

## Big Data in Nursing: A Commentary

George A Zangaro\*

School of Nursing, Walden University, Minneapolis, USA

### Commentary

Received date: 03/04/2018

Accepted date: 12/12/2018

Published date: 20/12/2018

#### \*For Correspondence

George A Zangaro, School of Nursing, Walden University, 100 Washington Ave S. Suite 900, Minneapolis, MN 55401, USA. Tel: 301 461 4829.

E-mail: George.zangaro@mail.waldenu.edu

Keywords: Big data, Nursing

### ABSTRACT

Today, one of the major challenges facing nursing and other health professions is the concept of “big data.” The purpose of this commentary is to briefly define big data, discuss the need to incorporate big data in nursing curriculum to improve outcomes in health care organizations, and propose the development of a Masterfile to more accurately examine and predict the needs of the nursing workforce. The evolution of big data will benefit nursing and the advancement of nursing science by inspiring new ways of thinking about data and improving patient outcomes. In order to meet big data demands of health care organizations, schools of nursing must be encouraged to incorporate Big Data into their curriculum at all levels. The development of a Nursing Masterfile has the potential to support national and state level initiatives, and increase the accuracy with which the supply of the future nursing workforce can be projected. A comprehensive understanding of the use of big data in nursing will enhance the development of nursing science.

### INTRODUCTION

Nursing continues to embrace new challenges in the health care system. Today, one of the major challenges facing nursing and other health professions is the concept of “big data.” Several nursing organizations are encouraging Nursing’s participation in initiatives to embrace data science and increase the use of advanced statistical approaches to analyze big data<sup>[1-4]</sup>. The health care sector collects an enormous amount of data on an annual basis, and that data generated/collected is estimated to double every 12 to 14 months. This massive increase in data creates a major challenge for individuals conducting analysis<sup>[5,6]</sup>. The explosion of data, as well as advancements in computerized clinical technologies, have created an urgent need for nursing schools to incorporate the concept of big data science into their curriculum. Educating the emerging nursing workforce in the use, interpretation and analysis of big data will optimize patient outcomes and increase nursing knowledge<sup>[7]</sup>.

The purpose of this commentary is to briefly define big data, discuss the need to incorporate big data in nursing curriculum to improve outcomes in health care organizations, and propose the development of a Masterfile to more accurately examine and predict the needs of the nursing workforce.

Big data, as it relates to health care, refers to the large size or volume of clinical data that can be analyzed, interpreted, and applied to health care organizations to improve patient outcomes and reduce costs. In addition to volume, big data includes additional characteristics, as described by Brennan and Bakken<sup>[8]</sup>: velocity, variety, veracity and value. **Table 1** provides definitions for these “5 Vs” of big data.

Nursing organizations, government agencies, academic institutions, and private industry are some of the organizations who possess large, complex datasets requiring advanced methodological and analytical skills to manage and analyze. The evolution of big data will benefit nursing and the advancement of nursing science by inspiring new ways of thinking about data and improving patient outcomes. Benefits to nursing also include providing new ways of defining common

data elements. Brennan and Bakken <sup>[8]</sup>, discuss the long history and success that nursing has had developing common data elements and standardizing definitions of variables, which has increased the accuracy of measurement in nursing research.

**Table 1.** Conceptual definition of the 5 Vs.

5 Vs	Conceptual Definition
Volume	Magnitude of the data.
Velocity	Speed at which data are generated and translated into practice.
Variety	Diversity of data types and sources.
Veracity	Uncertainty of data accuracy for its primary purpose, and the appropriateness of the data for secondary use.
Value	Information extracted from the data source(s) leading to the discovery and translation of knowledge.

Note: Conceptual definitions were adapted from "Nursing needs big data and big data needs nursing," by Brennan et al. <sup>[8]</sup>.

Health care organizations increasingly rely on nursing graduates to be more familiar with the management, analysis and interpretation of big data. This reliance will continue to grow and provide competitive advantages to health care organizations who integrate big data into their operating strategies. This is because those organizations who can successfully analyze and apply the plethora of available data will assist them tremendously in meeting the Quadruple Aim.

In order to meet big data demands of health care organizations, schools of nursing must be encouraged to incorporate Big Data into their curriculum at all levels. For example, Gephart and colleagues <sup>[7]</sup> recommend the integration of machine learning, artificial intelligence and advanced statistical techniques into the PhD curriculum. However, both PhD and Doctor of Nursing Practice (DNP) students should be exposed to these new methodologies. The PhD nurse scientist will use advanced data analysis techniques – to include the merging of multiple data sets, conducting data mining to identify trends, and performing survival analysis and structural equation modeling -to generate new knowledge. DNPs should be exposed to and understand these techniques because they are in a position to translate the outcomes from big data analysis to improve organizational and patient outcomes based on existing evidence. If DNPs are not familiar with the methodological techniques used to generate the evidence, they are not going to be well positioned to critique and translate the evidence into practice. From this perspective, it is critical to ensure that DNPs have a complete understanding of these new technologies, since they are on interdisciplinary teams that are making decisions to improve patient care based on existing evidence.

In addition to improving outcomes in health care organizations, big data principles should be applied to managing the nursing workforce. Currently, there are approximately 4 million licensed nurses in the U.S. <sup>[9]</sup>, and nursing would benefit from a big data initiative designed to create a Nursing Masterfile. The Masterfile would contain current and historical information on all licensed nurses in the United States. Currently, all 50 states and the District of Columbia have unique data systems to collect and report data on their nursing workforce. The data being collected in each state differs and this does not allow for adequate comparisons across states. Additionally, several professional nursing organizations develop databases that consist of demographic information on its members but do not include non-members. The American Community Survey, the Current Population Survey and the National Sample Survey of Registered Nurses are some of the more common databases used to research nursing trends and predict future supply needs of the nursing workforce. However, these are self-report surveys and they do not produce an inventory of all nurses in the United States. Because of this limitation, the estimates that are produced by different organizations or individuals may vary widely.

## DISCUSSION AND CONCLUSION

A Nursing Masterfile would standardize the data elements that are available for analysis across all states and DC and would require data sharing across all these jurisdictions. Increased data sharing would promote a more comprehensive understanding of the nursing workforce at the national and state levels, and provide clearer insights on distributional issues in nurse supply. Given these various methods of data collection, there is a need to bring together representatives from every state to develop a universally agreed upon minimum data set to standardize the collection of common data elements that would be included in a Nursing Masterfile. In addition to nurse scientists and directors of nursing workforce centers, the inclusion of economists, epidemiologists, data scientists and other experts would be needed to ensure the universally accepted variables to be included in the Masterfile would be the appropriate variables needed to conduct various types of research.

The development of a Nursing Masterfile has several advantages:

- (1) Develop an inventory of all licensed nurses in the U.S.
- (2) Apply a consistent methodological approach to collecting Masterfile data.
- (3) Improve the accuracy of forecasting the supply of nurses.
- (4) Maintain a current source of data to support state level initiatives.
- (5) Allow for states to make accurate comparisons on workforce issues.

The main question concerning the development of a Nursing Masterfile is: Who will initiate, maintain and update the file? One option may be for the National Forum of State Nursing Workforce Centers to lead the development of the Masterfile, but with state participation. Another option may be for a major professional nursing organization to lead a consortium with different nursing organizations contributing funding to initiate the development of the Masterfile. As part of determining who will own this process and move it forward, the costs and funding aspects of this initiative, both initial and ongoing, must be addressed. Additionally, there are ethical and privacy concerns when maintaining a file of this nature. These concerns will need to be addressed once the minimum data file is created and the variables in the file are universally agreed upon.

The era of big data and advanced statistical methods are significant factors in the advancement of nursing research and the translation of nursing research to practice. The nursing profession needs to maximize the benefits associated with analyzing Big Data, with a focus on improving outcomes for patients and health care organizations. Nursing is challenged with expanding or revising curriculum to integrate new methodological approaches to analyze big data and prepare the emerging nursing workforce to practice in a rapidly changing health care environment. The development of a Nursing Masterfile has the potential to support national and state level initiatives, and increase the accuracy with which the supply of the future nursing workforce can be projected.

### REFERENCES

1. Clancy TR, et al. A call to action: Engage in big data science. *Nursing Outlook*. 2014;62:64-65.
2. National Association of School Nurses. *NASN Research Priorities 2018-19*. Silver Spring, NASN. 2018.
3. National Institute of Nursing Research. *The NINR strategic plan: Advancing science, improving lives*. Bethesda, MD. 2018.
4. National League for Nursing. *NLN research priorities in nursing education: 2016-2019*. Washington, DC. 2016.
5. Dinov ID. Volume and value of big healthcare data. *J Med Stat Inform*. 2016;4:1-7.
6. Feinleib D. The big data landscape. In: *Big data bootcamp*. A press, Berkeley, CA. 2014.
7. Gephart SM. Perspectives on policy and the value of nursing science in a big data era. *Nurs Sci Q*. 2018;31:78-81.
8. Brennan PF, Bakken S. Nursing needs big data and big data needs nursing. *J Nurs Scholarsh*. 2015;47:477-484.
9. American Nurses Association. *About ANA*. 2018.