

NEONATE - A PROTOTYPE OF A COMPREHENSIVE EXPERT SYSTEM
EMBEDDED INSIDE A HOSPITAL INFORMATION SYSTEM

by

Arie Franco

A dissertation submitted to the faculty of
The University of Utah
In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Medical Informatics

The University of Utah

December 1990

Copyright © Arie Franco 1990

All Rights Reserved


THE UNIVERSITY OF UTAH GRADUATE SCHOOL

SUPERVISORY COMMITTEE APPROVAL

of a dissertation submitted by

Arie Franco

This dissertation has been read by each member of the following supervisory committee and by majority vote has been found to be satisfactory.









J. Haug





Fred L. Farr









Homer R. Warner

THE UNIVERSITY OF UTAH GRADUATE SCHOOL

FINAL READING APPROVAL

To the Graduate Council of the University of Utah:

I have read the dissertation of Arie Franco in its final form and have found that (1) its format, citations and bibliographic style are consistent and acceptable; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the supervisory committee and is ready for submission to The Graduate School.



Approved for the Major Department



Homer R. Warner
Chair/Dean

Approved for the Graduate Council



B. Gale Dick
Dean of The Graduate School

ABSTRACT

NEONATE is a prototype of an expert system for the Newborn Intensive Care Unit developed at the Primary Children's Medical Center in Salt Lake City, Utah. The system is designed to be embedded in the HELP Hospital Information System and as such has all the characteristics of a HIS (data acquisition, data analysis and data reporting). The decision module of NEONATE is designed to produce an admission problem list and to notify physicians if tests that prevent the system from confirming or eliminating diagnostic hypotheses are missing. The knowledge base mimics the cognitive model to make the system as close as possible to the physicians' methods to enhance the acceptance of the system in future.

The system currently in use at the Newborn Intensive Care Unit of the Primary Children's Medical Center is the CETUS-100 system that has data acquisition and data reporting capabilities, but lacks decision features. The physician problem list that appears in the CETUS report is typed manually into the system by the clerks. The admission problem list that NEONATE generates was compared to the physician admission problem list in the CETUS report. Measuring the agreement between both problem lists using Kappa statistic showed Kappa of 0.662. NEONATE was found to pick up more true positive problems than the physician. The mean of the true positive problems of NEONATE was 2.733 while the mean of the physician was 1.867 ($p < 0.01$). The mean of the false positive problems of NEONATE was 0.9 while the mean of the physician was 0.3 ($p < 0.01$). NEONATE sensitivity was

82.82% while the physician sensitivity was 53.77% ($p < 0.01$). NEONATE specificity was 97.19% while the physician specificity was 99.07% ($p < 0.01$).

The radiological frames, which are used by the Inference Engine to interpret radiological findings of newborn chest x-ray films, were used to prove that NEONATE helps physicians who are not trained in radiology to reach a better diagnosis. A system that compares the residents' problem list and the computer's problem list to a gold standard problem list was developed. It shows that the computer helps the first and the second year residents but not the third year residents.

NEONATE proved to be a potentially excellent prompting device. When physicians were prompted more true positive findings and problems were collected at the expense of more false positives.

Although NEONATE proved to have advantages over the present CETUS system and to be a system from which health care practitioners might benefit, further trials are necessary to establish its usability in real clinical settings.

This dissertation is dedicated to my parents, my wife Emma,
and my daughters Liran and Eynat.

TABLE OF CONTENTS

ABSTRACT.....	iv
LIST OF TABLES	x
LIST OF FIGURES.....	xi
ACKNOWLEDGMENTS.....	xii
Chapter	
I. INTRODUCTION.....	1
General Remarks on Expert Systems.....	1
Need for Expert Systems in Medicine.....	2
Categories of Application of Medical Expert Systems.....	4
Background - Review of Previous Diagnostic Expert Systems	5
Background - Computer Medical Systems for Newborn Intensive Care Unit.....	11
Acceptance of Computer Systems.....	13
Attributes Necessary for a Successful Expert System	15
Objectives of the Present Dissertation Work	17
II. THE DEVELOPMENT OF THE NEONATE SYSTEM.....	19
The NEONATE System	19
Tools for NEONATE Development.....	20
Developmental Rationale.....	21
The User Interface	21
The Data Base	23
The Knowledge Base.....	24
The Health Care Provider Component.....	32
The Report Generator.....	33

III. COMPARISON OF THE NEONATE TO THE CETUS ADMISSION REPORT	34
General Remarks	34
Comparison of the Completeness	34
Comparison of the Problem Lists	36
Methods	36
Results	43
Discussion	46
IV. EVALUATION OF THE RADIOLOGICAL FRAMES OF THE KNOWLEDGE BASE	51
General Remarks	51
Methods	52
Results	57
Discussion	64
V. EVALUATION OF NEONATE AS A POTENTIAL PROMPTING DEVICE FOR MEDICAL FINDINGS AND PROBLEMS.....	66
General Remarks	66
Methods	66
Results	68
Discussion	77
VI. CONCLUSIONS REGARDING NEONATE'S POTENTIAL AS A PRACTICAL SYSTEM.....	78
 Appendices	
A. NUMBER OF PROBLEMS THAT EXIST IN NEONATE PROBLEM LIST AND IN THE GOLD STANDARD LIST, BUT NOT IN THE PHYSICIAN PROBLEM LIST IN EACH OF THE 30 RECORDS.....	82
B. NUMBER OF TRUE POSITIVE, FALSE POSITIVE, FALSE NEGATIVE, TRUE NEGATIVE PROBLEMS AND THE SENSITIVITY AND THE SPECIFICITY VALUES OF NEONATE IN EACH OF THE 30 RECORDS	84

C.	NUMBER OF TRUE POSITIVE, FALSE POSITIVE, FALSE NEGATIVE, TRUE NEGATIVE PROBLEMS AND THE SENSITIVITY AND THE SPECIFICITY VALUES OF THE PHYSICIAN PROBLEM LIST IN EACH OF THE 30 RECORDS.....	86
D.	COMPARISON OF THE NUMBER OF TRUE POSITIVES BETWEEN NEONATE PROBLEM LIST AND THE PHYSICIAN PROBLEM LIST BY PROBLEM.....	88
E.	COMPARISON OF THE NUMBER OF FALSE POSITIVES BETWEEN NEONATE PROBLEM LIST AND THE PHYSICIAN PROBLEM LIST BY PROBLEM.....	90
	REFERENCES	92

LIST OF TABLES

Table	Page
1. Kappa Values of Fifty Records	44
2. Means of True Positive, False Positive, True Negative and False Negative Values of NEONATE Problem List and the Physician Problem List	45
3. Sensitivities and Specificities of NEONATE Problem List and the Physician Problem List	47
4. Matching Criteria for Attributes	56
5. Variation of Distance by Level of Residency	60
6. Variation of Number of Attributes that do not Match by Level of Residency	61
7. Mean of Number of Findings Identified by Level of Residency.....	69
8. Mean of Number of Findings Identified by Chest X-Ray Film	71
9. Mean of Number of Problems Identified by Level of Residency.....	73
10. Mean of Number of Problems Identified by Chest X-Ray Film	75

LIST OF FIGURES

Figure	Page
1. The Components of NEONATE.....	22
2. The Knowledge Base.....	27
3. List of Problems that are Entered into the Problem List by the Initiator if Tachypnea Exists.....	28
4. Inference Engine Frames.....	30
5. Mean of Kappa Agreement of the Ten Subgroups of Two Radiologists as Compared to the Mean of the Agreement Between Each Radiologist and the Computer's Output With His or Her Findings.....	58
6. Variation of Distance by the Level of Residency Without and With the Computer.....	62
7. Variation of the Number of Attributes that do not Match by the Level of Residency.....	63
8. Mean of Number of Findings Identified by Level of Residency.....	70
9. Mean of Number of Findings Identified by Chest X-Ray Film.....	72
10. Mean of Number of Problems Identified by Level of Residency.....	74
11. Mean of Number of Problems Identified by Chest X-Ray Film.....	76

ACKNOWLEDGMENTS

The author is grateful to the many people who assisted in this project. Special thanks are due to Dr. Justin S. Clark, my advisor and chairman of my supervisory committee, whose insights and experience allowed me to avoid many pitfalls, and whose interest and friendship provided encouragement along the way. Special thanks to Dr. Fred L. Farr who helped me cope with all the problems during the realization of this project. I am grateful to Mr. Alan Smith who gave me the best conditions for working on this project. Special thanks to other members of my supervisory committee, Dr. Jerald D. King whose input permitted me to develop the knowledge base and the user interface, Dr. Peter J. Haug who spent time with me on guiding me on the project, and Dr. Homer R. Warner for his support. My sincere appreciation is given to Dr. Henry Lundsgaarde whose insights were of real help to me at the starting point of this project. I recognize the assistance of all the Pediatric radiologists of the Primary Children's Medical Center, especially Dr. Richard Boyer and Dr. William Nixon for assisting me in developing the radiological frames and in their evaluation. Special thanks to all the Pediatric residents of the University of Utah who assisted in the evaluation of the radiological frames. I also express my sincere appreciation to my family for helping to make this project possible.

CHAPTER I

INTRODUCTION

General Remarks on Expert Systems

Expert systems are computer programs that are capable of performing tasks that normally only skilled and trained human beings are able to perform. The field of expert systems has been under investigation during the last two decades. Research has mainly focused on ways to represent knowledge in order to provide decisions as close as possible to those of the human experts. Medicine has been viewed as a vast field where experiments on expert systems could be conducted. The idea that such computer programs could be used to help and assist the medical staff in their clinical work subsequently led to research in computer simulation of the human reasoning process, coupled with efforts to assist physicians in coping with the explosive growth in medical knowledge.

A Hospital Information System (HIS) is a general data base management system for a hospital. The HIS consists of three modules: (i) The data acquisition module; (ii) The data analysis module; and (iii) The data reporting module. If the Hospital Information System contains decision logic in the analysis module at the level of an expert, this module can then be referred to as the expert system part of the HIS.

The traditional structure of medical expert system consists of three components: (i) The data base where the data on a particular patient are

stored. (ii) The knowledge base where the knowledge pertaining to a particular field of medicine is stored; and (iii) The inference engine which makes a link between the data base and the knowledge base. The inference process is done by specific mechanisms, such as rules-based etc., described below.

Today, medical expert systems are being developed in Medicine for practical use as an aid to the health care providers. However, state of the art, medical expert systems are far from their potential of serving the needs of the health professionals. Therefore, the attitudes of medical practitioners to computers are still ambivalent, a fact that can be an obstacle to the implementation of such systems. To achieve maximum success, research that focuses on the implementation of expert systems should be tailored to the health professionals' needs so that the systems will be attractive to them.

Need for Expert Systems in Medicine

Medicine is a field in which help is needed. If all-knowing, consistent, experts were ever-present in every medical care situation, there would be no need for medical expert systems. However, since such is far from being the case, medical expert systems can be of significant aid to patient management. This management aid can be either in diagnosis or in treatment.

The major need for expert systems comes from the fact that physicians are human beings and as such, they are not all-knowing, and suffer occasional memory lapses. Computers (which contain expert systems) can come close to matching the knowledge of an expert (or panel of experts) in a certain field. Such computers can also help the less expert physicians or

the out of field physicians to cope with the explosive growth of medical knowledge. Lawrence L. Weed [1] asserts that physicians can never remember enough, they can never retrieve from memory accurately and quickly enough to cover all possibilities: so they distort the original problems or lose sight of them altogether. He sees computers as an extension of the physician's memory, much as the automobile is an extension of one's muscles. He urges the use of expert systems to provide better ways of organizing the care of multiple-problem patients [2].

Computer based expert systems can also provide better management of science information. A study by Williamson et al [3], to identify self-perceived problems in managing science information, suggests that published reviews are the most useful means of fulfilling those information needs of practitioners and their opinion leaders. Less than 1 in 3 practitioners personally searched the literature when information was needed. Two in 3 claimed literature volume was unmanageable. The study concluded that 9 of 10 practitioners and opinion leaders assessed the scientific value of literature primarily from their own experience with less than 1 in 10 contacting research methods specialists. It was concluded that primary practitioners require substantial help in meeting current science information needs and that expert networks containing a distillation of current medical science might help meet these needs. This study further suggests that data that are stored in computer networks can be used for updating the knowledge base for quality assurance and for research.

Expert systems can help make medical care more consistent. As pointed out by Brent James [4] consistency or rather the elimination of inappropriate variation is the key to improved quality in medicine. Consistency is improved at several levels. At the level of the individual

expert clinician, inconsistencies due to time constraints or fatigue are eliminated. At another level, computers can also encourage consensus within a group of clinicians and on a still broader level, encourage consensus among hospitals in a hospital system. Experts have even discovered inconsistencies in the way they practice when they attempt to organize and structure their knowledge to meet the requirements of a computer resident knowledge-base.

Categories of Application of Medical Expert Systems

There are three categories of applications of expert systems in medicine: (i) alerting; (ii) diagnosis (prompting); and (iii) protocols management.

The alerting function of an expert system is for the purpose of telling clinicians that something is wrong. Examples might be that a chemistry is out of the normal range, or that the heart rate has changed significantly, or that a prescribed medication has not been given. While alerting is less sophisticated than the diagnostic function, it is very important for averting catastrophe. As the value of alerting is established, the debate over cost effectiveness of clinical computers will disappear. No hospital will dare to be without the protection of expert system alerting.

The diagnostic function of an expert system goes beyond a warning that something is wrong; it attempts to pinpoint the underlying illness or condition. Sometimes the diagnosis is based on a limited set of data. For example, a cardiac arrhythmia uses only the ECG. Diagnosing other conditions or illnesses may require lab tests, x-rays, patient history and physical exams. In such cases a practical expert system (like a good physician) should make judicious use of medical resources. It should prompt the user

for only that information that is best calculated to rule out remaining diagnostic possibilities.

The protocol function of an expert system when combined with the diagnostic function tells not only what is wrong with the patient, but also, what to do about it. If an expert system were interfaced directly to a respirator and IV pumps, aspects of the protocol could be implemented automatically. Even where expert system protocols are presented only as recommendations to clinicians, they offer greater inducements to utilization than paper protocols. For instance, where a computer terminal is also being used for placing orders and reviewing results, the system can make protocol difficult to ignore. Also, a computer system has the ability to make all members of the medical team aware of protocol violations immediately.

Background - Review of Previous Diagnostic Expert Systems

While there have been many successful expert system developments directed towards medical diagnosis, most of them are restricted to too small of a field (subspecialty) to serve the application need described above. For example MYCIN [5,6] is limited to advising physicians and medical students about the appropriate treatment of infection. CASNET [7] is restricted to cases of Glaucoma. The Present Illness Program (PIP) [8] deals with renal physiology and pathology, while ABEL [9] is restricted to acid electrolytes disorders. These programs were not intended to be used clinically; their intended purpose was that of demonstrating capabilities of expert systems in the medical field. Example of systems that do not have the subspecialty limitation are INTERNIST [10], CADIAG-2 [11], ILIAD [12] and HELP [13, 14,15].

Expert systems vary in the way that the knowledge base is represented and in their inference mechanism. Mycin was the first program that acquainted us with the rule-based method. The system uses a backward-chaining process that tries to satisfy the consequent part of the rules by the antecedent part. As medical rules involve uncertainty, Mycin associates a certainty factor with each rule, which is a number between 0 and 1.

INTERNIST [10] the program uses an ad hoc mechanism for generating the differential diagnosis in the domain of internal medicine. It creates the differential diagnosis list from the positive and the negative manifestations that are associated with each disease. The diagnoses in INTERNIST are classified in a disease hierarchy. This controls the proliferation of hypotheses during the diagnostic process. For each hypothesis, a score is computed that permits ranking diseases according to their likelihoods. The differential diagnosis list is sorted and refined till a final list is created or the computer cannot reach any conclusion. This system currently has 500 diseases in its data base with 2600 links and 3550 manifestations.

A difficulty with the hierarchical classification mechanism of INTERNIST is that problems have no substance outside the context of a particular case in which they emerge, leading to missed diagnoses. To overcome this limitation a more highly structured knowledge base, called CADUCEUS, was developed by Pople et al. [16]. In CADUCEUS a new type of network was added to the knowledge base called a causal network or also, a "deep network." The manifestations of a disease are organized on the basis of the pathological states giving rise to the observations, and these states are in turn organized into a causal network. The nodes have links that are labelled as "caused-by." CADUCEUS uses heuristic operators for combining

independent tasks of the system. As such, CADUCEUS represents the essential nature of the diagnostic reasoning process. A second example of a system that uses causal networks is ABEL, developed by Patil et al. [9]. This system differs from CADUCEUS in that associational links and grouping links are used. This system has its operators for constructing the patient model from the specific data about the patient. CASNET [7] is a third example of a system that uses causal network. Unique to CASNET is the use of a specific probabilistic method to describe a disease.

EXPERT [17] system developed by Kulikowski, Weiss and Kern at Rutgers University uses production rules in a causal-inferential network. AI/Rheum [18], which is a computer-based rheumatology consultant performing at the level of an expert, runs under the EXPERT system. EXPERT allows AI/Rheum to use patient findings to infer disease hypotheses.

The knowledge base of the Present Illness Program (PIP) [8] is based on a pattern matching method. If a finding matches one of the hypotheses then that hypothesis becomes active. The program has a scoring system that measures the likelihood of its leading hypothesis. The likelihood is estimated by combining a function that measures the fit of the observed findings to the expectations of the hypothesis with a function that is the ratio of the number of findings that are accounted for by the hypothesis to the total number of existing findings. These two components of the likelihood estimate are called the matching score and the binding score.

RECONSIDER [19] is an interactive program that produces a differential diagnosis given a list of patient attributes. Attributes that define each disease are partitioned into parts such as etiology, signs, laboratory data, etc. The program matches the attributes. A physician is able to enter a list of clinical findings and is then provided with a list of diagnoses in which those

clinical manifestations are attributes. The list is ordered so that the diagnosis that best matches the set of attributes is listed first. The user controls the search by selecting stronger or weaker definitions of the match between the patient's findings and disease attributes to broaden or narrow the search. The program includes 3,262 disease definitions. RECONSIDER was recently implemented at Georgetown University Medical Center. The system is used for teaching purposes, to expose medical students to special computer applications in their profession, to raise their level of computer competency and to compensate for the missing educational experience [20].

ILIAD [12] is a system that uses mainly statistical inference process. It uses knowledge frames for diseases encountered in internal medicine to teach medical students about differential diagnosis. Its knowledge is represented in both Bayesian and Boolean frames that permit the use of sensitivities and specificities to describe the relationship of a disease to its manifestations and to provide a basis for explaining its conclusions. In addition to differential diagnosis, Iliad provides advice regarding the most appropriate information to seek at each stage of the work-up. The Iliad knowledge base is also used to simulate patient cases and evaluate the problem solving performance of medical students.

ICON [21], developed by Swett and Miller is a computer-based expert system being developed to help radiologists with the process of differential diagnosis. It is built with the use of ESSENTIAL-ATTENDING, a computer program developed to help construct expert critiquing systems. ESSENTIAL-ATTENDING has three components: production rules, expressive frames and a prose generator. ICON has two components: a domain independent part (ESSENTIAL-ATTENDING) and a domain specific part. ICON focuses on the domain of lung disease as seen on a chest radiograph in patients with

lymphoproliferative disorders. ICON asks the radiologist to propose a diagnosis and then it gives the evidence supporting that diagnosis or competing diagnoses. The system's output is in the form of an English prose critique. This approach combines the computer's ability to recall detailed information with human reasoning skills.

CADIAG-2 [11] is an expert system that uses quantitative numbers to characterize the strength and association between symptoms and diseases. These quantitative numbers are interpreted and manipulated by means of fuzzy set theory and fuzzy logic.

CADIAG-2 is an example of a comprehensive system that is able to access patient data and laboratory test results already collected in the patient's data base. It has automatic screening procedures to detect abnormal findings and to propose further useful examination and an on-line consultation system for the clinician to assist him or her in clarifying the patient's disorder completely and in great detail. This system, unlike most of the systems, is not a stand alone system, but embedded in a HIS. It is a data-driven system. There are very few data-driven systems such as the HELP system [13 , 14 , 15] , PUFF [22] and INTERNIST. Very few of these systems are integrated into a central Hospital Information System.

Each of the medical expert systems has different characteristics: the way that data are coded and stored; the way that the knowledge base is structured and represented; the mode of the user-interaction with the system and the system output. Despite the tremendous efforts to computerize the medical practice, partially or completely, the success is very elusive. This might be due to several reasons: negative attitudes of practitioners to computers, unmet needs of medical care providers,

cumbersome hardware, or inability of computer systems to function and behave in the same way that physicians do.

A Hospital Information System that is currently in use with success and has expert system elements embedded in it is the HELP system [14,15]. The HELP (for Health Evaluation through Logical Processing) Hospital Information System consists of a comprehensive expert system for acquiring medical data and implementing decision logic. The development has been ongoing for over 20 years at the University of Utah and the LDS Hospital in Salt Lake City, Utah. This system is presently operational at LDS Hospital which is a 550-bed tertiary care hospital serving the needs of the Intermountain West. Four objectives were outlined in the development of HELP. They are: (1) the system must be able accommodate an ever-expanding medical data base; (2) the system has to process medical decision logic; (3) the system has to be capable of serving both the medical and administrative needs of the hospital; and (4) the system must provide effective research subsystems that facilitate clinical research on the large data base naturally acquired by the system.

The system characteristics are: (1) the decision logic is modular with control and interaction of modules determined by medical experts for a given module; (2) the processing of the decision criteria is data driven; (3) the system is designed to provide output to a variety of sources; and (4) the system has been designed to interact with the data acquisition modules of HELP.

Background - Computer Medical Systems

for Newborn Intensive Care Unit

Some of few computer systems found in the literature that were developed for the Newborn Intensive Care Unit are described below.

The system that was developed at the Perinatology Center of New-York Hospital - Cornell University [23], responds to two major problems that are often seen in Newborn Intensive Care Units: (i) the difficulty of accessing the avalanche of information generated by each patient in the Neonatal Unit; (ii) the need for long term storage of data permitting an easy retrieval in future.

Van der Lei et al. [24] described an experience at the Free University Hospital, Amsterdam, the Netherlands, of using fourth-generation software to develop, in a short time, departmental information system for the Neonatal Intensive Care Unit. This system is capable of giving an output of case summaries that are used for follow-up studies, list of diagnoses according to the H-ICDA classification, summaries of follow-up studies, and scheduling of follow-up studies.

Janik et al. [25,26] reported an IBM 370-based system for recording, reporting and researching data on infants admitted to newborn intensive care units. They concluded [27] that automatic computer production of admission/discharge documents and of letters to physicians and agencies accounted for: (i) 80% reduction in admission/discharge paperwork; (ii) 12-fold increase in completeness of medical data; (iii) 96% reduction in time from patient discharge to receipt of patient care information within the medical community. This software is currently in use at the Intermountain Newborn Intensive Care Center of the Primary Children's Medical Center and the University of Utah Hospital in Salt Lake City, Utah. The software

currently runs on the DEC-compatible hardware and is commercialized under the name CETUS-100.

A comprehensive computer-assisted diagnostic system of acquired and congenital diseases has been developed by Barnes et al. [28 , 29] . With information available at the time of the initial history and physical examination such as abnormal symptoms, physical findings, and laboratory results, this system can provide a list of diagnoses that will include the correct one in a high percentage of cases.

A recent prototype system for perinatal knowledge engineering using an artificial intelligence tool was reported by Sokol et al. [30]. This system mainly uses artificial intelligence techniques to develop a computer based "Perinatal Consultant."

Some other publications on limited applications of computers in the Newborn Intensive Care Unit were also reported. Walker [31] developed a computerized record for the Neonatal unit; Perlstein et al. [32] developed a computer program for on-line access to physiologic and environmental data and for control of heating the infant; Finner et al. [33] described the development of a computerized data base and a program to develop automated neonatal discharge summaries using a personal computer and proprietary software; Lindstrom et al. [34] documented the general use of computers in a Newborn Intensive Care Unit.

It was found that only the program described by Barnes et al. [28] and Swender et al. [29] is able to produce a differential diagnosis in the domain of congenital pediatric diseases. There was no report of a comprehensive and integrated computerized neonatal system that is able to make a differential diagnosis or to automatically produce a problem list.

Acceptance of Computer Systems

Some of the computer systems for medical use, such as the HELP system are used in real clinical settings, but most of them are not used and ended just as research product or a teaching tool [35]. Friedman et al. [36] made a survey of 32 clinical applications of computers and found that 51% of projects had been abandoned or suspended. They reported that only 19% of the systems are in routine use in the hospital surveyed.

Physicians like computer systems that will enhance their management capabilities. They like systems where access to data will be fast and convenient, systems that warn them of problems or potential problems, and systems that aid in diagnosis. They reject systems that they do not understand as jeopardizing their role and their status in front of their patients. Also, physicians do not like systems that are inconvenient to use or increase hospital or government control over them.

Surveys on the attitude of physicians towards computers were conducted by several researchers. Singer et al. [37] concluded that more than one third of the physicians who participated in a survey were not satisfied with their ability to keep up with new development and less than 10% were satisfied. Ninety percent thought that a computer data base would improve their access to information in the literature, and 85% thought it would improve their practice of medicine. Most physicians indicated a significantly greater preference for literature summary and patient registry features over probability estimation capabilities of a computer data base. Friedman and Gustafson [36] reported that the reason that most of the systems are not in routine use is due mainly to inconvenience for physicians in using such systems and to the fact that the systems do not meet their needs. Physicians today are reluctant to accept computers as diagnostic-

support or history-taking tools because it is perceived as a threat to the physician's status and relationship to patients. In a survey conducted by Anderson et al. [38] it was concluded that physicians recognize the potential of computers to improve patient care, but are concerned about the possibility of increased governmental or hospital control, threats to privacy, legal and ethical problems. All participants of the survey were uncertain as to the potential effects of computers on their traditional professional role as physicians and on the organization of practice. It might be that this attitude of students and physicians plays a major role for the slow introduction of clinical computer systems. Teach and Shortliffe [39] studied physician attitude toward computer-based clinical decision aids and the effect of a two day tutorial on medical computing. The results indicate that physician accept applications that enhance their patient management capabilities, but tend to oppose applications that infringe upon their management role. General expectation about the effect of computing on current medical practices is found to be generally favorable.

Physicians, as well as many other professionals, want incorporation of computers in their medical practice. Many practitioners cannot specify in great details their needs and expectations. The ones who are able to do so are less concerned with decision support than with acquisition of laboratory data and electronic data storage for future reference and report generation. The fact that more and more physicians, mainly in the younger generation, are becoming computer-literate will facilitate the implementation of computer systems in medical practice in the future. Hardware issues do not presently impose any obstacle to computers' use in medical practice. Software tools are also available to facilitate communications with computer systems. The main issue of the field of

medical computing, at present, is tailoring the purpose and behavior of systems to the physician's needs.

Attributes Necessary for a Successful Expert System

In order to create a decision support system that will meet most of the needs of the health care providers, comprehensive expert systems are necessary. These systems should support higher levels of application such as diagnosis and protocols. Expert systems should be developed in a broad field, as opposed to a narrow field, in order to be sufficiently useful to gain physician acceptance. Expert systems should not be regarded as stand alone systems that deal with a specific problem. It should be realized that a small program that calculates the ingredients of total parenteral nutrition, or a program that is designed only for interpretation of ECG or some other context that serves a very narrow or a limited need, will not attract the practitioners. An automated drug dosage calculation program appears to be readily accepted, but it is a stand alone system that focuses on a very limited need. The needs today are for integrating data from all sources, for formulating problems and for decision assistance. Today in the eve of the 21st century there is no shortage of specialists nor subspecialists in all fields of medicine in the United States. A small program that is designed to do a simple task will not help and will not attract the practicing specialist. Only a comprehensive expert system that is designed to assist the physician in the decision making process, a system that allows automation of such tasks as data collection and reporting, and a system whose decision support serves to alert and remind, has a chance of being accepted and used routinely by physicians and other health care providers.

The expert system should be embedded in a hospital information system that will help collect a comprehensive data base while minimizing physician data input. An example of an area in medicine where most computerization efforts have been invested is certainly the intensive care unit. The intensive care unit has an increasing number of sophisticated machines such as monitors, respirators, etc. that are readily interfaced to computers which can post alerts when ever one of the measured variable is out of its normal range. Most of the computers in intensive care units interface with the medical instrumentation, but do not deal directly with the health care providers. Computerized interfaces to monitors offer the acquisition of only a limited set of data. This consequently leads to limited decision capabilities if a decision aid tool is embedded into the system. Most of these computers [40,41] deal with subjects such as hemodynamic calculations or calculations of infusion rates of potent intravenous drugs. Some of these systems are designed as an educational tool to the health care providers [42], others are used as an aid in the complex task of the medical decision making [43,44]. The Intensive Care Unit is an example of a place that can benefit from a comprehensive expert system. In order to develop a computer system that will make really useful decisions, it has to be interfaced not only to monitors, but to all sources of data including the laboratory, the x-ray department and mainly to the one who has the final decision in managing and treating patients - the physician. For an integrated computer system that must acquire data and make decisions in an area of specialty such as the Newborn Intensive Care Unit to be effective, it has to acquire data that only physicians and nurses can provide such as physical examination data in conjunction with other data such as the laboratory, x-rays, and monitoring data.

Such a system should be friendly, not only in the way that the communication with users is being handled, but also in the way that the medical advice is being given. Physicians might be offended by systems that constantly critique their thinking process or systems that are used to control their performance. Computer systems should not be stand alone systems in giving their medical advice. Systems should elaborate the diagnosis and management with physicians and help them in those matters where their limitations as humanbeings exist. If possible, the reasoning process of an expert system should be as close as possible to the reasoning process of the physician humanbeing to facilitate his or her understanding of its operation. In this way he or she will feel confident with the system and his or her role as the physician will not be jeopardized. This will attract practitioners into using computer systems more and more in their practice and will facilitate the implementation of such systems in real clinical settings.

Objectives of the Present Dissertation Work

The goal of the present work was to develop a comprehensive expert system for the newborn intensive care unit of the Primary Children's Medical Center in Salt Lake City, Utah. The system, called NEONATE, was to be developed as part of the HELP system. NEONATE has the further requirement of replacing the CETUS system [27] that is currently used by the physicians in the Newborn Intensive Care Unit of the Primary Children's Medical Center to produce their admission report. As the medical center will adopt the HELP system [14,15] in the future, it was necessary to develop an application for the neonatal unit that will not be stand alone, but will be part of the Hospital Information System. Currently the Neonatal Unit lacks any expert system applications on the HELP system. Furthermore, it was

planned to add decision elements that are absent in the current CETUS system. For the purpose of this dissertation a prototype was developed that deals only with aspects of the admission process. The knowledge base is limited only to the differential diagnosis of tachypnea, which is the most common finding of the neonates admitted to the intensive care unit.

The evaluation of NEONATE was limited to testing the following three hypotheses

- a. NEONATE produces an admission report that is superior to the current admission report.
- b. NEONATE helps physicians who are not trained in radiology to make more correct diagnoses using x-ray films.
- c. NEONATE is a potentially effective prompting device.

CHAPTER II

THE DEVELOPMENT OF THE NEONATE SYSTEM

The NEONATE System

The NEONATE system is being developed at the Primary Children's Medical Center (PCMC) in Salt Lake City, Utah. PCMC is a 200-bed tertiary care children's hospital in the Intermountain West region. PCMC, as the LDS Hospital, is part of the Intermountain Health Care corporation and recently has been engaged in activities to develop the Pediatric module of the HELP system. PCMC plans to purchase shortly the TANDEM computer and start implementing existing HELP modules in its facilities.

The NEONATE system makes part of the HELP system and it was developed for the Newborn Intensive Care Unit. It was developed as a comprehensive decision support system that takes advantage of all aspects of the Hospital Information System: data acquisition, data analysis and data reporting. This system is designed not to be a stand alone consultation system, but a system that covers all aspects of clinical activities and will be used in the future by all medical professionals. The system should enable easy access to and efficient retrieval of data by all medical professionals. It is assumed that a comprehensive expert system that is designed to assist physicians and nurses in the decision making process, that allows automatization of tasks such as data collection and reporting, and whose decision support serves to alert and remind has a better chance of being

accepted and used routinely by the medical staff. Such a system should eventually produce a change in the behavior of the health care providers towards data collection. It should increase the number of items collected, help to focus the physical examination, increase awareness of problems otherwise ignored, and consequently improve patient care.

Tools for NEONATE Development

Neonate is part of the HELP system and as such it uses the tools that are available in HELP.

The data base consists of two elements: (1) a long term abstract of demographic and clinical information likely to be useful if the patient is readmitted to the hospital; and (2) a short term comprehensive collection of all data gathered during the current hospital admission. All data are stored in codes form (as opposed to free text) so that data can be retrieved and analyzed for use in research and decision logic. Hierarchic codes are defined using a system called PTXT (Pointer to Text) which is basically a computer based dictionary.

Programs that manipulate the data base are written in PTXT Application Language (PAL), a structured programming language. New tools were recently added to the language. Among them is a Report Generator which permits printing and sending reports to a specified destination within the system. A second tool is the HELP Video Control (HVC) which is terminal emulator software with the capability to program windows within the PAL language. A new version of the PAL language has been recently released. The new version of the software permits using modular frames.

The HELP system at LDS hospital presently includes 10 Tandem Computer fault tolerant processors using 3.4 gigabytes of disk storage

distributed over 14 disk drives. The 8 drives handling clinical data are mirrored to reduce the possibility of data corruption. Eighteen Charles River Data System (CRDS) minicomputers are interfaced to the Tandem serving as multiplexors and pre-processors [45].

Developmental Rationale

The system is designed in a way that will encourage a successful implementation such as: comprehensive data base; user friendly interface; and decision support that mimics the way physicians reason.

The prototype of the NEONATE System has five components as shown in Figure 1.

The User Interface

To make such a system adopted and utilized, the interface for data acquisition should be friendly, easily used and appealing to physicians. Data entry should be structured in a way that encourages completeness and discourages omissions frequently seen in traditional medical records. Laboratory data will be entered into the patient's computerized medical record by the laboratory technicians and shift data will be entered into the computer by the nursing staff. The HVC in the PAL programming language is used to create the windows where direct data acquisition is necessary. Two kind of windows can be displayed on screen: Multiple choice windows and Data entry windows. Both of these types of windows are used to create a series of menus for admission data entry by physicians. The main menu of the admission process contain ten items: Seven of these items are for data entry per se - Antepartum data acquisition, post-partum data acquisition, family history data, obstetrical data, physical examination and chest x-ray

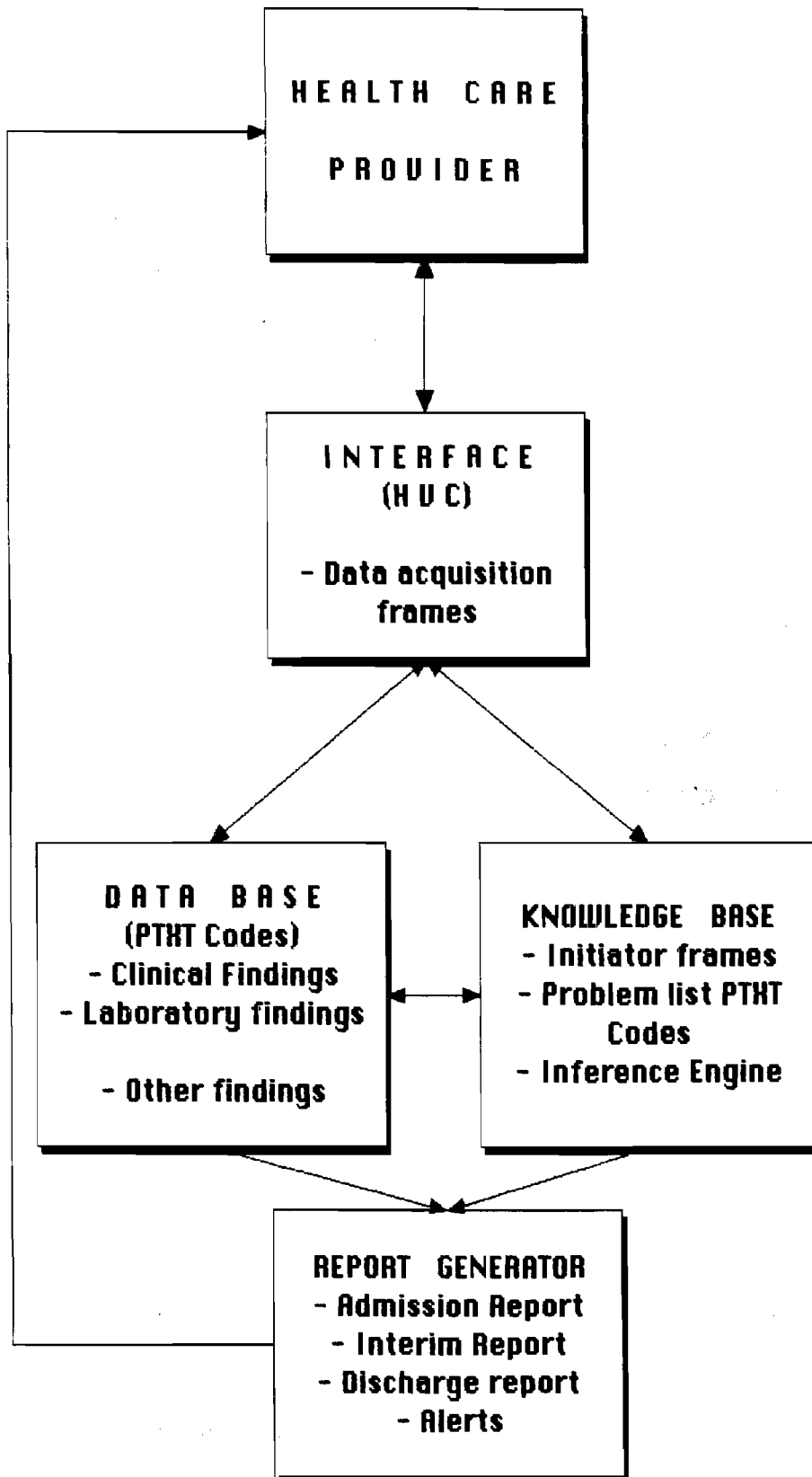


Figure 1. The Components of NEONATE

data acquisition. One item permits the user to print a report by simply pushing a button. There is only one item for entering a problem list. Currently the user has the option to enter problems in addition to the problems that the computer is able to diagnose by its decision mechanism. The physician may also alter the problem list and change attributes of problems (see knowledge base section).

The physician is led through a series of menus to complete the data acquisition process. The structured method of data collection prevents him or her from inadvertently skipping items in his or her data gathering. Furthermore, the system prompts the clinician for the information that is vital to decision making and thus to patient care. The process permits more complete data acquisition than the pen and paper process. Some positive findings (such as Hepatomegaly) might trigger more questions about findings otherwise ignored. (Hepatomegaly might trigger another window that consists of describing an enlarged liver. Four choices are offered to the user such as: Hepatomegaly, firm, tender; Hepatomegaly, firm, nontender; Hepatomegaly, soft, tender; Hepatomegaly, soft, nontender. The physician should choose only one of these possibilities; and, if he or she does not know the choice he or she should go back to the patient and re-evaluate the finding.)

The Data Base

The data base is the computerized medical record. It consists of strings that are stored as PTXT codes in the standard HELP clinical data base. The codes consist of clinical and laboratory findings as well as items in the problem list. PTXT versions of ICD-9 codes are used as the codes for the problems that reside in the problem list. The data dictionary has recently

been extended to include findings for the pediatric population. This was necessary to avoid ambiguities involved with the use of codes developed for adult findings. It will enhance the value of the data base for future research purposes. Codes are stored with the exact time of storage. This permits trending for the patient findings or laboratory results. It also permits easy temporal access to all the data.

The Knowledge Base

The knowledge base should have capabilities to mimic the way physicians reason. Extensive research to determine the way this is done was conducted by Elstein et al. [46]. They confirmed that diagnostic problems are solved through a process of hypothesis generation and verification. Hypotheses are consistently generated early in work-up when very limited data have been obtained. An early formulation may be revised or discarded if subsequent data fail to confirm it. The Present Illness Program [8] is generally associated with this cognitive model and many other systems use it as well. The differences among these systems are in the methods used for hypothesis verification or rejection. Some use rules of thumb, others use scoring methods, and others use statistical approaches. Although the choice of rules of thumb for NEONATE is consistent with the goal of mimicking the physician's approach, some additional justification and background is offered here. As documented by Szolovits and Pauker [47] the Present Illness Program and some of the other famous expert systems sometime fail to reach the correct diagnosis while using scores or probabilities. In some instances, in the case of the Present Illness Program, small variation in a borderline clinical case can push a score just above or below a threshold and can affect the program's conclusion significantly. While developing the

system, the reluctance of the Neonatal staff to add scores to the decision logic was obvious. When the clinicians were asked to establish sensitivities and specificities for the findings they declined - doubting that their colleagues would agree with their estimates. The rationale of Weed [48, 49] in developing the Problem-Knowledge Couplers was to overcome the need of using probabilities. According to him probabilities could be applied afterward using knowledge-couplers if uncertainty still remained, but most of the time it would not be necessary. Although practically all expert systems have some probabilistic mechanism embedded in them, it was desirable in the present development to put the probabilistic mechanism of NEONATE in a textual form ("possible," "suspected," etc., see below), that was familiar to the medical staff. The reasoning logic of the NEONATE system seeks to avoid mathematical computation, and come as close as possible to the physician's way of thinking.

The rule based approach to the cognitive model used by NEONATE uses at least two types of rules emphasized in other expert systems. Some rules are based on understanding of cause and effect. Causal networks [7,9] have been used in the fields of medicine where the pathophysiology is well-known. In other cases the underlining physiology is not well-understood and the practicing physicians use Pattern Matching methods [8] to reach a diagnosis. That is why a mechanism that can be described as a combination of a rule-based and a pattern matching mechanism to verify or eliminate the hypotheses was adopted.

While in some expert systems, such as MYCIN [5,6], the inference engine is domain independent and separated totally from the knowledge base, NEONATE embeds the inference engine in its knowledge base. The knowledge base/inference engine of NEONATE is considered as one unit. The

inference engine is considered as one of the components of the knowledge base. The domain dependency of the inference engine is a direct outcome of the pattern matching mechanism adopted in NEONATE. Each of the frames is developed individually with its own rules and its own textual probabilities. Each frame is very specific, and no common diagnostic mechanism for all diseases could be found.

The knowledge base has three components (Figure 2): the initiator, the problem list and the inference engine.

The initiator looks for abnormal findings in the data base. An abnormal finding might be an abnormal laboratory value such as a low level of blood glucose or an abnormal physical finding such as hepatomegaly. For each abnormal finding the initiator puts all related differential diagnostic problems in the problem list. If the problem is already in the list the initiator will not reenter it, but it will mark the problem with the new finding that triggered it. The initiator is equivalent to the hypotheses generator of the cognitive model [46]. It creates a differential diagnosis list for each abnormal finding and makes physicians aware of all possible problems including rare ones that might otherwise be overlooked. (Figure 3 is an example of the list triggered by tachypnea, when respiratory rate is above 60.)

The problem list is a list of all the potential problems under consideration for a given patient. Each problem has an associated status and a set of findings that caused this problem to be included in the list. The problem statuses are "suspected," "highly suspected," "confirmed" or "resolved." For problems that can be diagnosed by x-rays, the possible statuses are "possible," "suspicious," "highly suspicious" and "confirmed." When the initiator puts problems in the problem list, they are given the

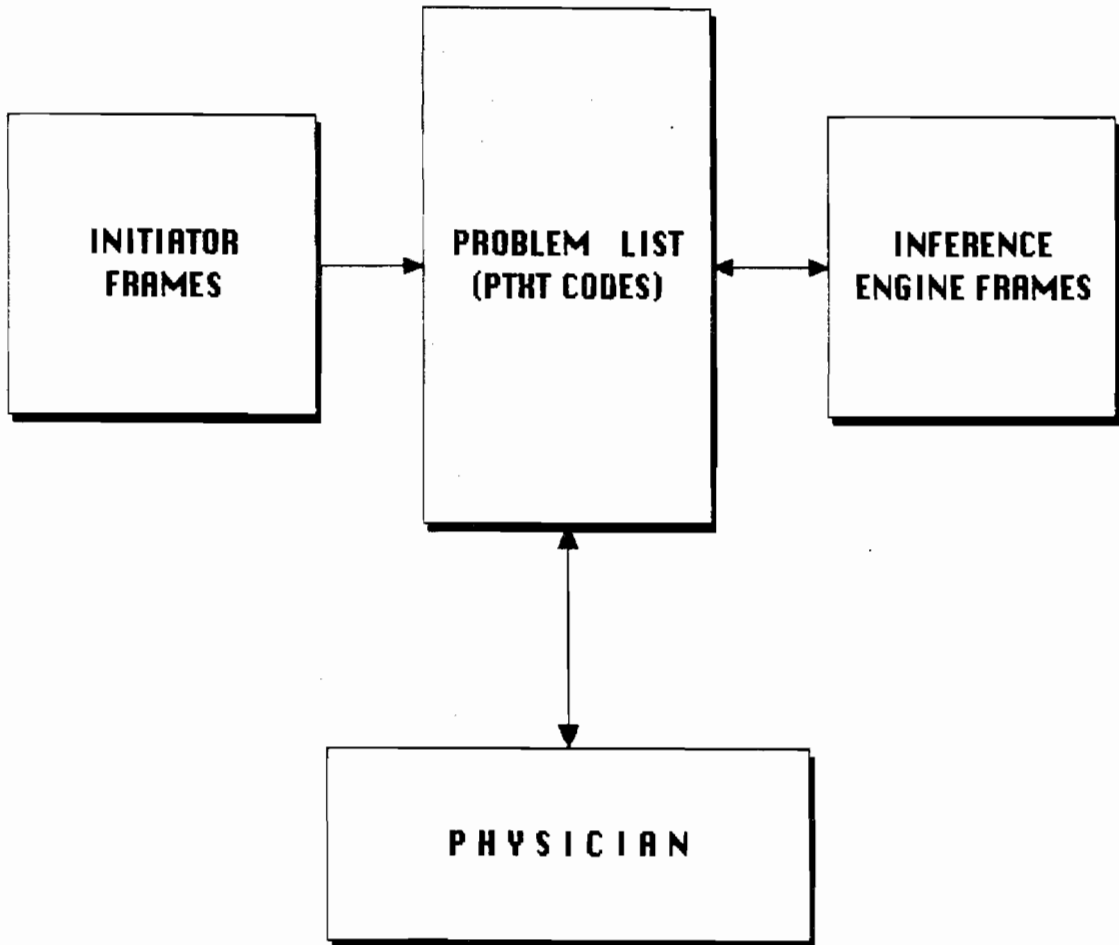


Figure 2. The Knowledge Base

1. Fever
2. Dehydration
3. Metabolic acidosis
4. Anemia
5. Hypovolemia
6. Hypovolemic shock
7. Heart failure
8. Cardiogenic shock
9. Congenital cyanotic heart disease
10. Neonatal hypoglycemia
11. Hyaline membrane disease
12. Transient tachypnea of the newborn
13. Aspiration pneumonia
14. Pneumothorax
15. Pneumonia
16. Broncho-pulmonary dysplasia
17. Polycythemia neonatorum
18. Neonatal hypocalcemia
19. CNS depression
20. Intraventricular hemorrhage
21. Subarachnoid hemorrhage
22. Diaphragmatic hernia
23. Esophageal atresia or T-E fistula
24. Sepsis
25. Pneumomediastinum
26. Pulmonary hemorrhage
27. Chylothorax
28. Hemothorax
29. Infectious pleural effusion
30. Pneumopericardium
31. Methemoglobinemia
32. Muscular dystrophy
33. Hypokalemia
34. Diaphragmatic paralysis
35. Persistent pulmonary hypertension

Figure 3. List of Problems that are Entered into the Problem List by the Initiator if Tachypnea Exists.

status of "suspected". The physician can subsequently override and change the status of the problems.

The inference engine consists of a series of frames - one frame for each problem or disease. If a particular problem exists in the problem list, the corresponding frame will be activated and may change the status of the problem or delete the problem from the list. Each frame in the inference engine contains rules for changing the status of the problem. It can be changed from "suspected" to "highly suspected," from "highly suspected" to "confirmed" or from "confirmed" to "resolved." In cases where the data necessary to satisfy the rules in the inference engine are lacking, the status will remain unchanged, but a query will be issued to the physician for the missing data. For example, in the case of newborn seizures, Hypoglycemia is one of the problems entered by the initiator. If at a time the Hypoglycemia frame is triggered and blood glucose has not been entered within the last 60 minutes, then a query will be issued in order to make the clinician aware of this problem. (Figure 4a. is the Hypoglycemia frame, 4b. is the Metabolic acidosis frame, 4c. is the Heart Failure frame and 4d. is the Cardiogenic Shock frame). All frames are triggered by findings that exist in the patient's data base; but, whereas the data required by some of the frames consist of only abnormal findings (such as Heart Failure or Metabolic Acidosis), other frames require the output of previously run frames. Each frame in the inference engine can remove its corresponding problem from the list if there is no evidence to confirm the problem. Once a problem has been confirmed and resolved it remains in the problem list for reporting purposes. Most of the problems can be removed from the problem list and reentered again by the initiator. Some problems, such as Diaphragmatic Hernia, once ruled out and removed from the list cannot reappear since congenital problems do not

a. NEONATAL HYPOGLYCEMIA

<u>HIGHLY SUSPECTED:</u>	Blood Glucose level < 40 mg%.
<u>CONFIRMED:</u>	Blood Glucose level < 30 mg%.
<u>RESOLVED:</u>	Confirmed clinical AND Blood Glucose level >= 40 mg%.

b. METABOLIC ACIDOSIS

<u>HIGHLY SUSPECTED:</u>	pH < 7.3 AND pCO ₂ < 45 AND Plasma Bicarbonates < 25 mEq/l.
<u>CONFIRMED:</u>	pH < 7.25 AND pCO ₂ < 40 AND Plasma Bicarbonates < 20 mEq/l.
<u>RESOLVED:</u>	Confirmed clinical AND (pH >= 7.3 OR pCO ₂ >= 45 OR Plasma Bicarbonates >= 25 mEq/l).

Figure 4. Inference Engine Frames

- a. Hypoglycemia
- b. Metabolic Acidosis
- c. Heart Failure
- d. Cardiogenic Shock

c. HEART FAILURE

X-RAYS FINDINGS:

HIGHLY SUSPECTED: Enlarged Heart OR
Normal, but Increased Heart.

CLINICAL FINDINGS:

HIGHLY SUSPECTED: Highly Suspected for Hypovolemia OR
(Weak pulse OR Decreased Pulse OR
Bounding Pulse) OR

CONFIRMED: Highly Suspected Chest X-rays film.
Confirmed for Hypovolemia AND
(Weak OR Decreased OR Bounding pulses) AND
Highly Suspected Chest X-rays film.

d. CARDIOGENIC SHOCK

HIGHLY SUSPECTED: (Confirmed Heart Failure AND
Highly Suspected Metabolic Acidosis) OR
(Highly Suspected Heart Failure AND
Highly Suspected Metabolic Acidosis) OR
(Highly Suspected Heart Failure AND

CONFIRMED: Confirmed Metabolic Acidosis).
Confirmed Heart Failure AND
Confirmed Metabolic Acidosis.

Figure 4 Cont.

reoccur. The inference engine is equivalent to the component that verifies hypotheses of the cognitive model.

Szolovits et al. [47] point out that it might be that rapid generation and equally rapid modification or elimination of many explicit hypotheses plays a significant role in the physician's reasoning process. That is exactly the way the system functions. The system is designed in a way that the inference engine will function immediately after the initiator has been active and several times a day. In such a way the system might generate hypotheses which are then immediately eliminated by the inference engine because of lack of evidence to support these hypotheses. This might be compared to the physician who rapidly generates and eliminates hypotheses in his or her subconsciousness.

The Health Care Provider Component

The physicians and the nurses as an integral part of the system do the data entry of the history and physical examination. They accept alerts, queries and reports. The physician can override the problem list, change attributes of problems, eliminate problems and add new ones. It is no longer a critiquing system that might give the physician a feeling of being offended and a system that questions his or her role. It is a system that works with the physicians, helps them throughout the patient's hospitalization, but the last word and the last decision is of course made by the physician who is free to act according to his or her best professional judgement.

The Report Generator

This tool is embedded within the PAL language and it permits printing the admission report, discharge report, and interim report. Alerts and queries as well as the problem list can be also printed.

CHAPTER III

COMPARISON OF THE NEONATE TO THE CETUS ADMISSION REPORT

General Remarks

The NEONATE and the CETUS systems produce an admission report that include items such as full history of the mother's pregnancy, delivery history and physical examination findings. Both systems also include a problem list as part of the admission summary.

This chapter compares:

1. The completeness of the NEONATE vs. the CETUS report (except for the problem list).
2. The problem list in the current CETUS admission report to the unedited problem list produced by NEONATE.

It should be pointed out that the problem list in the CETUS admission report is composed by clinicians without expert system support whereas the NEONATE problem list used for this comparison is entirely composed by the expert system without the human editing it might otherwise receive.

Comparison of the Completeness

The claim that NEONATE will produce a more complete admission report than the current CETUS report is based on three things. First, when comparing the information sheet filled out by clinicians which is used to

create the CETUS report with the input screens of NEONATE it is noted that the items on the CETUS information sheet are a subset of the items on the NEONATE screens. The main additions to NEONATE are in the area of more branching choices for findings in the physical exam section. (Both systems allow the addition of free text.) Second, the lab and x-ray data NEONATE uses for its inference process can easily be added to the admission report with no additional data entry required from the users. If the CETUS report were to include such data it would have to be gathered and hand entered. Third, NEONATE insists on answers to history questions, (such as prepregnancy weight of the mother) even if the answer has to be "unknown." In the admission report NEONATE lists an item for each question. In the CETUS report no item is listed for some history questions if the question was left unanswered on the information sheet. A study was conducted that primarily measures the impact of this difference between the systems.

A study was undertaken to determine the completeness of NEONATE. Fifty patient records each of which included admission notes and a CETUS admission report, were used to create 50 NEONATE admission reports. Each NEONATE report thus listed those items particular to the patient plus the items always included in the NEONATE report. The items in the NEONATE report which were not present in the admission notes of the medical chart or the CETUS admission report were counted for the 50 records and the mean was calculated. The mean of the number of items that exist in the CETUS admission reports or the admission notes of the medical charts but were not present in the NEONATE reports were also calculated for the 50 records. The two values were compared by the t-test. The mean of the number of items in the NEONATE reports that are not present in the admission notes of the CETUS system and the medical charts was 8.42. The mean of the number of

findings that exist in the CETUS admission report or in the admission notes of the medical chart but not in the NEONATE report was 0.52 ($p < 0.01$).

Comparison of the Problem Lists

The comparison study between CETUS's clinician composed problem list (hereafter called the physician problem list) vs. NEONATE's expert system composed (clinician unedited) problem list is aimed at proving the following hypotheses:

1. There is a good agreement between the two problem lists.
2. NEONATE's computer inference process picks up problems that are currently missed by the physician problem list.
3. NEONATE's expert system produces more true positive items on the problem list and fewer false positives than the physician problem list.

Methods

NEONATE was evaluated retrospectively for its agreement with the physician problem list and for its improvement over this list. This was performed by comparing the NEONATE problem list generated from chart data with the physician problem list. Fifty medical records of patients were pulled from the archive of the Primary Children's Medical Center with the initial finding of tachypnea. Ten records were selected with the admitting diagnosis of Hyaline Membrane Disease, 10 records with Transient Tachypnea of the Newborn, 10 records with Aspiration Pneumonia, 10 records with Infectious Pneumonia and 10 records with extra-ventilatory air (Pneumothorax or Pneumomediastinum). However, it became apparent that separation into groups was not possible because most of the records showed multiple problems listed in the admission problem list. It is not unusual that

an infant admitted to the Newborn Intensive Care Unit has multiple problems such as Infectious Pneumonia and Hyaline Membrane Disease. For this reason it was decided to treat the 50 medical records as one group. All the 50 records were used to prove the first hypothesis of this chapter (the good agreement between the two problem lists) and only 30 random records were picked up from this group of 50 records to prove the following two hypotheses of this chapter. Because many of the common problems that NEONATE is capable of diagnosing are closely related and because of the relative simplicity of diagnosing the rare diseases that exist in the NEONATE knowledge base, it was impractical to choose the rare diseases for evaluating the system. For example confirmation of esophageal atresia does not pose any difficulties to the physician or to the computer if the right x-ray data are entered. On the other hand differentiation of Aspiration Pneumonia from Infectious Pneumonia or from Transient Tachypnea of the Newborn is sometimes difficult for the physician and is also a challenge to the computer.

If tachypnea is present, the initiator inserts all 35 diseases that make up the differential diagnosis of tachypnea into the problem list. The inference engine is capable of eliminating a few of them, provided adequate data is present in the computerized data base. Some of the problems can be eliminated from the problem list only by the physician. Most of the time two to three problems are being promoted by the inference engine to "Highly Suspected" or "Confirmed" and about 10 diseases are left under "Suspected" waiting for more data or waiting to be eliminated by the physician. The neonatologists found this long list to be inappropriate in terms of size; therefore, it was decided not to display diseases in the problem list with the attribute of "Suspected" except for Infectious Pneumonia and Sepsis which are frequently seen among sick newborns. The undisplayed diseases are not

eliminated from the problem list, but remain for future reference. The retrospective studies were conducted by referring to the problems that are actively displayed on screen and are apparent to the user physician. The fact that the computer knows much more than what it shows only strengthen the capabilities of the computer to request information for ruling out these problems, if necessary. If more data are available these problems can be given a higher attribute that allows them to be displayed on the screen or on the print-out without overwhelming the users with a long list of problems.

1. The Methods for evaluating the agreement between NEONATE's problem list and the physician problem list:

The agreement between the NEONATE admission problem list and the physician admission problem list was measured by the Kappa statistic [50]. For each of the 50 records that were studied, a Kappa value was determined by using the following formula:

$$\text{Kappa} = \frac{P_o - P_e}{1 - P_e}$$

where:

$$P_o = (A + D) / 35$$

$$P_e = ((A + C) / 35) * ((A + B) / 35) + ((C + D) / 35) * ((B + D) / 35)$$

P_o is the observed probability.

P_e is the expected probability.

A - Number of problems that appear in the physician admission problem list and the NEONATE problem list that are part of the closed list of 35 problems NEONATE is able to diagnose.

B - Number of problems that appear in the physician admission problem list but not in the NEONATE admission problem list that are part of the closed list of 35 diseases NEONATE is able to diagnose.

C - Number of problems that appear in the NEONATE admission problem list but not in the physician admission problem list that are part of the closed list of 35 diseases NEONATE is able to diagnose.

D - Number of problems that do not appear in the physician admission problem list and do not appear in the NEONATE admission problem list that are part of the closed list of 35 diseases NEONATE is able to diagnose.

Obviously, the sum of A, B, C, and D is 35, which is the total number of problems/diseases that NEONATE is able to diagnose.

Kappa as a measure of agreement beyond chance only, does not give insights to the true positive value or the true negative value of the problem list that is produced by NEONATE. Different studies, to determine the advantage of NEONATE problem list over the physician problem list, were conducted.

2. Methods to prove that NEONATE picks up problems missed by the physician problem list:

For the purpose of this study, in addition to the NEONATE admission problem list that the computer generates, and in addition to the admission

problem list as it is written in the CETUS system admission report, a Gold Standard list was defined. This Gold Standard list includes all the problems (inferred from the patient's chart) that the patient should be suspected of having at the time of admission. Problems were entered into the Gold Standard list by the following rules:

- a. All diseases/problems that are mentioned in the record's discharge report, having their onset during the day of admission, and being part of the closed list of 35 diseases NEONATE is able to diagnose.
- b. All diseases that are explicitly mentioned in the handwritten admission notes by the admitting physician, and that are part of the list of 35 diseases NEONATE is able to diagnose.
- c. All diseases that are explicitly mentioned in the first chest x-ray report, if the film was taken during the first day of admission, and the diseases are part of the closed list of 35 diseases that NEONATE is able to diagnose. (Chest x-ray findings as mentioned in the report are not interpreted as a problem. The name of the disease should be specifically written. For example: If it is written in the report "The findings suggest Aspiration Pneumonia," then the Problem Aspiration Pneumonia will be inserted into the Gold Standard list.)
- d. For the problems: Anemia, Neonatal hypocalcemia, Metabolic acidosis and Hypokalemia, the problem was inserted to the gold standard list if the corresponding laboratory value (hemoglobin, total calcium, pH and plasma potassium level) during the first 24 hours of admission was below the normal range as defined by the rules of the corresponding frame for the attribute "highly suspected."

In this study the value of adding an expert system was shown. The mean of the number of problems that exist in both the NEONATE and the

Gold Standard list but not in the physician problem list was calculated from 30 medical records.

3. Methods that prove that NEONATE produces more true positive problems and less false positive problems than the physician problem list:

This study tests to see if NEONATE's expert system is an improvement over the physician problem list. Here, in addition to the NEONATE admission problem list and in addition to the physician admission problem list, five lists were defined:

- a. The Gold Standard list which is the same list as it was defined on the previous section that describes the method for proving that NEONATE picks up problems missed by the physician problem list.
- b. A list that includes all problems that exist in the physician admission problem list and in the gold standard list.
- c. A list that includes all problems that exist in the physician admission problem list, but not in the gold standard list.
- d. A list that includes all problems that exist in the NEONATE admission problem list and in the gold standard list.
- e. A list that includes all problems that exist in the NEONATE admission problem list, but not in the gold standard list.

The true positive value of NEONATE was defined as the number of problems that exist in list d. The true positive value of the physician problem list was defined as the number of problems that exist in list b.

The false positive value of NEONATE was defined as the number of problems that exist in list e. The false positive value of the physician problem list was defined as the number of problems that exist in list c.

The false negative value of NEONATE was defined as the number of problems that results from the subtraction of the number of problems that determines the true positive value of NEONATE from the number of problems that exist in the gold standard list. In the same way the false negative value of the physician problem list was defined.

The true negative value of NEONATE was determined from the subtraction of the true positive value of NEONATE, the false negative value of NEONATE and the false positive value of NEONATE from 35 (which is the total number of problems NEONATE is able to diagnose). The same applies to the definition of the true negative value of the physician problem list.

To demonstrate that NEONATE performs better than the current physician problem list (or the current problem list that is typed into the CETUS system manually, since CETUS has no decision capabilities) it must be shown that the true positive value of NEONATE is greater than the true positive value of the physician problem list and the false positive value of NEONATE is smaller than the false positive value of the physician problem list.

Each of the four parameters (true positive, false positive, true negative and false negative) was calculated for the physician problem list and NEONATE in each of 30 charts. The results were averaged to obtain a single value for each of the parameters. From these values the sensitivity was calculated in each of the 30 records as

$$\text{True positives} / (\text{True positives} + \text{False negatives})$$

for NEONATE and for the physician problem list and the results were averaged. The specificity for both problem lists was calculated in each of the 30 records as

$$\text{True negatives} / (\text{True negatives} + \text{False positives})$$

and the results were averaged.

Results

1. Results of the agreement measure between NEONATE and the physician problem list:

Table 1 represents Kappa values of the 50 records. The mean of Kappa values is 0.662 with standard error of margin 0.004748. This confirms that the mean value of the Kappas is within a confidence interval of 0.014.

2. Results of the study that shows that NEONATE picks up problems that are missed by the physician problem list:

It was shown that there are problems that exist in the NEONATE admission problem list and in the gold standard list, but not in the physician admission problem list. A mean of 1.033 problems that fits this criteria was found from the study of the 30 records. This mean is significantly greater than 0 ($p < 0.01$). Appendix A represents the number of these problems in each of the 30 records.

3. Results of the study that NEONATE produces more true positive problems and less false positive problems than the physician problem list:

Table 2 represents the means of the four parameters (true positive values, true negative values, false positive values and the false negative values). The difference in the values between the NEONATE problem list and the physician problem list for each of the four parameters was significant ($p < 0.01$).

Table 1. Kappa Values of 50 Records

1. 0.71	26. 0.72
2. 0.79	27. 0.64
3. 0.72	28. 0.72
4. 0.45	29. 0.37
5. 0.87	30. 0.35
6. 0.76	31. 0.63
7. 0.45	32. 0.78
8. 0.76	33. 0.78
9. 0.72	34. 1.00
10. 0.72	35. 0.63
11. 0.41	36. 0.84
12. 0.63	37. 0.63
13. 0.45	38. 0.84
14. 0.21	39. 0.64
15. 0.84	40. 0.87
16. 0.45	41. 0.64
17. 0.72	42. 0.64
18. 0.21	43. 0.64
19. 0.78	44. 0.84
20. 0.53	45. 1.00
21. 0.64	46. 1.00
22. 0.52	47. 0.64
23. 0.72	48. 0.78
24. 0.37	49. 0.64
25. 0.72	50. 0.72

Table 2. Means of True Positive, False Positive, True Negative and False Negative Values of NEONATE Problem List and the Physician Problem List

	NEONATE PROBLEM LIST	PHYSICIAN PROBLEM LIST
True Positives mean	2.733	1.867
False Negatives mean	0.633	1.500
False Positives mean	0.900	0.300
True Negatives mean	30.733	31.333

On studying 30 charts it was shown that the mean of the true positive values of NEONATE was 2.733 while the mean of the true positive values of the physician problem list was 1.867 ($P < 0.01$). The assumption that the mean of the false positive values of NEONATE is smaller than the mean of the false positive values of the physician problem list proved to be wrong since the mean of the false positive values of NEONATE was 0.900 while the mean of the false positive values of physician problem list was 0.300 ($P < 0.01$).

Table 3 represents the means of the sensitivity and the specificity for both problem lists. Here also the difference between the two problem lists is significant ($p < 0.01$). Appendix B represents a table with the number of true positive, false positive, false negative, true negative problems and the sensitivity and the specificity values of NEONATE in each of the 30 records that were studied. Appendix C represents a table with the number of the true positive, false positive, false negative, true negative problems and the sensitivity and the specificity values of the physician problem list in each of the 30 records that were studied.

Discussion

Kappa of 0.662 confirms that there is a good agreement between the NEONATE admission problem list and the physician admission problem list. This in turn infers that NEONATE's problem list is close to the physician problem list.

As it was shown explicitly that there are problems that exist in the NEONATE admission problem list and in the gold standard list, but do not appear in the physician admission problem list, it can be concluded that NEONATE's computer inference process picks up problems that are currently missed by the physician problem list. In this study the comparison is done

Table 3. Sensitivities and Specificities of NEONATE Problem List and the Physician Problem List

	NEONATE PROBLEM LIST	PHYSICIAN PROBLEM LIST
Sensitivity	82.82%	53.77%
Specificity	97.19%	99.07%

with the unedited version of the admission problem list of NEONATE. In the future the problem list is expected to be edited by physicians; therefore, it seems that NEONATE expert system edited by clinicians should perform better than clinicians alone.

NEONATE performs better than the physician admission problem list in a sense that it diagnoses more true positive problems, but the price to be paid is an increase in false positive problems. The sensitivity of NEONATE is far better than the sensitivity of the physician problem list. It is well shown that NEONATE catches more true positive problems than the physician admission problem list. It is difficult to conclude by the results whether the specificity of NEONATE is less than the specificity of the physician admission problem list. The computer deals only with a finite number of diseases (35 diseases in this case). In real clinical practice a much larger number of problems ought to be considered, an event that might increase the specificity value of NEONATE. Nevertheless, the values as they stand now (97% for NEONATE and 99% for the physician problem list) can be considered as the same. By relying on the experiment that demonstrates that the false positive value of the physician problem list is smaller than the false positive value of NEONATE, it can be concluded that the physician problem list does better in not listing problems that do not exist in the gold standard list than does NEONATE. A major contribution to this gap lies in the fact that NEONATE lists all cases of "Infectious Pneumonia" that get the status "suspected" if tachypnea is present. Fourteen cases of "Infectious Pneumonia" were true positive in NEONATE and 16 cases were false positive in NEONATE. Sepsis, the other "suspected" status problem that makes NEONATE's problem list, is also listed whenever tachypnea is present, but in this case 28 were true positive and only 2 were false positives. Appendix D displays a histogram

that compares the number of the true positives between NEONATE and the physician problem list by problem, whereas Appendix E displays a histogram that compares the number of the false positives between NEONATE and the physician problem list by problem. From these histograms it can be concluded that the difference in the true positive values, which is spread among several problems, contributes to the improvement of NEONATE over the physician problem list.

The significantly larger number of true positives between NEONATE and the physician problem list by problem probably results from several variables. First, and perhaps most importantly, it may be from inadequate documentation on the part of the admitting neonatologist. Common problems such as hypoglycemia are so frequently managed as part of routine care that they may be overlooked on the problem list even though there is laboratory proof, which the NEONATE system is designed to document. Secondly, the approach to some problems may not conform to a "standard." Choosing a hematocrit to define anemia is an example. The standard number is 40, below which a neonate is often transfused. However, in the setting of having an infant who is sick on a ventilator, one neonatologist may choose to transfuse when the hematocrit drops below 45, never reach the "standard" number of 40, and never thinking to enter anemia as a problem, even though management indicated anemia was present. Another neonatologist may define anemia by whether a transfusion is needed. He or she may tolerate a hematocrit down to 35 without transfusing, and because this hematocrit is not reached, and no transfusion is given, anemia is not entered. Finally, neonates may often have one or more problems that manifest by the same symptoms. For example, an infant may be delivered prematurely because of maternal infection and have sepsis with pneumonia, but because

of the prematurity, also have hyaline membrane disease. There are two diseases causing pulmonary disease. Often, the mothers are treated with antibiotics prior to delivery, neonatal blood cultures never grow, and sepsis and pneumonia are eventually dropped as a diagnoses even though the patient received a 7-10 day course of antibiotics, and may have indeed had both. NEONATE is designed to list both diagnoses, but some neonatologists may elect to list only the "proven" diagnosis of hyaline membrane disease. This may explain why sepsis and infectious pneumonia generate a particularly large number of differences between NEONATE and the physician problem list.

CHAPTER IV

EVALUATION OF THE RADIOLOGICAL FRAMES OF THE KNOWLEDGE BASE

General Remarks

The analysis module of NEONATE consists of a knowledge base that has four components: (1) The initiator that looks for abnormal findings. When they are found, then the differential diagnosis of each abnormality is stored in a computer-maintained problem list with an attribute (status) "suspected"; (2) The computerized problem list itself; (3) The inference engine that contains medical knowledge for each of the diseases that the computer is able to diagnose. The knowledge is organized in the form of frames. In each frame there are rules. These rules can change attributes (status) of diseases (from "suspected" to "highly suspected," from "highly suspected" to "confirmed," from "confirmed" to "resolved") or can eliminate a disease from the problem list if no evidence is found to support the disease. If a test is missing and the inference engine cannot make any decision, an information request will be posted; (4) The physician acts as part of the knowledge base. He or she can change the attribute of each disease if he or she feels that the decision of the inference engine is wrong. A reporting module is also included to produce reports of the known findings and of the status of the problems.

In some diseases the decision is partly based on an x-ray evaluation. In order to incorporate the x-ray knowledge in the frame of a particular disease, the frame could either use a radiologist's evaluation or (in the absence of a radiologist) it can invoke a secondary frame (radiological frame) that helps a nonradiologist physician to evaluate the x-ray. Based on findings entered by the nonradiologist, each radiological frame can generate an attribute for a particular disease that is found in the problem list. The possible radiological attributes that a particular disease can have are: "possible," "suspected," "highly suspected," "confirmed," "cannot make any comments," and "absolutely not." This radiological attribute is used by the corresponding primary frame of the disease to generate a final status/attribute for the disease.

A group of chest x-ray data acquisition frames function as a part of the data acquisition module. The x-ray findings, which are entered into the computer by the nonradiologist physician, are used by the radiological frames of the inference engine to generate radiological attributes for the disease frames. For instance, the radiological attribute together with other clinical data, are used by a disease frame to assign a final attribute to that disease.

An evaluation of this system was planned, in order to determine the accuracy of the decision logic of the radiological frames and to prove, by using the radiological frames, the hypothesis that NEONATE helps the nonradiologist physicians in making a more accurate diagnosis.

Methods

NEONATE is capable of making a diagnosis of 35 neonatal diseases, 17 of which diseases use results of radiological frames. The 17 diseases are:

Heart Failure, Hyaline Membrane Disease, Aspiration Pneumonia, Pneumothorax, Pneumomediastinum, Infectious Pneumonia, Transient Tachypnea of the Newborn, Broncho-Pulmonary Dysplasia, Diaphragmatic Hernia, Esophageal Atresia or T-E Fistula, Pulmonary Hemorrhage, Chylothorax, Hemothorax, Infectious Pleural Effusion, Pneumopericardium, Diaphragmatic Paralysis, and Muscular Dystrophy.

Five pediatric radiologists were asked to read 13 chest x-ray films. Each of the films was selected from the hospital teaching file and is known to represent one or more of the 17 diseases. The diagnoses of each film from film 1 to film 13 are: Pneumomediastinum, Normal film, Pneumothorax, Transient Tachypnea of the Newborn, Hyaline Membrane Disease, Infectious Pneumonia, Broncho-Pulmonary Dysplasia, Aspiration Pneumonia, Infectious Pneumonia, Pulmonary Hemorrhage, Diaphragmatic Hernia, Normal film, and Transient Tachypnea of the Newborn. The radiologists were asked to list the chest x-ray findings for each film (this activity simulated entering those findings via the computer data acquisition frames). Afterwards they were given a list of the 17 diseases and were asked to assign an attribute to each of these 17 problems/diseases for each film. The computer logic excludes some of the attributes as unnecessary for certain diseases. For example Pneumothorax does not have an attribute of "possible," "suspected" or "highly suspected," and the computer can only produce the attribute of "confirmed," "absolutely not" or "cannot make any comments." All the radiological frames are able to say "cannot make any comments" when they are unable to reach a decision but only some frames are capable of complete exclusion of a problem ("Absolutely not"). In contrast to the computer system the radiologists were able to choose any of the six attributes for each of the problems during this evaluation.

The Kappa statistic [50,51], which is a measure of agreement beyond chance, was used to evaluate the agreement of the computer generated attributes with the attributes chosen by the radiologists. A Kappa statistic was calculated for each problem for each radiologist or group of radiologists (17 diseases X 5 radiologists = 85 Kappas). In this approach a Kappa for a given problem and a given radiologist was calculated from the set of computer generated attributes for the 13 x-ray films and the corresponding set of radiologist chosen attributes. (The findings used as input to the computer to produce its set of attributes were the findings of the particular radiologist under consideration.) A mean Kappa was calculated from the 85 individual Kappas.

Next, the Kappa statistic was used to evaluate the agreement of one radiologist with another. Ten subgroups, each including two radiologists, were obtained from the main group of the five pediatric radiologists. Kappa agreement was calculated for each of the subgroups in the same way as it was calculated for each individual radiologist given attribute and the computer generated attribute. A mean Kappa was calculated from the 170 individual Kappas (17 diseases X 10 subgroups).

Initially, Kappa calculations were based on a direct match between attributes for each of the problems ("suspected" of the radiologist decision with "suspected" of the computer decision, "confirmed" with "confirmed," etc.). At a second time, Kappa calculations were based on less rigid criteria in that one unit distance of difference on agreement between the computer and the radiologist was considered as a perfect agreement. Rigid one unit distance of difference on agreement was considered to be when the two attributes from both raters (computer and radiologist) were either "possible" and "suspected," "suspected" and "highly suspected," "highly suspected" and

"confirmed," "possible" and "cannot make any comments," and "cannot make any comments" and "absolutely not." Table 4 summarizes the different ways of matching computer attributes and radiologist attributes.

To evaluate the radiological frames as an aid to nonradiologist physicians in evaluating an x-ray, the gold standard radiological attribute list was developed. It consists of the consensus of the radiologists on the attribute for each film for each disease. The consensus was established by the following two rules:

1. If at least three of the radiologists agree on an attribute, that attribute is considered as the consensus.
2. If there is no majority on agreement on an attribute, the attribute that is considered as the median after ranking the attributes in an order from least likely to most likely (Absolutely not; Cannot make any comments; Possible; Suspected; Highly suspected; and Confirmed;) is chosen.

Eighteen pediatric residents were selected in order to test the computer's ability to evaluate x-ray films. Six of them were during their first year of residency, six during the second year and six during the third year. Each of them was asked to read the 13 x-ray films, to write his or her findings, and to assign attributes for each disease in exactly the same way that it was done by the radiologists. At a second time the residents' findings were entered to the computer in order to obtain a computer generated attribute list. Both attribute lists (the residents' and the computer's using the residents findings) were compared to the gold standard attribute list.

The comparison to the gold standard attribute list was achieved by using a scoring system that reflects the distance between the resident's attribute list or the computer's attribute list and the gold standard list. For

Table 4. Matching Criteria for Attributes

	ABS	CAN	POS	SUS	HSUS	CON
ABS	X Y	Y				
CAN	Y	X Y	Y			
POS		Y	X Y	Y		
SUS			Y	X Y	Y	
HSUS				Y	X Y	Y
CON					Y	X Y

where:

X - Direct match

Y - Match by one unit distance

ABS - Absolutely not

CAN - Cannot make any comments

POS - Possible

SUS - Suspected

HSUS - Highly suspected

CON - Confirmed

each chest x-ray film a score was obtained by summing across all 17 diseases as follows:

1. If an attribute matches perfectly with the gold standard list attribute for that particular disease in the same x-ray film the score is 0. If there is no match but each of the two attributes in both lists are "confirmed" and "highly suspected," "highly suspected" and "suspected," "suspected" and "possible," "possible" and "cannot make any comments," "cannot make any comments" and "absolutely not," or "absolutely not" and "possible," then a score of 1 is given. Otherwise, any other mismatch receives a score of 2. For each x-ray film the sum of the scores of the attributes is obtained as the final x-ray score.
2. The second way of measuring the distance does not take into account more than one unit distance of mismatch and scores all mismatches as 1. For each x-ray film the distance score is simply the number of attribute mismatches.

Results

Kappa was evaluated to measure the agreement for each radiologist between his own attribute list and the attribute list from the computer using his own findings. The mean Kappa for the five groups and 17 diseases was 0.648, when a perfect match was used and 0.848 when one unit distance was used (Figure 5). Mean Kappa agreement of the 10 subgroups of two radiologists and 17 diseases was shown to be 0.688 ($p > 0.01$) when a perfect match in the attributes was used for the ten subgroups. Figure 5 also shows the increase in Kappa to 0.898 ($p < 0.01$) when using the one unit distance in attributes.

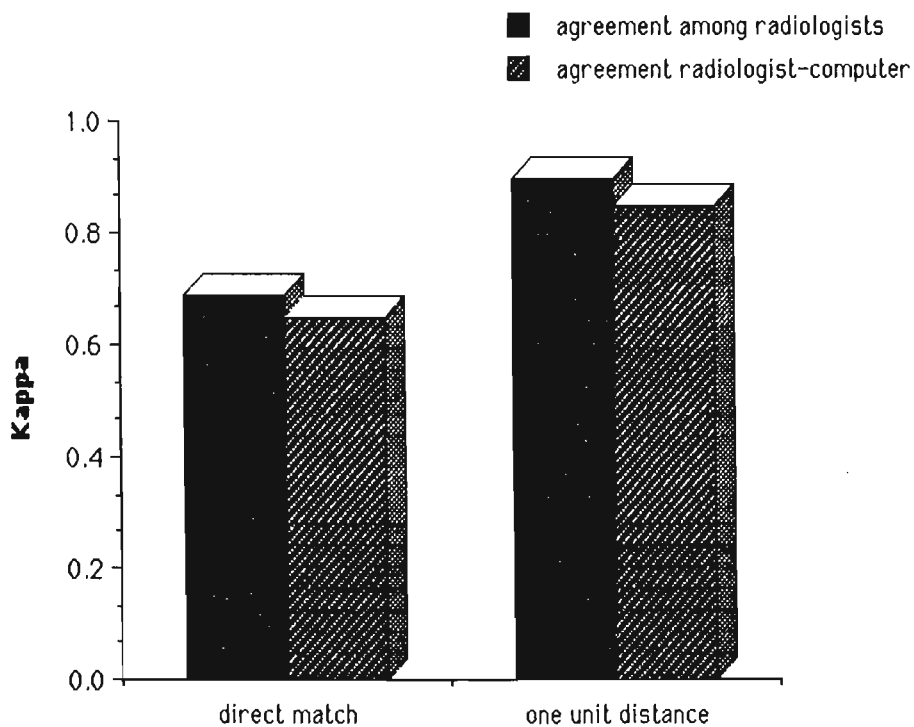


Figure 5. Mean of Kappa Agreement of the Ten Subgroups of Two Radiologists as Compared to the Mean of the Agreement Between Each Radiologist and the Computer's Output With His or Her Findings.

The mean of the distances, as explained in the method section, between the residents' attribute list without using the computer and the gold standard attribute list was 7.299 while the mean of the distances between the computer's attribute list using the resident data and the gold standard attribute list was 6.637 ($p < 0.01$). The number of attributes that do not match between the residents' attribute list without using the computer and the gold standard attribute list was 6.402 while the numbers of attributes that do not match between the computer's list with the resident data and the gold standard list was 5.248 ($p < 0.01$). Table 5 displays the variation of the distance by residency level. Table 6 displays variation of the number of items that do not match by residency levels.

The means of the distances without and with the computer in the first year of residency are significantly different ($p < 0.01$). The means of the number of items that do not match without and with the computer in the first year of residency are also significantly different ($p < 0.01$). The same applies to the second year of residency. The means of the distances and the means of the number of items that do not match without and with the computer respectively in the third year of residency are not significantly different ($p > 0.01$).

Figure 6 is a histogram of the variation of the distances by residency levels. Figure 7 is a histogram that displays the variation in the number of attributes that do not match by residency level. In these figures each residency level is represented by 4 bars. The left one represents the mean of the distance or the mean of the number of attributes that do not match (distance in Figure 6 and number of attributes in Figure 7) between the radiologists' attribute list and the consensus attribute list which is the gold standard list. These values are clearly different from the global

Table 5. Variation of Distance by Level of Residency

Residency level	Resident's distance	Computer's distance
PL - 1	8.128	6.551
PL - 2	7.500	6.462
PL - 3	6.269	6.897

Table 6. Variation of Number of Attributes that do not Match by Level of Residency

Residency level	Resident's number	Computer's number
PL - 1	7.128	5.282
PL - 2	6.577	5.077
PL - 3	5.500	5.385

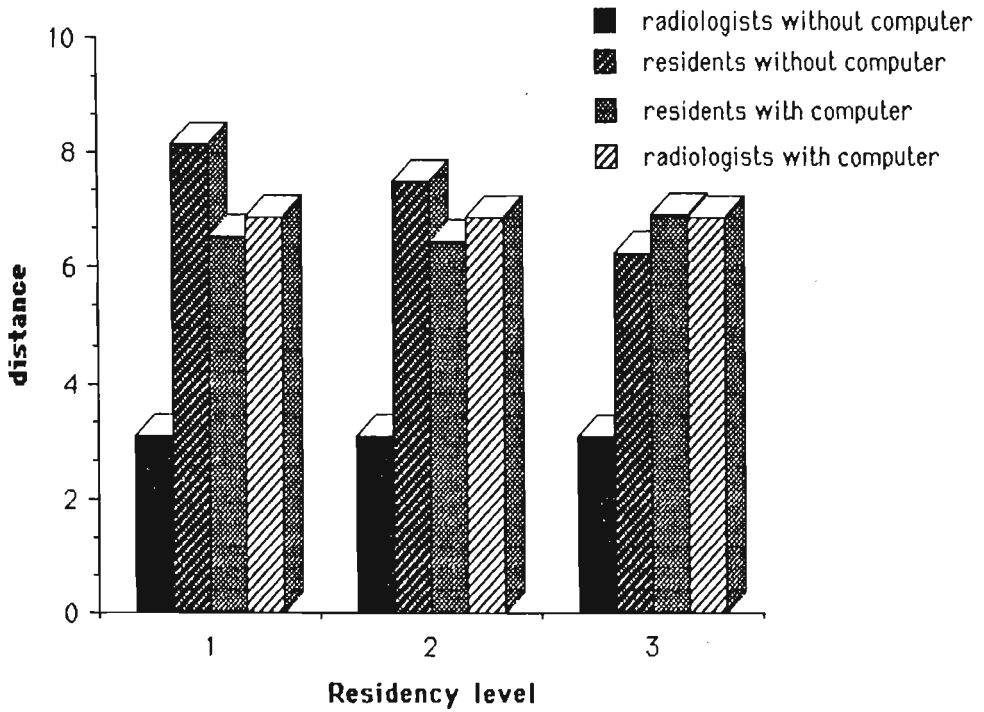


Figure 6. Variation of Distance by the Level of Residency Without and With the Computer. The Three Levels are Compared to the Radiologists' Performance

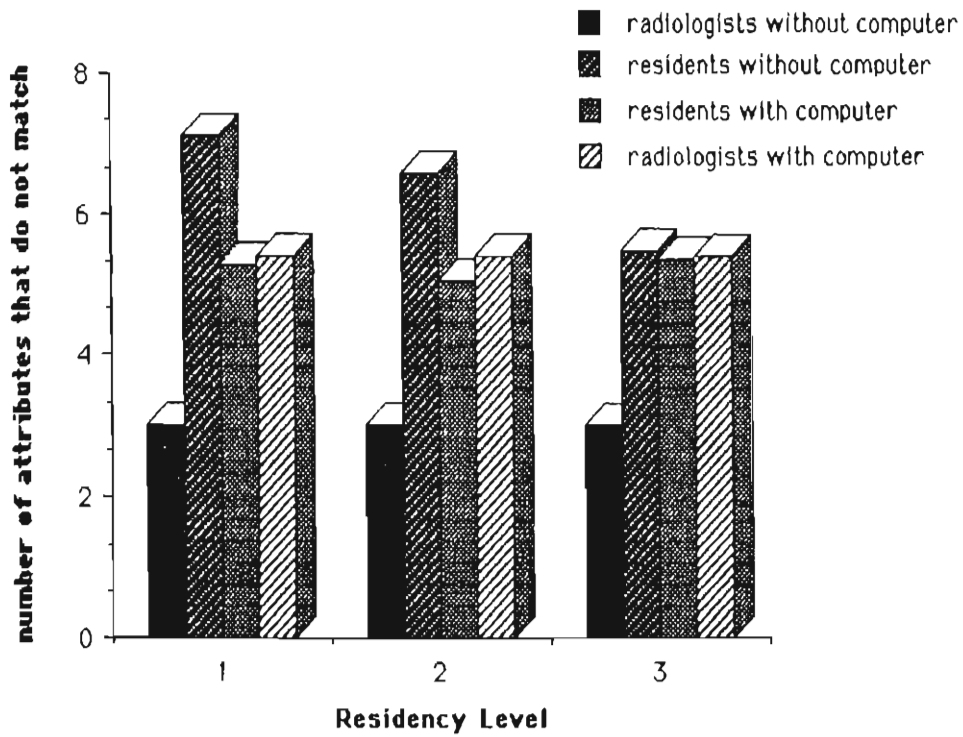


Figure 7. Variation of the Number of Attributes that do not Match by the Level of Residency Without and With the Computer. The Three Levels are Compared to the Radiologists' Performance.

corresponding values of the residents. (The distance of the radiologist without computer is 3.086 and the number of attributes that do not match between the radiologists attribute list and the consensus is 2.999.) The right bar in the two figures represents the mean of the distance or the mean of the number of attributes that do not match of the computer's attribute list using the radiologists' findings to the gold standard list. The distance was 6.860 and it is not significantly different from the global distance of the residents ($P > 0.01$). The number of attributes that do not match was 5.399. Also, it is not significantly different from the mean of the number of attributes that do not match of the global residents ($P > 0.01$).

Discussion

Validation of the radiological frames was performed by comparing the agreement measure between each radiologist's attribute list without the computer and the computer's attribute list using that radiologist's findings to the agreement among the radiologists themselves without the computer. Although Kappa obtained when the computer was used is slightly lower than Kappa of the radiologists among themselves, statistically it was shown that the two means are not significantly different. This signifies that the radiological frames perform the decisions about as well as the radiologists.

A measure of distance between two attribute lists was defined. It was shown that the mean distance of attribute lists of first year residents, obtained without using the computer, to the consensus attribute list of all five radiologists was significantly higher than the distance of the list obtained by the computer using their findings. The same results were obtained for the second year residents. The same applies to the number of attributes that do not match. These were significantly lower when the first

and second year residents used the computer. The distance and the number of attributes that do not match were statistically the same for the third year residents. These results suggest that the computer can help novices to reach a better diagnosis, but is not yet helpful for more expert physicians. It suggests that the computer may be used as an educational tool. If it is used as an aid tool in real clinical settings, the less experienced physicians may benefit more from its usage.

The findings were analyzed to observe the relationship between the distance and the findings. The consensus of the findings among the five radiologists was determined in the same way that the consensus of the attributes for each problem was determined. The number of mismatches between each radiologist's finding list and the consensus finding list was counted. The mean of the number of mismatches between the radiologists' finding list and the consensus finding list was 0.1519. The means for the first, second and third year of the residents were 0.3290, 0.2492 and 0.2233 respectively. The distance of the computer attribute list using the radiologists' findings is statistically the same as the distance of the global computer attribute list using the residents' findings. This means that the distances of the attribute lists that were obtained when the computer was used are unrelated to the findings, but mainly it is due to the algorithm. The knowledge base has undergone two iterations. It seems that further improvement in the rules of the knowledge base will decrease the distances of the computer attribute lists from the consensus attribute list.

CHAPTER V

EVALUATION OF NEONATE AS A POTENTIAL PROMPTING DEVICE FOR MEDICAL FINDINGS AND PROBLEMS

General Remarks

A survey to determine if physicians need to be prompted was conducted. A system that is designed to be an extension of the physician's memory must have prompting capabilities. The study describes how well the physicians perform with a prompting mechanism. It describes the effect of prompting on the capture of x-ray findings and the effect on the physician diagnosis list. This survey proves the hypothesis that the NEONATE is a potentially effective prompting device.

Methods

Assessment of NEONATE as a potential prompting device was accomplished, using a paper based simulation of the Neonate system. Twenty-eight pediatric residents (six first year residents, 12 second year residents and 10 third year residents) were asked to read the 13 chest x-ray films and to write on a blank sheet of paper the findings which they detected and a differential diagnosis for each film. Subsequently, they were given a questionnaire simulating the data entry screens. They were asked to select the true findings and give an attribute to each of the seventeen problems (diseases). The findings captured in the freetext format and those

captured on the questionnaire were counted. The same approach was used for the problems. A problem was considered as existent if the resident mentioned it on the blank sheet or listed it as "possible," "suspected," "highly suspected" or "confirmed" on the questionnaire.

The difference between the number of true positive findings without being prompted and the number of true positive findings with a prompting mechanism served as the measure of help by prompting. Also the difference in the number of true positive problems with and without prompting was evaluated.

A finding was considered to be the gold standard, if at least three of the radiologists agreed on the existence of the finding. In one case where this definition did not provide a gold standard, the radiologists' choices were arranged in hierarchical order and the median choice of the five radiologists was selected. This case was the finding that describes the expansion of the lungs where the radiologists had to choose one out of five options: hypoexpanded, focal; hypoexpanded, diffuse; normal; hyperexpanded, focal; and hyperexpanded, diffuse. Two radiologists concluded this finding as hypoexpanded, one concluded it as normal, and two concluded it as hyperexpanded. It was decided that the "truth" for this particular case would be the normal. Likewise, a problem was considered as a true positive problem if at least three of the five radiologists gave to that problem an attribute of possible, suspected, highly suspected or confirmed.

"Findings true positive ratio" and "Problems true positive ratio" were defined as the number of true positive findings or problems over the total number of findings or problems respectively. The "Findings true positive ratio" and the "Problems true positive ratio" were measured for each resident on each film.

Results

Counting the number of findings that the pediatric residents identified without being prompted resulted in a mean of 2.83; with a prompting mechanism (the questionnaire) it was 4.604 ($p < 0.01$). Table 7 and Figure 8 display the means of the number of findings by level of residency. Table 8 and Figure 9 display the means of the number of findings by chest x-rays film which are significantly different ($p < 0.01$). The mean number of findings did not differ significantly by level of residency ($p > 0.05$) but did differ significantly by chest x-ray film ($p < 0.01$).

The mean of the number of the true positive findings without being prompted was 1.585 while the mean with prompting was 2.412 ($p < 0.01$). These means were not statistically different when they were compared by level of residency but were statistically different when they were compared by chest x-ray film. The mean of the "Findings true positive ratio" without prompting was 0.537 while the mean of the "Findings true positive ratio" with prompting was 0.483 ($p < 0.01$).

Counting the number of the problems that the pediatric residents identified without being prompted resulted in a mean of 1.81 while the mean of the number of problems with a prompting mechanism was 4.887 ($p < 0.01$). Table 9 and Figure 10 display the means of the number of problems by level of residency ($p > 0.05$). Table 10 and Figure 11 display the means of the number of problems by chest x-ray film which are significantly different ($p < 0.01$).

The mean of the number of the true positive problems without being prompted was 1.178 while the mean with prompting was 2.359 ($p < 0.01$). These means were not statistically different when they were compared by

Table 7

Means of Number of Findings Identified
by Level of Residency

Level of residency	without prompting	with prompting
PL - 1	2.900	4.785
PL - 2	2.878	4.615
PL - 3	2.838	4.282

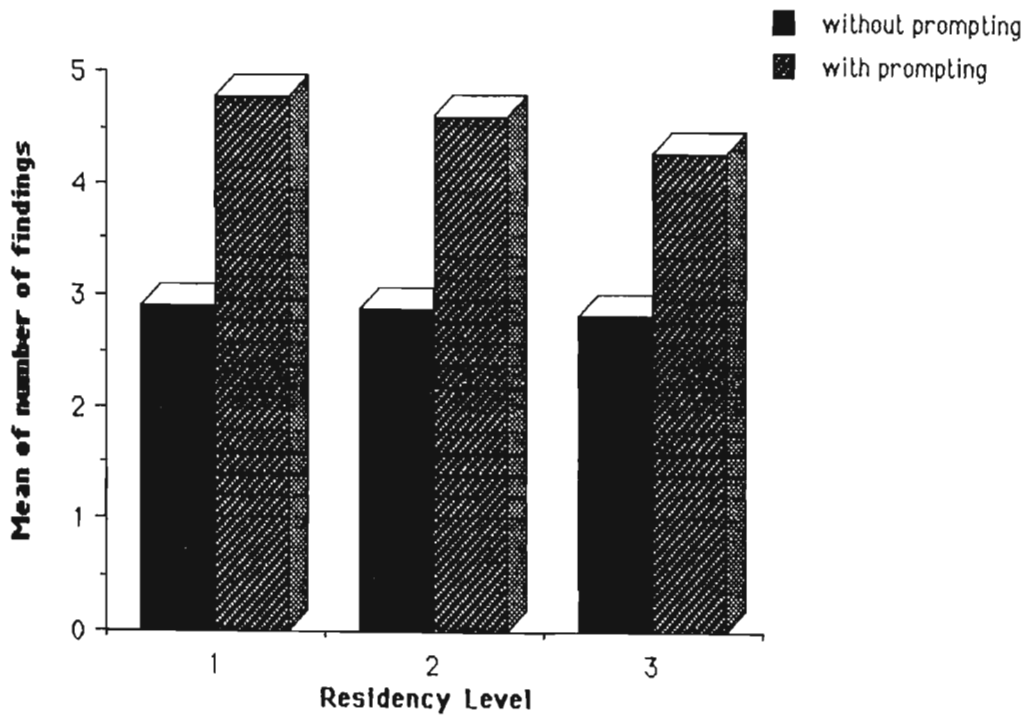


Figure 8. Mean of Number of Findings Identified by Level of Residency

Table 8
 Means of Number of Findings Identified
 by Chest X-ray Film

Chest x-ray film	without prompting	with prompting
Film 1	3.821	6.571
Film 2	2.107	2.321
Film 3	2.857	4.893
Film 4	2.536	3.143
Film 5	2.750	4.750
Film 6	2.321	3.750
Film 7	4.857	7.464
Film 8	1.964	3.321
Film 9	3.786	6.071
Film 10	3.357	6.214
Film 11	3.857	7.643
Film 12	1.107	0.964
Film 13	1.571	2.750

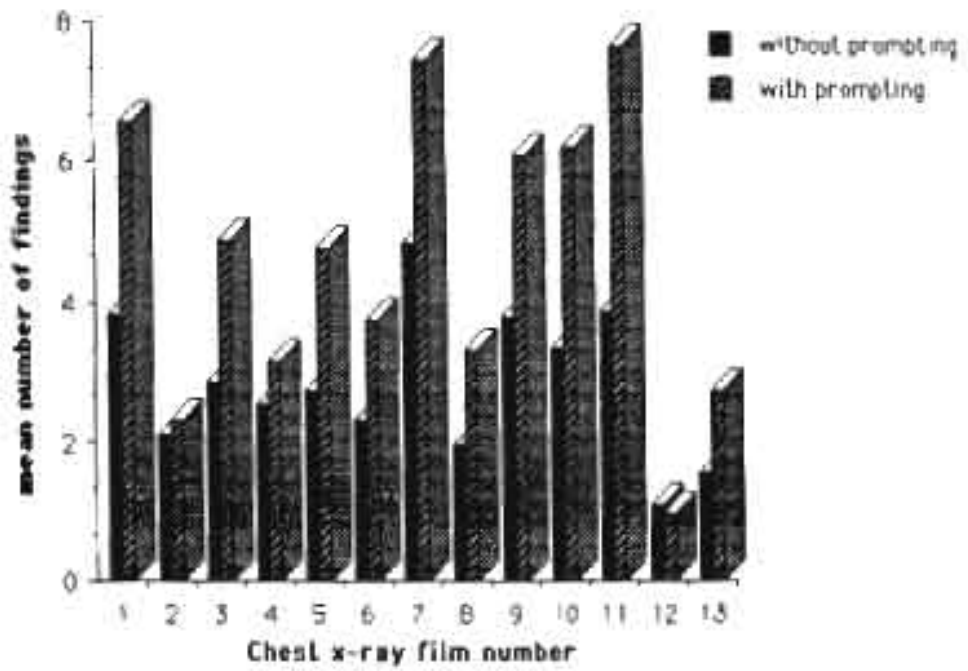


Figure 9. Mean of the Number of Findings Identified by Chest X-Ray Film

Table 9

Means of Number of Problems Identified
by Residency Level

Level of residency	without prompting	with prompting
PL - 1	1.662	5.085
PL - 2	1.846	5.026
PL - 3	1.987	4.282

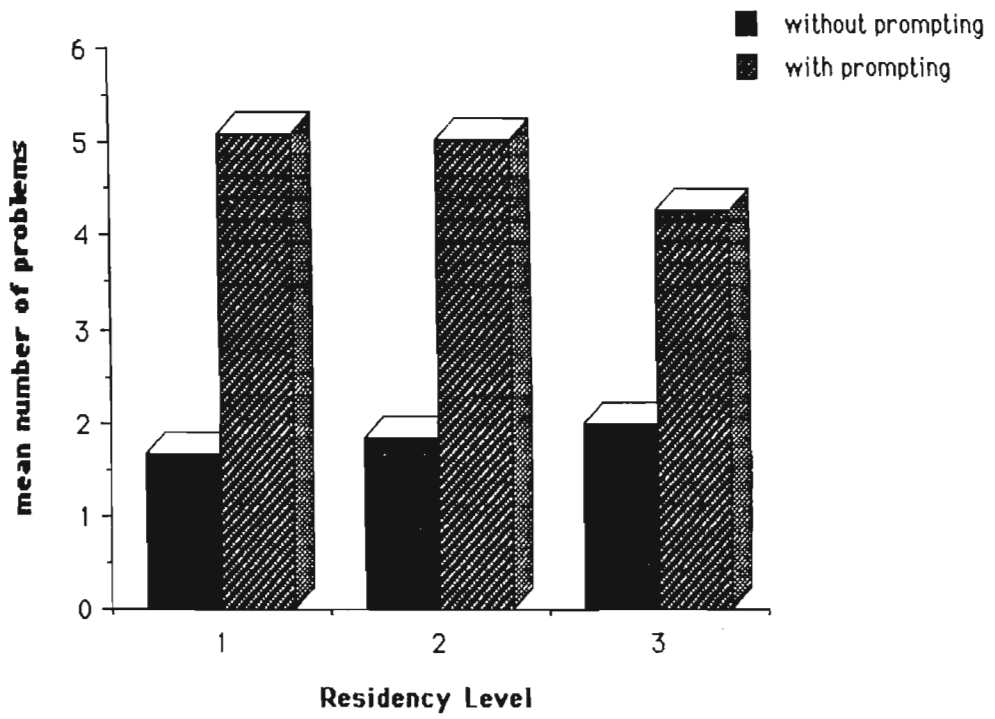


Figure 10. Mean of the Number of Problems Identified by Level of Residency

Table 10
Means of Number of Problems Identified
by Chest X-ray Film

Chest x-ray film	without prompting	with prompting
Film 1	2.500	6.607
Film 2	2.036	3.821
Film 3	1.286	4.571
Film 4	2.393	5.179
Film 5	1.500	4.214
Film 6	2.286	5.643
Film 7	2.143	5.536
Film 8	1.179	4.214
Film 9	2.107	5.500
Film 10	1.786	7.071
Film 11	1.429	4.643
Film 12	0.929	1.607
Film 13	1.964	4.929

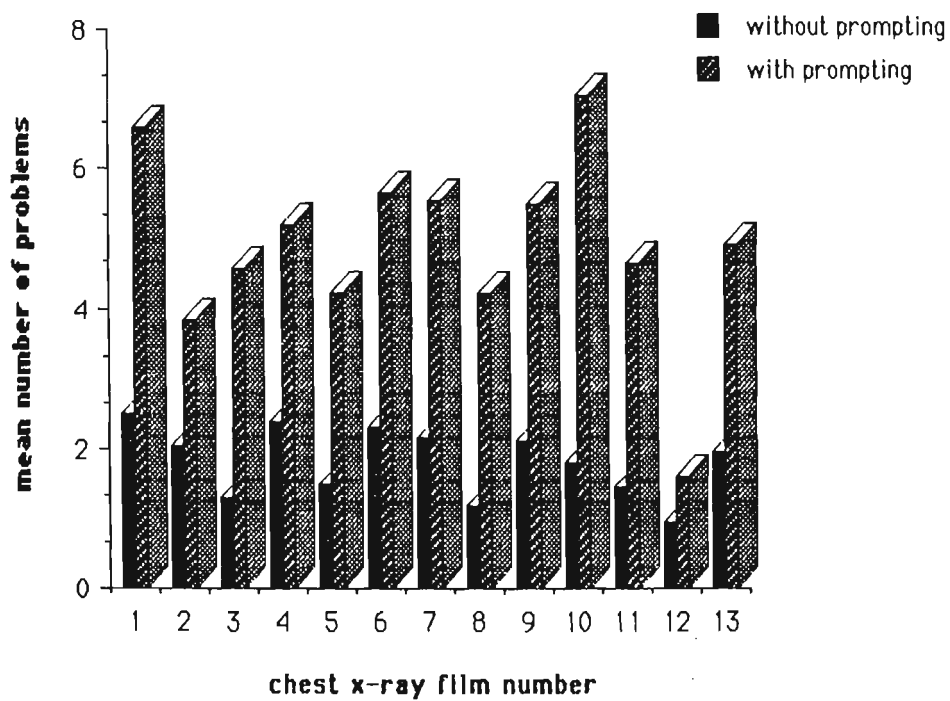


Figure 11. Mean of the Number of Problems Identified by Chest X-Ray Film

level of residency but were statistically different when they were compared by chest x-ray film. The mean of the "Problems true positive ratio" without prompting was 0.644 while the mean of the "Problems true positive ratio" with prompting was 0.515 ($p < 0.01$).

Discussion

This study proves the potential effectiveness of NEONATE as a prompting device. It was asserted [1] that physicians can never remember enough, they can never retrieve from memory accurately and quickly enough to cover all possibilities, so they distort the original problems or lose sight of them altogether. The computer is seen as an extension of the physician's memory much as automobile is an extension of one's muscles. This was concluded on the number of findings and problems that were collected with and without a prompting mechanism. The number of true positive findings and the number of true positive problems are increased with prompting, but the rate of "Findings true positive ratio" as well the rate of "Problems true positive ratio" has a tendency to decrease.

CHAPTER VI

CONCLUSIONS REGARDING NEONATE'S POTENTIAL AS A PRACTICAL SYSTEM

A prototype of an expert system that is embedded inside a Hospital Information System has been developed. This system has three advantages over old methods of expert systems:

1. This system is not a stand alone system as many others, but is an integral part of the Hospital Information System. This permits direct access to all data that exist in the computerized medical record and eliminates unnecessary manual data entry that is sometimes needed in stand alone expert systems in order to give these systems the ability to make decisions.
2. The system provides alerting and prompting capabilities.
3. It has all the features of an expert system and a Hospital Information System. It has the ability to acquire data, generate reports and make decisions.

The knowledge base is designed in a very simple way. It is designed using a rule based mechanism without relying on numerical probabilities. The knowledge base was organized in this way to accommodate the physicians that cooperated during this development. These physicians were reluctant to use the Bayesean approaches common to the HELP system. To

overcome this problem and to be able to deal with uncertainty, probabilities were used in a textual form rather than numerical. The modular way that the knowledge base and all other components are constructed permits easy modifications and expansions with no need to modify the core program.

The knowledge base was designed only to aid the physician in pursuing the correct diagnosis. The literature indicates that replacing the physician's diagnostic abilities with a computer is not feasible with conventional hardware and software. Medicine is an art, with no clear criteria for the diagnostic process, and cannot be duplicated by computers. Computers can help physicians only with their human frailties such as a shortage of memory. So, the goal of the expert part of the system has been to serve only in alerting and reminding. Although the results that were obtained are encouraging, an expert system cannot be used as a substitute for the diagnostic process of physicians. The real value of it is in reminding. This is why the decisions have to be reviewed by the health care providers (physicians or nurses) who must edit the problem list at regular intervals.

It was hypothesized that NEONATE's computer inference process would occasionally pick up problems that are missed on the current, physician composed problem list (CETUS admission summary). The results would be an admission problem list that lacked none of what is currently there while adding problems sometime overlooked (assuming of course, that the physicians would edit the problem list, would add their problem to it, and would not delete valid problems posted by the inference engine). While it was hypothesized that NEONATE's computer generated problem list would have more true positives than the current problem list (CETUS), it was expected that this hypothesis would be rejected. The real surprise in this study has been that in the 30 patient charts considered, NEONATE's inference

engine has found a total of 31 problems missed by the physicians (absent from the physician admission problem list) whereas the physicians have only found 5 problems missed by NEONATE's inference engine. While the author claims a clear victory for NEONATE over the physician in producing an admission problem list, he does not conclude that computers are better at diagnosis than physicians. The results are more likely to be explained by the difficulty in conscientiously completing documentation in the busy world of clinical medicine. In any case there is a powerful argument to be made for an expert system being embedded in a Hospital Information System.

It was hypothesized that NEONATE's radiological frames could help physicians not trained as radiologists to interpret x-ray films. While this hypothesis has been affirmed in the case of first and second year residents, it was rejected for third year residents. The data for radiologists with and without the computer show considerable room for improvement of the computer algorithm. It remains an open question as to whether or not additional effort on the computer algorithm would make enough difference for them to be useful to a neonatologist.

It was hypothesized that prompting helps physicians to find more of the information in an x-ray film. The hypothesis was affirmed for all physicians. The drawback with prompting is that it not only helps users to find information that is in a film, but it sometimes causes them to "see" things that are not present.

The road until this system will be fully operational is still long and filled with obstacles. This prototype deals only with a small aspect of the routine work in the Neonatal unit (admission process). Before it can be tested in a real clinical environment all other features of the routine clinical work in a Neonatal unit, such as interim reporting and the discharge process

should be added to the system. Many more diseases and more alerts should be incorporated in order to make the system functional. The major obstacle will be to get the physicians using the system on a routine basis and to get all physicians treating the patients to agree on the computer decisions. It might be that as Neonatology is a limited subspecialty with a very limited problem set, physicians will not benefit from using this system. The residents and the inexperienced physicians might benefit more from the system. Such a prototype, if developed in other departments such as internal medicine or general pediatrics, might be more beneficial due to the larger number of problems with which clinicians deal. Further development and expansion of NEONATE in a real clinical setting is necessary in order to establish its usability.

APPENDIX A

NUMBER OF PROBLEMS THAT EXIST NEONATE PROBLEM LIST AND IN THE
GOLD STANDARD PROBLEM LIST, BUT NOT IN THE PHYSICIAN
PROBLEM LIST IN EACH OF THE 30 RECORDS

Record	Problems in NEONATE and in Gold Standard, but not in CETUS
1	1
2	0
3	2
4	2
5	1
6	1
7	1
8	1
9	1
10	2
11	0
12	0
13	1
14	2
15	0
16	1
17	0
18	1
19	1
20	1
21	1
22	0
23	2
24	1
25	2
26	1
27	2
28	1
29	1
30	1

APPENDIX B

NUMBER OF TRUE POSITIVE, FALSE POSITIVE, FALSE NEGATIVE, TRUE
NEGATIVE PROBLEMS AND THE SENSITIVITY AND THE
SPECIFICITY VALUES OF NEONATE IN EACH
OF THE 30 RECORDS

Record	TP	FP	FN	TN	Sensitiv.	Specific.
1	1	2	1	31	0.5000	0.9394
2	1	2	1	31	0.5000	0.9394
3	5	0	2	28	0.7143	1.0000
4	3	0	1	31	0.7500	1.0000
5	4	0	0	31	1.0000	1.0000
6	2	1	0	32	1.0000	0.9697
7	3	1	0	31	1.0000	0.9688
8	2	1	2	30	0.5000	0.9677
9	3	1	0	31	1.0000	0.9688
10	3	4	0	28	1.0000	0.8750
11	2	0	1	32	0.6667	1.0000
12	3	1	0	31	1.0000	0.9688
13	1	2	0	32	1.0000	0.9412
14	3	2	0	30	1.0000	0.9375
15	3	1	1	30	0.7500	0.9677
16	3	0	0	32	1.0000	1.0000
17	1	1	3	30	0.2500	0.9677
18	2	1	2	30	0.5000	0.9677
19	3	0	0	32	1.0000	1.0000
20	3	1	0	31	1.0000	0.9688
21	3	1	1	30	0.7500	0.9677
22	2	1	0	32	1.0000	0.9697
23	2	1	0	32	1.0000	0.9697
24	2	1	1	31	0.6667	0.9688
25	4	0	1	30	0.8000	1.0000
26	4	0	0	31	1.0000	1.0000
27	5	2	0	28	1.0000	0.9333
28	3	0	0	32	1.0000	1.0000
29	3	0	1	31	0.7500	1.0000
30	3	0	1	31	0.7500	1.0000

TP - True positive
 FP - False positive
 FN - False negative

TN - True negative
 Sensitiv. - Sensitivity
 Specific. - Specificity

APPENDIX C

NUMBER OF TRUE POSITIVE, FALSE POSITIVE, FALSE NEGATIVE, TRUE
NEGATIVE PROBLEMS AND THE SENSITIVITY AND THE
SPECIFICITY VALUES OF THE PHYSICIAN
PROBLEM LIST IN EACH
OF THE 30 RECORDS

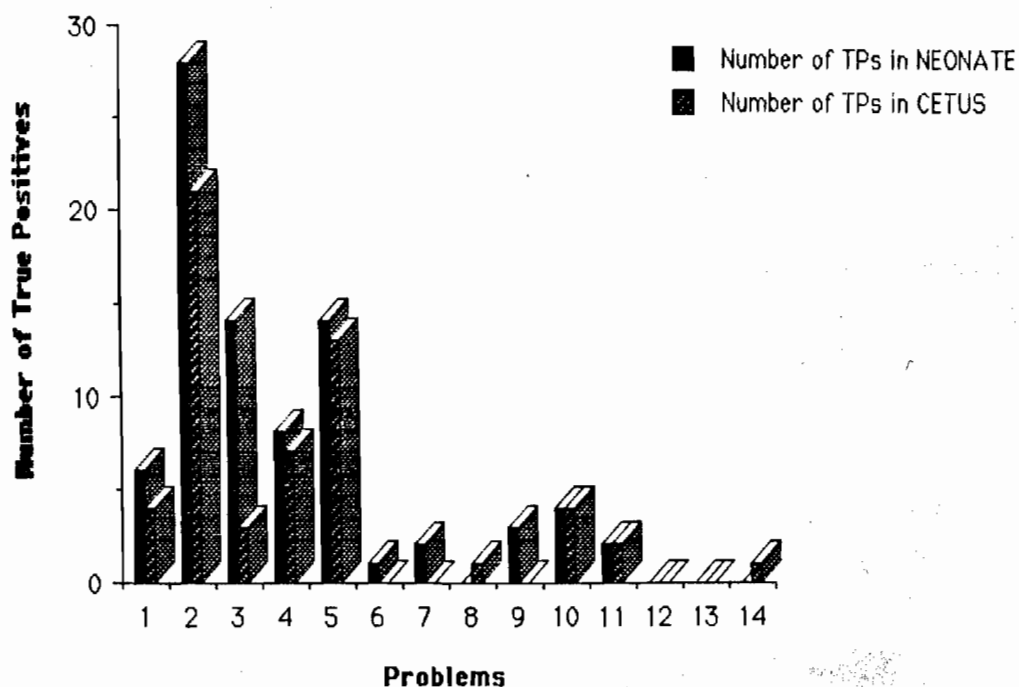
Record	TP	FP	FN	TN	Sensitiv.	Specific.
1	1	0	1	33	0.5000	1.0000
2	1	1	1	32	0.5000	0.9697
3	5	0	2	28	0.7143	1.0000
4	1	0	3	31	0.2500	1.0000
5	3	0	1	31	0.7500	1.0000
6	1	1	1	32	0.5000	0.9697
7	2	1	1	31	0.6667	0.9688
8	1	0	3	31	0.2500	1.0000
9	2	0	1	32	0.6667	1.0000
10	1	0	2	32	0.3333	1.0000
11	2	1	1	31	0.6667	0.9688
12	3	1	0	31	1.0000	0.9688
13	0	1	1	33	0.0000	0.9706
14	1	0	2	32	0.3333	1.0000
15	3	0	1	31	0.7500	1.0000
16	2	0	1	32	0.6667	1.0000
17	1	1	3	30	0.2500	0.9677
18	2	0	2	31	0.5000	1.0000
19	2	0	1	32	0.6667	1.0000
20	2	0	1	32	0.6667	1.0000
21	2	0	2	31	0.5000	1.0000
22	2	0	0	33	1.0000	1.0000
23	0	1	2	32	0.0000	0.9697
24	1	1	2	31	0.3333	0.9688
25	2	0	3	30	0.4000	1.0000
26	3	0	1	31	0.7500	1.0000
27	3	0	2	30	0.6000	1.0000
28	2	0	1	32	0.6667	1.0000
29	2	0	2	31	0.5000	1.0000
30	3	0	1	31	0.7500	1.0000

TP - True positive
 FP - False positive
 FN - False negative

TN - True negative
 Sensitiv. - Sensitivity
 Specific. - Specificity

APPENDIX D

COMPARISON OF THE NUMBER OF TRUE POSITIVES BETWEEN NEONATE
PROBLEM LIST AND THE PHYSICIAN PROBLEM LIST BY PROBLEM

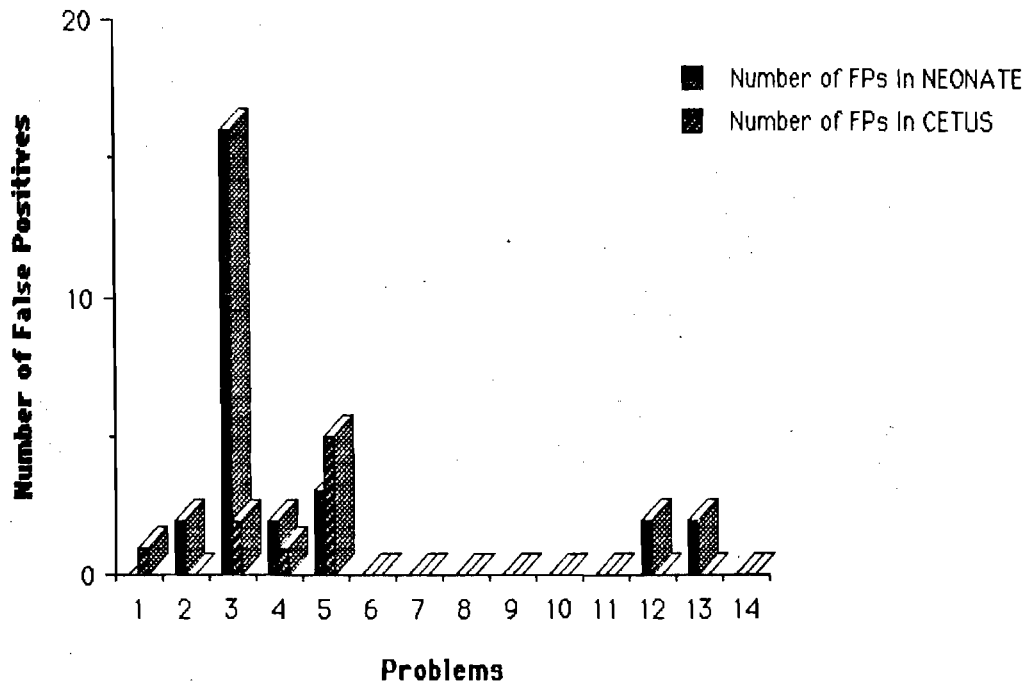


problems:

- 1 - Transient Tachypnea of the Newborn
- 2 - Sepsis
- 3 - Infectious Pneumonia
- 4 - Aspiration Pneumonia
- 5 - Hyaline Membrane Disease
- 6 - Metabolic Acidosis
- 7 - Hypokalemia
- 8 - Septic Shock
- 9 - Anemia
- 10 - Pneumothorax
- 11 - CNS Depression
- 12 - Subarachnoid Hemorrhage
- 13 - Intraventricular Hemorrhage
- 14 - Persistent Pulmonary Hypertension

APPENDIX E

COMPARISON OF THE NUMBER OF FALSE POSITIVES BETWEEN NEONATE
PROBLEM LIST AND THE PHYSICIAN PROBLEM LIST BY PROBLEM



problems:

- 1 - Transient Tachypnea of the Newborn
- 2 - Sepsis
- 3 - Infectious Pneumonia
- 4 - Aspiration Pneumonia
- 5 - Hyaline Membrane Disease
- 6 - Metabolic Acidosis
- 7 - Hypokalemia
- 8 - Septic Shock
- 9 - Anemia
- 10 - Pneumothorax
- 11 - CNS Depression
- 12 - Subarachnoid Hemorrhage
- 13 - Intraventricular Hemorrhage
- 14 - Persistent Pulmonary Hypertension

REFERENCES

1. Weed LL. Technology is a link, not a barrier, for doctor and patient. *Mod Hosp*, February 1970; 80 - 83.
2. Weed LL. What physicians worry about: how to organize care of multiple-problem patients. *Mod Hosp*, June 1988; 90 - 94.
3. Williamson JW, German SG, Weiss R, Skinner EA, Bowes III F. Health science information management and continuing education of physicians. *Ann Intern Med* 1989; 110: 151 - 160.
4. Brent CJ. *Quality Management for health care delivery*. The Hospital Research and Educational Trust, Chicago Illinois, 1989.
5. Davis R, Buchanan B, Shortliffe E. Production rules as a representation for a knowledge-based consultation program. *Artif Intell* 1977; 8: 15 - 45.
6. Shortliffe EH, Buchanan BG. A Model of inexact reasoning in medicine. *Math Biosci* 1975; 23: 351 - 379.
7. Kulikowski CA, Weiss SM. Representation of expert knowledge for consultation: The CASNET and EXPERT projects. In: Szolovits P, ed. *Artificial intelligence in medicine*. Boulder, Colorado: Westview Press, 1982: 21 - 55.
8. Pauker SG, Gorry SA, Kassirer JP, Schwartz WB. Towards the simulation of clinical cognition. Taking a present illness by computer. *Am J Med* 1976; 60: 981 - 996.
9. Patil RS, Szolovits P, Schwartz WB. Modeling knowledge of the patient in acid-base and electrolyte disorders. In: Szolovits P, ed. *Artificial intelligence in medicine*. Boulder, Colorado: Westview Press, 1982: 191 - 226.
10. Miller AM, Pople Jr HE, Myers JD. An experimental diagnostic consultant for general internal medicine. *N Engl J Med* 1982; 307: 468 - 476.

11. Aldasng KP, Kolorz G, Scheithauer W, Grabner H. Approach to a hospital-based application of a medical expert system. *Med Inform* 1986; 11 (3): 205 - 223.
12. Warner HR, Haug P, Bouhaddou O, et al. Iliad as an expert to teach differential diagnosis. *Proc SCAMC* 1988: 371 - 376.
13. Warner HR, Olmsted CM, Rutherford BD. HELP - A program for medical decision-making. *Comput Biomed Res* 1972; 5: 65 - 74.
14. Pryor TA, Gardner RM, Clayton PD, Warner HR. The HELP system. *J Med Sys* 1983; 7 (2): 87 - 101.
15. Pryor TA. The HELP medical record system. *MD Comput* 1988; 5 (5): 22 - 33.
16. Pople HE Jr. Heuristic methods for imposing structure on ill-structured problems. The structuring of medical diagnosis. In: Szolovits P, ed. *Artificial intelligence in medicine*. Boulder, Colorado: Westview Press, 1982: 119 - 190.
17. Weiss SM, Kulikowski CA. Expert. A system for developing consultation models. *Proc sixth IJCAI, Tokyo, 1979*: 942 - 947.
18. Kingsland LC III, Lindberg DA, Sharp GC. Anatomy of a knowledge-based consultant system. *MD Comput* 1986; 3 (5): 18 - 26.
19. Blois MS, Tuttle MS, Sheretz DD. Reconsider: a program for generating differential diagnosis. *Proc SCAMC* 1981: 263 - 266.
20. Broering NC, Corn M, Ayers WR, Mistry P. Implementing Reconsider, a diagnostic prompting computer system, at the Georgetown University Medical Center. *Bull Med Libr Assoc* 1988; 76 (2): 155 - 158.
21. Swett HA, Miller PL. ICON: A computer-based approach to differential diagnosis in radiology. *Radiology* 1987; 163: 555 - 558.
22. Aikins JS, Kunz JC, Shortliffe EH, Fallat RJ. PUFF: an expert system for interpretation of pulmonary function data. *Comput Biomed Res* 1983; 16: 199 - 208.
23. Frayer WW. Patient data management in neonatal intensive care. *Clin Perinatol* 1980; 7 (1): 145-154.

24. Van der Lei J, Derksen-Samson JF. Neonatal ICU system: experience with AIDA. *Comput Methods Programs Biomed* 1987; 25: 315 - 320.
25. Janik DS, Swarner OW, Henriksen KM, Wyman ML. A computerized single entry system for recording and reporting data on high-risk newborn infants. *J Pediatr* 1978; 93: 519-523.
26. Janik DS, Swarner OW, Henriksen KM, Wyman ML. Computerized newborn intensive care data recording and reporting II; an online system. *J Pediatr* 1979; 94: 326 - 330.
27. Janik DS, Sharp EM, Forbush L, Wyman ML, Jung AL. Computerized newborn intensive care data recording, reporting and research III; a practical microcomputer system. *J Pediatr* 1980; 97: 497 - 500.
28. Barness LA, Tunnessen WW Jr, Worley WE, Simmons TL, Ringe TBK. Computer-assisted diagnosis in pediatrics. *Am J Dis Child* 1974; 127: 852 - 858.
29. Swender PT, Tunnessen WW Jr, Oski FA. Computer-assisted diagnosis. *Am J Dis Child* 1974; 127: 859 - 861.
30. Sokol RJ, Chik L. A prototype system for perinatal knowledge using an artificial intelligence tool. *J Perinat Med* 1988; 16: 273 - 281.
31. Walker CHM. Neonatal records and the computer. *Arch Dis Child* 1977; 52: 452 - 461.
32. Perlstein PH, Edwards NK, Atherton HD, Sutherland JM. Computer-assisted newborn intensive care. *Pediatrics* 1976; 57 (4): 494 - 501.
33. Finner NN, Fraser AJ. Neonatal data base and automated discharge summary using a personal computer and proprietary software. *Pediatrics* 1985; 76: 296 - 273.
34. Lindstorm D, Cotton R. Use of computers in the newborn intensive care unit. *Clin Perinatol* 1983; 10: 195 - 203.
35. Lundsgaarde HP. Evaluating medical expert systems. *Soc Sci Med* 1987; 24 (10): 805 - 829.

36. Friedman JG, Gustafson DH. Computers in clinical medicine: a critical review. *Comput Biomed Res* 1977; 10: 199 - 204.
37. Singer J, Sacks HS, Lucente F, Calmers TC. Physician attitude towards applications of computer database systems. *JAMA* 1983; 249: 1610 - 1614.
38. Anderson JG, Jay SJ, Schweer HM, Anderson MM. Why doctors don't use computers: some empirical findings. *J R Soc Med* 1986; 79: 142 - 144.
39. Teach RL, Shortliffe EH. An analysis of physicians attitude regarding computer-based clinical consultation systems. *Comput Biomed Res* 1981; 14: 542 - 558.
40. Oh Te, Black JL, De Roach JN, Davis NJ. Computerised patient monitoring in intensive care. *Anaesth Intensive Care* 1982; 10 (3): 229 - 232.
41. Cotterill JJ, Pennock BE, Grenvik A, Rogers RM. Critical care computing. *JAMA* 1982; 248: 2289 - 2291.
42. Greenburg AG, McClure DK, Fink R, Stubbs JA, Peskin GW. Computerization of the surgical care unit: improvement of patient care via education. *Surgery* 1975; 77 (6): 799 - 806.
43. Gardner RM. Computerized data management and decision making in critical care. *Surg Clin North Am* 1985; 65 (4): 1041 - 1051.
44. Gardner RM. Computerized management of intensive care patients. *MD Comput* 1986; 3 (1): 36 - 51.
45. Kuperman JG, Gardner RM. The HELP system. A snapshot of time. Internal Publication, September 1988.
46. Elstein AS, Shulman LA, Sprafka SA. Medical problem solving: an analysis of clinical reasoning. Harvard University Press, Cambridge Mass. 1978.
47. Szolovits P, Pauker SG. Categorical and probabilistic reasoning in medical diagnosis. *Artif Intell* 1978; 11: 115 - 144.
48. Weed LL. The computer as a new basis for analytic clinical practice: coupling individual problems with medical knowledge. *Mt Sinai J Med* 1985; 52 (2): 94 - 98.

49. Weed LL. Problem-knowledge coupling. *Med Instrum* 1987; 21 (5): 284 - 287.
50. Spitzer RI, Cohen J, Fleiss JI, Endicott J. Quantification of agreement in psychiatric diagnosis. *Arch Gen Psychiat* 1967; 17: 81 - 87.
51. Fleiss JI. Measuring nominal scale agreement among many raters. *Psychol Bull* 1971; 76 (3): 378 - 382.