

Case Based Medical Diagnosis of Occupational Chronic Lung Diseases From Their Symptoms and Signs

Prem Pal Singh Tomar

Faculty of Engineering,
Dayalbagh Educational Institute,
Uttar Pradesh,
Agra, India

singhppst@rediffmail.com

Dr. Ranjit Singh

Faculty of Engineering,
Dayalbagh Educational Institute,
Uttar Pradesh,
Agra, India

rsingh_dei@yahoo.com

Dr. P K Saxena

Faculty of Engineering,
Dayalbagh Educational Institute,
Uttar Pradesh,
Agra, India

premkumarsaxena@gmail.com

Jeetu Sharma

Electronics & Electrical engineering
Anand Engg. College
Agra

jitusharma_19@rediffmail.com

Abstract

The clinical decision support system using the case based reasoning (CBR) methodology of Artificial Intelligence (AI) presents a foundation for a new technology of building intelligent computer aided diagnoses systems. This Technology directly addresses the problems found in the traditional Artificial Intelligence (AI) techniques, e.g. the problems of knowledge acquisition, remembering, robust and maintenance. In this paper, we have used the Case Based Reasoning methodology to develop a clinical decision support system prototype for supporting diagnosis of occupational lung diseases. 127 cases were collected for 14 occupational chronic lung diseases, which contains 26 symptoms. After removing the duplicated cases from the database, the system has trained set of 47 cases for Indian Lung patients. Statistical analysis has been done to determine the importance values of the case features. The retrieval strategy using nearest-neighbor approaches is investigated. The results indicate that the nearest neighbor approach has shown the encouraging outcome, used as retrieval strategy. A Consultant Pathologist's interpretation was used to evaluate the system. Results for Sensitivity, Specificity, Positive Prediction Value and the Negative Prediction Value are 95.3%, 92.7%, 98.6% and 81.2% respectively. Thus, the result showed that the system is capable of assisting an inexperience pathologist in making accurate, consistent and timely diagnoses, also in the study of diagnostic protocol, education, self-assessment, and quality control. In this paper, clinical decision support system prototype is developed for supporting diagnosis of occupational lung diseases from their symptoms and signs through employing Microsoft Visual Basic .NET 2005 along with Microsoft SQL server 2005 environment with the advantage of Object Oriented Programming technology

Key words: Clinical Support System, Artificial Intelligence, Case-Based Reasoning, Pathologist

1. INTRODUCTION

The use of artificial Intelligence (AI) technique i.e case-based reasoning (CBR), in the the development of Clinical Support System has a relatively young history, arose out of the research in cognitive science. The earliest contributions in this area were from Roger Schank and his colleagues at Yale University [1],[2]. During the period 1977–1993, CBR research was highly regarded as a plausible high-level model for cognitive processing. It was focused on problems such as how people learn a new skill and how humans generate hypotheses about new situations e cognitive-based researches were to construct decision based on their past experiences. Many prototype of decision support system based on CBR technique were built during this period: for example, Cyrus [3],[4], Mediator [5], Persuader [6], Chef [7], Julia [8], Casey, and Protos [9]. Three CBR workshops were organized in 1988, 1989, and 1991 by the U.S. Defense Advanced Research Projects Agency (DARPA). These formally marked the birth of the discipline of Decision Support System using case-based reasoning.

Computerized evidence-based guidelines and Clinical decision support systems (CDSS) have been promoted as the key to improving effectiveness and efficiency of clinical decisions [14]. Although the use of decision support systems (DSS) in the field of medicine has accelerated in recent years [15]. Many researchers are working on Clinical Support System using CBR with many diverse applications ranging from psychiatry and epidemiology to clinical diagnosis. Most of them aim for a successful implementation of CBR methods to enhance the work of health experts to improve the efficiency and quality of health care. Researchers who have contributed substantially to CBR in medicine include Gierl Schmidt and their colleagues who focused on a range of applications including children dysmorphic syndromes, antibiotics therapy advising for intensive care and monitoring emerging diseases (Gierl, 1993 Schmidt & Gierl, 2001) Notable is their ICONS system Gierl, 1993),first applied to the determination of antibiotic therapy treatment for intensive care then to the prognosis of kidney function defects, For this latter application ICONS learned prototypes associated with graded levels of severity through temporal abstraction Gierl, 1993),and matched new cases with these prototypes to predict the severity of a renal disease [16].

Some real Clinical Support Systems based on CBR technique are: CASEY that gives a diagnosis for the heart disorders [10], GS.52 which is a diagnostic support system for dysmorphic syndromes, NIMON is a renal function monitoring system, COSYL that gives a consultation for a liver transplanted patient [11] and ICONS that presents a suitable calculated antibiotics therapy advise for intensive care patients [12]. Computerized evidence-based guidelines and Clinical decision support systems (CDSS) have been promoted as the key to improving effectiveness and efficiency of clinical decisions [17]. Although the use of decision support systems (DSS) in the field of medicine has accelerated in recent years [18].

However, none of the aforementioned studies presented results that showed evidence of first, the inclusion of all the 14 occupational lung diseases perspective; and secondly a system capable of assisting a Pathologist who is not specialized in the aspect of occupational lung diseases diagnosis. Thus, this system proposes a medical decision support system for diagnosis of occupational lung diseases as an improvement of earlier works.

2. JUSTIFICATION FOR STUDY

In clinical practice, making decision involves a careful analysis of harms and benefits associated with different treatment options. These decisions, often associated with high stake and important long term consequences, are frequently made in presence of limited resources and information and an incomplete clinical picture. Under such circumstances, a rigorous and objective analysis of outcomes and probabilities is essential to achieve the best possible decision given a specific clinical situation.

Therefore, Pathologist is required to be fully conversant with the diversity of possible patterns, recognize and diagnose them, timely and accurately. Hence, a Pathologist who is not a specialist in the pathology of the occupational lung diseases has to refer to textbooks and study past diagnosis before concrete diagnosis can be made and conclusion reached. Hence, there is the need for a system, which can assist the Pathologist to reach timely and accurate decision.

3. KNOWLEDGE ENGINEERING TASKS IN DEVELOPING CBR BASED SYSTEM

The problem-solving life cycle in a CBR system consists essentially of the following :

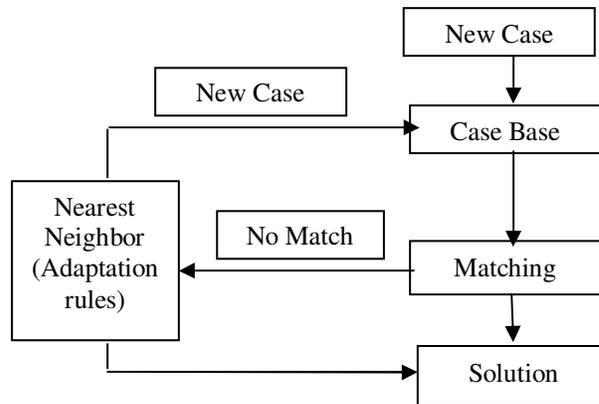


FIGURE 1: Case Base Reasoning Technique

The figure shows that when a new case comes to the system, the system carries out the work of matching. Upon getting exact match same result is displayed while in case no exact match is found the nearest neighbor is looked for whose result is adjusted according to the new case using the adaptation rules and the result is displayed. Such a new case is also saved in the case base for future assistance. Accordingly the methodology of developing Clinical Support System, CBR-based systems in specific domain can be summarized in the following steps:

1. Retrieving :The system will search its Case-Memory for an existing case that matches the input problem specification.
2. Reusing :If we are lucky (our luck increases as we add new cases to the system), we will find a case that Exactly matches the input problem and goes directly to a solution.
3. If we are not lucky, we will retrieve a case that is similar to our input situation but not entirely appropriate to provide as a completed solution.
4. Revising: The system must find and modify small portions of the retrieved case that do not meet the input specification. This process is called "case-adaptation".
5. Retaining :The result of case adaptation process is (a) Completed solution, and (b) generates a new case that can be automatically added to the system's case- memory for future use

4. KNOWLEDGE ACQUISITION

Knowledge acquisition is a process of acquiring, organizing and studying knowledge for the lung diseases. The data and knowledge of Clinical Support System based on Case-Base technique are collected from different sources. The first primary source is, acquired from a physician (Domain Expert). The second source is from specialized databases like lung disease diagnostic laboratory at Agra, Uttar Pradesh, India, books and a few electronic websites. This knowledge can be divided by important fact into 26 facts, which are shown on table 1.

No.	Chief symptoms	No.	Chief symptoms
1.	Cough	14	Alcohol use
2.	Dyspnea	15	Heart rhythm problem
3.	Chest Discomfort	16	Abdominal pain
4.	Malaise	17	Shoulder pain
5.	Fever	18	difficulty in swallowing
6.	Wheezing	19	Pain under rib cage
7.	Hemoptysis	20	Chemicals exposure
8.	Persistent cough	21	Fungi exposure
9.	Fever with chill	22	humidifiers exposure
10	Night sweat	23	Coke oven emissions
11.	Asbestosis exposure	24	Silica exposure
12.	Excessive sweating	25	Coal dust
13.	Smoking	26	Cotton Dust

TABLE 1: Fact List of Symptoms

5. ASSIGNING IMPORTANCE VALUE TO CASE SYMPTOMS

Features weights for most problem domains are context dependent. The weight assigned to each feature of the case tells how much attention to pay to matches and mismatches in the field when computing the distance measure of a case. Those that are good predictors are then assigned higher importance for matching [10].

The importance of the feature depends upon its prevalence among the diseases. If a feature is common among all diseases like Cough, then it will have the least importance in leading to a diagnosis.

6. CASE INDEXING AND RETRIEVAL

Here, we focus our discussion on case indexing and retrieval strategy. Case indexing and retrieval are two separate but closely related processes. Since a case memory may contain thousands of cases, case indices organize their key features to expedite the search process. Case retrieval searches the case base to find candidate cases that share significant features with the new case. Existing literature in case-based reasoning has proposed several mechanisms for case indexing and retrieval. A good review of early literature can be found in [13].

7. RETRIEVAL USING NEAREST-NEIGHBOR TECHNIQUE

If however an exact match is not found, which can be the case many times, nearest neighbor technique (ref. table 2) and adaptation rules have to be used. Let T is new case and C1 and C2 are old cases then Nearest neighbor formula = sum of (weight*similarity)/sum of weight

$$(T, C1)=72/97=0.74$$

$$(T, C2)=65/97=0.67$$

So, C1 is the nearest neighbor.

Then the presence of the symptoms in the new and the old case is listed in the next two columns. Local similarity is given in Clinical Decision support system. The total of all the weights is calculated by adding them which is 97. Then the

sum of weight*similarity is calculated by adding all the products of weight*similarity. In the first comparison the sum is 72 while in the second comparison it is 65. The sum of weights in the first comparison is: 97

The nearest neighbor value is: $72/97 = 0.74$

In the second comparison:

The sum of weights is: 65

The sum of weights in the first comparison is: 67

The nearest neighbor value is: $38/67 = 0.57$

Therefore, the first comparison, which is case C1, is the nearest neighbor for the new case T.

The system will use the result of the nearest match found and use adaptation rules to 'revise' this result according to the demands of the novel situation. The system uses the Nearest-neighbor algorithm that finds the closest matches of the cases already stored in the database to the new case using a distance calculation, which determines how similar two cases are by comparing their features, the pseudo code of this algorithm [10] can be written as follows:

For each feature in the input case:

Find the corresponding feature in the stored case

Compare the two values to each other and compute the degree of match

Multiply by a coefficient representing the importance of the feature to the match

Add the results to derive an average match score

This number represents the degree of match of the old case to the input.

A case can be chosen by choosing the item with the largest score.

Nearest-neighbor techniques applied to the retrieval phase of a CBR system (i.e., measuring similarity among cases). The equation

$$\text{Similarity (T, S)} = \frac{\sum (T_i S_i) \times W_i}{\sum W_i}$$

represents a typical nearest-neighbor technique that describes a situation for which T(Target case) and S (Source case) are two cases compared for similarity, n is the number of attributes in each case, i is an individual attribute from 1 to n, and W_i is the feature weight of attribute. Similarities are usually normalized to fall within the range 0 to 1, where 1 means a perfect match and 0 indicates a total mismatch.

8. DEVELOPMENT AND RESULTS

Development of clinical decision support system prototype is through employing Microsoft Visual Basic .NET 2005 environment with the advantage of Object Oriented Programming technology. The Microsoft SQL server 2005 was used to develop the database module.

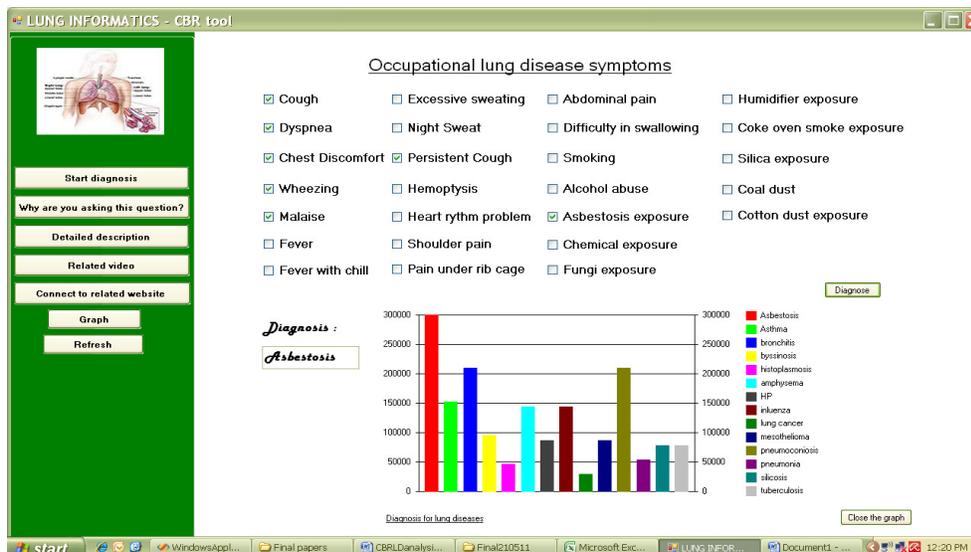


FIGURE 2: Diagnostic window

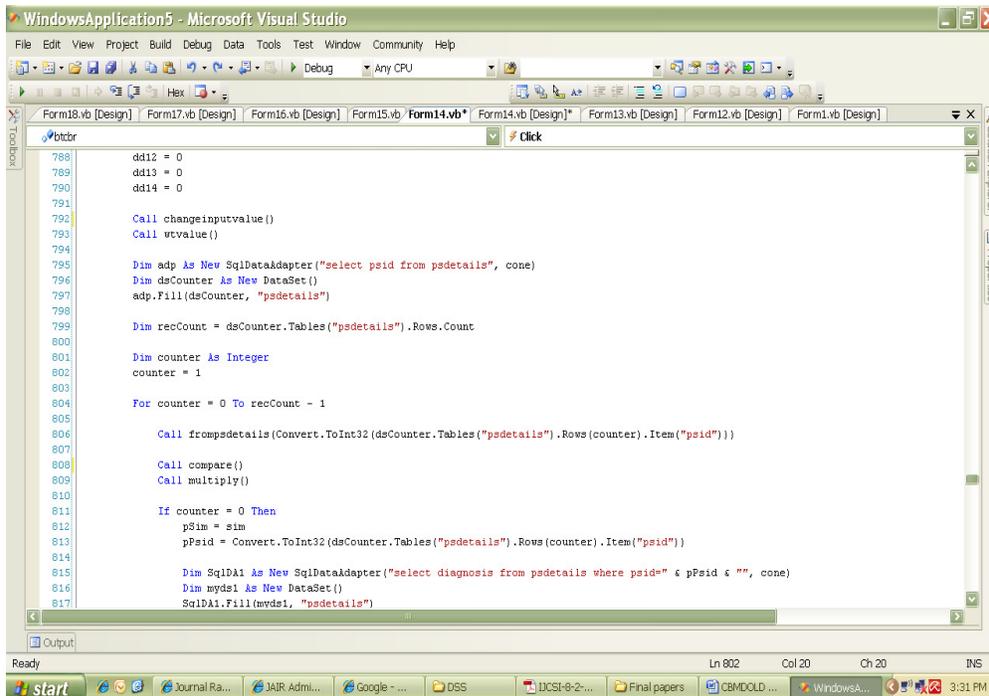


FIGURE 3 : Code window

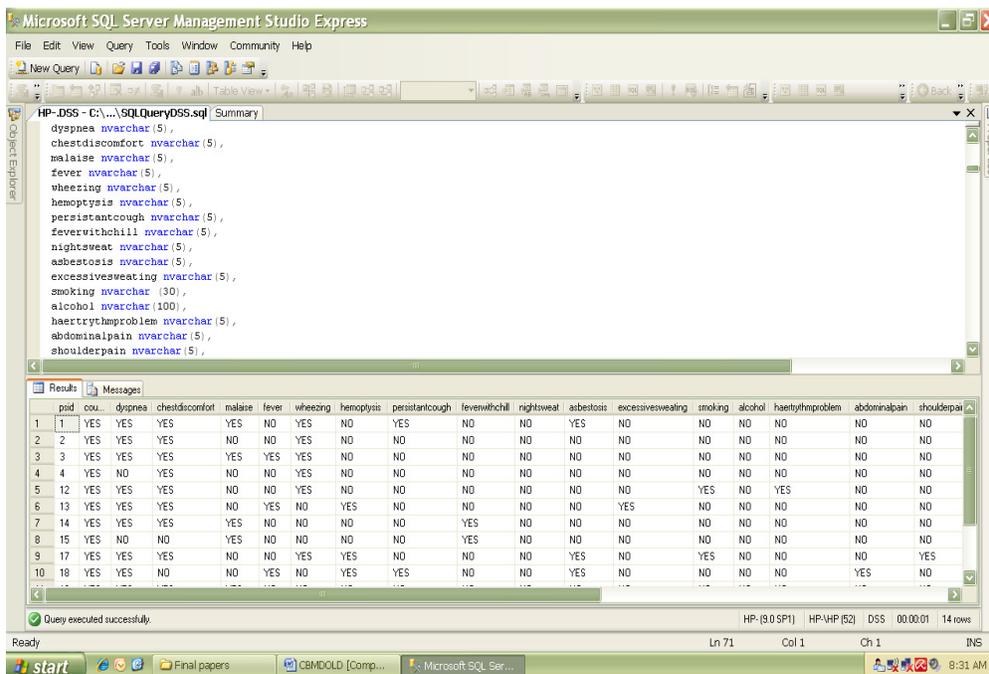


FIGURE 4: SQL server window

In this paper, the architecture, and the implementation of a prototype of a Clinical Support System using case-based technique that supports diagnosis of occupational lung diseases was developed. Knowledge structure was represented via a formalism of cases. The system used nearest-neighbor techniques technique for the retrieval process.

Using a Consultant Pathologist's interpretation as a "gold standard" (reference test), the system's parameters for diagnosing occupational lung diseases were calculated.

(i) True positive (TP):

The diagnostic system yields positive test result for the sample and thus the sample actually has the disease;

(ii) False positive (FP):

The diagnostic system yields positive test result for the sample but the sample does not actually have the disease;

(iii) True negative (TN):

The diagnostic system yields negative test result for the sample and the sample does not actually have the disease; and

(iv) False negative (FN):

The diagnostic system yields negative test result for the sample but the sample actually has the disease.

The formulas for used for calculating Sensitivity, Specificity, PPV and NPV are:

$$\text{Sensitivity} = [\text{TP}/(\text{TP}+\text{FN})] \times 100\% \dots\dots\dots (1)$$

$$\text{Specificity} = [\text{TN}/(\text{TN}+\text{FP})] \times 100\% \dots\dots\dots (2)$$

$$\text{PPV} = [\text{TP}/(\text{TP}+\text{FP})] \times 100\% \dots\dots\dots (3)$$

$$\text{NPV} = [\text{TN}/(\text{TN}+\text{FN})] \times 100\% \dots\dots\dots (4)$$

Using equations (1), (2), (3) and (4), respectively, the Sensitivity, Specificity, Positive Prediction Value (PPV) and the Negative Prediction Value of the system are:

Sensitivity = 95.3%;

Specificity = 92.7%;

PPV = 98.6%

NPV = 81.2%.

8. SUMMARY AND CONCLUSION

In this paper we presented a clinical support system, which could be used by stakeholders for arriving at very vital decisions regarding the diagnosis of 14 occupational chronic lung diseases. The focus was on the development of a clinical support system that can assist Pathologist, especially those who may not be specialist in the area of occupational chronic lung diseases treatment. Thus, the system attempts to improve the effectiveness of diagnosis (in relation to accuracy, timeliness and quality) that is performed by a human pathologist, rather than improve their efficiency with respect to decision making. Therefore, the diagnoses made by the system are at least as good as those made by a human expert.

From the development and analysis of Clinical Support System, it is evident that CBR technique of Artificial Intelligence (AI) is appropriate methodology for all medical domains and tasks for the following reasons: cognitive adequateness, explicit experience and subjective knowledge, automatic acquisition of subjective knowledge, and system integration. CBR technique presents an essential technology of building intelligent Clinical Support System for medical diagnoses that can aid significantly in improving the decision making of the physicians. the

proposed method gives an Sensitivity = 95.3%; and PPV = 98.6% which is better than the existing methods. Future research involves more intensive testing using a larger occupational chronic lung disease database to get more accurate results.

The use of Microsoft Visual Basic .NET 2005 along with Microsoft SQL server 2005 as database is found to be very effective in producing the system under windows environment. For future work, more cases will be added to the case memory.

REFERENCES

- [1] R. Schank and R. Abelson (eds.), *Scripts, Plans, Goals and Understanding*, Lawrence Erlbaum Associates, Hillsdale, NJ, 1977
- [2] R. Schank, *Dynamic Memory: A Theory of Reminding and Learning in Computers and People*, Cambridge University Press, New York, 1982
- [3] J. L. Kolodner, Maintaining organization in a dynamic long term memory, *Cognitive Science*, vol. 7, pp. 243–280, 1983.
- [4] J. L. Kolodner, Reconstructive memory, a computer model, *Cognitive Science*, vol. 7, pp. 281–328, 1983.
- [5] R. L. Simpson, A computer model of case-based reasoning in problem solving: an investigation in the domain of dispute mediation, Ph.D. dissertation, School of Information and Computer Science, Georgia Institute of Technology, Atlanta, GA, 1985.
- [6] K. Sycara, Using case-based reasoning for plan adaptation and repair, in *Proceedings of the Case-Based Reasoning Workshop*, DARPA, Clearwater Beach, FL, Morgan Kaufmann, San Francisco, pp. 425–434, 1988.
- [7] K. J. Hammond, *Case-Based Planning*, Academic Press, San Diego, CA, 1989.
- [8] T. R. Hinrihs, *Problem Solving in Open Worlds*, Lawrence Erlbaum Associates, Hillsdale, NJ, 1992.
- [9] R. Bareiss, *Exemplar-Based Knowledge Acquisition: A Unified Approach to Concept Representation, Classification, and Learning*, Academic Press, San Diego, CA, 1989.
- [10] J. L. Kolodner (Ed.), *Case-Based Reasoning*, Morgan Kaufmann Publishers: California, 1993
- [11] M. Lenz, S. Wess, H. Burkhard and B. Bartsch, *Case-Based Reasoning Technology: From Foundations to applications*, Springer, 1998.
- [12] B. Heindl. Et al., A Case-Based Consiliarius for Therapy Recommendation (ICONS): Computer-Based Advice for calculated Antibiotic Therapy in Intensive Care Medicine. *Computer Methods and Programs in Biomedicine* 52 , pp. 117-127, 1997.
- [13] E. L. Rissland, J. Kolodner, & D. Waltz, Case-based reasoning from DARBA: “Machine learning program plan”. *Proceedings of the case-Based Reasoning Workshop*. San Mateo, CA: Morgan Kaufmann, pp. 1-13, 1998.
- [14] Fieschi M, Dufour J-C, Staccini P, et al. Medical decision support systems: old dilemmas and new paradigms? Tracks for successful integration and implementation. *Methods Inform Med* 2003; 42:190–198.8
- [15] Decision support systems for antibiotic prescribing, Vitali Sintchenko, b, c, Enrico Coierac and Gwendolyn L. Gilberta, b , *Current Opinion in Infectious Diseases* 2008, 21:573–579
- [16] Medical applications in case based reasoning , ALEC HOLT, ISABELLE BICHINDARITZ, RAINER SCHMIDT, and PETRA PERNER , *The Knowledge Engineering review*, Vol.20:3, 289-292 – 2006, Cambridge University Press

- [17] Graham B, Detsky AS, the application of decision analysis to the surgical treatment of early osteoarthritis of wrist. J Bone Joint Surg Br 2001;83:650-4
- [18] Detsky AS, Nalie G, Krahn MD, Naimark D, Redelmeier DA. Primer on medical decision analysis : Part 1- Getting started. Med Decision Making 1997 : 17:123-5