

# Nursing and Data: Powering Nursing Leaders for Big Data Science

**Lisiane Pruinelli<sup>1</sup>**

ORCID: 0000-0002-1046-6037



<sup>1</sup>PhD, MS, RN, Fellow of the American Medical Informatics Association (FAMIA). Assistant Professor, School of Nursing and Affiliate Faculty, Institute for Health Informatics, University of Minnesota. Minnesota, United States.

## How to cite this article:

Pruinelli L. Nursing and Data: Powering Nursing Leaders for Big Data Science. Rev Bras Enferm. 2021;74(4):e740401. <https://doi.org/10.1590/0034-7167.2021740401>

More comprehensive knowledge and improved data science skills are needed by nurses and nursing leaders who conduct analytics using large sets of nursing data with the goal of improving population outcomes. Nursing informatics competencies, not just for nurse informaticians, but for nurse leaders as well, need to move beyond the current formal training to a more competency-based model, including education and practice. Very few efforts have been made to advance nursing understanding and skillsets required for using data for clinical applications in real-world scenarios. Below are some concepts and current initiatives to address, inform, and discuss future directions on nursing big data science are presented.

## Big Data Science

Although big data has been used and is a well-accepted concept in the nursing field, moving beyond big data to data science applications is needed. The concept of big data has been defined by several authors, who have described the several “Vs” comprising its definition. Over the years, more “Vs” have been added to that definition; however, the basic concepts include *Volume, Velocity, Variety, Veracity, and Value*, with more details found elsewhere<sup>(1)</sup>. On the other hand, data science has a broader definition, including all the processes around the use of big data. It’s defined as the “field with a broad scope, encompassing approaches for generation, characterization, management, storage, analysis, visualization, integration and use of large, heterogeneous data sets that have relevance to population health”<sup>(1)</sup>. Best practices for conducting data science research and initiatives has been defined by several institutions, such as the United States National Institute of Health (NIH - <https://datascience.nih.gov/>), National Institute of Nursing Research (NINR)<sup>(2)</sup>. Globally, the European Data Initiative (EUDAT - <https://eudat.eu/european-data-initiative>), the Brazilian GO FAIR Brazil, and *Brazil Open Data Portal*<sup>(3)</sup> have the common goal of making safe use and sharing of data available to everyone. These institutions have laid the foundation for promising directions for workforce development and new partnerships to advance the field of data science and its use.

## Nursing Data Science Initiatives

Several national organizations in the United States are pursuing initiatives to improve health outcomes that would benefit from big data science and analytics, such as the Personalized Medicine Initiative (PMI - <https://www.pmi.gov/>), the Symptom Science Model (SSM)<sup>(2)</sup>, and the Nursing Knowledge Big Data Science (NKBDS) conference ([z.umn.edu/bigdata](http://z.umn.edu/bigdata)). The NKBDS has pointed out that “extract[ing] meaningful insights from Big Data in order to transform health care, while improving clinical operations, care quality, cost-effectiveness, patient safety, and outcomes” is needed. Under this initiative, the *Data Science and Clinical Analytics Workgroup* has developed a framework for educating nurse leaders on the use of Data Science<sup>(4)</sup>. This work is a continuous commitment to advance the understanding of data science, where members from around the world

participate in discussions, hands-on learning, and the application of the life-cycle framework process.

What all of these initiatives have in common is the adoption of data science techniques to identify populations that would benefit from interventions resulting in better outcomes. However, many nurse leaders, nurse informaticians and health care leaders lack a broad understanding of the concepts and resources available to conduct real-world analytic projects that would have the potential to inform health conditions and outcomes. While there are some early theoretical initiatives focused on a data science roadmap (i.e. Big Data to Knowledge), currently, few available and practical frameworks using data science on large data sets are suitable and replicable for nurse leaders working in clinical practice. One example is the just mentioned Roadmap Data Science for Nursing Leaders<sup>(4)</sup>. This framework (or roadmap) was built with the goal of educating nurses on using data science principles and tools to inform decision-making. It is envisioned that this framework can be adopted and used by nurse leaders in the continuum of care to address critical health and population issues.

### Nursing Data Science Education

When building data science curriculum, several aspects need to be considered. There is a lack of educational resources, from textbooks to trained faculty, on data science for clinical applications that incorporate a holistic (or patient-centric) approach to health. Many new textbooks have been published with no success in addressing the whole picture of a life-cycle data science project. As a result, there is a greater need of faculty with experience in the application of such methodologies. With the combination of both appropriate textbooks and further faculty training, it is possible to deliver a comprehensive curriculum on the use and application of data sciences. However, data science principles have not traditionally used nursing data, family data, nor continuum of care data to improve patient outcomes<sup>(5-6)</sup>. Neither have nurses widely used such principles, which are heavily reliant on computing knowledge, to implement data-driven studies and projects within research, education, and practice to improve patient outcomes. Based on the current state of the nursing data science field, the first step is to develop methodologies to further educate nurses and nurse leaders, aiming ultimately to build a community of nurses focused on data science projects that use multiple types of data to improve population health and tailor interventions to individual patients in the continuum of care.

Several universities have developed focused curricula and serve as exemplars on data science education, including making data science courses required, specifically for graduate-level nursing informatics students. Recently, several universities, such as the University of Minnesota School of Nursing (<https://www.nursing.umn.edu/>), the NIH Big Data 2 Knowledge program (<https://commonfund.nih.gov/bd2k>), the University of San Diego (<https://www.sandiego.edu/nursing/>), and the Columbia School of Nursing, (<https://www.nursing.columbia.edu/>) have updated their curriculum to meet this need.

nursing.umn.edu/), the NIH Big Data 2 Knowledge program (<https://commonfund.nih.gov/bd2k>), the University of San Diego (<https://www.sandiego.edu/nursing/>), and the Columbia School of Nursing, (<https://www.nursing.columbia.edu/>) have updated their curriculum to meet this need.

### Nursing Data Science Research

The adoption of electronic health records (EHR) around the globe and the increasing use of data science and advanced analytics approaches, as well as the surge of several national and international incentives, have set the pace for knowledge discovery and machine-learning applications applied to large sets of healthcare data. There are few studies that have successfully taken advantage of large sources of data with the goal of personalizing the treatment of individual patients by utilizing a wide range of predictors to anticipate, and even prevent, the development of certain health conditions<sup>(7)</sup>. There is a need to go beyond the investigation of comorbidities that are disease-centric and move towards a holistic approach to health. These will allow researchers to use data science for discovering new directions for major population health concerns, such as the COVID-19 pandemic. The inclusion of other determinants of health, such as environmental, economic, educational, geographic, and quality of life, are some examples of data available and ready to be used for nursing big data science research.

In closing, although many advances have been made, and the promises are enormous for nursing, care delivery, and better patient outcomes, some cautions should also be considered and acknowledged. The emergence of Artificial Intelligence (AI) in healthcare applications raises concerns about the use of such “smart” approaches to healthcare delivery. If developed and applied responsibly, AI technologies have the potential to support evidence-based nursing practice through provision of cognitive insights and clinical decision support at the point of care. The ultimate goal would be to build AI-enabled care delivery tools to foster understanding of a patient’s conditions and augment clinical-decision making<sup>(8)</sup>. But this will only be possible once investigators and AI developers take a responsible AI approach and work together with care deliverers. Machine learning, specifically deep-learning techniques and methods, should be able to represent the multi-faceted complexity of conditions that patients experience over time. However, AI models need to be fair and responsible (e.g. responsible AI). As a result, these models could also be applied to examine the complex temporal relationships presented by health conditions and how they influence health progression; thus, aiding to care delivery. Nevertheless, AI-enabled technology needs to be fair in the sense that it should be unbiased and even decrease the already existing health disparity differences between patients, care delivery, and outcomes.

---

## REFERENCES

1. Topaz M, Pruinelli L. Big Data and Nursing: Implications for the Future. *Stud Health Technol Inform[Internet]*. 2017[cited 2021 May 30];232:165-71. Available from: <https://pubmed.ncbi.nlm.nih.gov/28106594/>

2. Cashion AK, Grady PA. The National Institutes of Health. National Institutes of Nursing Research intramural research program and the development of the National Institutes of Health Symptom Science Model. *Nurs Outlook*. 2015;63(4):484-7. <https://doi.org/10.1016/j.outlook.2015.03.001>
  3. Sales L, Henning P, Veiga V, Costa MM, Sayão LF, Santos LS, et al. GO FAIR Brazil: a challenge for Brazilian Data Science. *Data Intelligence*. 2020;2(1-2):238-45. [https://doi.org/10.1162/dint\\_a\\_00046](https://doi.org/10.1162/dint_a_00046)
  4. Pruinelli L, Johnson SG, Fesenmaier B, Winden TJ, Coviak C, Delaney CW. An Applied Healthcare Data Science Roadmap for nursing leaders: a workshop development, conceptualization, and application. *Comput Inform Nurs*. 2020;38(10):484-9. <https://doi.org/10.1097/CIN.0000000000000607>
  5. Brennan PF, Bakken S. Nursing needs big data and big data needs nursing. *J Nurs Scholarsh*. 2015;47(5):477-84. <https://doi.org/10.1111/jnu.12159>
  6. Zhu R, Han S, Su Y, Zhang C, Yu Q, Duan Z. The application of big data and the development of nursing science: a discussion paper. *Int J Nurs Sci*. 2019;6(2):229-34. <https://doi.org/10.1016/j.ijnss.2019.03.001>
  7. Schaefer GO, Tai ES, Sun S. Precision Medicine and Big Data: The Application of an Ethics Framework for Big Data in Health and Research. *Asian Bioeth Rev*. 2019;11(3):275-88. <https://doi.org/10.1007/s41649-019-00094-2>
  8. Nyrup R, Ronquillo C, Bakken S, Beduschi A, Cato K, Chu C, et al. Artificial intelligence in nursing: priorities and opportunities from an international invitational think-tank of the Nursing and Artificial Intelligence Leadership Collaborative. *J Adv Nurs*. 2021. <https://doi.org/10.17863/CAM.66238>
-