Diagnosing Community-Acquired Pneumonia with a Bayesian Network

Dominik Aronsky, MD and Peter J. Haug, MD
Dept. of Medical Informatics, LDS Hospital/University of Utah, Salt Lake City, Utah

We present the development and the evaluation of a Bayesian network for the diagnosis of community-acquired pneumonia. The Bayesian network is intended to be part of a larger decision support system which assists emergency room physicians in the management of pneumonia patients. Minimal data entry from the nurse or the physician, timely availability of clinical parameters, and high accuracy were requirements we tried to meet. Data from more than 32,000 emergency room patients over a period of 2 years (June 1995–June 1997) were extracted from the clinical information system to train and test the Bayesian network. The network performed well in discriminating patients with pneumonia from patients with other diseases. The Bayesian network achieved a sensitivity of 95%, a specificity of 96.5%, an area under the receiver operating characteristic of 0.98, and a predictive value positive of 26.8%.
Our feasibility study demonstrates that the proposed Bayesian network is an appropriate method to detect pneumonia patients with high accuracy. The study suggests that the proposed Bayesian network may represent a successful component within a larger decision support system for the management of community-acquired pneumonia.

INTRODUCTION

Community-acquired pneumonia (CAP) is the sixth leading cause of mortality in the US and the leading cause of death in patients with infectious diseases.[1] The cost of CAP is estimated to be $4 billion per year.[2] The diagnosis and the management of CAP involves much uncertainty when a patient presents to the Emergency Room (ER). At this point, however, important decisions about the empiric antibiotic selection and the admission to the hospital have to be made. Making decisions under uncertainty results in practice variation.

To reduce practice variation guidelines for the management of patients with CAP have been developed.[3,4] One of them has been successfully implemented in the medical delivery systems of Intermountain Health Care.[4] The guideline is paper based and requires additional time to be filled out. Therefore the physicians' compliance varies. Computerizing the guideline may increase the compliance. However, a computerized guideline may only be successful if a sensitive and specific trigger mechanism that accurately identifies patients with CAP is present.

As the quality of computerized patient records improve, decision support systems represent a promising method to improve patient outcomes and cost-effectiveness. Most of the real-time decision support systems are rule-based. Probabilistic methods such as Bayesian networks still need to demonstrate their value and applicability in an integrated clinical environment.

A Bayesian network (BN) is a graphical representation that is based on probability theory, primarily on Bayes' theorem.[5,6] A BN is a directed acyclic graph with nodes, arcs and tables. Each node represents an uncertain variable and is associated with a table representing a probability distribution. The estimation of the conditional probabilities by literature review or with the help of domain experts is tedious and time consuming. In particular, the probabilities of findings in the population without the target disease are difficult to assess. Although clinical databases can potentially provide accurate probabilities, they have not been deployed for the development of a BN, because they often lack the required detail.

CAP is a good candidate for the probabilistic nature of a BN because uncertainty is involved in both the diagnosis and in the management of the disease. The physician encounters much variation in symptoms, findings, and laboratory and blood gas test results. Even in the chest x-ray, which is considered the gold standard for the diagnosis of CAP, the interpretations may vary. The sputum and blood cultures take one or two days to be completed and are often negative. At the time the reports become available, most important decisions have already been made.

In our feasibility study we present the development and the evaluation of a BN for the diagnosis of community-acquired pneumonia in the emergency room.

METHODS

We identified 41,371 patients who presented to the ER of LDS Hospital, Salt Lake City, during a 25 months period (June 1995-June 1997). A primary discharge diagnosis of viral or bacterial pneumonia

1091—8280/98/$5.00

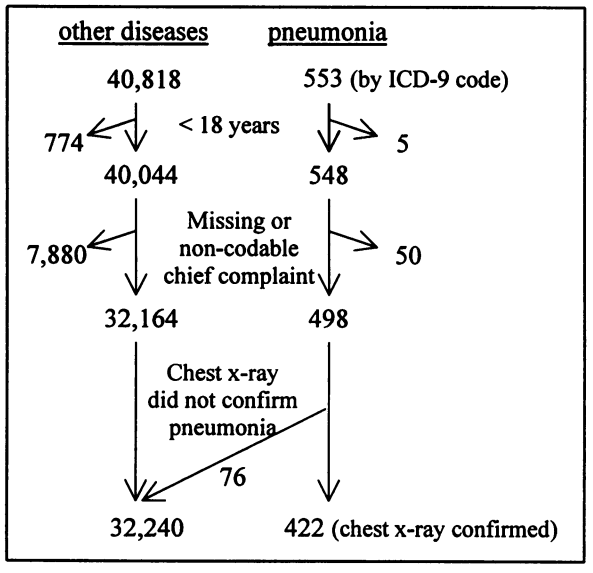

Figure 1: inclusion and exclusion criteria of patients with and without CAP

(ICD-9 code: 480-486) was the inclusion criterion for patients with CAP (553 patients).

Adapting our guideline criteria, we excluded 779 patients who were younger than 18 years. We excluded 4,540 patients who had a chief complaint that has been removed from the current list of coded chief complaints. Due to a change of charting practice in free text chief complaints[7], we excluded 3,390 patients with a free text chief complaint. We excluded a total of 8,709 patients. Fig. 1 illustrates the inclusion and exclusion criteria.

For each of the remaining 32,662 patients, we extracted a total of 65 variables from the HELP System[8], of which 59 were coded and 6 free text. Only the first incidence of the data elements were considered. These data elements originated from different sources. A triage nurse captured the chief complaint, the current and past history, the current medication, allergies, and the vital signs. The nurse who took care of the patient during the encounter entered the patient's assessment. Lab values entered the HELP system through a laboratory interface. Free text information such as the current or past history was parsed for keywords. For all patients we calculated a risk factor similar to the algorithm used in our practice guideline[4].

The chest x-ray reports were extracted for all patients who had one or more chest x-rays taken within the first 72 hours of their encounter in the ER. The chest x-ray interpretation of the ER physician was generally available at a time when relevant decisions were made. However, the chest x-ray interpretation of the radiologist was considered the gold standard for diagnosing CAP. For all the patients with an ICD-9 code of CAP, we extracted both the dictated radiologist's chest x-ray reports and the ER physician's dictated clinical reports. We manually reviewed the reports for all the 498 patients with CAP. We applied the gold standard criteria and only included the 422 patients who had chest x-ray confirmed CAP. For the 32,163 patients without CAP, we identified 8,102 patients who had at least one chest x-ray taken within the 72-hour period. Their chest x-ray reports were parsed for keywords that were suggestive of CAP (e.g. "infiltrate", "consolidation", "no evidence of"). Based on a conservative algorithm we identified 995 patients who actually had other diseases than CAP, but whose dictated chest x-ray reports were compatible with CAP.

We developed the BN with Netica™, a software that performs Bayesian parameter learning.[9] Different network structures were manually developed according to medical knowledge. The distributions for a given network structure were derived from the training set. A 300 MHz PC with 64 MB RAM was used for training and testing. We randomly assigned each of the 32,662 patients to one of three different subsets. We tested the BN with each of the three subsets while the two remaining subsets represented the training set.

An evaluation of the accuracy and the performance of the BN was determined applying measures that are typically used for clinical tests. These included the sensitivity, the specificity, and the positive predictive value.[10] The sensitivity and the specificity are important descriptive characteristics of a diagnostic test. To the clinician, however, the predictive value has a more clinically oriented meaning. In a patient with a positive test, it indicates, how many times a true or false positive result can be expected. Unfortunately, the prevalence of a disease influences the predictive value, whereas the sensitivity and the specificity are more consistent in the face of varying prevalence.

We calculated the receiver operating characteristic (ROC) curve to refine and evaluate different versions of the BN. The ROC curve is a graphic measure that plots corresponding pairs of the true positive rates (sensitivity) and the false positive rates (1-specificity).[11] The area under the ROC curve is a standard measure indicating the overall performance of a diagnostic test.[12] A lack of discriminatory ability exists when the sensitivity equals the specificity in which case the ROC curve is a 45° line and the corresponding area under the curve equals 0.5. Perfect discrimination exists when the sensitivity and the specificity equal 100% which yields an area under the ROC curve of 1.0.

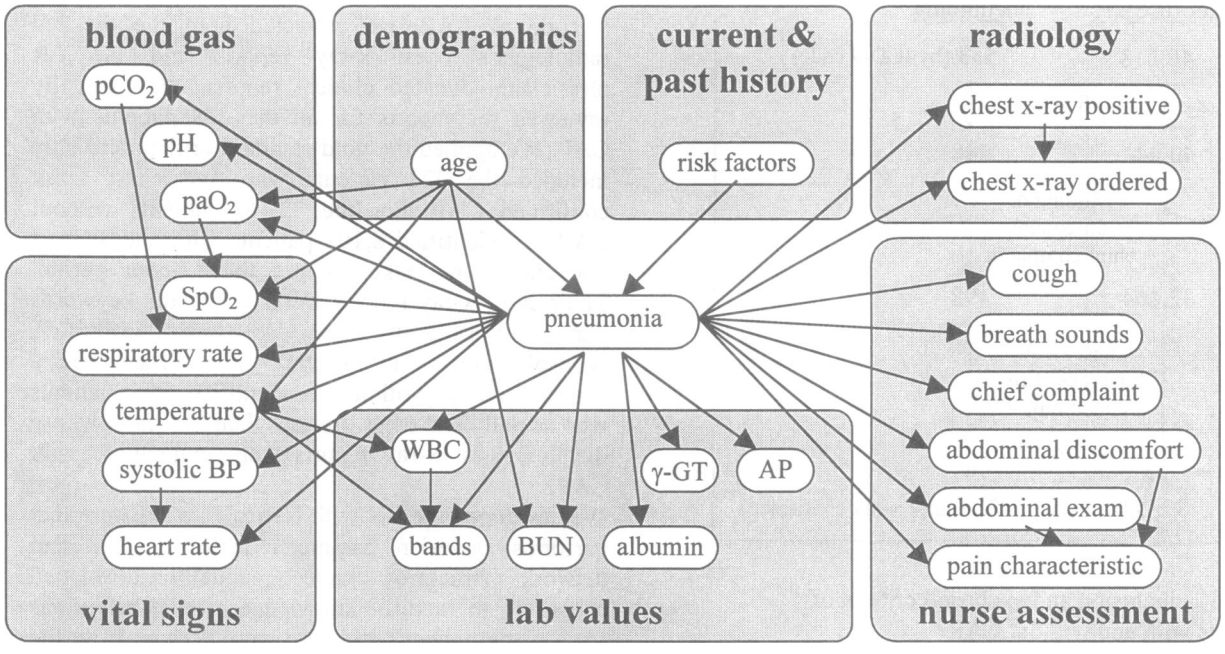

Figure 2: Structure of the Bayesian network. All variables are available in the HELP system during a patient's encounter in the emergency room with the exception of the chest x-ray information ("chest x-ray positive").

RESULTS

We implemented and evaluated more than 50 different network structures. The number of nodes in the different network structures ranged from 20 to 77 nodes. The size of the BN ranged from 262 kB to 8.6 MB and the run time for 100 cases ranged from 6 to 46 seconds.

The most parsimonious and most accurate BN contained 25 nodes, 38 links, and 10,100 conditional probabilities (Fig. 2). There were 3 dichotomous, 6 categorical and 16 continuously valued nodes. The node "chief complaint" contained 60 different states. The BN was 262 kB large and required 6 seconds to compute a probability of CAP for 100 cases.

The results of the three different test subsets are presented in Fig. 3. When the sensitivity was fixed at 95%, the corresponding specificity averaged 96.5%.

set	specificity (sensitivity fixed at 95%)	positive predictive value	area under the ROC curve
1	97.3 %	30.1 %	0.991
2	95.6 %	21.2 %	0.977
3	96.6 %	29.1 %	0.979

Figure 3: Results for the three testing set.

The mean predictive value positive was 26.8% and the average area under the ROC curves of the three subsets was 0.9825. For subset 3 we show the ROC curve (Fig 4), the 2x2 table (Fig. 5), and the most frequent discharge diagnosis (ICD-9) for the false positive group (Fig. 6.)

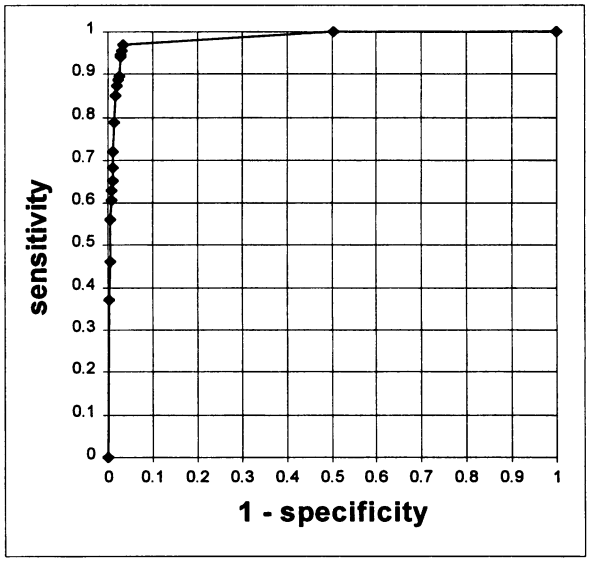

Figure 4: The ROC curve for test set 3. The area under the curve is 0.979

	patient with CAP	patients without CAP	total
BN positive	155	378	533
BN negative	8	10,622	10,630
total	163	11,000	11,163

Figure 5: 2x2 table of subset 3. The sensitivity is 95%, the specificity is 96.6%, and the predictive value positive is 29.1%.

congestive heart failure	32
aspiration pneumonia	19
urinary tract infection	17
fever of unknown origin	17
acute bronchitis or bronchiolitis	10
acute upper respiratory infections	9
other symptoms involving respiratory tract	8
status asthmatics	8
unspecified viral infections	6
painful respiration	6
chest pain	6
acute respiratory distress	6
respiratory failure	5
pulmonary embolism	5
chronic obstructive asthma	5
chronic bronchitis with acute exacerbation	5
asthma unspecified	5
acute pyelonephritis	5

Figure 6: Discharge diagnosis (ICD-9) of false positive patients in subset 3. Only diseases with more than four patients are listed. They account for 46.0% of all the 378 false positive cases in this subset.

DISCUSSION

As part of the feasibility study we have developed a BN for diagnosing community-acquired pneumonia in patients who present at the ER. The BN is being designed to screen every patient who presents to ER and to alert the ER physician about the possible presence of a patient with CAP. If the ER physician acknowledges the alert and confirms the diagnosis, the BN may trigger the computerized practice guideline for the management of the patient.

In our opinion three important requirements are important when the feasibility of a probabilistic real-time decision support system is evaluated. First, a sensitivity of 95% or higher combined with a very high specificity is mandatory. Second, most data elements in the BN and the updated probability have to be available while the patient is in the ER and before the physician has made his final decisions. Third, any additional data entry beyond the normal charting practices of the nurse or the physician should be eliminated or kept to an absolute minimum. Considering the clinical application of the BN as a screening and alerting tool, the predictive value positive requires special attention. Although the combination of a 95% sensitivity with a 96.5% specificity is excellent, the predictive value is clinically more informative for the ER physicians. The predictive value specifies how many times the BN will alert the physician in patients with and without CAP. Our BN averages a predictive value positive of 26.8% which indicates that out of 4 issued alerts, only one would actually be a patient with CAP. The three false positive alerts, however, represent in large valid differential diagnosis to CAP (Fig. 6). Considering that the ER physicians see about 55 patients every day and one patient with CAP about every second day, the BN would alert them twice a day. Improving the predictive value will be difficult given the current level of specificity. An area under the ROC curve of 0.98 demonstrates that the overall accuracy and the discriminatory ability are exceptionally high.

To achieve this high level of accuracy, the clinical variables must be available during the patient's encounter in the ER. Clearly, an alert assists the physician only while the patient is in the ER. During the training and testing phase the BN has been presented at one single point with all the available data elements of a patient. The clinical work flow, however, is much different. The data are not present at one single point, but are gathered and charted over the entire time period that a patient is in the ER. In a BN, however, not all of the data have to be present. The BN accounts for the presence of uncertainty and can operate with varying amounts of missing data. Whenever the clinical information system records a new piece of data, the BN can incorporate the new evidence and update the joint probability. Experience suggests that the physician's compliance with a computerized decision support system depends on the amount of additional data elements that have to be entered. Currently, the BN incorporates variables that are part of the nurses' charting practices or originate from the laboratory. From both the nurse and the physician the BN does not need additional information, with the chest x-ray being the only exception.

The radiologists' chest x-ray interpretation is currently not available in the clinical information system during the patient's encounter in the ER. However, the chest x-ray is important for the diagnosis of CAP. For the following pilot study, we will need to prompt the ER physician to indicate whether the chest x-ray is positive or negative, a task which they have agreed to do. Since the HELP

System records the time when a chest x-ray has been ordered, the BN will recognize for which patients it should prompt the ER physician. Although the radiologists' interpretation would be preferred, the current work flow does not provide a feasible procedure to include their interpretation while the patient is in the ER.

There are limitations in our study. First, the BN was designed for CAP in the ER of a tertiary care hospital. However, most of the patients with CAP are treated by their primary care physician. The representativeness of our database is therefore limited as our population represents a selected group of patients only. Patients that enter our ER may be more seriously ill and have a unrepresentative spectrum of causative organisms. Second, it is difficult to predict which data elements are gathered in a specific patient, and in which order they enter the HELP system. For instance, a blood gas sample is not obtained in every patient, and if it were, the moment will greatly vary at which the results will become available to the BN. With incomplete evidence the BN may cross the threshold and generate an alert, but may then drop the probability after more evidence becomes available. Defining a minimal set of instantiated variables may help to prevent premature alerts. Third, the database contains retrospective data, but a prospective evaluation in our ER is required to demonstrate whether the BN will perform as expected.

CONCLUSION

The results of our study have general significance for the application of Bayesian networks in a clinical environment. Clinical information systems are an accurate source for the assessment of probabilities that are required for the development of a probabilistic decision support system. The results obtained are encouraging and suggest that a Bayesian network may provide a promising method as a real-time decision support system in a clinical environment. We feel confident to perform a pilot study in the ER and test whether the Bayesian network may be an accurate component within a larger decision support system that assists emergency room physicians in the management of community-acquired pneumonia.

Acknowledgments:

The author is supported with a grant from the Swiss National Science Foundation (31-44406.95). I thank my advisors, Reed M. Gardner and R. Scott Evans, for suggestions on the manuscript.

References

1. Garibaldi RA. Epidemiology of community-acquired respiratory tract infections in adults: incidence, etiology, and impact. *Am J Med*, 1985; 78:32-37.
2. Medicare and Medicaid statistical supplement. *Health Care Financ Rev*, 1995; 16 (September).
3. Fine MJ, Auble TE, Yealy DM, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med*, 1997; 336: 243-50.
4. Dean NC, Bateman KA, Hadlock CJ, McKinstry CA, James BC.Use of Care Process Model for Community Acquired Pneumonia in Three Rural Counties. American Thoracic Society, Proceedings of the 1997 International Conference, San Francisco, CA.
5. Jensen FV. An introduction to Bayesian Networks, Springer-Verlag New York Inc., New York, 1997.
6. Pearl J. Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann, San Mateo, California, 1988.
7. Aronsky D, Merkley K, Haug PJ, James BC. Reducing Free Text Entries: A Continuous Quality Improvement Project. AMIA 1998 Annual Symposium (submitted).
8. Kuperman GJ, Gardner RM, Pryor TA: HELP: A Dynamic Hospital Information System. New York: Springer-Verlag, 1991.
9. Netica, Application for Belief Networks and Influence Diagrams, User's Guide, Norsys® Software Corp., 1997.
10. Sox HC, Blatt MA, Higgins MC, Marton KI. Medical Decision Making. Butterworth-Heinemann, Newton MA, 1988.
11. Metz CE. Basic principles of ROC analysis. *Seminars in Nuclear medicine*, 1978;8:283-98.
12. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 1982;143:29-36.