaman et al. (2017) mentioned data should be Findable, Accessible, Interoperable and Reproducible (FAIR). In addition, the data would need to be stored as open source, where researchers, stakeholders and even patients have access to those data, while also ensuring data protection and privacy. As an example, data storage in a protected environment after a proper de-identification of clients' ID to maintain a privacy law.

4 Conclusion And Future Recommendations

The UAE government remains aware of the power of big data, as shown by their establishment of the UAE Strategy for Artificial Intelligence. The Dubai government policy framework is intended to develop and implement a culture of data sharing and evidence-based decision-making in Dubai (U.ae, 2019). The study concludes that the healthcare systems in the UAE can be enhanced through big data; however, the authorities within the UAE must acknowledge that the development of efficient frameworks for the performance and quality assessment of the new health care system is significant. The said goal can be achieved via the integration of big data and health informatics with the help of IT specialists, health care managers, and stakeholders.

4.1 Recommendations to use Big Data in the Health Sector in the UAE

4.1.1 Specific Recommendation for Big Data Handling in the Health care Sector

- Formulate a unified EMR standardization for the medical note to be able to process medical note data and transfer it quickly.
- Incentivization of the health care provider to ensure they are following high standardization of EMR, as data entry is very important and this was a significant challenge in previous studies.
- Public-private partnership is essential and there should be incentivization of the private sector to share their data. The UAE has the big challenge of population diversity and most of the population are not seeking medical care in the government health care sectors. To overcome population discrepancy and ensure the data represents the whole population, data from the private sector should be accessible as well.
- Agree on data needed for real-time monitoring, such as infectious diseases or surveillance that needs intervention. The big data analysis will simplify and quicken the intervention decision, consequently offering a better response from authorities to an emergency.
- Utilize big data to take a proactive measurement and to ensure timely involvement of stakeholders to prevent a disease occurrence in the case of non-communicable diseases. That will ultimately reduce the cost of management and the burden of chronic diseases.

4.2 Conclusion & Recommendations for Future Research

The digital revolution has arrived, and it is impacting everyday life. There is a dire need to utilize existing health care system-related data which is automatically generated daily. This big data can transform the healthcare system to improve patient care, proactively envisaging disease origins and implementing timely solutions to bridge the gaps in the existing healthcare system. It is time to think about innovative, technological solutions to link and analyze data faster, understand the disparities between community health and public health alongside enhancing the overall health care system, and implement SDG Goal 3, ensuring healthy lives and promoting well-being for all ages.

Declarations

Funding

The authors disclose receipt of the following financial support for the authorship, research, and/or publication of this article: The authors would like to acknowledge Mohammed Bin Rashid School of Government (MBRSG), Dubai, UAE, and the Alliance for Health Policy and Systems Research at the World Health Organization for financial support as part of the Knowledge to Policy (K2P) Center Mentorship Program [BIRD Project].

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Ethical approval

Studies involving animal subjects: No animal studies are presented in this manuscript.

Studies involving human subjects: No human studies are presented in this manuscript.

Inclusion of identifiable human data: No potentially identifiable human images or data is presented in this study.

Author contributions

KH & IAM: Made a substantial contribution to all the sections and participated in the review, analysis, and interpretation. Involved in drafting the manuscript and revising it critically for important intellectual content. KH: Made a substantial contribution to study design. Participated in review, analysis, and interpretation. KH: Made a considerable contribution to background and method sections. KH: Made a significant contribution to the background and discussion sections. KH & IAM: Made a significant contribution to the discussion and conclusion sections. IAM: Made a substantial contribution to background, methods, and discussion sections. All authors give final approval for the version to be published and agree to be accountable for all aspects of the work.

References

1. U.ae . (2021). *The United Arab Emirates Government portal*. Retrieved 1 19, 2021, from https://u.ae/en/information-and-services/g2g-services/open-government-data
2. Adibuzzaman, M., DeLaurentis, P., Hill, J., & Benneyworth, B. D. (2017). Big data in healthcare–the promises, challenges and opportunities from a research perspective: A case study with a model database. In AMIA Annual Symposium Proceedings (Vol. 2017, p. 384). American Medical Informatics Association.
3. Auffray, C., Balling, R., Barroso, I., Bencze, L., Benson, M., Bergeron, J., Bernal-Delgado, E., Blomberg, N., Bock, C., Conesa, A. and Del Signore, S., (2016). Making sense of big data in health research: towards an EU action plan. Genome medicine, 8(1), 1-13.
4. Bani-issa, W., Eldeirawi, K., & Al Tawil, H. (2014). Perspectives on the attitudes of healthcare professionals toward diabetes in community health settings in United Arab Emirates. *Journal of Diabetes Mellitus, 5(*01), 1.
5. Goldstein, B. A., Navar, A. M., Pencina, M. J., & Ioannidis, J. (2017). Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *Journal of the American Medical Informatics Association, 24(*1), 198-208.
6. Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. Science, 350(6264), 1073-1076.
7. Catalyst, N. E. J. M. (2018). Healthcare big data and the promise of value-based care. *NEJM Catalyst*, *4*(1).

8. Coorevits, P., Sundgren, M., Klein, G.O., Bahr, A., Claerhout, B., Daniel, C., Dugas, M., Dupont, D., Schmidt, A., Singleton, P. and De Moor, G., (2013). Electronic health records: new opportunities for clinical research. *Journal of internal medicine,* 274(6), 547-560.
9. Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: management, analysis and future prospects. *Journal of Big Data*, *6*(1), 1-25.
10. Demchenko, Y., Zhao, Z., Grosso, P., Wibisono, A., & De Laat, C. (2012, December). Addressing big data challenges for scientific data infrastructure. In *4th IEEE International Conference on Cloud Computing Technology and Science Proceedings* (pp. 614-617). IEEE.
11. Effoe, V.S., Katula, J.A., Kirk, J. et al. (2016) The use of electronic medical records for recruitment in clinical trials: findings from the Lifestyle Intervention for Treatment of Diabetes trial. Trials 17, 496.https://doi.org/10.1186/s13063-016-1631-7
12. Gamache, R. K. (2018). Public and population health informatics: the bridging of big data to benefit communities. *Yearbook of medical informatics, 27*(1), 199.
13. Leyens, L., Reumann, M., Malats, N., & Brand, A. (2017). Use of big data for drug development and for public and personal health and care. *Genetic epidemiology, 41(*1), 51-60.
14. Barrett, M. A., Humblet, O., Hiatt, R. A., & Adler, N. E. (2013). Big data and disease prevention: from quantified self to quantified communities. *Big data*, *1*(3), 168-175.
15. Mittelstadt, B. D., & Floridi, L. (2016). The ethics of big data: current and foreseeable issues in biomedical contexts. *Science and engineering ethics, 22*(2), 303-341.
16. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group*. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*, 151(4), 264-269.
17. Muni Kumar, N., & Manjula, R. (2014). Role of Big data analytics in rural health care-A step towards svasth bharath. *International Journal of Computer Science and Information Technologies, 5*(6), 7172-7178.
18. Pastorino, R., De Vito, C., Migliara, G., Glocker, K., Binenbaum, I., Ricciardi, W., & Boccia, S. (2019). Benefits and challenges of Big Data in healthcare: an overview of the European initiatives. *European journal of public health*, *29*(Supplement_3), 23-27.
19. Paxton, C., Niculescu-Mizil, A., & Saria, S. (2013). Developing predictive models using electronic medical records: challenges and pitfalls. AMIA ... Annual Symposium proceedings. AMIA Symposium, 2013, 1109–1115.
20. Shelton, T., Poorthuis, A., Graham, M., & Zook, M. (2014). Mapping the data shadows of Hurricane Sandy: Uncovering the sociospatial dimensions of 'big data'. *Geoforum*, *52*, 167-179.
21. Srikanth Thudumu, P. B. (2020). A comprehensive survey of anomaly detection techniques for high dimensional big data. *Journal of Big Data, 7*(42). doi:https://doi.org/10.1186/s40537-020-00320-x
22. Steele, J.E., Sundsøy, P.R., Pezzulo, C., Alegana, V.A., Bird, T.J., Blumenstock, J., Bjelland, J., Engø-Monsen, K., De Montjoye, Y.A., Iqbal, A.M. and Hadiuzzaman, K.N.,. (2017). Mapping poverty using mobile phone and satellite data. *Journal of The Royal Society Interface*, *14*(127), 20160690.

23. Tricco, A.C., Lillie, E., Zarin, W., O'Brien, K.K., Colquhoun, H., Levac, D., Moher, D., Peters, M.D., Horsley, T., Weeks, L. and Hempel, S., 2018. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Annals of internal medicine*, *169*(7), pp.467-473.
24. UAE (2019). *U.ae*. Retrieved 12 11, 2020, from https://u.ae/en/about-the-uae/strategies-initiatives-and-awards/local-governments-strategies-and-plans/dubai-data-strategy
25. UAE GOVERNMENT. (2020). *Vision 2021 and health*. Retrieved April 05, 2021, from https://u.ae/en/information-and-services/health-and-fitness/vision-2021-and-health
26. United Nations. (2015). SDG, The Sustainable Development Goals,Retrieved April 05, 2020, from https://sdgs.un.org/goals
27. United Nations. (2018). *UN World Data Forum 2018 wraps up with launch of Dubai Declaration*. Retrieved April 05, 2020, from https://www.un.org/sustainabledevelopment/blog/2018/10/un-world-data-forum-2018-wraps-up-with-launch-of-dubai-declaration/
28. Wang H, X. Z. (2016). Towards felicitous decision making: an overview on challenges and trends of Big Data. *Inf Sci.* , 367–368:747–765.
29. Wu, P. a. (2018). Unstructured big data analytics for retrieving e-commerce logistics knowledge. . *Telematics and Informatics, 35*(1), 237-244.
30. Young, S. D., Rivers, C., & Lewis, B. (2014). Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. *Preventive medicine*, 63, 112–115. https://doi.org/10.1016/j.ypmed.2014.01.024
31. Zeng, X., Zhang, Y., Kwong, J.S., Zhang, C., Li, S., Sun, F., Niu, Y. and Du, L., (2015). The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. *Journal of evidence-based medicine*, *8*(1), 2-10.

Figures

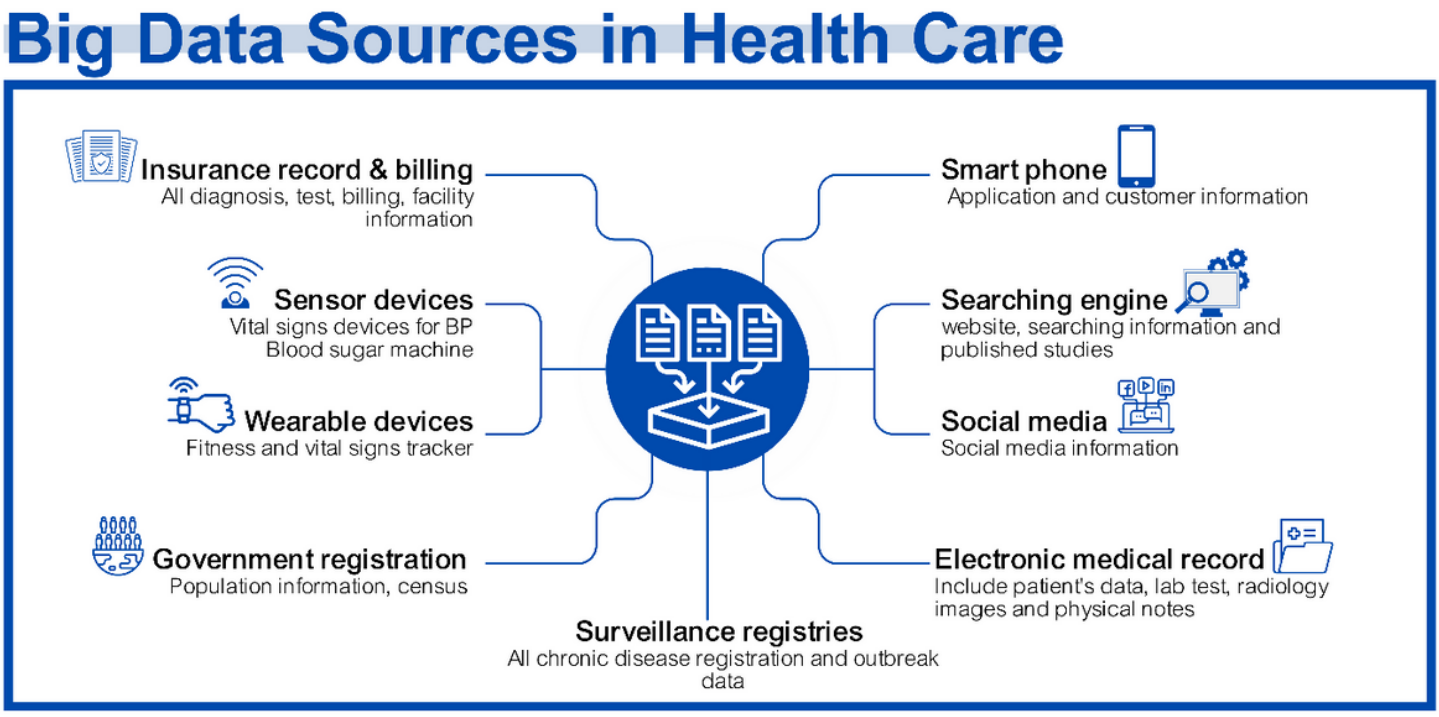

Figure 1

Please See image above for figure legend.

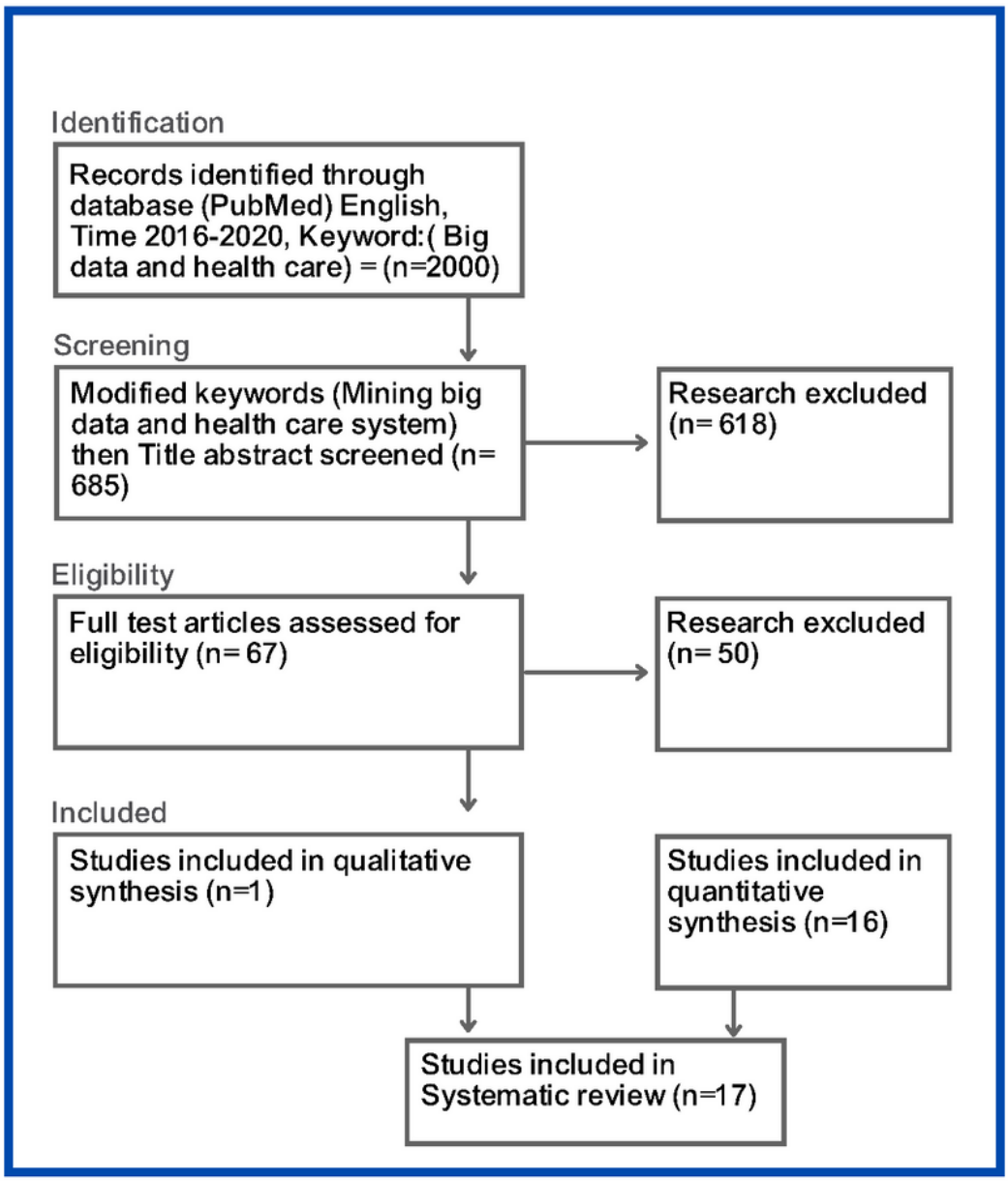

Figure 2

Please See image above for figure legend.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AppendixIPRISMAchecklist.docx