



Nursing Orientation to Data Science and Machine Learning

A primer on how ‘big data’ and new analytic models are transforming nursing.

ABSTRACT: Nurses collect, use, and produce data every day in countless ways, such as when assessing and treating patients, performing administrative functions, and engaging in strategic planning in their organizations and communities. These data are aggregated into large data sets in health care systems, public and private databases, and academic research settings. In recent years the machines used in this work (computer hardware) have become increasingly able to analyze large data sets, or “big data,” at high speed. Data scientists use machine learning tools to aid in analyzing this big data, such as data amassed from large numbers of electronic health records. In health care, predictions for patient outcomes has become a focus of research using machine learning. It’s important for nurses and nurse administrators to understand how machine learning has changed our ways of thinking about data and turning data into knowledge that can improve patient care. This article provides an orientation to machine learning and data science, offers an understanding of current challenges and opportunities, and describes the nursing implications for nurses in various roles.

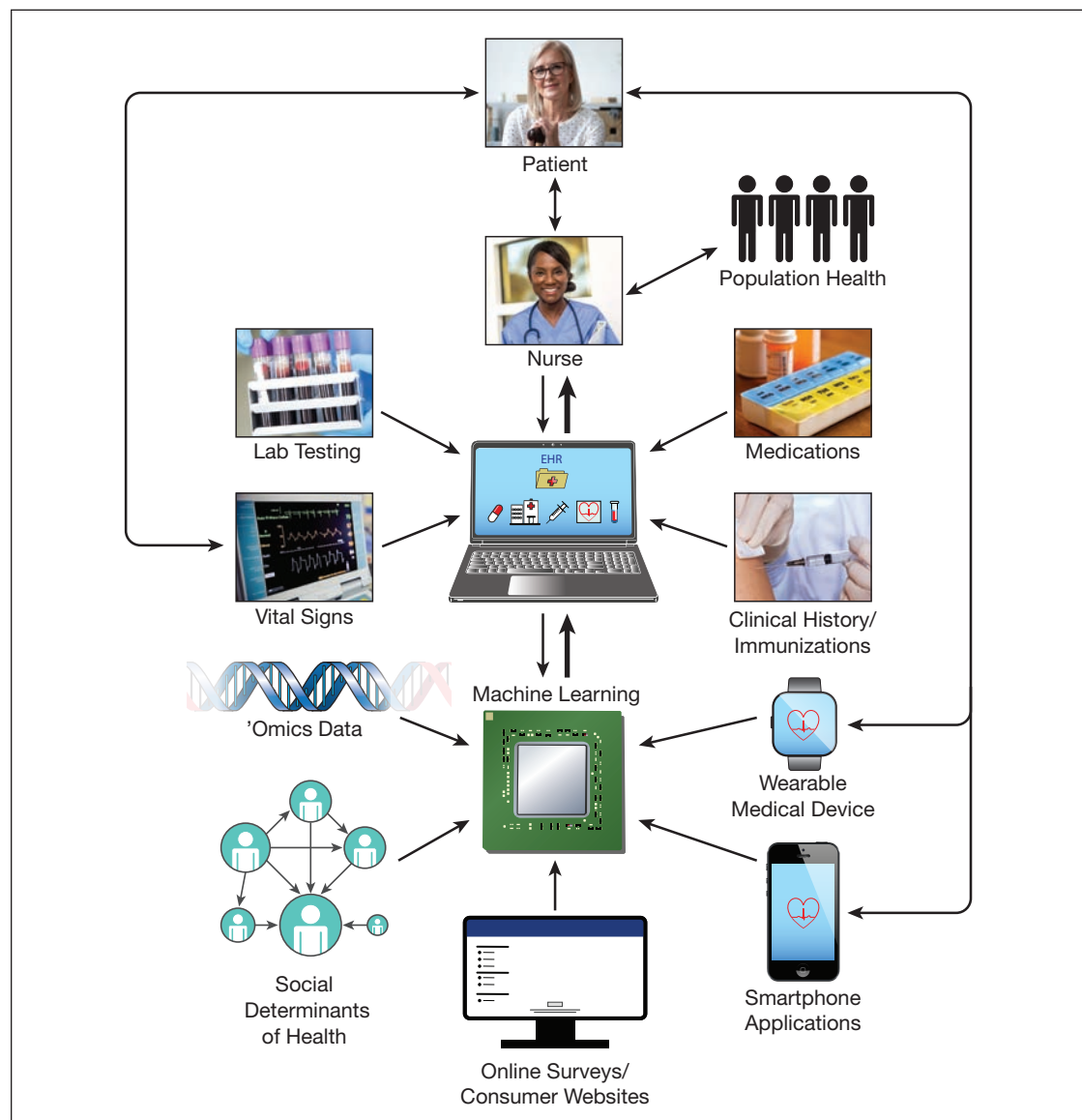
Keywords: algorithm, bias, big data, data science, electronic health record, ethics, evidence-based practice, machine learning

It’s said that change is the one constant in life, and for many people the constant need to adapt and stay current can feel both exciting and daunting. In the health care arena, the fields of data science and machine learning are revolutionizing the ways information is used, and it’s essential for nurses to understand how this impacts nursing practice and how the nursing profession can benefit. In this article, we provide frontline nurses and nurse administrators with an orientation to data science and machine learning, discuss current challenges and opportunities, and describe the implications for various nursing roles. Throughout, we also discuss relevant literature. (For details on the search method, see *Literature Search*.)

A SHORT OVERVIEW

Rapid advances in computers’ capacity to analyze data and reductions in hardware costs have allowed the creation of very large databases, which in turn contain multiple data sets. Many industries stand to gain—including the health care industry, which can use these large databases to inform data-driven, evidence-based care decisions.¹⁻³ In hospitals, health care systems, and other health-related entities, data scientists might be employed to manage large databases compiled from electronic health records (EHRs).¹ They might also use machine learning to develop predictive models that influence clinical decision-making and patient outcomes.

Nursing practice can benefit from these models. For example, predictive models identifying patients

Figure 1. Data Science and Health Care

A depiction of the relationship between data science and machine learning and how it can improve the practice of health care, with an emphasis on bidirectional information exchange and data application for patients, nurses, and the health system. Illustration by Sara Jarret.

at risk for specific postoperative complications could help nurses in evaluating a given patient's discharge readiness and identify appropriate predischarge educational materials.⁴ Indeed, when Alderden and colleagues used machine learning to develop a predictive model for pressure ulcer development in critical care patients, their model performed comparably to the widely used Braden Scale.⁵ Predictive models facilitate preplanning, as appropriate interventions

and specialty equipment can be ordered earlier in a patient's stay, leading to better patient outcomes.

UNDERSTANDING THE BASICS

First, it's important to understand some essential terms and definitions. The following is a short guide, with examples relevant for nurses.

A *data set* is a discrete set of data, whereas a *data-base* comprises multiple data sets. The distinction

Literature Search

We conducted a search of CINAHL and PubMed using the keywords *big data*, *data science*, *informatics*, *machine learning*, and *nursing*. The specified date range was 2009 to 2019. Sixty-nine articles in that range were identified as potentially relevant; of these, 30 were selected for inclusion in the article. We also included one website and made an exception for an older article published in 2003 because of its relevance.

Of the 30 articles published within the specified date range, 20 had been published between 2017 and 2019, reflecting the increasing interest and research in the topic areas. It was noted that data science as a discrete topic was not singled out in many studies, although machine learning is a core function within this field. The role of the data scientist was also not addressed in most studies. Nurses were seldom identified as data scientists; the term “nurse informaticist” was more commonly used.

isn’t always clear, and the terms are sometimes used interchangeably. *Big data* can be defined as digital data that are generated in such high volumes, varieties, and speeds that the resulting data sets are too large for the usual data-processing systems.³ Big data originally emerged from commercial shopping and entertainment sites such as Amazon and Netflix. In health care, big data now comes from multiple sources such as EHRs, biometric medical devices such as “smart” infusion pumps and cardiac monitors, and “smart” watches and other wearable devices (see Figure 1).^{6,7} Data sets compiled from the EHRs of a health care system can be sorted in countless ways according to the patient populations of interest, such as all patients with stroke or all female African American patients with diabetes.

Data science refers both to the field of the study of data and to the principles that guide such study.³ Data scientists seek to extract information from data that can be turned into knowledge. The data scientist drives this process by controlling the “data life cycle,” which has five stages: data capture, maintenance, processing, analysis, and communication.⁸ *Data capture* occurs when data are gathered from one or more sources such as EHRs (data of interest might include laboratory results, imaging data, medications, and other treatment data) and wearable devices. *Maintenance* generally refers to the stage at which data are inserted into an infrastructure (a database or other computational environment). *Processing* refers to the stage when the data are organized into a manageable form. This might include dimensionality reduction (determining the minimum number of random variables to include in analysis), summarization (such as identifying the mean and median, preparing a bar graph), and transformation (such as by displaying data in a normal distribution curve). *Analysis* involves apply-

ing models to the data using machine learning tools (such as statistical analytics software) and identifying patterns and relationships. *Communication*, the last stage, refers to the act of presenting the results of the analysis to appropriate stakeholders, whether through publications, presentations, or internally within an organization.

Machine learning refers to a kind of artificial intelligence in which a machine (a computer or array of computers) applies statistical algorithms to the data set of interest. The machine “learns” to apply the algorithms to the data and builds a statistical model. This model then combs through hundreds of variables from thousands of individual sources, such as patients, to find patterns and groupings. Machine learning can also refer to the field of study in which data scientists explore how computers learn from data and develop appropriate algorithms to facilitate such learning.³ The machine learning toolbox consists of many classic statistical methods of analyses such as linear regression. In addition, machine learning can create specialized models capable of analyzing complex multidimensional data, such as magnetic resonance imaging or ultrasound results. For example, using thousands of X-rays of a specific bone cancer, researchers can use machine learning to “train” a model. The model can then analyze new, unread X-rays and predict which patients are more likely to have the same cancer.

Unlike hypothesis testing, which involves making observations about variables and testing possible causal relationships among them, machine learning allows researchers to predict future outcomes without testing for causality. For example, a public health researcher might want to test the hypothesis that children with certain nutritional deficits are at higher risk for a specific childhood cancer. Yet machine learning might identify children with this disease clustered by zip code, pointing to environmental conditions that warrant further investigation.

PREDICTING PATIENT OUTCOMES

Researchers are already using machine learning to develop predictive models for hospital readmissions, patient mortality, and patient morbidity. Examples include using machine learning to predict pressure injuries in critical care patients,⁵ delirium in hospitalized patients,⁹ early readmissions for patients with diabetes,¹⁰ and the likelihood of sepsis in ED patients.¹¹ These studies used differing sources of data, including condition-specific registries and hospital-specific EHRs.

It’s important to note that machine learning can only use—and thus is limited by—the data that humans have provided. Missing clinical data and errors in patient administration data (such as transfer and discharge data, and data from classification

systems such as diagnosis-related groups) all affect machine learning findings. Moreover, some nursing assessments aren't captured in EHR data fields, resulting in critical gaps in predictive models used for practice recommendations.

Stafos and colleagues compared nurses' assessment of patient risks discussed during daily huddles with data gathered using an electronic risk assessment tool.¹² The tool assigned a score to each patient for risks such as skin breakdown, falls, need for restraints, and others, based on EHR data. During the huddles, nurses verbally shared concerns about psychosocial, behavioral, and other risk factors (such as confusion, family dynamics, and drug or alcohol withdrawal) that were not captured in the EHRs. The researchers found that nurses perceived additional risks besides those reflected in EHR data, while the tool assigned risk based solely on EHR data. As this finding shows, the limitations of machine learning predictive models should be taken into account in clinical decision-making.

in 25,000 hospitalized patients, nine were features of nursing assessment.⁹ In descending order of significance, these were disorientation, ability to dress self in the hospital, lifestyle, number of days since last discharge, ability to walk, excretion, nursing support, ability to dress self at home, and fluid intake. The researchers reached two important conclusions: that data from nursing assessments made at and before the current hospitalization can potentially predict delirium; and that nursing data add independent information, outside of clinical and demographic data, that can strengthen the predictions of machine learning models.

Given that nurses' narrative notes often contain rich details and insights, it's clear that finding a way to translate unstructured into structured data is highly desirable.¹³ Currently, narrative data aren't likely to be included in machine learning processes because standardized ways to translate that data haven't been created. Researchers are advocating the development of national EHR templates that

The importance of the accuracy of clinical documentation cannot be overemphasized.

In another study, Horng and colleagues looked at records from nearly 231,000 ED patient visits to discover whether sepsis could be predicted earlier in a patient's stay.¹¹ Machine learning models were applied to both *structured data* (such as diagnostic codes, physiological data, and patient demographics) and *unstructured data* (free-text data such as triage notes, responses to cognitive status questions, and patients' statements of their chief complaint). The researchers concluded that, in addition to using structured data, adding unstructured data to their predictive model made earlier identification of sepsis significantly more likely. In the future, EHR vendors may embed predictive models in their software to support clinical decision-making. It's important for vendors to consider how unstructured data collected by nurses can be captured and included in those models.

NURSING DOCUMENTATION

To date, few studies have addressed the input of the vast amount of information that nurses routinely gather for every hospitalized patient—much of it in free-text form—into predictive models. Veeranki and colleagues found that, of the 20 most important data points useful in predicting delirium

include both structured and unstructured data entry options.¹⁴⁻¹⁶ As Moon and colleagues have stated, "Creating a standardized approach to measuring individual nurse contribution in patient outcomes is innovative" and will make the value of nursing more recognizable.¹⁵

One promising option for translating unstructured data is natural language processing (NLP). NLP methods take narrative entries and process the text to capture its data in machine-usable form.¹⁷ Researchers have used NLP in conjunction with EHRs in disease-specific studies; but many aspects of free text such as abbreviations, shorthand, condition-specific vocabularies, and spelling errors or variations remain challenging.¹³ In both the private and public sectors, software system developers are working to expand the use of NLP in EHRs and make it routinely available to researchers.¹⁷ This would facilitate analyses of subjective data such as the reasoning behind a given action, the extent of communication among providers, and a patient's experience of treatment.

POTENTIAL FOR BIAS IN MACHINE LEARNING MODELS

The presence of bias in data analysis is well known, and machine learning is subject to the same pres-

ence.¹⁸ Common reasons for bias include missing data and poor data quality.

Missing data. Bias occurs when characteristics of the phenomenon being studied are not captured in the data. The problem of missing data pervades all fields of research. Various machine learning methods have been developed to provide substitutes, thereby allowing the completion of statistical analyses.^{19,20} But while such methods might reduce bias, they can't eliminate it, and machine learning processes can still produce biased predictive models. Busy clinicians may not understand how missing data points in the EHR might affect the inferences drawn from data analyses, which are then used to create policy and guide practice.

record of the care provided, it communicates the decisions made about a patient's care strategy, and it's the basis for billing and reimbursement.¹⁶ As more and more researchers use machine learning models to analyze EHR data, missing variables and values threaten the integrity of the results.

Data quality. As health care increasingly makes use of data science, the importance of working with high-quality data will also rise. In clinical documentation, there will always be competing pressures between the need for accuracy and the need for completeness. With regard to the EHR, *accuracy* means that the correct information is documented in the correct place. *Completeness* means that enough variables have been recorded (enough data

Machine learning has the potential to advance the understanding, treatment, and prevention of disease.

Missing data occurs in two forms: missing variables and missing data values. *Missing variables* refers to instances when a variable, or variables, needed for inclusion in a predictive model is missing from the data set. For example, if the variable "confusion" was absent from an aggregated data set used to develop a predictive model for delirium, the data analysis could be rendered incomplete. *Missing data values* refers to instances when the data entry field for a specific variable is left blank. For example, if clinical assessment values for blood pressure and fluid volume infusion were consistently lacking from a data set used to develop a predictive model for preeclampsia or preterm labor, those omissions would skew the results.

Regarding nursing practice, the impact of missing data is twofold: specific variables in the EHR that reflect nursing interventions might be lacking, and patient records might be incomplete. The first issue lies beyond the scope of this article. Briefly, the omission of variables that reflect nursing interventions occurs when the preselected variables in the EHR don't accurately represent those interventions.²¹ Many nurse researchers and informaticists have called for nurses to be involved in defining the nursing actions placed in EHRs in order to build nursing knowledge and link nursing interventions to patient outcomes.^{22,23}

The importance of the accuracy of clinical documentation cannot be overemphasized. It's the legal

fields have been filled in) to ensure a meaningful data set for research.¹

In patient care, data entry into the EHR generally happens at the point of care. By performing EHR data entry accurately and with integrity, nurses have the power to influence how big data and machine learning affect clinical decisions. Thus, orientation programs should teach this rationale to new nurses, and in-service trainings that incorporate illustrative case studies should reinforce it for current staff. Links to, or printouts of, an institution's EHR templates and flow sheets will also help new hires become familiar with data expectations.

ETHICAL CONCERNS

In biomedical and nursing research, machine learning has the potential to advance the understanding, treatment, and prevention of disease, as well as of trauma resulting from accident or injury. But ethical concerns have been raised regarding the use of big data in practice change and policy development.

Discrimination. *Discriminatory data mining* refers to looking for data to support certain assumptions. This is a risk in any research, including research that uses large data sets.²⁴ For example, selectively omitting certain demographic data (such as age, gender, and race or ethnicity) could result in clinical recommendations or policies that adversely affect an entire population.

Anonymity and confidentiality. *Reidentification of patients* may also be a concern. Atreya and col-

leagues reviewed laboratory results from a data set that included 8.5 million laboratory results for over 61,000 inpatients, using a “threat model” to test reidentification risk.²⁵ They concluded that although laboratory data alone held minimal risk, when combined with other pieces of information normally submitted with a sample or specimen, the potential for reidentification was higher. They cautioned researchers to be aware of this risk and recommended further research into reducing it.

What about HIPAA? Historically, machine learning has been used in commercial settings that don’t fall under the purview of the Health Insurance Portability and Accountability Act (HIPAA). Data collected and stored or shared by health care insurers, providers, and their business partners are protected through HIPAA, which regulates access to and storage of that data. But today’s health-conscious consumers often allow their data to be gathered through wearable devices such as smart watches and heart monitors, smartphone applications, and online surveys and consumer-oriented forums.²⁶ Data collected by these routes aren’t protected by HIPAA regulations, which have no control over how aggregated data are used by entities outside the health care industry, such as smartphone providers, credit card companies, social media platforms, and search engines.^{24, 27} Although an employer-based wellness program might be covered as a health care business partner under HIPAA, many other types of individual, consumer, and community-based wellness programs are not. Even when people are asked to accept and sign a terms and conditions agreement, they may not understand what their data are being used for and whether their data are being sold or exchanged with third parties.^{24, 27, 28} Anonymity, confidentiality, informed consent, and database security are all areas of concern that warrant more research.

NURSING IMPLICATIONS

As big data and machine learning become more widely used in describing and predicting health care outcomes, it’s ever more important to find ways to fully integrate nursing interventions into EHRs. All research must address concerns about data quality, bias, patient privacy and confidentiality, and data discoverability. With regard to machine learning analyses, particular concerns can include the generalizability of findings, the trade-off between data variance (complexity) and bias, the time and expense involved in working with large data sets, and the availability of computer access and storage. As the field of big data research expands through machine learning models, risks for unintended consequences will have to be resolved, and rules for ethics review established.

Recommended Reading^a

On Machine Learning

Hardy LR. Using big data to accelerate evidence-based practice. *Worldviews Evid Based Nurs* 2018;15(2): 85-7.

A short discussion of evidence-based nursing practice and using large data sets to move evidence-based practice forward.

Meyfroidt G, et al. Machine learning techniques to examine large patient databases. *Best Pract Res Clin Anaesthesiol* 2009;23(1):127-43.

General overview of machine learning techniques, with discussion of some specific algorithms.

National Academies of Sciences, Engineering, and Medicine. *Artificial intelligence and machine learning to accelerate transitional research: proceedings of a workshop—in brief*. Washington, DC: National Academies Press; 2018.

Brief discussion of artificial intelligence and machine learning in the contexts of diagnosis, treatment, and drug discovery; emerging opportunities; and the future of the workforce.

Studies Applying Machine Learning to Health Care

Gao C, et al. Predicting length of stay for obstetric patients via electronic medical records. *Stud Health Technol Inform* 2017;245:1019-23.

Using three years of billing and electronic health record (EHR) data for 9,188 patients, the researchers created a machine learning prediction model to see whether they could more accurately predict patients’ lengths of stay.

Gundlapalli AV, et al. Detecting the presence of an indwelling urinary catheter and urinary symptoms in hospitalized patients using natural language processing. *J Biomed Inform* 2017;71S:S39-S45.

Using natural language processing, researchers developed a lexicon from free-text nursing notes in the EHR and used it in large-scale chart review aimed at improving the identification of catheter-related urinary tract infections.

Jeffery AD, et al. Advancing in-hospital clinical deterioration prediction models. *Am J Crit Care* 2018;27(5):381-91.

The researchers compared two machine learning algorithms (random forest and random survival forest) and two statistical models (logistic regression and Cox proportional hazards regression) regarding their accuracy in predicting inpatient cardiopulmonary arrest.

^a Some articles are free; some are available for a small fee.

The following are the implications that data science and machine learning have for frontline, administrative, and academic and research nurses.

Frontline nurses drive the quality of documentation. It’s nursing data that are used in predictive models to develop interventions for improving patient care. As care plans, nursing interventions, and discharge planning guides become increasingly embedded in EHRs, nurses will be able to spend less time searching for these resources and more time interacting with patients. To this end, documentation by frontline nurses should be assessed weekly

Databases of Interest to Nurse Researchers

National Database of Nursing Quality Indicators (NDNQI)

<https://nursingandndnqi.weebly.com/ndnqi-indicators.html>

Hospitals from all 50 states and the District of Columbia voluntarily contribute selected unit-level nursing performance indicators to the NDNQI for comparison across regional, state, and national levels.

National Health and Nutrition Examination Survey (NHANES)

www.cdc.gov/nchs/nhanes/index.htm

The NHANES is an ongoing program of studies that assess the health and nutritional status of adults and children in the United States. The interview component includes demographic, socioeconomic, dietary, and health-related questions. The examination component consists of medical, dental, and physiological assessments.

Nursing Value Research Data Warehouse (NVRDW)

<https://news.cuanschutz.edu/news-stories/research-initiative-uses-big-data-to-improve-patient-care>

The NVRDW, which is based at the University of Colorado, pools data collected from various sources within multiple organizations. It's designed to facilitate research and quality improvement efforts to transform health care systems and improve patient outcomes.

Patient-Centered Outcomes Research Institute (PCORI)

www.pcori.org

PCORI funds projects that seek to answer patient-centered questions about health and health care and highlights findings that matter in order to enhance decision-making by patients, caregivers, clinicians, and other stakeholders.

(or more often) by unit leaders, who can check for redundancies and share software shortcuts and tips to encourage efficiency. Accurate documentation also supports frontline nurses in cost containment by facilitating the appropriate distribution of specialty equipment and treatments and potentially reducing lengths of stay. Job satisfaction could rise as individual patient plans of care and discharge readiness improve and the overall documentation burden lessens. Lastly, frontline nurses should seek opportunities to participate in EHR software selection, implementation, and updates, in order to ensure EHR purchases meet their evolving needs.

Nurse administrators. Improving patient care and controlling costs will continue to be top priorities. Machine learning can facilitate better real-time nurse staffing, creating predictive models for ED surges, surgical scheduling, and staffing needs in many other settings. Nurse administrators can work with quality and risk management departments, unit councils, and individual champions to integrate predictive models into daily work and use such models to develop quality improvement projects that meet regulatory and health system requirements. Having higher quality data will also mitigate

liability risks. Opportunities for staff nurses to participate on informatics committees or councils should be supported.

For administrators and frontline nurses, the heightened focus on data quality invites collaboration. Nurse administrators should meet frequently with nursing staff to discover barriers to accurate and complete documentation and find solutions. They can reinforce the meaning and value of such documentation by providing practical examples of how the data nurses input are used. When nurses understand this better, the sense that charting is just “busy work” might change.

Academic and research nurses are well positioned to advocate for the accurate representation of nursing care in machine learning analyses and predictive models. Nursing care-sensitive indicators must be included, not only to improve predictive accuracy but also to build nursing knowledge, which then informs practice, policies, state regulations, and nursing school curricula. The ability to link specific nursing actions to reimbursement will improve when nurse educators and researchers collaborate with developers to ensure that the EHR better captures these actions. At the national level, new entities such as the Center for Nursing Informatics at the University of Minnesota School of Nursing (www.nursing.umn.edu/centers/center-nursing-informatics) offer academic and research nurses a voice in developing the standards for and content of EHR software.

For more on machine learning and its applications in health care, see *Recommended Reading* and *Databases of Interest to Nurse Researchers*.

CONCLUSIONS

The application of data science to health care, in particular the use of machine learning predictive models, shows great promise and suggests several areas for further research. Such models must be capable of fully capturing nursing interventions if they are to complement face-to-face nurse-patient interactions.²⁹ In providing patient care, nurses first assess a patient's health status and needs, then implement evidence-based interventions to help the patient heal, understand their treatments, and plan for the future. This holistic, individualized approach can be enhanced through the use of machine learning methods to identify patterns and trends that improve patient outcomes in discrete populations.^{4, 22} Through a better understanding of how data science and machine learning impact health care, nurses in all roles can ensure that their contributions will drive improvements in patient care. ▼

For 61 additional nursing continuing professional development activities on the topic of electronic information in nursing, go to www.nursingcenter.com.

Roxanne L. O'Brien is retired from her position as lecturer at the School of Nursing, California State University, Fullerton. Matt W. O'Brien is a data scientist at the Center for Digital Health Innovation, University of California San Francisco. Contact author: Roxanne L. O'Brien, roxanne.obrien@gmail.com. The authors and planners have disclosed no potential conflicts of interest, financial or otherwise. A podcast with the authors is available at www.ajnonline.com.

REFERENCES

- Al-Rawajfah OM, et al. Use of electronic health-related datasets in nursing and health-related research. *West J Nurs Res* 2015;37(7):952-83.
- Provost F, Fawcett T. Data science and its relationship to big data and data-driven decision making. *Big Data* 2013;1(1):51-9.
- Sanchez-Pinto LN, et al. Big data and data science in critical care. *Chest* 2018;154(5):1239-48.
- Byrne MD. Machine learning in health care. *J Perianesth Nurs* 2017;32(5):494-6.
- Alderden J, et al. Predicting pressure injury in critical care patients: a machine-learning model. *Am J Crit Care* 2018;27(6):461-8.
- Darcy AM, et al. Machine learning and the profession of medicine. *JAMA* 2016;315(6):551-2.
- Westra BL, et al. Nursing knowledge: big data science-implications for nurse leaders. *Nurs Adm Q* 2015;39(4):304-10.
- UC Berkeley School of Information. *What is data science?* 2020. <https://datascience.berkeley.edu/about/what-is-data-science>.
- Veeranki SPK, et al. Effect of nursing assessment on predictive delirium models in hospitalised patients. *Stud Health Technol Inform* 2018;248:124-31.
- Kwon JY, et al. Nurses "seeing forest for the trees" in the age of machine learning: using nursing knowledge to improve relevance and performance. *Comput Inform Nurs* 2019;37(4):203-12.
- Horng S, et al. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. *PLoS One* 2017;12(4):e0174708.
- Stafos A, et al. CE: Original research: identifying hospitalized patients at risk for harm: a comparison of nurse perceptions vs. electronic risk assessment tool scores. *Am J Nurs* 2017;117(4):26-31.
- Kreimeyer K, et al. Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. *J Biomed Inform* 2017;73:14-29.
- Keenan GM. Big data in health care: an urgent mandate to CHANGE nursing EHRs! *Online J Nurs Inform* 2014;18(1).
- Moon LA, et al. Nursing value user stories: a value measurement method for linking nurse contribution to patient outcomes. *Comput Inform Nurs* 2019;37(3):161-70.
- Willbanks BA, Moss J. Evidence-based guidelines for interface design for data entry in electronic health records. *Comput Inform Nurs* 2018;36(1):35-44.
- Demner-Fushman D, Elhadad N. Aspiring to unintended consequences of natural language processing: a review of recent developments in clinical and consumer-generated text processing. *Yearb Med Inform* 2016(1):224-33.
- Crown WH. Potential application of machine learning in health outcomes research and some statistical cautions. *Value Health* 2015;18(2):137-40.
- Bose E, et al. Machine learning methods for identifying critical data elements in nursing documentation. *Nurs Res* 2019;68(1):65-72.
- Shivasabesan G, et al. Missing data in trauma registries: a systematic review. *Injury* 2018;49(9):1641-7.
- Goodwin L, et al. Data mining issues and opportunities for building nursing knowledge. *J Biomed Inform* 2003;36(4-5):379-88.
- Brennan PF, Bakken S. Nursing needs big data and big data needs nursing. *J Nurs Scholarsh* 2015;47(5):477-84.
- Jenkins P, et al. Academic/practice/industry collaboration to develop nursing value research data warehouse governance. *Nurs Econ* 2018;36(5):207-12; 251.
- National Committee on Vital and Health Statistics. *Health information privacy beyond HIPAA: a 2018 environmental scan of major trends and challenges*. Washington, DC: U.S. Department of Health and Human Services; 2017 Dec 13. https://ncvhs.hhs.gov/wp-content/uploads/2018/05/NCVHS-Beyond-HIPAA_Report-Final-02-08-18.pdf.
- Atreya RV, et al. Reducing patient re-identification risk for laboratory results within research datasets. *J Am Med Inform Assoc* 2013;20(1):95-101.
- Milton CL. The ethics of big data and nursing science. *Nurs Sci Q* 2017;30(4):300-2.
- Parasidis E, et al. A Belmont report for health data. *N Engl J Med* 2019;380(16):1493-5.
- Cohen IG, Mello MM. HIPAA and protecting health information in the 21st century. *JAMA* 2018;320(3):231-2.
- Foundations S. Systems biology for nursing in the era of big data and precision health. *Nurs Outlook* 2018;66(3):283-92.

Lippincott®
NursingCenter®

TEST INSTRUCTIONS

- Read the article. Take the test for this nursing continuing professional development (NCPD) activity online at www.nursingcenter.com/ce/ajn. Tests can no longer be mailed or faxed.
- You'll need to create an account (it's free!) and log in to your personal NCPD planner account before taking online tests. Your planner will keep track of all your Lippincott Professional Development (LPD) online NCPD activities for you.
- There's only one correct answer for each question. The passing score for this test is 7 correct answers. If you pass, you can print your certificate of earned contact hours and access the answer key. If you fail, you have the option of taking the test again at no additional cost.
- For questions, contact LPD: 1-800-787-8985.
- Registration deadline is March 3, 2023.

NCPD

Nursing Continuing
Professional Development

PROVIDER ACCREDITATION

LPD will award 2 contact hours for this NCPD activity.

LPD is accredited as a provider of NCPD by the American Nurses Credentialing Center's Commission on Accreditation.

This activity is also provider approved by the California Board of Registered Nursing, Provider Number CEP 11749 for 2 contact hours. LPD is also an approved provider of continuing nursing education by the District of Columbia, Georgia, and Florida, CE Broker #50-1223. Your certificate is valid in all states.

PAYMENT

The registration fee for this test is \$21.95.