Using Data Mining Strategies in Clinical Decision Making

A Literature Review

Lu-Yen A. Chen, MN, BSc, RN, Tonks N. Fawcett, MSc, BSc (Hons), RN, RNT, PFHEA

Several data-mining models have been embedded in the clinical environment to improve decision making and patient safety. Consequently, it is crucial to survey the principal data-mining strategies currently used in clinical decision making and to determine the disadvantages and advantages of using these strategies in data mining in clinical decision making. A literature review was conducted, which identified 21 relevant articles. The article findings showed that multiple models of data mining were used in clinical decision making. Although data mining is efficient and accurate, the models are limited with respect to disease and condition.

KEY WORDS: Clinical decision making, Data mining, Nursing

For those working in the nursing profession, decision making for caregiving is constantly required in real time in the clinical environment. Therefore, achieving a correct understanding of the decision-making process, as well as a high level of quality in clinical decision making, is crucial to minimize risk and error in the press of real time for the purpose of patient safety.

Decision Making

Theories and models of decision making have been presented and studied over decades. Paley et al[1] and Bjørk and Hamilton[2] categorized decision making as having two kinds of validity: rational systematic-positivist (decision making is analytical and logical) and phenomenological (interpretive, with a more intuitive approach). Concept attainment theory was discussed by Aitken[3] as a rationalist way to make decisions. Aitken presented the concept attainment theory as a four-stage process, from generating attributes to developing hypotheses.[2–4] This type of step-by-step information process is considered a linear process. It is arguable whether a decision-making process is linear. Social judgment theory is another of the important theories in clinical decision making. It explains the way in which the decision maker places values on the information. Consequently, information that is applied in different circumstances can result in a different judgment.[5] Probability theory, also known as Bayesian theory, is a statistical model that is used to calculate the probability of an event in order to make a decision with a normative approach.[6] Decision trees (a functional strategy for complicated situations) demonstrate the utilities and outcomes of each option.

In addition to the above rationally based theories, intuition plays an important role in decision making. However, "intuition has seldom been granted legitimacy as a sound approach to clinical judgment."[7] Benner and Tanner[7] emphasize the importance of intuition, which distinguishes expert judgment from inexpert, mechanical judgment. It is experts who can complete a clinical picture with efficiency and validity.[1] The development of dual process theory effectively combines the reality and relative virtues of both intuitive and rational theory.[8] There are two systems in dual process theory: System 1 is described as fast, holistic, and unconscious, whereas System 2 is slow, analytic, and conscious reasoning. Although System 1 and System 2 cooperate with each other, they do not appear necessarily to be operating at the same time.[8] Finally, it is worth noting a seven-stage theory of decision making (Table 1) suggests that certain stages in the decision-making process are often overlooked, including the recognition and formulation of the problem, action, and feedback.[9]

Data Mining

The term "data mining" encompasses understanding and interpreting the data by computational techniques from statistics, machine learning, and pattern recognition, in order to predict other variables or identify relationships within the information. According to Finlay,[10(p2)] data mining is commonly used to "identify relationships in data that give an insight into individual preferences," especially "what someone is likely to do in a given scenario." As far back as 1970, decision theory, combined with probabilistic criteria, was

Author Affiliations: The School of Health in Social Science, The University of Edinburgh, Edinburgh, United Kingdom

The authors have disclosed that they have no significant relationship with, or financial interest in, any commercial companies pertaining to this article.

Corresponding author: Tonks N. Fawcett, MSc, BSc (Hons), RN, RNT, PFHEA, The School of Health in Social Science, The University of Edinburgh, Medical School (Doorway 6), Teviot Place, Edinburgh, EH8 9AG, United Kingdom (T.Fawcett@ed.ac.uk).

Table 1. The Seven-Stage Theory of Decision Making[9]

1. Recognition of the problem
2. Formulation of the problem
3. Alternative generation of hypotheses
4. Information search
5. Judgment or choice
6. Action
7. Feedback

Used with permission.

implemented to diagnose renal disease. Several approaches were introduced in the 1990s to support the diagnosis of lymph node disease. Probability systems with decision theory were developed within artificial intelligence, based on approaches for making clinical decisions.[11] Several models apply to the process of data mining. Table 2 presents a summary of these models.

The 20th century was seen as the era of technology. The knowledge explosion continues in medical science and the clinical field. However, according to Yildirim et al,[12] diagnosis of many diseases involves a substantial degree of uncertainty. Pattern recognition and generating provisional hypotheses based on patient symptoms, medical history, physical examination, and some tests are used by physicians for diagnosis. However, "clinical decision making is prone to error," especially in a complicated scenario with a great deal of information. Therefore, data mining has been developed as one way to minimize errors in decision making.[2,3] Data mining is an effective method for extracting valuable knowledge from data. The models that are used in data mining are designed to resemble clinical decision-making strategies.[10] Accordingly, a few data-mining methods are already being used to improve the process of decision making in various fields within the clinical area. Several applications and types of software have been developed to support clinical decision making, for example, a software application for detecting septic shock. The goal of data mining in clinical decision making is to recognize the pattern and relationships in attributes of the clinical setting and to estimate the outcome, to support clinicians when making decisions. Based on the above considerations, the aim of this literature review is to survey the diverse data-mining strategies used in real-time clinical decision making. A search strategy for the literature review is presented in the following sections. The article first demonstrates the strategies of data mining involved in the clinical field and then discusses the advantages and disadvantages of data mining in clinical decision making.

METHODS

A literature review was undertaken to survey the strategies used in data mining in clinical decision making. An integrated review was conducted with four databases: CINAHL, Cochrane Library, PubMed, and MEDLINE. Database searches of titles, abstracts, and key words were performed, using the following search terms: *data mining*, *clinical decision making*, and *decision making*, to expand the sensitivity and specificity of the results and ensure that all potentially applicable articles were included. The findings were limited to English-language publications with full text available; the chosen time period was January 2004 to November 2014. The review included articles connecting data mining with clinical environment, which influences the process of decision making. Articles on coding or informatics data-mining technologies were excluded.

Search Outcome

A total of 323 articles were found in the results from the four databases. Using the exclusion criteria, 254 articles were identified, with duplicates removed. The titles and abstracts

Table 2. Models in Data Mining

Model	Definition of Model
Decision trees	A flowchart-like structure for predicting the value of objects in different categories by using classification algorithms.
Neural network	The model is a two-stage process. First, calculate each attribute into values, then the interconnection pattern between the different layers will be computed. The learning process will be converted.
Support vector machine	Generates a hyperplane with supportive components. In addition, support vector machines tend to produce the closest pattern with supportive data.
Linear regression	The model value is calculated by multiplying the predictor variable. A final model is computed by summing up the contribution, for example, $Y = XB + U$.
Cluster models	A data-mining technique that groups similar data together.
Expert systems	By using a knowledge base, an inference engine that gathers data to focus on the most possible result and an interface, which provides the interaction between users and the system, the model tries to replicate an expert in decision making.
Bayesian classifiers	A probability theory, based on the calculation of probability in an event.
Genetic algorithms	The algorithm will decide the data sequence with a group. After evaluating the similarities between the data and the group, an adjustment will be made in order to create a better conclusion.
Rough sets	A process of spontaneous data transformation; includes modifying the vagueness concepts. Accordingly, the information together creates rules for defining the boundary region of ultimate goals.

were assessed, and 44 articles were recognized as relevant to data mining and decision making in general. After a screening of the full texts, 21 articles were eventually selected as relevant to *clinical* decision making.

Seven of the selected articles discussed data-mining models for the purpose of diagnosis. Of these seven articles, two were related to radiological imaging interpretations, and one related to nursing diagnosis. Four of these articles recommended the identification of an appropriate treatment for the patient. Some articles focused on advancing data-mining strategies, whereas others concentrated on improvements in dealing with adverse drug events. Three of the articles examined the prediction of a certain type of patient by data-mining models (Table 3).

Data-mining theories were considered in all the articles. Among all of the theories, six used decision trees, and three used artificial neural networks. A cluster model was applied in two of the articles. A Bayesian network was identified in two of the articles.

The decision tree is an approach that classifies samples and has a flowchart-like structure. It is for predicting or presenting the value of objects in different categories by using classification algorisms.[19,30,31]

Worachartcheewan et al[13] identified a metabolic syndrome by using a decision tree model. The construction of the decision tree could define the metabolic syndrome by categorized attributes, with accuracy exceeding 99.8%. Through the decision tree strategy, three combinations of variables (triglyceride + blood pressure, fasting plasma glucose + blood pressure, and triglyceride + blood pressure + fasting plasma glucose) were extracted as the most important evidence for classifying metabolic syndrome.

Takada et al[14] used decision trees to predict axillary lymph node metastases in breast cancer patients, in order to determine future treatment. A large dataset of breast cancer patients was provided. Although the model overestimated the proportion of patients at risk of axillary lymph node metastasis, it was accurate with respect to patients with lower risk of metastases.

Finlay et al[19] illustrated the use of an artificial neural network. An artificial neural network appears as a two-stage process.

Table 3. Aim of the Articles

Aim of Articles	No. of Articles/Articles
Supporting diagnosis	7 (Including imaging interpretation: 2 and nursing diagnosis:1)[12–18]
Improving data-mining strategies	6[11,19–23]
Determining treatment	4[24–27]
Reducing adverse drug events	3[4,28,29]
Predicting outcome	1[22]

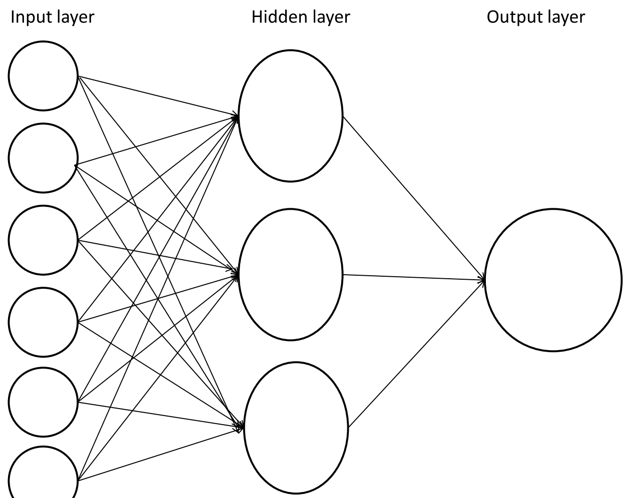

FIGURE 1. An artificial neural network.

In the first stage, each attribute is calculated into values and generates a "neuron." The interconnection pattern between the different layers of neurons will compute the weights of the interconnections. The learning process will convert an output produced by neurons in the second stage. Figure 1 shows an artificial neural network model, which is inspired by the action of the human brain.[10,19]

In a study by Lu et al,[15] the patient was classified as "tear" or "no tear" in the rotator cuff. The orthopedists made a diagnosis of a rotator cuff tear based on a physical examination. This had a high false-positive rate and could result in unnecessary imaging tests, such as magnetic resonance imaging (MRI), which incurs considerable cost. However, this neural network combined clinical examination with multiple personal characteristics (such as age, gender) and symptom history (such as pain index) to make a diagnosis and recommend a treatment plan. Doctors could use the results of this predictive data in making diagnostic decisions, especially if the pretest of a rotator cuff tear was intermediate.[15] Tam et al[24] presented the data-mining strategy of artificial neural networks with respect to osteoarthritic knees. Neural network training was conducted with input attributes, and the estimated outcome was then compared with the real outcome. Consequently, adjustments were made automatically. In the article, three treatments were chosen: transcutaneous electrical nerve stimulation, exercise, and transcutaneous electrical nerve stimulation with exercise. The authors applied the artificial neural network programming techniques with limited attributes to predict the appropriate treatment protocol. Finally, a suggestion would be made by the program (Figure 2).[24]

A Bayesian network is associated with probability distribution. For example, from the probability distribution, the fever and PO_2/FIO_2 ratio for a mechanically ventilated

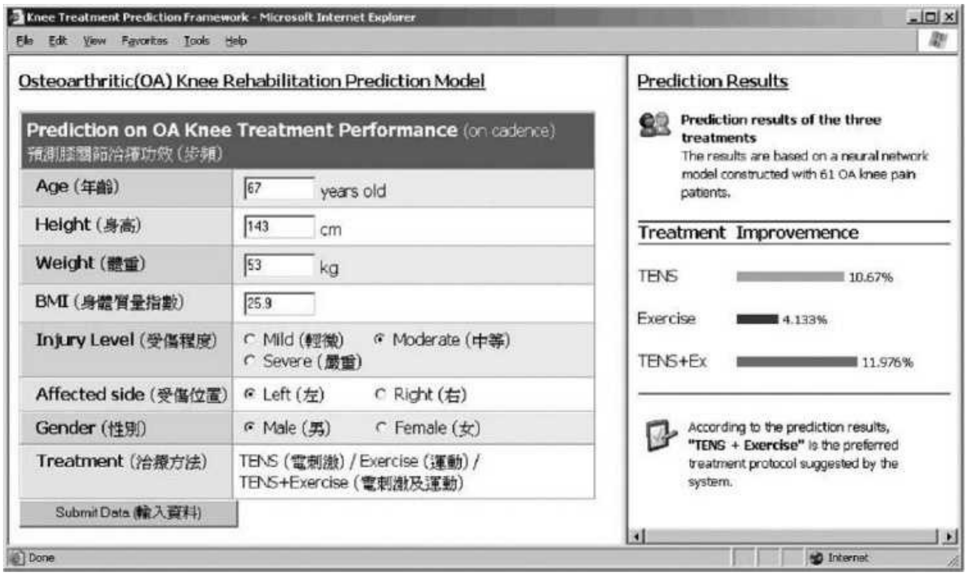

FIGURE 2. Prediction system's Web-based user interface.[19] Used with permission.

patient with pneumonia can be found with a pattern. Through these individual variables, the network will express this probability of an ICU patient with pneumonia.[20] An example of another model—the cluster model—was given by Almasalha et al.[16] Similar attributes were grouped, and historic data were gathered to search for patterns. The pattern generated a nursing diagnosis with nursing interventions suggested (Figure 3).

According to Baceanu et al,[4] the Expert Explorer is a Web-based data visualization tool that can generate reports, update datasets, import rules, and load rules. Experts view the interpretation of the data and give their advice. The data-mining software then learns the rules, which could help to alert clinical staff to adverse drug events. Research by Bowles et al[25] compared the decisions made by a human expert and a data-mining expert model, which judged a patient according to six factors. The data-mining expert model produced 87.6% accuracy. In a review of the literature, Wagholikar et al[11] described an EXPERT model, which is a rule-based model, built with hypotheses, findings, or observations; decision rules were set for the logical relationships between variables and the database. The system tried to give interactive advice for the users. It was used within rheumatology, ophthalmology, and endocrinology.[11] Like heuristics in clinical decision making, heuristic algorithms are the fastest strategies for data mining, but may not be the best method for decision making. Wagholikar et al[11] found the first heuristic system was developed in 1980s by Kulilowski, who interpreted the disease process with a descriptive model and developed consultation systems for neuro-ophthalmology, eye infections, rheumatology, and pathology. A knowledge discovery database (KDD) could be functional in various fields, such as the interpretation of a radiology examination, determination of uncertainty, and clinical care decision making. One of the clinical decision-making scenarios involving KDD illustrated by Reiner[17] was the interpretation of computed tomography and MRI of an emergency patient, which can help clinicians to distinguish strokes.

One of the advantages of using data mining is to increase both computational and diagnostic efficiency.[18,21] According to Orthuber and Sommer,[26] the time for calculating 1 million vectors within double accuracy was between 0.20 and 0.21 seconds. Batal and Hauskrecht[21] developed a model that was able to minimize the predictive rules and attributes and to lessen time for decision making. It is evident that data-mining technology can manage a great amount of data efficiently. Wagholikar et al[11] mention, interestingly, that the progress of data mining for complex problems was better than for simpler problems. In addition, using data-mining strategies in clinical decision making can be accurate, especially when forecasting or diagnosing.[15,19] Lu et al[15] indicate that compared with physical examination, with 40% to 98% accuracy, the data-mining tool can detect a rotator cuff tear with 83% to 95% accuracy. Morrison et al[18] state that the accuracy of the probability of malignancy for pulmonary disease decreased unnecessary computed tomography or pulmonary angiography. Several comparative studies have also highlighted the accuracy of data-mining models such as decision trees.[14,25] Accuracy can eventually improve patient safety and reduce medical errors. However, Takada et al[14] point out that, compared with the diagnostic performance of human experts, data-mining strategies are not accurate enough.

Although data mining can be useful and efficient, it has a few limitations. First, a huge database is required to build

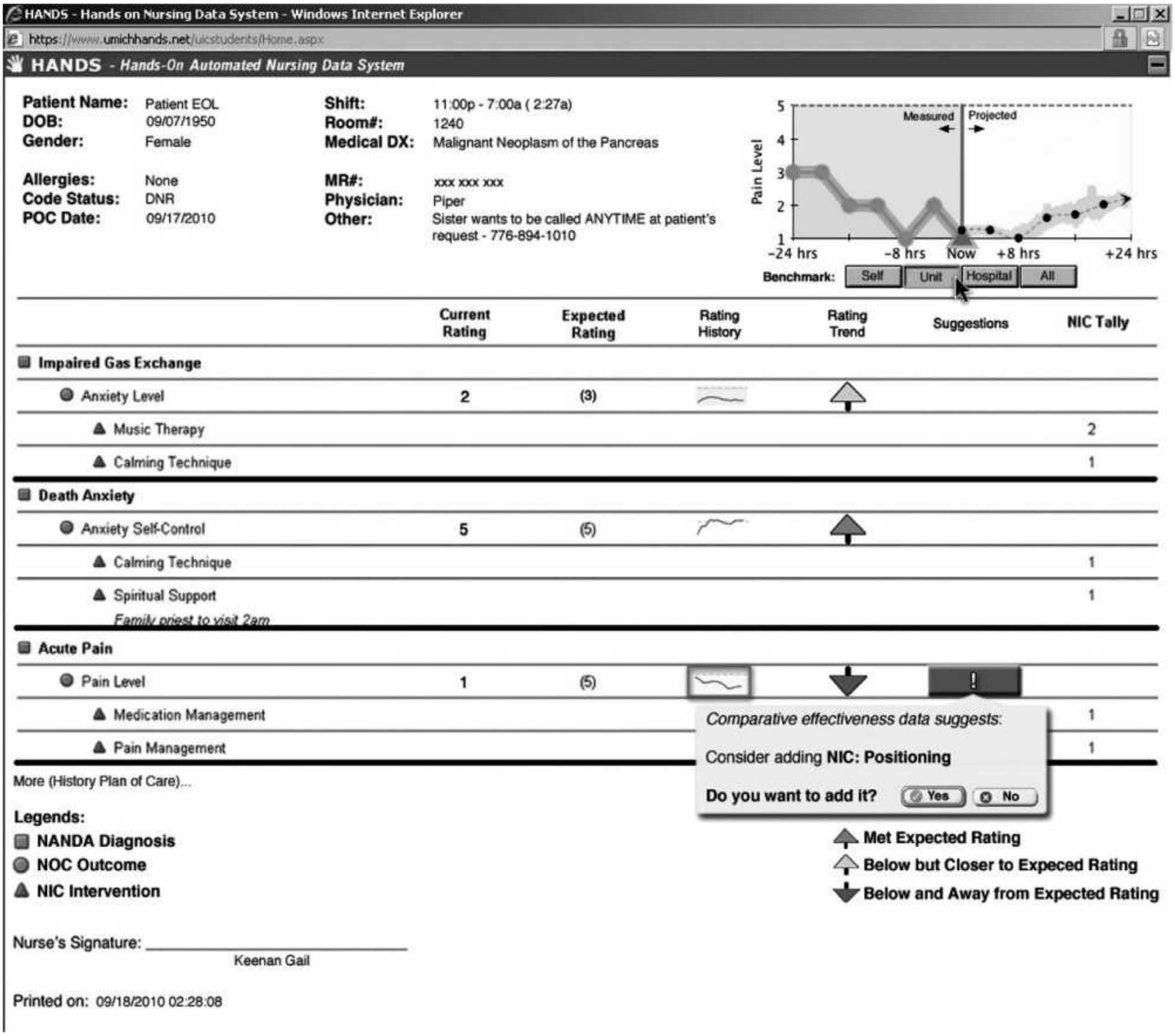

FIGURE 3. Data mining of nursing care plans.[21] Used with permission.

up a data-mining model or to define the patterns.[22,27] For example, a tool for determining treatments for breast cancer patients used the database built by gathering data of 474 breast cancer patients over 5 years.[14] Nevertheless, the use of a data-mining model might be restricted to a specific disease under a certain condition, which means that the tools can only help certain groups of patients with limited conditions, and some of the data-mining strategies might not lead to an interpretation if there were a missing attribute.[22,25–27] Moreover, even though the pattern between the decision and the attribute is found, explanations are seldom provided. A few factors were relatively important for the decision, according to Lu et al.[15] However, it was not clear how the correlations between the diagnosis and these factors were established. This can create an uncertain environment for clinicians to make a judgment with data-mining strategies.[11,20]

DISCUSSION

Data-mining theories and models are similar to the clinical decision-making model; for instance, decision trees occur in both fields, and neural networks with concept attainment theory and heuristics are used in both areas. In decision tree theory, both data mining and clinical decision making use the branches in a decision tree to classify the options of various decisions. Neural networks are similar to concept attainment theory, as they are both linear processes with step-by-step approaches. They generate all attributes to form a hypothesis and then evaluate hypothesis to generate a final score. In data-mining heuristic algorithms, they are the fastest strategies, but may not provide the best decision. Similarly, heuristics in decision-making theory represent an immediate decision that may not be ideal.[32] Nevertheless, one of the most beneficial issues of data mining, compared with clinical decision making, is the feedback system, emphasized in every model. In contrast, clinicians seldom receive any feedback for the judgments they make.

Data-mining theories are a more rational system of decision making. Although data mining is powerful at directing complex situations, it is limited by current technology. Compared with humans, who have, in general, a "limited channel capacity,"[18] an expert nurse may be able to decide an appropriate plan of care by intuition with only a few attributes, whereas data mining is restricted by the database and the conditions of the disease. As nurses, we use our decision-making theories to take care of individual patients in order to provide personal care. On the other hand, data-mining strategies are designed for a group of people. It is arguable whether data mining can come to a decision by itself for individuals.

The results of data mining will affect a clinician's decision making. However, current studies do not mention the correlation in decision making between human beings and information technology. Future studies are required to examine the effect of embedding data-mining models for clinicians.

CONCLUSION

Data mining is an information technology with an innovative effect on the way that people live, communicate, and learn. The technology aims to assist clinicians in clinical decision making and promote patient safety. Several data-mining models have been embedded in the clinical environment to improve decision making and patient safety. This review surveyed the data-mining strategies in clinical decision making and also assessed the disadvantages and advantages of using data mining in clinical decision making, through a literature review of 21 articles. Various aims of data mining were identified, and different data-mining models were introduced in the articles, including a decision tree model for presenting the value of objects in different classifications, a neural network model that gathered attributes and then performed a comparison of the outcomes, a Bayesian network that focused on probability between attributes, an expert model that had a high degree of accuracy, and a KDD that assisted with interpreting imaging. Data mining is efficient with high accuracy. On the other hand, lack of explanation, such as inadequate scientific evidence, is one of the disadvantages because of the problem of working with missing attributes and its limitation with certain conditions or diseases. Predictive data mining is becoming an essential instrument for researchers and clinical practitioners in medicine. Understanding the main issues underlying these methods and the application of agreed and standardized procedures are mandatory for their effective deployment and the proper dissemination of results.

References

1. Paley J, Cheyne H, Dalgleish L, Duncan EA, Niven CA. Nursing's ways of knowing and dual process theories of cognition. *J Adv Nurs*. 2007;60(6): 692–701.
2. Bjørk IT, Hamilton GA. Clinical decision making of nurse working in hospital settings. *Nurs Res Pract*. 2011: 1–8.
3. Aitken L. Critical care nurses' use of decision–making strategies. *J Clin Nurs*. 2003;12(4): 476–483.
4. Baceanu A, Atasiei I, Chazard E, Leroy N. Detection and prevention of adverse drug events: information technologies and human factors. The expert explorer: a tool for hospital data visualization and adverse drug event rules validation. *Stud Health Technol Inform*. 2009; 148: 85–94.
5. Dowding D, Thompson C. Using judgment to improve accuracy in decision-making. *Nurs Times*. 2004;100(22): 42.
6. Elstein AS, Schwarz A. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ*. 2002;324(7339): 729–732.
7. Benner P, Tanner C. How expert nurses use intuition. *Am J Nurs*. 1987;87(1): 23–34.
8. Croskerry P. Context is everything or how could I have been that stupid? *Healthc Q*. 2009;12: e171–e176.
9. Bryans A, McIntosh J. Decision making in community nursing: an analysis of the stages of decision making as they relate to community nursing assessment practice. *J Adv Nurs*. 1996;24(1): 24–30.
10. Finlay S. *Predictive Analytics, Data Mining and Big Data Myths, Misconceptions and Methods*. Basingstoke, Hampshire, UK: Palgrave Macmillan; 2014. http://www.palgraveconnect.com.ezproxy.is.ed.ac.uk/pc/doifinder/10.1057/9781137379283. Accessed July 28, 2015.
11. Wagholikar KB, Sundararajan V, Deshpande AW. Modeling paradigms for medical diagnostic decision support: a survey and future directions. *J Med Syst*. 2012;36(5): 3029–3049.
12. Yildirim P, Majnarić L, Ekmekci O, Holzinger A. Knowledge discovery of drug data on the example of adverse reaction prediction. *BMC Bioinform*. 2014;15(suppl 6): S7–S7.
13. Worachartcheewan A, Nantasenamat C, Isarankura-Na-Ayudhya C, Pidetcha P, Prachayasittikul V. Identification of metabolic syndrome using decision tree analysis. *Diabetes Res Clin Pract*. 2010;90(1): e15–e18.
14. Takada M, Sugimoto M, Naito Y, et al. Prediction of axillary lymph node metastasis in primary breast cancer patients using a decision tree-based model. *BMC Med Inform Decis Making*. 2012;12(1): 54.
15. Lu HY, Huang CY, Su CT, Lin CC. Predicting rotator cuff tears using data mining and Bayesian likelihood ratios. *PLoS One*. 2014;9(4): e94917–e94917.
16. Almasalha F, Xu D, Keenan GM, et al. Data mining nursing care plans of end-of-life patients: a study to improve healthcare decision making. *Int J Nurs Knowl*. 2013;24(1): 15–24.
17. Reiner B. Uncovering and improving upon the inherent deficiencies of radiology reporting through data mining. *J Digital Imagine*. 2010;23(2): 109–118.
18. Morrison JJ, Hostetter J, Wang K, Siegel EL. Data-driven decision support for radiologists: re-using the national lung screening trial dataset for pulmonary nodule management. *J Digital Imaging*. 2015;28(1): 18–23.
19. Finlay DD, Nugent CD, Wang H, Donnelly MP, McCullagh PJ. Mining, knowledge and decision support. *Technol Health Care*. 2010;18(6): 429–441.

20. Lucas P. Bayesian analysis, pattern analysis, and data mining in health care. *Curr Opin Crit Care*. 2004;10(5): 399–403.

21. Batal I, Hauskrecht M. Mining clinical data using minimal predictive rules. *AMIA Annu Symp Proc*. 2010: 31–35.

22. Glover S, Rivers PA, Asoh DA, Piper CN, Murph K. Data mining for health executive decision support: an imperative with a daunting future. *Health Serv Manage Res*. 2010;23(1): 42–46.

23. Rothman B, Leonard JC, Vigoda MM. Future of electronic health records: implications for decision support. *Mount Sinai J Med N Y*. 2012;79(6): 757–768.

24. Tam S, Cheing GLY, Hui-Chan CWY. Predicting osteoarthritic knee rehabilitation outcome by using a prediction model developed by data mining techniques. *Int J Rehab Res*. 2004;27(1): 65–69.

25. Bowles KH, Holmes JH, Ratcliffe SJ, Liberatore M, Nydick R, Naylor MD. Factors identified by experts to support decision making for post acute referral. *Nurs Res*. 2009;58(2): 115–122.

26. Orthuber W, Sommer T. A searchable patient record database for decision support. *Stud Health Technol Inform*. 2009;150: 584–588.

27. Çakır A, Demirel B. A software tool for determination of breast cancer treatment methods using data mining approach. *J Med Syst*. 2011;35(6): 1503–1511.

28. Iyer SV, Harpaz R, Lependu P, Bauer-Mehren A, Shah NH. Mining clinical text for signals of adverse drug-drug interactions. *J Am Med Inform Assoc*. 2014;21(2): 353–362.

29. Woosley RL, Romero K. Assessing cardiovascular drug safety for clinical decision-making. *Nat Rev Cardiol.* 2013;6: 330.

30. Holmes DE, Jain LC. *Data Mining. Volume 2, Statistical, Bayesian, Time Series and Other Theoretical Aspects Foundations and Intelligent Paradigms*. Berlin, Germany : Springer-Verlag Berlin Heidelberg; 2012.

31. Paprotny A, Thess M. *Realtime Data Mining*. Cham: Birkhäuser; 2013.

32. Thompson C. Clinical experience as evidence in evidence-based practice. *J Adv Nurs*. 2003;43(3): 230–237.

For 14 additional continuing education articles related to data analysis in nursing, go to NursingCenter.com/CE/CIN

Instructions for Taking the CE Test Online
Using Data Mining Strategies in Clinical Decision Making: A Literature Review

- Read the article. The test for this CE activity can be taken online at www.nursingcenter.com/ce/CIN. Tests can no longer be mailed or faxed.
- You will need to create a free login to your personal CE Planner account before taking online tests. Your planner will keep track of all your Lippincott Williams & Wilkins online CE activities for you.
- There is only one correct answer for each question. A passing score for this test is 13 correct answers. If you pass, you can print your certificate of earned contact hours and the answer key. If you fail, you have the option of taking the test again at no additional cost.
- For questions, contact Lippincott Williams & Wilkins: 1-800-787-8985.

Registration Deadline: October 31, 2018

Disclosure Statement:

The authors and planners have disclosed that they have no financial relationships related to this article.

Provider Accreditation:

Lippincott Williams & Wilkins, publisher of *CIN, Computers Informatics Nursing*, will award 2.0 contact hours for this continuing nursing education activity.

Lippincott Williams & Wilkins is accredited as a provider of continuing nursing education 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 2.0 contact hours. Lippincott Williams & Wilkins is also an approved provider of continuing nursing education by the District of Columbia and in all states.

Payment:

- The registration fee for this test is $21.95.