A QUANTITATIVE APPROACH TO MEDICAL DECISION MAKING

DISSERTATION

Presented to the Graduate Council of the
North Texas State University in Partial
Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

John W. Meredith, B.B.A., M.B.A.
Denton, Texas

May, 1975

Meredith, John W., A Quantitative Approach to Medical Decision Making. Doctor of Philosophy (Management Science), May, 1975, 101 pp., 4 tables, 11 figures, bibliography, 38 titles.

The purpose of this study is to develop a technique by which a physician may use a predetermined data base to derive a preliminary diagnosis for a patient with a given set of symptoms. The technique will not yield an absolute diagnosis, but rather will point the way to a set of most likely diseases upon which the physician may concentrate his efforts. There will be no reliance upon a data base compiled from poorly kept medical records with non-standardization of terminology.

The significance of this study is that it frees the physician from a diagnosis based upon his subjective intuition. This type of diagnosis has been useful in the past, but with new techniques and treatments becoming available practically every day, the physician is in a position where he is finding it increasingly difficult to keep up with the changes. If the patient is to receive the best possible care, there must be a method available to the doctor which will enable him to provide that care in the shortest possible time. It is not intended that this replace the doctor, but that it will furnish him with a tool which will aid him in providing quality patient care.

Evaluation of test results and consideration of alter-

natives concerning diseases, prognosis, and treatment are among the most time-consuming of a doctor's tasks. Any reduction in this time would be of great value to doctors and their patients.

The system as developed in the study utilizes the ROCOM Health History Questionnaire, copyright by Patient Care Systems, Inc., Darien, Connecticut, as a means of collecting symptom data for each patient. For each of the 129 symptoms contained in the questionnaire a list of diseases was compiled from a standard medical reference. When the initial list was completed a second list was produced from it which was a cross-reference list containing each disease and its associated symptoms.

Conceptually, the system is a type of tree structure. It begins with one symptom in the patient's symptom set. Each disease which has that symptom is listed. As the tree structure branches, the symptoms for each of the diseases are next noted. At each symptom level in the tree, symptoms are eliminated according to a set pattern. The original symptom is eliminated as it reappears at each level, as are any symptoms which are not present in the patient's original set. This process continues until all symptoms are eliminated. With the complete tree a linking of diseases from one level to the next is attempted and the number of matches is accumulated. A separate tree is constructed for each symptom exhibited by the patient. When all possible trees are completed and all disease-to-disease matches accumulated, the disease with the

greatest frequency is considered to be the most likely. Output from the system consists of a listing for each patient containing the ten most likely diseases, their International Classification of Diseases code numbers, and the relative frequency of the disease.

While this study produces a workable tool for the physician to use in the process of medical diagnosis, the ultimate responsibility for the patient's welfare must still rest with the physician.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER I

INTRODUCTION

Background

In the area of medical diagnosis of diseases, it is not known, with any degree of certainty, exactly what process a doctor utilizes in reaching his decision. It is thought by some that the process is one of pattern recognition. The doctor notes certain signs, symptoms, and results of clinical tests and recognizes a pattern which indicates the proper classification. Another technique which may be used is the "multiphasic screening" process, a process of elimination by means of which the diagnostician eliminates possibilities until the proper diagnosis is achieved. Yet another approach to the diagnostic problem is through the use of probabilities.

There are two primary methods of approaching probabilistic diagnosis. The first of these, described by Ledley and Lusted (2,pp.13-15), is a purely Bayesian

approach which utilizes a data base of historical cases from hospital records to produce the initial prior probabilities. The Bayesian approach in its form for medical diagnosis is

$$P(D|S) = \frac{P(S|D)\ P(D)}{P(S)}$$

where

$P(D|S)$ is the probability of a disease (D) given that the patient has a symptom, or symptom set (S).

$P(S|D)$ is the probability of a symptom, or symptom set (S) given that a patient has a specific disease (D). Found from the initial data base.

$P(D)$ is the probability of a disease (D) occurring in a given population. Found from the initial data base.

$P(S)$ is the probability of a symptom or symptom set (S) occurring in a given population. Derived from the Bayesian theorem.

This approach assumes a large, available, and valid source of medical data about which more will be said later.

The second approach to probabilistic diagnosis, developed by Gorry and Barnett (1,pp.492-501), utilizes a "sequential" format in conjunction with the basic Bayesian analysis. In this sequential process the diagnostician begins with one symptom, or symptom set, which produces basic probabilities of certain classifications of diseases given that the patient has that particular symptom. Selecting the branch with the greatest probability requires the input of another symptom or symptom set which yields yet another set of probabilities. The procedure may be terminated at any point at which the cummulative probability reaches a certain predetermined level.

Thus far the majority of studies done in these areas use the purely Bayesian approach. They make the assumption that a valid, usable data base exists. In interviews and correspondence with a cardiologist, Donald Ramsegram of Dallas, Texas, a pathologist, John Childers of St. Paul's Hospital, Dallas, Texas, a radiologist, Lee Lusted of The University of Chicago Radiology Department, Chicago, Illinois, a psychologist, Edwin Mosely of National Aeronautics and Space Administration, Johnson Spacecraft

Center, Houston, Texas, there seems to be the general agreement that no such data base exists. If these men are correct, this means the work of Overall and Williams "Models for Medical Diagnosis" in Behavioral Sciences , April, 1961, pp. 134-141 (3), and Winkler, Reichertz and Kloss "Computer Diagnosis of Thyroid Diseases: Comparison of Incidence Data and Considerations of the Problem of Data Collection" in The American Journal of the Medical Sciences , January, 1967, pp. 27-33 (4), and others may be of questionable value because of the use of inadequate data bases. The major difficulty is a relatively loose and non-standard system of medical record keeping. For example, in the Winkler, Reichertz and Kloss study, some difficulty arose as to the proper definition of the symptom "lethargy". One group considered the term to mean "tiredness" while another took it to mean "apathy". The result was a significant difference in frequency of this symptom in two different populations.

Purpose

The purpose of this study is to develop a technique by which a physician may use a predetermined data base to derive a preliminary diagnosis for a patient with a given

set of symptoms. The technique will not yield an absolute diagnosis, but rather will point the way to a set of most likely diseases upon which the physician may concentrate his efforts. There will be no reliance on a data base compiled from poorly kept medical records with non-standardization of terminology.

Significance

The significance of this study is that it frees the physician from a diagnosis based upon his subjective intuition. This type of diagnosis has been useful in the past. But with new techniques and treatments becoming available practically every day, the physician is in a postion where he is finding it increasingly difficult to keep up with the changes. If the patient is to receive the best possible care , there must be a method available to the doctor which will enable him to provide that care in the shortest possible time. It is not intended that this replace the doctor, but to provide him with a tool which will aid him in providing quality patient care.

Evaluation of test results and consideration of alternatives concerning diseases, prognoses and treatment

are among the most time-consuming of a doctor's tasks. Any reduction in this time would be of great value to doctors and their patients. In any case the ultimate responibility for the patient's welfare rests with the physician.

Limitations

Because of availability of computer hardware, this study was limited to programs designed for a system no larger than the IBM 360 model 50. Few hospitals, with the exeption of some major medical research centers, have more sophisticated systems. The programs developed during the course of this study will be of use in the average hospital computer facility.

CHAPTER BIBLIOGRAPHY

1. Gorry, G. Anthony and G. Octo Barnett, "Experience with a Model of Sequential Diagnosis," Computers and Biomedical Research, Vol. I., New York, Academic Press, 1968.

2. Ledley, Robert S. And Lee B. Lusted, "Reasoning Foundations of Medical Diagnosis," Science CXXX(July, 1959), 9-21.

3. Overall, John E., and Clyde M. Williams, "Models for Medical Diagnosis," Behavioral Sciences, VI (April, 1961), 134-141.

4. Winkler, C.,P. Reichertz and G. Kloss, "Computer Diagnosis of Thyroid Diseases, Comparison of Incidence Data and Considerations on the Problem of Data Collection," The American Journal of the Medical Sciences, CCXLI (January, 1967), 27-33.

CHAPTER II

THEORETICAL BACKGROUND

The Beginning of Inquiry

Until the late 1950's there was but passing interest in the field of mathematical or quantitative techniques in medical diagnosis. At that time Ledley and Lusted wrote their classic work "Reasoning Foundations of Medical Diagnosis" in which they attempted to describe the processes which a physician utilizes in reaching a diagnosis, including intuitive reasoning.

> If a physician is asked , "How do you make a medical diagnosis?" his explanation of the process might be as follows. "First , I obtain the case facts from the patient's history, physical examination, and laboratory tests. Second, I evaluate the relative importance of the different signs and symptoms. Some of the data may be of first-order importance and other data of less importance. Third, to make a differential diagnosis I list all the diseases which the specific case can reasonably resemble. Then I exclude one disease after another from the list until it becomes apparent that the case can be fitted into a definite disease category, or that it may be one of several possible diseases, or

> else that its exact nature cannot be determined." This, obviously, is a greatly simplified explanation of the process of diagnosis, for the physician might also comment that after seeing a patient he often has a "feeling about the case." this "feeling", although hard to explain, may be a summation of his impressions concerning the way the data seem to fit together, the patient's reliability, general appearance, facial expression, and so forth; and the physician might add that such thoughts do influence the considered diagnosis. No one can doubt that complex reasoning processes are involved in making a medical diagnosis. The diagnosis is important because it helps the physician to choose an optimum therapy, a decision which in itself demands another complex reasoning process.....
>
> Medical diagnosis involves processes that can be systematically analyzed, as well as those characterized as "intangible." For instance, the reasoning foundations of medical diagnostic procedures are precisely analyzable and can be separated from certain considered intangible judgements and value decisions. Such a separation has several important advantages. First, systematization of the reasoning process enables the physician to define more clearly the intangibles involved and therefore enables him to concentrate full attention on the more difficult judgements. Second, since the reasoning processes are susceptible to precise analysis, errors from this source can be eliminated. (7,p.9)

The article covered the areas and techniques which have been in use from that time to the present.

> Two well-known mathematical disciplines, symbolic logic and probability, contribute to our understanding of the reasoning foundations of medical diagnosis; a third mathematical discipline, value theory, can aid the choice of an

optimum treatment. These three basic concepts are inherent in any medical diagnostic procedure, even when the diagnostician utilizes them subconsciously, or on an "intuitive" level.

....the logical concepts inherent in medical diagnosis emphasize the fundamental importance of considering combinations of symptoms or symptom complexes in conjunction with combinations of diseases or disease complexes . This point is emphasized because often an evaluation is made of a sign or symptom by itself with respect to each possible disease by itself, whereas consideration of the combinations of signs and symptoms that the patient does and does not have in relation to possible combinations of diseases is of primary importance in diagnosis

The probabilistic concepts inherent in medical diagnosis arise because a medical diagnosis can rarely be made with absolute certainty; the end result of the diagnostic process usually gives a "most likely" diagnosis. The logical considerations present alternative possible disease complexes that the patient can have; the purpose of the probabilistic considerations is to determine which of these alternative disease complexes is "most likely" for this patient.

The value theory concepts inherent in medical diagnosis and treatment are concerned with the important value decisions that the diagnostician frequently faces when he is choosing between alternative methods of treatment. The problem facing the physician is to choose that treatment which will maximize the chance of curing the patient under the ethical, social, economic, and moral constraints of our society. (7,p.10)

Development of Theory

The majority of work to date has been concerned with studies in applications of formal probability theory to medical diagnosis. Overall and Williams in an article entitled "Models for Medical Diagnosis" combined the concepts of probability theory with discriminant analysis to arrive at their final recommendations.

> The general approach which the writers consider to be practical involves the making of a series of decisions, each decision delimiting further the number of diagnostic classifications to be considered in subsequent decisions. At each stage the amount of information to be considered must be reduced to reasonable magnitude by searching for those few signs and symptoms which provide maximum discrimination between all of the diagnostic groups to which the patient might reasonably belong at that stage in the process of diagnosis. As a result of successive decisions, the number of diagnostic categories is reduced and new measures are sought which maximize discrimination between the remaining groups. The total problem in establishing such a comprehensive system reduces to that of deciding how many decisions are necessary and what information is necessary for each.
>
> ....The first step in developing a comprehensive diagnostic system involves the identification of a small number of signs and symptom measures which maximizes discrimination between all of the diagnostic groups. It should not be difficult to determine empirically which single signs and symptoms are of most use in discriminating between the many disease groups.

> Once the set of measures to be used in making the initial diagnostic decision has been selected, either the multiple discriminant function model or the frequency model can be employed to yield the probability values actually used in the initial classification of a patient. (9,pp.139-140)

During the mid 1960's there was a noticable decrease in interest in quantitative diagnosis as evidenced by the lack of publication in this area. However in 1967, three German physicians, Winkler, Reichertz and Kloss published the results of a study which touched off a flurry of articles and books on the subject. Their study "Computer Diagnosis of Thyroid Disease, Comparison of Incidence Data and Considerations on the Problem of Data-Collection" (15) used as its base an earlier article by Fitzgerald and Williams, "Computer Diagnosis of Thyroid Disease" (2). Both studies were strictly Bayesian in nature. That is, they relied heavily on the frequency theory of probability in determining prior probabilities for use with the Bayesian theorem. The study by Winkler, Reichertz and Kloss used a population of 974 cases from which the results of seventeen clinical signs and five laboratory tests were collected. From the frequencies of occurance of each of the signs and test results, probabilities were calculated. The results of these studies were duplicated by this writer in a preliminary study for the present project. Results from

this preliminary study are compatible with those of the Overall and Williams and Winkler, Reichertz and Kloss studies. It was these studies which prompted the present study.

The use of Bayesian probability was taken to task in mid-1967 by Vanderplas in an article in Computers and Biomedical Research entitled "A Method for Determining Probabilities for Correct Use of Bayes' Theorem in Medical Diagnosis" in this article Vanderplas attacked not the use of Bayes' theorem, but the inappropriate use of it. He

> ....outlines an empirical procedure, easily implemented by computer, for (1) determining the number of mutually exclusive [symptom sub set] in a given [disease set] for a patient population and (2) deriving from those subsets the correct probabilities necessary for proper application of Bayes's Theorem to the determination of posterior probabilities of diseases, given a particular [symptom sub set] for members of the population. (12,p.215)

He emphasizes the proper use of the procedure by presenting several assumptions which most of the researchers have used which violate the concepts inherent in the theorem.

A book by Bailey entitled The Mathematical Approach to Biology and Medicine deals in a chapter on "Mathematical

Methods of Medical Diagnosis", with the necessity for the development of mathematical techniques for medical diagnosis.

> Few would deny that the whole process of diagnosis is of central importance in medicine, and that the process demands, moreover, a high degree of skill, knowledge and imagination in its practitioners. It is also universally recognized that the accuracy of a diagnostic decision and speed with which it can be reached vary enormously according to the patient's condition; the data available on his individual symptoms, signs and laboratory tests; the general body of medical information about the occurence of such observable material over a wide range of alternative diseases; and the ability of the clinician himself. Moreover, the medical treatment adopted and the patient's expectation of recovery frequently depend on early and accurate diagnosis. With such tremendous variations in the effective efficiency of diagnostic procedures it is quite natural to try to determine the circumstances under which optimal results can be achieved. Physicians have, of course, been making this attempt, with different degrees of success, for centuries. But with modern methods of diagnosis and treatment, using all the special skills of science and technology, the potentialities for success have been greatly enhanced in recent years. It is therefore important to have precise methods for discussing, investigating, evaluating and controlling the process of diagnosis....the best way of achieving precise logical thought is by a mathematical approach. This approach can, in principle, be adopted no matter how difficult and complicated the subject. Indeed, if there are a large number of interdependent factors all showing appreciable natural variation, it is <u>only</u> by using an appropriate statistical method that the complex pattern of correlated effects can be handled with a reasonable degree of efficiency. And if the numbers of factors or items of data are very large it may be desirable, or even necessary, to use an

> automatic computer in order that the required results can be obtained in a sufficiently short peroid of time. This attitude by no means depreciates the value of insight and imagination. On the contrary, it allows these faculties greater scope by arranging for all tasks that <u>can</u> be couched in numerical and logical terms to be handled by mathematical and computer techniques. (1,pp.238-239)

He continues in the chapter to recommend the use of Bayes' theorem for the solution to the problem of medical diagnosis.

The following year, 1968, an article in <u>Computers and Biomedical Research</u>, "Experience with a Model of Sequential Diagnosis" by Gorry and Barnett for the first time took exeption to the use of Bayes' theorem as the only model to be used.

> In recent years, a number of studies of the use of computer programs in diagnosis have been performed. Central to each of these efforts has been the development of an explicit, precisely formulated procedure for diagnosis. Such a development is a prerequisite for computer programs of this type. In general, attention has been focused on models of the <u>inference function</u> of diagnosis, the development of a diagnosis from the given set of clinical signs. Some interesting probabilistic models have been developed which employ Bayes rule.
>
> Bayes rule has understandable appeal for use in such a model. First, it permits the use of probabilities in inference. This is preferable to

a deterministic approach, because it reflects some of the basic uncertainties of diagnosis. Also, Bayes rule provides a rational means for considering both a priori belief about the incidence of various diseases and the evidence embodied in the clinical signs in a given case. Finally, the formulation of the inference function in terms of Bayes rule is particularly suited for incorporation into a computer program. Given the necessary statistical data, the problem of inference is thereby reduced to a problem of computation.

While the Bayesian model is well suited for computer diagnosis, there are certain problems associated with its use. First, the model requires that extensive statistical data be available for the given area. The collection and processing of this data may be a very formidable task. There are also problems in properly accounting for the dependence of various signs and the possibility of the simultaneous occurence of more than one disease. In spite of these difficulties, a number of investigations of the use of the Bayesian model have obtained encouraging results.

....a given set of tests are performed on the patient, and the test results become the input to the program. Through the use of Bayes rule, the program computes the conditional probability distribution for the diseases in question. This distribution constitutes the diagnosis provided by the program.

....diagnosis, then, consists of two major functions, inference and test selection.....a more complete model of diagnosis must provide for the interaction of these two functions. (3,pp.490-492)

Gorry and Barnett include in their discussion a procedure for employing the sequential test selection function based on the cost of misdiagnosis, in an attempt to find which of several different techniques would provide the best diagnosis in thyroid patients.

The three models used by Nordyke, Kulikowski and Kulikowski in their study "A Comparison of Methods for the Automated Diagnosis of Thyroid Dysfunction" in Computers and Biomedical Research were the Bayesian model, the linear discriminant model and the pattern recognition method of class.

> A simulation of the doctor's diagnostic process is probably the first approach that appeals to the designer of a computer program for medical diagnosis. This is impossible, since the induction performed by a physician follows no clearly defined rules and the computer needs rules to classify patients into disease categories. But when diagnosis is posed as a problem of classification the methods of decision theory, discriminant analysis, and pattern recognition become immediately applicable. The computer can use a fixed decision rule based on a wide "experience" of many relevant case histories, in contrast to the flexible recognition of the doctor based on more limited experience.....the advantages of speed and consistency make such programs good screening tools and could supplement the short supply of specialists in various fields of medicine.
>
> ....to do this, three mathematical models were compared: a simple Bayes model, a linear

discriminant function, and a pattern recognition method. Such a systematic comparison was needed before it could be decided which methods were the most valuable clinically.

1. The Bayes Model. When this model is used with a first-order approximation to conditional probability, it is the simplest and most popular mathematical model applied to automated diagnosis. Its principal feature is that it relies only on the first-order probabilities of patient characteristics (symptoms, signs and laboratory tests), rather than on higher-order joint probabilities of characteristics. Computer programs using this method, therefore, require less storage and computer time than those using other methods.

2. The Linear Discriminant Model. This method incorporates in its classification rule the effects of correlation or second-order interdependence between characteristics. The approach is to find a weighted sum or linear combination of the measurements, such that the sum will take on very different values for members of different diagnostic categories. It is best suited for continuous data sampled from multivariate normal distributions.

3. The Pattern Recognition Method of Class Featuring Information Compression. This method attempts to extract the most characteristic features of each diagnostic category, rather than trying to discriminate directly between categories. A patient is then classified into the category with which his data shares the most features. (8,pp.375-377)

This study reached the conclusion that the Bayesian method performed significantly better than the others when laboratory tests are included in the initial symptom set.

Other Studies

Since 1971 there has been a lack of any new or different models proposed for quantitative diagnosis. Several studies have proposed aids to diagnosis and general models for better utilization of hospital services, such as room occupancy and housekeeping chores.

Some of the studies were directed toward systems of diseases classification as in the Hurtado and Greenlick study "A Disease Classification System for Analysis of Medical Care Utilization, with a Note on Symptom Classification" their study

> ....seeks to identify significant determinants of medical care utilization by investigating the relationships among background characteristics of patient populations, disease patterns, and medical care utilization. It was posited that different sets of background characteristics are significant determinants of medical care utilization in different disease situations.
>
> The emphasis is on studying the full range of medical care services, including professional visits in the clinic, home, or emergency room, hospital services, telephone calls and letters, and laboratory and x-ray services. Since the vast bulk of morbidity in our society is treated outside of hospitals, the disease classification

> system was designed primarily to reflect this broad spectrum of conditions and does not provide a fine breakdown of diseases for which institutional care is required.
>
> Since the general analytic framework of the study is based on the hypothesis that different sets of background characteristics are significant determinants of medical care utilization in different situations, it was necessary to design a disease classification system focusing on the impact of diseases on individuals' utilization behavior. (5,p.236)

In a study by Whinery (14) which was prepared at the University of Texas, M.D.Anderson Hospital and Tumor Institute, Houston, Texas for the National Aeronautics and Space Administration, Manned Spacecraft Center, a program was developed which was of the diagnosis aiding type. This program is a comprehensive analytical program for the acquisition, analysis and display of electrocardiogram data. While this program does a vast amount of analysis, any decision making is left entirely up to the physician.

More recently, another of the diagnosis-aiding type of system was developed which could have been useful, but stopped short of its full potential. Warner, Olmstead and Rutherford's study "HELP-A Program for Medical Decision Making" in Computers and Biomedical Research provides a base from which the physician may make his diagnosis.

> The program (HELP) is designed to help the physician or nurse with the intellectual task of recognizing the occurrence of preset conditions which indicate nodes of decision points in a patient's illness. Such a node may arise from one or more new entries into the patient's record. Each of these decisions must represent the best in current medical knowledge and be easily modified without alteration of the program itself as new information becomes available. The form in which these decisions are specified must be understandable to the physician. The data base upon which decisions are made may originate from entries, automated reading of physiological transducers or laboratory devices (autoanalysers, etc) or as the result of prior decisions made by HELP itself, thus establishing a hierachial data structure and information system. The system also provides ready access, not only to raw data and trends in any variable, but to all currently relevant decisions previously made on a patient.... The basis for each decision is displayed to the physician on request so that he may review the pertinent data himself. (13,pp.65-66)

Completion of the Cycle

As reported in Time magazine, January 28, 1974, a team of physicians and scientists from Tufts University School of Medicine and Massachusetts Institute of Technology have taken a new look at one aspect of the medical decision making problem.

> When the researchers began trying to write a decision-making computer program three years ago, says Dr. William Schwartz, chairman of the department of medicine at Tufts Univeristy School

> of Medicine and spokesman for the group, they discovered that little was known about how physicians arrived at their complex decisions. "Medical school emphasis emphasizes the acquisition of specific factual data," Schwartz says, " but it has paid remakably little attention to the decision-making process." Thus he and his colleagues are analyzing how a doctor decides on such serious treatment as abdominal surgery.
>
> A first-stage result of their study is a computer program that duplicates some of the mental processes of a highly skilled physician. Using acute kidney failure as an experimental model, the research group programmed the machine to weigh the risks and benefits of various tests and treatments and to consider such factors as the patient's attitude toward surgery. "We find it is like playing chess," says Schwartz. "Doctors don't make just one isolated move, they have to look ahead at what else is likely to happen." (10,p.49)

The original articles, published as a series in the American Journal of Medicine (4;6;11) have completed the cycle from the first of the articles cited which dealt with the attempt to discover how physicians reason, to attempts to develop specific techniques of diagnosis and back to the problem of how decisions are made.

CHAPTER BIBLIOGRAPHY

1. Bailey, Norman T.J., The Mathematical Approach to Biology and Medicine, New York, John Wiley & Sons, 1967.

2. Fitzgerald, L. T. And C. M. Williams, Computer Diagnosis of Thyroid Disease, Gainesville, University of Florida Printing Office, 1964.

3. Gorry, G. Anthony and G. Octo Barnett, "Experience with a Model of Sequential Diagnosis," Computers and Biomedical Research, Vol. I., New York, Academic Press, 1968.

4. Gorry, G. Anthony, Jerome P. Kassirer, Alvin Essig and William B. Schwartz, "Decision Analysis as the Basis for Computer-Aided Management of Acute Renal Failure," The American Journal of Medicine, LV(October, 1973), 473-484.

5. Hurtado, Arnold V. And Merwyn R. Greenlick, "A Disease Classification System for Analysis of Medical Care Utilization, with a Note of Symptom Classification," Health Services Research, VI(Fall, 1971), 235-250.

6. Jelliffe, Roger W., "Quantitative Aspects of Clinical Judgement," The American Journal of Medicine, LV(October, 1973), 431-433.

7. Ledley, Robert S. And Lee B. Lusted, "Reasoning Foundation of Medical Diagnosis," Science CXXX(July, 1959), 9-21.

8. Nordyke, Robert A. , Casimar A. Kulikowski and C. Wilk Kulikowski, "A Comparison of the Methods for the Automated Diagnosis of Thyroid Dysfunction," Computers and Biomedical Research, Vol. IV, New York, Academic

Press, 1971.

9. Overall, John E. And Clyde M. Williams, "Models for Medical Diagnosis," Behavioral Sciences, VI(April, 1961) , 134-141.

10. "Prescription by Computer," TIME, CIII(January 28, 1974), 48-49.

11. Schwartz, William B., G. Anthony Gorry, Jerome P. Kassirer and Alvin Essig, "Decision Analysis and Clinical Judgement," The American Journal of Medicine, LV(October, 1973), 459-472.

12. Vanderplas, James M. , "A Method for Determining Probabilities for Correct Use of Bayes's Theorem in Medical Diagnosis," Computers and Biomedical Reasearch, Vol. I, New York, Academic Press, 1967.

13. Warner, Homer R., Charles M. Olmstead and Barry D. Rutherford, "HELP-A Program for Medical Decision-Making, " Computers and Biomedical Research, Vol. V, New York, Academic Press, 1972.

14. Whinery, D. Gaye, E.C.G. Data Acquisition and Analysis Package, Houston, National Aeronautics and Space Administration, 1971.

15. Winkler, C. P. Reichertz and G. Kloss, "Computer Diagnosis of Thyroid Diseases, Comparison of Incidence Data and Considerations on the Problem of Data Collection," The American Journal of the Medical Sciences, CCXLI(January, 1967), 27-33.

CHAPTER III

EVOLUTION OF THE PRESENT STUDY

The present study began in 1970 as an investigation of the applicability of the Bayesian theorem to medical decision making. It was conceded that the Bayesian theorem had application to this area and work began toward the end of developing programs and procedures for its implementation.

A pilot study, utilizing the Bayesian Theorem and based on two previously cited papers, by Overall and Williams and Riechertz, Winkler and Kloss was conducted which reproduced the results of both studies. The pilot study utilized the data collected by the two studies as the basis for computations. Flowcharts, using American National Standards Institute (ANSI) symbols, for the pilot study are included as Appendix A. The computer program, written using the ANSI FORTRAN language, is included as Appendix B.

The study produced the same results as those of the Overall and Williams and Reichertz, Winkler and Kloss

studies in that a diagnosis was reached for thyroid disease in each of the test cases used. The decision table which was calculated using the program is included as Appendix C. Actual output from the program showing Bayesian probabilities of diagnoses are found in Table I.

The methodology of the pilot study was to first establish the set of relevant symptoms, signs and clinical tests which are applicable to thyroid disease. Table II is a listing of the set. Symptoms and signs 1 through 17 are observable through a physical examination of the patient, 18 through 42 are four clinical tests and their ranges based on the test results. Using then the prior probabilities from the previous studies for P(S|D) for each of the symptoms and diseases and P(D) for each disease, P(S) was computed from the basic formulation: $P(Si) = \sum_{j=1}^{n} P(Si|Dj)\ P(Dj)$. The necessary elements were then at hand for the computation of P(Dj|Si) which is found by the Bayesian Theorem for medical diagnosis:

$$P(Dj|Si) = \frac{P(Si|Dj)\ P(Dj)}{P(Si)}$$

For a patient with a given symptom set the probability of each disease given that symptom set is the multiplicative probability of the P(Dj|Si) for each of the Si present in

TABLE I

BAYESIAN DIAGNOSIS OUTPUT

THE PROBABILITY THAT PATIENT NUMBER 1
HAS HYPERTHYROIDISM IS 1.00000
THAT HE HAS HYPOTHYROIDISM IS 0.00000
AND THAT HE HAS EUTHYROIDISM IS 0.00000

THE PROBABILITY THAT PATIENT NUMBER 2
HAS HYPERTHYROIDISM IS 0.00000
THAT HE HAS HYPOTHYROIDISM IS 1.00000
AND THAT HE HAS EUTHYROIDISM IS 0.00000

THE PROBABILITY THAT PATIENT NUMBER 3
HAS HYPERTHYROIDISM IS 0.00000
THAT HE HAS HYPOTHYROIDISM IS 0.00000
AND THAT HE HAS EUTHYROIDISM IS 1.00000

TABLE II

THYROID SIGNS, SYMPTOMS
AND CLINICAL TESTS

1.	Nervousness	
2.	Heat sensitivity	
3.	Perspiration	
4.	Appetite gain	
5.	Weight loss	
6.	Hyperkinetic movements	
7.	Warm, moist skin	
8.	Light finger tremor	
9.	Lethargy	
10.	Cold sensitivity	
11.	Decreased perspiration	
12.	Appetite loss	
13.	Weight gain	
14.	Slower movements	
15.	Dry, rough skin	
16.	Face edema	
17.	Eye symptoms	
18.	BMR	under -40
19.		-40 to 0
20.		1 to 20
21.		21 to 45
22.		over 45
23.	PBI	under 2.3
24.		2.3 to 4.0
25.		4.1 to 7.0
26.		7.1 to 9.6
27.		over 9.6
28.	131I uptake, 6 hour value	under 4
29.		4 to 6
30.		7 to 20
31.		21 to 37
32.		over 37
33.	131I uptake, 24 hour value	under 6
34.		6 to 13
35.		14 to 34
36.		35 to 55
37.		over 55
38.	Serum T3	under 25
39.		25 to 27
40.		27 to 34
41.		34 to 36
42.		over 36

the symptom set.

The tentative decision was made to expand Bayesian probability to a general diagnostic procedure. A search through a source of disease classifications, International Classification of Diseases, Adopted showed the necessity for providing for diagnosis of several thousand possible diseases. Relevant symptoms involved in such a procedure would be on the same order or greater. While the task looked massive, it was felt that the project could be handled with appropriate use of magnetic tape and disk units.

A search was then begun to locate a data base, or a group of data which might be adapted for use as a data base. In an effort to discover the proper method for data collection, a trip to John Sealey Hospital, University of Texas Medical Branch in Galveston, Texas was made. It was there that Reichertz did much of the preliminary work for his study on thyroid disease. Reichertz had returned to Germany prior to the visit and an interview was arranged with Robert Peake, a researcher in thyroid disease in the Division of Endocrinology. Peake was most helpful in revealing the difficulties involved in assembling data for the proposed study. A number of patient data files were

provided and a detailed search made through each in an attempt to isolate the pertinent signs and symptoms as listed in Table II. Little difficulty was encountered in noting clinical test results as they were listed in patient files on separate laboratory report forms. Signs and symptoms however posed quite another problem. It was found that in medical schools in general, no standardization of notation for signs and symptoms is taught. Each physician uses his own style of notation and frequently uses his own shorthand method for write up of narrative case histories. Additionally, patient files are maintained for considerable periods of time, often covering several volumes of material amounting to over one thousand pages per patient. Data thus compiled would be of little use to anyone other than the physician who originally wrote it. It soon became apparent that the task at hand was to find a source of data which had been kept in a more orderly fashion, perhaps with a view toward a computerized base of some type, or alternatively, to collect data at the source after having designed the input format.

Because of the lack of uniformity of medical records as found at John Sealey Hospital, a trip was made to the IBM Corporation offices in Houston, Texas where an interview with George Junkin of the Data Processing Division was

conducted. Information concerning an experimental system developed by IBM known as the "Clinical Decision Support System (CDSS)" was provided by their office. IBM's CDSS was designed to have the capabilities:

> ....to acquire, record, organize and summarize a good general medical history without effort on the physician's part.
> ....to assure good medical records with a minimum effort by the physician or his aides in creating, maintaining, summarizing, organizing, indexing, or retrieving them.
> ....to expand the range of tasks he can responsibly delegate without loss of authority or control.
> ....to capture patient data in machine-readable form at the outset, obviating the need for ex post facto conversion of conventional records for administrative, financial, billing, scheduling, statistical and other purposes in his office and/or hospital.
> ....to provide himself unfailing reminders of diagnostic possibilities, and of procedural or treatment caveats.
> ....to perform calculations such as, for example, chemical and fluid dosages in metabolic problems, radiation dosages in tumor therapy, ECG measurements and interpretations, or cardio-pulmonary function measurements and interpretations, using appropriate algorithms the physician has specified.
> ....to furnish patients with individualized written instructions, designed by the physician, regarding medications, diets, way of life, etc., with maintenance of the record of such advice.
> ....to communicate more fluently with patients when there is a physician-patient language barrier.
> ....to make appropriate reports of a given item to different recipients....

It is important to note that CDSS computer programs are not intended to handle all aspects of

medical data processing. (2,pp.3-6)

Specifically excluded from CDSS capabilities are the functions of administrative jobs, instruction, medical library, thesaurus of medical terms, medical logic and diagnosis.

Due, in part, to a lack of funding and the nonstandardization of medical records the CDSS project is currently discontinued.

Further searches were made to find a data base in Dallas, Texas at Southwestern Medical School and at University of Oklahoma Medical School Computing Center in Oklahoma City, Oklahoma. At both locations data has been collected for statistical use, but was kept in aggregate form rather than for individual patients. In that form, which could not be modified, the data was not useable.

Additional checks were made at the Nacogdoches Memorial Hospital where the Administrator, Glenn Heifner, cooperated in screening several sets of patient records. Here too, the same problem which was found to exist at John Sealey Hospital, that of lack of uniformity of medical records, was encountered.

A visit to the Stephen F. Austin State University Student Health Clinic revealed that data was being actively kept for about ten thousand students currently enrolled in the University. Data was collected for statistical reporting by patient visit to the clinic. For each patient visit a card was punched which had data for identification, sex, classification, marital status, date of birth, type of service, new or repeat visit, diagnostic code, operation code, outside visit and a clinician's code.

Using two simple programs the data contained in card form was transferred to magnetic tape, sorted by disease classification code (DCC) and by student identification number, and matched against the University's Student Master File to provide a listing by which data could be collected from Student Health Clinic files, which are filed by student name.

The time period selected for study was April 1, 1973 through August 31, 1973. During June, 1973 many of the Disease Classification Codes were changed or extended by University Physician, Ralph Bailey, for clarification. A short program was written which converted DCC's from the old system to the new system so the current DCC's would be compatible with those used earlier.

Work then proceded toward collecting data from patient files. With best effort, working as steadily as possible, the maximum number of records obtainable was on the order of about five per hour. It was estimated that full time work to cover the available records would amount to several thousand man hours of labor. Because of the time required and the other factors mentioned, the hand collection of data was deemed infeasible. These findings are supported by those of Gustafson, Kestly, Greist and Jensen in their paper "Initial Evaluation of a Subjective Bayesian Diagnostic System" in Health Services Research .

> Computer-aided medical diagnosis using Bayes' theorem (a formally optimal method of revising prior opinion in the light of new evidence) has been a promising area of research for some time but has had little real impact on the practice of medicine. Among the reasons for this ... May be mentioned insufficient data bases resulting from the inaccessibility and poor quality of medical records; incorrect aggregation of data resulting from conditional dependence of data; and an inability to incorporate new information into the diagnostic model because of the difficulty of collecting sufficient data to develop new likelihood estimates.
>
> One solution to some of these problems would be to delay further research and implementation of computer-aided diagnostic systems unitl adequate data bases have been developed. Then conditional dependence could be identified and accounted for by existing statistical methodologies, although the addition of new symptoms would still not be feasible. Research being conducted in this area

> is essentially focusing on the development of better medical information systems, including computerized interviews and record systems. While their potential and need are apparent, these systems have generally not been implemented outside their research-based environments; thus the data bases collected are quite small and often describe special populations.
>
> Another solution to the data-base problem is to obtain the likelihood estimates required for Bayes' theorem through sources other than medical records. (1,pp.204-205)

The decision was then made to develop a method of medical diagnosis which was not dependent, to so great an extent, upon Bayes theorem. The research approach to be used would be one in which no compilation of medical history data would be required. A large data base would be necessary, but it was to be developed from data gathered from sources other than medical histories. Constraining factors for the new method were that it provide a list of diagnoses which would be ranked, if possible, and that the method be compatible with existing medical record-keeping practices to keep it within a reasonable cost range and that it be useful in practice as well as in theory.

CHAPTER BIBLIOGRAPHY

1. Gustafson, David H., John J. Kestly, John H. Greist and Norman M. Jensen, "Initial Evaluation of a Subjective Bayesian Diagnostic System," Health Services Research, VI(Fall, 1971), 204-213.

2. International Classification of Diseases, Adopted, 8th ed., Washington, D. C., U.S.Government Printing Office, 1968.

3. Moore, Frederick J., Implementation of an Experimental Clinical Decision Support System, New York, International Business Machines Corporation, 1971.

CHAPTER IV

CONCEPTION AND EXECUTION

Development of a Data Base

The decision to change to a method other than the Bayesian theorem was influenced in part by a switchover made by the Stephen F. Austin State University Student Health Clinic. University Physician Ralph Bailey adopted the RCCCM Health History Questionaire (2) for use in collecting patient data. The complete questionaire is reproduced in Appendix D by permission of the publisher, Patient Care Systems, Inc., Darien, Connecticut. Each patient entering the clinic for treatment completes the questionaire prior to seeing one of the physicians. Completed questionaires then become part of the patient's permanent file in the Clinic and is used by the physician as a basis for diagnosis of patient's complaint.

It was felt that data collected by the RCCCM questionaire might serve as a basis for quantitative

diagnosis. As a result of previous experience in collecting symptom data from patient files, it was considered infeasible to attempt to build a data base for prior probability computations from the ROCOM form.

There being no viable method available for utilizing the data at hand, an attempt was made to develop a method which would be independent of prior probabilities.

The new method began as two separate lists, one listing each symptom and the diseases which have that symptom in common, and another listing each disease and the symptoms which are found in it. A merging of these two lists into a common list with the patient's symptom set as an index should provide a set of possible diseases which the patient could have.

Initial compilation of the diseases found for each of the symptoms on the ROCOM questionaire was a manual listing made during the course of several weeks. A standard medical reference, Symptom Diagnosis , (3) and a medical dictionary, Taber's Medical Dictionary , (1) were used as source material for this step. The resulting list contained each of the symptoms from the ROCOM form, numbered by corresponding questions, and the Disease Classification Code (DCC) for each of the diseases which have that symptom.

Appendix E is a listing of the DCC's used by the Student Health Clinic for classification of diagnoses. When the list was completed a short program was written, which provided a complete listing of the symptoms and their common diseases, sorted through the entire group of data and printed a listing of each disease and its symptoms by appropriate code numbers. Sample output from the program is provided in Table III.

Using the symptom codes from the ROCOM questionnaire and the DCC's from the Student Health Clinic, a compilation was made for each symptom, shown in the left column of Table III, and all of the common diseases which have that symptom. DCC's appear in the rows to the right of each symptom code. Zeros in the table represent an absence of common diseases and are used as fillers for the array only. A test in the main program halts a row search when the first zero in encountered. In the second part of the table, after a comparison and search have been made, a listing is generated by the program which contains the compliment of the first table. This table is a listing of each disease with all of its common symptoms. To the left is the DCC of each disease, followed by an indented list of symptom codes. These two lists form the basis for the data base which is used in the diagnostic program discussed below.

TABLE III

SYMPTOM/DIAGNOSIS TABLE

Symptom	Diagnosis				
87	4619	5040	3759	3640	3510
	2859	5000	0010	5350	2740
	0849	1910	2259	3460	5259
	3839	3879	4470	3068	5810
	5830	5932	0459	3459	5640
	4409	2919	3039	0910	0912
	0971	0950	4339	0	0
88	6809	6829	7179	7150	7172
	7123	7259	0370	5110	0320
89	6809	6829	7062	0720	2149
	0220	4920	5010	5239	0320
	0342	0550	0560	0509	6829
	7827	6900	0912	0950	1459
	1499	1609	1619	1740	0750
	2010	0210	2049	2001	2409
	2420	2459	0	0	0
91	3600	3639	3710	3640	3759
	3749	3460	3772	9809	0
92	3600	3639	3710	3640	3759
	3749	3460	3772	9809	9613
	2259	0	0	0	0
93	2919	3039	0369	0459	3200
	3400	7330	3749	0	0
94	3772	3760	0	0	0
95	3600	3639	3630	3640	3759
	3460	4700	7812	0	0

TABLE III--Continued

DIAGNOSIS/SYMPTOM TABLE

4619	
	87
5040	
	87
	107
3759	
	87
	91
	92
	95
3640	
	87
	91
	92
	95
	96
3510	
	87
	105
	106
2859	
	87
	126
	129
	51
5000	
	87
	107
	109
	113
	116
0010	
	87
	54
	58
	65
	62
	7

Conceptual Discussion

Conceptual logic for the main program may be illustrated by means of an oversimplified example: If a person enters the clinic with two symptoms, S1 and S3, a decision tree may be constructed for each. Beginning with either one of the symptoms exhibited by the patient, the tree will go from that symptom as the root, and branch to each of the several diseases which have that symptom as in Figure 1. Using S1 as the root yields diseases D1 and D2 which both have symptom S1. Continuing from diseases D1 and D2 it is found that they have symptoms S1 and S3, and S1, S2 and S3 respectively. At this second symptom level, superflouous symptoms are eliminated. On the D1 branch S1 is eliminated because the tree began with S1 and any further level using S1 would only result in a repetition of that branch, which is unnecessary. On the D2 branch S1 is eliminated for the same reason as before and S2 is eliminated beacuse it is not among the symptoms which the patient has. With these eliminations the only remaining symptom is S3 for each branch. The process now continues to the next level where the diseases which have S3 in common are D1, D2 and D3 for each branch. Moving to the third symptom level yields for each of the diseases D1, D2 and D3

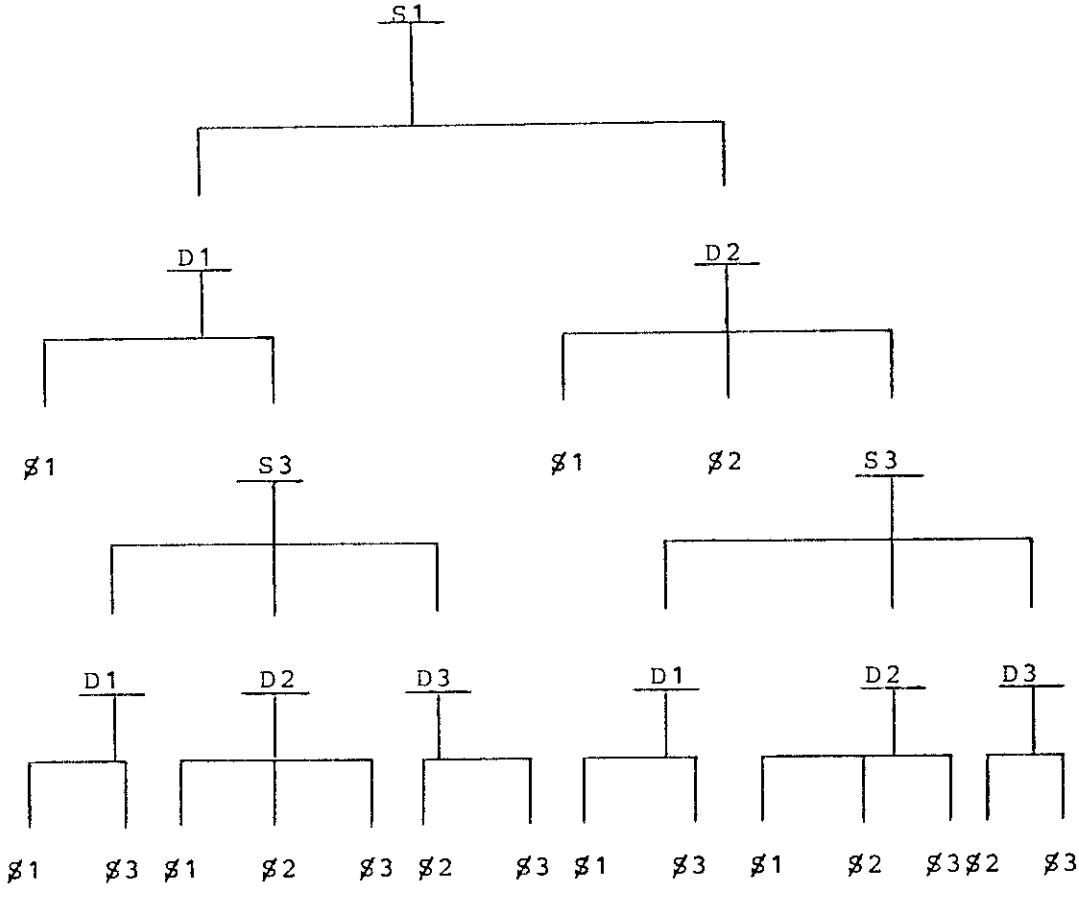


Fig. 1--Diagnosis decision tree

the symptoms S1 and S3, S1, S2 and S3, and S2 and S3 respectively. Again eliminating symptoms at the third level, S1 is eliminated throughout the level because retaining it would only cause a repetition of the entire branch. S2 is eliminated beacuse it is not part of the patient's symptom set and S3 is eliminated because it was the starting point for the second level and retention of S3 would cause repetition of that portion of the branch. At the third symptom level, all of the symptoms have been eliminated and the tree is complete. The next step is to begin with the last disease level and try to trace back through the tree to the top with no breaks in diseases. D1 traces back from the lowest disease level to the highest in the left branch while D2 and D3 do not. In the right branch D2 traces back from the lowest to highest with no breaks while D1 and D3 fail to trace. In order to keep track of the most likely diseases, the concept of a match is used. A match occurs when a disease is traced from a low level to the next highest level in the tree. Thus, for diseases D1 and D2, there is one match each, while for D3 there are no matches. The interperetation of this tree is that D1 and D2 are the patient's most likely diseases.

The process then continues to the next of the patient's symptoms, S3, and constructs a new tree as shown in Figure

2. Using the same logic as for the first tree, matches are accumulated. Overall interperetation of the process at its conclusion is that diseases D1 and D2 are the most likely with a relative frequency of one-half each and D3 is the least likely with no matches.

Discussion of the Algorithm

In practice, the actual programming of the problem could not be done as conceived. Space limitations prevented the construction of all possible symptom trees and their subsequent search. Therefore an algorithm was derived, Figure 3, which yielded the same results as the conceptual approach, but with the advantage that it could be stored in far less core space. The algorithm provides for the simultaneous construction and search of each tree structure with no need to store each completed tree. By this means each individual symptom tree may be searched, eliminated and another begun in its place. The algorithm may be termed recursive and utilizes the concept of the "stack", which may be defined as a list to and from which additions and deletions are both made to and from the same end, or top, of the stack. The stack concept was used in programming to provide a means by which each node, or branch point, in the

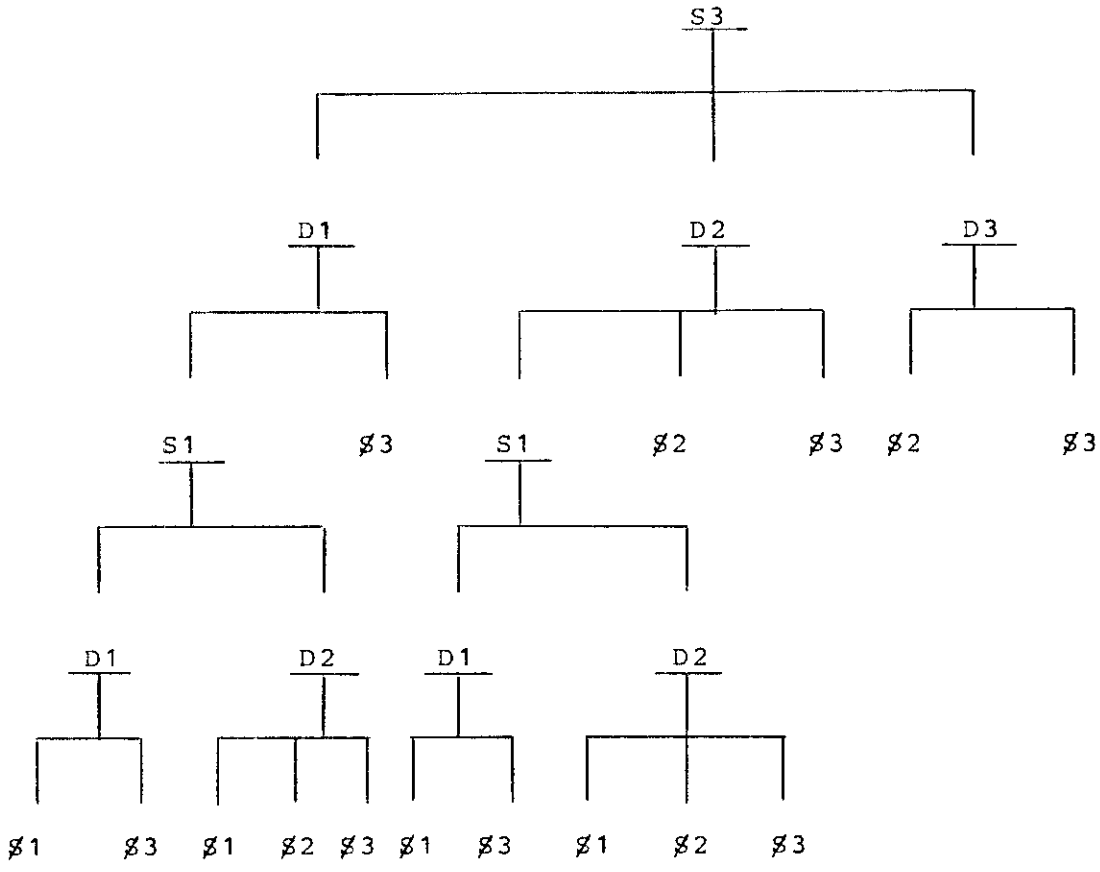


Fig. 2--Diagnosis decision tree

```
For A = 1 to N by 1

      Put P(A) as first STACK element

Find B ∋ SYM(B,1) = P(A)

For C = 2 to S by 1, do

      If SYM(B,C) = 0, go to next A

Find D ∋ DIS(D,1) = SYM(B,C)

For E = 2 to R by 1, do

      If DIS(D,E) = 0 go to next E
                    (pop stack)

           If pointer ≤ 1, go to next C
              and reset pointer to 1

      If DIS(D,E) ∉ P | DIS(D,E) ∈ STACK

           Increment E

           Increment pointer and put E
              and DIS(D,E) on STACK

Find F ∋ DIS(D,E) = SYM(F,1)

For G = 2 to S by 1, do

      If SYM(F,G) = 0, go to next C and
           reset pointer

      If SYM(F,G) ≠ SYM(B,C), go to next G

           Add 1 to counter

Go to restart loop (E = 2)
```

Fig. 3--Tree algorithm

structure could be addressed. Thus if a branch point were saved in the stack and the remainder of the branch searched until all possibilities were exhausted, it would be possible to backtrack to the branch point by "popping" the stack to bring back the address of the branch point in order to continue down another branch. The algorithm begins by placing the first of the patient's symptoms, P(A), on the top of the stack. Because the symptom list is stored with symptom codes in the first column of its array, a search is made to match the first patient symptom with the symptom list SYM. The next step is to find the corresponding first disease in the SYM list in the DIS list. If at any time a SYM element is zero, the algoritm moves to the next of the patient's symptoms because all of the branches of that tree have been searched. If a zero element is encountered in the DIS list, the indication is that the end of a branch has been reached and the algorithm backs up, or pops the stack, to the previous branch point and continues down the next branch.

The process continues, placing DIS codes on the stack as they are searched. When a branch being tranversed is found to contain a match between disease codes, the counter for that code is incremented. The search goes on until there are no more symptoms remaining in the patient's

symptom set. The main program flowchart is provided as Figure 4. When the main program logic, flowcharting and coding were completed, subroutines were developed which provided for a means of calculating relative frequencies, for sorting and printing of results.

Relative frequency may be defined as the frequency with which a disease appears in a patient's symptom set relative to the number of total occurances of all diseases in the sets. It provides the physician with some idea of how likely a particular disease is in relation to all of the diseases searched. The computation of relative frequency uses:

$$Rj = Mj/\sum_{i=1}^{n} Mi,$$ where M is the number of matches per disease.

Because of the large number of diseases searched, the relative frequencies tend to appear rather low, but the leading diseases are a great deal higher than the reamining list. The flowchart for relative frequency is Figure 5.

By sorting the possible diseases into descending order by relative frequency, the physician is provided a list with the most likely diseases first. The sort used for this task is the "tag" sort. Because of the large size of the data base used, it was considered infeasible to manipulate the

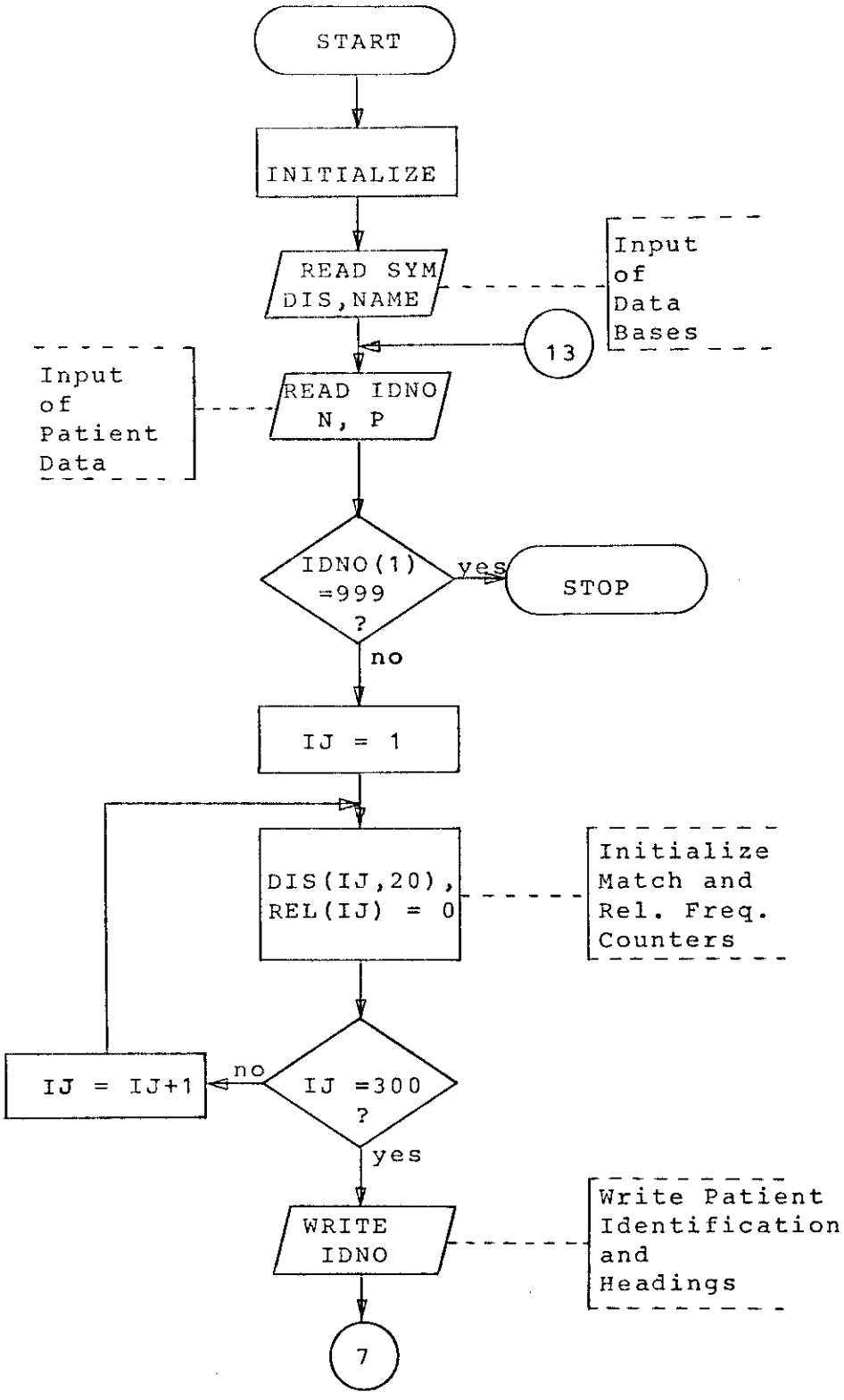


Fig. 4--Tree main program flowchart

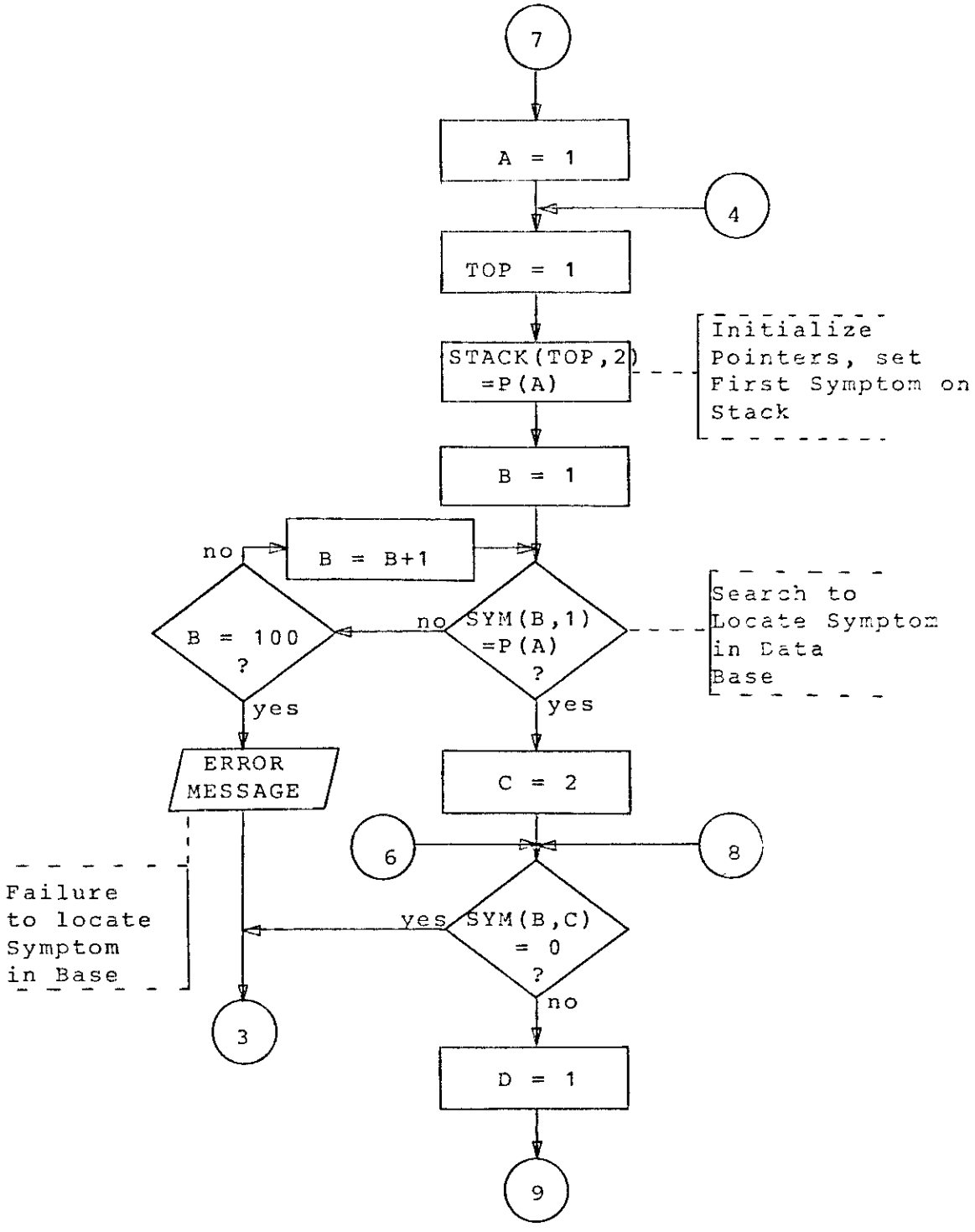


Fig. 4--Continued

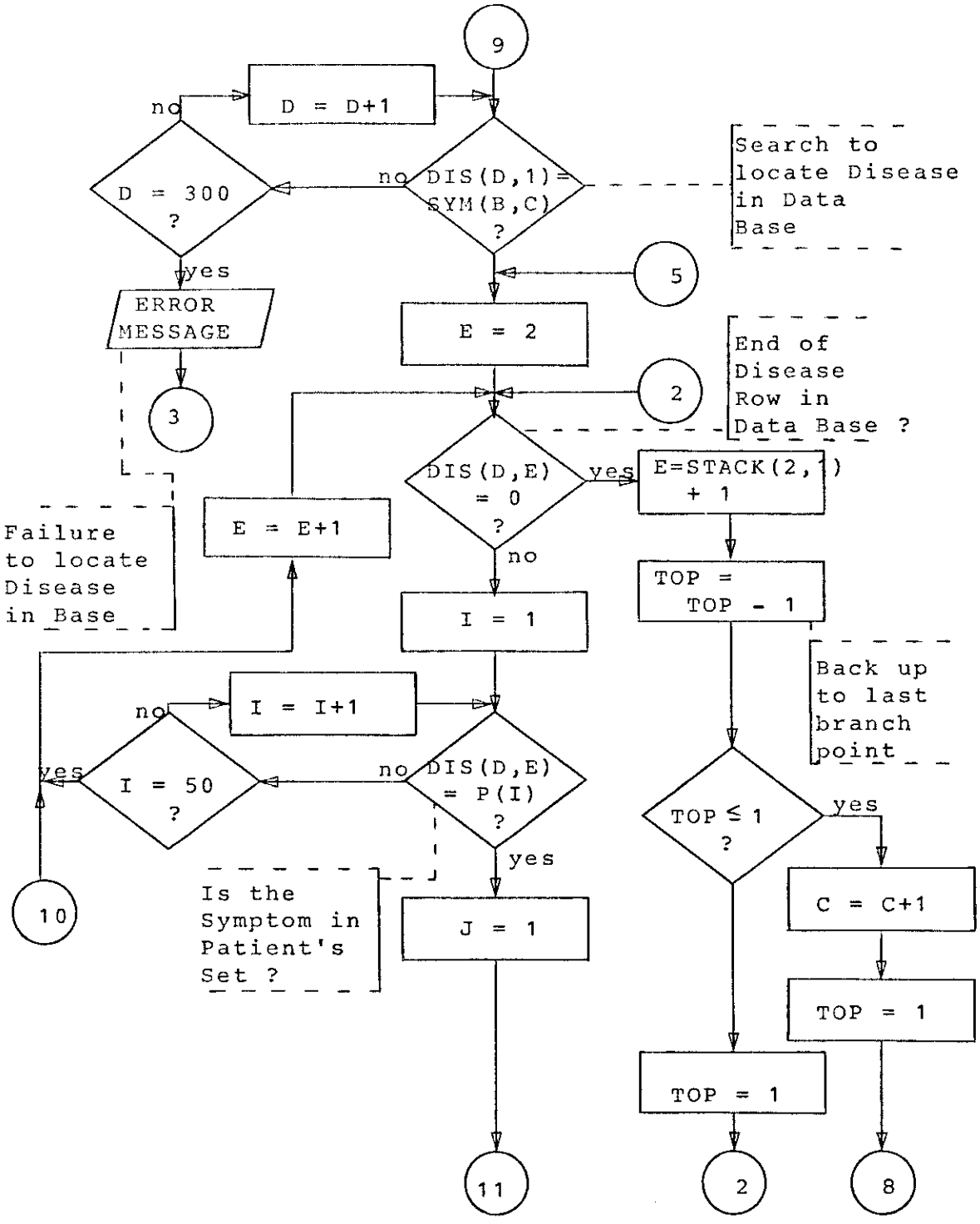


Fig. 4--Continued

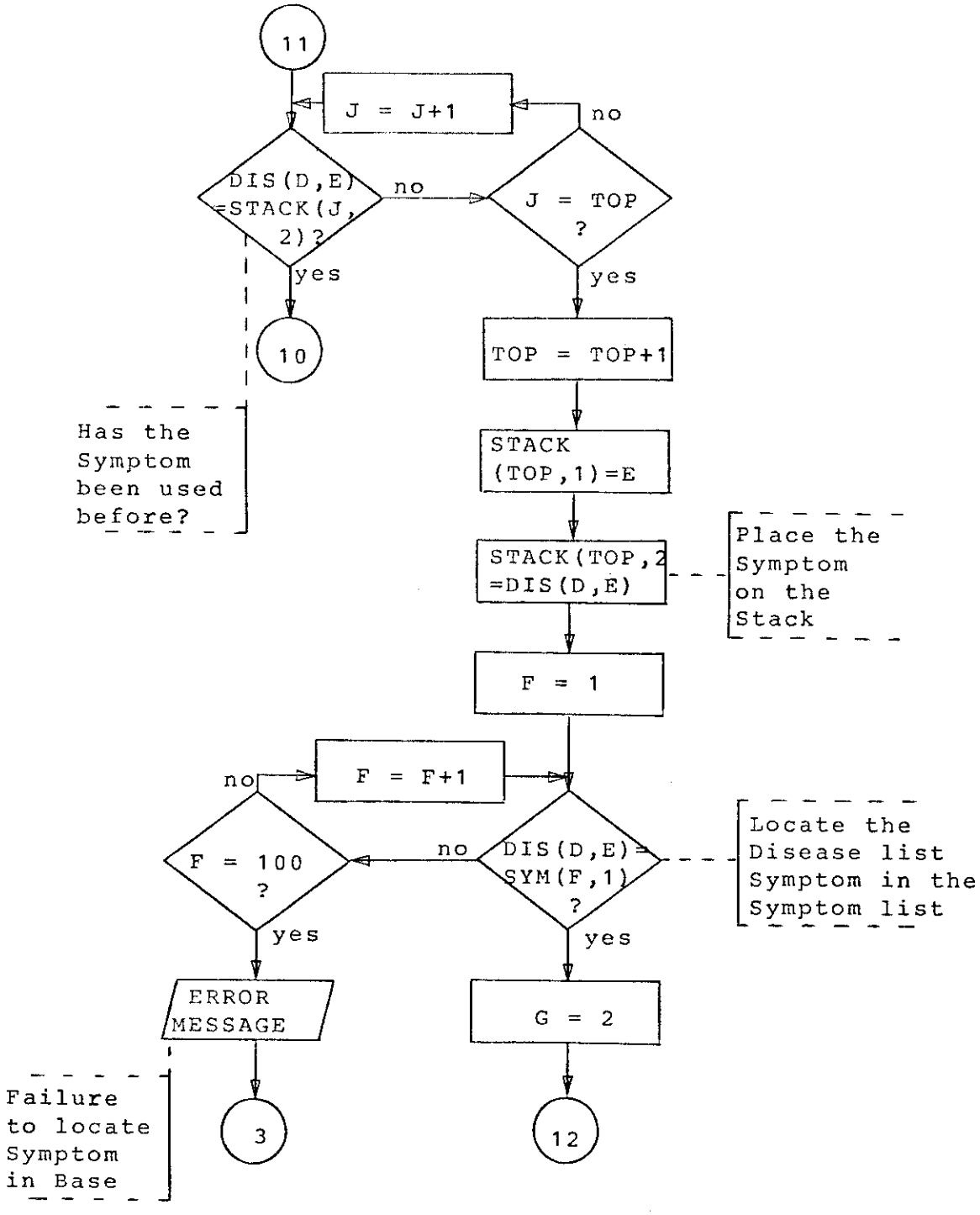


Fig. 4--Continued

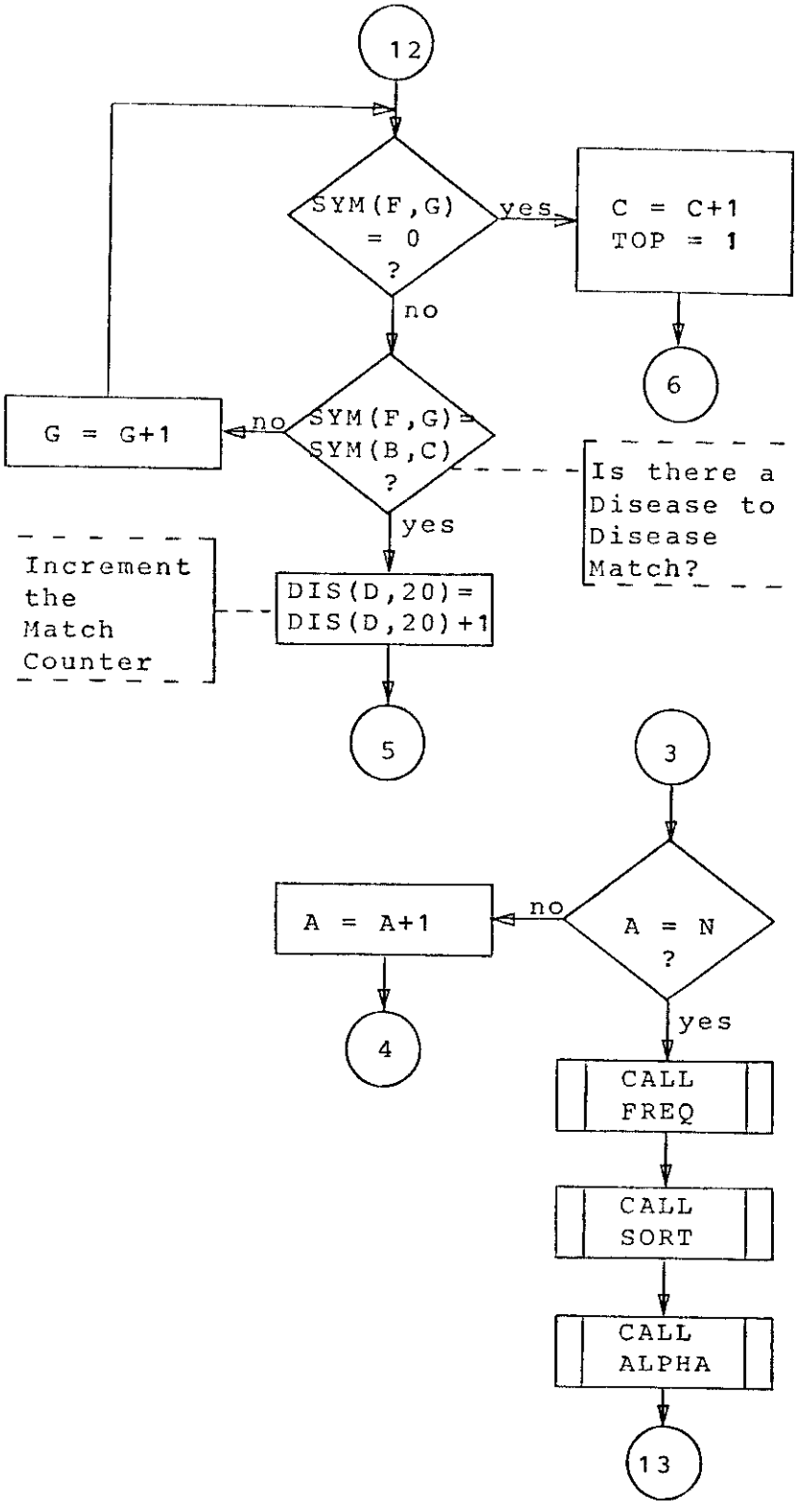


Fig. 4--Continued

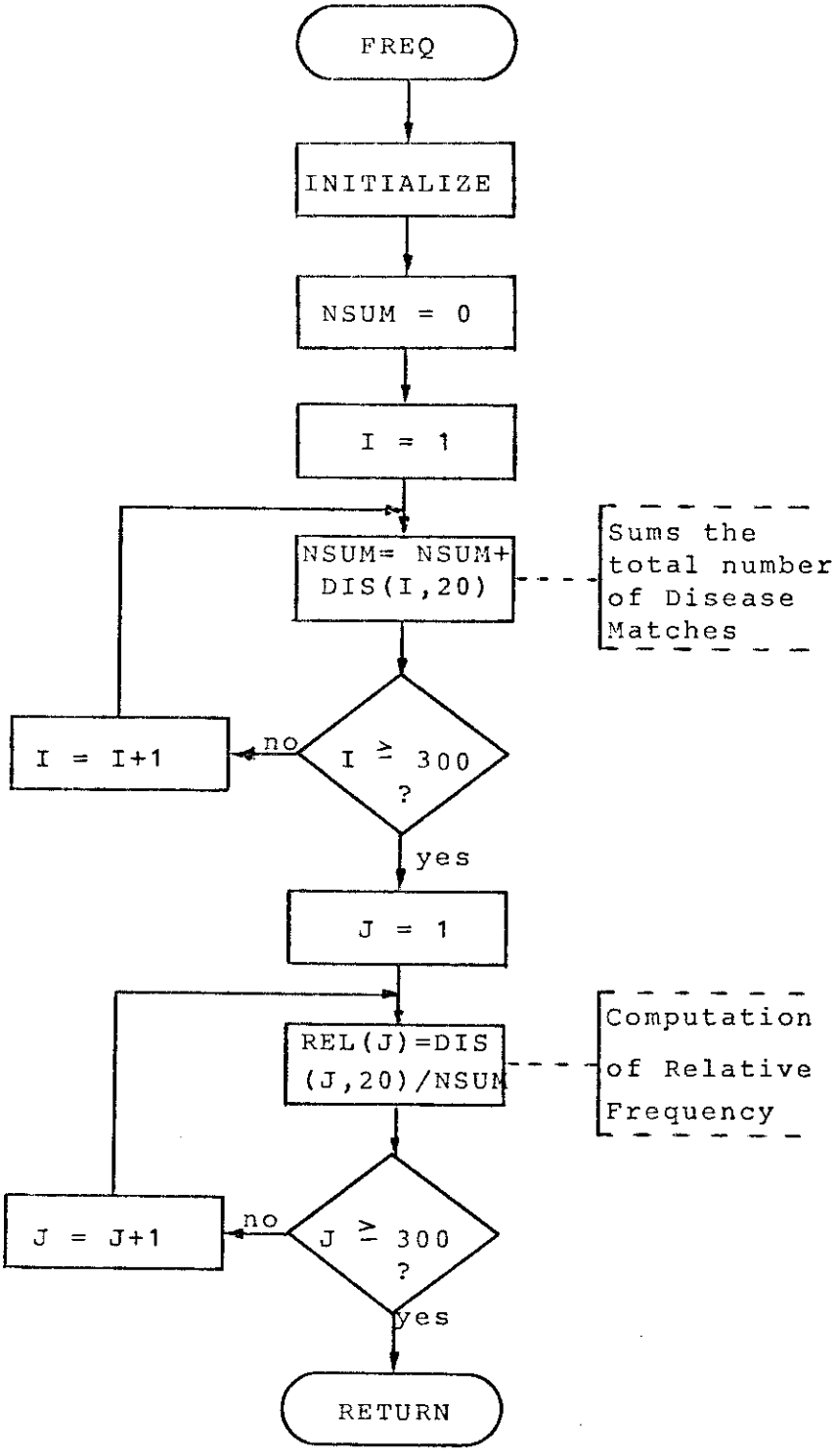


Fig. 5--Subroutine FREQ flowchart

base itself into a sorted form. The tag sort uses a special array in which are stored the subscrpits of the data base. A second subroutine is used to locate the largest element in the match column of the data base. When the largest element is found, its subscript is stored in the special index array, which when written out in order, automatically lists the diseases in their proper sequence. The flowcharts for the sort routine and its subroutine are Figures 6 and 7.

It was also considered desireable to provide for the alphabetic name of each disease in the list rather than a code number alone. Another subroutine was written which searches for a match between code numbers in the data base and code numbers in a name master file. This subroutine, Figure 8, also performs the task of writing out the sorted data in report form.

Detailed Discussion of Programs

The main program, Appendix F, written in FORTRAN IV, begins by initializing all necessary variables. Because the algorithm was written using the set of subscripts from A through G, it was decided to use the same subscripts in the program for uniformity. Data bases for symptoms, diseases

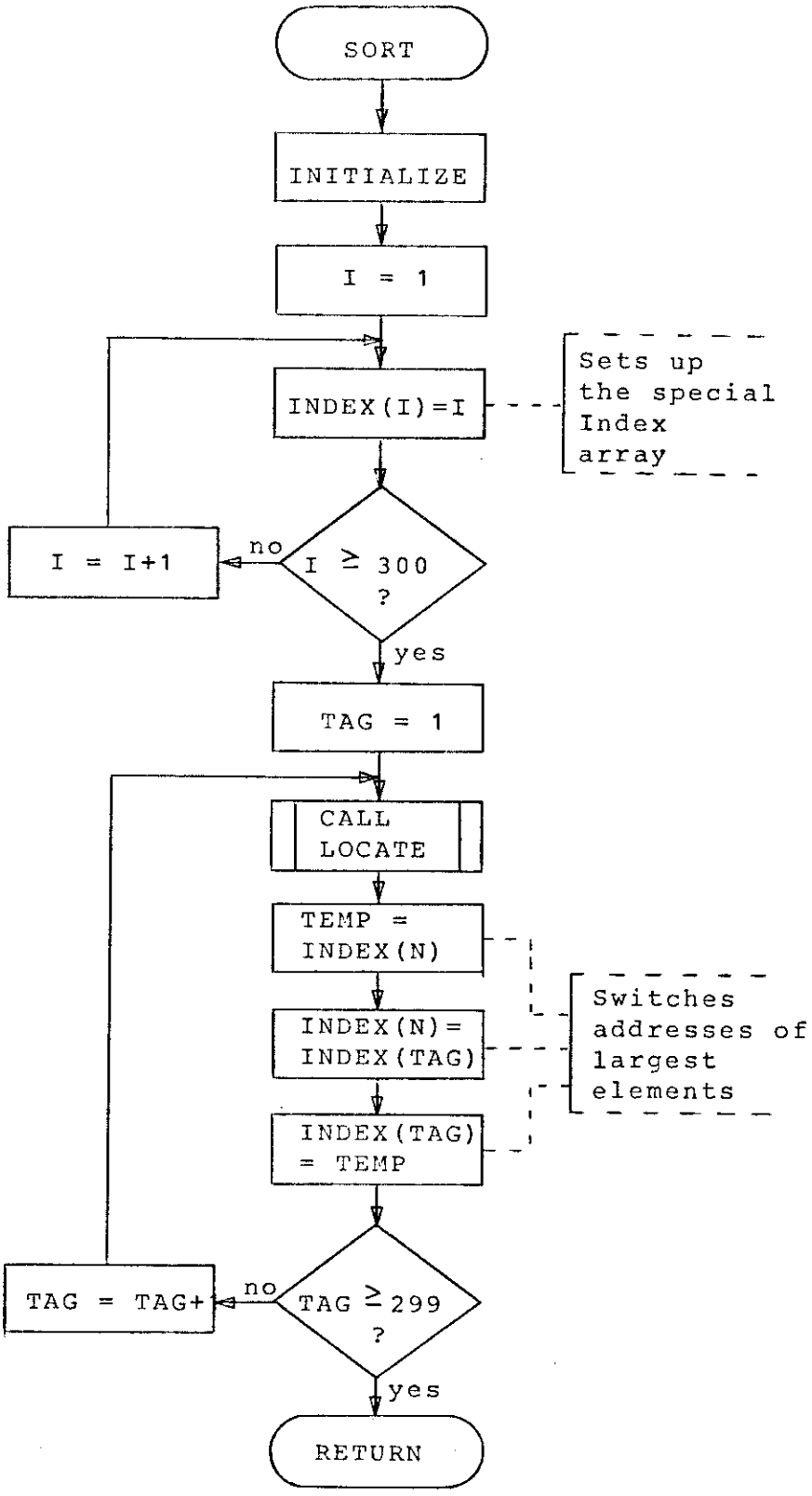


Fig. 6--Subroutine SORT flowchart

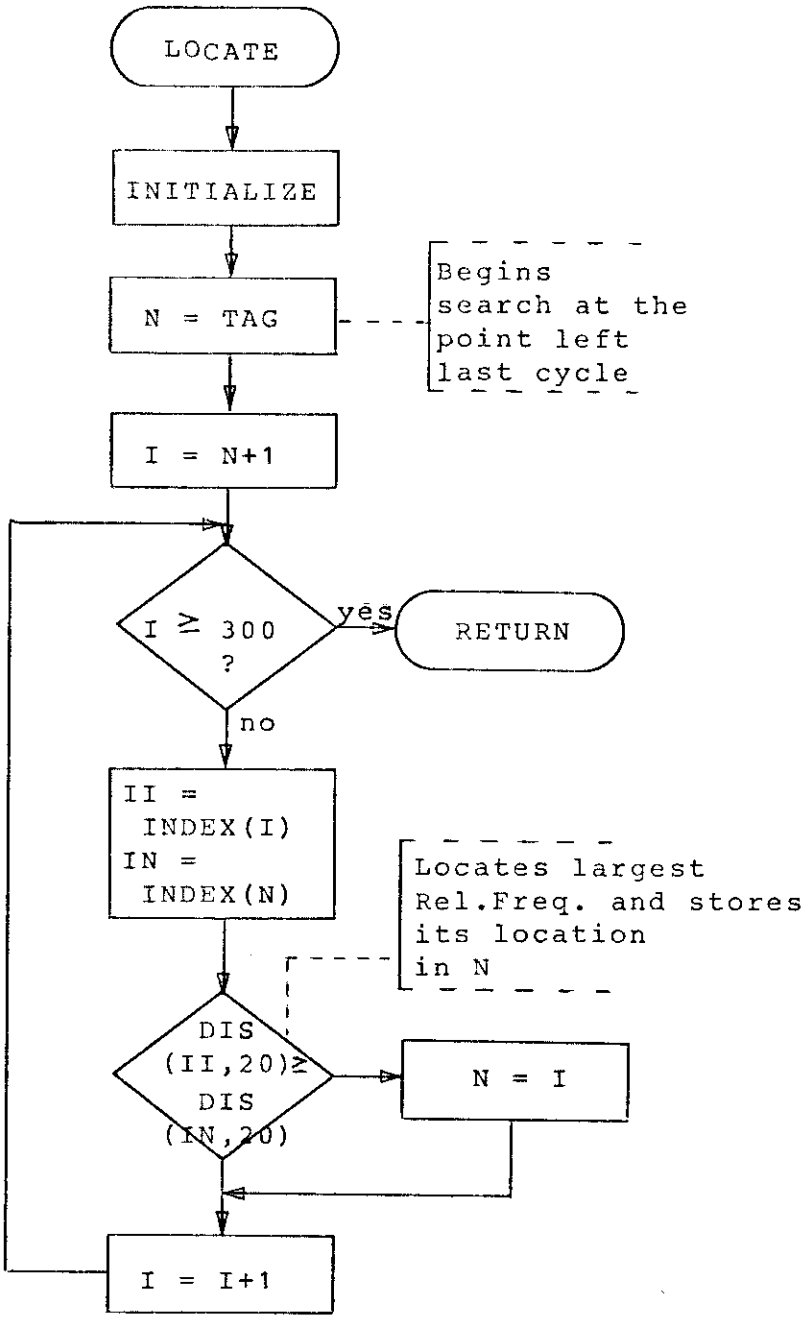


Fig. 7--Subroutine LOCATE flowchart

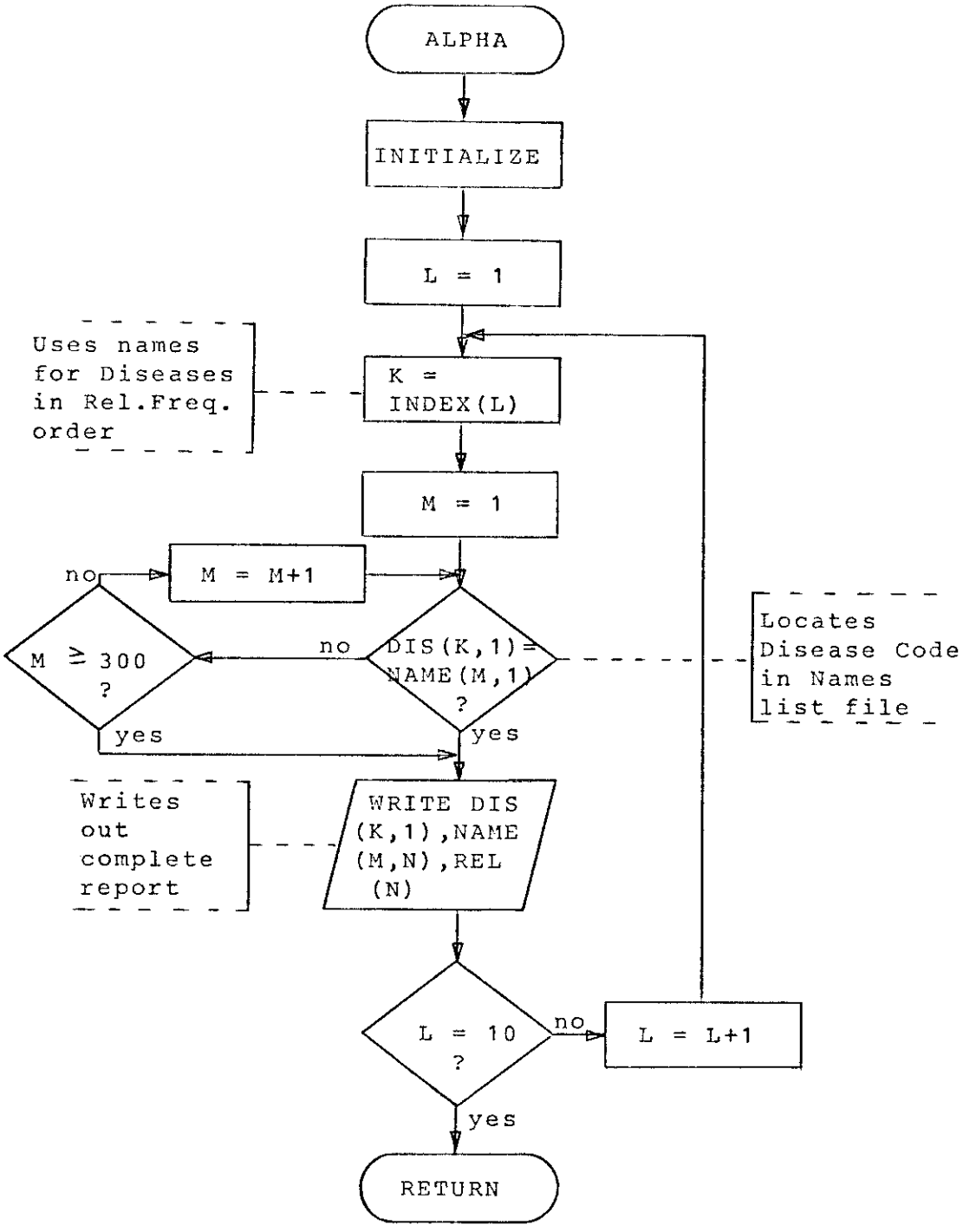


Fig. 8--Subroutine ALPHA flowchart

and disease names, called SYM, DIS and NAME respectively, are then read into storage.

The structure of SYM is a two-dimensional array in which the first column contains symptom codes. Following each symptom code, each row contains possible disease codes. The DIS array is a cross reference for the SYM array in that the first column contains disease codes and the row following each disease code has corresponding symptom codes. NAME is a two dimensional array in which the first column is a disease code and the row following is the alphabetic name of that disease.

The next read statement reads in data for a particular patient in the form of an identification number (IDNO), the number of symptoms exhibited (N) and his symptom set (P). Multiple Card Layout Forms for SYM, DIS, NAME, IDNO, N and P are included as Appendix G.

The last column of the DIS file, which has purposely been left blank and is to be used for accumulating the number of disease matches, and the relative frequency array (REL) are both initialized with values of zero. Headings and the patient's identification number are written next.

Main looping in the program is accomplished by means of a counter, A, which will count the number of iterations from 1 to the number of symptoms exhibited by the patient, N. A pointer, TOP, is initialized at 1 and the first element in the stack is given the code number of the first of the patient's symptoms, P(A). This is to insure that this symptom is not included in searches further down in the tree structure.

The first major searching loop uses B as its index and attempts to find the first symptom in the patient's set in the symptom master file, SYM. If no match is found, indicating the symptom is not in the symptom master file, an error message is printed and the program moves to the next symptom.

When a match is found the entire file of diseases for that symptom is made available as the row from column two to its end. After checking to be certain that the end of the row has not been reached, indicated by a zero, the first disease is located in the disease master file by means of a loop with D as its index. Again, if there is a failure to find this disease in the master file, an error message is printed and the program moves to the next symptom in the patient's set.

With a match, the disease file is made available as the row of that disease from column two to next to the end of the row. The last element in the row being reserved for a counter for the number of disease to disease matches. After determining that the end of the row has not been reached, indicated by a zero, two key tests are made. First, is the symptom, DIS(D,E) a part of the patient's symptom set, and second, has the symptom been searched in this branch before? If either of the tests fail, the program continues to loop until it reaches the end of the disease file row, at which point it backs up its disease pointer, E and its row counter for STACK, TOP until TOP is less than or equal to one, increments C, the disease index for the symptom master file and begins again with the next disease for the symptom in question.

If both of the conditions are met, the program moves to the next level of its STACK by incrementing TOP, and stores the current value of E and the symptom code for future reference in the previous tests. Two more loops are entered into in which a counter, F, is used to find the cross reference symptom in the symtom file and determine if a match can be made between the disease SYM(B,C) and its cross reference SYM(F,G). If a successful match is made, a counter, the last element in the disease row is incremented

and the process continues. If no match is found for F, an error message is printed and the program moves to the next symptom. If no match is found between SYM(E,C) and SYM(F,G), the disease pointer C is incremented, TOP is reset and the process begins again with a new disease.

When all of the symptoms in the patient's symptom set have been searched, the program calls for a series of subroutines. All of the subroutines and main programs have four arrays in COMMON, DIS, the disease master file, REL and INDEX which will be discussed later, and NAME, the disease description file.

The first subroutine called is FREQ, which calculates the relative frequency of each of the diseases tested. The first step is to add up all of the matches that occured during the search of the patient's symptom set. This is accomplished with an accumulator, NSUM and a loop which totals the last element of each row in the DIS array. When NSUM has been completed, it is used as the denominator in the computation for relative frequency, the numerator being each individual diseases number of matches. These relative frequencies are stored in REL, a one dimensional array which has the same subscript as each corresponding row of DIS.

The second subroutine to be called is SORT which sorts the number of matches per disease into descending order. Since it was considered impractical to move large amounts of data physically in memory, a method was selected for sorting which is known as the "tag" or "label" sort. This method uses the number of matches as the sort key but does not move the array DIS internally. The subscript for the DIS row is stored in INDEX which is moved. Therefore, instead of moving an entire row in a large array, only a single element is moved. A one dimensional array, INDEX, is first initialized with consecutive digits beginning with one. A pointer, TAG, is used to indicate which element is involved in the sort, as in INDEX(TAG). Another subroutine is called upon, LOCATE, to search the last column of DIS to find the largest element each time it is called and return the row subscript of that element to SORT as the variable N. In SORT an exchange between the largest element and the element occupying its place in INDEX is made such that the final product is a sequentially ordered list of subscripts stored in INDEX.

The last subroutine to be called is ALPHA. In this subroutine use is made of the sorted subscripts stored in INDEX from subroutine SORT. A variable K, is set equal to the first subscript stored in INDEX and a search is made

through the NAME array unitl disease DIS(K,1) is matched to disease code NAME(M,1) at which time the disease code, disease description and relative frequency are printed. From this subroutine, data for the ten most likely diseases is printed, although this could easily be modified to include any number of diseases.

After printing data for a case, the program returns to read another set of patient data. If there is no more data to be read, the last data card has three nines for the first group in patient identification number. When the nines are detected, control of the program is transfered to STOP.

In testing the program for accuracy in diagnosis, several cases of known diagnosis were chosen from the Student Health Clinic files to be run with the program. Two of these cases are reproduced as Figures 9 and 10. Original data sets for each of the patients is found in Table IV. In each case, the computer diagnosis proved to be either the same as the physicians or at least to point to an area upon which the physician would need to concentrate.

PRELIMINARY DIAGNOSIS BASED ON THE ROCOM HEALTH HISTORY QUESTIONNAIRE

PATIENT IDENTIFICATION NUMBER

XXX XX XXXX

DISEASE CLASSIFICATION CODE	DISEASE DESCRIPTION	RELATIVE FREQUENCY
5350	GASTRITIS-DUODENITIS	0.269
5410	APPENDICITIS	0.121
5339	PEPTIC ULCER NOS	0.121
0010	TYPHOID FEVER NOS	0.040
4700	INFLUENZA & FLU SYNDROME	0.040
4409	ARTERIOSCLEROSIS NOS	0.040
3640	IRITIS	0.040
3460	MIGRAINE	0.040
3039	ALCOHOLISM	0.040
2919	ALCOHOLIC, INCL DELERIUM TREMENS	0.040

Fig. 9--Tree sample output-Case 1

PRELIMINARY DIAGNOSIS BASED ON THE ROCOM HEALTH HISTORY QUESTIONNAIRE

PATIENT IDENTIFICATION NUMBER

XXX XX XXXX

DISEASE CLASSIFICATION CODE	DISEASE DESCRIPTION	RELATIVE FREQUENCY
4409	ARTERIOSCLEROSIS NOS	0.248
0113	TUBERCULOSIS, PUL, ACTIVE	0.111
0110	TUBERCULOSIS, PUL, ACTIVE, MIN.	0.111
0010	TYPHOID FEVER NOS	0.037
4270	CONGESTIVE HEART FAILURE NOS	0.037
3039	ALCOHOLISM	0.037
2919	ALCOHOLIC, INCL DELERIUM TREMENS	0.037
2510	OTHER PANCREATIC DISORDERS	0.037
0532	HERPES ZOSTER, OF TRUNK	0.037
7882	RASH NOS	0.006

Fig. 10--Tree sample output-Case 2

TABLE IV

DATA SETS FOR SAMPLE CASES

Case 1 Symptoms	Case 2 Symptoms
3	3
7	6
10	7
38	10
46	14
47	34
48	37
50	38
51	46
53	87
54	91
87	93
89	94
95	120
96	121
	128
Program Diagnosis:	
Gastro-Duodenitis	Arteriosclerosis
Actual Diagnosis:	
Gastroenteritis	Arteriosclerosis

CHAPTER BIBLIOGRAPHY

1. Taber, Clarence W., Taber's Cyclopedic Medical Dictionary, Philadelphia, F. A. Davis Co., 1958.

2. ROCOM Health History Questionnaire, Darien, Conn., Patient Care Systems, Inc., 1971.

3. Yater, Wallace M. And William F. Oliver, Symptom Diagnosis, New York, Appleton-Century-Crofts, Inc., 1961.

CHAPTER V

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Summary

The purpose of this study was to develop a technique or method by which a physician using a predetermined data base, could derive a preliminary diagnosis for a patient with a given set of symptoms. The system described in the previous chapter is able to accomplish this goal. Another benefit of the system comes from its ability to aid in the increasing of efficiency and speed of handling patients.

According to Stephen F. Austin State University Student Health Clinic records, the three full time physicians on the staff see an average of thirty to thirty five patients per day per physician during the school year. Their work load would be the equivalent to that of a city of ten thousand population which had only three physicians available. Any means which would be able to either reduce the number of patients or to speed up the processing of each patient would increase the efficiency of the clinic. Figure 11 shows how such a system is implemented in the clinic.

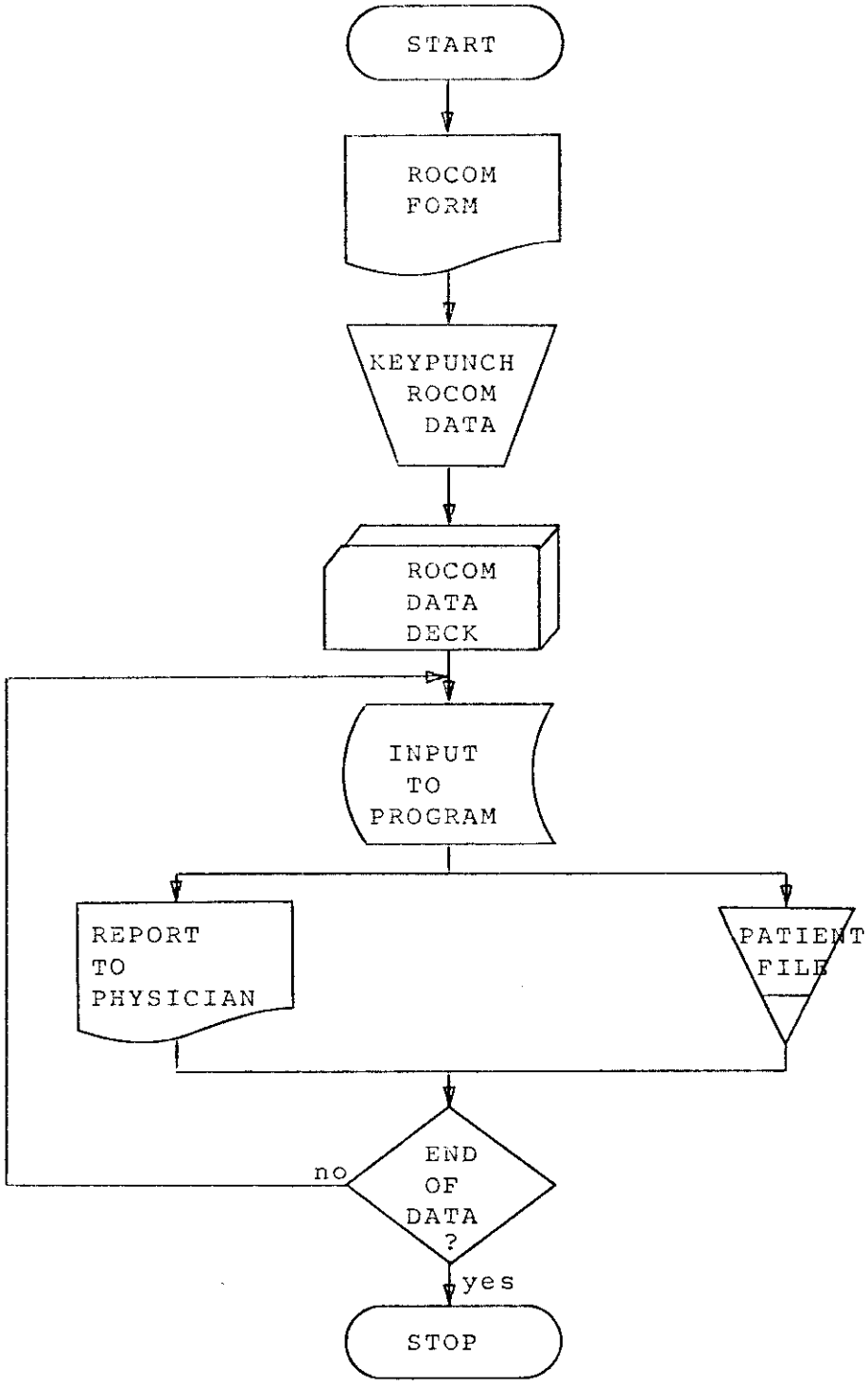


Fig. 11--Student Health Clinic system flowchart

When a patient enters the Clinic, he checks in with the receptionist on duty and is given a copy of the ROCOM Health History Questionnaire to be filled out. Average waiting time per patient is from fifteen to thirty minutes which allows ample time to complete the form. Upon completion of the questionnaire, the patient returns it to the receptionist who then keypunches the necessary data onto cards. The card deck is input to the system through a remote card reader. Report forms generated by the program are returned to the Clinic by a teleprinter for use by the physician and for inclusion in the patient's medical file.

The problem of erroneous source data resulting from a patient's failure to recognize a symptom was considered. Two points kept this from being a factor in this study. First, this study is intended to produce a preliminary diagnosis only and not the absolute diagnosis. Therefore the physician is directed toward possible diseases to be considered, or rejected as he desires. Secondly, the patient has physician contact and any symptoms not readily apparent or which the patient considers irrelevant, may be noted by the physician at that point who may modify the input data.

At present the University Computing Center uses an IBM 360 model 50 computer for all processing. Plans are for the installation of a XEROX 560 series computer with terminal capabilities early in 1975. The system presented here has been installed with the necessary programming changes incorporated to allow the use of a terminal in the Clinic when it becomes available.

Physicians will now be able to arrive at a diagnosis for any given patient faster than in the past. When a patient arrives in the Clinic he first registers with the receptionist and is given a ROCOM questionaire to be completed before seeing a physician. The questionaire becomes a part of the patient's permanent file. Depending upon the final configuration of hardware, the questionaire data will either be keypunched or directly input by means of a teletype keyboard in the Clinic. If data has been keypunched, it will be loaded into the computer through a card reader. Processing of one complete set of patient data takes on the order of one and one-half minutes, but with the new system and programs stored as object codes, it is estimated that this will require only one-half minute. Processing time per set of patient data will be reduced further if data sets are batch processed. When the physician sees the patient, he will have before him a

preliminary diagnosis which will enable him to concentrate on a particular area rather than wasting time eliminating other possibilities.

With all its ease of operation and flexibility, the system must be used as a tool and not as an ultimate solution for diagnosis. The ultimate responsibility for diagnosis, regardless of the means by which it is achieved, rests with the physician in charge of the case at hand.

In the words of Alvin Tofler, author of Future Shock ,

> Only when decision-makers are armed with better forecasts of future events, when by successive approximation we increase the accuracy of forecast, will our attempts to manage change improve perceptibly. For reasonably accurate assumptions about the future are a precondition for our understanding the potential consequences of our own actions. And without such understanding, the management of change is impossible. (2,p.470)

Conclusions

The basic system as described here, may be used in any installation which has, or has access to, a medium size computer such as the IBM 360 model 40 or 50, IBM 370 model 145, XEROX 560, Digital Equipment Company's PDP 10 or similar hardware, and thus has application in practically

any medical center for general practice.

The general approach to the problem of quantitative diagnosis has, with this study, been provided a viable alternative to the use of diagnosis based upon data bases and sources which leave doubt as to their validity. No reliance has been made here upon historical data which may be biased by a number of factors including interperetations of various symptoms and geographical bias.

Methodology used here reflects an attempt to force quantitative thinking into a new path with the overall objective of an upgrading in the process of medical diagnosis.

A by-product of the study has been that the basic diagnostic process is applicable to any specialized type of diagnosis and not only to a general clinical type application. For example, there are over one thousand diseases of the cornea of the eye alone with more than two thousand symptoms. (1,p.320) use of the system with a data base constructed for eye disease only could be of value to remote eye clinics where specialized diagnosis might not be available. Other specialized bases could be constructed on the same basis. As written the system is easily adaptable to any type of data base desired. Data bases of the type

required are simply read into the basic program. Processing and logic in the program, format of input and design of output remain the same.

Of particular value would be such a system used in large population centers where there are limited facilities and a high incidence of disease. With new hardware becoming available almost daily for the transmission and reception of data by telephone lines, microwave, radio and even satellite, possibilities for vast applications are great. Foreign applications in underdeveloped nations becomes feasible. Areas such as India, the Middle East and Africa where there are very few physicians per capita, but where communications lines exist or are being installed, could use the system to great advantage.

Shipboard applications are possible with the use of radio data transceivers which are already in use by the Navy. Hospital ship Hope would be able to use the system with a self-contained ship-board computer, if necessary, as would the large naval vessels such as aircraft carriers and cruisers.

The most obvious application is in a hospital, particularly in an out-patient department rather than for in-patients or emergency rooms. There would be limited use

for in-patients exept for preliminary diagnosis prior to more extensive operative techniques or laboratory procedures. Emergency patients most often require physician contact in a situation which demands immediate medical measures.

For the system detailed here, the ROCOM Health History Questionaire would not be essential for successful use. A questionaire, or check list of symptoms may be designed for use with a more specialized data base, thus giving the system a great deal more flexibility.

Recommendations

While the present system is able to achieve the goals specified, it should be possible to expand these goals to include a broader range of applications.

In the future, research may be channeled into the area of improving upon the concept of quantitative diagnosis. The combining of this method with a probabilistic technique may yield a more reliable diagnosis. Perhaps a two stage system may be produced which will give a preliminary diagnosis plus a series of recommended laboratory tests and examinations to be performed as a first stage. The second

stage would involve a reevaluation of the first stage diagnosis in terms of the test results which would produce a more specific diagnosis.

The possibility also exists for use of the system in non-medical decision situations. In cases where a cause-effect relationship is evident, a questionnaire or check list similar to that used for medical diagnosis may be developed which would lead to a decision point. This in combination with probabilities or weights and payoff tables should produce some interesting areas for future investigations.

Another area of investigation may be in the use of the system as a teaching tool for medical students. Given a set of symptoms, the student's diagnosis could be compared with that of the system.

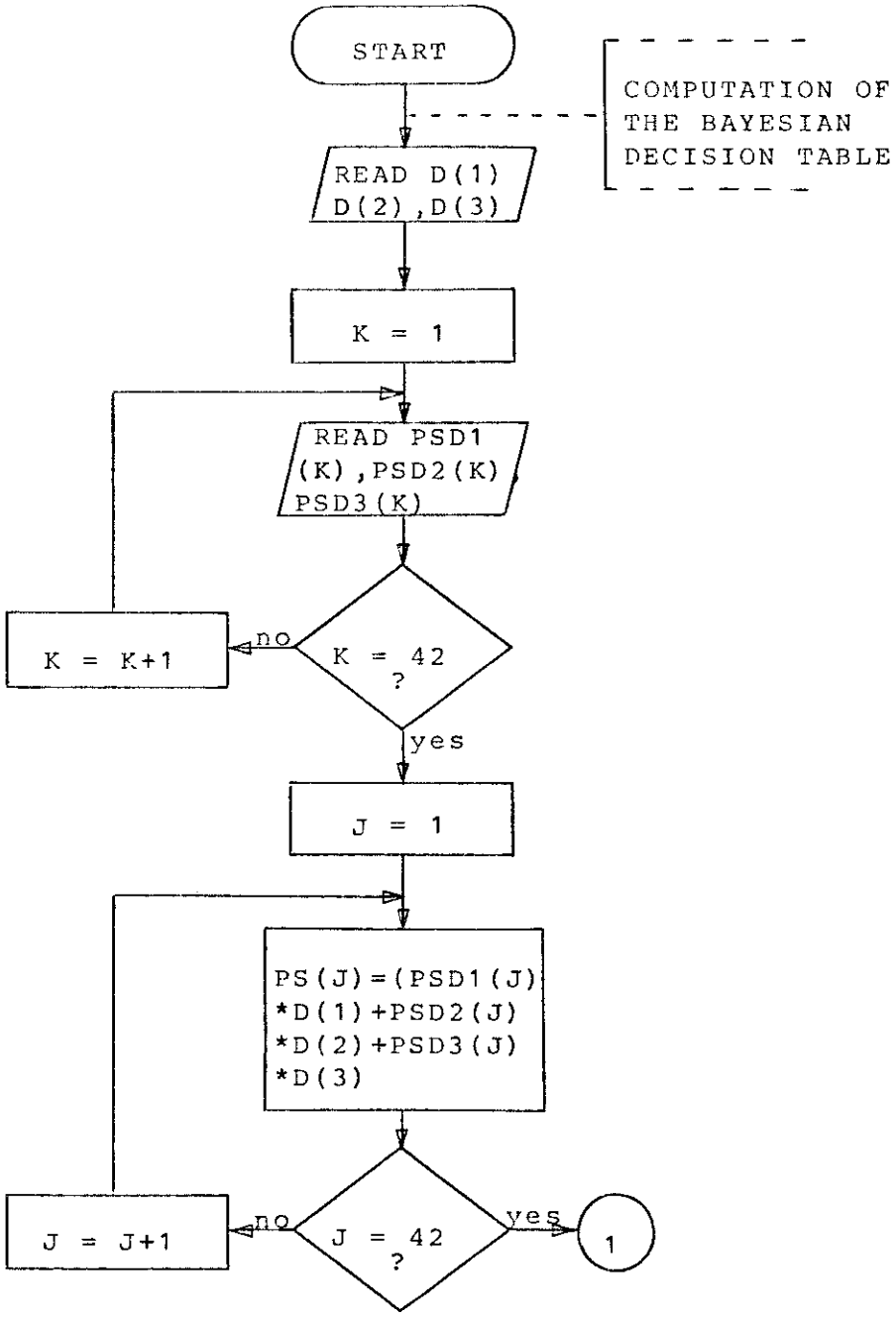
A great deal of additional research will be required to accomplish the goals suggested and to investigate further applications and validations of the system.

CHAPTER BIBLIOGRAPHY

1. Lusted, Lee B., "Computer Techniques in Medical Diagnosis," Computers in Biomedical Research, Vol. 1, New York, Academic Press, 1965.

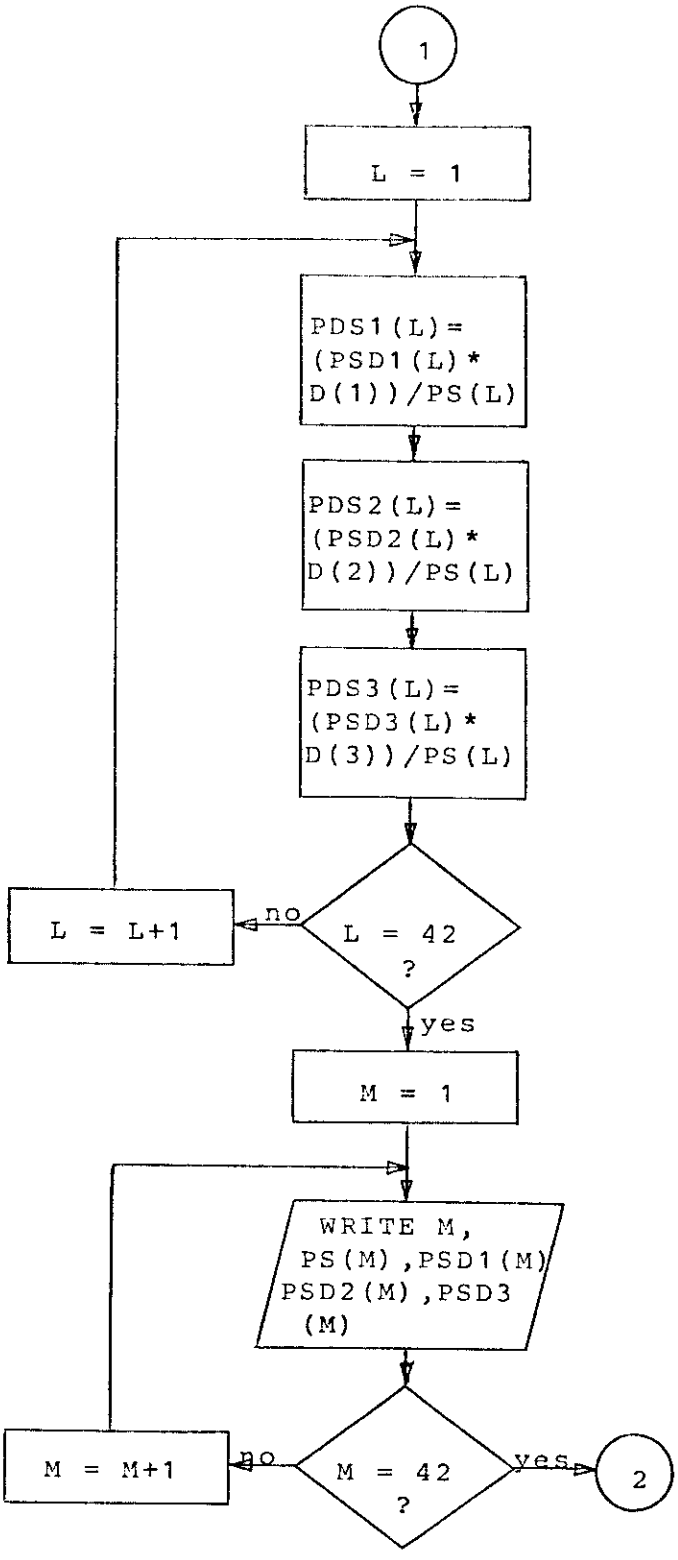
2. Tofler, Alvin, Future Shock, New York, Bantom Books, 1974.

APPENDICES

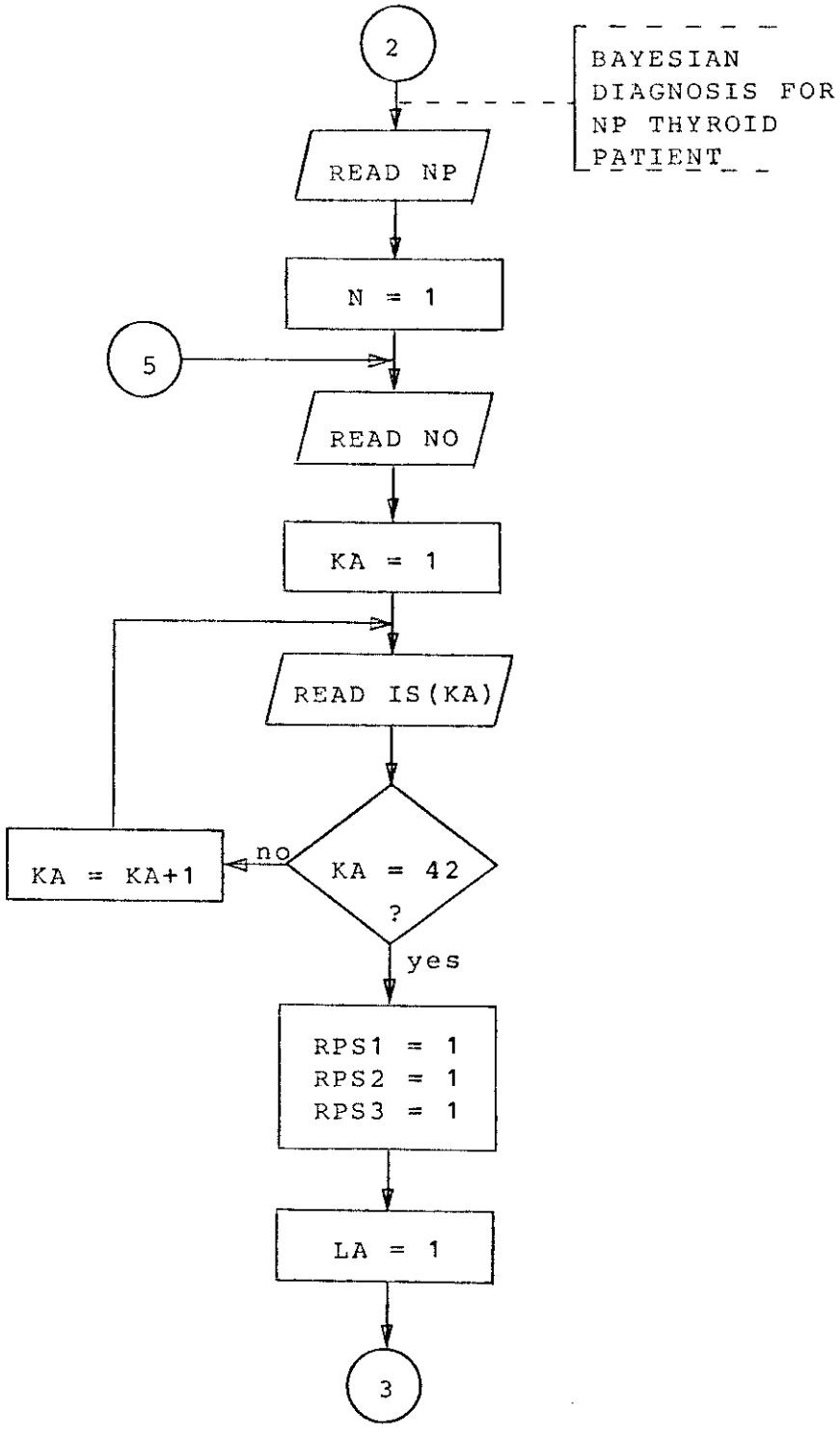


Thyroid Diagnosis Flowchart

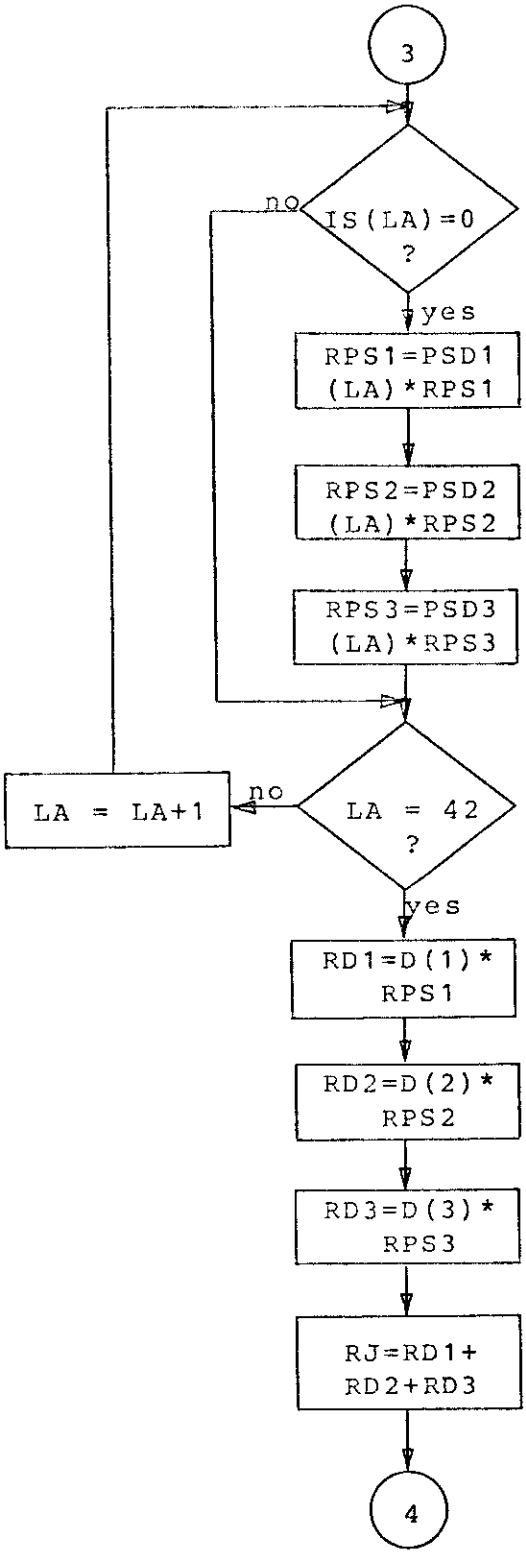
Appendix A



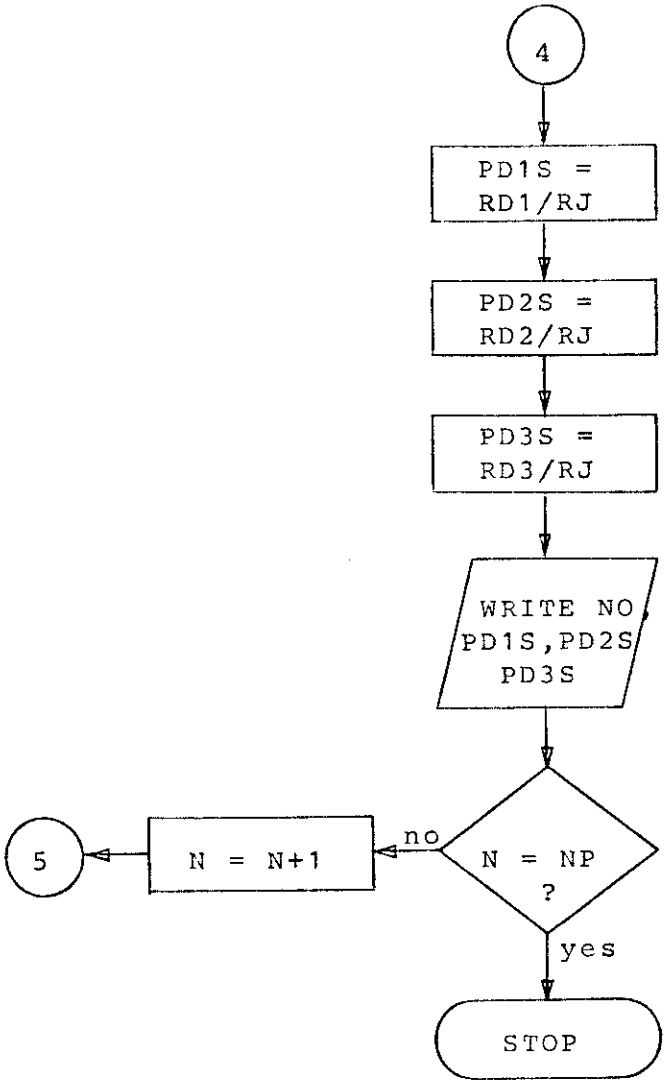
Appendix A--Continued



Appendix A--Continued



Appendix A--Continued



Appendix A--Continued

```
C  COMPUTATION OF A BAYESIAN DECISION TABLE
C  FOR THYROID DISEASE
C
      DIMENSION PS(42),PSD1(42),PSD2(42),PSD3(42)
      DIMENSION PDS1(42),PDS2(42),PDS3(42)
      DIMENSION D(3),IS(42)
      READ(5,1)D(1),D(2),D(3)
      READ(5,2)(PSD1(K),PSD2(K),PSD3(K),K=1,42)
      DO 10 J=1,42
   10 PS(J)=(PSD1(J)*D(1))+(PSD2(J)*D(2))+(PSD3(J)*D(3))
      DO 11 L=1,42
      PDS1(L)=(PSD1(L)*D(1))/PS(L)
      PDS2(L)=(PSD2(L)*D(2))/PS(L)
   11 PDS3(L)=(PSD3(L)*D(3))/PS(L)
      WRITE(6,6)
      WRITE(6,4)
      WRITE(6,5)
      DO 12 M=1,42
   12 WRITE(6,3)M,PS(M),PSD1(M),PSD2(M),PSD3(M),
     *PDS1(M),PDS2(M),PDS3(M)
      WRITE(6,6)
C
C COMPUTATION OF BAYESIAN PROBABILITIES FOR A
C PATIENT WITH A SET OF FIFTEEN POSSIBLE
C COMBINATIONS OF SYMPTOMS
C
      DO 35 N=1,3
      READ(5,8)NO,(IS(KA),KA=1,42)
      RPS1=1.
      RPS2=1.
      RPS3=1.
      DO 23 LA=1,42
      IF(IS(LA)-1)23,22,22
   22 RPS1=PSD1(LA)*RPS1
      RPS2=PSD2(LA)*RPS2
      RPS3=PSD3(LA)*RPS3
   23 CONTINUE
      RD1=D(1)*RPS1
      RD2=D(2)*RPS2
      RD3=D(3)*RPS3
      RJ=RD1+RD2+RD3
      PD1S=RD1/RJ
      PD2S=RD2/RJ
      PD3S=RD3/RJ
      WRITE(6,7)NO,PD1S,PD2S,PD3S
      STOP
      END
```

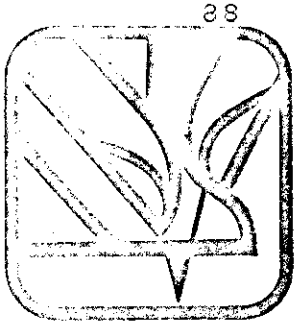
Bayesian Diagnosis Program Listing

Appendix B

SYMPTOM	P(S)	P(D1/S)	P(D2/S)	P(D3/S)
1	.437	.421	.018	.562
2	.261	.566	.001	.434
3	.331	.446	.056	.497
4	.186	.616	.022	.363
5	.484	.436	.036	.617
6	.209	.653	.001	.346
7	.229	.656	.001	.343
8	.326	.545	.006	.449
9	.172	.118	.690	.193
10	.238	.078	.479	.443
11	.043	.005	.841	.154
12	.200	.158	.247	.595
13	.176	.036	.399	.565
14	.149	.016	.816	.168
15	.186	.012	.672	.317
16	.129	.040	.806	.154
17	.141	.624	.142	.234
18	.036	.006	.976	.018
19	.477	.000	.220	.780
20	.190	.060	.001	.940
21	.203	.506	.001	.493
22	.096	.875	.001	.124
23	.095	.002	.942	.056
24	.124	.002	.309	.690
25	.532	.043	.024	.933
26	.109	.357	.001	.642
27	.142	.962	.001	.037
28	.093	.022	.934	.064
29	.099	.022	.336	.662
30	.533	.000	.038	.962
31	.101	.293	.001	.706
32	.174	.965	.001	.034
33	.091	.002	.933	.065
34	.114	.009	.307	.685
35	.520	.028	.038	.934
36	.136	.372	.001	.627
37	.138	.946	.001	.053
38	.115	.002	.809	.189
39	.091	.022	.515	.483
40	.554	.000	.000	.999
41	.055	.205	.003	.793
42	.188	.996	.001	.004

Decision Table for Thyroid Diseases

Appendix C



ROCOM

Health History Questionnaire

The health history questionnaire you are about to fill out is important to your doctor. It gives him information he needs about your health which only you can tell him.

The questionnaire is divided into sections. Please read the instructions given with each section before answering the questions. Please PRINT, using a ballpoint pen, when you are asked to complete information. Make an (X) where you are asked to do so, pressing firmly on the pen.

Take the time you need to finish the questionnaire. Do not worry about questions you cannot answer. If you are not sure how a question should be answered, place a solid circle (●) in the "Yes" column or in the space provided if it is not a "Yes/No" question. You will have a chance to go over these questions with the doctor during your appointment.

NOTE: Spread the questionnaire out flat on a hard surface when filling it out. Do not fold or double it back on itself.

Created and Developed by Patient Care Systems, Inc., Darien, Conn.

Used by permission.

Please answer each of the following questions by placing an (X) in the "Yes" blank at the r
if your answer to the questions is yes, or by placing an (X) in the "No" blank at the right if y
answer to the question is no. If you are unable to answer a question for any reason, place a s
circle (●) in the "Yes" blank.

1. Are you troubled with stiff or painful muscles or joints?
2. Are your joints ever swollen?
3. Are you troubled by pains in the back or shoulder?
4. Are your feet often painful?
5. Are you handicapped in any way?

6. Do you have any skin problems?
7. Does your skin itch or burn?
8. Do you have trouble stopping even a small cut from bleeding?
9. Do you bruise easily?

10. Do you ever faint or feel faint?
11. Is any part of your body always numb?
12. Have you ever had fits or convulsions?
13. Has your handwriting changed lately?
14. Do you have a tendency to shake or tremble?

15. Are you very nervous around strangers?
16. Do you find it hard to make decisions?
17. Do you find it hard to concentrate or remember?
18. Do you usually feel lonely or depressed?
19. Do you often cry?
20. Would you say you have a hopeless outlook?
21. Do you have difficulty relaxing?
22. Do you have a tendency to worry a lot?
23. Are you troubled by frightening dreams or thoughts?
24. Do you have a tendency to be shy or sensitive?
25. Do you have a strong dislike for criticism?
26. Do you lose your temper often?
27. Do little things often annoy you?
28. Are you disturbed by any work or family problems?
29. Are you having any sexual difficulties?
30. Have you ever considered committing suicide?
31. Have you ever desired or sought pyschiatric help?

32. Have you gained or lost much weight recently?
33. Do you have a tendency to be too hot or too cold?
34. Have you lost your interest in eating lately?
35. Do you always seem to be hungry?
36. Are there any swellings in your armpits or groin?
37. Do you seem to feel exhausted or fatigued most of the time?
38. Do you have difficulty either falling or staying asleep?
39. Do you fail to get the exercise you should?
40. Do you smoke?
41. Do you take two or more alcoholic drinks a day?
42. Do you drink more than six cups of coffee or tea a day?
43. Have you ever used marijuana?
44. Have you ever used heroin, LSD or similar drugs?
45. Do you bite your nails?

TURN TO THE NEXT PAGE

46. Are you troubled by heartburn? .....
47. Do you feel bloated after eating? .....
48. Are you troubled by belching? .....
49. Do you suffer discomfort in the pit of your stomach? .....
50. Do you easily become nauseated (feel like vomiting)? .....
51. Have you ever vomited blood? .....
52. Is it difficult or painful for you to swallow? .....
53. Are you constipated more than twice a month? .....
54. Are your bowel movements ever loose for more than one .....
55. Are your bowel movements ever black or bloody? .....
56. Are your bowel movements ever grey in color? .....
57. Do you suffer pains when you move your bowels? .....
58. Have you had any bleeding from your rectum? .....

59. Do you frequently get up at night to urinate? .....
60. Do you urinate more than five or six times a day? .....
61. Do you wet your pants or wet your bed? .....
62. Have you ever had burning or pains when you urinate? .....
63. Has your urine ever been brown, black or bloody? .....
64. Do you have any difficulty starting your urine flow? .....
65. Do you have a constant feeling that you have to urinate? .....

For Men Only

66. Is your urine stream very weak and slow? .....
67. Has a doctor ever told you that you have prostate trouble? .....
68. Have you had any burning or discharge from your penis? .....
69. Are there any swellings or lumps on your testicles? .....
70. Do your testicles get painful? .....

For Women Only

71. Are you having trouble with your menstrual periods? .....
72. Have you ever had bleeding between your periods? .....
73. Do you have heavy bleeding with your periods? .....
74. Do you feel bloated and irritable before your period? .....
75. Have you ever taken any birth control pills? .....
76. Have you ever had any lumps in your breasts? .....
77. Have you had any excessive discharges from your vagina? .....
78. Please print the month and year of your last PAP smear. .....
79. Please print the date your last menstrual period began. .....

Print the following information in the spaces at the right:

80. Number of pregnancies. .....
81. Number of miscarriages. .....
82. Number of stillbirths. .....
83. Number of premature births. .....
84. Number of children born alive. .....
85. Number of cesarean operations. .....
86. Have you ever had an abortion? .....

TURN TO BACK OF THIS PAGE

Name _____ Date _____

7. Yes ___ No ___
8. Yes ___ No ___
9. Yes ___ No ___

0. Yes ___ No ___
1. Yes ___ No ___
2. Yes ___ No ___
3. Yes ___ No ___
4. Yes ___ No ___
5. Yes ___ No ___
6. Yes ___ No ___
7. Yes ___ No ___

8. Yes ___ No ___
9. Yes ___ No ___
0. Yes ___ No ___
1. Yes ___ No ___
2. Yes ___ No ___

3. Yes ___ No ___
4. Yes ___ No ___
5. Yes ___ No ___
6. Yes ___ No ___

7. Yes ___ No ___
8. Yes ___ No ___
9. Yes ___ No ___
0. Yes ___ No ___
1. Yes ___ No ___
2. Yes ___ No ___
3. Yes ___ No ___
4. Yes ___ No ___

5. Yes ___ No ___
6. Yes ___ No ___
7. Yes ___ No ___
8. Yes ___ No ___
9. Yes ___ No ___
0. Yes ___ No ___

1. Yes ___ No ___
2. Yes ___ No ___
3. Yes ___ No ___
4. Yes ___ No ___
5. Yes ___ No ___
6. Yes ___ No ___
7. Yes ___ No ___
8. Yes ___ No ___
9. Yes ___ No ___

46. Yes ___ No ___
47. Yes ___ No ___
48. Yes ___ No ___
49. Yes ___ No ___
50. Yes ___ No ___
51. Yes ___ No ___
52. Yes ___ No ___
53. Yes ___ No ___
54. Yes ___ No ___
55. Yes ___ No ___
56. Yes ___ No ___
57. Yes ___ No ___
58. Yes ___ No ___

59. Yes ___ No ___
60. Yes ___ No ___
61. Yes ___ No ___
62. Yes ___ No ___
63. Yes ___ No ___
64. Yes ___ No ___
65. Yes ___ No ___

66. Yes ___ No ___
67. Yes ___ No ___
68. Yes ___ No ___
69. Yes ___ No ___
70. Yes ___ No ___

71. Yes ___ No ___
72. Yes ___ No ___
73. Yes ___ No ___
74. Yes ___ No ___
75. Yes ___ No ___
76. Yes ___ No ___
77. Yes ___ No ___
78. ___ ___
79. ___ ___

80. ___ ___
81. ___ ___
82. ___ ___
83. ___ ___
84. ___ ___
85. ___ ___
86. Yes ___ No ___

1. Yes ___ No ___
2. Yes ___ No ___
3. Yes ___ No ___
4. Yes ___ No ___
5. Yes ___ No ___

6. Yes ___ No ___
7. Yes ___ No ___
8. Yes ___ No ___
9. Yes ___ No ___

10. Yes ___ No ___
11. Yes ___ No ___
12. Yes ___ No ___
13. Yes ___ No ___
14. Yes ___ No ___

15. Yes ___ No ___
16. Yes ___ No ___
17. Yes ___ No ___
18. Yes ___ No ___
19. Yes ___ No ___
20. Yes ___ No ___
21. Yes ___ No ___
22. Yes ___ No ___
23. Yes ___ No ___
24. Yes ___ No ___
25. Yes ___ No ___
26. Yes ___ No ___
27. Yes ___ No ___
28. Yes ___ No ___
29. Yes ___ No ___
30. Yes ___ No ___
31. Yes ___ No ___

32. Yes ___ No ___
33. Yes ___ No ___
34. Yes ___ No ___
35. Yes ___ No ___
36. Yes ___ No ___
37. Yes ___ No ___
38. Yes ___ No ___
39. Yes ___ No ___
40. Yes ___ No ___
41. Yes ___ No ___
42. Yes ___ No ___
43. Yes ___ No ___
44. Yes ___ No ___
45. Yes ___ No ___

Special problems or symptoms: _____

Name _____ Date _____ Patient no. _____

Doctor's notes _____

1. HEAD and NECK
- ___ frequent headaches
- ___ neck pains
- ___ neck lumps or swelling

2. EYES
- ___ wears glasses
- ___ blurry vision
- ___ eyesight worsening
- ___ sees double
- ___ sees halos
- ___ eye pains or itching
- ___ watering eyes
- ___ eye trouble

3. EARS
- ___ hearing difficulties
- ___ earaches
- ___ running ears
- ___ buzzing in ears
- ___ motion sickness

4. MOUTH
- ___ dental problems
- ___ swellings on gums or jaws
- ___ sore tongue
- ___ taste changes

5. NOSE and THROAT
- ___ congested nose
- ___ running nose
- ___ sneezing spells
- ___ headcolds
- ___ nose bleeds
- ___ sore throat
- ___ enlarged tonsils
- ___ hoarse voice

6. RESPIRATORY
- ___ wheezes or gasps
- ___ coughing spells
- ___ coughs up phlegm
- ___ coughed up blood
- ___ chest colds
- ___ excessive sweating

7. CARDIOVASCULAR
- ___ high blood pressure
- ___ racing heart
- ___ chest pains
- ___ dizzy spells
- ___ shortness of breath
- ___ swollen feet or ankles
- ___ leg cramps
- ___ hot flashes
- ___ heart murmur

8. DIGESTIVE
- heartburn ___
- bloated stomach ___
- belching ___
- stomach pains ___
- nausea ___
- vomited blood ___
- difficulty swallowing ___
- constipation ___
- loose bowels ___
- black stools ___
- grey stools ___
- pain in rectum ___
- rectal bleeding ___

9. URINARY
- night frequency ___
- day frequency ___
- wets pants or bed ___
- burning on urination ___
- brown, black or bloody urine ___
- difficulty starting urine ___
- urgency ___

10. MALE GENITAL
- weak urine stream ___
- prostate trouble ___
- burning or discharge ___
- lumps on testicles ___
- painful testicles ___

11. FEMALE GENITAL
- menstrual trouble ___
- breakthrough bleeding ___
- heavy bleeding ___
- premenstrual tension ___
- birth control pill ___
- lumps in breasts ___
- vaginal discharge ___
- PAP smear _____
- last period _____

12. PREGNANCIES
- gravida _____
- miscarriages _____
- stillbirths _____
- premature births _____
- para _____
- cesareans _____
- abortion ___

13. MUSCULOSKELETAL
- ___ aching muscles or joints
- ___ swollen joints
- ___ back or shoulder pains
- ___ painful feet
- ___ handicapped

14. SKIN
- ___ skin problems
- ___ itching or burning skin
- ___ bleeds easily
- ___ bruises easily

15. NEUROLOGICAL
- ___ faintness
- ___ numbness
- ___ convulsions
- ___ change in handwriting
- ___ trembles

16. MOOD
- ___ nervous with strangers
- ___ difficulty making decisions
- ___ lack of concentration or memor
- ___ lonely or depressed
- ___ cries often
- ___ hopeless outlook
- ___ difficulty relaxing
- ___ worries a lot
- ___ frightening dreams or thoughts
- ___ shy or sensitive
- ___ dislikes criticism
- ___ loses temper
- ___ annoyed by little things
- ___ work or family problems
- ___ sexual difficulties
- ___ considered suicide
- ___ desired psychiatric help

17. GENERAL
- ___ weight changes
- ___ tends to be hot or cold
- ___ loss of interest in eating
- ___ always hungry
- ___ armpits or groin swelling
- ___ fatigue
- ___ sleeping difficulties
- ___ lack of exercise
- ___ smokes
- ___ drinks alcohol daily
- ___ heavy coffee or tea drinker
- ___ marijuana
- ___ heroin, LSD, similar drugs
- ___ bites nails

Special problems or symptoms: _____

Disease Code (DCC)	Description
0000-1399	Infective and Parasitic Diseases
1400-2099	Malignancies
2100-2299	Benign Tumors
2300-2399	Neoplasms
2400-2799	Endocrine and Metabolic Diseases
2800-2899	Diseases of Blood
2900-2999	Psychoses
3000-3199	Neuroses
3200-3499	Central Nervous System Diseases
3500-3599	Peripheral Nerves and Ganglia
3600-3799	Diseases of eye
3800-3899	Diseases of Ear and Mastoid
3900-4599	Diseases of Circulatory System
4600-5199	Diseases of Respiratory System
5200-5269	Dental Conditions
5270-5799	Diseases of Digestive System
5800-5999	Diseases of Kidney and Bladder
6000-6099	Diseases of Male Genitalia
6100-6199	Diseases of Breast and Ovaries
6200-6299	Diseases of Female Genitalia
6300-6799	Complications of Pregnany
6800-7099	Diseases of Skin
7100-7399	Diseases of Musculo-Skeletal System
7400-7599	Congenital Anomolies
7600-7799	Not Used
7800-7999	Ill Defined Conditions
8000-8299	Fractures
8300-8399	Dislocations
8400-8499	Strains and Sprains
8500-8599	Intracranial Injury
8600-8699	Internal Injury
8700-9099	Lacerations
9100-9199	Superficial Injury
9200-9299	Contusions
9300-9399	Foriegn Body Entering Orifice
9400-9499	Burns
9500-9599	Nerves and Spinal Chord
9600-9799	Effects of Chemicals
9800-9899	Effects of Non-Medicinal Compounds
9900-9999	Effects of Physical Substances

Disease Classification Codes

Appendix E

```
      INTEGER SYM(100,50),P(50),STACK(50,2)
      INTEGER A,B,C,D,E,F,G,TOP
      INTEGER DIS(300,20)
      DIMENSION REL(300),INDEX(300),NAME(300,41)
      DIMENSION IDNO(3)
      COMMON DIS,REL,INDEX,NAME
      READ(5,1)((SYM(I,J),J=1,50),I=1,100)
      READ(5,2)((DIS(K,L),L=1,20),K=1,300)
      READ(5,8)((NAME(I,J),J=1,41),I=1,300)
150   READ(5,3)IDNO,N,P
      IF(IDNO(1) .EQ. 999)GO TO 200
      DO 17 IJ=1,300
      DIS(IJ,20)=0
 17   REL(IJ)=0.
      WRITE(6,4)IDNO
      DO 99 A=1,N
      TOP=1
      STACK(TOP,2)=P(A)
      DO 10 B=1,100
      IF(SYM(B,1) .EQ. P(A))GO TO 15
 10   CONTINUE
      WRITE(6,5)P(A)
      GO TO 99
 15   C=2
 32   IF(SYM(B,C) .EQ. 0)GO TO 99
      DO 20 D=1,300
      IF(DIS(D,1) .EQ. SYM(B,C))GO TO 25
 20   CONTINUE
      WRITE(6,6)SYM(B,C)
      GO TO 99
 25   E=2
 71   IF(DIS(D,E) .EQ. 0)GO TO 30
 33   CONTINUE
      DO 35 I=1,50
      IF(DIS(D,E) .EQ. P(I))GO TO 40
 35   CONTINUE
      E=E+1
      GO TO 71
 40   DO 45 J=1,TOP
      IF(DIS(D,E) .EQ. STACK(J,2))GO TO 50
 45   CONTINUE
      TOP=TOP+1
      STACK(TOP,1)=E
      STACK(TOP,2)=DIS(D,E)
```

Tree Main Program Listing

Appendix F

```
      GO TO 55
   50 E=E+1
      GO TO 71
   30 E=STACK(2,1)+1
      TOP=TOP-1
      IF(TOP .LE. 1)GO TO 70
      TOP=1
      GO TO 71
   55 DO 60 F=1,100
      IF(DIS(D,E) .EQ. SYM(F,1))GO TO 65
   60 CONTINUE
      WRITE(6,7)DIS(D,E)
      GO TO 99
   65 G=2
   76 IF(SYM(F,G).EQ. 0)GO TO 70
      IF(SYM(F,G).NE.SYM(B,C))GO TO 75
      DIS(D,20)=DIS(D,20)+1
      GO TO 25
   70 C=C+1
      TOP=1
      GO TO 32
   75 G=G+1
      GO TO 76
   99 CONTINUE
      CALL FREQ (NSUM)
      CALL SORT
      CALL ALPHA
      GO TO 150
  200 STOP
      END
```

Appendix F - continued

Symptom/Disease Data Cards

Card	Description	Card Columns
1	3 digit symptom code	1-3
	4 digit disease code (19)	4-79
2	4 digit disease code (20)	1-80
3	4 digit disease code (10)	1-40

Disease/Symptom Data Cards

1	4 digit disease code	1-4
	3 digit symptom code (18)	5-58

Disease Name Cards

1	4 digit disease code	1-4
	Blank (Not Used)	5-17
	Disease Description	18-57

Patient Data Cards

1	Patient I.D. Number	1-9
	Number of Symptoms	10-15
2	3 digit symptom code (25)	1-75
3	3 digit symptom code (25)	1-75

Data Card Layout

Appendix G

BIBLIOGRAPHY

Books

Armitage, P., Statistical Methods in Medical Research, Oxford, Blackwell Scientific Publications, 1971.

Bailey, Norman T. J., The Mathematical Approach to Biology and Medicine, New York, John Wiley & Sons, 1967.

Fitzgerald, L. T. And C. M. Williams, Computer Diagnosis of Thyroid Disease, Gainsville, University of Florida Printing Office, 1964.

Hyman Harold T., Differential Diagnosis: An Integrated Handbook, Philadelphia, J. B. Lippincott Co., 1965.

International Classification of Diseases, Adopted, 8th ed., Washington, D. C., U. S. Government Printing Office, 1968.

Lusted, Lee B., "Computer Techniques in Medical Diagnosis," Computers in Biomedical Research, Vol. 1, New York, Academic Press, 1965.

----------, Introduction to Medical Decision Making, Springfield, Charles C. Thomas, 1968.

Magalini, Sergio I., Dictionary of Medical Syndromes, Philadelphia, J. B. Lippincott, 1971.

Matousek, William C., Differential Diagnosis, Chicago, Yearbook Medical Publishers, Inc., 1967.

Moore, Frederick J., Implementation of an Experimental Clinical Decision Support System, New York, International Business Machines Corporation, 1971.

Nordyke, Robert A., Casimar A. Kulikowski and C. Wilk Kulikowski, "A Comparison of Methods for Ohe Automated Diagnosis of Thyroid Dysfunction, " Computers in Biomedical Research, Vol. 4, New York, Academic Press, 1971.

Procedings of the Fifth I.B.M. Medical Symposium, New York, Endicott, 1963.

Taber, Clarence W., Taber's Cyclopedic Medical Dictionary, Philadelphia, F.A.Davis Company, 1958.

The Merck Manual of Diagnosis and Therapy, Rahway, Merck, Sharp and Dohme Research Laboratories, 1966.

Tofler, Alvin, Future Shock, New York, Bantom Books, 1974.

Whinery, D. Gaye, E.C.G. Data Acquisition and Analysis Package, Houston, National Areonautics and Space Administration, 1971.

Yater, Wallace M. And William P. Oliver, Symptom Diagnosis, New York, Appleton-Century-Crofts, Inc., 1961.

Articles

"Analyzing Patterns of Illness," I.B.M. Computing Report, (Spring, 1973), 8-12.

Cady, Lee D., Menard M. Gertter, Lida G. Gottsch, Max A. Woodbury, "The Factor Structure of Variables Concerned with Coronary Artery Disease," Behavioral Sciences, VI (June, 1961), 37-41.

Fitzgerald, Lawrence T., John E. Overall and Clyde M. Williams, "A Computer Program for Diagnosis of Thyroid Disease," The American Journal of Roentgenology, Radium Therapy and Nuclear Medicine, XCVII (August, 1966), 901-905.

Gorry, G. Anthony and G. Octo Barnett, "Experience With a Model of Sequential Diagnosis," Computers and Biomedical Research, Vol. I, edited by Ralph W. Stacy and Bruce D. Waxman, New York, Academic Press, 1968.

Gorry, G. Anthony, Jerome P. Kassirrer, Alvin Essig and William B. Schwartz, "Decision Analysis as the Basis for Computer-Aided Management of Acute Renal Failuee Computer-Aided Management of Acute Renal Failure," The American Journal of Medicine, LV (October, 1973), 473-484.

Gustafson, David H., John J. Kestly, John E. Greist and Norman M. Jensen, "Initial Evaluation of a Subjective Bayesian Diagnostic System," Health Services Research, VI (Fall, 1971), 204-213.

Hurtado, Arnold V. And Merwyn R. Greenlick, "A Disease Classification System for Analysis of Medical Care Utilization, with a Note on Symptom Classification," Health Services Research, VI (Fall, 1971), 235-250.

Jelliffe, Roger W., "Quantitative Aspects of Clinical Judgement," The American Journal of Medicine, LV (October, 1973), 431-433.

Ledley, Robert S. And Lee B. Lusted, "Reasoning Foundations of Medical Diagnosis," Science, CXXI (July, 1959), 9-21.

Overall, John E. And Clyde M. Williams, "Models for Medical Diagnosis," Behavioral Sciences, VI(April, 1961), 134-141

----------, "Conditional Probability Program for Diagnosis of Thyroid Function," The Journal of The American Medical Association, CLXXXI(February, 1963), 307-313.

Parker, R. D. And T. L. Lincoln, "Medical Diagnosis Using Bayes Theorem," Health Services Research, II (Spring, 1967), 34-45.

"Prescription By Computer," TIME, CIII (January 28, 1974), 48-49.

Schwartz, William B., G. Anthony Gorry, Jerome P. Kassirer and Alvin Essig, "Decision Analysis and Clinical Judgement," The American Journal of Medicine, LV (October, 1973), 459-472.

Toronto, A. F., L. G. Veasy and H. R. Warner, "Evaluation of a Computer Program for Diagnosis of Congenital Heart Disease," Progressive Cardiovascular Diseases, V (January, 1963), 362-377.

Vanderplas, James M., "A Method for Determining Probabilities for Correct Use of Bayes's Theorem in Medical Diagnosis," Computers in Biomedical Research, Vol. 3, New York, Academic Press, 1967.

Ward, H., and Marvin E. Hook, "Use of Regression Analysis and Electronic Computers in the Practice of Coronary Artery Disease," Beahvioral Sciences, VII (January, 1962), 120-126.

Warner, Homer R., Charles M. Olmstead and Barry D. Rutherford, "HELP - A Program for Medical Decision-Making," Computers in Biomedical Research, Vol. 5, New York, Academic Press, 1972.

Winkler, C., P. Reichertz and G. Kloss, "Computer Diagnosis of Thyroid Diseases, Comparison of Incidence Data and Considerations on the Problem of Data Collection," The American Journal of the Medical Sciences, CCLIII (January, 1967), 27-33.

Reports

Reitman, Judith S., Computer Simulation of an Information Processing Model of Short Term Memory, Mental Health Research Institute Communication Number 226 and Information Processing Working Paper Number 8, The University of Michigan, 1968.

Townsend, John C., "Medical Usage of Computer Science," Biomedical Research and Computer Application in Manned Space Flight: A Report, Houston, National Aeronautics and Space Administration, 1972.