Application of an Information-Theoretic Method for Efficacy Assessment

David A. Gift, William R. Schonbein, Eugene L. Saenger, and E. James Potchen

Department of Radiology, Michigan State University, East Lansing, Michigan

An information-theoretic pattern recognition method was used to construct descriptive models of data related to 1,674 radioisotope lung scan referrals for the purpose of assessing lung scan influence on diagnosis and management of pulmonary embolism. It was observed that, relative to other clinical information available prior to the scan, the lung scan significantly improved the ability of the models to predict diagnostic and management outcomes, implying that the lung scan has significant influence on these clinical decisions.

J Nucl Med 26:807–811, 1985

A common approach to assessing the efficacy of a diagnostic test involves descriptive modeling of the test's use in clinical practice. Models are constructed which relate attributes describing the patient (e.g., signs, symptoms, test results) to various outcomes (e.g., final diagnosis, choice of therapeutic management or some measure of health status subsequent to management). The investigator first attempts to construct the most accurate model relating attributes to outcomes without using the result of the test-of-interest (TOI) as an attribute, and then attempts a similar model including the TOI as an attribute. The difference in accuracy of the two models provides a measure of the diagnostic information provided, or the influence on clinical decisions exerted, by the TOI relative to the other patient attributes used in the models.

Many analytical modeling techniques may be employed for this type of study. This paper reports results of a study of radioisotope lung scanning for the diagnosis and management of pulmonary embolism (PE) in which Christensen's entropy-minimax pattern detection method (*1*) was used to construct the required models. This work is a part of the Society of Nuclear Medicine Efficacy Study of lung scanning in which two modeling techniques were used. Saenger et al. report the results of logistic regression modeling in a related paper (*2*).

Received Aug. 10, 1984; revision accepted Dec. 6, 1984.

For reprints contact: David A. Gift, MS, Dept. of Radiology, B-220 Clinical Ctr., Michigan State University, East Lansing, MI 48824-1315.

METHODS

Multivariate data regarding 2,023 lung scan (LS) cases were collected over the course of two years from 22 hospitals distributed across the U.S. The cases were divided into two groups by a mid-way date of receipt of the data collection forms. Models were constructed for the Group I data (n = 1,065), and the Group II data (n = 958) were used to validate the performance of models "trained" on Group I data.

Data collected for each LS case included 24 signs, symptoms and history variables, the referring physicians' pre-LS and post-LS choice of the most likely and most important diagnoses for the patient (most important diagnosis is that which is most critical to the patient's clinical management, although it may not be the most likely diagnosis) and the likelihood estimates for each of these diagnostic possibilities, the attending physicians' pre-LS and post-LS choice of therapeutic management strategy, the patients' discharge diagnosis, and, from the nuclear medicine specialist who interpreted the LS, the scan technique, three potential diagnoses indicated by the scan (in order of likelihood), and a description of defects seen in the scan, if any. All information regarding the patient was obtained from the referring physician, with the exception of the items identified in the preceding sentence which were provided by the nuclear medicine specialist. The data collection instrument and its design and testing are described by Saenger et al. (*2*).

Since this study was intended to investigate LS refer-

rals prompted by concern regarding PE, the Group I and Group II data were analyzed in subgroups defined by the referring physician's pre-LS choice of most-important (MI) and most likely (ML) diagnosis. The MI/ML diagnosis subgroup PE/PE accounted for 275 cases or 26% of Group I and 245 cases or 26% of Group II. The MI/ML diagnosis subgroup PE/OTHER accounted for 640 cases or 60% of Group I and 514 or 54% of Group II. The OTHER/OTHER and OTHER/PE subgroups, representing 14% of Group I and 20% of Group II, were not used in these analyses.

The influence of the LS on determination of discharge diagnosis and on the decision to use anticoagulant therapy (ACT) were selected as foci of this efficacy assessment. For analysis of the former, two outcomes were employed: discharge diagnosis including PE, and discharge diagnosis excluding PE. Two outcomes also were employed for analysis of the latter: management including ACT, and management excluding ACT.

Selection of attributes from those available for use in modeling was accomplished in two steps designed to select those attributes which would be most likely to yield the best models with respect to the selected outcomes. First, attributes having extremely low frequency, that is, occuring in less than 3% of cases, were excluded due to the absence of discriminatory power. No attributes occured with very high frequency, that is, in nearly all cases; they also would have been excluded. Second, an information entropy measure (described below) was computed for the outcomes with the Group I data partitioned in a univariate manner for each attribute (attribute present, attribute absent). Those 13 attributes producing the smallest entropy values were selected for use in pattern recognition analyses. The computer program is limited to 14-dimensional data; the fourteenth attribute position was reserved for adding the lung scan diagnosis to the models. Results reported in this paper were derived from analyses using the 13 binary attributes (yes/no, present/absent) listed in Table 1 and the primary nuclear medicine diagnosis regarding PE (secondary and tertiary LS diagnoses were rarely reported, and mention of PE always occured in the primary LS diagnosis).

TABLE 1
Patient Attributes Used in Modeling

Abnormal chest radiograph	Hypoxemia
Arrythmia	Prior pulmonary disease
Chest pain	Rales
Cough	Sex (female/male)
Cyanosis	Syncope
Dyspnea	Tachypnea
Fever	

DATA ANALYSIS

Attribute-outcome modeling of the data was performed using the entropy-minimax pattern detection method of Christensen (*1*). The method operates by systematically searching the entire data set for that subset of cases having a common pattern of attribute values and having the least information entropy, H, estimated from the proportion of outcome occurrences within the subset. The data in the subset are then removed from further consideration and the process is repeated until all of the original data have been partitioned into such subsets, or until a subset is identified for which the entropy is not significantly different from the entropy of the original unpartitioned data. Attribute patterns defining data subsets consist of logical conjunctions of attribute values (e.g., cough *and* no dyspnea *and* hypoxemia), and may involve any number of the analyzed attributes.

The measure of information entropy used is Shannon's (*3*),

$$H = -\Sigma p_i \log_2 p_i,$$

where p_i is the probability of the i-th outcome occurring in the data subset, and the sum is over all outcomes. H is maximized when all outcomes are represented in equal proportions. As the proportion of the subset data representing one specific outcome goes to unity and the proportion representing all other outcomes goes to zero, H also goes to zero. For this reason, H may be interpreted as a measure of the uncertainty associated with using an attribute pattern as a decision rule in an attempt to predict the outcome of a randomly-chosen sample for which the outcome is unknown. Thus, the entropy-minimax algorithm models the data set by partitioning it into subsets such that the H of the model is minimized. The model H is equal to the sum of all subset H's, each weighted by that fraction of the original data set samples which belong to the subset. That is, the uncertainty associated with the set of decision rules comprising the data set model is equal to the weighted average of the individual decision rule uncertainties, where the weighting factor is an estimate of the likelihood of having to use that individual decision rule in an attempt to predict the outcome of a randomly-chosen sample.

MODELING RESULTS

Results of the assessment of LS influence on determination of discharge diagnosis are shown in Table 2. Two entropy-minimax models were constructed on each of the Group I data subgroups, PE/PE and PE/OTHER. In the first model, the thirteen attributes listed in Table 1 were related to the discharge diagnosis outcomes PE and No PE. The information entropy of a data subgroup before modeling is computed from the prevalence

TABLE 2
Lung Scan Influence on Discharge Diagnosis: Group I Modeling Results

Data subgroup	PE/PE		PE/OTHER	
Attributes	Table 1	Table 1 + scan	Table 1	Table 1 + scan
Subgroup H (bits)	0.99	0.99	0.56	0.56
Model H (bits)	0.82	0.57	0.51	0.45
REDUCTION IN H	17%	42%	9%	20%

TABLE 3
Lung Scan Influence on Anticoagulant Therapy: Group I Modeling Results

Data subgroup	PE/PE		PE/OTHER	
Attributes	Table 1	Table 1 + scan	Table 1	Table 1 +scan
Subgroup H (bits)	1.0	1.0	0.64	0.64
Model H (bits)	0.82	0.66	0.61	0.57
REDUCTION IN H	18%	34%	5%	11%

of these outcomes in the subgroup. Before modeling the relationship of attributes to outcomes, one would have only this prevalence information to use in predicting the outcome for a randomly-selected patient. With respect to this discharge diagnosis in the Group I subgroups PE/PE and PE/OTHER, H = 0.99 and 0.56 bit, respectively. The radix of the logarithm in Shannon's expression for H determines the unit of information entropy; base-2 logarithms make the unit of the measure bits. As a reference, a prediction involving two equally-likely outcomes, such as calling the flip of a fair coin, has H = 1 bit. The entropy-minimax models of the Group I subgroups PE/PE and PE/OTHER for diagnostic outcome have H = 0.82 and 0.51 bit, respectively. Thus, modeling the relationship between Table 1 patient attributes and diagnostic outcome in the Group I data subgroups results in reductions of predictive uncertainty regarding discharge diagnosis of 17 and 9%.

In the second such model of the Group I data subgroups the primary LS diagnosis was added to the Table 1 attributes. The resulting models for the PE/PE and PE/OTHER subgroups have H = 0.57 and 0.45 bit, respectively. The incremental reductions in model H, relative to the Table 1 attributes alone, due to addition of the LS diagnosis to the models are 25% and 11%, more than doubling the reduction due to the Table 1 attributes alone.

Similar pairs of models were constructed for the management-strategy outcomes ACT and No ACT, and the results are shown in Table 3. With respect to these outcomes, the Group I PE/PE and PE/OTHER subgroups have H = 1.0 and 0.64 bit, respectively. The models relating Table 1 attributes to these management-strategy outcomes have H = 0.82 and 0.61 bit, respectively, representing 18% and 5% reductions of H. Adding the LS diagnosis to the Table 1 attributes results in models having H = 0.66 and 0.57 bit, respectively, and 16% and 6% incremental reductions of H, again roughly doubling the reduction due to the Table 1 attributes alone.

VALIDATION RESULTS

As described earlier, entropy-minimax models consist of a hierarchical set of attribute-outcome relations which may be thought of as a set of hierarchical decision rules for making predictions of outcome based on knowledge of attributes. The rule set is hierarchical in the sense that the rules must be applied in the same order as that in which they were derived originally during modeling. The pattern detection algorithm attempts to determine the least-H new attribute-outcome relationship at each step in the modeling process, so the rules tend to be derived in order of increasing information entropy. At some point in the modeling process, after, say, k rules have been derived, a new rule may be determined for which the predicted outcome probabilities are not significantly different for the original outcome prevalences. At this step, the decision rules are no better at predicting outcomes than is prediction using knowledge only of outcome prevalence. Hence, in validating the performance of Group I models on Group II data, only the performance of those first k rules ("k-rule set") is noted, where the (k + 1)-th rule is the first for which the rule's predicted outcome probabilities are not significantly different from the pre-model data set outcome prevalences.

This practice provides another measure of model performance in addition to predictive accuracies. The proportion of the validation data set in which the k-rule set makes predictions is a measure of the extent of the model's descriptive ability in the validation data.

The validation results of Group I PE/PE and PE/OTHER subgroup models on the associated Group II subgroup data are shown in Table 4 in terms of the predictive value positive (PVP), predictive value negative (PVN), and proportion of the Group II subpopulation to which the k-rule set applies (% POP). In every analysis but one (PVN for predicting ACT use in the PE/PE subgroup), adding the LS result to the Table 1 attributes significantly ($p < 0.03$) enhanced the predictive performance of the models, as well as significantly ($p < 0.0001$) extended the descriptive domain of the k-rule set.

All four models of the PE/OTHER subgroup had k-rule sets comprised solely of "negative" decision rules (i.e., predict No PE or predict No ACT), thus, only the PVN of these rules may be calculated. This particular result may imply that, among those patients for whom

TABLE 4
Performance of Group I Models in Group II Data

Subgroup	Outcomes	Attributes	PVP*	PVN†	% POP‡
PE/PE	PE; No PE	Table 1	0.29	0.75	40
		Table 1 + scan	0.58	0.83	71
			($p < 0.0001$)§	($p < 0.03$)	($p < 0.0001$)
PE/OTHER	PE; No PE	Table 1	Undef.¶	0.83	30
		Table 1 + scan	Undef.	0.92	46
				($p < 0.0001$)	($p < 0.0001$)
PE/PE	ACT; No ACT	Table 1	0.50	0.63	42
		Table 1 + scan	0.82	0.65	46
			($p < 0.0001$)	(N.S.)	(N.S.)
PE/OTHER	ACT; No ACT	Table 1	Undef.	0.78	24
		Table 1 + scan	Undef.	0.85	35
				($p < 0.02$)	($p < 0.0001$)

* Predictive Value Positive = $\frac{(\#\text{ correct PE or ACT predictions})}{(\#\text{ PE or ACT predictions})}$.

† Predictive Value Negative = $\frac{(\#\text{ correct No PE or No ACT predictions})}{(\#\text{ No PE or No ACT predictions})}$.

‡ Percentage of the subgroup population for which the k-rule set makes predictions.

§ Numbers in parenthesis are probabilities that difference in values occurred by chance; "N.S." means "not significant."

¶ Undefined.

PE is a possible but not most likely diagnosis, the LS has its greatest influence by diminishing the likelihood of PE and discouraging use of ACT.

DISCUSSION

As assessment of efficacy such as that described herein involves observing the details of many *individual* instances of clinical decision-making, then modeling the data to determine if a particular decision input noticeably affects *aggregate* decision-making behavior. Here, data regarding 1,674 lung scan (LS) referrals were analyzed in an attempt to determine the influence of the LS result on attending physicians' final diagnoses regarding the presence of PE, and on their decisions to anti-coagulate patients. Modeling of the first 915 LS cases by entropy-minimax pattern detection demonstrated that models which include the LS result have significantly reduced predictive uncertainty regarding a discharge diagnosis of PE or the use of anti-coagulant therapy (ACT) than do models constructed without the LS result. Validation of model performance in the remaining 759 LS cases confirmed this result. Models incorporating the LS result had greater accuracy, and improved accuracy over a greater proportion of the sample cases, than did models without the LS result. The conclusion is that the lung scan significantly influences aggregate clinical decision-making with respect to the diagnosis and management of PE.

Discharge diagnosis and choice of therapeutic management are subjective decision outcomes which were chosen as endpoints for this study, rather than a "gold standard" such as the result of pulmonary angiography. This was done because of an intent to study the influence of LS results in the context of typical clinical practice. Angiography was performed in 5% of the cases obtained for analysis. Limiting analysis to only those cases for which angiographic results were available would have resulted in a small and potentially very biased sample (*4*). Furthermore, it was considered inappropriate to require pulmonary angiograms of all LS referrals at participating hospitals as part of the study design.

It is important to note that the method for efficacy assessment presented in this paper does not limit the nature of the investigation. If subjective outcomes are used, one may assess the relative *influence* of the test-of-interest (TOI) on these outcomes. If a "gold standard" outcome is used, one may assess the relative *information* provided by the TOI. In any case, this method provides a means to assess the efficacy of the TOI relative to other information available within the context of TOI use, which is a critical part of an efficacy study design.

The entropy-minimax method of pattern detection is suited specifically to analysis of the "noisy" qualitative data typical of clinical practice research, does not require assumptions regarding the underlying statistical distribution of the observed data, does not require subjective likelihood estimates (although they may be used), and provides a reasonable means of dealing with missing data values (not a problem in this particular study) (*5*). This and other combinatorial pattern recognition methods (*6, 7*) can be very useful as adjuncts to traditional statistical methods of modeling for the study of efficacy.

ACKNOWLEDGMENT

Research supported by U.S. Department of Energy Contract No. DE-AC02-77EV04533.

REFERENCES

1. Christensen R: *Entropy Minimax Sourcebook,* Vol. 1. Lincoln MA, Entropy Limited, 1981
2. Saenger EL, Buncher CR, Specker BL, et al: Determination of clinical efficacy: Nuclear medicine as applied to lung scanning. *J Nucl Med* 26:793-806, 1985
3. Shannon CE: A mathematical theory of communication. *Bell System Tech J* 27, 1948
4. Philbrick JT, Horwitz RI, Feinstein AR, et al: The limited spectrum of patients studied in exercise test research: Analyzing the tip of the iceberg. *J Am Med Assoc* 248:19, 2467-2470, 1982
5. Gift DA, Schonbein WR: An introduction to entropy-minimax pattern detection and its use in the determination of diagnostic test efficacy. In *Entropy Minimax Sourcebook,* Vol. 4 (Chap. 27), Christensen R, ed. Lincoln MA, Entropy Limited, 1981
6. Duda RO, Hart PE: *Pattern Classification and Scene Analysis,* New York, John Wiley and Sons, 1973
7. Stoffel JC: A classifier design technique for discrete variable pattern recognition problems. *IEEE Trans on Comp* C23:4, 428-441, 1974