CHAPTER 9

BIG DATA AND MEDICAL SOCIOLOGY RESEARCH

[1]Kolawole Lateef KAZEEM, [2]Musediq Olufemi LAWAL

[1]Fountain University, Nigeria,
[2]Osun State University, Nigeria

Chapter Overview

In this chapter we shall offer a brief overview of big data challenges in medical sociological research. Our contribution will reflect on the subject of big data in a general context and the field of social sciences and medical sociology in particular. We shall discuss this title under 7 themes with the following learning objectives: explain what big data is in general context including the field of social sciences and sociology of health in particular, describe the concept of big data and the determinants of its various applications; describe the applications of big data in social research; discuss issues related to big data in medical sociology; offer an explanation of how big data can be used to improve the health research outcomes.

GLOBAL HEALTH INSIGHT

Health sociologists are expanding their perspectives beyond border thanks to data analytics. Through meticulous analysis of cross-cultural data, social health experts are gaining invaluable insights into worldwide health disparities and similarities. They are stepping into the future of health with better knowledge.

Health experts are embracing the transformative synergy of Big Data and Health sociological research to elevate medical practices, drive groundbreaking research, and influence global health policies.

Introduction

The vital and effective tool of research is advancing humankind. There would have been very little advancement without organized research. Science has been credited with advancing society. The ever-increasing significant research in the physical, biological, social, and psychological fields has led to the development of new products, facts, concepts, and methods of doing things (Diaz-Bone et al., 2020, Gray, 2010). Research can lead to the discovery of information that is accurate, real, true, and genuinely helpful. According to Namanji and Ssekyewa (2012), studies benefit both researchers and the general public because the data and findings are trustworthy and supported by science. With solutions in place, the underlying causes of a social issue could be better understood. New findings and ways of thinking in social research keep it relevant as a source of knowledge. It enlightens the researcher about hidden truths, dispels ignorance, and offers fresh guidance for social interaction. The choice of information sources for a study depends on the researcher's understanding of the kind of information required (Salganik, 2018).

Most information is made up of data, which is a group of facts that have been given context. The data might be unprocessed and disorganized. Data points are unique and occasionally unrelated. After being organized and processed, data becomes information. A broad perspective of how everything fits together is provided with this data transformation into information, allowing for meaningful application for developmental purposes. The necessity of data in human beings' daily activities is thus demonstrated by this (DiMaggio, 2015; Gandomi and Haider, 2015). The idea of "Big Data" has emerged as a result of developments in computing technology. Large amounts of data, typically on the petabyte scale, are referred to as "big data" in scientific endeavors (Escobar et al. (2017) and Edwards (2016).

Big data has received a lot of attention from academics and practitioners as a vital element of innovation (Baig et al., 2020, Wu and Lin, 2018, and Wolfert et al., 2017). There has been a lot of research done to date to understand how big data is used in various fields for a variety of goals (Zheng and Bender, 2019, Shahat, 2019, Yang and Du, 2016, Wassan, 2015). Determining how big data is used in the field of medical sociology is the goal of the current effort. This is because, among other things, medical sociology contributes to our understanding of how social factors affect people's health. The structure and socioeconomic contexts of the health care system can therefore be effectively examined using large sets of diverse heterogeneous data. Applying big data methodologies to the understanding of how cultural influences affect attitudes toward illness and wellness will be extremely beneficial in light of this.

Big Data, Uniqueness and Features

The data definition uses the term "Big Data" to refer to data that is at least one petabyte in size. According to Ishwarappa and Anuradha (2015), big data is also referred to as having five characteristics: variety, volume, value, veracity, and velocity. In the modern era, web-based e-commerce has become widely prevalent, business models based on big data have developed, and they now view data as an asset in and of itself. Big Data has a wide range of advantages, including decreased costs, improved productivity, increased sales, etc. Big data typically involves data sets that are too large or intricate for use with conventional data-processing software (Riahi and Riahi, 2018, Williamson, 2014). Data with more fields (rows) provide more statistical power, whereas data with more attributes (or columns) may result in a higher false discovery rate (Gao et al., 2013, Gray, 2012).

Structured, Unstructured, and Semi-Structured Data are the three categories into which big data is classified (Allen, 2023; Mishra and Misra, 2017). Big data that can be processed, accessed, and stored in a fixed format is known as "structured" data. Over time, computer science talent has had more success creating methods for handling this type of data (where the format is well known in advance) and also extracting value from it. However, in the modern era, we are anticipating problems as the size of such data increases to enormous proportions; typical sizes are now in the range of multiple zettabytes (Praveen and Chandra, 2017, Eberendu, 2016).

Big data that is unstructured consists of information that is either undefined in terms of form or structure (Mishra and Mishra, 2017, Wieringa, 2016). Unstructured data is enormous in size and presents a number of processing challenges that must be overcome in order to extract value from it. A heterogeneous data source with a mix of simple text files, images, videos, etc. is an example of unstructured data in practice. Organizations today have a wealth of data at their disposal, but since this data is in its raw or unstructured form, they are unable to value-add from it (Kumar et al., 2021). Both types of data are compatible with semi-structured data. Although semi-structured data is not defined, it could be categorized as structured. Therefore, big data could be summed up as enormously larger datasets (volume), more diversified, including structured, semi-structured, and unstructured (variety) data, and arriving faster (velocity) than before, making what is known as the 3V (Banik and Bandyopadhyay, 2016).

Applications of Big Data

Through the analysis of vast amounts of data and the discovery of hidden patterns, big data applications offer enormous benefits to the organizations that adopt them for better business decisions (Cappa et al., 2021; Hashem et al., 2015). Therein, data sets from social media, sensor data, website logs, customer reviews, and other sources may be included. Businesses are investing a lot of money in big data applications to find hidden patterns, undiscovered associations, market style, consumer preferences, and other important business data. The following industries may benefit from the use of big data: (a) healthcare; (b) media and entertainment; (c) Internet of Things; (d) manufacturing; and (e) government.

Health care: Thanks to the use of big data systems, personalized medicine and prescriptive analytics have significantly improved the state of healthcare. To determine the best course of treatment for a specific disease, drug side effects, future health risks, etc., researchers analyze the data. Data is expanding exponentially thanks to mobile applications for wearables and the health sector. By mapping healthcare and geographic data, it is possible to anticipate the spread of diseases. When an outbreak is anticipated, it can be controlled and plans for eradication of the illness can be made (Haleem et al. Changwon and others, 2020, 2014; Kaplan et al. Hilbert and Lopez (2011), 2014).

The media and entertainment sectors are using new business models to produce, market, and distribute their content. Due to customer demands, digital content can now be viewed at any time and from any location. New Netflix channels, online TV programs, etc. is demonstrating that new customers are interested in accessing data from anywhere, in addition to watching TV. By anticipating what audiences will find interesting, determining how to target advertisements, monetizing content, etc., media companies are able to target specific audiences. By examining viewer behavior, big data systems are boosting these media companies' earnings (Katyal, 2019, Saleh et al., 2018).

Internet of Things: Everyday, devices from this network send data to a server in a continuous stream. These data are extracted to enable device interoperability. Numerous businesses and government organizations can benefit from using this mapping to improve their competence (Miller and Mork, 2013, Minelli et al., 2013).

Manufacturing: By reducing machine downtime, predictive manufacturing can help produce more goods and boost efficiency. For such industries, a significant amount of data is involved. The exploration of useful information for these data is done using sophisticated forecasting tools that follow a structured process. Some of the key benefits of using big data applications in the manufacturing sector include the following: (a) high product quality; (b) tracking faults; (c) supply planning; (d) predicting the output; (e) increasing energy efficiency; (f) testing and simulating new manufacturing processes; and (g) large-scale customization of manufacturing (Kimble and Milolidakis, 2015; Minelli et al., 2013).

Government: The government can achieve efficiencies in terms of cost, output, and novelty by implementing big data systems. Since the same data set is utilized by numerous applications, numerous departments can collaborate with one another. By intervening in each of these areas, the government contributes significantly to innovation. Big data can be used in every industry, which is also important to note. Agriculture, aviation, cyber security and intelligence, crime prediction and prevention, e-commerce, fake news detection, fraud detection, pharmaceutical drug evaluation, scientific research, weather forecasting, and tax compliance are just a few of the major fields where big data finds applications (Brayne, 2017, Breur, 2016).

Big Data and the Determinants of its Various Applications

Big Data can be categorized using the four Vs. These Vs stand for the four dimensions that include: Volume, Velocity, Variety and Veracity (Arun and Jabasheela, 2014). According to estimates, 2.3 trillion gigabytes of data are produced every day. The volume and the demand for new database management systems and IT personnel both continue to increase quickly. Millions of new jobs in information technology have been created as a result of this volume, and more are expected to be created in the coming years to keep up with the expansion of the Big Data flow. Data are still available in real time, which is a result of both the presence of Big Data and the speed of the internet itself (Zakir et al., 2015). More data created requires more methods to monitor all this data. Software with high velocity has emerged as a result, greatly accelerating the rate at which data is generated and processed (Gray, 2012).

The variety of data types is related to the high speed and sizeable volume. After all, smart Information Technology (IT) solutions are available today for all sectors, from the medical world to construction and business. Consider, for example, the electronic patient records in healthcare (Au-Yong-Oliveira, 2021), which contribute to many trillions of gigabytes of data. And that's not even talking about the videos we watch on Youtube, the posts we share on Facebook and the blog articles we write. The volume and variety will only grow once the internet is widely available throughout the entire planet in the future.

Big data ecosystems' accessibility opens up a wide range of possibilities for high-end services (Calzada, 2018, Edwards, 2016, Wamba et al., 2015). Systems that are trained by the availability of big data are also a possibility thanks to the evolution of artificial intelligence (machine learning techniques).

Big Data Applications in Social Research

Digital systems continue to serve as a medium for social life, which takes place in more digital settings (DiMaggio et al., 2013). Big data represents the data being generated by the digitization of social life, which in the words of Lazer and Radford (2017) was broken down into three domains such as digital life, digital traces, and digitalized life. Big Data promises to revolutionise the production of knowledge within and beyond science, by enabling novel, highly efficient ways to plan, conduct, disseminate and assess research. Scholars in the fields of social sciences, statistics, and computer science as well as the field of data science are not relenting on unique approaches to applying modern social science research principles and current analytical and computational tools (Foster et al., 2020). In the past few decades, new methods for producing, storing, and analyzing data have been developed. This has led to the emergence of the field of data science, which combines computational, algorithmic, statistical, and mathematical techniques in order to extrapolate knowledge from big data (Klochikhin and Boyd-Graber, 2020).

As Computer Scientists in their creativity continues to dominate the world of data science, they have generated new ways of creating and collecting data. They equally succeeded in developing new analytical and statistical techniques and provided new ways of visualizing and presenting information (Biemer, 2020, Foster and Heus, 2020). These new sources of data and techniques have the potential to transform the way applied social science is done. In the recent literature of big data research, an increasing section is dedicated to the capacity of big data to support social sciences research (Lytras et al., 2015a). There is the anticipation that big data is potentially a social good that must be secured and be used for the transparency of services, and for the evolution of a user-centric new culture for sustainable computing. In parallel, several concerns have been documented, mostly related to trust, privacy and the protection of personalities in the new technology-driven domain of services and applications (Lytras and Visvizi, 2019, Lytras et al., 2015b).

Big data were first created on the Internet. They enable behavior analysis without affecting individuals (Webb et al., 1999). This demonstrates how advancements in data analysis are made possible by digital web platforms. One instance is the information that users voluntarily upload to websites, blogs, and social media platforms. According to Escobar et al (2007), these are primarily images and videos that are frequently unstructured. The Internet of things and transactional data are two additional examples. The use of digital devices, such as smartphones, scanners, tablets, credit cards, and shopping cards with embedded chips, to conduct transactions falls under this category. These devices produce data with some structure. These data include transaction-related metadata (date, time, duration, or costs). The Internet of Things' (IoT) connected objects typically produce structured data that can be arranged and analyzed, such as sensors for home automation, driving assistance, and health monitoring (Marres, 2017).

Big Data and Health Sociology Research

Big data has been portrayed in promotional and positivist discourses as a phenomenon that fundamentally alters all facets of social and personal life, as well as conventional ways of knowing the world (Rieder and Simon, 2016). These changes are typically summed up around three ideas: establishing correlations can take the place of looking for causality; data's sheer volume determines its heuristic value; and aggregated datasets from diverse sources can be analyzed independently of the contexts in which they were created, disseminated, and consumed (DiMaggio, 2015; Gandomi and Haider, 2015; Shaw, 2015). The logical conclusion is that computational approaches can be used to analyze social and cultural processes as well as human behavior because digital data about various facets of human life are readily available. Nevertheless, despite the fact that these techniques are regarded as being highly effective in some fields of science and business, they appear difficult in social research and many fields of practice (Resnyansky, 2019).

One of the social sectors where the disruptive transformation associated with turning every problem into data is most advanced and already perceptible in institutions and policies as well as daily routines is the health sector. We therefore use the field of health as the main example throughout this article to illustrate the changing arrangements of actors, data infrastructures, social and economic valorization of data, and epistemic practices, values, and orientations. The challenge of determining what constitutes "good data" and the universal principles that serve as the foundation for data is one that affects both the social sciences and the field of health (Diaz-Bone et al., 2020).

There are many datasets in healthcare. Even the smallest hospitals capture a lot of X-rays, scans, test records, etc. Hospitals can analyze the data and produce valuable information in a number of useful scenarios. Data capturing is not an issue for them as they get real-time patient data, test results while the patient is undergoing care, and impact of medicines. These can be compared to information about the patient's social, environmental, and economic circumstances. One example of how healthcare uses this is real-time devices that record patient health parameters using a wearable device that records important health parameters like blood sugar levels, blood pressure, heart rate, etc. from time to time along with the body activity and environmental conditions, i.e. weather conditions. This data is automatically uploaded on the servers and doctors at remote locations can help the patient with medication without the need to travel to the hospital (Mukherjee et al, 2012).

The same data sets can be employed to ascertain the various diseases that people may occasionally experience under various environmental circumstances. The scope can be expanded to produce better results by correlating these with additional data from social media. Hospitals can better understand the effects of their medications on their patients as well as comparable results from other doctors and hospitals to make an informed decision on the next level of medication and treatment (Priyanka and Kulennavar, 2014; Youssef, 2014; Iroju and Ikono, 2013). In order to make an informed decision, it is important to gather the experiences of various hospitals and doctors (Iroju and Olaleke, 2015; Raghupathi and Raghupathi, 2014). This adds a more scientific perspective to treatment over and above the doctor's experience.

Challenges of using Big Data in Health Research

In medical field, data analysts gather and examine information from many sources in order to gain insights into prevailing health issues. Electronic patient records, clinical decision support systems, medical imaging, doctor's written notes and prescriptions, pharmacies, laboratories, clinical data, and machine-generated sensor data are among the several sources (Raghupathi and Raghupathi, 2014). Clinical, public-health, and behavioural data are used to create a more effective treatment system that can simultaneously lower costs and raise treatment quality (Brown et al. 2011). The symptoms of individual patients were examined by the Rizzoli Orthopedic Institute in Bologna, Italy, to better understand the clinical variances within a family. This assisted in lowering the number of hospital admissions and imaging by 60 percent and 30 percent, respectively (Raghupathi and Raghupathi, 2014).

Getting information from outside sources, like social media, aids in early epidemic detection and preventative measures. Following the Haitian earthquake in January 2010, analysis of tweets assisted in monitoring the development of cholera in the area (Raghupathi and Raghupathi, 2014). The sensors' data are tracked and examined for safety monitoring and adversity event prediction (Mukherjee et al., 2012). Blount et al. (2010) created the Artemis system, which tracks and examines physiological data from sensors in intensive care units to identify the beginning of medical issues, particularly in the case of neo-natal care. Fraud can be reduced through real-time examination of a large number of claim requests. In spite of this, certain challenges are noted as serving constraints on big data. These challenges have the capability of affecting the general outcome of big data analysis. The various ways by which these challenges manifest are as discussed below:

Lack of adequate professionals with the necessary knowledge: It takes skilled data professionals to effectively handle contemporary technologies and large data tools in order to produce the desired results. To work with the tools and make sense of enormous data sets, professionals such as data scientists, data analysts, and data engineers are required. Despite playing a crucial role in this task, they lack a sufficient number of people. This frequently occurs as a result of the rapid evolution of data handling tools. This implies that there may be times when you must improvise. When this is the case, especially in medical research, the results may not accurately reveal the current state of health and may then present inaccurate projections and predictions (Estrada and Ruiz, 2016, Jin et al., 2015).

Lack of adequate understanding of Big Data: Not all people working with big data do so with sufficient understanding. Employees might not be aware of the storage, processing, significance, and sources of certain types of data, for instance. Data experts might understand what is going on while others may not have a clear picture. The employees who don't recognize the value of knowledge storage might not be able to maintain the backup of sensitive data. They might not use databases for storage in the right way. As a result, when this crucial information is needed, it is difficult to retrieve (Bhadani and Jothimani, 2016; Daniel. 2015).

Data growth issues: Properly storing these enormous sets of data is one of the most urgent problems with big data. Data centers and corporate databases are storing an ever-growing amount of knowledge. It becomes difficult to manage as these data sets grow exponentially over time. Documents, videos, audio, text files, and other sources provide the majority of the unstructured information. According to (DiMaggio, 2015, Kalra *et al.,* 2014), this indicates that you cannot find them in the database.

Confusion with Big Data tool selection: Just like any type of growing technology, the market for big data analytics has so many options that users find it confusing to settle for one or a few that they need. With too many choices, it can be challenging to identify the exact big data technologies that are in line with the stated research goals. This mistake happens even to individuals and many businesses that are relatively well-grounded on data administration. Thus, data scientists use many strategies and methods to collect, safeguard, analyze, and interpret data. Hence, there isn't a one-size-fits-all approach (Cheung and Jak, 2018, Al-Jarrah et al., 2015). The market for big data analytics has so many options that users find it difficult to choose just one or a few that they need, similar to how users find it difficult to choose a tool for any other type of growing technology. It can be difficult to pinpoint the precise big data technologies that are in line with the stated research goals when there are too many options. Even those who are reasonably knowledgeable about data issues can experience this confusion (Cheung and Jak 2018, Al-Jarrah et al., 2015).

Data Security and Integrity: Big data also faces issues with data security and integrity. One of the challenging issues with massive data is securing enormous bodies of knowledge. Data handlers frequently place the greatest emphasis on the need to comprehend, store, and analyze the data sets. Data security is always pushed to later stages. The abundance of channels and nodes that connect to one another makes it more likely that hackers will exploit any system weaknesses. The more important the data is, the more likely it is that a small error will result in significant losses (Al-Jarrah et al., 2015; Changwon et al., 2014).

Accuracy and Data Governance: Handling data from numerous sources is a requirement of big data. Most of these sources make use of distinctive formats and special data collection techniques. As a result, it is common to find inconsistencies in data even when the variables' values are similar, and making adjustments is very difficult. It is unclear how accurate big data is. It's never entirely accurate. Data may be contaminated during this process with duplicates and contradictions in addition to inaccurate information. This is done while keeping in mind that low-quality data is unlikely to provide any insightful information or assist in pinpointing opportunities for handling medical research tasks (Changwon et al., 2015; Kalra et al., 2014).

Data integration: Data integration is the process of combining data from various sources into a single, accurate version of the truth that is accessible to all organization members. Dealing with big data that comes from numerous different software and hardware platforms and takes on all possible forms is typically very difficult for the information technology (IT) team. In an effort to simplify their IT infrastructure and achieve simple data handling and big data process flows, organizations find it challenging due to the prevalence of numerous and distinctive data processing platforms. The IT departments view this as a significant challenge. For those who are using the process (like those in the medical field) to advance their own profession, the challenging situation may not bode well. This ultimately has the effect of making it difficult to promote the kind of behavioral change that is expected in the field of medical sociology (Bhadani and Jothimani, 2016, Daniel, 2015, Jothimani et al., 2015,).

Big Data Handling Costs: From the moment big data is adopted, there are significant costs associated with managing it. If you decide to use an on-premises solution, for instance, you will have to pay for new hardware, electricity, new hires like developers and administrators, as well as other costs. Although the required frameworks are open source, this will also necessitate paying the costs associated with creating, installing, configuring, and maintaining new software (Li and Lu, 2014; Kaisler et al., 2013). In cases where a cloud-based solution is chosen, costs will be incurred for things like hiring new employees (developers and administrators), cloud services, development, and covering the expenses related to the creation, installation, and upkeep of the necessary frameworks. In the end, the data may become so large that there is no more room for expansion, necessitating unforeseen costs to handle the situation (Sabharwal and Miah, 2021; Jin et al., 2015) and Galbraith, 2014).

Problems of Big Data in Medical Sociology Research

Medical sociology is concerned with how society, social structures, and population health are related. Emphasizing the interactions between patients and doctors within a healthcare system or organization is beneficial (Steve, 2015). It demonstrates how members of a society view the concepts of health and healing. The experts in this field help people see the healthcare system as a component of society and work to improve it on all fronts. Professionals in this field of social endeavors are frequently interested in social environments and how they affect human health. (Hafferty and Castellani, 2019). They use theoretical models as tools to identify the components of the complex social reality that are in charge of either increasing or decreasing the health risks in communities.

In order to examine and interpret their four main areas of interest (a) the social construction of health and illness, (b) the social production of health and illness, (c) the study of healthcare systems and facilities, and (d) the postmodern perspective on health and illness, they rely on data from the study participants. Similar to other social scientists, medical sociologists value the use of big data in their research (Hafferty and Castellani, 2019). However, data frequently reduces the subject (the context of the information) to an object and commercializes personal knowledge. When data is viewed as an object, the subjective, contextual, relational, and situational elements that go into its creation, management, use, and interpretation are disregarded (Saleh et al. , 2018).

The idea of reliability is closely related to the issue of truthfulness because big data is primarily derived from the following two major channels. (a) Self-production associated with online movements due to the traceability provided by digital technology, which allows every action (consumption, buying, and seeing) to leave distinctly distinct traces (Gray, 2010), (b) The propensity to share personal information through access to both public and private sites and/or resources, whether voluntarily or involuntarily (such as cookies, privacy management, likes, etc.). Due to the increasingly subtle, pervasive, and ubiquitous nature of the digital presence, people hardly realize how pervasive digital technology is in their daily activities.

Another aspect that is directly related to the issue of data authenticity is the randomness with which personal information is disclosed. Even with the voluntary disclosure of information taken into account (as in the case of social media sites like Facebook, Instagram, etc.), this is still easily falsifiable). It is obvious that the information is filtered according to what the person wants to say and show at the moment, i.e. the portrayal of their ideal selves. Another essential element is the finding's ability to be replicated, which is known to be essential to the scientific method (Popper, 2002). This is an important point because every digital tool, platform, and resource should be considered a social artifact that represents a specific worldview. As a result, it is never neutral; rather, it is a "black box" that always contains an asymmetry relationship that is understood by the fact that the user, even a researcher, is unaware of and lacks access to the logic, goals, constraints, and various rationales that have influenced specific decisions. There is a persistent reality of an inescapable lack of clarity that runs counter to the fundamental assumption of reproducibility and control along the entire chain of theoretical-methodological choices that serve as the foundation for data collection.

Social research loses control over data formats, construction methods, potential distortions, and/or other sources of inaccuracy even in a learning environment. Social scientists neither create the technologies used for data collection nor do they have control over the data chain (Salganik, 2018). Unlike conventional surveys, they do not have any control over how variables' dimensions of constructs are operationalized. It is managed by engineers, computer scientists, statisticians, and other experts with experience in the regular implementation of hardware, or even by sensors and software (Apps, etc.).) made to automatically feed the survey platform. The theme of validity, which is illustrated by the concept's coherence and the instrument's ability to measure what it is intended to detect, introduces the relationship between theory and empiricism. One common criticism is that big data-driven research methods were developed rather than theory-driven research (Kar and Dwivedi, 2020). This runs the risk of dealing with superficial information if the influences derived from the context, the environment, the same data creation tool, and from the specific fields of action within which it is developed are not properly understood.

How Health Sociologists Navigate Big Data Challenges

The management of the big data problem utilizes both qualitative and quantitative methods of collecting, analyzing, and encoding data in accordance with sociologists' professional calling. This makes methodological approaches like exploratory data analysis (which includes visualizing), connecting big data (analysis) to well-established social research data formats (like survey data), and fusing qualitative digital research techniques with statistical tools effectively (Lupton 2015; Marres 2017). Important choices are made as a result regarding how to conceptualize the social phenomenon in the treatment of illness.

In order to address systematic irregularities and biases present in social media data collection processes, data are "algorithm ready" during the cleaning process (Gillespie, 2014; Gitelman, 2013). This is due to the fact that user behavior on social media platforms reflects technical potential, particularly with regard to the algorithms on those platforms (Boyd and Crawford, 2012; Marres, 2012; Ruths and Pfeffer, 2014). The improvements made to deal with this situation included the inclusion of qualitatively established insights on motives and logics among others and how they are connected to one another across texts (Mohr et al., 2013). This can also be viewed from the perspective of the rapidly growing discipline known as socio-semantic network analysis. The inherent benefit of this is that cultural sociology's established methods and big data's emerging methods can work brilliantly together to produce successful outcomes in sociological research, including medical sociology (Bail, 2014).

Social science methods are now more in and of social worlds rather than being seen as objects or subjects of inquiry that are separate from them, as has been noted by research on the social life of methods (Mützel, 2015a; Ruppert, 2013). Big Data is being used to address the difficulties of fusing theory and methods in a more practical way. Students majoring in sociology at the undergraduate level are urged to become interested in conducting empirical research using online social media resources. Beyond the conventionally constrained fixed tools for statistical analysis, method training is being expanded. The training now extends beyond data construction, large data set manipulation, and machine learning tool usage for large text corpora analysis. In order to train researchers in both technical skills and theoretical approaches, practice is now under strict and thorough theoretical guidance (Mützel, 2015b). Students who are proficient in coding and modeling are encouraged to pursue academic careers while also participating in medical sociological research.

Conclusion

Much has been written on the benefits of big data for sustainable human livelihood. Typical in this situation is the area of healthcare and the efforts at improving patient outcomes, public health surveillance, and healthcare policy decisions. Over the years, Big Data, and the data sciences field in general, has been hyped as the expected solution for the healthcare industry. It has the potential of guaranteeing better healthcare outcomes and a more effective healthcare system (Househ and Aldosari, 2017). However, Househ *et al.* (2017) pointed out that small data techniques using traditional statistical methods are, in many cases, more accurate and can lead to more improved healthcare outcomes than Big Data methods. These scholars submitted further that Big Data for healthcare may cause more problems for the healthcare industry than solutions. Their position therefore corroborated the general direction against absolute reliance on Big Data, that when it comes to the use of data in healthcare, size should not be everything. This therefore underscore the necessity of making use of qualitative data to support data in medical research for appreciation of the general populace the research is intended to benefit. In order to understanding the interpretative, negotiating, and implementation perspectives of the actors involved in a research, it is important to overcome the challenges associated with the use of big data in social research. In this regard, understand the different justifications that set up the processes and their implications for group action and the common good is possible.

References

Al-Jarrah, O.Y., Yoo, P.D., Muhaidat, S., Karagiannidis, G.K. and Taha (2015). Efficient machine learning for big data: A review, Big Data search, 2(3), 87-93.

Allen, R. (2023). What are the Types of Big Data? Select Hub; available at https://www.selecthub.com/big-data-analytics/types-of-big-data-analytics/

Arun, K. and Jabasheela, L. (2014). Big Data: Review, Classification and Analysis Survey. International Journal of Innovative Research in Information Security (IJIRIS), 1(3): 17-23,

Au-Yong-Oliveira, M., Pesqueira, A., Sousa, M.J., Dal Mas, F. and Soliman, M. (2021). The Potential of Big Data Research in HealthCare for Medical Doctors' Learning. *Journal of Medical System,* 7; 45(1):13.

Baig, M.I., Shuib, L. and Yadegaridehkordi, E. (2020). Big data in education: a state of the art, limitations, and future research directions. *International Journal of Educational Technology and Higher Education* **17:** 44 https://doi.org/10.1186/s41239-020-00223-0

Bail, C.A. (2014). The Cultural Environment: Measuring Culture with Big Data. *Theory and Society*, 43(3–4): 465–482.

Banik, A. and Bandyopadhyay, S.K. (2016). Big Data – A Review on Analysing 3Vs. *Journal of Scientific and Engineering Research,* 3(1): 1-4.

Bhadani, A. and Jothimani, D. (2016), Big data: Challenges, opportunities and realities, In Singh, M.K. and Kumar, D.G. (eds.), Effective Big Data Management and Opportunities for Implementation, Pennsylvania, USA, IGI Global, pp. 1-24.

Biemer, P.P. (2020). Data Quality and Inference Errors in Foster, I., Ghani, R., Jarmin, R.S., Kreuter, F. and Lane, J. (eds.)(2020). *Big Data and Social Science: Data Science Methods and Tools for Research and Practice.* Routledge: Chapman and Hall Books/CRC

Blount, M., Ebling, M., Eklund, J., James, A., McGregor, C., Percival, N., Smith, K., Sow, D. (2010). Real-time analysis for intensive care: Development and deployment of the Artemis Analytic System. *IEEE Engineering in Medicine and Biology Magazine,* 29(2): 110-118.

Boyd, D. and Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, Technological, and Scholarly phenomenon. *Information, Communication, and Society*, 15(5), 662–679.

Brayne, S. (2017). Big Data Surveillance: The Case of Policing. American Sociological Review. 82 (5): 977–1008.

Breur, T. (2016). Statistical Power Analysis and the Contemporary Crisis in Social Sciences. Journal of Marketing Analytics (London, England: Palgrave Macmillan) 4(2–3): 61–65

Brown, B., Chui, M. and Manyika, J. (2011). Are you ready for the era of big data? *McKinsey Quarterly,* data, Available at http://www.mckinsey.com/insights/strategy/are-you-ready-for-the-era-of-big-data.

Burrows, R. and Savage, M. (2014). After the crisis? Big Data and the Methodological Challenges of empirical sociology. *Big Data & Society* 1(1): 1–6.

Calzada, I. (2018). Citizens from Data Providers to Decision-Makers? The Case Study of Barcelona. *Sustainability*, 10, 3252.

Cappa, F., Oriani, R., Peruffo, E. and McCarthy, I. (2021). Big Data for Creating and Capturing Value in the Digitalized Environment: Unpacking the Effects of Volume, Variety and Veracity on Firm Performance. Journal of Product Innovation Management, 38 (1): 49–67.

Changwon. Y., Ramirez, L. and Liuzzi, J. (2014). Big data analysis using Modern Statistical and Machine Learning Methods in Medicine, *International Neurourology Journal,* 18:50-57.

Cheung, M.W.L. and Jak, S. (2018). Challenges of big data analyses and applications in psychology. *Zeitschrift Fur Psychol. J. Psychol.* 226, 209–211. doi: 10.1027/2151-2604/a000348

Daniel, B. (2015). Big data and analytics in higher education: opportunities and challenges. *Br. J. Educ. Technol.* 46, 904–920. doi: 10.1111/bjet.12230

Diaz-Bone, R., Horvath, K. and Cappel, V. (2020). Social Research in Times of Big Data: The Challenges of New Data Worlds and the Need for a

Sociology of Social Research. *Historical Social Research,* 45(3): 314-341. https://doi.org/10.12759/hsr.45.2020.3.314-341

DiMaggio, P. (2015). Adapting computational text analysis to Social Science (and vice versa). *Big Data & Society,* 2(2): 1–5.

DiMaggio, P., Nag, M. and Blei, D. (2013). Exploiting affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding. *Poetics,* 41(6): 570–606.

Duan, Y., Shao, L., Yang, X., Sun, X., Zhou, Z. and Yu, L. (2017). Data, information and Knowledge-driven manipulation between strategical planning and technical implementation for Wireless Sensor Network Construction. *International Journal of Distributed Sensor Networks,* 13(11):1-15

Eberendu, A.C. (2016). Unstructured Data: An overview of the data of Big Data. *International Journal of Computer Trends and Technology (IJCTT),* 38(1): 46-50.

Edwards, L. (2016). Privacy, Security and Data Protection in Smart Cities: A Critical EU Law Perspective. *European Data Protection and Law Review*, 2: 28–58. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2711290

Escobar, M., Prieto, C., Barrios, D. and Martinez-Uribe, L. (2017). netCoin: interactive networks with R. available at https://cran.r-project.org/web/packages/netCoin/index.html

Estrada, R. and Ruiz, I. (2016). Big Data, Big Challenges, Big data SMACK, pp. 3–7,

Foster, I. and Heus, P. (2020). Databases, in Foster, I., Ghani, R., Jarmin, R.S., Kreuter, F. and Lane, J. (eds.)(2020). Big Data and Social Science: Data Science Methods and Tools for Research and Practice. Routledge: Chapman and Hall Books/CRC

Foster, I., Ghani, R., Jarmin, R.S., Kreuter, F. and Lane, J. (eds.)(2020). *Big Data and Social Science: Data Science Methods and Tools for Research and Practice.* Routledge: Chapman and Hall Books/CRC

Galbraith, J.R. (2014). Organizational design challenges resulting from big data. *Journal of Organisational Design*.;3(1):2–13.

Gandomi, A. and Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 35(2): 137–144.

Gao, H., Miao, H. and Zeng, H. (2013). Predictive web service monitoring using probabilistic model checking. *Applied Mathematics and Information Science,* 7(1L): 139–148.
Gao, H. and Miao, H. (2013). Research on the dynamic reconfiguration of web application using two-phase compatibility verification. *International Journal of Computing Mathematics,* 90(11): 2265–2278.

Gillespie, T. (2014). The relevance of algorithms. in Gillespie, T, Boczkowski, PJ, Foot, K (eds) *Media Technologies,* Cambridge, MA: MIT Press, pp. 167–193.

Gitelman, L. (2013). *Raw Data is an Oxymoron*, Cambridge, MA: MIT Press.

Gray, H. (2010). *Toward a Sociology of the Trace Edition.* NED, University of Minnesota Press.

Gray, P. (2012). Software economics, History, Theology. *Information and System Management,* 29(2): 161–164.

Hafferty, F.W. and Castellani, B. (2019). Medical Sociology, in Book: Handbook of Twenty-first Century Sociology. Sage, pp. 331-338 10.4135/9781412939645.n33

Haleem, A., Javaid, M., Khan, I. and Vaishya, R. (2020). Significant Applications of Big Data in COVID-19 Pandemic. Indian Journal of Orthopaedics. 54 (4): 526–528.

Hashem, I.A.T., Ibrar, Y., Nor, B.A., Salimah, M., Abdullah, G. and Samee, U.K. (2015). The rise of "big data" on Cloud Computing: Review and Open Research issues. *Information Systems,* 47: 98-115.

Hilbert, M. and López, P. (2011). The World's Technological Capacity to Share, Communicate and Compute Information. Science. 332 (6025): 60–65.

Househ, M. and Aldosari, B. (2017). The Hazards of Data Mining in Healthcare. *Studies in Health Technology Information,* 238:80-83.

Househ, M.S., Aldosari, B., Alanazi, A., Kushniruk, A.W. and Borycki, E.M. (2017). Big Data, Big Problems: A Healthcare Perspective. Studies on Health Technology Information, 238:36-39.

Iroju, O. and Ikono, R. (2013). A security based framework for interoperability of healthcare systems. *International Journal of Applied Information Systems*, 1: 23-31.

Iroju, O.G. and Olaleke, J.O. (2015). A systematic review of natural language processing in healthcare. *International Journal of Information Technology and Computer Science,* 8: 44-48.

Ishwarappa, K. and Anuradha, J. (2015). A brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology, *Procedia - Computer Science,* 48:219-324.

Jin, X., Wah, B.W., Cheng, X. and Wang, Y. (2015). Significance and Challenges of Big Data Research, *Big Data Research,* 2(2): 59–64,

Jothimani, D., Bhadani, A.K. and Shankar, R. (2015). Towards Understanding the Cynicism of Social Networking Sites: An Operations Management Perspective, *Procedia - Social and Behavioral Sciences,* 189, 117–132

Kaisler, S., Armour, F., Espinosa, J. and Money, W. (2013). Big data: Issues and challenges moving forward. In: 46th Hawaii International Conference on System Sciences (HICSS), Hawaii. pp. 995-1004.

Kakhani, M.K., Kakhani, S. and Biradar, S.R. (2015). Research issues in big data analytics, *International Journal of Application or Innovation in Engineering and Management,* 2(8): 228-232.

Kalra, B., Yadav, S. and Chauhan, D.K. (2014). A Review of Issues and Challenges with Big Data. *International Journal of Computer Science and Information Technology Research*. 2(4): 97-101

Kaplan, R.M., Chambers, D.A. and Glasgow, R.E. (2014). Big data and large sample size: a cautionary note on the potential for bias. *Clinical and Translational Science,* 7:342-346.

Kar, A.K. and Dwivedi, Y.K. (2020). Bit by bit: Social Research in the Digital Age, *International Journal of Information Management,* 54(3):102205, doi:10.1016/j.ijinfomgt.2020.102205

Katyal, S.K. (2019). Artificial Intelligence, Advertising and Disinformation. Advertising & Society Quarterly. 20 (4). doi:10.1353/asr.2019.0026

Kimble, C. and Milolidakis, G. (2015). Big Data and Business Intelligence: Debunking the Myths. Global Business and Organizational Excellence, 35 (1): 23–34. arXiv:1511.03085.

Klochikhin, E. and Boyd-Graber, J. (2020). Text Analysis in Foster, I., Ghani, R., Jarmin, R.S., Kreuter, F. and Lane, J. (eds.)(2020). *Big Data and Social Science: Data Science Methods and Tools for Research and Practice.* Routledge: Chapman and Hall Books/CRC

Kumar, R., Kumar, S. and Wasim, J. (2021). Big Data Analytics Approach for Structured and Unstructured Data, *Turkish Online Journal of Qualitative Inquiry (TOJQI),* 12 (1), 404-412.

Lazer, D. and Radford, J. (2017). Data ex Machina: Introduction to Big Data. *Annual Review of Sociology,* 43:19-39.

Li, H. and Lu, X. (2014). Challenges and trends of big data analytics. In: Ninth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), Guangzhou China. pp. 566-567.

Lupton, D. (2015). *Digital Sociology.* London: Routledge.

Lytras, M.D. and Visvizi, A. (2019). Big Data and Their Social Impact: Preliminary Study, *Sustainability,* 11, 5067.