CJNR 2013, Vol. 45 N° 1, 98–112

Résumé

La découverte de connaissances fondée sur la pratique à des fins de recherche sur l'efficacité comparative : un cadre organisationnel

Robert J. Lucero, Suzanne Bakken

Les systèmes d'information électroniques sur la santé peuvent accroître la capacité des organismes de soins de santé à étudier les effets des interventions cliniques. Le présent article propose un cadre organisationnel qui intègre les paradigmes de la recherche en informatique et de la recherche sur les résultats afin de faciliter la création de connaissances à l'aide de bases de données cliniques. Afin d'illustrer le cadre proposé, les auteurs l'appliquent à l'exemple des plaies de pression. Ce cadre de création de connaissances à l'aide de l'informatique aux fins de la recherche comparative sur l'efficacité des traitements (CCI-RCET) a été conçu dans le but de servir d'outil heuristique pour la conceptualisation des modèles d'étude et de surmonter les contraintes méthodologiques que peut éventuellement imposer toute perspective de recherche unique. Les percées de la recherche en informatique appliquée peuvent jouer un rôle complémentaire dans le développement du champ de la recherche sur les résultats, y compris de la recherche comparative sur l'efficacité des traitements. Le cadre de CCI-RCET peut être utilisé afin de favoriser la production de connaissances à partir des données cliniques électroniques qui sont recueillies de façon systématique.

Mots clés : bases de données cliniques, recherche comparative sur l'efficacité des traitements, modèles d'étude, production de connaissances, informatique, recherche

CJNR 2013, Vol. 45 N° 1, 98–112

Practice-Based Knowledge Discovery for Comparative Effectiveness Research: An Organizing Framework

Robert J. Lucero, Suzanne Bakken

Electronic health information systems can increase the ability of health-care organizations to investigate the effects of clinical interventions. The authors present an organizing framework that integrates outcomes and informatics research paradigms to guide knowledge discovery in electronic clinical databases. They illustrate its application using the example of hospital acquired pressure ulcers (HAPU). The Knowledge Discovery through Informatics for Comparative Effectiveness Research (KDI-CER) framework was conceived as a heuristic to conceptualize study designs and address potential methodological limitations imposed by using a single research perspective. Advances in informatics research can play a complementary role in advancing the field of outcomes research including CER. The KDI-CER framework can be used to facilitate knowledge discovery from routinely collected electronic clinical data.

Keywords: comparative effectiveness, knowledge generation, informatics, health services, research, data mining

Introduction

As health-care organizations increase their ability to collect and store electronic clinical data, outcomes researchers continue to develop and test novel research designs and analytic methods to better understand quality of care. Researchers have continually demonstrated a link between health-care organizational factors and patient outcomes, which theoretically supports the investigation of the effectiveness of clinical care (Aiken, Clarke, Cheung, Sloane, & Silber, 2003; Estabrooks, Midodzi, Cummings, Ricker, & Giovannetti, 2005; Friese, Lake, Aiken, Silber, & Sochalski, 2008; Tourangeau et al., 2007). Electronic databases compiled by health-care organizations will be important sources of information for evaluating the effects of practice-based interventions.

Informatics approaches and resources are becoming critical to the field of outcomes research. Informatics approaches, such as data mining, are important for enabling knowledge-driven health care based on a solid research foundation (Embi, Kaufman, & Payne, 2009). The use of electronic clinical databases can provide opportunities to identify comprehensive new evidence from clinical practice and to accelerate knowledge generation. Discovering effective practice patterns through the use of electronic databases will provide empirical evidence of what clinical interventions constitute safe, efficient, high-quality care for patients at risk for problematic conditions such as pressure ulcers. As digital information and communication technologies overtake manual data collection and storage, we need to leverage the strengths of outcomes and informatics research.

This article presents an organizing framework that integrates the Quality Health Outcomes Model (QHOM) (Mitchell, Ferketich, & Jennings, 1998) with the Knowledge Discovery in Databases (KDD) process (Fayyad, Piatetsky-Shapiro, & Smyth, 1996) to guide knowledge discovery from electronic clinical databases to support the conduct of comparative effectiveness research (CER). Our framework is based on the notion that expert practice-based knowledge is critical to (a) identifying clinically relevant interventions based on patient and system characteristics; and (b) facilitating rigorous and efficient knowledge discovery, using electronic clinical databases, that is relevant to the field of outcomes research. We describe the framework and illustrate its application using the example of hospital acquired pressure ulcer (HAPU) prevention.

Outcomes: The End Results of Clinical Care

Outcomes research examines the end results of health-care delivery that takes into account patients' experiences, preferences, and values (Clancy & Eisenberg, 1998). A major challenge in outcomes research has been to balance what researchers *want to* measure and what they *can* measure using existing information systems. Outcomes research involves a range of statistical methods and data collection, including collecting data de novo and drawing on primary studies. Two broad categories of data are relevant to the conduct of outcomes research: (1) patient-specific information (laboratory results, health-care-provider assessments, and other information that can be found in a medical record); and (2) system-specific information (health-care expenditures, staffing, hospital size, etc.). Both patient- and system-level data are needed to contextualize and evaluate relevant measures such as complications, length of hospital stay, health status, and mortality. The goal of outcomes research is to generate evidence pertaining to decisions made by those who participate in health care, including administrators, providers, and patients. Associating differences in the process of care with differences in outcomes can serve to clarify what care is effective or worthwhile as well as where improvements can be made by clinicians and organizations.

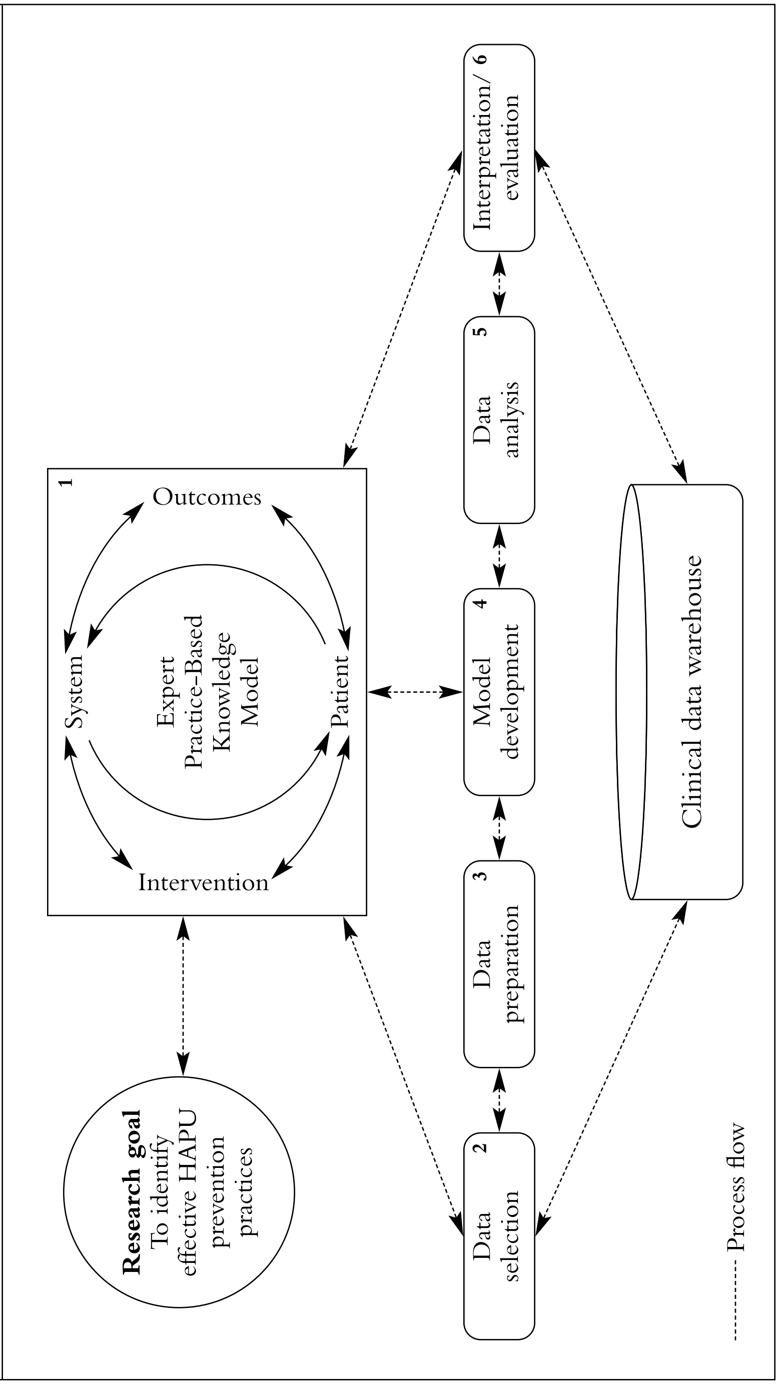
Informatics: Facilitating the Use of Electronic Clinical Databases in Outcomes Research

Recent developments and advances in informatics research can enable rigorous outcomes research. A central concern for the outcomes researcher is the aggregation of data from multiple disparate information resources. Informatics platforms and resources used in clinical practice can potentially address data concerns. Integrated electronic health records (EHRs) can be used to collect data on potential research participants and reduce reliance on redundant and error-prone paper-based collection methods (Bates, Ebell, Gotlieb, Zapp, & Mullins, 2003). Moreover, EHR data collected through routine health-care processes can be reused in outcomes research. These data can be stored and maintained in a clinical data warehouse, a type of database or data repository that is designed for data reuse in research (Dewitt & Hampton, 2005). This type of data storage facilitates longitudinal or episodic queries based on more than one criterion of interest, including laboratory, radiology, and pathology results; surgical reports; discharge summaries; demographic information; diagnostic and procedural codes; and operations data, which are especially useful in outcomes research. Initiatives aimed at supporting reuse of data stored in clinical and research warehouses have resulted in the development of query tools to facilitate researcher exploration and extraction of data (Sittig et al., 2012). Moreover, the application of data mining and statistical methods to identify or test hypotheses is common in large-scale data extraction from data repositories. These automated methods can complement rigorous outcomes research by enabling the use of electronic clinical data as well as enhancing the efficiency of large-scale practice-based studies.

Integrating Outcomes and Informatics Perspectives

The Knowledge Discovery through Informatics for Comparative Effectiveness Research (KDI-CER) organizing framework is one approach to the integration of outcomes (i.e., QHOM) and informatics (i.e., KDD) paradigms for the purpose of guiding knowledge discovery for CER. Our approach is sufficiently broad to (1) guide the development of research on what works in clinical practice, (2) provide a framework for outcomes research and knowledge discovery that involves clinicians and researchers across multiple disciplines, and (3) support the diversity of clinical outcomes research. The KDI-CER framework suggests that discovering knowledge about the relative effects of practice-based interventions is an interactive and iterative process informed by expert practice-based knowledge at every stage (see Figure 1).

Figure 1 *Organizing Framework*



QHOM

The Quality Health Outcomes Model depicts the relationships of four constructs: patients, systems, interventions, and outcomes (see Figure 1, box 1). Mitchell et al. (1998) suggest reciprocal directions of influence between constructs with no single direct connection between interventions and outcomes. The effect of interventions on outcomes is thought to be mediated by patient and/or system characteristics. Interventions are considered direct or indirect patient care, such as the administration of total parenteral nutrition or the provision of culturally appropriate education to increase adherence to a therapeutic regimen. Patient characteristics can include traditional demographic variables, existing health problems, and socioeconomic status measures. The system includes organizational and provider characteristics — for example, hospital bed size or nurses' education level. Thus outcomes are "the results of care structures and processes that integrate functional, social, psychological, physical, and physiological aspects of people's experience in health and illness," which may be individual or organizational measures (Mitchell et al., 1998, p. 45).

KDD

Knowledge Discovery in Databases refers to a human (e.g., researchers and/or clinical experts) supported interactive and iterative process of discovering useful, nontrivial contextualized knowledge from large electronic databases (Fayyad et al., 1996). KDD has typically focused on financial and other business-related databases but is being used increasingly with health-care databases. It involves a number of steps with many decisions made by researchers and/or clinical experts throughout the process. Generally, KDD consists of (1) developing a knowledge model of the clinical domain; (2) creating a target data set; (3) data cleaning and pre-processing; (4) data reduction and projection; (5) choosing the data-mining task (e.g., description or prediction); (6) choosing data-mining methods; (7) identifying and interpreting patterns (or reiterating any of the previous procedures based on the preliminary findings); and (9) documenting and/or incorporating discovered knowledge into practice or policy (Fayyad et al., 1996). KDD can involve numerous iterations of any of the procedures. In health-care research its goal is to generate and/or specify a model that can be tested to confirm the association between explanatory and response variables.

The KDI-CER framework (see Figure 1) harnesses the conceptual and methodological maturity of both the QHOM and the KDD to facilitate knowledge discovery in large clinical databases. Practice-based expert knowledge is one of its critical components. Researchers have demonstrated that theoretical and pragmatic expert practice-based suggestions in outcome studies enhance knowledge discovery through electronic clinical databases (Gaines, 1989; Wilcox & Hripcsak, 2003). The QHOM is used to motivate clinicians to identify key variables in intervention research and provide direction to clinicians and researchers in developing hypotheses to test during the knowledge discovery process. While there are no theoretical links between the QHOM and KDD processes, the KDI-CER approach depends on data/information about patient and system characteristics and interventions to generate practice-based knowledge from existing data. In contrast to the traditional approach to knowledge discovery (Fayyad et al., 1996), the KDI-CER groups several procedures within one (i.e., *data preparation* includes data cleaning, pre-processing, reduction, and projection; *model development* includes choosing the data-mining task and methods (see Figure 1, boxes 3 and 4). The KDI-CER framework is intended to guide practice-based observational studies of electronic clinical data.

Example: Hospital Acquired Pressure Ulcers

The following describes the process of conducting knowledge discovery based on the KDI-CER organizing framework. The major assumptions of this approach are that (a) knowledge discovery using electronic clinical databases requires the inclusion of clinical expertise about practice-based interventions, and (b) the relative effects of interventions cannot be evaluated without practice-based data.

The Problem

Hospital acquired pressure ulcers are a major cause of morbidity that significantly increases average length of stay, human suffering, and financial costs (Reynolds, 2008). The 2011 HealthGrades Patient Safety in American Hospitals Study reports that HAPU is the second most common patient safety event, after death from treatable surgical complications, among Medicare patients (Reed & May, 2011). Russo et al. (2008) report a 78.9% increase in HAPU occurrence from 1993 (N = 281,300) to 2006 (N = 503,300) for adults 18 years and older. The associated cost in 2006 was $11 billion. The estimated incidence rate for HAPU ranges from 7% to 9% (Whittington & Briones, 2004) and the documented incidence rate for ICU HAPU ranges from 3.8% to 12.4% (Vangilder, Amlung, Harrison, & Meyer, 2009).

HAPU prevention is based largely on expert opinion and/or consensus panels, and not on empirical evidence (European Pressure Ulcer Advisory Panel & National Pressure Ulcer Advisory Panel, 2009). The most conclusive findings on HAPU are related to the identification of risk factors or validation of pressure ulcer risk-assessment tools (Lyder & Ayello, 2008). A challenge in patient safety research is discovering effective HAPU-prevention interventions. While comprehensive prevention programs based on clinical practice guidelines can reduce overall prevalence rates (Lyder, Grady, Mathur, Petrillo, & Meehan, 2004; Rich, Shardell, Margolis, & Baumgarten, 2009), there is insufficient evidence linking individual risk assessment and clinical care to HAPU prevention.

Discovering HAPU Prevention Interventions in Electronic Databases

In this example, a mixed-method approach is used to discover empirical relationships between practice-based interventions and the prevention of HAPU in electronic clinical databases. A multidisciplinary clinical team comprising practice-based experts (i.e., nurses, physicians, nutritionists, and physical therapists) collaborates throughout all phases of the knowledge discovery process. The research team uses the QHOM constructs as guideposts to engage the clinical experts in an iterative interchange to develop a practice-based-knowledge model of HAPU prevention based on the patient (e.g., comatose) and system characteristics (e.g., nurse staffing) and interventions (e.g., skin care) that are considered important in the prevention of HAPU (see Figure 1, box 1). The robust characterization of patient and system characteristics and interventions can enhance the external validity of empirical findings (Horn & Gassaway, 2007). The research team extracts hospitalizations from the clinical data warehouse based on selection criteria agreed upon by the clinical and research teams. The data from these hospitalizations are used to create a target data set of the patient and system characteristics and interventions (see Figure 1, box 2).

Data-Preparation Stage

The research team evaluates the quality of the data during the data-preparation stage (see Figure 1, box 3). First, the research team creates decision rules for automated methods of converting textual information from unstructured sources such as narrative clinical notes or reports into structured, coded descriptions (Friedman, Alderson, Austin, Cimino, & Johnson, 1994). Next, it creates decision rules to extract structured coded data from its original representation (i.e., multiple database tables) to create a single database table. Each patient record for a hospitalization is represented by a row and the features (i.e., variables) are represented in columns. Organizing the data in this structure reduces hierarchical and nested structuring (Adriaans & Zantinge, 1996). Third, the research team uses various methods to transform the data into a useable form, including data imputation, time-lag specification, and data reduction. Data transformation minimizes the potential for over-fitting the analytic models

(Kohavi, 1995). Many of the cases in the target data set will have missing features. The research team addresses missing data by choosing from among the following data-imputation methods: (1) procedures based on quasi-randomization modes of inference, (2) model-based approaches, and (3) machine learning methods (Lakshminarayan, Harp, & Samad, 1999). Additionally, clinical practices may include time-dependent interventions. In such cases, the research team specifies a time lag to represent the number of previous values that could influence the outcome (Kkantardzic, 2003). Depending on the dimensionality of the data set, statistical (i.e., one-dimension) or visualization (i.e., multiple-dimension) methods are used to conduct outlier analyses (Kkantardzic, 2003). The research team considers whether some data can be discarded to increase computational efficiency without reducing the quality of the data set (Koller & Sahami, 1996). Variable reduction can be based on principal component analysis, values reduction, variable discretization, and case reduction (Kkantardzic, 2003). Finally, the research team considers whether the target data set should be limited further using the information captured in the Expert Practice-Based Knowledge Model (see Figure 1, box 1). However, the team is careful not to over-limit the data set, thereby allowing potentially unknown empirical relationships to be discovered using automated data-mining methods.

After preparing the data for analysis, the clinical practice and research teams combine automated data-mining methods and Bayesian network analyses to estimate the relative effects of prevention interventions. There are a number of factors that make Bayesian networks ideal for knowledge discovery, including use of expert domain knowledge in the discovery process; increased precision and accuracy through the use of encoded knowledge to predict an outcome; adaptability to rapidly evolving interventions (e.g., devices, procedures, and provider or system interventions); and ease of interpretation as compared to other network structure data-mining techniques (Lee & Abbott, 2003). Adaptive approaches, such as Bayesian networks, can be useful for answering pragmatic questions within and across care units and patient populations — for example, "What is the probability that turning a patient every 2 hours is more effective than turning a patient every 4 hours?" or "What is the probability that Practice Model A is more effective than Practice Model B, and vice-versa, for pressure ulcer prevention in Population A versus Population B?"

The research team develops a Bayesian network structure that depicts probabilistic relationships among the patient and system characteristics and interventions identified by the clinical practice team and based on the associations posited in the QHOM. The structure reflects the conceptual relationships of probabilistic knowledge that take the form of a diagram with nodes (i.e., variables) and arrows (i.e., relationship directions). Based on the specifications of the network structure, the research team applies data-mining methods to a subset of the target data set (i.e., training data) to generate a relational analytic model. The resultant model (see Figure 1, box 4) reveals estimates of the relative effects of interventions and patient and system characteristics on HAPU prevention. The research team cross-validates this analytic model using a random subset of the target data (i.e., test data; box 5). The significance of the analytic model depends on its ability to detect at-risk patients who did not develop a HAPU (i.e., evaluation) and whether the clinical team thinks (i.e., interpretation; box 6) the results are clinically significant. In the long run, the research team will conduct a comparative effectiveness study to examine the differences between HAPU prevention based on the findings from the knowledge discovery process and the usual prevention model.

Implications of the KDI-CER Organizing Framework

The KDI-CER framework is intended to guide investigations into the effects of clinical interventions using electronic clinical databases. There has been limited use of informatics methods for knowledge discovery in health research (Chae, Ho, Cho, Lee, & Ji, 2001; Goodwin et al., 2001; Jakkula & Cook, 2008; Poynton & McDaniel, 2006; Sokol, Garcia, Rodriguez, West, & Johnson, 2001; Zhu, Zhang, Hirdes, & Stolee, 2007), and even less in nursing outcomes research. Knowledge discovery and the conduct of rigorous CER focused on nursing interventions are dependent on high-quality electronic nursing data. The integration of standardized nursing languages (SNLs) in EHRs can support robust knowledge discovery and CER by making available nurses' documentation of interventions and outcomes (Bakken, Stone, & Larson, 2008; Institute of Medicine, 2011). However, nurses must engage in reliable documentation to capture nursing's contribution to patient outcomes. When nurses are not fully engaged in reliable documentation, the analyses of the resulting data for CER are impeded by extensive data preparation, potentially biased information, and increased research costs. The use of SNLs in EHRs can support data collection across populations and settings as well as yield the evidence necessary to support practice-based knowledge discovery and CER (Lunney, Delaney, Duffy, Moorhead, & Welton, 2005; Westra, Delaney, Konicek, & Keenan, 2008).

The ongoing development of health information infrastructures in several countries, including Australia (Department of Health and Ageing, Australian Government, 2010), Canada (Canada Health Infoway, 2010), the United Kingdom (National Health Service, 2010), and the United

States (Yasnoff et al., 2004), as well as the development of EHR databases by large health-care delivery systems, will increase the capacity of informatics and outcomes research (Häyrinen, Saranto, & Nykänen, 2008). These developments will make available integrated health-care information on millions of patient episodes and increase the potential for discovering the effects of clinical interventions on patient outcomes (Liang, 2007). Our framework builds on the strengths of the QHOM and KDD processes by reflecting the various components of care and incorporating practice-based expertise throughout the knowledge discovery process. The KDI-CER framework provides a way not only to conceptualize study designs but also to address methodological limitations imposed by the use of a single research perspective. Organizing frameworks are frequently developed by experts and should be put to the test in terms of practice. Although we have illustrated application of the framework to the example of HAPU, the usefulness of the KDI-CER framework will be fully appreciated when it is applied to real-world electronic clinical data to discover knowledge about the effects of clinical interventions for use in CER.

Conclusion

The KDI-CER framework was conceived as a heuristic for knowledge discovery to support CER. It encourages clinicians and researchers to conceptualize clinical practice from a complex perspective that suggests there is an indirect influence of interventions on outcomes and system and patient characteristics that mediate the effects of interventions on the outcomes of care delivery. This can stimulate the identification of relevant practice-based interventions as well as patient and system characteristics to facilitate knowledge discovery in electronic clinical databases.

References

Adriaans, P., & Zantinge, D. (1996). *Data mining.* Harlow: Addison Wesley Longman.

Aiken, L. H., Clarke, S. P., Cheung, R. B., Sloane, D. M., & Silber, J. H. (2003). Educational levels of hospital nurses and surgical patient mortality. *Journal of the American Medical Association, 290*(12), 1617–1623.

Bakken, S., Stone, P. W., & Larson, E. L. (2008). A nursing informatics research agenda for 2008–18: Contextual influences and key components. *Nursing Outlook, 56*(5), 206–214.

Bates, D. W., Ebell, M., Gotlieb, E., Zapp, J., & Mullins, H. C. (2003). A proposal for electronic medical records in US primary care. *Journal of the American Medical Informatics Association, 10*(1), 1–10.

Canada Health Infoway. (2010). *EHRS blueprint: An interoperable EHR framework.* Retrieved April 22, 2011, from http://www.infoway-inforoute.ca/lang-en/.

Chae, Y. M., Ho, S. H., Cho, K. W., Lee, D. H., & Ji, S. H. (2001). Data mining approach to policy analysis in a health insurance domain. *International Journal of Medical Informatics, 62*(2/3), 103–111.

Clancy, C. M., & Eisenberg, J. M. (1998). Outcomes research: Measuring the end results of health care. *Science, 282*(5387), 245–246.

Department of Health and Ageing, Australian Government. (2010). *HealthConnect.* Retrieved April 22, 2011, from http://www.health.gov.au/internet/main/publishing.nsf/Content/EHeath+Healthconnect.

Dewitt, J. G., & Hampton, P. M. (2005). Development of a data warehouse at an academic health system: Knowing a place for the first time. *Academic Medicine, 80*(11), 1019–1025.

Embi, P. J., Kaufman, S. E., & Payne, P. R. (2009). Biomedical informatics and outcomes research: Enabling knowledge-driven health care. *Circulation, 120*(23), 2393–2399.

Estabrooks, C. A., Midodzi, W. K., Cummings, G. G., Ricker, K. L., & Giovannetti, P. (2005). The impact of hospital nursing characteristics on 30-day mortality. *Nursing Research, 54*(2), 74–84.

European Pressure Ulcer Advisory Panel & National Pressure Ulcer Advisory Panel. (2009). *Prevention and treatment of pressure ulcers: Quick reference guide.* Retrieved June 1, 2010, from from http://www.npuap.org/Final_Quick_Prevention_for_web_2010.pdf.

Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery: An overview. In U. S. Fayyad, G. Piatetsky-Shapiro, P. Smyth, & R. Uthurusamy (Eds.), *Advances in knowledge discovery and data mining* (pp. 1–34). London: AAAI Press/MIT Press.

Friedman, C., Alderson, P. O., Austin, J. H., Cimino, J. J., & Johnson, S. B. (1994). A general natural-language text processor for clinical radiology. *Journal of the American Medical Informatics Association, 1*(2), 161–174.

Friese, C. R., Lake, E. T., Aiken, L. H., Silber, J. H., & Sochalski, J. (2008). Hospital nurse practice environments and outcomes for surgical oncology patients. *Health Services Research, 43*(4), 1145–1163.

Gaines, B. R. (1989, June 26–27). An ounce of knowledge is worth a ton of data: Quantitative studies of the trade-off between expertise and data based on statistically well-founded empirical induction. In A. M. Segre (Ed.), *Proceedings of the Sixth International Workshop on Machine Learning, Cornell University, Ithaca, New York* (pp. 156–159). San Mateo, CA: Morgan Kaufmann.

Goodwin, L. K., Iannacchione, M. A., Hammond, W. E., Crockett, P., Maher, S., & Schlitz, K. (2001). Data mining methods find demographic predictors of preterm birth. *Nursing Research, 50*(6), 340–345.

Häyrinen, K., Saranto, K., & Nykänen, P. (2008). Definition, structure, content, use and impacts of electronic health records: A review of the research literature. *International Journal of Medical Informatics,* 77(5), 291–304.

Horn, S. D., & Gassaway, J. (2007). Practice-based evidence study design for comparative effectiveness research. *Medical Care, 45*(10), S50–S57.

Institute of Medicine. (2011). *The future of nursing: Leading change, advancing health.* Washington: National Academies Press.

Jakkula, V., & Cook, D. J. (2008). Anomaly detection using temporal data mining in a smart home environment. *Methods of Information in Medicine, 47*(1), 70–75.

Kkantardzic, M. (2003). *Data mining: Concepts, models, methods, and algorithms.* Hoboken, NJ: Wiley-Interscience/IEEE Press.

Kohavi, R. (1995, April 25–27). The power of decision tables. In N. Lavrac & S. Wrobel (Eds.), *Proceedings of the 8th European Conference on Machine Learning, Heraclion, Greece* (pp. 174–189). Berlin/Heidelberg: Springer-Verlag.

Koller, D., & Sahami, M. (1996, July 3–6). Toward optimal feature selection. In L. Saitta (Ed.), *Proceedings of the Thirteenth International Conference on Machine Learning, Bari, Italy* (pp. 284–292). San Mateo, CA: Morgan Kaufmann.

Lakshminarayan, K., Harp, S. A., & Samad, T. (1999). Imputation of missing data in industrial databases. *Applied Intelligence, 11,* 259–275.

Lee, S. M., & Abbott, P. A. (2003). Bayesian networks for knowledge discovery in large datasets: Basics for nurse researchers. *Journal of Biomedical Informatics, 36*(4/5), 389–399.

Liang, L. (2007). The gap between evidence and practice. *Health Affairs, 26*(2), W119–W121.

Lunney, M., Delaney, C., Duffy, M., Moorhead, S., & Welton, J. (2005). Advocating for standardized nursing languages in electronic health records. *Journal of Nursing Administration, 35*(1), 1–3.

Lyder, C. H., & Ayello, E. A. (2008). Pressure ulcers: A patient safety issue. In R. G. Hughes (Ed.), *Patient safety and quality: An evidence-based handbook for nurses.* Rockville, MD: Agency for Healthcare Research and Quality. Retrieved June 1, 2010, from http://www.ahrq.gov/qual/nurseshdbk/.

Lyder, C. H., Grady, J., Mathur, D., Petrillo, M. K., & Meehan, T. P. (2004). Preventing pressure ulcers in Connecticut hospitals by using the plan-do-study-act model of quality improvement. *Joint Commission Journal on Quality and Patient Safety, 30*(4), 205–214.

Mitchell, P. H., Ferketich, S., & Jennings, B. M. (1998). Quality health outcomes model. American Academy of Nursing Expert Panel on Quality Health Care. *Image: Journal of Nursing Scholarship, 30*(1), 43–46.

National Health Service. (2010). *Connecting for health.* Retrieved April 22, 2011, from http://www.connectingforhealth.nhs.uk/.

Poynton, M. R., & McDaniel, A. M. (2006). Classification of smoking cessation status with a backpropagation neural network. *Journal of Biomedical Informatics, 39*(6), 680–686.

Reed, K., & May, R. (2011). HealthGrades eighth annual Patient Safety in American Hospitals Study. *AHRQ PSNet,* March. Denver: HealthGrades.

Reynolds, T. M. (2008). Risk assessment for prevention of morbidity and mortality: Lessons for pressure ulcer prevention. *Journal of Tissue Viability, 17*(4), 115–120.

Rich, S. E., Shardell, M., Margolis, D., & Baumgarten, M. (2009). Pressure ulcer preventive device use among elderly patients early in the hospital stay. *Nursing Research, 58*(2), 95–104.

Sittig, D. F., Hazlehurst, B. L., Brown, J., Murphy, S., Rosenman, M., Tarczy-Hornoch, P., et al. (2012). A survey of informatics platforms that enable distributed comparative effectiveness research using multi-institutional heterogenous clinical data. *Medical Care, 50*(Suppl 1), S49–S59.

Sokol, L., Garcia, B., Rodriguez, J., West, M., & Johnson, K. (2001). Using data mining to find fraud in HCFA health care claims. *Topics in Health Information Management, 22*(1), 1–13.

Tourangeau, A. E., Doran, D. M., McGillis Hall, L., O'Brien Pallas, L., Pringle, D., Tu, J.V., et al. (2007). Impact of hospital nursing care on 30-day mortality for acute medical patients. *Journal of Advanced Nursing, 57*(1), 32–44.

Vangilder, C., Amlung, S., Harrison, P., & Meyer, S. (2009). Results of the 2008–2009 International Pressure Ulcer Prevalence Survey and a 3-year, acute care, unit-specific analysis. *Ostomy Wound Management, 55*(11), 39–45.

Westra, B. L., Delaney, C. W., Konicek, D., & Keenan, G. (2008). Nursing standards to support the electronic health record. *Nursing Outlook, 56*(5), 258–266, e251.

Whittington, K. T., & Briones, R. (2004). National Prevalence and Incidence Study: 6-year sequential acute care data. *Advances in Skin and Wound Care, 17*(9), 490–494.

Wilcox, A. B., & Hripcsak, G. (2003). The role of domain knowledge in automating medical text report classification. *Journal of the American Medical Informatics Association, 10*(4), 330–338.

Yasnoff, W. A., Humphreys, B. L., Overhage, J. M., Detmer, D. E., Brennan, P. F., Morris, R. W., et al. (2004). A consensus action agenda for achieving the national health information infrastructure. *Journal of the American Medical Informatics Association, 11*(4), 332–338.

Zhu, M., Zhang, Z., Hirdes, J. P., & Stolee, P. (2007). Using machine learning algorithms to guide rehabilitation planning for home care clients. *BMC Medical Informatics and Decision Making, 7,* 41.

Acknowledgements

Preparation of the manuscript was supported by the Reducing Health Disparities Through Informatics Training Program (NIH T32-NR007969, Suzanne Bakken, PI), Columbia Center for the Health of Urban Minorities, Northern Manhattan Center for Comparative Effectiveness Research to Eliminate Health Disparities (NIH NCMHD, 3P60MD000206-08S1, José Luchsinger, PI), and the Washington Heights/Inwood Informatics Infrastructure for Comparative Effectiveness Research (R01 HS019853, Suzanne Bakken).

Conflict of interest statement: The authors declare no conflicts of interest.

Robert J. Lucero, RN, PhD, MPH, is Assistant Professor of Nursing, Center for Evidence-Based Practice in the Underserved, School of Nursing, Columbia University, New York, New York, United States. Suzanne Bakken, RN, DNSc, FAAN, is Alumni Professor of Nursing, Professor of Biomedical Informatics, and Director, Center for Evidence-Based Practice in the Underserved, School of Nursing, Columbia University.