LOGICAL ANALYSIS IN MEDICAL DIAGNOSIS

LEE B. LUSTED
UNIVERSITY OF OREGON MEDICAL SCHOOL
and
OREGON REGIONAL PRIMATE RESEARCH CENTER

1. Introduction

As medical data has increased in volume the physician has had difficulty in using the data effectively. One explanation seems to be that the large volume of data makes it a formidable task for the physician to remember the interrelationships among signs, symptoms, and laboratory tests for a wide variety of diseases. Thus, it seems that studies need to be undertaken which will attempt to analyze and synthesize medical data in order to permit a reorganization of the data in a more meaningful, more manageable, and more useful form. The logical analysis procedures which I wish to discuss are presented with this objective in mind. Some of the procedures were developed quite a few years ago and have recently been reinvestigated with renewed interest because of the availability of computers. In this paper I have attempted to summarize and to place in new perspective some studies which I hope will stimulate further research on techniques that will make medical data more manageable for the physician in his efforts to maintain health and treat disease in his patients.

Research projects on logical analysis in medical diagnosis seem to be grouped in two main areas neither of which has a very distinct boundary. One area of research might be called studies on the diagnostic process and a second area has been referred to frequently as computer aided medical diagnosis. These two areas of research are complementary if research on the diagnostic process is defined as an effort to understand and to predict what the human diagnostician can accomplish, and if computer diagnosis studies are defined as attempts to construct mathematical or statistical models of medical diagnosis which will perform the most accurate job possible. The discovery of functional relationships between signs, symptoms, and diagnoses in a computer diagnosis study may give clues to better methods in human diagnosis. Similarly, greater understanding of human judgment processes involved in diagnosis may enable the investigator to produce these processes more exactly on a computer. An interesting group of papers which discusses the diagnostic process will be found in the pro-

Publication No. 99, Oregon Regional Primate Research Center, supported in part by a research grant (FR 00163) of the National Institutes of Health.

ceedings of a Conference on the Diagnostic Process [37] and in a series of articles by Engle [38].

2. Studies on the diagnostic process

The clinician in his role as a decision maker has become a subject for study by physicians who are interested in improving medical school teaching methods and by psychologists concerned with problem solving and judgment [39], [40]. The interest in the decision maker in his role as a medical teacher was expressed recently by Adams [41] at a Conference on the Diagnostic Process when he noted that "the questions the teachers asked the students may represent the kinds of questions that he asks himself when he solves problems alone. This, in fact, may be a possible way into the problem solving process of the teacher. We will have to come to grips with the question of what goes on in the teacher's mind. One of the big questions of the symposium seems to be 'What can we learn about the processes that actually take place in the mind of the person solving problems of medical diagnosis through the study of computer processes or by other means?' "

The problem of how the clinician processes information is clearly within the domain of psychological studies that are concerned with problem solving and judgment. Kleinmuntz [42] refers to a number of recent reports in which researchers have proposed that the clinician rather than the clinical data be considered as an instrument to be analyzed. Most of these researchers were concerned with the clinical interpretation of various types of psychological tests and the material is of interest to us because some of the same techniques are being used in studying medical decision making processes.

A particular approach which uses information processing computer languages to simulate the clinical decision making processes has been selected for further discussion. This information processing approach is one way of looking at psychological activity and the use of information processing computer languages encourages the investigator to specify precisely and explicitly the systems of cognitive structures he induces from the behavior he observes [43]. Furthermore, this type of study deals with processes and functions and it is concerned with the fine structure of behavior.

The concept of simulating human behavior with a computer program may be unfamiliar to some readers. Newell and Simon [44] have provided considerable evidence that an electronic computer can carry out complex patterns of processes that parallel closely the processes observable in humans who are thinking. These investigators collect data from subjects who are asked to "think aloud" during problem solving sessions. The subjects' comments are tape recorded and they then try to write computer programs that will simulate the behaviors observed in their data. They view a program capable of simulating behavior as a theory of the system of psychological processes and structures underlying the behavior. Such theories are held to have a status comparable to those framed in words

or in mathematical symbols and to be subject to the same theory of adequacy [43]. The significance of adopting this approach to the study of clinical decision making is that it forces us to specify with complete rigor the processes which are involved.

Kleinmuntz has combined heuristic and information processing language approaches with the "think aloud" technique previously described to investigate MMPI (Minnesota Multiphasic Personality Inventory) profile interpretation [42] and clinical decision making processes of neurologists [45]. At this point it will be helpful to note the distinction between three types of problem solving computer programs, namely, algorithmic simulating and heuristic [42]. An *algorithmic* solution to the medical diagnosis problem would consist of putting all information on symptoms, signs, and so forth in the computer and instructing the computer to combine these variables in every possible manner and to stop when it has reached a certain decision or diagnosis. This is a task of considerable magnitude even for the fastest computer and "computer diagnosis" for certain limited areas such as congenital heart disease works as well as it does because logical grouping of signs and symptoms has eliminated impossible combinations which the computer otherwise would be forced to process. This point is considered in greater detail in the discussion of computer aided diagnosis.

Direct simulation of the clinical decision making process would consist of a computer program that followed closely all of the consistencies, inconsistencies, and errors that the clinician would verbalize during the time he was making a diagnosis. The heuristic approach attempts to construct a program with the best possible combination of decision rules in order to maximize the success rate of the diagnostic process. This combination may include interpretive rules of thumb and procedural tricks of the trade which are used by an expert diagnostician. The simulation approach although a larger task than the heuristic programming seems more likely to lead to a hypothesis of pattern analytic behavior of the clinical decision maker.

To illustrate how the "think aloud" and information processing computer programs are used in the study of medical diagnosis, I have selected an example from the work of Kleinmuntz [45] on clinical decision making in neurology. The technique used is a diagnostic game played somewhat in the form of "Twenty Questions" with one player, called the experimenter, thinking of a disease and a second player, called the subject, trying to diagnose the disease the first experimenter has in mind. The experimenter must be an expert neurologist if the game is to be meaningful because in the various roles he may assume he must be able to recognize the appropriateness and validity of the questions which the subject asks him in trying to make a diagnosis. The subject's questions and the experimenter's answers are tape recorded for further analysis and for use in developing a computer program. An end product of the diagnostic game may be pictured as a tree structure which represents the search strategy which the subject used in arriving at the diagnosis.

For a particular diagnostic game a decision tree structure might result as

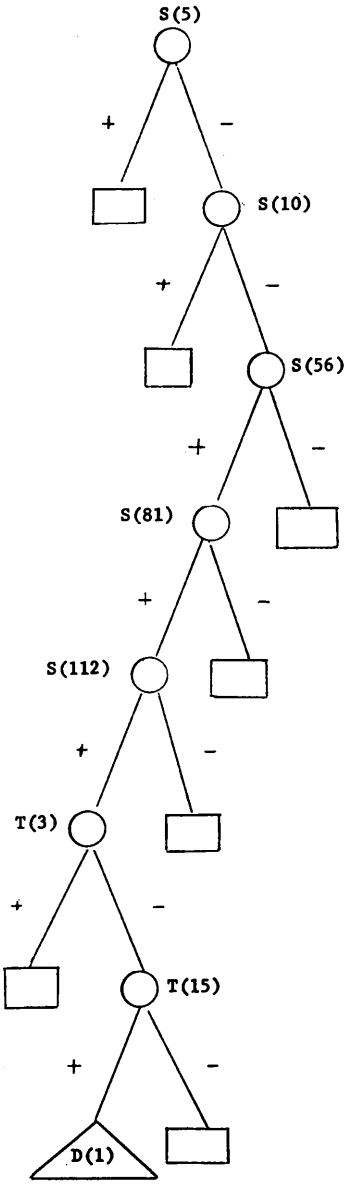

FIGURE 1

The process of diagnostic decision making may be visualized as a tree structure which represents the search strategy of a diagnostician. The circles represent nodes at which the physician may give a differential diagnosis or he may elect to proceed by asking for more information. The plus or minus branches represent the presence or absence of the preceding sympton, for example, $S(56)$. (Personal communication from B. Kleinmuntz.)

shown in figure 1. The experimenter asserts he has a disease in mind (shown as $D(1)$) and the subject asks whether a certain symptom $S(5)$ is present. He receives a negative reply. Then the subject asks whether symptom $S(10)$ is present and again receives a negative reply. Based on information he has so far he asks whether symptom $S(56)$ is present and affirmative replies lead him to ask about $S(81)$ and $S(112)$. He then asks for evidence in terms of laboratory tests $T(3)$ and $T(15)$. The experimenter replies that $T(3)$ results are negative and $T(15)$ yields positive results. On the basis of the cumulative information the subject suggests diagnosis $D(1)$, the correct diagnosis, and the game is concluded. If $D(1)$ had been incorrect the game would have continued. The circles in the tree represent decision nodes and the number of nodes which designate decision levels plus the number of tree branches may be considered as a rather crude model of the fine structure of the decision process.

The number of branches in a particular decision tree seems to be a function of the experience and ingenuity of the diagnostician. For the same diagnostic problem one diagnostician may use a decision tree with four branches whereas another diagnostician may use forty. By varying the amount of initial information supplied to the subject and, by controlling the kinds and amount of information available during the game it should be possible to make a systematic study of diagnostic decision trees and thereby to determine optimum decision tree structure and to evaluate the effects of various signs, symptoms, and so forth on diagnostic accuracy. The use of one type of diagnostic decision tree has been described for a programmed computer system to be used in teaching medical diagnosis. The decision tree developed by Swets and Feurzeig [46] illustrates the manner in which a student can be led through a differential diagnosis to a diagnosis which in the example was pneumococcal pneumonia.

If one attempts to improve upon the expert's diagnostic performance it will probably be necessary to combine various techniques developed from studies of computer diagnosis and diagnostic judgment. On this point Kleinmuntz [42] says with regard to MMPI profile analysis, "In an effort to improve upon the experts Q-sort performance and to sharpen the existing decision rules, the method used had a strong algorithmic flavor. The method consisted mainly of statistical searching and shuttling back and forth between intuitive hunches about combinations of various scales and their effect on the hit per cents. Also a pattern analytic approach to the decision rules themselves was used by writing the programmed instructions in such a way so that MMPI rules, i.e., rules that rarely misclassified a profile, received greater weight than weak rules. Finally, the completed set of MMPI rules included the original test interpreter's information, a number of intraprofile slope characteristics that the expert failed to observe and an optimal ordering of the various decision rules to yield the highest possible hit per cent. The revised rules did considerably better than the Z-sorter and achieved valid positive and valid negative rates of 91 and 84 per cents, respectively."

It will be admitted that clinical decision making may be a more complex

process than personality test interpretation but the studies already accomplished can serve as a basis for further work on diagnosis decision making. Many intriguing questions remain to be investigated. For instance, are there diagnostic problem solving behaviors which are common to the various clinical areas and which underlie decision making in most or all of the clinical specialties?

The investigator who has started research on the use of logical analysis in medical diagnosis must sooner or later consider how his data are influenced by observer variation and the accuracy of diagnostic procedures. This statement applies whether the original information on incidence of signs and symptoms comes from "personal" probability estimates, textbooks, or medical records. Several hundred articles [47] have been published on the subject of observer variation. Although many investigators already are familiar with some of the work in this field a short discussion of the topic will be presented to act as a bridge between the subjects of studies on the diagnostic process and computer aided medical diagnosis.

3. Observer variation

Suppose that a physician has a group of 100 patients from whom he has collected an extensive amount of information on signs, symptoms, laboratory data, and so forth, and after a study of this information he reaches a judgment on each patient concerning whether the patient is ill or not-ill. We allow the physician only one of two possible choices for each patient. Now suppose we present this same group of patients to an additional nine physicians of equal competence. There will be some difference of opinion among the ten physicians concerning the number of ill and not-ill patients in the group. We suppose also that a proven diagnosis is available on each patient.

The judgments of the ten physicians can be divided into the four categories true positive (patient truly ill), true negative (patient truly not-ill), false positive (not-ill patient judged to be ill) and false negative (ill patient judged to be not-ill). The patients who are found to be in the false negative (type I error) and false positive (type II error) categories represent variation in physician judgment and this group therefore requires further study to determine the possible sources of variation [48].

It is important for the physician to understand that the demarcation line between the ill and not-ill population will vary depending on the criteria (signs and symptoms) he uses to make his decision. No matter what signs, symptoms or tests he uses to define an ill patient he will find these same signs and symptoms occurring in some not-ill patients. Part of the uncertainty in diagnosis results from this overlap of the populations of truly ill patients and truly not-ill patients [49]. Collen, Rubin, Neyman, Dantzig, Baer, and Siegelaub [23] have shown how this uncertainty is treated in the screening procedure for asthma by making use of the likelihood ratio.

When signs and symptoms are chosen to define an ill patient the physician

has established two things. First, he has set a demarcation line between ill and not-ill populations. Second, he has established a ratio of false negative to false positive cases. This is illustrated in figures 2 and 3 in which a population has

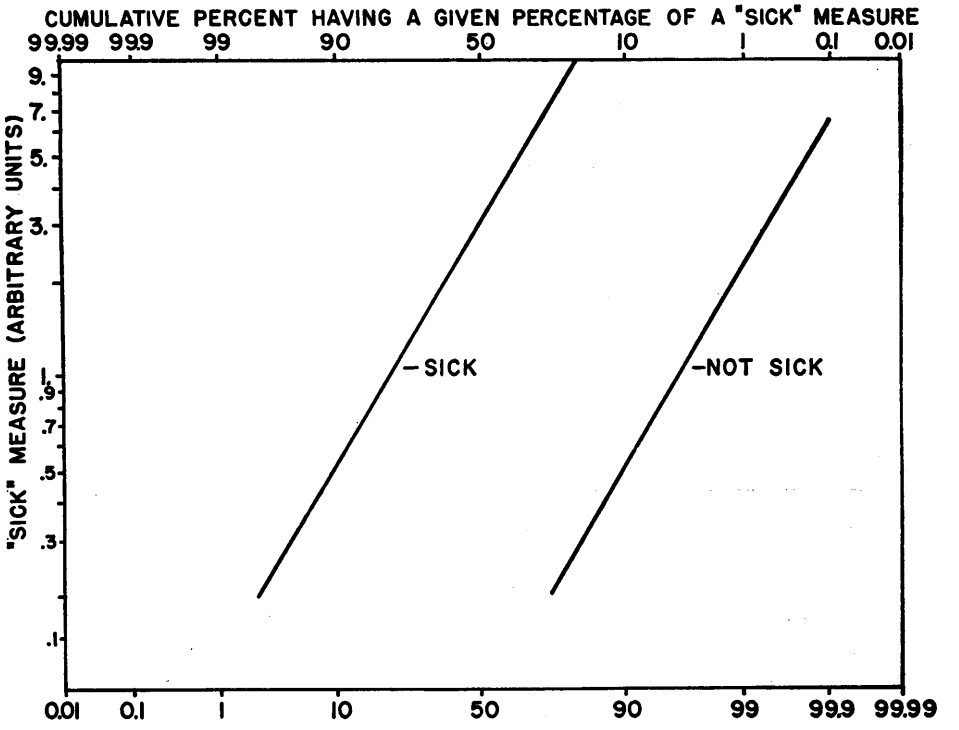

FIGURE 2

A population has been divided into two groups, "ill" and "not-ill" on the basis of arbitrary criteria for "ill." The plot is made on logarithmic probability paper [17].

been divided into two subpopulations (ill and not-ill) on the basis of a set of criteria chosen (in this case arbitrarily) to define "ill."

Frieden, Shapiro, and Feinstein [50] have given an interesting example to illustrate these points. In their study of radiologic evaluation of heart size in rheumatic heart disease they initially considered significant indentation on the anterior wall of the esophagus to be adequate evidence of left atrial enlargement.

The authors then noted: "It is apparent from the large number of false positive readings in normal hearts that this criterion is overly sensitive.

"It is our impression that the posterior wall of the esophagus as well as the anterior wall should be significantly indented before the left atrial enlargement is diagnosed. By using this criterion in the re-evaluation of our films, the incidence of false positive readings in normal patients could be reduced from

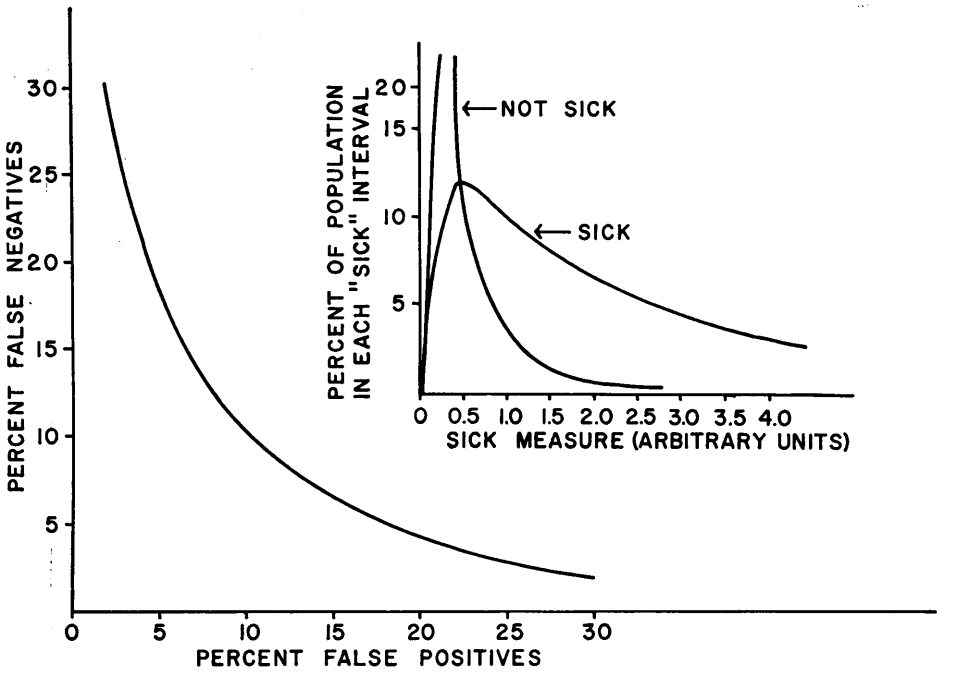

FIGURE 3

Upper right graph: the population distribution from figure 2 is replotted to show that the curves are lognormal.
Lower left graph: the reciprocal false positive-false negative relationship for the population is very near a hyperbola and is a good fit for the experimental data curve in figure 4 [17].

approximately 10% to almost none, and, in addition, the disagreements in left atrial interpretation of patients with rheumatic heart disease could be dramatically reduced.

"It is possible that the use of these rigorous criteria for determining left atrial enlargement may diminish the accuracy of interpretation in patients with rheumatic heart disease resulting in the underreading of some cases of left atrial enlargement. However, we feel that there is less harm to be done in underreading than in overreading, since the latter interpretation may result in needless restriction of activity with impairment of the emotional and psychological development of the patient."

Here is the crux of the problem. False negatives and false positives are errors which should be recognized as a "cost" to a system and it is important to find a combination of false positives and false negatives which represent a minimum total cost. The "cost" may actually represent dollars in the case of false positives where follow up studies are required. The "cost" of the false negative case to

the individual and to society is perhaps harder to evaluate and requires further study.

Numerous experimental studies [51] on the accuracy of cytological diagnosis of Papanicolaou smears and on the interpretation of chest roentgenograms have shown that a reciprocal relationship exists between the number of false negatives and the number of false positives. This inverse relationship which is demonstrated in figure 4 for the reading of chest X-rays [52] shows that as the

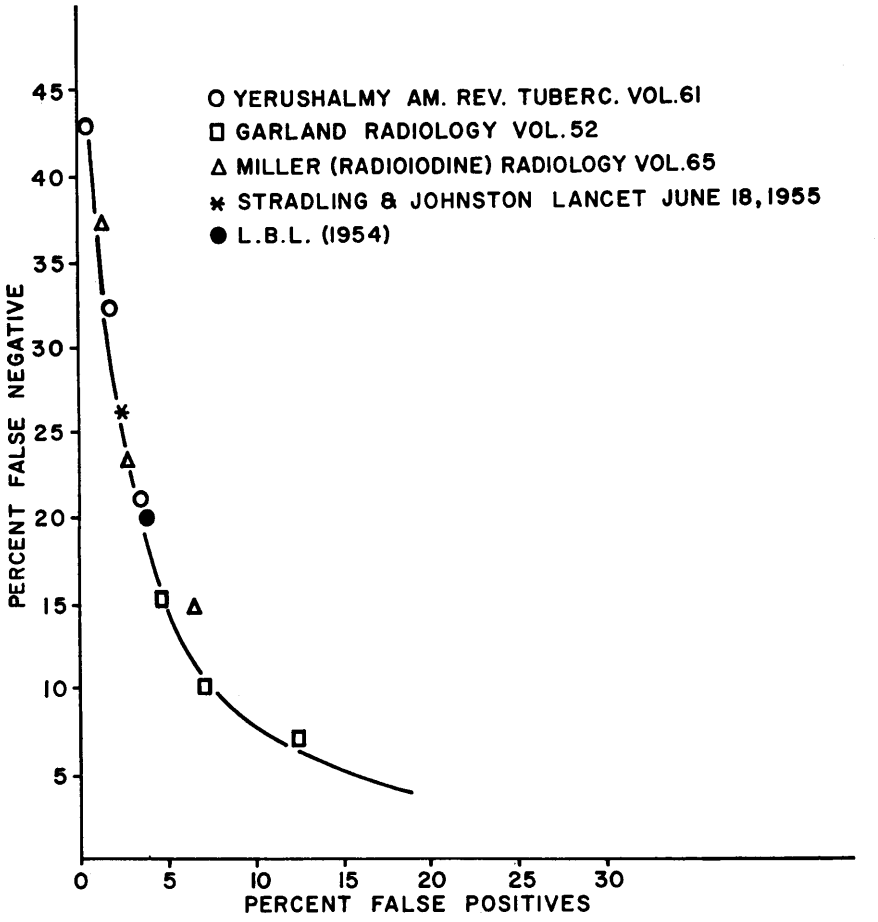

FIGURE 4

Graph showing the reciprocal relationship between the percentage of false positive and false negative readings obtained from studies on the accuracy of chest film interpretation. Miller's study on the interpretation of the radiation effect on thyroid gland and other studies on the accuracy of diagnosis of Papanicolaou smears have shown a similar reciprocal relationship.

number of false negatives decreases the number of false positives increases. As the criteria (signs and symptoms) are made more stringent for diagnosing the ill patients, the number of false negatives will decrease but at the expense of increasing the false positives.

From considerations of statistical decision theory it can be shown that a family of error probability curves may be developed which are similar in shape to the curve shown in figure 4. Within this family each curve represents the results obtained with a particular set of diagnostic criteria. One curve will represent a lower, left boundary and along this curve there is no simple and categorical basis for choosing the "better" of two points of operation [18] (equivalent to choosing a "better" set of diagnostic criteria) since each of the two points is better with respect to one of its error probabilities (that is, better with respect to either the false negatives or the false positives). Thus, the physician is given simultaneously both the problem of setting his diagnostic criteria to obtain the diagnostic accuracy he desires and the opportunity to gain further insight into the diagnostic process. It should be possible in future studies of computer aided diagnoses to evaluate the relative contribution of various types of errors which now are lumped together under the heading of observer variation. In a study of congenital heart disease diagnosis, Toronto, Veasy, and Warner [53] were able to demonstrate that participating clinicians made two types of errors: namely, errors in symptom recognition and errors in logic (coming to a wrong conclusion from valid available data), and that the two types of errors could be investigated separately. The physician working on computer aided diagnosis needs to have some appreciation of how observer bias effects his original probability estimates and his subsequent observations on signs and symptoms. As Grosz and Grossman [54] have pointed out there is a much greater degree of observer variation in clinical judgment when data are subjective and ambiguous than when the data are objective and factual.

It might be argued that physicians are usually not required to make judgments in terms of categorical alternatives such as ill and not-ill and that the preceding discussion of false positives and false negatives is not very pertinent. To a certain extent this may be true but it does seem that carefully controlled experiments in terms of a limited number of alternatives (diagnoses) may lead to the use of new mathematical procedures which would help to relate qualitative and quantitative criteria to medical diagnosis. One of the interesting studies to appear recently was concerned with the consistency of physician judgment in ranking patients with myocardial infarction on the basis of prognosis. In this project King and Manegold [55] presented twenty cardiologists with twenty case histories and asked each physician to make a judgment on the continuous dimension, that is, favorability of prognosis. Thus, they avoided the criticism of testing categorical alternatives.

Two of the conclusions from this study are worth particular attention. The first is that apparently the time taken to make the prognostic judgments did not influence their consistency. The implications of this finding are not clear

and similar studies should be made in other areas of medicine. It is interesting to note, however, that in the studies where categorical alternatives were presented (ill and not-ill in the cytological screening procedure and chest X-ray screening procedure) the amount of time spent on the decision was related to the judgments which were made [51]. These results suggest that in the screening procedures the observers were using a search pattern recognition process and that the cardiologists in making prognostic judgments about myocardial infarction cases were using "some other" process. A second conclusion of King and Manegold is that cardiologists are able to make extremely consistent prognostic judgments in a disease of their specialty. Although this finding is reassuring it raises questions about further investigations which might help to reveal the logical processes used by the clinicians in reaching a decision. These processes are usually not apparent to the clinician himself and as a result the study of the physician and how he uses diagnostic decision trees has been undertaken as I pointed out earlier in this article.

To conclude the discussion on observer variation, two points can be made which emphasize the contribution that observer variation studies can make to analysis and synthesis of medical data. First, the studies can help to indicate which signs, symptoms, and laboratory tests are really most useful in terms of repeatability and validity [49] for making a diagnosis. The real benefit, however, can come from studies which show the effect of different diagnostic criteria on the number of false negatives and false positive diagnoses. The work cited earlier of Frieden, Shapiro, and Feinstein [50] on the radiologic evaluation of heart size in rheumatic heart disease is an example of such a study. Signs and symptoms that seem to have great diagnostic importance may not be as important as we think and we need not continue to carry the symptoms which are low in diagnostic value. The second thing that observer variation studies can do is to help us to formulate odds concerning how sure we are about a sign or symptom with respect to a diagnosis. How these odds may be used in Bayesian inference will be discussed in section 5.

4. Computer aided diagnosis

Medical diagnosis as a logical process by which the physician follows a chain of events from cause to effect and, thereby, reaches a decision has stimulated the interest of investigators for a long time. For many areas of medicine and perhaps well over half, the chain of events is short and causal relationships are obvious enough so that mathematical models of diagnosis and computer techniques will never be needed. However, computer diagnosis studies already have demonstrated the statistical significance of various clinical findings for several areas of diagnosis. For these areas diagnosis has become less cumbersome because the most important diagnostic signs have been separated from signs of less importance. At present the entire diagnostic process cannot be described in terms of a computer program but some parts of the process can be expressed

in terms of mathematical models. Some of these models will be discussed in the material which follows and the first part of the references will give the reader an impression about the mathematical models which have been applied to a group of medical specialties.

5. Bayes' theorem and conditional probability

The eminent English physician, Sir George W. Pickering [56] wrote recently, "Diagnosis is a matter of probability, as those of us who follow the fate of our patients to post-mortem room know only too well. Prognosis is a matter of probability, and in judging treatment we have to base our judgment on knowledge of probabilities." Only in the past five years have studies in a number of medical specialties indicated the usefulness of conditional probability models to diagnosis.

Conditional probabilities are the probabilistic expression of learning from experience and a particular form of conditional probability known as Bayes' theorem is a useful mathematical device that combines evidence from data with prior information. Bayes' theorem [57], [58] has been used extensively in experimental studies of medical diagnosis with favorable results but its application to diagnosis has been disputed on the grounds that conditions under which the theorem is meaningful are rarely met and that sources of the probabilities often are vague. On the other hand Bayes' theorem is universally accepted as mathematically correct. Some of the problems which arise in the application of Bayes' theorem will be discussed later. An excellent general discussion of the use of Bayes' theorem in medical diagnosis has been given by Cornfield [59], [60]. Edwards, Lindman, and Savage [61] have published an interesting report on the use of Bayesian statistical inference in experimental psychology. These articles and the work by Mosteller and Wallace [62], [63] on a problem of disputed authorship of twelve Federalist papers using Bayesian analysis will be helpful to the investigator who is interested in the use of Bayes' theorem in diagnosis.

Bayes' theorem applied to diagnosis may be written

$$P\{D|S\} = \frac{P\{D\}P\{S|D\}}{\sum_j P\{D_j\}P\{S|D_j\}}. \tag{5.1}$$

This expresses the conditional probability of disease D, given symptom complex S as a function of the unconditional probability of disease D and the conditional probability of symptom complex S, given disease D.

An expandable form of Bayes' theorem which has been used by Warner, Toronto, and Veasy [2] for the diagnosis of congenital heart disease is

$$A_{jS} = \prod_{i=1}^{M} (P_{ij})^{a_i}(1 - P_{ij})^{b_i}(EX_{ij})^{c_i} \tag{5.2}$$

and

$$P\{D|S\} = \frac{A_{jS}}{\sum_{j=1}^{N} A_{jS}}, \tag{5.3}$$

where P_{ij} is a matrix of dimensions M by N where i identifies a symptom and j a disease, P_{ij} is the probability of a patient with the jth disease having the ith symptom, P_{ij} is the *a priori* incidence of the jth disease in the population of patients having one of the diseases to be considered. The term EX is used to handle mutually exclusive symptoms and S is any particular array of M digits each of which is either 1, 0, or -1 and labeled S_i.

Bruce [8], Lodwick, Haun, Smith, Keller, and Robertson [15], and Overall and Williams [11] have used similar computational forms of Bayes' theorem in the diagnosis of acquired heart disease, bone tumors, and thyroid disease. All of these models assume that the symptoms are independent of one another within a given disease and that the diseases are mutually exclusive. Independence of symptoms can be tested by chi square analysis if sufficient data are available on the coincidence of symptoms in each disease.

The condition of symptom independence has not always been met and yet the Bayesian analysis has worked quite well. This seems to bear out Cornfield's statement [59], "there is no assumption . . . that the components of the multidimensional event are independent. We stress this obvious point only because there seems to be a widespread misconception in the medical literature that the application of Bayes' theorem to the diagnosis problem requires the assumption of independence of the various symptoms."

The assumption that diseases are mutually exclusive has been handled by Warner by noting that if a patient is found to have more than one disease, the new combination of diseases is considered as a new single disease entity. This point needs further study.

However, the most controversial issue concerns the assignment of the unconditional or prior probabilities and on this point a distinction between frequency and nonfrequency interpretation of probability must be carefully considered. Probability may be considered a "degree of belief" or as relative frequency and the two interpretations are both used in the application of Bayes' theorem to medical diagnosis. As Mosteller [63] has pointed out the degree of belief interpretation is more widely applicable, but specifying the probabilities is often difficult in the absence of conditions of symmetry or long run relative frequencies. When the two interpretations are simultaneously applicable, the same numerical values would normally be assigned in each interpretation.

The prior probability $P\{D\}$ may be constructed from a variety of sources such as medical textbook, patient medical records, or a subjective probability [61] from the judgment of an experienced clinician. Since the sources and prior distributions of such data are not well documented the correctness of $P\{D\}$

may be questioned. Of course the best available data or estimates should be used to determine conditional probabilities and prior probabilities and when this has been done the investigator may derive some satisfaction from a methodological conclusion of Mosteller and Wallace [62], "prior distributions are not of major importance. While choice of underlying constants (choice of prior distributions) matters, it doesn't matter very much, once one is in the neighborhood of a distribution suggested by a fair body of data. We conclude from this that the emphasis on the difficulty, even impossibility of choosing prior distributions as a criticism of the use of Bayes' theorem is not well placed."

Various techniques have been used by investigators to obtain data for $P\{D\}$ and $P\{S|D\}$. Nugent, Warner, Dunn, and Tyler [10] working in Salt Lake City discusses the population from which Cushing's syndrome patients were derived and he has constructed a frequency distribution curve of individuals with Cushing's syndrome as a function of the measured variables (symptoms). In order to enlarge the amount of statistical information he obtained data from Liddle working in Nashville. Comparison of the two sets of data revealed some interesting differences such as the effect of altitude on the volume of packed red blood cells. With further study essential agreement was reached on the $P\{D\}$ and $P\{S|D\}$ factors to be used for Cushing's syndrome.

This example is cited to emphasize that broadening the data base can give insight into the relative contribution of symptoms and to help determine whether there is substantial agreement among investigators about $P\{D\}$ and $P\{S|D\}$. Only when data is collected or an estimate is made will the investigator find whether his model for these probabilities is "public" [61]. That is to say, whether there is substantial agreement among investigators that $P\{D\}$ and $P\{S|D\}$ are derived from a model that scientists will agree is an acceptable description about the situation in which the data were obtained. With the data collection organized it is possible to expand the data by using a computer program which adds each new case as it is diagnosed. Such a computer program is now used by Williams [11] for the diagnosis of thyroid disease.

The experience with Bayes' theorem in medical diagnosis during the past five years can be summarized by noting that for certain diseases new insight has been gained into the relative contribution a symptom or symptom complex will make to a diagnosis. For instance in the clinical and radiologic diagnosis of benign gastric ulcer versus malignant gastric ulcer it was shown by Wilson, Templeton, Turner, and Lodwick [14] that 17 statistically valid signs and symptoms would make the diagnosis with a high degree of certainty. These 17 variables were selected from a group of 70 signs and symptoms listed in medical textbooks as significant variables in the diagnosis of gastric ulcer.

Further studies are needed to determine how to handle low frequency symptoms and diseases. We need to investigate the effect on $P\{D|S\}$ of various prior distributions for $P\{D\}$ and $P\{S|D\}$ and we need simple Bayesian methods that can be applied without appeal to high speed computation. The latter need previously was noted by Mosteller and Wallace [63]. Perhaps the most impor-

tant need if we are to use Bayesian methods for realistic diagnostic problems in a large medical clinic or hospital is to develop better methods for obtaining $P\{D\}$ and $P\{S|D\}$. One possible solution is to use likelihood ratios. Cornfield [60] has shown recently that diagnostic procedures based upon either likelihood ratios or posterior probabilities are closely related rather than as sharply opposed as has been suggested in the past. Because of the importance of likelihood ratios a discussion is given in the following material.

6. Likelihood ratio

A consideration of Bayes' theorem will be used to develop the likelihood principle and the likelihood ratio. Suppose that two possible observations on a symptom s_1 and s_1'—not necessarily from the same experiment—can have the same bearing on your opinion about a diagnosis d_j. Another way to state this is that $P\{s_1|d_j\}$ can equal $P\{s_1'|d_j\}$ for each s_1. The question may be asked, when are s_1 and s_1' equivalent or of the same importance with respect to your opinion about the diagnosis d_j. Remember the example of Nugent and Liddle pooling the sign and symptom data of Cushing's syndrome.

The conclusion concerning the question above is given in terms of the likelihood principle which says that s_1 and s_1' are of the same import if the following equation is valid.

$$P\{s_1|d_j\} = KP\{s_1'|d_j\}. \tag{6.1}$$

For the purpose of making a diagnosis this equation says that the sequence of numbers $P\{s_1|d_j\}$ is, according to the likelihood principle, equivalent to any other sequence obtained from it by multiplication by a positive constant K. We will see shortly how Collen, Rubin, Neyman, Dantzig, Baer, and Siegelaub [23] has used this principle as an aid in diagnosis.

Equation (6.1) may be rewritten in the form

$$K = \frac{P\{s_1|d_j\}}{P\{s_1'|d_j\}}. \tag{6.2}$$

This equation, known as a likelihood ratio, is an application of the likelihood principle and it has been used as an aid to diagnosis in a multiphasic screening program by Collen, Rubin, Neyman, Dantzig, Baer, and Siegelaub [23] as follows.

A given population is to be screened for the presence of certain stated diseases. The only data needed for the likelihood ratio method are the determinations for each disease $(d_1, d_2, \cdots, d_j)$ to be screened for the proportion of individuals with a disease d_j with certain sets of symptoms s and the proportion of individuals not having the disease d_j but with the same set of symptoms s.

The statistical procedure used to implement the likelihood ratio method of screening was suggested by Neyman [64] in 1950 and is based upon selecting sets of symptoms whose relative frequency in diseased and nondiseased patients will define a region having a maximum likelihood of detecting the disease, if

present, and at the same time, a predetermined likelihood of indicating that a disease is present when actually it is absent. The ratio of the frequency of a set of selected symptoms S in the diseased patients D to the frequency of this same set of symptoms S in the nondiseased patients N when written in the form of the likelihood ratio defined in the preceding equation, and using the notation of Collen is

$$\theta = \frac{P\{S|D\}}{P\{S|N\}}. \tag{6.3}$$

This method offers a simplification in the theory of testing because it does not require that the prevalence of a disease in the population be known and it is possible, therefore, to use this method at the outset of a diagnosis study when a large amount of data is not yet available. The method may also be used in the presence of changing incidence of disease, in various geographical areas where disease prevalence may differ and under circumstances where it is not necessary to prove symptom independence. From the mathematical point of view the question of mutually exclusive diseases does not need to be considered when using the likelihood ratio but from a medical point of view the question cannot be ignored. Likelihood ratio studies may be used to determine the sensitivity and specificity of a diagnostic test and to show observer variation, a subject which was discussed earlier in this paper.

7. Other statistical methods

Overall has made a comparison of alternative statistical models and clinical diagnosis in the diagnosis of thyroid function. The computer based statistical models used were a Bayesian conditional probability approach [11], factor analysis [27], and discriminant function analysis [26]. With the Bayes' model the investigators were able to classify patients in agreement with clinical diagnosis more than 97 per cent of the time and the factor analysis results agreed with clinical opinion in over 95 per cent of the cases. An additional group of studies best categorized as matching procedures which are related to multivariate analysis are listed in the references.

8. Summary

In this article I have attempted to draw together some examples of research activity which represent various approaches to a subject I have called logical analysis in medical diagnosis. Studies on computer aided diagnosis and on the diagnostic process complement each other and share a common interest in problems of observer bias. The statistical information gleaned from research on computer aided diagnosis can supply data to be used in the diagnostic process studies and these studies in turn may supply new methods for testing and utilizing the statistically based diagnostic algorithms.

I appreciate discussions with Professor B. de Finetti and Professor Lyle D. Calvin concerning some of the material presented in this article.

REFERENCES TO MATHEMATICAL AND STATISTICAL PROCEDURES USED IN COMPUTER AIDED DIAGNOSIS

A. Bayes' Theorem and Conditional Probability

Congenital Heart Disease

[1] H. R. Wagner, A. F. Toronto, and L. G. Veasy, "A mathematical approach to medical diagnosis. Application to congenital heart disease," *J. Amer. Med. Assoc.*, Vol. 177 (1961), pp. 177–183.

[2] ———, "Experience with Bayes' theorem for computer diagnosis of congenital heart disease," *Ann. New York Acad. Sci.*, Vol. 115 (1964), pp. 558–567.

[3] N. M. Amosov and Ye. A. Shkabara, "Experience in formulating a diagnosis by means of diagnostic machine," *Eksp. Khir. Anest.*, Vol. 4 (1961), pp. 15–22. (In Russian; English translation available JPRS 11815, January 3, 1962.)

[4] M. L. Bykhovskiy, A. A. Vishnevskiy, and S. Shr. Kharnas, "Problems in the development of a diagnostic process by means of mathematical machines," *Eksp. Khir. Anest.*, Vol. 4 (1961), pp. 3–15. (In Russian; English translation available JPRS 11814, January 4, 1962.)

[5] J. E. Gustafson, "The computer—for use in private practice," *Proceedings of the Fifth IBM Medical Symposium*, Endicott, IBM, 1963, pp. 99–112.

[6] J. E. Gustafson and T. D. Throckmorton, "The computer as a consultant," *Med. Times*, Vol. 93 (1965), pp. 115–120.

Acquired Heart Disease

[7] R. A. Bruce and A. R. Cox, "Medical diagnosis by computer," *Northwest Med.*, Vol. 60 (1961), pp. 1075–1082.

[8] R. A. Bruce, "Computer diagnosis of heart disease," *Proceedings of the Fifth IBM Medical Symposium*, Endicott, IBM, 1963, pp. 77–98.

Electrocardiography

[9] E. Kumura, Y. Mibukura, and S. Miura, "Statistical diagnosis of electrocardiogram by theorem of Bayes," *Jap. Heart J.*, Vol. 4 (1963), pp. 469–488.

Cushing's Syndrome

[10] C. A. Nugent, H. R. Warner, J. T. Dunn, and F. H. Tyler, "Probability theory in diagnosis of Cushing's Syndrome," *J. Clin. Endocr.*, Vol. 24 (1964), pp. 621–627.

Thyroid Function

[11] J. E. Overall and C. M. Williams, "Conditional probability program for diagnosis of thyroid function," *J. Amer. Med. Assoc.*, Vol. 183 (1963), pp. 307–313.

Note. A booklet entitled *Computer Diagnosis in Thyroid Disease*, has been prepared by Dr. Williams. This excellent booklet contains an interesting and detailed description of items used in the study, that is, case histories, data collection sheets, data card layout, data matrices, computer programs, program flow charts. The physician should be able to start his own project with the use of this booklet. A copy may be obtained by writing to Dr. Clyde M. Williams, Department of Radiology, University of Florida College of Medicine, Gainesville, Florida.

[12] K. SILINK, "A new mathematical approach to endocrinological diagnosis," *Progress in Biocybernetics* (edited by N. Weiner and J. P. Schade), New York, Elsevier, 1964, Vol. 1, pp. 20–35.

Gastrointestinal Tract

[13] J. A. RINALDO, P. SCHEINAK, and C. E. RUPE, "Symptom diagnosis. A mathematical analysis of epigastric pain," *Ann. Intern. Med.*, Vol. 59 (1963), pp. 145–154.

[14] W. J. WILSON, A. W. TEMPLETON, A. H. TURNER, and G. S. LODWICK, "The computer analysis and diagnosis of gastric ulcers," *Radiology*, Vol. 85 (1965), pp. 1064–1073.

Bone Tumors

[15] G. S. LODWICK, C. L. HAUN, W. E. SMITH, R. F. KELLER, and E. R. ROBERTSON, "Computer diagnosis of primary bone tumors," *Radiology*, Vol. 80 (1963), pp. 273–275.

[16] G. S. LODWICK, "Computer analysis of tumor roentgenograms," *Proceedings of the Fifth IBM Medical Symposium*, Endicott, IBM, 1963, pp. 213–224.

[17] G. S. LODWICK, A. H. TURNER, L. B. LUSTED, and A. W. TEMPLETON, "Computer-aided analysis of radiographic images," *J. Chron. Dis.*, Vol. 19 (1966), pp. 485–496.

Neuropathology and Psychiatry

[18] A. BIRNBAUM and A. E. MAXWELL, "Classification procedures based on Bayes' formula," *Appl. Statist.*, Vol. 9 (1961), pp. 152–169.

[19] N. S. MISYUK, N. A. LEPESHINSKIY, O. A. LISKOVETS, and A. S. MASTYKIN, "Diagnosis of brain tumors with the help of the Universal 'Ural-1' computer," *Zhur. Nevr. i Psikh. (im Korsakava)*, Vol. 64 (1964), pp. 453–458. (In Russian; English translation available JPRS 24759, May 26, 1964.)

[20] J. E. OVERALL and J. E. GORHAM, "A pattern probability model for the classification of psychiatric patients," *Behavl. Sci.*, Vol. 8 (1963), pp. 108–116.

B. LIKELIHOOD RATIO

Hematology

[21] M. LIPKIN, "The likelihood concept and data processing in the diagnostic process," *The Diagnostic Process*, Proceedings of a Conference held at the University of Michigan, May 1963 (edited by J. A. Jacquez) Ann Arbor, Malloy Lithographing, 1964, pp. 255–273.

[22] ———, "The likelihood concept in differential diagnosis," *Perspec. Biol. Med.*, Vol. 7 (1964), pp. 485–497.

Screening and Diagnosis

[23] M. F. COLLEN, L. RUBIN, J. NEYMAN, G. B. DANTZIG, R. M. BAER, and A. B. SIEGELAUB, "Automated screening and diagnosis," *Amer. J. Publ. Health*, Vol. 54 (1964), pp. 741–750.

[24] R. M. BAER, "A note on computer diagnosis," *IEEE Trans. Biomed. Engin.*, Vol. ME-11 (1964), pp. 8–12.

C. DISCRIMINANT FUNCTION

Diabetes

[25] T. BOGDANEK, "The application of mathematical discrimination in the differential diagnosis of two types of diabetes," *Progress in Biocybernetics* (edited by N. Weiner and J. P. Schade), New York, Elsevier, 1964, Vol. 1, pp. 45–54.

Thyroid Function

[26] J. E. OVERALL and C. M. WILLIAMS, "Models for medical diagnosis. Part one: theoretical," *Med. Dok.*, Vol. 5 (1961), pp. 51–56; "Part two: experimental," *Med. Dok.*, Vol. 5 (1961), pp. 78–80.

D. FACTOR ANALYSIS

Thyroid Function

[27] ———, "Comparison of alternative computer models for thyroid diagnosis," *Proceedings of the San Diego Symposium for Biomedical Engineering*, La Jolla, Symposium for Biomedical Engineering, 1963, Vol. 3, pp. 141–145.

Psychiatry

[28] J. E. OVERALL, "A configural analysis of psychiatric diagnostic stereotypes," *Behavl. Sci.*, Vol. 8 (1963), pp. 211–219.

Use of Cornell Medical Index

[29] A. J. VAN WOERKAM and K. BRODMAN, "Statistics for a diagnostic model," *Biometrics*, Vol. 17 (1961), pp. 229–318.

Urology

[30] H. H. ZINSSER, "Patient derived autodefinition of a disease: pyelonephritis," *The Diagnostic Process*, Proceedings of a Conference held at the University of Michigan, May 1963 (edited by J. A. Jacquez), Ann Arbor, Malloy Lithographing, 1964, pp. 103–128.

E. MATCHING PROCEDURES

Dermatology

[31] A. L. NORINS, "Computers in dermatology," *Arch. Dermatol.*, Vol. 90 (1964), pp. 506–511.

Hematology

[32] M. LIPKIN, R. ENGLE, B. J. DAVIS, V. K. ZWORYKIN, R. EBALD, and M. SENDROW, "Digital computer as an aid to differential diagnosis," *Arch. Int. Med.*, Vol. 108 (1961), pp. 56–72.

Neurology

[33] V. I. SHTABTSOV and YU. A. OSTROUMOV, "Experience with the diagnostics of nervous diseases using an electronic digital 'Ural-2' (ETsVM) computer," *Zhur. Nevr. i Psikh.*, Vol. 64 (1964), pp. 515–519 (in Russian).

Personality Assessment

[34] H. P. ROME, "Automation techniques in personality assessment," *J. Amer. Med. Assoc.*, Vol. 182 (1962), pp. 1069–1072.

[35] J. S. PEARSON, W. M. SWENSON, H. P. ROME, P. MATAYA, and T. L. BRANNICK, "Further experience with the automated Minnesota Multiphasic Personality Inventory," *Proc. Mayo Clin.*, Vol. 39 (1964), pp. 823–829.

Surgery

[36] O. L. PETERSON and E. BARSAMIAN, "Diagnostic performance," *The Diagnostic Process*, Proceedings of a Conference held at the University of Michigan, May 1963 (edited by J. A. Jacquez), Ann Arbor, Malloy Lithographing, 1964, pp. 347–362.

ADDITIONAL REFERENCES

[37] J. A. JACQUEZ (ed.), *The Diagnostic Process*, Proceedings of a Conference held at the University of Michigan Medical School, May 1963, Ann Arbor, Malloy Lithographing, 1964, 391 pp. (Library of Congress Catalog Card: 64-21210.)

[38] R. L. ENGLE and B. J. DAVIS, "Medical Diagnosis: present, past and future. I. Present concepts of the meaning and limitations of medical diagnosis. II. Philosophical foundations and historical development of our concepts of health, disease and diagnosis. III. Diagnosis in the future, including a critique of the use of electronic computers as diagnostic aids to the physician," *Arch. Int. Med.*, Vol. 112 (1963), pp. 512–543.

[39] P. J. SPROSTY, "The use of questions in the diagnostic problem solving process," *The Diagnostic Process* (edited by J. A. Jacquez), Ann Arbor, Malloy Lithographing, 1964, pp. 281–308.

[40] H. J. A. RIMOLDI, "Testing and analysis of diagnostic skills," *The Diagnostic Process* (edited by J. A. Jacquez), Ann Arbor, Malloy Lithographing, 1964, pp. 315–343.

[41] W. R. ADAMS, "Studies of teaching of the diagnostic process," *The Diagnostic Process* (edited by J. A. Jacquez), Ann Arbor, Malloy Lithographing, 1964, pp. 367–385.

[42] B. KLEINMUNTZ, "Profile analysis revisited: a heuristic approach," *J. Counsel. Psychol.*, Vol. 10 (1963), pp. 315–321.

[43] W. R. REITMAN, "Information processing models in psychology," *Science*, Vol. 144 (1964), pp. 1192–1198.

[44] A. NEWELL and H. A. SIMON, "Computer simulation of human thinking," *Science*, Vol. 134 (1961), pp. 2011–2017.

[45] B. KLEINMUNTZ, "Computer simulation of diagnostic judgments," presented at the 36th annual meeting of the Midwestern Psychological Association, 1964.

[46] J. A. SWETS and W. FEURZEIG, "Computer-aided instruction," *Science*, Vol. 150 (1965), pp. 572–576.

[47] L. B. LUSTED, "Logic of the diagnostic process," *Methods of Information in Medicine*, Vol. 4 (1965), pp. 63–68.

[48] T. J. SCHEFF, "Decision rules, types of error and their consequences in medical diagnosis," *Behavl. Sci.*, Vol. 8 (1963), pp. 97–107.

[49] S. NISSEN-MEYER, "Evaluation of screening tests in medical diagnosis," *Biometrics*, Vol. 20 (1964), pp. 730–755.

[50] J. FRIEDEN, J. H. SHAPIRO, and A. R. FEINSTEIN, "Radiologic evaluation of heart disease," *Arch. Int. Med.*, Vol. 111 (1963), pp. 44–50.

[51] W. J. HORVATH, "The effect of physician bias in medical diagnosis," *Behavl. Sci.*, Vol. 9 (1964), pp. 334–340.

[52] L. B. LUSTED, "Logical analysis in roentgen diagnosis," *Radiology*, Vol. 74 (1960), pp. 178–193.

[53] A. F. TORONTO, L. G. VEASY, and H. R. WARNER, "Evaluation of a computer program for diagnosis of congenital heart disease," *Progr. Cardiovas. Dis.*, Vol. 5 (1963), pp. 362–377.

[54] H. J. GROSZ and K. G. GROSSMAN, "The sources of observer variation and bias in clinical judgments. I. The item of psychiatric history," *J. Nerv. Ment. Dis.*, Vol. 138 (1964), pp. 105–113.

[55] D. J. KING and R. F. MANEGOLD, "Consistency of cardiologists' prognostic judgments in cases of myocardial infarction," *Amer. J. Cardiol.*, Vol. 15 (1965), pp. 27–32.

[56] SIR G. W. PICKERING, *Conclusion: The Physician. Controlled Clinical Trials—A Symposium*, Oxford, Blackwell, 1960, pp. 163–167.

[57] R. S. LEDLEY and L. B. LUSTED, "Reasoning foundations of medical diagnosis," *Science*, Vol. 130 (1959), pp. 9–21.

[58] ———, "Medical diagnosis and modern decision making," *Mathematical Problems in the Biological Sciences;* Proceedings of Symposia in Applied Mathematics, Providence, American Mathematical Association, 1962, pp. 117–158.

[59] J. CORNFIELD, "Bayes' theorem," *Proceedings of the Sixth IBM Medical Symposium*, Endicott, IBM, 1964, pp. 163–196.

[60] ———, "Discriminant functions," *Proceedings of the Sixth IBM Medical Symposium*, Endicott, IBM, 1964, pp. 567–589.

[61] W. EDWARDS, H. LINDMAN, and L. J. SAVAGE, "Bayesian statistical inference for psychological research," *Psychol. Rev.*, Vol. 70 (1963), pp. 193–242.

[62] F. MOSTELLER and D. L. WALLACE, "Inference in an authorship problem," *J. Amer. Statist. Assoc.*, Vol. 58 (1963), pp. 275–309.

[63] ———, *Inference and Disputed Authorship: The Federalists*, Reading, Addison-Wesley, 1964.

[64] J. NEYMAN, *First Course in Probability and Statistics*, Chapter 5, New York, Ho lt, 1950