Application of Big Data in Public Health

Muxi Zheng

University of California, Irvine, CA, USA

Keywords: Big data, Public health, Stakeholders, Governance frameworks and regulations

Abstract: Big Data is the massive information from various sources that can be analyzed and is characterized by data growing in dimensions. It can be applied in the field of public health. The application of Big Data in the public health domain would affect the whole population, especially in this era of global public health. Meanwhile, there are challenges in implementing Big Data in public health surveillance. Therefore, there must be appropriate governance frameworks and regulations in the use of Big Data to ensure the safe storage and use of personal data. The government plays an important role in it. In this paper, the application of Big Data in Public Health field is been discussed with both the advantages and challenges. In the end of the paper, the countermeasures and suggestions are also given as references for further study.

1. Introduction

Big Data is the massive information from various sources that can be analyzed and is characterized by data growing in dimensions including volume, velocity, and variety, surpassing traditionally used amounts of storage (Laney et al.). When coupled with artificial intelligence algorithms, the information can be used for decision making and prediction of the future (Benke and Benke). Big Data can be applied in the field of public health, which is the “art or science of preventing disease, prolonging life, and promoting health and efficiency through organized community effort” (Winslow). Public health surveillance is an essential tool in public health to understand the burden in healthcare and to inform actions to offer better overall health to the community (“Surveillance Strategies for Improvement”). This is a growing field in healthcare that would rely largely on the Big Data and technology of AI (Thiébaut et al.).

2. The Application of Big Data

The population can be benefited in many ways through the advance of technology. This could include better population health from surveillance with sources like social media that can collect massive amounts of information, and thus enable events like seasonal flu and epidemics to be better tracked and contained to ensure the well-being of more people (Strachan). With the implementation of Big Data in healthcare facilities, there would be improvements in the consistency of care (Strachan). For example, the comparison in medical imaging with the technology of Big Data would increase the precision in treatment and diagnosis, which would be more reliable than depending on the experience of a physician, and this is very likely to benefit the population on an individual level (Hulsen et al.). Also, uses of Big Data would benefit the research field, where diseases could be better studied with large samples and data collected from variable sources and would benefit many people from more information on health-related issues (Hulsen et al.).

A case of using Big Data for surveillance purposes is the use of contact tracing applications on mobile during the COVID-19 pandemic. Companies including Google and Apple have launched contact tracing software, which uses Bluetooth signals to connect with other users. The information of

DOI: 10.25236/cseem.2020.021

contact would be sent to local health facilities if the individual tests positive, which acts as a surveillance tool for detecting cases of COVID-19 (Graham). Experts have positive attitudes toward the use of such applications, suggesting that this tool would help prevent the spread of the virus and save lives (Perrigo). Although government officials have stated that the applications would de-identify personal information and the data would be deleted if the software is deleted, it is still hard to build trust in the public while there has been a history of the data breach (Perrigo). As there are larger volumes of data, a single breach could expose a massive amount of information, which has occurred in the past (Groot). Also, there is a question about whether such tools for surveillance should be coercion. While the use of such technology can potentially provide information to preserve public health and stop the spread of the virus, it is questionable whether the government can force people to share their data for the public good (Lanzing).

Previously, the common practice for analysis and storage of patient information used physical or typed data, and in 2003, electronic health records first started to be put into use (Laney et al.). According to the CDC, the modernization of implementing informatics into the surveillance system started in 2014, with strategies such as encouraging the use of electronic reporting and statistical systems (“Surveillance and Data Strategy - Notable Milestones”). Later in 2015 and 2016, there was the development of using open data and enhancement in cloud-based platforms in data analysis and storage, which helped in identifying health threats as well as monitoring (“Surveillance and Data Strategy - Notable Milestones”). In 2018, new strategies have been proposed in emphasizing the use of data from non-traditional sources, the use of new technologies, and communication among different stakeholders (“Improving Disease Surveillance”). Current researches on surveillance focus on areas including mobile health, social media, and population health, with the discussion of Big Data and new technologies such as artificial intelligence (Gamache et al.). The characteristics of Big Data allow the implication into the public health domain, including the high volume of data, the high velocity which allows computation and communication, and also highly variable as the data is collected from various sources (Fig 1), such as from traditional electronic health records as well as from search engine data (“Healthcare Big Data and the Promise of Value-Based Care”).

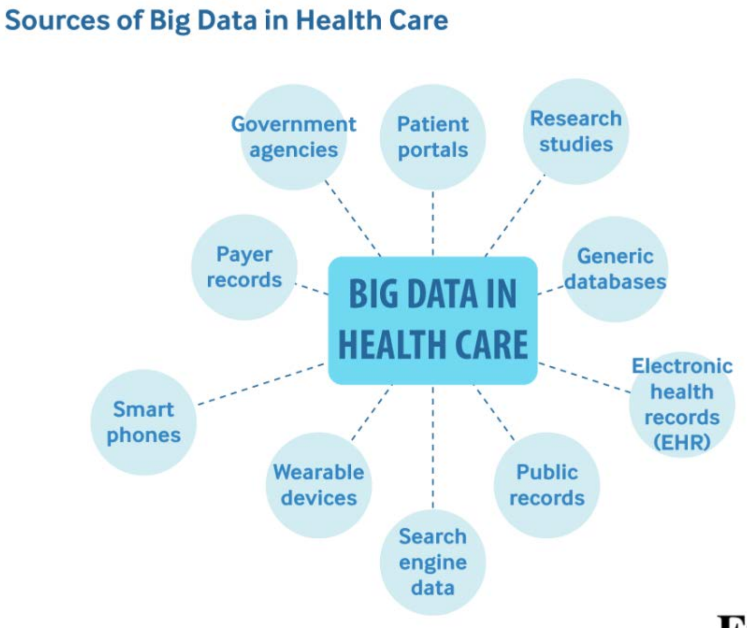

Fig 1

NEJM Catalyst (catalyst.nejm.org) © Massachusetts Medical Society

Fig.1 Sources of Big Data in Health Care

Machine learning is a technology that is vital in the application of Big Data in public health. With the massive amount of data, machine learning can help researchers analyze the existing data (M Bublitz et al.). The benefit of using such technology has already been shown in many cases, such as that AI algorithms can extract useful information from noise efficiently, which helps in providing integrated information for public health (Brownstein et al.).

As the sources of data used in surveillance are expanding, social media has become a source that is believed to contribute to public health greatly in the future. There can be many applications with the data from population-based sources, including communication surveillance and situational awareness; for epidemiologic monitoring, data from social media can be used to detect disease and analyze the pattern of the disease using artificially intelligent algorithms(Fung et al.).

3. The Impact of Big Data in Public Health

3.1 The Advantages

While Big Data is still a new field in public health, it is suggested that information from computer chips on objects such as cars, watches, and other common objects might be better implemented with an internet connection, which can generate large amounts of health-related data that can be used in surveillance (Laney et al.). This “Internet of Things” is believed to contribute to public health surveillance in monitoring the spread of disease as well as individual health status (Laney et al.). The potential advantages also include real-time and continuous acquisition of patient information, which can support the monitoring and tracking of diseases like chronic illness (Meinert et al.).

The application of Big Data in the public health domain would affect the whole population, especially in this era of global public health. The new technology would likely improve the overall health in low- and middle-income countries, where there is currently a need for good information feedback mechanisms in the public health domain (Wyber et al.). With Big Data, there might be more access to health information from sources like mobile devices, which could stimulate better coverage of surveillance and improve the overall quality of health surveillance (Wyber et al.).

An important issue related to the use of health records and Big Data is The Health Insurance Portability and Accountability Act (HIPAA), which was first established on August 21, 1996 (“HIPAA History”). This act was first introduced to improve the accountability of health coverage through approaches including promoting the use of medical savings accounts (“HIPAA History”). Throughout history, there have been major additions to this act related to securing health information. The Privacy and Security Rules were introduced in 2006, where officials including the Department of Health and Human Services and the Department's Office for Civil Rights were given the right to enforce the rules (“HIPAA History”). In 2009, the Health Information Technology for Economic and Clinical Health Act (HITECH) was established as an extension of the HIPAA with the goal of implementing the use of Electronic Health Records in healthcare (“HIPAA History”). This included incentivizing protection of data of patient’s personal data and reporting of data breaches, and the act was also applied to third party suppliers to the healthcare industry (“HIPAA History”). In 2013, there was the implementation of The Final Omnibus Rule, which cleared up the uncertainties in the previous amendments, taking into account technological advances such as data collection from mobile devices (“HIPAA History”). Although the HIPAA has been amended several times, there are still grey areas regarding which organizations are covered entities, which could still cause problems in the use of data, especially in the era of Big Data (“What Is a HIPAA-Covered Entity?”). According to the Department of Health and Human Services, the use of data for public health purposes would be allowed without authorization to public health authorities, while the HIPAA regulation is still relevant in preventing the breach of personal information of patients (“Public Health”).

3.2 The Challenges

There are challenges in implementing Big Data in public health surveillance. One of the challenges is fragmentation. This is when data is not able to communicate with each other, which prevents the integration of Big Data to provide useful information. This might be caused by different forms of storage and data used by different stakeholders, which suggests the need for global standardization and consensus to provide a better platform for Big Data (Agrawal and Prabakaran). Data ownership and sharing are also a challenge, and this leads to the lack of regulation to manage the data which can also cause problems in data sharing (Agrawal and Prabakaran). Regarding data collection from population-based sources, there might be challenges in ensuring the quality of information collected, and it also is hard to obtain data with high reliability since the authors online cannot be identified (Moorhead et al.).

The debate around the issue of using Big Data in public health surveillance is around ethical considerations. Even though regulations like HIPAA have been established, there are still concerns about the privacy and security of data due to the grey areas in regulating the use of personal information. The concerns of privacy include if personal data would be collected without being acknowledged when using online tools, whether personal data is de-identified sufficiently, and whether the data would be used properly in ethical research (Mooney and Pejaver). One example of the debate can be seen in a study by Facebook where the company manipulated the news feed and measured the mood change in different groups of the population (Meyer). This used Big Data and the research was on the populational level, and the users agreed to the term of use, which gave the company access to their information (Hu). This has led to debates over the ethical issues, including whether informed consent would be necessary for such a study. Some suggest that this study is appropriate from the regulation perspective, and the study itself serves the purpose of science and beneficence; others disagree with the collection of personal data for research purposes even though it was mentioned in the term of use (Meyer). This might also be an ethical issue related to the collection of data for surveillance in public health.

4. Conclusion and Countermeasures

There must be appropriate governance frameworks and regulations in the use of Big Data to ensure the safe storage and use of personal data, and close monitoring would also be required (Wyber et al.). This can be a barrier for implementing Big Data related technologies in low- and middle-income countries, where there might be poor infrastructure preventing them from having the ability to best ensure the confidentiality and privacy of data, leading to ethical questions (Wyber et al.).

In the future, the stakeholders for Big Data in healthcare would likely include not just the healthcare system, but also other entities that can provide related services, such as businesses and development agencies (Vayena et al.). Policies would be needed to provide clear guidelines about the exchange of data and to ensure the dynamic relationships (Fig 2) among all the stakeholders involved (Vayena et al.). There also need to be regulations to prevent the misuse of personal data, and issues including to what extent can personal lives be monitored for the greater good and the line between preserving public health and personal freedom need to be addressed (Vayena et al.).

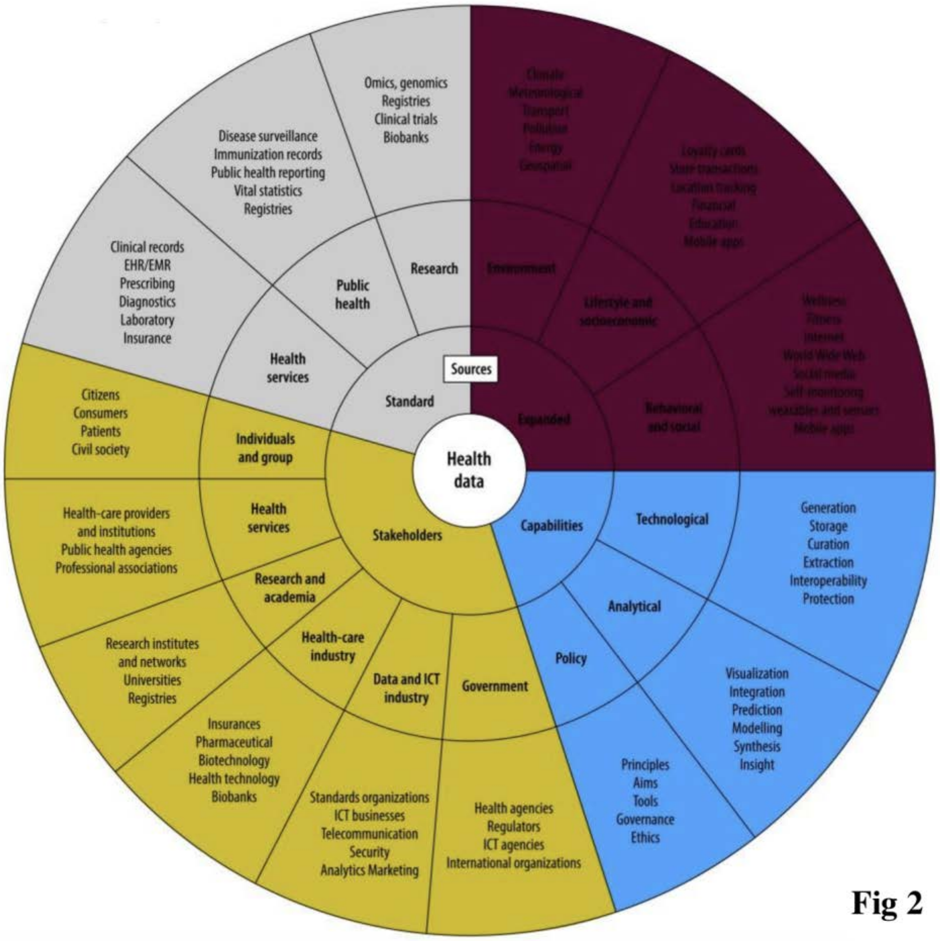

Fig 2

Fig.2 Dynamic Relationships in Health Data

Despite the researches have shown a promising future in public health surveillance with the aid of Big Data, clear regulations with consideration on ethics would be needed in implementing on a large scale in the field of global surveillance (Choi).

References

[1] Agrawal, Raag, and Sudhakaran Prabakaran. “Big Data in Digital Healthcare: Lessons Learnt and Recommendations for General Practice.” Nature News, Nature Publishing Group, 5 Mar. 2020, www.nature.com/articles/s41437-020-0303-2.

[2] Benke, Kurt, and Geza Benke. “Artificial Intelligence and Big Data in Public Health.” International Journal of Environmental Research and Public Health, MDPI, 10 Dec. 2018, www.ncbi.nlm.nih.gov/pmc/articles/PMC6313588/.

[3] Brownstein, John S, et al. “Surveillance Sans Frontières: Internet-Based Emerging Infectious Disease Intelligence and the HealthMap Project.” PLoS Medicine, Public Library of Science, 8 July 2008, www.ncbi.nlm.nih.gov/pmc/articles/PMC2443186/.

[4] Choi, Bernard C K. “The Past, Present, and Future of Public Health Surveillance.” Scientifica, Hindawi Publishing Corporation, 2012, www.ncbi.nlm.nih.gov/pmc/articles/PMC3820481/.

[5] Fung, Isaac Chun-Hai, et al. “The Use of Social Media in Public Health Surveillance.” Western Pacific Surveillance and Response Journal : WPSAR, World Health Organization, 26 June 2015, www.ncbi.nlm.nih.gov/pmc/articles/PMC4542478/.

[6] Gamache, Roland, et al. “Public and Population Health Informatics: The Bridging of Big Data to Benefit Communities.” Yearbook of Medical Informatics, Georg Thieme Verlag KG, Aug. 2018, www.ncbi.nlm.nih.gov/pmc/articles/PMC6115205/.

[7] Graham, Jefferson. “Tracking Coronavirus: Are Apple and Google Contact Tracing Apps Available in Your State?” USA Today, Gannett Satellite Information Network, 5 Oct. 2020, www.usatoday.com/story/tech/2020/10/02/apple-google-coronavirus-contact-tracing-apps/3592355001/.

[8] Groot, Juliana De. “The History of Data Breaches.” Digital Guardian, 5 Oct. 2020, digitalguardian.com/blog/history-data-breaches.

[9] “Healthcare Big Data and the Promise of Value-Based Care.” NEJM Catalyst, 1 Jan. 2018, catalyst.nejm.org/doi/full/10.1056/CAT.18.0290.

[10] “HIPAA History.” HIPAA Journal, 19 May 2020, www.hipaajournal.com/hipaa-history/.

[11] Hu, Elise. “Facebook Manipulates Our Moods For Science And Commerce: A Roundup.” NPR, NPR, 30 June 2014, www.npr.org/sections/alltechconsidered/2014/06/30/326929138/facebook-manipulates-our-moods-for-science-and-commerce-a-roundup.

[12] Hulsen, Tim, et al. “From Big Data to Precision Medicine.” Frontiers in Medicine, Frontiers Media S.A., 1 Mar. 2019, www.ncbi.nlm.nih.gov/pmc/articles/PMC6405506/.

[13] “Improving Disease Surveillance.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 30 Aug. 2018, www.cdc.gov/surveillance/projects/Improving-Disease-Surveillance.html.

[14] Laney, D., et al. “Big Data in Healthcare: Management, Analysis and Future Prospects.” Journal of Big Data, SpringerOpen, 1 Jan. 1970, link.springer.com/article/10.1186/s40537-019-0217-0.

[15] Lanzing, Marjolein. “Contact Tracing Apps: an Ethical Roadmap.” Ethics and Information Technology, Springer Netherlands, 29 Sept. 2020, www.ncbi.nlm.nih.gov/pmc/articles/PMC7524380/.

[16] M Bublitz, Frederico, et al. “Disruptive Technologies for Environment and Health Research: An Overview of Artificial Intelligence, Blockchain, and Internet of Things.” International Journal of Environmental Research and Public Health, MDPI, 11 Oct. 2019, www.ncbi.nlm.nih.gov/pmc/articles/PMC6843531/.

[17] Meinert, Edward, et al. “The Internet of Things in Health Care in Oxford: Protocol for Proof-of-Concept Projects.” JMIR Research Protocols, JMIR Publications, 4 Dec. 2018, www.ncbi.nlm.nih.gov/pmc/articles/PMC6299230/.

[18] Meyer, Robinson. “Everything We Know About Facebook's Secret Mood Manipulation Experiment.” The Atlantic, Atlantic Media Company, 9 Sept. 2014, www.theatlantic.com/technology/archive/2014/06/everything-we-know-about-facebooks-secret-mood-manipulation-experiment/373648/.

[19] Mooney, Stephen J, and Vikas Pejaver. “Big Data in Public Health: Terminology, Machine Learning, and Privacy.” Annual Review of Public Health, U.S. National Library of Medicine, 1 Apr. 2018, www.ncbi.nlm.nih.gov/pmc/articles/PMC6394411/.

[20] Moorhead, S Anne, et al. “A New Dimension of Health Care: Systematic Review of the Uses, Benefits, and Limitations of Social Media for Health Communication.” Journal of Medical Internet Research, JMIR Publications Inc., 23 Apr. 2013, www.ncbi.nlm.nih.gov/pmc/articles/PMC3636326/.

[21] Perrigo, Billy. “Will COVID-19 Contact Tracing Apps Protect Privacy?” Time, Time, 9 Oct. 2020, time.com/5898559/covid-19-contact-tracing-apps-privacy/.

[22] “Public Health.” HHS.gov, US Department of Health and Human Services, 16 June 2017, www.hhs.gov/hipaa/for-professionals/special-topics/public-health/index.html.

[23] Strachan, Meredith. Big Data Means Big Benefits for Healthcare Providers and Patients. 22 Sept. 2020, www.trapollo.com/big-data-means-big-benefits-for-healthcare/.

[24] “Surveillance and Data Strategy - Notable Milestones.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 6 Aug. 2018, www.cdc.gov/surveillance/improving-surveillance/milestones.html.

[25] “Surveillance Strategies for Improvement.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 7 Aug. 2018, www.cdc.gov/surveillance/surveillance-data-Strategies/index.html.

[26] Thiébaut, Rodolphe, et al. “Artificial Intelligence in Public Health and Epidemiology.” Yearbook of Medical Informatics, Georg Thieme Verlag KG, Aug. 2018, www.ncbi.nlm.nih.gov/pmc/articles/PMC6115208/.

[27] Vayena, Effy, et al. “Policy Implications of Big Data in the Health Sector.” Bulletin of the World Health Organization, World Health Organization, 1 Jan. 2018, www.ncbi.nlm.nih.gov/pmc/articles/PMC5791870/.

[28] Winslow, C.-E A. The Untilled Fields of Public Health. S.n., 1920.

[29] “What Is a HIPAA-Covered Entity?” HIPAA Journal, 8 Apr. 2019, www.hipaajournal.com/hipaa-covered-entity/.

[30] Wyber, Rosemary, et al. “Big Data in Global Health: Improving Health in Low- and Middle-Income Countries.” Bulletin of the World Health Organization, World Health Organization, 1 Mar. 2015, www.ncbi.nlm.nih.gov/pmc/articles/PMC4339829/.