



Review

The Need for a Definition of Big Data for Nursing Science: A Case Study of Disaster Preparedness

Ho Ting Wong^{1,2,3}, Vico Chung Lim Chiang^{1,*}, Kup Sze Choi⁴ and Alice Yuen Loke¹

¹ Disaster Nursing Task Force, School of Nursing, the Hong Kong Polytechnic University, Hong Kong, China; frankwong@connect.hku.hk (H.T.W.); alice.yuen.loke@polyu.edu.hk (A.Y.L.)

² Department of Geography, National Taiwan University, Taipei City 10617, Taiwan

³ Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

⁴ School of Nursing, the Hong Kong Polytechnic University, Hong Kong, China; thomask.choi@polyu.edu.hk

* Correspondence: vico.chiang@polyu.edu.hk; Tel.: +852-2766-6683

Academic Editor: Jason K. Levy

Received: 20 July 2016; Accepted: 12 October 2016; Published: 17 October 2016

Abstract: The rapid development of technology has made enormous volumes of data available and achievable anytime and anywhere around the world. Data scientists call this change a data era and have introduced the term “Big Data”, which has drawn the attention of nursing scholars. Nevertheless, the concept of Big Data is quite fuzzy and there is no agreement on its definition among researchers of different disciplines. Without a clear consensus on this issue, nursing scholars who are relatively new to the concept may consider Big Data to be merely a dataset of a bigger size. Having a suitable definition for nurse researchers in their context of research and practice is essential for the advancement of nursing research. In view of the need for a better understanding on what Big Data is, the aim in this paper is to explore and discuss the concept. Furthermore, an example of a Big Data research study on disaster nursing preparedness involving six million patient records is used for discussion. The example demonstrates that a Big Data analysis can be conducted from many more perspectives than would be possible in traditional sampling, and is superior to traditional sampling. Experience gained from the process of using Big Data in this study will shed light on future opportunities for conducting evidence-based nursing research to achieve competence in disaster nursing.

Keywords: Big Data; disaster preparedness; research method

1. Introduction

Computing technology has further advanced in recent years, and the availability and popularity of social media and smart phones has led to new sources for collecting research data. These technological advancements have caused the world to evolve to an “era of data”. Enormous amounts of data can now be generated anytime and anywhere around the world [1,2]. This is certainly good news for scientists, including nursing scholars, as their research is no longer limited to conventional data collection methods.

The popularity and accessibility of social media and smart phones have led to revolutionary changes in numerous disciplines, including journalism [3,4], geography [5,6], and emergency medicine [7,8]. It is now common for journalists to search and analyze the postings on social media, resulting in a new professional area called data journalism, which emphasizes the role of data in news reports [9]. Similarly, Big Data is also facilitating research relating to disaster management among geographers. For example, the built-in Global Positioning System (GPS) in a typical smart phone is a valuable source of data. A study was conducted making use of the smart phone GPS

data of 1.6 million people to develop a spatio-temporal model that simulated and predicted the moving patterns of people in Tokyo after the 11 March earthquake in 2011 [10]. Similar to smart phones, social media such as Twitter can generate tens of thousands of tweets every second around the world. The United States Geological Survey demonstrated the feasibility of monitoring the number of earthquake-related posts in tweets immediately after an earthquake to detect earthquakes in 2009. This method is particularly convenient for areas without earthquake sensors [11]. Apart from personal communication devices and online social media platforms, the government is one of the major owners of Big Data, given the tremendous amounts of data held by different public organizations. Abundant data are readily available in public hospitals regarding the health of citizens and their patterns of service usage, meaning that nursing and related research can likely benefit by utilizing such data. As revolutionary changes have already been observed in other disciplines, it is envisioned that Big Data will also have a revolutionary impact on nursing science.

According to Google Trends [12], the number of Google searches using the query “Big Data” began increasing in 2011 (Figure 1). This suggests that the concept and term are becoming more popular among the public, and that the situation among nursing scholars is similar. The same search among PubMed Nursing Journals without year restrictions returned 38 publications. The first publication related to Big Data in PubMed Nursing Journals appeared in 2013 [13]. The number of such papers increased to 2 in 2013, 7 in 2014, and 19 in 2015. In 2016, a total of 10 publications were identified up to the month of June. Although the trend of increase and the total number of publications related to Big Data in PubMed Nursing Journals are not as impressive as the popularity of “Big Data” in Google searches, they show that Big Data is already having a definite impact on nursing science. Such an impact is one day expected to be revolutionary.

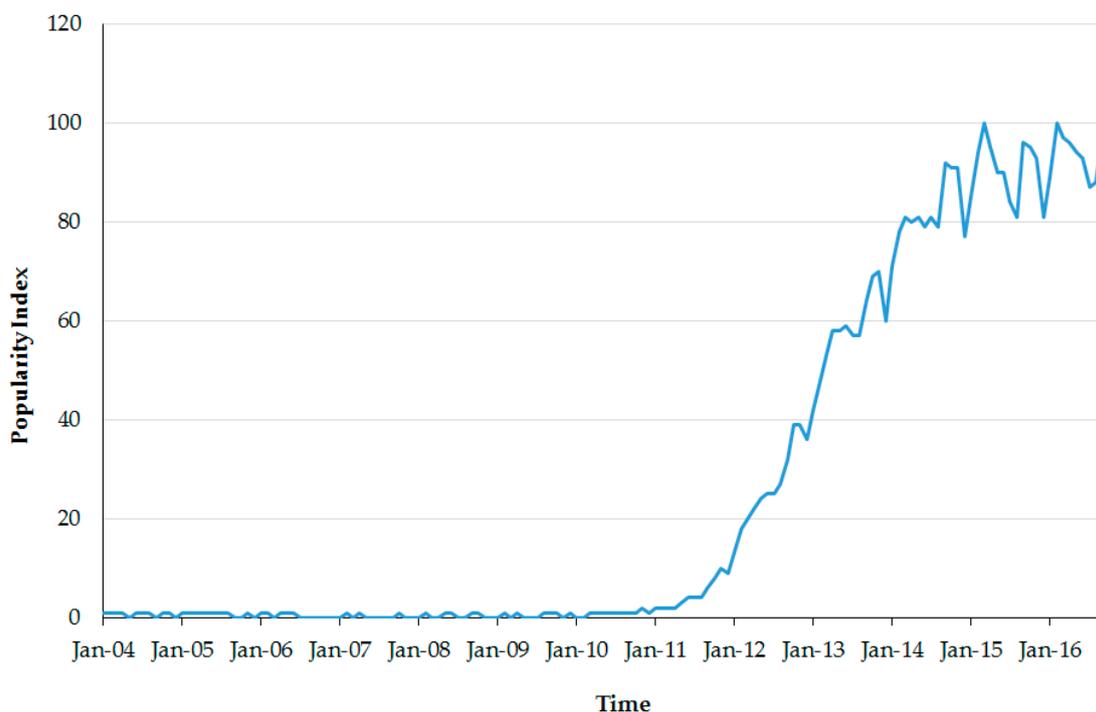


Figure 1. The popularity of the search query “Big Data” in Google Trends.

2. The Challenge of Big Data to Scholars in Nursing

Although the concept of Big Data has been attracting the attention of scholars in nursing since the birth of the era of data, the use of Big Data in nursing research remains limited. One reason for this could be the lack of a unified definition of “Big Data”. In the 38 publications in PubMed Nursing Journals (identified using the PubMed Nursing Journals full text keyword search function and the

keyword “Big Data”; a full list of the 38 publications can be found in Table A1), scholars in nursing have adopted very different definitions of Big Data [14–16], while some made no attempts to define the term before using it [17,18]. Among the 38 publications related to Big Data in PubMed Nursing Journals, 22 explicitly defined Big Data. Nine of the 22 adopted the popular 3Vs or 4Vs definition [19], while the remaining publications adopted different, less structured, definitions. One referred to Big Data as “any collection of data that is large and complex enough to become difficult to process” [20]. In fact, unstandardized or imprecise definitions cause confusion to scholars, as many of them perceive Big Data as equivalent to big datasets, with a sample size “significantly bigger” than those in a “typical study”. De Mauro et al. [21] (a frequently cited article, 53 times in two years, according to Google Scholar) also mentioned that “The lack of a formal definition has led research to evolve into multiple and inconsistent paths” and “It is necessary to invest time and effort in the proposition and the acceptance of a standard definition of Big Data that would pave the way to its systemic evolution and minimize the confusion related to its usage”. A summary of the definitions used in the Big Data related PubMed Nursing Journals can be found in Table A2.

Whether the existing definitions are suitable to nursing scholars is also questionable since the definitions were created by data scientists who were interested in developing complicated algorithms to handle Big Data. In contrast, nursing scholars are more interested in analyzing Big Data in order to obtain better research evidence to better inform nursing practice. Nevertheless, the discussion on Big Data in nursing literature is small. When conducting the same keyword search on “Big Data” from all PubMed Journals, almost 2500 publications can be found in addition to the 38 PubMed Nursing Journal publications. Although it is not surprising that significantly higher number of papers about Big Data are published when not restricting to nursing journals, a difference of over 65 times in number is much greater than expected. Out of all, in view of the dynamic nature of the definition for Big Data and the intrinsic needs of nursing and different professions for the advancement in Big Data research, an attempt is made in this paper to discuss the issue using a case study for illustration, and to identify a definition that may be suitable to nursing research.

3. A Widespread Definition of Big Data: 4Vs

The definition of Big Data has been described as a “moving target” because there are at least 32 such definitions available from different sources [21,22]. The most popular and widespread of these is the “4Vs” because of its clear structure [19] and because the concept originated from a data scientist [23]. According to this definition, Big Data should fulfill four characteristics that can be represented by four keywords starting with the letter “V”. These characteristics are “Volume”, “Velocity”, “Variety”, and “Veracity”.

The term “Volume” refers to the size of the dataset measured by the unit byte. The second keyword “Velocity” refers to the speed at which new data is being generated. A source of Big Data that generates data at a high speed would naturally produce large volumes of data. Social media, such as Facebook, are examples of such sources of data. Facebook users upload enormous numbers of postings and photographs; with around 1.6 billion active Facebook users in 2016, the speed requirement for Big Data is clearly being fulfilled [24]. The use of social media postings as a source of data has become popular [25,26].

The third keyword “Variety” refers to the number of data types, sources, or attributes. Numeric value is the most popular type of data in a typical nursing study; however, data can also appear in different formats. For example, a database from hospitals can appear in formats other than numeric value, such as images, videos, and text. Image data are generated when patients have X-rays taken in the radiology department, while video data can be generated from gastroscopy examinations in the endoscopy department, and text data are commonly found in patients’ hospital records, generated by physicians, nurses, and other healthcare professionals. Data from the finance department can also be used to conduct health economic studies. Studies that are wider in scope may also be conducted if data from observatories and police departments are used to study the relationship between changes in

weather, car accidents, and visits to the accident and emergency department of hospitals. The variety of sources of data is indirectly related to requirements relating to the variety of data attributes. This is because datasets with multiple sources of data are generally expected to have more attributes, and to have a larger capability and more flexibility to answer different research hypotheses. Nevertheless, the variety of data attributes makes Big Data different from traditional small data.

The final keyword “Veracity” refers to the accuracy of a database. Based on the “garbage in garbage out” principle, if a database contains too much inaccurate data, the analysis conducted using this database will generate inaccurate results. In the content of Big Data, a certain degree of inaccuracy is acceptable because the volume of data is much larger than in traditional small datasets, offsetting the potential errors.

Although the 4Vs definition of Big Data is popular, it is still difficult for scholars to judge whether a dataset has fulfilled the 4Vs definition. In practical terms, the definition is also difficult to use. With regard to the term “Volume”, the concept of big is difficult to quantify and it is difficult to come up with an objective threshold value to classify whether a dataset is big or small. The concept depends on the times and on the discipline of research. For example, personal computers were far less common two decades ago and a 3.5 inches floppy disk was the standard medium for transferring and storing data. The storage capacity of a 3.5 inches floppy disk was only 1.44 MB. Nowadays, a USB thumb drive has replaced the floppy disk as the standard medium for managing data. The storage capacity of a typical 64 GB USB thumb drive is 44 thousand times greater than the storage capacity of a 3.5 inches floppy disk. In this connection, the definition of “big” will change from generation to generation, depending on advancements in the technologies available to enable Big Data research. Even if only the current generation of technologies is considered, the concept of big is also related to research disciplines and context. In analyses of data, the unit of size most popular among computer scientists is probably the terabyte, since much of their work is related to handling images and sound tracks. In contrast, quantitative nursing scholars are more familiar with the megabyte, as they normally deal with observational (e.g., questionnaire surveys) or interventional studies. As traditional personal computer hardware and software may not be able to handle data at the gigabyte level or larger, scholars in nursing may encounter a barrier when attempting to analyze the so-called Big Data.

A similar situation of the lack of a benchmark threshold value also exists for the characteristics of “Velocity” and “Variety”. Speed is also affected by advancements in technologies. In the old days, the process of generating data using personal computers normally involved only one party at a time and took place only when people intended to create data, such as data inputs or measurements. The process of generating data is now much more complicated and normally involves multiple parties. Data are not only generated because of deliberate data entries; rather, people’s daily activities indirectly create large amounts of unintentional data. For example, a person’s physical location can be captured by telecommunications companies through the GPS system of that person’s smart phone. Even without any communication device, data can also be generated and captured when people use tap-and-go cards to make small payments. As with the issue of how big is big in terms of “Volume”, it is also difficult for those in different research disciplines to come to a consensus on “Velocity” and “Variety”. Although a social media researcher and a health administration researcher may both claim that they are conducting research using “Big Data”, the former, who would be conducting research using Facebook postings, would have a very different perception of data speed from the latter, who would be conducting research using hospital records, as it is obvious that the speed with which 1.6 billion active Facebook users generate data has to be much faster than that of a hospital or even a hospital network.

The characteristic of “Variety” also presents a challenge for those attempting to come up with a universal standard on how many types of data would be enough to fulfill the criteria of variety. In fact, there are many techniques for analyzing different types of data including texts and images, which require transforming non-numeric types of data back to a numeric value format. For example, linguists may conduct a Big Data study using Facebook posts in text format. They may need to transform the texts to numeric values by counting the number of keywords in the posts in order to

conduct statistical comparison tests. Hence, the necessity for Big Data to consist of data in a variety of formats is also debatable. Even shifting the discussion from the number of data formats to data sources, it is also questionable if multiple sources of data are necessary. This is because the number of data sources required should depend on the research hypothesis and on the aims of the study. Having extra sources of data will generally not increase the robustness of a study, except in the case where the data is cross-checked to improve its accuracy. Moreover, having too many attributes could also be troublesome in a data analysis unless the attributes are linearly independent of each other. This is because linearly dependent attributes will not be able to provide additional information as they are directly related to each other [27].

Unlike “Volume”, “Velocity”, and “Variety”, defining the criterion of “Veracity” presents less of a challenge because it has long been an important requirement in data analyses, even before the birth of Big Data. Although it can be difficult to say what level of inaccuracy is considered to be acceptable, it is commonly assumed that the more accurate the data is, the better. Furthermore, this requirement has been relatively stable across different disciplines and generations.

4. One Definition of Big Data for Nursing Scholars

It is worth exploring the possibility of having a suitable definition of Big Data for nursing scholars. If, despite the challenges involved, the 4Vs definition of Big Data is adopted in the area of nursing science, nursing scholars may have to face the problem that some of the studies that claim to have adopted Big Data have actually not done so. Moreover, the 4Vs definition may become a barrier to nursing scholars to joining the Big Data research family because of the independent nature of nursing scholars, who are not unique in this regard [6]. In particular, nursing scholars do not typically anticipate themselves to have the relevant computational skills, software, and hardware equipment to handle big data of tremendous volume and at a very fast high speed. It is because these computational skills are expected to be the expertise of computer scientists [6].

In the book “Big Data: A Revolution That Will Transform How We Live, Work and Think” [2], the concept of “sample equals to population” is discussed. There, a more suitable definition of Big Data for nursing scholars is proposed. More criteria could be added to improve the definition, such as by crystallizing the concept of big through imposing an objective threshold, even though that threshold may not perfectly suit all situations. Basically, the concept of “sample equals to population” simply suggests that, for Big Data studies, various sampling processes are no longer required and that the study sample is simply equivalent to the study population in the Big Data world. Thus far, this concept is obviously unfeasible for nursing research. For example, it is not cost-effective to conduct a population census in order to answer a research question for a typical nursing research study. Hence, nursing researchers utilize various sampling methods to select the appropriate subjects to represent the population for research. However, with the world now evolving to enter the data era and with data covering nearly the whole population becoming increasingly available, more opportunities have arisen for nursing scholars to conduct large population-based studies with Big Data. If the concept of “sample equals to population” is adopted as the definition of Big Data in nursing research, it will be easier to classify in a more objective way whether such a nursing study is one that involves Big Data. Conducting Big Data research will become more conceptually clear to nursing scholars. The situation that nursing scholars are excluded from the Big Data world of research is preventable if they can identify a suitable Big Data definition.

5. A Case Study of Disaster Preparedness Research Using Big Data

According to the International Council of Nurses (ICN) Framework of Disaster Nursing Competencies, preparedness competence is one of the four major disaster nursing competencies [28]. Although disaster preparedness is considered to be the most critical phase in the disaster management continuum, none of the publications related to disaster preparedness in nursing can be identified in the PubMed Nursing Journals as having used Big Data. Disaster preparedness research using Big Data

has, however, been found in the publications of health geographers [5,29], an indication that health geographers are among the first to use Big Data in such research.

An example is a Big Data Analytic study exploring the relationship between changes in weather and the demand for ambulances, which concerns the development of evidence-based practices relating to disaster preparedness in health services and nursing [30]. The study utilized a big dataset of the Hong Kong Hospital Authority (HA—a public hospital system in Hong Kong), which contains six million records of patient visits to the Accident and Emergency Departments (AED) of the hospitals under the HA from May 2006 through April 2009 (3 years). The aim of the study, which was to develop a spatio-temporal daily emergency ambulance demand forecast system, contributes to the knowledge base on the disaster preparedness of emergency ambulances and related health services [30]. If the 4Vs Big Data definition is adopted, since the dataset from the HA can be stored in a compact disk in the format of a single Microsoft Access file of around 500 MB in size, the study would not have been classified as Big Data in terms of the “volume” of the data involved. However, the dataset of the HA contains the total number of AED patient visits with emergency ambulance access in Hong Kong, and covers the whole population of Hong Kong, which meets the criterion for the definition of big data as “sample equals to population” [2]. At the same time, we have to emphasize that the definition of “sample equals to population” is an example to demonstrate the need of a suitable Big Data definition for nursing scholars. In principle, the definition still needs to be improved based on the realistic situation of nursing scholars’ comments and general consensus. For example, some scholars may not agree that the above mentioned dataset is large enough to be called “Big Data”. It is because the dataset with six million records could be considered as big in Hong Kong but small in big country like USA. Moreover, the concept of “Big” could also be changed due to technology advancement such as the change from storing data in floppy disks to USB thumb drives. Hence, the concept of Big Data is complicated and clearly depends on both social and technical aspects [31]. Nursing scholars from different specialties should be involved in formulating the Big Data definition for nursing scholars.

The Big Data Analytic study was divided into three phases of data analysis to establish the weather-ambulance demand relationship and the socio-demographic characteristics of users, so as to forecast the short-term and long-term demand for ambulances. In phase one, the daily ambulance demand time-series data by the different socio-demographic characteristics of ambulance users (e.g., age and gender) were extracted to correlate with different meteorological time series data (e.g., temperature and relative humidity) in order to identify the meteorological variables that were significantly correlated with ambulance demand by people’s socio-demographic characteristics. The research results indicated that older people, females, those with a lower income, and patients with more severe conditions were more sensitive to extreme weather conditions [8].

In phase two, a spatio-temporal daily emergency ambulance demand forecast system using the Hong Kong Observatory’s seven-day weather forecast report was developed. The model validation results showed that it is feasible to use a weather forecast report to forecast the future daily demand for ambulances [30]. The accuracy of the forecast was further improved by introducing the seven-day average temperature forecast into the emergency ambulance demand forecast model [29].

In phase three, a long-term projection algorithm to project the demand for ambulances until 2036 was developed, using the corresponding relationship between weather and ambulance demand and the population projection report compiled by the Hong Kong Planning Department. The projection showed that, if the factor of an aging population was neglected, both the yearly demand for ambulances and the sensitivity of the daily demand for ambulances under extreme weather conditions would be severely underestimated [32]. One implication for disaster preparedness during extreme weather conditions that are potentially or actually disastrous is that older people have a higher demand for ambulances. There should be a greater focus on offering disaster preparedness education and relevant health services to this group of citizens.

With the success of this Big Data Analytic study on the relationship between changes in weather and emergency ambulance demand, the same concept and procedures were translated into

another study using the health-related help-seeking behavior data of the “personal emergency link” (PE-Link) service for older people. The PE-link service is a telehealth system for older citizens operated by the Senior Citizens Home Safety Association (SCHSA), which has more than 80,000 active users in Hong Kong [33]. The research results also identified a certain group of PE-link users who are relatively sensitive to extreme weather conditions that can be potentially disastrous, e.g., extremely cold days, which are not uncommon during winter in Hong Kong. That study suggested that it is feasible to predict the health-related help-seeking behavior of PE-Link users using Big Data from official government weather forecast reports. The study provided nursing researchers with insights on the benefits of strengthening the use of the PE-link system for disaster preparedness involving older people [34].

The above multi-phased process studies demonstrated that “sample equals to population” Big Data can be analyzed from different perspectives (i.e., the effects of weather on ambulance demand and the effects of population changes on ambulance demand). The use of Big Data has demonstrated its superiority in public health nursing to the sampling approach adopted in traditional nursing research that usually includes only the selected research participants for particular research topics. The data collection process of such traditional research studies is also often time-consuming and demands much effort. The “sample equivalent to population” type of Big Data approach can be used with patient records, which can be analyzed using datasets other than weather reports and can generate results in other aspects that contribute information on public health. In fact, the case study demonstrated that conventional data collection methods (such as questionnaire surveys and in-depth interviews) are not the only ways to conduct disaster preparedness research. When the Big Data approach is used appropriately, more opportunities will be available to nursing scholars to conduct disaster preparedness research because they are no longer limited under the conventional approach. The given example in disaster nursing is only the tip of the iceberg. Other examples of Big Data analytic applications include clinical operations, patient profile analytics, device/remote monitoring, and pre-adjudication fraud analysis. A good review on Big Data analytics in healthcare can be found from Raghupathi [35].

6. The Use of Big Data from Disaster Preparedness to Public Health

The above example demonstrated that the Big Data approach can be used in disaster preparedness research, contributing to public health and to research on public health. Big Data research is a good match with public health, as the scope of analysis encompasses the whole population [36]. For example, infectious disease surveillance, which is an important area of public health, was limited in the past by passive methods of collecting data, such as notifiable cases of disease reported by physicians or laboratories [37]. In particular, in the case of influenza, it is likely that the passive reporting approach would have resulted in a significant number of missing cases, due to false negative diagnoses or to the recovery or death of people who had become infected but who had not undergone any medical consultation. If Big Data is adopted in public health surveillance, the reported number of infected cases could be improved. This is because the sources of data will not only be limited to medical records, but also include indirect data such as sales records for flu medicines or records of Internet searches on flu-related subjects, which could be used to estimate the number of infections. In fact, the occurrence of Big Data has helped public health to evolve into public health 2.0, which describes public health activities that are completely user-driven by the use of social media, search engines, mobile technologies, or other technologies [38]. In the new version of public health, the data to conduct infectious disease surveillance are obtained directly from public entities, such as Google Flu Trends and the Global Public Health Intelligence Network [20,39].

7. Future Challenges of Using Big Data in Nursing Research

Despite the advantages of conducting nursing research using Big Data, there are some foreseeable challenges that may inhibit its development. The governments of major cities in the world are now

emphasizing the use of Big Data to manage cities (e.g., the smart city) and promoting open data policies to increase the public's involvement in formulating policies. When such ideas are promoted, the issue of data privacy should not be neglected. The public is becoming increasingly concerned about personal privacy. It is projected that it will not be enough to employ data anonymization in connection with the use of Big Data. More rigorous and standardized measures need to be developed to meet the public's expectations of personal privacy.

In the case of sampling in survey research designs, data anonymization is generally sufficient to assure confidentiality and privacy, as the datasets are specific and the number of research participants is limited. In Big Data analysis, computer algorithms are available to perform data de-anonymization [40]. However, there have been reports showing that merging multi-anonymous Big Data can generate personally identifiable data [1]. For example, assume that there is simple, anonymous Big Data containing all of the records of patients who have visited AEDs in Hong Kong. The dataset contains only six variables (date of visit, time of visit, age, gender, hospital name, and diagnosis). The structure of the data is illustrated in Table 1. If there is a newspaper report about a politician's visit to an AED, their medical diagnosis can then be easily identified based on the time of the visit, the specific hospital that was visited, and the politician's gender and age. In this connection, the accessibility of Big Data from both the public and private sectors can be unfavorable to privacy.

Table 1. Part of the simple, anonymous AED Big Data related to the suspected patient.

Date	Time	Gender	Age	Hospital Name	Diagnosis
20 May 2016	14:57	Male	35	A *	Flu
	15:00	Male	60	A	Schizophrenia
	15:05	Female	17	A	Tuberculosis

*: "A" is a label for the hospital.

In order to have a sustainable environment for encouraging the use of Big Data in nursing research, it is essential to devise a set of standard operating procedures (SOPs) on security measures to ensure privacy and confidentiality when using Big Data for research. For example, researchers can only be allowed to access the data in a dedicated room, all data should be password protected, all data should be destroyed within a certain period after the end of a project, and so on. In fact, the Health Insurance Portability and Accountability Act Privacy Rule has a guideline that includes a list of variables that should be removed from a dataset so that the dataset can be considered to have been de-identified [41]. It is also necessary to invite the public to join in the development of SOPs through open discussion forums.

In fact, the future challenges of using big data in nursing research is not only limited within the issues of privacy and confidentiality. There are many future challenges in Big Data such as (1) deciding the acceptable amount of missing data; (2) addressing the issue of data accuracy and validity; (3) handling large number of confounders; and (4) the impact of using proxy variable. Unlike the issue of privacy and confidentiality which should directly be handled by nursing scholars, other professions such as data scientists and statisticians will play an important role in handling these issues. Hence, there is room for them to multi-disciplinarily collaborate with nursing scholars in formulating a suitable and the best Big Data definition for nursing research in public health.

8. Conclusions

Big data is an emerging area of research that could lead to the development of the next generation of more sophisticated and compelling research for evidence-based nursing practices. However, for the time being, in nursing research as in many other disciplines, the existing definitions of Big Data may not be suitable. In this paper, the use and limitations of the 4Vs definition of Big Data for nursing research was introduced and discussed. The suitability of the concept of the whole population as the sample for the definition of Big Data for nursing research was proposed. A case study of using Big Data

to conduct disaster preparedness research was also provided to demonstrate that conventional data collection methods (such as sampling surveys and in depth interviews) are not the only ways to conduct disaster preparedness research. Finally, nursing scholars should come to a consensus on the definition of Big Data and establish widely accepted SOPs on security measures to address the issue of the privacy and confidentiality of Big Data, so that research using Big Data in nursing science can be further developed and sustained.

Acknowledgments: A previous version of this paper was published and presented at the 20th International Geographic Conference of Taiwan [6]. A short review article discussing the same issue from the perspective of emergency medicine was published in the Journal of Acute Disease and is cited in this paper [7].

Author Contributions: Ho Ting Wong conceived and wrote the review; Alice Yuen Loke led the review; Vico Chung Lim Chiang, Kup Sze Choi, and Alice Yuen Loke revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. The 38 Big Data related publications in PubMed Nursing Journals.

No.	Author	Year	Manuscript Title	Journal Name
1	Bakken and Reame [42]	2016	The Promise and Potential Perils of Big Data for Advancing Symptom Management Research in Populations at Risk for Health Disparities	Annual Review of Nursing Research
2	Barton [43]	2016	Big Data	Journal of Nursing Education
3	Blair [44]	2016	Publicly Available Data and Pediatric Mental Health: Leveraging Big Data to Answer Big Questions for Children	Journal of Pediatric Health Care
4	Booth [45]	2016	Informatics and Nursing in a Post-Nursing Informatics World: Future Directions for Nurses in an Automated, Artificially Intelligent, Social-Networked Healthcare Environment	Nursing Leadership
5	Broome [46]	2016	Big Data, Data Science, and Big Contributions	Nursing Outlook
6	Clancy and Reed [15]	2016	Big Data, Big Challenges: Implications for Chief Nurse Executives	Journal of Nursing Administration
7	Linnen [47]	2016	The Promise of Big Data: Improving Patient Safety and Nursing Practice (Part 1)	Nursing
8	Linnen [48]	2016	The Promise of Big Data: Improving Patient Safety and Nursing Practice (Part 2)	Nursing
9	Remus [49]	2016	The Big Data Revolution: Opportunities for Chief Nurse Executives	Nursing Leadership
10	Zurlinden [50]	2016	Health Professions Education: Make Big Data Meaningful	The Journal of Continuing Education in Nursing
11	Brennan and Bakken [14]	2015	Nursing Needs Big Data and Big Data Needs Nursing	Journal of Nursing Scholarship
12	Cohen et al. [51]	2015	Challenges Associated with Using Large Data Sets for Quality Assessment and Research in Clinical Settings	Policy, Politics & Nursing Practice
13	Cohen et al. [52]	2015	Implementing Common Data Elements across Studies to Advance Research	Nursing Outlook
14	Elgin and Bergero [53]	2015	Technology and the Bedside Nurse: An Exploration and Review of Implications for Practice	Nursing Clinics of North America
15	Fingeld-Connett [54]	2015	Twitter and Health Science Research	Western Journal of Nursing Research
16	Harper and Parkerson [55]	2015	Powering Big Data for Nursing Through Partnership	Nursing Administration Quarterly
17	Henly et al. [56]	2015	Emerging Areas of Science: Recommendations for Nursing Science Education from the Council for the Advancement of Nursing Science Idea Festival	Nursing Outlook
18	Jackson [57]	2015	Teaching and Learning About Big Data: Start Small	Nurse Educator
19	McCartney [58]	2015	Big Data Science	The American Journal of Maternal/Child Nursing
20	McConnell et al. [59]	2015	The Future of Big Data: Innovative Methodological Approaches	Issues in Mental Health Nursing
21	Price et al. [60]	2015	The Veterans Affairs's Corporate Data Warehouse: Uses and Implications for Nursing Research and Practice	Nursing Administration Quarterly

Table A1. Cont.

No.	Author	Year	Manuscript Title	Journal Name
22	Samuels et al. [61]	2015	Using the Electronic Health Record in Nursing Research: Challenges and Opportunities	Western Journal of Nursing Research
23	Sensmeier [62]	2015	Big Data and the Future of Nursing Knowledge	Nursing Management
24	Simpson [63]	2015	Big Data and Nursing Knowledge	Nursing Administration Quarterly
25	Stifter et al. [64]	2015	Nurse Continuity and Hospital-Acquired Pressure Ulcers: A Comparative Analysis Using an Electronic Health Record “Big Data” Set	Nursing Research
26	Stifter et al. [16]	2015	Proposing a New Conceptual Model and an Exemplar Measure Using Health Information: Technology to Examine the Impact of Relational Nurse Continuity on Hospital-Acquired Pressure Ulcers.	Advances in Nursing Science
27	Welton and Harper [65]	2015	Nursing Care Value-Based Financial Models	Nursing Economics
28	Westra et al. [66]	2015	Nursing Knowledge: 2015 Big Data Science	CIN: Computers, Informatics, Nursing
29	Westra et al. [67]	2015	Nursing Knowledge: Big Data Science-Implications for Nurse Leaders	Nursing Administration Quarterly
30	Henly et al. [68]	2014	Mother Lodes And Mining Tools: Big Data For Nursing Science	Nursing Research
31	Lee [17]	2014	A Systems Science Approach to Fatigue Management in Research and Health Care	Nursing Outlook
32	Lyon et al. [18]	2014	Biobehavioral Examination of Fatigue across Populations: Report from a P30 Center of Excellence	Nursing Outlook
33	Maughan et al. [69]	2014	Standardized Data Set for School Health Services: Part 1—Getting to Big Data	NASN School Nurse
34	Rashbass and Peake [70]	2014	The Evolution of Cancer Registration	European Journal of Cancer Care
35	Skiba [71]	2014	The Connected Age: Big Data & Data Visualization	Nursing Education Perspectives
36	Westra and Choromanski [72]	2014	Amazing News for Sharable/Comparable Nursing Data to Support Big Data Science	CIN: Computers, Informatics, Nursing
37	Harper [13]	2013	The Economic Value of Health Care Data	Nursing Administration Quarterly
38	Salcido [73]	2013	Big Data and Disruptive Innovation in Wound Care	Advances in Skin & Wound Care

Table A2. The Big Data definition appeared in the 38 Big Data related publications PubMed Nursing Journals.

No.	Big Data Definition	Reference
1	3Vs or 4Vs	[14,42,45,47,48,55,58,60,61]
2	Large pools of data that can be captured, communicated, aggregated, stored, and analyzed, are now part of every sector and function of the global economy	[13]
3	The formation of large electronic repositories of healthcare data	[15]
4	The complexity, challenges, and new opportunities presented by the combined analysis of data. In biomedical research, these data sources include the diverse, complex, disorganized, massive, and multimodal data being generated by researchers, hospitals, and mobile devices around the world.	[43]
5	It is diverse, complex, disorganized and multimodal data generated by hospitals, researchers, and individuals who wear mobile devices and sensors that provide real time data about health status and parameters	[46]
6	The exponential growth of information and knowledge emerging from sources.	[49]
7	Any collection of data that is large and complex enough to become difficult to process.	[52]
8	Techniques developed to manage, analyze, and translate the vast and expanding amount of patient data elements	[53]
9	Data sets that perform beyond the capabilities of traditional software programs to store, manage, and analyze	[57]
10	Datasets whose sizes are beyond the ability of typical database software tools to capture, store, manage, and analyze	[62]
11	Data sets that are of a size or complexity that is beyond the ability of typical database software tools	[64]
12	High volume or complex data that originates from a variety of sources.	[67]
13	The use of ginormous bits or sets of data in the terabyte-and-beyond range, which are exponentially increasing by the minute, owing to the capacity of data collection, storage, and analysis	[73]

References

1. Chartier, T. *Big Data: How Data Analytics Is Transforming the World*; The Teaching Company: Chantilly, VA, USA, 2014.
2. Mayer-Schönberger, V.; Cukier, K. *Big Data: A Revolution that Will Transform How We Live, Work, and Think*; Mariner Books: Boston, MA, USA, 2014.
3. Auer, M.; Fu, K.W. Clearing the air: Investigating weibo censorship in china: New research to show censorship of microbloggers who spoke out about pollution documentary. *Index Censorsh.* **2015**, *44*, 76–79. [[CrossRef](#)]
4. Liang, H.; Shen, F.; Fu, K.W. Privacy protection and self-disclosure across societies: A study of global twitter users. *New Media Soc.* **2016**, in press. [[CrossRef](#)]
5. Kitchin, R. Big data and human geography: Opportunities, challenges and risks. *Dialogues Hum. Geogr.* **2013**, *3*, 262–267. [[CrossRef](#)]
6. Wong, H.T.; Chiang, V.C.L.; Choi, K.S.; Loke, A.Y. The opportunity of big data research for health geographers. In Proceedings of the 20th International Geographic Conference of Taiwan, Taipei, Taiwan, 20–21 May 2016.
7. Wong, H.T.; Yin, Q.; Guo, Y.Q.; Murray, K.; Zhou, D.H.; Slade, D. Big data as a new approach in emergency medicine research. *J. Acute Dis.* **2015**, *4*, 178–179. [[CrossRef](#)]
8. Wong, H.T.; Lai, P.C. Weather inference and daily demand for emergency ambulance services. *Emerg. Med. J.* **2012**, *29*, 60–64. [[CrossRef](#)] [[PubMed](#)]
9. The Data Journalism Handbook. Available online: Datajournalismhandbook.org/ (accessed on 10 June 2016).
10. Song, X.; Zhang, Q.; Sekimoto, Y.; Horanont, T.; Ueyama, S.; Shibasaki, R. Intelligent system for human behavior analysis and reasoning following large-scale disasters. *IEEE Intell. Syst.* **2013**, *28*, 35–42. [[CrossRef](#)]
11. Earle, P.; Guy, M.; Buckmaster, R.; Ostrum, C.; Horvath, S.; Vaughan, A. OMG earthquake! Can twitter improve earthquake response? *Seismol. Res. Lett.* **2010**, *81*, 246–251. [[CrossRef](#)]
12. Google Trends. Available online: Newslab.withgoogle.com/data (accessed on 10 June 2016).
13. Harper, E.M. The economic value of health care data. *Nurs. Adm. Q.* **2013**, *37*, 105–108. [[CrossRef](#)] [[PubMed](#)]
14. Brennan, P.F.; Bakken, S. Nursing needs big data and big data needs nursing. *J. Nurs. Scholarsh.* **2015**, *47*, 477–484. [[CrossRef](#)] [[PubMed](#)]
15. Clancy, T.R.; Reed, L. Big data, big challenges: Implications for chief nurse executives. *J. Nurs. Adm.* **2016**, *46*, 113–115. [[CrossRef](#)] [[PubMed](#)]
16. Stifter, J.; Yao, Y.; Lopez, K.D.; Khokhar, A.; Wilkie, D.J.; Keenan, G.M. Proposing a new conceptual model and an exemplar measure using health information: Technology to examine the impact of relational nurse continuity on hospital-acquired pressure ulcers. *ANS Adv. Nurs. Sci.* **2015**, *38*, 241–251. [[CrossRef](#)] [[PubMed](#)]
17. Lee, K.A.; Dziadkowiec, O.; Meek, P. A systems science approach to fatigue management in research and health care. *Nurs. Outlook* **2014**, *62*, 313–321. [[CrossRef](#)] [[PubMed](#)]
18. Lyon, D.; McCain, N.; Elswick, R.; Sturgill, J.; Ameringer, S.; Jallo, N.; Menzies, V.; Robins, J.; Starkweather, A.; Walter, J. Biobehavioral examination of fatigue across populations: Report from a p30 center of excellence. *Nurs. Outlook* **2014**, *62*, 322–331. [[CrossRef](#)] [[PubMed](#)]
19. Oracle: Big Data for the Enterprise. Available online: www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf (accessed on 10 June 2016).
20. Dion, M.; AbdelMalik, P.; Mawudeku, A. Big data and the global public health intelligence network (GPHIN). *CCDR* **2015**, *41*, 209–214.
21. De Mauro, A.; Greco, M.; Grimaldi, M. What is big data? A consensual definition and a review of key research topics. *AIP Conf. Proc.* **2015**, *1644*, 97–104.
22. Definitions of Big Data. Available online: www.opentracker.net/article/definitions-big-data (accessed on 10 June 2016).
23. 3D Data Management: Controlling Data Volume, Velocity and Variety. Available online: blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf (accessed on 10 June 2016).
24. By the Numbers: 200 Surprising Facebook Statistics (April 2016). Available online: expandedramblings.com/index.php/by-the-numbers-17-amazing-facebook-stats/ (accessed on 10 June 2016).
25. Fu, K.W.; Chan, C.H.; Chau, M. Assessing censorship on microblogs in China: Discriminatory keyword analysis and the real-name registration policy. *IEEE Int. Comput.* **2013**, *17*, 42–50. [[CrossRef](#)]

26. Fu, K.W.; Cheng, Q.; Wong, P.W.C.; Yip, P.S.F. Responses to a self-presented suicide attempt in social media. *Crisis* **2013**, *34*, 406–412. [[CrossRef](#)] [[PubMed](#)]
27. Hair, J.F. *Multivariate Data Analysis*; Prentice Hall: Upper Saddle River, NJ, USA, 2010.
28. ICN Framework of Disaster Nursing Competencies. Available online: www.icn.ch/images/stories/documents/networks/DisasterPreparednessNetwork/Disaster_Nursing_Competencies_lite.pdf (accessed on 13 July 2016).
29. Wong, H.T.; Lai, P.C. Weather factors in the short-term forecasting of daily ambulance calls. *Int. J. Biometeorol.* **2014**, *58*, 669–678. [[CrossRef](#)] [[PubMed](#)]
30. Wong, H.T. Biometeorological Modelling and Forecasting of Ambulance Demand for Hong Kong: A Spatio-Temporal Approach. Ph.D. Thesis, The University of Hong Kong, Hong Kong, China, February 2012.
31. Kuo, M.H.; Sahama, T.; Kushniruk, A.W.; Borycki, E.M.; Grunwell, D. Health Big Data Analytics: Current Perspectives, Challenges and Potential Solutions. *Int. J. Big Data Intell.* **2014**, *1*, 114–126. [[CrossRef](#)]
32. Lai, P.C.; Wong, H.T. Weather and age-gender effects on the projection of future emergency ambulance demand in Hong Kong. *Asia Pac. J. Public Health* **2015**, *27*, NP2542–NP2554. [[CrossRef](#)] [[PubMed](#)]
33. Senior Citizen Home Safety Association Website. Available online: www.schsa.org.hk/en/home/index.html (accessed on 10 June 2016).
34. Wong, H.T.; Chiu, M.Y.L.; Wu, C.S.T.; Lee, T.C.; Senior Citizen Home Safety Association. The influence of weather on health-related help-seeking behavior of senior citizens in Hong Kong. *Int. J. Biometeorol.* **2015**, *59*, 373–376. [[CrossRef](#)] [[PubMed](#)]
35. Raghupathi, W.; Raghupathi, V. Big data analytics in healthcare: Promise and potential. *Health Inf. Sci. Syst.* **2014**, *2*, 3. [[CrossRef](#)] [[PubMed](#)]
36. Public Health. Available online: www.who.int/trade/glossary/story076/en/ (accessed on 10 June 2016).
37. Tomines, A.; Readhead, H.; Readhead, A.; Teutsch, S. Applications of electronic health information in public health: Uses, Opportunities and Barriers. *eGEMs* **2013**, *1*, 1019. [[CrossRef](#)] [[PubMed](#)]
38. Wilson, K.; Keelan, J. Coping with public health 2.0. *CMAJ* **2009**, *180*. [[CrossRef](#)] [[PubMed](#)]
39. Carneiro, H.A.; Mylonakis, E. Google trends: A web-based tool for real-time surveillance of disease outbreaks. *Clin. Infect. Dis.* **2009**, *49*, 1557–1564. [[CrossRef](#)] [[PubMed](#)]
40. Narayanan, A.; Shmatikov, V. Robust de-anonymization of large sparse datasets. In Proceedings of the 2008 IEEE Symposium on Security and Privacy, Oakland, CA, USA, 18–22 May 2008.
41. Guidance Regarding Methods for De-Identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. Available online: www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#standard (accessed on 10 June 2016).
42. Bakken, S.; Reame, N. The Promise and Potential Perils of Big Data for Advancing Symptom Management Research in Populations at Risk for Health Disparities. *Annu. Rev. Nurs. Res.* **2016**, *34*, 247–260. [[CrossRef](#)] [[PubMed](#)]
43. Barton, A.J. Big Data. *J. Nurs. Educ.* **2016**, *55*, 123–124. [[CrossRef](#)] [[PubMed](#)]
44. Blair, L.M. Publicly Available Data and Pediatric Mental Health: Leveraging Big Data to Answer Big Questions for Children. *J. Pediatr. Health Care* **2016**, *30*, 84–87. [[CrossRef](#)] [[PubMed](#)]
45. Booth, R.G. Informatics and Nursing in a Post-Nursing Informatics World: Future Directions for Nurses in an Automated, Artificially Intelligent, Social-Networked Healthcare Environment. *Nurs. Leadersh.* **2016**, *28*, 61–69. [[CrossRef](#)]
46. Broome, M.E. Big data, data science, and big contributions. *Nurs. Outlook* **2016**, *64*, 113–114. [[CrossRef](#)] [[PubMed](#)]
47. Linnen, D. The promise of big data: Improving patient safety and nursing practice. *Nursing* **2016**, *46*, 28–34. [[CrossRef](#)] [[PubMed](#)]
48. Linnen, D. The promise of big data: Improving patient safety and nursing practice. *Nursing* **2016**, *46*, 34–45. [[CrossRef](#)] [[PubMed](#)]
49. Remus, S. The Big Data Revolution: Opportunities for Chief Nurse Executives. *Nurs. Leadersh.* **2016**, *28*, 18–28. [[CrossRef](#)]
50. Zurlinden, J. Health Professions Education: Make Big Data Meaningful. *J. Contin. Educ. Nurs.* **2016**, *47*, 199–200. [[CrossRef](#)] [[PubMed](#)]

51. Cohen, B.; Vawdrey, D.K.; Liu, J.; Caplan, D.; Furuya, E.Y.; Mis, F.W.; Larson, E. Challenges Associated with Using Large Data Sets for Quality Assessment and Research in Clinical Settings. *Policy Polit. Nurs. Pract.* **2015**, *16*, 117–124. [[CrossRef](#)] [[PubMed](#)]
52. Cohen, M.Z.; Thompson, C.B.; Yates, B.; Zimmerman, L.; Pullen, C.H. Implementing common data elements across studies to advance research. *Nurs. Outlook* **2015**, *63*, 181–188. [[CrossRef](#)] [[PubMed](#)]
53. Elgin, K.H.; Bergero, C. Technology and the bedside nurse: An exploration and review of implications for practice. *Nurs. Clin. N. Am.* **2015**, *50*, 227–239. [[CrossRef](#)] [[PubMed](#)]
54. Finfgeld-Connett, D. Twitter and Health Science Research. *West J. Nurs. Res.* **2015**, *37*, 1269–1283. [[CrossRef](#)] [[PubMed](#)]
55. Harper, E.M.; Parkerson, S. Powering Big Data for Nursing Through Partnership. *Nurs. Adm. Q.* **2015**, *39*, 319–324. [[CrossRef](#)] [[PubMed](#)]
56. Henly, S.J.; McCarthy, D.O.; Wyman, J.F.; Heitkemper, M.M.; Redeker, N.S.; Titler, M.G.; McCarthy, A.M.; Stone, P.W.; Moore, S.M.; Alt-White, A.C.; et al. Emerging areas of science: Recommendations for Nursing Science Education from the Council for the Advancement of Nursing Science Idea Festival. *Nurs. Outlook* **2015**, *63*, 398–407. [[CrossRef](#)] [[PubMed](#)]
57. Jackson, A. Teaching and Learning about Big Data: Start Small. *Nurse Educ.* **2015**, *40*, 297. [[CrossRef](#)] [[PubMed](#)]
58. McCartney, P.R. Big data science. *MCN Am. J. Matern. Child Nurs.* **2015**, *40*, 130. [[CrossRef](#)] [[PubMed](#)]
59. McConnell, E.A.; Birkett, M.; Shattell, M. The Future of Big Data: Innovative Methodological Approaches. *Issues Ment. Health Nurs.* **2015**, *36*, 478–480. [[CrossRef](#)] [[PubMed](#)]
60. Price, L.E.; Shea, K.; Gephart, S. The Veterans Affairs's Corporate Data Warehouse: Uses and Implications for Nursing Research and Practice. *Nurs. Adm. Q.* **2015**, *39*, 311–318. [[CrossRef](#)] [[PubMed](#)]
61. Samuels, J.G.; McGrath, R.J.; Fetzer, S.J.; Mittal, P.; Bourgoine, D. Using the Electronic Health Record in Nursing Research: Challenges and Opportunities. *West. J. Nurs. Res.* **2015**, *37*, 1284–1294. [[CrossRef](#)] [[PubMed](#)]
62. Sensmeier, J. Big data and the future of nursing knowledge. *Nurs. Manag.* **2015**, *46*, 22–27. [[CrossRef](#)] [[PubMed](#)]
63. Simpson, R.L. Big data and nursing knowledge. *Nurs. Adm. Q.* **2015**, *39*, 87–89. [[CrossRef](#)] [[PubMed](#)]
64. Stifter, J.; Yao, Y.; Lodhi, M.K.; Lopez, K.D.; Khokhar, A.; Wilkie, D.J.; Keenan, G.M. Nurse Continuity and Hospital-Acquired Pressure Ulcers: A Comparative Analysis Using an Electronic Health Record “Big Data” Set. *Nurs. Res.* **2015**, *64*, 361–371. [[CrossRef](#)] [[PubMed](#)]
65. Welton, J.M.; Harper, E.M. Nursing Care Value-Based Financial Models. *Nurs. Econ.* **2015**, *33*, 14–19. [[PubMed](#)]
66. Westra, B.L.; Pruinelli, L.; Delaney, C.W. Nursing Knowledge: 2015 Big Data Science. *Comput. Inform. Nurs.* **2015**, *33*, 427–431. [[CrossRef](#)] [[PubMed](#)]
67. Westra, B.L.; Clancy, T.R.; Sensmeier, J.; Warren, J.J.; Weaver, C.; Delaney, C.W. Nursing Knowledge: Big Data Science-Implications for Nurse Leaders. *Nurs. Adm. Q.* **2015**, *39*, 304–310. [[CrossRef](#)] [[PubMed](#)]
68. Henly, S.J. Mother lodes and mining tools: Big data for nursing science. *Nurs. Res.* **2014**, *63*. [[CrossRef](#)] [[PubMed](#)]
69. Maughan, E.D.; Johnson, K.H.; Bergren, M.D.; Wolfe, L.C.; Cole, M.; Pontius, D.J.; Mendonca, L.L.; Watts, E.; Patrick, K. Standardized data set for school health services: Part 1—Getting to big data. *NASN Sch. Nurse* **2014**, *29*, 182–186. [[CrossRef](#)] [[PubMed](#)]
70. Rashbass, J.; Peake, M. The evolution of cancer registration. *Eur. J. Cancer Care* **2014**, *23*, 757–759. [[CrossRef](#)] [[PubMed](#)]
71. Skiba, D.J. The connected age: Big data & data visualization. *Nurs. Educ. Perspect.* **2014**, *35*, 267–268. [[PubMed](#)]
72. Westra, B.L.; Choromanski, L. Amazing news for sharable/comparable nursing data to support big data science. *Comput. Inform. Nurs.* **2014**, *32*, 255–256. [[CrossRef](#)] [[PubMed](#)]
73. Salcido, R.S. Big data and disruptive innovation in wound care. *Adv. Skin Wound Care* **2013**, *26*. [[CrossRef](#)] [[PubMed](#)]

