B. J. Flehinger
R. L. Engle, Jr.

HEME: A Self-Improving Computer Program for Diagnosis-Oriented Analysis of Hematologic Diseases

Abstract: HEME, a computer program for diagnosis-oriented analysis of hematologic diseases, accepts as input information about a patient and provides as output an ordered list of suggested diagnoses, an analysis of the logic behind these diagnoses, and a list of tests relevant to these diagnoses and not yet performed. The decision algorithm is based on Bayes' Theorem. Each disease in the system is individually analyzed, and the probability that the patient has the disease vs the probability that he does not is calculated. Bayesian methods of statistical inference are utilized in that the prior probabilities of the diseases and the probabilities of findings in given diseases were initially estimated from the judgment of experienced hematologists with the intention that they be modified automatically as data are accumulated. This program is intended for use in teaching hematology, as an aid to diagnosis, and as a means for studying the diagnostic process.

Introduction

HEME is a computer program that was conceived and developed to use Bayes' Theorem and Bayesian methods of inference for diagnosis-oriented analysis of hematologic diseases. It may be used by physicians to suggest diagnoses, to remind them of available test procedures, and to check their thinking at each stage of the diagnostic process. HEME is also useful as a teaching tool to train students in hematology and in methods of interacting with a computer. The program is a vehicle for identifying those features that are vital to the diagnostic process.

For more than twenty years physicians and computer scientists have been trying to produce useful computer aids to diagnosis. Our work on aids to diagnosis of hematologic diseases was initiated in 1952 when Lipkin and Hardy used McBee marginal punched cards to match the findings about a patient with a predetermined set of findings thought to characterize each of the 27 hematologic diseases in the system. A hand-operated mechanical sorting process selected those diseases for which the findings best matched those of a patient, and a score was calculated for the patient's findings in relation to each of the selected diseases [1-3]. Using these same principles our group of investigators organized by Zworykin and Lipkin developed a computer program to sort the data and print out information about the matching of a patient's findings with findings characteristic of diseases. The first demonstration of this system on a computer in 1957 applied to 20 hematologic diseases [4-8]. Later, the group developed a larger system. A thorough literature search resulted in the tabulation of incidence figures for the important findings in 75 hematologic diseases [9]. We used the incidence figures to estimate weights for each of 540 findings in each disease, and we developed algorithms to utilize these weights to calculate scores for a patient's findings in relation to each disease [10-11]. The incidence and weight tables from these previous studies were used to help in arriving at the value judgments required for the Bayesian program initiated in 1967 [11] and reported in this paper. Portions of this work have been presented at meetings [12-14].

Many investigators have used decision theory in medical diagnosis. Ledley and Lusted in 1959 [15] in a major contribution discussed the logic of medical diagnosis and the application of symbolic logic, probability, and value theory in medical decision making. They introduced the use of decision trees in diagnosis and suggested the use of Bayes' Theorem in medicine. Their paper provided considerable stimulus to us and other workers. Much subsequent work has been the elaboration and practical application of their ideas. By 1961 Warner, Toronto, Veasey, and Stephenson [16] had applied Bayes' Theorem to the diagnosis of 33 congenital heart diseases using 50 findings [17-18]. In addition to the contributions of many other investigators, significant ideas in the field of medical decision making came from Gorry and Barnett [19], who applied sequential decision theory to diagnostic problems, and from Gustafson [20], who experimented with subjective judgment in the estimation of probabilities.

Description of HEME

The HEME program performs probabilistic calculations that relate a patient's findings to the diagnosis of hematologic diseases. It contains the following elements:

Table 1 Hematologic diseases presently used in HEME program.

01	AGRANULOCYTOSIS
02	ANEMIA OF MALABSORPTION SYNDROME
03	APLASTIC ANEMIA
04	CHRONIC MYELOGENOUS LEUKEMIA
05	IRON DEFICIENCY ANEMIA
06	MULTIPLE MYELOMA
07	POLYCYTHEMIA VERA
08	PERNICIOUS ANEMIA
09	MEGALOBLASTIC ANEMIA OF PREGNANCY
10	INFECTIOUS MONONUCLEOSIS
11	DRUG INDUCED HEMOLYTIC ANEMIA
12	SICKLE CELL ANEMIA
13	SICKLE CELL TRAIT
14	HODGKIN'S DISEASE
15	ACUTE LEUKEMIA
16	LYMPHOSARCOMA
17	IDIOPATHIC THROMBOCYTOPENIC PURPURA
18	SECONDARY POLYCYTHEMIA
19	ANEMIA OF LIVER DISEASE
20	CHRONIC LYMPHATIC LEUKEMIA
21	RETICULUM CELL SARCOMA
22	GAUCHER'S DISEASE
23	FACTOR VIII DEFICIENCY
24	HEREDITARY SPHEROCYTOSIS
25	ERYTHROBLASTOSIS FETALIS
26	ANEMIA OF INFECTION
27	HEMOCHROMATOSIS
28	LUPUS ERYTHEMATOSUS
29	THALASSEMIA MAJOR
30	THALASSEMIA MINOR
31	NIEMANN-PICK DISEASE
32	GIANT FOLLICULAR LYMPHOMA
33	CONGENITAL AFIBRINOGENEMIA
34	CONGENITAL SEX-LINKED AGAMMAGLOBULINEMIA
35	CONGENITAL SWISS-TYPE AGAMMAGLOBULINEMIA
36	ANEMIA OF HYPOTHYROIDISM
37	NONHEMOLYTIC ANEMIA OF MALIGNANCY
38	THROMBOTIC THROMBOCYTOPENIC PURPURA
39	PRIMARY IDIOPATHIC NON-TROPICAL HYPERSPLENISM
40	ACQUIRED IDIOPATHIC REFRACTORY SIDEROBLASTIC ANEMIA

1. A list of diseases that may be diagnosed along with the frequency Φ_i with which these diseases occur in the population under consideration. The present list of 40 diseases appears in Table 1. The frequencies are those that are thought to occur on the hematology service of the New York Hospital; they may be modified through Bayesian inference as data are accumulated.

2. A list of descriptors considered relevant to the diagnosis of hematologic diseases. A descriptor is defined as a property to be described by the patient (demographic characteristic or symptom), observed by the physician (sign), or measured in the laboratory (laboratory characteristic). Typical descriptors are sex, blood pressure, hepatosplenomegaly, abdominal pain, and leukocyte count.

3. For each descriptor a finite number of findings that are thought to characterize the clinically significant states that the descriptor may assume. These findings must be mutually exclusive and exhaustive, so that each patient may be characterized by one and only one finding for each descriptor. The absence of a sign or symptom is itself a finding. Findings corresponding to the above descriptors are:

Descriptor	*Findings*
Sex	Male; female
Blood pressure	Normal; high; low
Hepatosplenomegaly	Neither; liver only; spleen only; both (enlarged)
Abdominal pain	Absent (−); present (+)
Leukocyte count	<3000; 3000-4999; 5000-9999; 10000-49999; 50000-99999; ≥100000

In Table 2, 58 of the 585 findings presently used are listed.

4. For each disease and for each finding relevant to that disease, there are two probabilities, also known as likelihoods, defined as follows:

p_{ij} is the probability that a ptaient with disease i has finding j at the time when the disease is diagnosed.

q_{ij} is the probability that a patient who does *not* have disease i, but for whom the descriptor corresponding to finding j is observed during the diagnostic process, does have finding j at the time of observation.

Both these probabilities refer to patients on the hematology service at the New York Hospital.

Initally, the p_{ij} and q_{ij} were estimated from the judgment of the clinicians responsible for the program, based on frequency data that were previously collected [9]. Each estimate has been recorded as the ratio of two integers, parameters of a prior distribution from the beta class, i.e.,

$$p_{ij} = m_{ij}/n_i; \qquad q_{ij} = r_{ij}/s_i.$$

In this scheme the clinician not only guesses values for p and q but he also indicates numerically how sure he is of his guess. The larger the values of the numerator and denominator the more certain is the estimate of p or q, i.e., the smaller is the variance of the prior distribution. Suppose, for example, the frequency of some finding in a given disease is thought to be one in ten. If the physician were quite confident of that, he might set m_{ij} at 1000 and n_i at 10000. On the other hand, if he were very unsure of his guess, he might set m_{ij} at 1 and n_i at 10. It is intended that these estimates be modified automatically from information in the Patient Data File described in item 6 below.

5. Bayes' Theorem, as used in the HEME program, computes, for each disease, the probability that the patient has that disease vs the probability that he does not, i.e.,

$$\begin{aligned}\text{Prob (disease } i/\text{findings)} &= \\ &\Phi_i \text{ Prob (findings/disease } i) \\ &\div [\Phi_i \text{ Prob (findings/disease } i) \\ &+ (1-\Phi_i) \text{ Prob (findings/not disease } i)].\end{aligned}$$

Assuming that the findings are mutually independent, we set

$$\text{Prob (findings/disease } i) = \prod_{\text{findings}} p_{ij};$$

$$\text{Prob (findings/not disease } i) = \prod_{\text{findings}} q_{ij}.$$

The question of independence and our efforts to approximate it are discussed later under the heading "Independence."

We now have the version of Bayes' Theorem that is incorporated into the HEME program:

$$\text{Prob (disease } i/\text{findings)} = \frac{\Phi_i \Pi p_{ij}}{\Phi_i \Pi p_{ij} + (1-\Phi_i) \Pi q_{ij}}.$$

Each product is taken over all the patient's findings.

Dividing both numerator and denominator in the previous expression by $(1-\Phi_i)\ \Pi\ q_{ij}$ we find:

$$\text{Prob (disease } i/\text{findings)} = \frac{[\Phi_i/(1-\Phi_i)]\ \Pi\ (p_{ij}/q_{ij})}{[\Phi_i/(1-\Phi_i)]\ \Pi\ (p_{ij}/q_{ij}) + 1}.$$

We can observe from this equation that it is the quantity p_{ij}/q_{ij}, the likelihood ratio of disease i in relation to finding j, that indicates the effect of the observation of a finding on the diagnosis of a disease. Thus, if p_{ij}/q_{ij} is much larger than 1, the observation of the finding tends to lead to the diagnosis of the disease; if p_{ij}/q_{ij} is much smaller than 1, the finding tends to rule out the disease; and if p_{ij}/q_{ij} is close to 1, the finding has little relevance to the diagnosis of the disease.

The version of Bayes' Theorem presented here was used by Nugent and co-workers for the decision of whether or not a patient had Cushing's Syndrome [21]. Our group has initiated its use in making simultaneous decisions about whether or not a patient has each of a wide range of diseases. In this version Bayes' Theorem is applied separately for each disease, and each time it is used it refers to a universe that consists of only two groups, patients who have the given disease and patients who do not have the disease. The patients who do not have the disease belong to the specified population of patients under consideration, and they may have some

Table 2 Portion of list of findings presently used in HEME program.

PERIPHERAL BLOOD (PB) MORPHOLOGY

	268+	RBC INDICES HGB	<7 , MCV >94, MCH >30
	269+	RBC INDICES HGB	<7 , MCV 80-94, MCH >30
	270+	RBC INDICES HGB	<7 , MCV <80, MCH >30
	271+	RBC INDICES HGB	<7 , MCV <80, MCH <30
	272+	RBC INDICES HGB	7-12.9, MCV >94, MCH >30
	273+	RBC INDICES HGB	7-12.9, MCV 80-94, MCH >30
	274+	RBC INDICES HGB	7-12.9, MCV <80, MCH >30
	275+	RBC INDICES HGB	7-12.9, MCV <80, MCH <30
	276+	RBC INDICES HGB	13-17 RBC 3.5-6, PCV 30-55
	277+	RBC INDICES HGB	>17 , RBC >5, PCV >50
	278+	LEUKOCYTE COUNT	<3000
	279+	LEUKOCYTE COUNT	3,000-4,999
	280+	LEUKOCYTE COUNT	5,000-9,999
	281+	LEUKOCYTE COUNT	10,000-49,999
	282+	LEUKOCYTE COUNT	50,000-99,999
	283+	LEUKOCYTE COUNT	≥100,000
	284+	LYMPHOCYTES	<20%
	285+	LYMPHOCYTES	20-39%
	286+	LYMPHOCYTES	40-59%
	287+	LYMPHOCYTES	60-79%
	288+	LYMPHOCYTES	≥80%
	289+	LYMPHOCYTES ATYP IN PB-NONE	
	290+	LYMPHOCYTES ATYP IN PB < 10% TOTAL LYMPHS	
	291+	LYMPHOCYTES ATYP IN PB ≥ 10% TOTAL LYMPHS	
292−	293+	MONOCYTES > 5%	
294−	295+	EOSINOPHILS > 3%	
	296+	GRANULOCYTES (NEUT, EOSIN, BASOPHILS)	<2%
	297+	GRANULOCYTES (NEUT, EOSIN, BASOPHILS)	2-49%
	298+	GRANULOCYTES (NEUT, EOSIN, BASOPHILS)	50-69%
	299+	GRANULOCYTES (NEUT, EOSIN, BASOPHILS)	≥70%
300−	301+	NEUTROPHILS HYPERSEGMENTED	
302−	303+	GRANULOCYTES IMMATURE IN PB > 4%	
	304+	BLASTS, PROMYEL., PROLYMPH. IN PB	NONE
	305+	BLASTS, PROMYEL., PROLYMPH. IN PB	0- 9%
	306+	BLASTS, PROMYEL., PROLYMPH. IN PB	10-49%
	307+	BLASTS, PROMYEL., PROLYMPH. IN PB	≥50%
308−	309+	ANISOCYTOSIS & POIKILOCYTOSIS	
310−	311+	TARGET CELLS	
312−	313+	NUCLEATED ERYTHROID CELLS IN PB	
	314+	SPHEROCYTES IN PB—ABSENT	
	315+	SPHEROCYTES IN PB—OCCASIONAL	
	316+	SPHEROCYTES IN PB—FREQUENT	
	317+	PLATELET COUNT	<150,000
	318+	PLATELET COUNT	150,000-399,999
	319+	PLATELET COUNT	400,000-999,999
	320+	PLATELET COUNT	≥1,000,000
	321+	RETICULOCYTE COUNT	<1%
	322+	RETICULOCYTE COUNT	1-1.9%
	323+	RETICULOCYTE COUNT	2-4.9%
	324+	RETICULOCYTE COUNT	5-9.9%
	325+	RETICULOCYTE COUNT	≥10%

other disease or they may be normal. Patients who have the disease might also have one or more other diseases, and allowances for this possibility are made in the estimated values of p. In the usual version the universe consists of patients each of whom has one and only one disease in the system, so that the estimated probabilities of all diseases must sum to one. In our application, the probabilities of all diseases do not add up to one. This version allows for the very real possibility that a patient has more than one disease. Whereas this may not be significant if the system is limited to hematologic diseases, it might be very important in a broader diagnostic scheme. Furthermore, the usual version requires that probabilities be estimated for every finding in every disease, whereas our program contains entries only for those findings considered *relevant* for each disease. Five-hundred eighty-five findings in 40 diseases would require that 23400 probabilities be estimated in the usual formulation. Our program uses only about 4000 p_{ij}/q_{ij} ratios. In the following sections the importance of the p_{ij}/q_{ij} quantities in the use of the program are explained. The significance of these quantities is unique to this version of Bayes' Theorem.

6. It is planned that a Patient Data File be maintained in the computer for the purpose of improving the estimates of the probabilities of findings in diseases and the probabilities of diseases in the population from the feedback of information accumulated through the use of the program. It is intended that, whenever a final diagnosis for a patient is reached by a consensus of physicians responsible for the program, an edited list of all the patient's findings be entered into the computer. The p_{ij} tables are updated automatically at regular intervals from this file as follows: Suppose the previous value of p_{ij} was m_{ij}/n_i and y_i new patients with disease i are subsequently diagnosed, of whom x_{ij} have finding j. Then the new value of p_{ij} is set at $(m_{ij}+x_{ij})/(n_i+y_i)$. This procedure provides a rational framework in which to combine clinical judgment with data and is based on standard methods of Bayesian inference [22]. As more and more data are accumulated, the facts tend to outweigh the initial clinical judgment in the estimated value of p_{ij}. Analogous methods may be used to revise values of q_{ij} and Φ_i.

Independence of findings

A critical step in most applications of Bayes' Theorem is the assumption that findings are independent of one another. This is equivalent to expressing the probability of a complex of findings as the product of the probabilities of the individual findings. This assumption is virtually never strictly fulfilled. Dependence arises in a number of different ways in the medical area. Primarily, it must be realized that all the findings observed in a single patient arise from interrelated mechanisms that developed in the same genetic and environmental setting. It is common for two or more findings that are measured quite separately to be controlled by the identical or closely linked physiological mechanisms. On the one hand, when these mechanisms function normally, many findings are simultaneously maintained in their normal ranges, whereas a single malfunction can be reflected in a number of apparently independent abnormal findings. Pulse rate and body temperature are obvious examples of independently observed descriptors with this kind of physiological dependence.

A second form of dependence occurs when certain descriptors are observed only when specified findings are present. Questions about pregnancy are asked only of females over the age of ten, and lymph node biopsy results are obtained only when other findings lead to the decision to perform a biopsy of the lymph node.

Finally, there is a type of dependence that arises from the man-made organization of diseases. Even if findings are independent in the context of one disease framework, a slight change in the organization of diseases destroys the independence. An example illustrates this point. Consider the disease acute leukemia and its two subdivisions, acute lymphatic leukemia (ALL) and acute myelogenous leukemia (AML). Within each of the two subdivisions considered as separate diseases, it is reasonable to assume that the two findings "Age <20" and "Bone Marrow Shift to Left, Myeloblasts" are independent, so that the probability that a patient has both findings is equal to the product of the probabilities of the individual findings. Now the patient with ALL is most likely younger than 20, say with probability 0.80, and is unlikely to have increased myeloblasts found, say with probability 0.05. The probability that he is younger than 20 *and* shows increased myeloblasts is then 0.80 × 0.05 or 0.04. On the other hand, in AML the patient is likely to be older and almost surely increased myeloblasts are found in his bone marrow, so we assume the probability of age less than 20 is 0.15 and of increased myeloblasts found is 0.95, so that both are present with a probability of 0.15 × 0.95 or 0.1425.

Now suppose that, instead of considering ALL and AML as distinct disease entities, we merge them and consider only the single disease acute leukemia. We will demonstrate that the two findings "Age <20" and "Bone Marrow Shift to Left, Myeloblasts" are no longer independent in the merged disease. Let 2/3 of the acute leukemics have the lymphatic type and 1/3 have the myelogenous type. We calculate the probabilities of Age <20, increased myeloblasts, and both findings in the mixed disease by taking weighted averages of their probabilities in the individual diseases, i.e., we add 2/3 of the probability in ALL to 1/3 of the probability in AML.

For example, for age <20 we add 2/3 of 0.80 to 1/3 of 0.15 to obtain 0.5833. In similar fashion we find that the probability of increased myeloblasts in the mixed population is 0.35 and of both findings is 0.0742. If age and myeloblasts were independent, the probability of both findings would be equal to the product of 0.5833 and 0.35 or 0.2042, which is 275% of the true probability. Not unexpectedly we find that in a mixed population of people with acute leukemia, ALL or AML, age and bone marrow myeloblasts are not independent. The older members of the population are more likely to have AML and are therefore more likely to have myeloblasts increased.

In general, let us consider two diseases and two findings that are independent within each of these diseases. By a generalization of the reasoning used above it may be proved that, if the two diseases are considered as one, the findings are independent in the merged disease only if the frequency of one of the findings is the same in both diseases [23]. We must bear in mind that the classification of a population into diseases is man-made and is subject to change. Therefore, in any diagnostic system based on Bayes' Theorem an assumption of independence between findings is only an approximation that ordinarily yields results of the correct order of magnitude. In our version of Bayes' Theorem, which uses quantities like q_{ij}, the probability of finding j in the absence of disease i, the lack of independence between findings could be particularly troublesome, because absence of one disease is a merger of many others. This problem of lack of independence plagues any diagnostic scheme, Bayesian or not, intended to relate a large number of diseases to a large number of findings. The problem of investigating all possible dependencies assumes a magnitude well beyond human understanding or computer capabilities.

One approach to eliminating obvious dependencies is the construction of separate disease-finding tables for different demographic groups, distinguished by sex, age, ethnic group, or geographic location. The more nearly uniform a population with respect to these characteristics, the fewer problems arise due to dependency.

In the HEME program an attempt has been made to eliminate dependencies between findings by coupling descriptors that are known to be highly dependent. For example, instead of considering the liver and spleen separately, we have one descriptor called hepatosplenomegaly, with findings of neither, liver only, spleen only, and both (enlarged). The various measurements of red blood corpuscles (HGB, RBC, and PCV) are considered together and divided into categories considered to have diagnostic significance. An examination of Table 2 shows that we consider normal blood; macrocytic, normocytic, and microcytic anemias of two degrees of severity; and polycythemia as the meaningfully different categories of these correlated measurements. A similar approach to handling dependencies was recently suggested by Norusis and Jacquez [24, 25].

Operation of the program

The HEME program has been written in FORTRAN to operate interactively at a computer terminal. We present a typical, though oversimplified, exchange between a physician or student and the computer.

At the start of the program the computer asks which function the physician wishes to exercise. The physician indicates by entering a 1 that he wishes to enter a series of findings on a patient. After referring to a code list, he enters, by code number, those findings that have already been determined.

```
FUNCTION?
1
ENTER SXS
7, 12, 14, 21, 56, 64, 74, 76, 89, 105, 134, 140, 150, 200,
220, 275, 280, 284, 289, 292, 298, 304, 309, 318, 491,
495, 497, 501, 503
ENTER SXS
```

If the physician wishes to check his input and obtain a list of the findings he entered, he requests function 4. He is given a choice of a complete list or, if he enters a 1, a list of demographic and all abnormal findings. In this case the physician has asked for the latter.

```
FUNCTION?
4
ENTER "1" FOR ABNORMAL HX FORM
1
  HIST
   7  AGE 40-49 YRS
  12  SEX MALE
  14  RACE WHITE
  56  FATIGUE, LETHARGY OR MALAISE
  74  PALPITATION
  76  PRECORDIAL PAIN
  89  BOWEL FUNCTION-DIARRHEA
PE
XRAY
LAB
  275  RBC INDICES HGB 7-12.9, MCV <80,
                            MCH <30
  284  LYMPHOCYTES <20%
  309  ANISOCYTOSIS & POIKILOCYTOSIS
```

The physician then requests the listing of differential diagnoses and probabilities by entering function 5. The computer calculates the probability that the patient has each of the 40 diseases currently in the system and lists those with probability greater than 1%. The physician is able to compare this list of probabilities with his own clinical judgment. At this stage there is not enough information to give a high probability of any disease.

```
FUNCTION?
5
DIFFERENTIAL DIAGNOSIS
#  5 IRON DEFICIENCY ANEMIA        38.7%
# 30 THALASSEMIA MINOR              9.1%
#  2 ANEMIA OF MALABSORPTION        2.8%
# 37 ANEMIA OF MALIG., NON-HEM.     1.3%
```

Because iron deficiency anemia has a relatively high score in the differential diagnosis, the physician asks the computer for the rationale behind that diagnosis by entering function 6 and the disease code number 5 for iron deficiency anemia. The computer prints out a list of the findings supporting the diagnosis and those opposing the diagnosis, in order of their significance. In this example there is only one finding in each category. However, the weight of each finding is shown as *p/q* for making the diagnosis or *q/p* for ruling out the diagnosis. For convenience, when *p/q* is less than one, the inverse, *q/p*, is displayed. If either ratio is greater than 1000, its value is shown as ************.

```
FUNCTION?
6
ENTER DISEASE NUMBER FOR P/Q RATIOS?
5
RECORDED SYMPTOM P/Q RATIOS FOR
                  5 IRON DEFICIENCY ANEMIA
P/Q FOR DIAGNOSIS
  12.0 # 275 RBC INDICES HGB 7-12.9, MCV
                              <80, MCH <30
Q/P AGAINST DIAGNOSIS
  2.5 #  12 SEX MALE
```

If the physician thinks there is enough evidence to pursue the diagnosis of iron deficiency anemia, he may request a list of suggested findings to investigate. He does this by entering function 9 and the code number for the disease in question. Unrecorded findings that support or oppose the diagnosis are listed in order of *p/q* or *q/p*. The physician compares this list with his own judgment and decides on the priorities for further examinations.

```
FUNCTION?
9
ENTER DISEASE NUMBER FOR P/Q RATIOS?
5
UNRECORDED SYMPTOM P/Q RATIOS FOR
                  5 IRON DEFICIENCY ANEMIA
P/Q FOR DIAGNOSIS
 980.0  # 345  BM IRON-ABSENT
  80.0  # 427  SERUM COPPER HIGH
  33.0  # 395  RESPONSE TO IRON-
                              POSITIVE
  19.8  # 435  SERUM IRON BINDING CAP
                          (TOTAL) HIGH
  17.4  # 340  BM CELLULARITY-
                             INCREASED
  16.5  # 176  FINGERNAILS-SPOONED
                            OR BRITTLE
  10.0  # 311  TARGET CELLS
  10.0  # 321  RETICULOCYTE COUNT <1%
   9.3  # 430  SERUM IRON LOW <70
   8.3  # 197  TONGUE SMOOTH OR SORE
   3.0  #  72  DYSPNEA
   2.7  # 564  ACHLORHYDRIA-PRESENT
Q/P AGAINST DIAGNOSIS
******  # 433  SERUM IRON BINDING CAP
                           (TOTAL) LOW
******  # 432  SERUM IRON HIGH >130
******  # 348  BM IRON-INCREASED
 100.0  # 325  RETICULOCYTE COUNT
                                >= 10%
  97.0  # 394  RESPONSE TO IRON-
                              NEGATIVE
  50.0  # 338  BM CELLULARITY-
                             DECREASED
  10.0  # 337  BM MEGALOBLASTIC
   5.0  # 426  SERUM COPPER NORMAL
```

After any or all of the additional tests suggested, or any other tests the physician wishes, have been performed, the physician may enter the additional finding codes into the computer through function 2.

```
FUNCTION?
2
ENTER SXS
435, 175, 311, 321, 430
ENTER SXS
```

A revised differential diagnosis may then be requested through function 5. The findings, now including low serum iron and high iron binding capacity, have made the diagnosis of iron deficiency anemia virtually certain. At the same time thalassemia minor has appeared on the list with relatively high probability. Further studies and interaction with the computer would be required to pursue this possibility.

```
FUNCTION?
5
DIFFERENTIAL DIAGNOSIS
#  5 IRON DEFICIENCY ANEMIA        100.0%
# 30 THALASSEMIA MINOR              89.9%
# 37 ANEMIA OF MALIG., NON-HEM.     28.0%
#  2 ANEMIA OF MALABSORPTION         2.8%
```

From examples such as these we have concluded that HEME is useful in teaching hematology and has potential as an aid in diagnosis and as a means of studying the diagnostic process itself.

Testing of program

Immediately after the HEME program was written, it was exercised on a series of 31 cases of hematologic disease selected from the medical record library of New York Hospital. After each case was analyzed by the work group responsible for HEME, a subjective judgment was made about how well the program would have performed as an aid to a physician in the diagnostic process. In this study, 14 cases were rated excellent, seven good, one fair, three poor, and six not evaluable because the correct diagnoses were not yet in the system.

Encouraged by these results, we proceeded to seek experience with HEME as an aid to diagnosis and as a teaching tool. R. Strauss, who was a fourth-year medical student at the time, followed 30 hematology patients at the New York Hospital from time of admission until diagnosis was accepted. Taking information from the charts and from conferences with responsible physicians, he entered data into the program at frequent intervals, monitored the output of suggestions of diagnoses and findings to test, and compared these results with the decisions and procedures specified by the responsible physicians. In the course of this project, he found that in many instances there was a good correlation between the results coming out of the computer and the thinking of the physicians responsible for the patients. Other cases exposed correctable errors and the absence of important diseases or findings from the system.

Most recently HEME has been used experimentally under the direction of Dr. R. Friedman at the University of Wisconsin School of Medicine as an aid in the teaching of hematology. Students have found it instructive and useful, and staff hematologists have been impressed with its teaching potential. They find that it encourages students to take account of previously unconsidered diagnoses and often leads to lively teaching sessions. In the course of this experiment, the HEME analyses of 44 cases were evaluated. A record was made of the diagnoses suggested by the program after findings observed in the initial study were entered. This process was repeated after the results of all relevant tests had been reported and a "correct" diagnosis was accepted by the physicians responsible for the patient. For each case, we observed the ranking of the "correct" diagnosis by HEME, and these observations are summarized in Table 3.

Table 3 Results of trial of HEME as teaching aid at University of Wisconsin School of Medicine.

Ranking of "correct" diagnosis by HEME	*Initial study*	*Final decision*
1	16	25
2	7	8
3	8	1
4	0	2
5	5	0
>5	3	2
Correct diagnosis not in HEME	5	5
Case not yet complete		1

In all these experiments, it was demonstrated that HEME has fine potential as an aid to diagnosis and as a teaching device. In almost every case it stimulated thought in the right direction and jogged the memory about diagnoses and tests to consider. The p/q ratio was found to be a natural and intuitively comfortable way to think about the relevance of a finding to a disease, both in the initial assignment of p and q and in the interpretation of the analysis by users of the program. Exercising the program revealed a number of correctable errors as well as diagnoses and additional findings that need to be added to the system. These results lead to the conclusion that more effort is required and is definitely justified.

Discussion

In this paper we have described a Bayesian program for the diagnosis of hematologic diseases. It is distinguished from other Bayesian programs by the fact that each disease is analyzed individually to determine the probability that the patient has the disease vs the probability that he does not and by the property of combining initial clinical judgment with accumulating data in a self-improving mode. This program provides a framework whereby the intensive work of a few able physicians concentrating on a group of medical diagnostic problems can produce a system of value to many physicians and the communities they serve. The system can be used to train medical personnel, assist in diagnostic decisions, and record patient data in coherent form.

The version of Bayes' Theorem used in HEME requires far fewer probabilities of findings in diseases than does the usual version. Since each disease is analyzed separately, the p and q need be entered only for those findings relevant to the diagnosis of that disease. Inherent in the system is the capacity to grow and improve itself in three ways. New diseases may be added without changing the rest of the system; new findings relevant to one or more diseases may be added with only minor changes; and the probabilities required for Bayes' Theorem may be modified automatically as data are accumulated.

Acknowledgments

The work described in this paper was supported in part under Contract PH-43-67-1337 with the National Institutes of Health, Public Health Service, Department of Health, Education, and Welfare. Dr. S. Allen, Dr. M. Lipkin, Dr. B. J. Davis, Dr. L. Leveridge, Dr. R. Friedman, Dr. R. Craig, Dr. K. Peele, Dr. G. Ubogy, Dr. L. Frenkel, Dr. R. Strauss and Mr. G. Novick also contributed to the development of this program. A part of this paper was presented before the American Clinical and Climatological Association on October 21, 1974, in Williamsburg, Virginia.

References

1. M. Lipkin and J. D. Hardy, "Differential Diagnosis of Hematologic Diseases Aided by Mechanical Correlation of Data," *Report NADC-MA-5505*, U.S. Naval Air Development Center, Johnsville, Pa., 1955.
2. M. Lipkin and J. D. Hardy, "Differential Diagnosis of Hematologic Diseases Aided by Mechanical Correlation of Data," *Science* **125,** 551 (1957).
3. M. Lipkin and J. D. Hardy, "Mechanical Correlation of Data in Differential Diagnosis of Hematological Diseases," *J. Am. Med. Assoc.* **166,** 113 (1958).
4. V. K. Zworykin, R. Lane, M. Lipkin, R. Ebald, R. Engle, and B. Davis, "A Bizmac Program for Medical Data Processing," *Medical Data Processing Demonstration*, Radio Corporation of America, Camden, NJ, December 13, 1957.
5. R. Ebald and R. Lane, "Digital Computers and Medical Logic," *IRE Trans. Med. Electron* **7,** 283 (1960).
6. M. Lipkin, "Correlation of Data with a Digital Computer in the Differential Diagnosis of Hematologic Diseases," *IRE Trans. Med. Electron* **7,** 243 (1960).
7. M. Lipkin, R. L. Engle, Jr., B. J. Davis, V. K. Zworykin, R. Ebald, M. Sendrow, and C. Berkley, "Digital Computer as Aid to Differential Diagnosis," *Arch. Int. Med.* **108,** 56 (1961).
8. M. Lipkin, "Digital and Analogue Computer Methods Combined to Aid in the Differential Diagnosis of Hematological Diseases," *Circ. Res.* **11,** 607 (1962).
9. M. A. Atamer, *Blood Diseases*, Grune and Stratton Inc., New York, 1963.
10. R. L. Engle, Jr., M. Lipkin, B. J. Flehinger, B. J. Davis, and L. L. Leveridge, "Computer-Assisted Differential Diagnosis of Hematologic Diseases," Report on work accomplished under Contract #PH 43-67-1337 with NIH, PHS, HEW, 1969.
11. M. Lipkin, R. L. Engle, Jr., B. J. Flehinger, L. J. Gerstman, and M. A. Atamer, "Computer-Aided Differential Diagnosis of Hematologic Diseases, *Proc. NY Acad. Sci.* **161,** 670 (1969).
12. R. L. Engle, Jr., M. Lipkin, and B. J. Flehinger, "Computer-aided Differential Diagnosis of Hematologic Diseases (A Bayesian probability model)," *Computer Diagnosis and Diagnostic Methods*, edited by J. A. Jacquez, Charles C. Thomas, Springfield, Ill., 1972, p. 279.
13. R. L. Engle, Jr. and B. J. Flehinger, "Computer-oriented Thinking in Hematology," *Clinically Oriented Documentation of Laboratory Data*, edited by E. R. Gabrieli, Academic Press, New York, 1972, p. 311.
14. R. L. Engle, Jr. and B. J. Flehinger, "Computer-aided Differential Diagnosis of Hematologic Diseases: A Bayesian Probability Model," *Modern Concepts in Hematology*, edited by G. Izak and S. M. Lewis, Academic Press, New York, 1972, p. 265.
15. R. S. Ledley and L. B. Lusted, "Reasoning Foundations of Medical Diagnosis: Symbolic Logic, Probability, and Value Theory Aid Our Understanding of How Physicians Reason," *Science* **130,** 9 (1959).
16. H. R. Warner, A. F. Toronto, G. Veasey, and R. Stephenson, "A Mathematical Approach to Medical Diagnosis: Application to Congenital Heart Disease," *J. Am. Med. Assoc.* **177,** 177 (1961).
17. H. R. Warner, A. F. Toronto, and L. G. Veasey, "Experience with Bayes' Theorem for Computer Diagnosis of Congenital Heart Disease," *Ann. NY Acad. Sci.* **115,** 558 (1964).
18. A. F. Toronto, L. G. Veasey, and H. R. Warner, "Evaluation of a Computer Program for Diagnosis of Congenital Heart Disease," *Progr. Cardiovascular Diseases* **5,** 362 (1963).
19. G. A. Gorry and G. O. Barnett, "Sequential Diagnosis by Computer," *J. Am. Med. Assoc.* **205,** 849 (1968).
20. D. H. Gustafson, J. J. Kestly, J. H. Greist, and N. M. Jensen, "Initial Evaluation of a Subjective Bayesian Diagnostic System," *Health Services Res.* **6,** 204 (1971).
21. C. A. Nugent, H. R. Warner, J. T. Dunn, and F. H. Tyler, "Probability Theory in Diagnosis of Cushing's Syndrome," *J. Clin. Endocrinol Metab.* **24.** 621 (1964).
22. H. Raiffa and R. Schlaifer, *Applied Statistical Decision Theory*, Harvard Business School, Boston (1961).
23. M. L. Bykhovskiy, "Probability Dependence of Symptoms in the Problem of Diagnosis," *Machine Diagnosis and Information Retrieval in Medicine in the USSR*, edited by A. A. Vishnevskiy, I. I. Artobolevskiy, and M. L. Bykhovskiy, *DHEW Pub No (NIH) 73-424*, US Dept. of Health, Education, and Welfare, 1973.
24. M. J. Norusis and J. A. Jacquez, "Diagnosis, I. Symptom Nondependence in Mathematical Models of Diagnosis," *Comput. and Biomed. Res.* **8,** p. 156 (1975).
25. M. J. Norusis and J. A. Jacquez, "Diagnosis, II. Diagnostic Models Based on Attribute Clusters: A Proposal and Comparisons," *Comput. and Biomed. Res.* **8,** p. 173 (1975).

Received May 1, 1975

B. J. Flehinger is located at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York. Dr. R. L. Eagle, Jr., is located at Cornell University Medical College and the New York Hospital, New York, N.Y.