Nursing Informatics
U. Gerdin et al. (Eds.)
IOS Press 1997

A Collaborative International Nursing Informatics Research Project: Predicting ARDS Risk in Critically Ill Patients

L. Goodwin[a], J. Saville[a], B. Jasion[a], B. Turner[a], J. Prather[a], T. Dobousek[b] and S. Egger[b]
[a] *Box 3322 Duke University Medical Center, Durham, NC, USA 27710,*
[b]*University of Vienna - Allgemeine Krankenhaus General Hospital, 18-20 Wohringer-Gurtel, A-1090, Vienna Austria*

An international nursing informatics research collaboration between Duke University Medical Center in Durham, North Carolina, USA and Allgemeine Krankenhaus Hospital in Vienna, Austria used data mining techniques called Knowledge Discovery in Databases (KDD) to explore the relationship between clinical data variables and adult respiratory distress syndrome (ARDS) in critically ill patients. Results of the study and logistics of international research collaboration will be presented at NI '97. The conceptual model, data mining methodology, and objectives for the collaboration are described here.

Desciption of study

A collaborative research effort between nurses at Allgemeine Krankenhaus (AKH) in Vienna, Austria and Duke University Medical Center (DUMC) in Durham, North Carolina USA studied differences in positioning patients with Adult Respiratory Distress Syndrome (ARDS). The overall goals of this collaboration were:

1. To compare outcomes of different nursing interventions used to treat the same conditions.
2. To utilize data and information to improve patient outcomes.
3. To test the capability of a common clinical information system to collect data.
4. To utilize teleconferencing as a medium to support nursing research.
5. To explore transcultural delivery of care.
6. To transcend the geographic boundaries of healthcare.
7. To demonstrate a more efficient method of data collection.

ARDS has a high mortality rate of 50-60% despite recent advances in supportive care. In the United States, ARDS patients are placed in the supine and lateral position, however in Austria ARDS patients at AKH receive radical positioning as a standard of care. Radical positioning involves placing the critically ill patient in a prone position (proning) for 4-12 hours. It was proposed that proning the patient might increase oxygenation by decreasing intrapleural pressure and result in recruitment of previously atalectic alveoli. Complications associated with proning include pressure ulcer development on the face, chest, and pelvis. For the duration of this collaborative study, AKH patients were proned for 12 hours and supine for 12 hours on a low air loss therapy bed. DUMC patients were never proned but received continuous lateral rotation therapy (CLRT) and were turned 60-90 degrees as frequently as tolerated. This difference in standards of practice offered the opportunity to study differences in patient positioning on ARDS patient outcomes.

Two patient outcomes were studied. First, resolution of ARDS, was measured as scores for chest x-ray infiltration, hypoxemia score (PaO2/FiO2), use of PEEP, and static lung

compliance. These were described in the literature as indicators of the severity of lung injury [1] The other outcome studied was the presence or development of pressure ulcers. Except for the radiology information, data were captured on Hewlett Packard CareVue 90001 bedside workstations.
Supporting the DUMC-AKH collaborative research project entailed several interesting challenges from a technical perspective. While both institutions were served by the same type of computing system, each institution configured and used their system differently. Because of these differences, one of the initial tasks in supporting the research effort, was to review each institution's information system to find their similarities and differences.

Clearly, many of the patient care issues that spark research questions are the same on both sides of the Atlantic Ocean. The points of interest are how these problems are addressed and determining which care practices yield better results for the patient. Another interesting challenge was that the Duke database was written in English while the AKH database was written in German. When trying to coordinate data points between the two different databases this difference posed potential research problems. Fortunately, our Austrian colleagues are fluent in English and tolerant of our rudimentary attempts to speak and understand German.

One of our early tasks was to develop translation tables of the data elements in the AKH database and an equivalent phrase in English. Once a common understanding was achieved, the individual datapoints began to have meaning and we began to understand what data to extract from the databases to support the research effort. As we reviewed the differences and similarities that each institution brought to this research effort, we learned that hardware platform and computer operating systems were the same. We also encountered similar patient care situations and patient care problems. That is to say, each institution had the challenge of managing ARDS patients who suffered from sepsis, multi-organ failure and complicated traumatic events. In spite of using the same computer hardware and software tools, each institution used their clinical information system in different ways. At Duke, for example, the information system had a rather extensive physical assessment tool while AKH used a comprehensive order entry component.

Description of informatics tools that supported the study
Once research questions were formulated and the appropriate data points in the databases identified, the challenge turned to extracting data from the clinical databases. This proved to be a nontrival task. The clinical information system was designed to collect and present data about a single patient. Collecting aggregate data and information about an entire population of ARDS, or other types of patients, was not included as a part of the system design. Thus extracting aggregate ARDS patient data proved difficult. The single-patient bedside clinical information record design yielded massive volumes of aggregate data that required significant amounts of manipulation and "cleaning" before data analysis could be conducted. Several software tools were used to extract, manipulate, and archive the desired dataset [2]. This data extraction and archiving was automated so that it could be executed daily and delivered to researchers for further manipulation and analysis.

Once data were available for research, data mining methodologies were used to find clinically relevant patterns from massive datasets that were helpful in predicting patients at risk for ARDS.

1CareVue 9000 version G0.38 is a product of the Hewlett Packard Company, Palo Alto, California, USA

The explosive growth of health care databases has outpaced our ability to interpret and utilize the data we collect and store in massive volumes. New tools and techniques for 'data mining' are the foundation of an emerging field known as knowledge discovery in databases (KDD). Data mining techniques that accomplish knowledge discovery in databases (KDD) have been successfully used in financial and other nonclinical data domains to identify important relationships between data elements. The power of KDD methodologies is their potential ability to identify relationships between data elements and patient outcomes that are hidden in vast amounts of data [3]. The idea of discovering knowledge in massive amounts of patient clinical data is very appealing. Our limited human information processing capacities cannot detect the patterns and trends in massive volumes of data, but data mining software uses a variety of mathematical techniques to sift through mountains of both qualitative and quantitative data to find the repeated themes and matched variables that often escape our human brains. The paradox is that KDD is simultaneously simple and complex. Matching data elements and variables is simple, but matching hundreds of data elements for thousands of patients is complex.

Data mining begins with acquisition of the database(s) containing the relevant input variables as well as the outcome variable. For example, our collaborative study generated a table that includes a Yes/No field to indicate whether or not the patient had ARDS. This is the 'decision', or outcome variable in the data, which is essential both for training and evaluation of new data. The data is then split into two sets - a training set and a testing set. The training data set is mined for classification rules (or whatever chosen knowledge representation schema is used) and then the testing data set is given to the system to evaluate performance of the classification rules. A key advantage of our selected methodology is that classification rules can be inspected by expert nurses and health care administrators in order to check the rules for plausibility and to gain decision-making insight. Classification rules explain the rationale behind each decision. KDD methodologies, using inductive machine learning, produced results that were 53-89% accurate in predicting birth outcomes in pregnant women while humans are only 17-38% accurate at the same predictions [4,5] The same KDD methodologies are in progress for the ARDS data.

Results of this international collaborative study will be completed and reported in Stockholm at the NI'97 meeting. The monthly teleconferences that were conducted to coordinate the research will be described. In addition, more details will be provided concerning the efforts required to coordinate and communicate research methodology, technical problems, human factors and informatics solutions over thousands of distant miles and across language and cultural barriers that made this a particularly interesting and challenging project.

References

1. Heffner J, Brown, L, Barbiere E. Harpel K, DeLeo J. Propsective Validating of an Acute Respiratory Distress Syndrome Predictive Score. *Am J Respir Crit Care Med* 1995; 152: 1518-1526.
2. Saville J et.al. Data Extraction and Archiving for Nursing Research Using a Bedside Clinical Information. *submitted for review for NI '97*, 1996.
3. Fayyad, U. et. al. Advances in Knowledge Discovery & Data Mining. Boston, MA: AAAI Press - The MIT Press, 1996.
4. Woolery L, Grzymala-Busse, J. Machine learning and preterm birth risk assessment. *J Am Medical Informatics Assn* 1994; 1(6):439-446.
5. Woolery L., Grzymala-Busse J. Machine learning for development of an expert system to predict premature birth. *Biomed Sci Instrumentation* 1995; 31:29-34.