Clustered knowledge representation:
Increasing the reliability of computerized expert systems

Hong Yu MD, Peter J Haug MD, Michael J Lincoln MD,
Charles Turner PhD, Homer R Warner MD

University of Utah Departments of
Medical Informatics, Medicine, and Psychology

Abstract

Accuracy and reliability are major concerns for the developers of expert systems. We have developed a Bayesian expert system for medicine, Iliad. We have included in Iliad new types of decision frames, called clusters. Clusters are frames (usually Boolean) that contain groups of conditionally dependent findings. These groups often describe pathophysiologic concepts.

Because clusters encapsulate groups of conditionally dependent findings, we expected clusters might increase Iliad's accuracy and reliability. We tested this hypothesis by measuring the reliability of pairs of clustered and nonclustered frames using real patient data.

Introduction

We have developed a new model of knowledge representation, called "clusters," for use in the Iliad[1] medical education system. Before this new model was developed, Iliad's old knowledge base contained only sequential Bayesian decision frames. Because of conditional dependencies between patient findings in this old knowledge base, Iliad previously produced overconfident decisions. The new, clustered knowledge model has substantially reduced these overconfident tendencies. This paper compares the reliability of Iliad's decisions using the new clusters and the old, purely Bayesian knowledge base.

Iliad is a frame based system for medical education that runs on Macintosh computers. This system performs two major functions, consultations and simulations. In the consultation mode, a medical student presents a real case to Iliad and Iliad generates a differential diagnosis. Iliad's teaching tools allow the consulting student to pose "what if" scenarios, examine the expert logic in the knowledge base, ask for explanations of diagnoses or findings, and ask Iliad to reveal the next most pertinent information to seek at any point in the case. In the simulation mode, Iliad simulates an unknown, unique case. These cases are generated anew from data in HELP's clinical patient database each time an Iliad simulation is requested. Iliad can select the simulation topic by comparing a student's logged case experience against the official list of third-year student clerkship goals. Alternatively, the student can request a particular type of consultation. Iliad then presents the student with the chief complaint and allows the student to question the "patient." Iliad tracks the student's query strategy and periodically prompts the student to postulate a differential diagnosis. The student's strategy and diagnoses are compared to what Iliad would have asked and concluded given the same information.

In both modes, the need for accurate, reliable diagnoses is acute. Obviously inaccurate consultations will be rejected out of hand, causing students to lose interest in the system. Worse, in the simulation mode, an unreliable Iliad could actually teach students inaccurate diagnostic strategies. But the diagnoses produced by the old knowledge model were overconfident, as is typical of strictly Bayesian frames.[2] This overconfidence was caused by conditionally dependent, co-occurring findings in the knowledge frames. The old knowledge base attempted to account for conditionally dependent findings in one of two ways. The first solution was to use conditionally dependent findings as alternatives (using Boolean logic). The second solution was to eliminate most conditionally dependent findings, leaving only "key" findings. Unfortunately, these solutions were not completely successful.

The first solution restricted the Bayesian statistical analysis to *either* one conditionally dependent finding *or* another in any given patient. One advantage of this approach was that conditionally dependent findings could remain in the frame. This strategy worked best when a small number of conditionally dependent findings represented diagnostic alternatives. The Boolean statements could arbitrate between the findings so that the most powerful finding available was used in each individual case. But these statements became increasingly complex when multiple conditional dependencies were represented.

The second solution eliminated all but "key" findings from the Bayesian frame, producing "sparse" frames. For example, the sparse frame for *Lung Consolidation* deleted the less specific physical findings of lung consolidation in favor of the "key" finding, radiographic lung infiltrate. Sparse frames handled large numbers of conditionally dependent findings very well: they simply eliminated them. But this had significant educational disadvantages. Sparse frames were terse, non-informative to students exploring the knowledge base, and could not respond appropriately when non-key findings were entered during a patient consultation. Actual patients present with rich, diverse sets of findings, not just "key" findings.

Because the solutions for conditional dependence provided by the old knowledge model were inadequate, we decided to develop a new model of knowledge representation. "Clusters" are frames containing groups of highly conditionally dependent disease findings that describe pathophysiologic states. Clusters generally use Boolean decision logic to return outcomes that can be used in Bayesian frames. These outcomes can range from "confirmed" through "probable" or "suggested" to "denied" (Figure 1). The frames developed using the new knowledge

0195-4210/88/0000/0126$01.00

model are Bayesian, but contain clusters as subroutines (Figure 2).

Title: Chronic Airways Inflammation
Type: Boolean
Variables: prolonged cough **as** (have you coughed daily for more then 2 months?)
winter cough **as** (do you cough daily during the winter months?)
cough last year **as** (did you have a similar cough a year age?)
recurring cough **as** (do you have spells of increased cough and sputum?)
morning cough **as** (is your cough usually worse in the morning?)
rhonchi **as** (rhonchi)

LOGIC: confirmed — if [exist (prolonged cough) **or** exist (winter cough)] **and** [exist (cough last year) **or** exist (recurrent cough sputum)] then **true** else **false**.

suggested — if exist (winter cough) **or** exist (prolonged cough) **or** exist (recurring cough sputum) **or** exist (morning cough) **or** exist (rhonchi) then **true** else **false**.

denied — if not exist (prolonged cough) **and** not exist (recurring cough) **and** not exist (winter cough) **and** not exist (cough last year) **and** not exist (morning cough) **and** not exist (rhonchi) then **true** else **false**.

Figure 1. An example of the cluster *Chronic Airways Inflammation*

We have compared the reliability of decision frames built using the old knowledge model with that of frames using the clustered model. This comparison was conducted using a group of real patients from the HELP system patient database.[3] Because attempts had already been made to minimize overconfidence in the old frames, our procedure of comparing them to the new, clustered frames provides a conservative test of the hypothesized benefits of clustering. The present paper examines two inter-related hypotheses. First, substantial conditional dependence exists between findings in non-clustered Bayesian frames and this dependence leads to inaccuracies in assigning probabilities. These inaccuracies typically consist of a tendency to overestimate the probabilities of likely diagnoses and underestimate the probabilities of unlikely diagnoses. This behavior is referred to as "overconfidence". Second, a clustered knowledge representation can reduce these inaccuracies by reducing the effects of conditional dependence in Bayesian frames. If these hypotheses are proven, clusters provide a method of improving Iliad's diagnostic reliability. This improvement will facilitate the use of Iliad as a clinical consulting and teaching tool.

Method

The description of our method will be comprised of three parts. First, we will describe our approach to frame development using clusters model. Second, we will describe the patient population within which the two knowledge models were compared. Third, we will describe the statistical procedures used to assess diagnostic reliability.

Frame Development

The original, non-clustered knowledge frames were developed by Haug and others as part of the HELP knowledge engineering project.[4] The new, clustered frames are direct descendents of these original frames, and deliberately contain exactly the same patient findings. The only difference is that some findings are now contained in clusters. Three of the authors, a general internist (PH), a pulmonary internist (MJL), and a pediatrician (HY), developed these new frames. These new frames have the same apriori prevalences as their older counterparts. When individual findings are not clustered, these findings have the same statistical prevalence in diseased and non-diseased populations in both the old and new frames. New frames using clusters were developed for chronic bronchitis, bacteria pneumonia, pulmonary embolism, and asthma. Clustered frames are currently under development for six additional pulmonary diseases.

The Test Population

Our test population was comprised of 517 patients hospitalized in 1985 at the LDS Hospital in Salt Lake. We selected patients who had received a chest radiograph to ensure an adequate sample of patients in our test population with lung disease; some patients had received incidental radiographs and did not have pulmonary disease. Each patient had a HELP database file that contained historical, radiographic, and laboratory data gathered during the hospitalization. The HELP system also contained the final ICD-9 diagnosis assigned to each patient. The ICD-9 code was used to indicate the "gold standard" diagnosis.

We measured the prevalence of each disease, finding, and cluster outcome in our patient database. These measurements provided exact apriori disease prevalences, sensitivities, and specificities for findings and cluster outcomes. One might argue that we should have derived these statistics in a separate

Title: Chronic Bronchitis diagnosis
Type: Sequential Bayesian
Apriori: 0.0619

Cluster variables:	**confirmed TPR/FPR**	**probable TPR/FPR**	**suggested TPR/FPR**	**denied TPR/FPR**
Chronic airways inflammation	.44 / .14	---	.69 / .32	.31 / .71
Cigarette exposure	.63 / .21	---	---	.38 / .79
generalized airway obstruction	.09 / .02	---	.31 / .23	.34 / .31

Non clustered:	**TPR**	**FPR**
pulmonary toxin exposure **which is** (Have you been exposed to large amounts of dust or fumes in the work place?)	0.41	0.18
hx chronic bronchitis **as** (Do you have chronic bronchitis?)	0.38	0.03
xray COPD **as** (CHEST XRAY: Emphysema/COPD)	0.59	0.06

Figure 2. An example of the *Chronic Bronchitis* Bayesian frame with clusters

population and then applied them to our test population. However, we were primarily concerned with the effect of the clustered knowledge model on diagnostic reliability. We wished to eliminate the any confounding variables that might influence diagnostic reliability, such as inaccurate probability estimates. By providing perfect population statistics we were able to provide the opportunity for each knowledge model to achieve perfect diagnostic reliability, if only overconfidence did not occur.

Assessing reliability:

Hilden, et. al., have defined a series of statistics for assessing the reliability of probabilistic medical expert systems.[5,6,7] (see appendix). Hilden describes two different ways to assess reliability. In the first case, the expert system provides a continuous estimate of diagnostic probability. For example, the system provides a specific estimate of the probability of disease in a patient. In the second case, the expert system provides a dichotomous estimate of disease probability. In this case, the expert system provides a decision that the disease is present or absent. There is evidence that information is lost in dichotomous probability assessments.[8] We chose to assess reliability both ways because doctors are often forced to make dichotomous decisions.

The goal of Hilden's reliability assessment procedure is to determine whether the computer-based frames provide an overconfident, an underconfident (diffident), or an accurate (i.e., reliable) estimate of the rate of disease in the test population. Overconfidence refers to the tendency to assign probabilities too high to relatively likely diseases and probabilities too low to relatively unlikely diseases. Underconfidence refers to the opposite tendency, namely, the tendency to assign probabilities too low to relatively likely diseases and probabilities too high to relatively unlikely diseases. An overconfident physician would constantly conclude that his patients had specific diseases when there was in fact insufficient evidence. A diffident physician would continue to require additional testing after sufficient information was present to conclude a diagnosis. Reliability is the ability to assign to the diseases in a differential diagnostic list probabilities consistent with the evidence available to support them.

Hilden defines 10 reliability statistics, arbitrarily denoted as Q1 through Q10. He divides them into two groups. Q1 through Q5 measure the diagnostic reliability of a probabilistic expert system over a continuous scale of diagnostic certainty from 0% to 100%. Q6 through Q10 are analogous to Q1 through Q5, but measure diagnostic reliability when the diagnosis is made in a dichotomous fashion.

Q1 is the actual mean probability (summed over all patients in the test population) that the computer-based frame has assigned to the real diagnosis for each patient. In the present study, the real diagnosis is defined as the ICD-9 discharge code assigned by the medical staff. Q2 is defined as the expected mean probability of the diagnosis made by the system. It is derived from the probabilities assigned to all of the diseases in all of the patients for whom the system has been run.

The difference (Q1 minus Q2) between actual and expected mean diagnostic probabilities, called Q3, reflects the discrepancy between the computer's average estimate of the probability of the disease and the actual estimate of the probability of disease in the test population. If the expected mean value (Q2), based on the system's behavior over all of the possible diseases, is higher than the actual mean value (Q1), then the system is overconfident. Alternatively, if the expected mean is lower than the mean of the actual, then the system is diffident. Finally, if Q1 is not significantly different from Q2, then the system provides reliable estimates.

Apart from random fluctuations, Q3 averages zero for perfectly reliable systems. Q3 can be conceptualized as a statistic sampled from a normal distribution. Q3 can be converted into a standard score so that the value can be compared to a standard normal distribution. The statistic, Q4, is simply the standard deviation for the distribution of Q3. When Q3 is divided by Q4, the resulting value, called Q5, can be treated as a standard score (or Z-score) from a standard normal distribution. 95% of sample values of Q3 from a perfectly reliable system should be within 1.96 (2 standard deviation units) from zero. If the absolute value of a sample of Q5 is greater than 1.96, then one must reject the null hypothesis that the computer produces reliable decisions.

Hilden gives an example demonstrating the interpretation of negative and positive values of Q3. Let us suppose the system sometimes unwarrantedly stakes a 100% certainty on pneumonia in certain patients. We'll examine the effects of this decision on Q3 when the patient actually does or actually does not have the disease. Whether the patient has pneumonia or not, that patient's contribution to the Q2 score would always be $(1.0^2 + 0^2) = 1.0$. Now in the patient who really has pneumonia, that patient's contribution to Q1 would be 1.0. In this case the net Q3 is zero (1.0 - 1.0 = 0). The system has behaved reliably. But in the patient without pneumonia the expert system was mistakenly overconfident in assigning a 100% certainty of disease. In this patient the Q1 contribution would be zero. Because Q2 is still 1.0 the net contribution to Q3 (Q1 - Q2) for the non-diseased patient is (0 - 1.0) = -1.0. This demonstrates that mistakenly overconfident systems tend to make Q3 negative. A similar analysis can be used to demonstrate that Q3 tends to be positive for underconfident (diffident) systems.

The Q6 through Q10 statistics are directly analogous to Q1 through Q5 except that the Q6 through Q10 statistics compare non-error rates (NERs). Each patient is assigned a discrete diagnosis (present/absent), rather than a probabilistic diagnosis (like P[chronic bronchitis] = 0.8). The diagnosis assigned is the one with the highest probability.

Q6 is the actual NER (the frequency with which the system has assigned the patient's real diagnosis), Q7 is the expected NER (assuming the null hypothesis of perfect reliability), and Q8 is the difference between actual and expected NERs. Hilden has shown that Q8, like Q3, will be negative in overconfident systems, positive in diffident systems, and zero (apart from random fluctuations) in perfectly reliable systems. In similar fashion, Q9 is the standard deviation of Q8, and Q10 is the number of standard deviations Q8 varies from the mean. Hilden has demonstrated that if the absolute value of Q10 is greater than 1.96 (2 standard deviation units), one must reject the null hypothesis of system reliability. Like Q5, Q10 is also positive in underconfident systems and negative in overconfident systems.

This description makes it clear that Q5 and Q10 are the "key" statistics. They indicate both the direction of the unreliable tendency (positive or negative; underconfident or overconfident) and the magnitude of the unreliability. If Q5 and Q10 exceed an absolute value of 1.96, then the system is significantly unreliable. We hypothesized that the Q5 and Q10 statistics derived from our unclustered diagnostic frames would be significantly negative, denoting overconfidence. We also hypothesized that clustering these same frames would reduce overconfidence and bring the values for Q5 and Q10 within the limits of plus or minus 1.96.

Table 1

The statistics of four Pulmonary diseases

	Pneumonia		Chronic Bronchitis			Embolism		Asthma		
	Unclustered	Clustered	Unclustered	First Clustered	Second Clustered	Unclustered	Clustered	Unclustered	First Clustered	Second Clustered
Q1	0.913	0.831	0.881	0.912	0.920	0.957	0.921	0.938	0.924	0.889
Q2	0.926	0.837	0.934	0.927	0.930	0.956	0.910	0.947	0.938	0.896
Q3	-0.013	-0.006	-0.053	-0.015	-0.011	-0.010	0.010	-0.009	-0.014	-0.007
Q4	0.005	0.007	0.005	0.006	0.006	0.004	0.006	0.004	0.006	0.006
Q5*	**-2.54**	**-0.76**	**-10.6**	**-2.70**	**-1.85**	**-2.53**	**1.60**	**-2.16**	**-2.59**	**-1.21**
Q6	0.932	0.882	0.901	0.936	0.942	0.967	0.954	0.965	0.946	0.919
Q7	0.950	0.885	0.951	0.950	0.950	0.979	0.943	0.960	0.959	0.928
Q8	-0.018	-0.003	-0.049	-0.014	-0.012	-0.012	0.011	0.005	-0.013	-0.009
Q9	0.009	0.013	0.008	0.008	0.008	0.006	0.009	0.007	0.008	0.010
Q10**	**-2.07**	**-0.26**	**-6.17**	**-1.67**	**-1.51**	**-2.17**	**1.16**	**0.66**	**-1.73**	**-0.89**

Results

Table 1 summarizes the results of the reliability analysis for the *Bacterial Pneumonia, Chronic Bronchitis, Pulmonary Embolism,* and *Asthma* frames. Q5 is the summary statistic for the reliability analysis of the continuous (0 to 100%) diagnostic probabilities. Q10 is the summary statistic for the reliability analysis of non-error rates using 0.5 as threshold required to conclude a diagnosis. Because some information is always lost by classifying into two bins rather than by a continuous distribution,[9] the Q5 and Q10 reliability statistics sometimes diverge. For instance, the unclustered *Asthma* frame looks reliable according to Q10 (+0.66) but the Q5 statistic (-2.16) indicates significant unreliability. We do not assume reliable behavior unless the absolute values of both Q5 and Q10 are less than 1.96.

For each unclustered frame, the valve of Q5 denoted significant unreliability due to overconfidence. The Q10 statistic for unclustered Bayesian frames denoted similar significant overconfidence in each case excepting *Asthma*. In the cases of *Pulmonary Embolism* and *Bacteria Pneumonia*, the initial round of clustering removed all statistically significant overconfidence (Table 1). But in the cases of *Chronic Bronchitis* and *Asthma*, the initial round of clustering did not produce reliable behavior. In *Asthma*, frame reliability actually deteriorated with clustering. A second round of clustering corrected the mistakes and eliminated the overconfident tendencies (see third column, table 1).

Discussion

That data clearly show that the clustered frames exhibited significantly less diagnostic overconfidence than corresponding non-clustered frames. In two cases, *Pulmonary Embolus* and *Bacterial Pneumonia*, it was possible to achieve complete reliability after a single round of clustering. In contrast, *Chronic Bronchitis* and *Asthma* remained overconfident after a single round of clustering. In the case of *Asthma*, the overconfidence was actually worse. We suspected that the *Asthma* and *Chronic Bronchitis* frames contained residual conditional dependence between findings that had been initially overlooked.

As we re-examined the *Chronic Bronchitis* frame, we found several errors that had been overlooked. The most obvious error was using the finding "cough" in two clusters, *Chronic Cough* and *Increased Airways Secretions*. Because "cough" was, in effect, counted double in any coughing patient, the diagnosis of *Chronic Bronchitis* tended to be overconfident. We combined the *Chronic Cough* and the *Increased Airways Secretions* clusters into a new cluster, *Signs of Airway Inflammation*, that only used cough once. Several smaller errors were also fixed. The re-clustered *Chronic Bronchitis* frame had a Q5 value of -1.85, indicating that reclustering had produced acceptably reliable behavior.

We found two types of problems in the *Asthma* frame. The first problem was that the Boolean logic for producing a "denied" result in the cluster *Generalized Airways Obstruction* could never come true. Hence, every patient in the test population was diagnosed as having at least "suggested" *Generalized Airways Obstruction*. We found two instances of the second type of problem, which was failing to appropriately include conditionally non-independent findings in clusters. In each case we simply placed these findings in the appropriate clusters, where they should have been in the first place. The revised *Asthma* frame had a value for Q5 of -1.21, indicating reliable performance.

We were concerned that "over-clustering" might produce underconfident frames. In this experiment we did not discover significant underconfidence (Q5 or Q10 > +1.96) in any of the clustered frames. However, this does not preclude discovery of underconfident frames in the future. Fortunately, reliability analysis is a powerful tool to detect both overconfident and underconfident frames. Reliability analysis was able to detect two unreliable clustered frames, *Asthma* and *Chronic Bronchitis*, that we had not suspected were faulty. We propose that reliability analysis based an actual patient test populations be routinely used in the development of probabilistic decision frames so that reliable behavior can be guaranteed.

The clustered knowledge model offers major advantages to a frame-based, Bayesian decision system like Iliad. The first, and most important, advantage is that clusters increase diagnostic reliability, as clearly indicated by our data. Reliability is particularly important when a Bayesian decision system like Iliad is used for teaching. Students instinctively reject obviously inaccurate consultations and will lose interest in the system if

* If $|Q5| > 1.96$ then the continuous estimated probability is statistically significant unreliability; exists ($p < 0.05$).

** If $|Q10| > 1.96$ then the dichotomous estimated probability is statistically significant unreliability; exists ($p < 0.05$).

inaccurate consultations are commonly encountered. In Iliad's simulation mode, students are graded on how closely their diagnostic strategy and differential diagnosis match that of Iliad's. These grades cannot be fair unless Iliad can produce reliable diagnoses against which the students' diagnoses may be compared.

Clusters provide at least four additional advantages. This experiment was not designed to test these other advantages, but we will enumerate them and briefly discuss their importance. First, clusters modularize the knowledge base. By acting as common "subroutines" in more than one Bayesian frame, clusters save substantial amounts of time during knowledge development. Second, clusters explicitly describe pertinent patterns of findings can be used to diagnose pathophysiologic processes. The ability to recognize patterns of findings that constitute a diagnosis is a subtle skill that is quite important but rarely explicitly taught.[10] Third, clusters allow a rich knowledge representation. Sparse frames are unnecessary when the overconfidence resulting from conditional dependence can be avoided. In the consultation mode, a rich knowledge representation allows Iliad to respond appropriately when non-key findings are entered. Rich knowledge representations also promote more realistic simulations. A sparse knowledge base would produce unrealistic simulations consisting only of key findings. Fourth, clusters that reflect physiologic groupings can be used to provide more realistic diagnostic explanations.

Our future work will continue with testing of six additional pairs of pulmonary disease frames. We expect we will find these six frames to be significantly overconfident before clustering. Because the clustered frames used in this particular study were based on sparse frames, we may have provided a conservative test of clustering. We are planning a test of rich, clustered frames against rich, unclustered frames. This test will determine whether the clustered knowledge model is robust enough to defeat strong overconfident tendencies.

Some people may object to the clustered knowledge model in Iliad because cluster development is apparently so subjective. In fact, cluster development need not be subjective. We have already used a mathematical technique, called cluster analysis, to examine other, non-clustered knowledge bases for implicitly contained clusters. In the QMR[11] (Quick Medical Reference) knowledge base, we have found implicit clusters that are quite similar to their independently derived Iliad counterparts.[12] A similar cluster analysis technique will be applied to actual patient data from the HELP database in an attempt to discover whether real patient data is clustered. Thus, we will validate the clustered knowledge model by comparing Iliad clusters, QMR clusters, and clusters derived from actual patient data.

In summary, clusters are a new model of knowledge representation used in the Iliad expert system. Clusters encapsulate conditionally dependent findings and usually describe pathophysiologic entities. By reducing the effects of conditional dependence on Bayesian analysis, clusters increase the reliability of Iliad's decisions. Clustering does not seem to produce underconfident decisions, a side effect we had feared. Reliability analysis can discover unexpected overconfidence or underconfidence in newly developed knowledge frames. Knowledge frames should be debugged by analyzing reliability before being used clinically. Other advantages of clusters include modular knowledge representation, explicit teaching models for pattern recognition skills, and rich knowledge representations. We plan future work to further validate the clustered knowledge model. This work will include examining other expert knowledge bases (like QMR) for the presence of hidden clusters and attempting to cluster real patient data from the HELP database.

[1] Hukill MJ, Ward KM, Haug, PJ, Warner HR. HELP decision support on the Macintosh. Proceedings of the eleventh annual Symposium on Computer Applications in Medical Care. IEEE Computer Society Press 1987:155-157.

[2] Weinstein MC, Fineberg, HV. Clinical decision analyis. 1st ed. W.B. Saunders Company 1980; pgs. 156-158.

[3] Bouhaddou P, Haug PJ, Warner HR. Use of the HELP clinical database to build and test medical knowledge. Proceedings of the eleventh annual Symposium on Computer Applications in Medical Care. IEEE Computer Society Press 1987:64-67.

[4] Haug P, Clayton PD, Shelton P, Rich T, Tocino I, Fredrick PR, Crapo RO, Morrison WJ. Revision of diagnostic logic usig a clinical data base. Proceedings of AAMSI Congress 1987:238-242.

[5] Habbema JDF, Hilden J, Bjeeregaard B. The measurement of performance in probabilistic diagnosis: The problem, descriptive tools, and measures based on classification matrices. Methodik der Information in der Medizin 1978; 17(4):217-226.

[6] Hilden J, Habbema JDF, Bjerregaard B. The measurement of performance in probabilistic diagnosis: Trustworthiness of the exact values of the diagnostic probabilities. Methodik der Information in der Medizin 1978; 17(4):227-237.

[7] Hilden J, Habbema JDF, Bjerregaard B. The measurement of performance in probabilistic diagnosis: Methods based on continuous function of the diagnostic probabilities. Methodik der Information in der Medizin 1978; 17(4):227-237.

[8] Rifkin, RD. Maximum Shannon information content of diagnostic medical testing. Medical Decision Making 1985; 5(2):179-189.

[9] Shapiro, AR. The evaluation of clinical predictions: A method and initial applications. New England Journal of Medicine 1977; 296:1509-1514.

[10] Schwartz S, Griffin T. Medical thinking: The psychology of medical judgment and decision making. 1st ed. Springer-Verlag 1986; 178-179.

[11] QMR and Quick Medical Reference are registered trademarks of the Univerisity of Pittsburg. The QMR knowledge base has also been copyrighted in the years 1986, 1987, 1988 by the University of Pittsburg

[12] Michael Lincoln MD, Charles Turner PhD, Brad Hesse MS, Homer R Warner MD, and Randolph Miller, MD. Discovering Clustered Disease Findings: Prospects for Enhancing Expert Systems. Accepted for the twelfth annual proceedings of the Symposium on Computer Applications in Medical Care.