1975

A Bayesian Utility Model of Physicians' Diagnostic Processes

Robert S. Woodward

20245

Research Report 7507

A BAYESIAN UTILITY MODEL OF PHYSICIANS' DIAGNOSTIC PROCESSES

Robert S. Woodward*

March 1975

I. Introduction

This paper develops a model of physicians' diagnostic processes which demonstrates the importance of each doctor's preferences as a determinant of his diagnostic pathways.[1] Using "normal days lost" as an index of disease seriousness and "normal days saved" as an index of treatment effectiveness, the paper argues that each diagnostic question or test can be characterized by a unique combination of different kinds of expected information. The hypothesis that physicians have preferences for these alternative informations is supported by the results of a diagnostic simulation from which a utility map is derived.

Several policy implications follow from recognizing the importance of preferences in diagnostic decisions. Since physicians are unlikely to have identical preferences for the alternative kinds of expected information available from diagnostic tests, the diagnostic pathways of physicians may differ widely. Thus, no standard diagnostic process will be discovered by recording the actions of even the most senior or learned physicians.[2] Nor are the Professional Standards Review Organizations going to be able to define diagnostic pathways which are optimal, or even appropriate, for all physicians.[3] Nor can the diagnostic computer programs be truly objective since they must incorporate the preferences of the programmer.[4]

This paper proceeds according to the following outline. Section II describes the diagnostic process in general terms. Section III specifies the model. Section IV reports the results of the application of the model to the diagnostic decisions of a general pracdioner for a patient with backache and dysuria. Section V examines what implications the model has for policy.

II. Medical Diagnosis

The diagnostic process is assumed to begin when a patient enters his physician's office complaining of a particular symptom. On the basis of the symptom, the patient's medical history and socio-demographic characteristics, disease incidence and the way the patient entered the office, doctors form an initial hypothesis about the patient's condition (Elstein, et al.). This hypothesis may be expressed as a judgment about the most likely condition and a list of less-likely possibilities. At one extreme, the doctor might be absolutely certain that the patient has a well-defined disease. At the other extreme, the doctor might conclude that the patient could have one of some number of diseases but might be totally unsure of which it is. In fact, the initial hypothesis may only be the incidence of diseases among the physician's patients.

A number of simplifications have been incorporated into this model. First, the patient is assumed to have no more than one disease or condition. Second, the presenting symptom is assumed to be accurately expressed and not to be an inconsequential false start from which the patient hopes the physician will discover the true and more serious problem. Third, the physician is assumed to consider only a limited and known number, n, of diseases at any stage of the diagnostic process. Conditions too rare to be actively investigated may be grouped into an "other" category. Since patients may complain of symptoms which have no organic disfunctional origin, another of the n conditions might be "psychological" disorders.

In practice, physicians appear to form their initial hypotheses in terms of very aggregated groups of diseases. For the symptom couplet, backache and dysuria, a general practitioner's first concern might be whether the patient has a self-limiting condition or a more serious, perhaps life-threatening, disease. For the same couplet, a urology specialist might form his initial hypothesis in terms of the aggregate classifications: stone, tumor or infection. As the diagnosis continues, and as tests suggest the presence of a disease within an aggregate group, the physician then considers the more detailed conditions within the indicated group. But since the computer can arrive at empirical judgments quickly without aggregating, this model considers only the "n" detailed diseases.

For the purposes of this model, the doctor's judgments are interpreted as a vector of subjective or "personal" (Lusted, 1968, p. 5) probabilities, $[P(d_1), P(d_2),\ldots,P(d_n)]$, where $P(d_i)$ is the probability of the patient having disease i. Perfect certainty would be indicated by assigning a probability of one to a single disease and zero to all others. Perfect uncertainty is indicated by assigning a value of 1/n to each disease. For example, a doctor who believes his patient has condition A, but recognizes B, C and D as possibilities, might assign probabilities of [.4, .3, .2, .1].

After forming an initial hypothesis, the doctor is likely to proceed with the diagnosis which may include: 1) some historical questions, such as "how long has the symptom been present?"; 2) a search for observable physical signs; 3) at least some parts of a general physical examination; 4) clinical or laboratory tests; 5) the confirming opinion of a colleague or specialist; 6) trial treatments; and 7) time. For purposes of this paper, these seven components are called "diagnostic procedures."

Although physicians often follow the above sequence, it is not necessarily optimal. For example, doctors whose case mix is dominated by minor or psychological conditions may initially prescribe aspirin and two days' rest to separate self-limiting diseases from persistent and perhaps more serious conditions.

Each diagnostic procedure is associated with a degree of specificity and reliability. A test is specific if the results confirm or reject the existence of a particular condition. A test is reliable if negative results do not occur when the patient has the condition and if positive results do not occur when the patient does not have the condition. For example, a rabbit test for pregnancy is relatively specific but is unreliable during the first few weeks since during that time negative results occur even if the woman is pregnant.

The data used to simulate the diagnostic process in this paper specifically estimate the probabilities of negative results when the patient actually has the disease. The possibility of positive results when patients do not have the disease is not considered in this simulation.

Additionally, each diagnostic procedure is associated with explicit and implicit costs. Explicit costs include the payments made by the patient and insurance programs to the doctor, the laboratory, and/or hospital. Implicit costs include the value of the doctor's and the patient's time. Since the vast majority of patients in Canada are covered by insurance, the implicit costs are assumed to be the binding constraints which serve to limit the diagnostic process.

III. Model of Diagnostic Decisions

In this section, the diagnostic process is described as the physician's effort to maximize his utility subject to a time constraint. In particular, the doctor's diagnostic decisions are described as the attempt to maximize his satisfaction from the mix of alternative kinds of expected information associated with each diagnostic procedure. The section begins in Part A with a discussion of "days lost" as a measure of disease seriousness. A description of the utility function in Part B is followed by an explanation of the expected information associated with each diagnostic test in Part C.

A. Days Lost

A measure of the seriousness of each disease is critically important to the empirical estimation of this paper. Most fundamentally, patients seek to avoid death, debility, impairment, pain, and anxiety. In the attempt to avoid these consequences of ill-health, patients must have preferences regarding the trade-off between each variable. For example, each patient has a willingness to accept some level of pain to avoid death or impairment. In their efforts to provide quality care, physicians must implicitly make some assumption about each patient's relative aversion to each of these variables.

As first approximation, and for the purposes of this paper, the physician is assumed to make health delivery decisions as if he perceived patient's satisfaction, and/or quality care, as inversely proportional to the loss of normal working days. This characterization is judged better than alternative measures and sufficiently accurate to facilitate a simulation of the diagnostic process. It is not recommended as a complete description or measure of either patients' satisfactions or quality care.

Within this context, the seriousness of an untreated disease is indicated by the expected loss of the patient's normal days. Similarly, the effectiveness of any treatment is measured by the additional days of normal life which the treatment is expected to add to the patient's life. Thus, the most serious disease would be associated with instant death, and the most effective treatment would be the one which would be expected to cure the patient most rapidly and completely.

Alternative measures of disease seriousness and treatment effectiveness include indices of utility and health status indices. Indexing the doctor's perception of a patient's disutility caused by each disease either directly, by asking the physician to index the disease discomfort, or indirectly with methods such as the Von Neumann-Morgenstein standard gambles (Torrence, G. W., et al., 1973), would have the advantage of including the influence of variables such as pain and debility along with days lost as the consequences of each disease.

The advantage of days lost over utility is, at least in part, attributable to the more objective characteristic of days lost. Physicians grapple with the problem of calculating expected days lost by dividing up patients with a given disease into the percent that will get better of their own accord in one or two weeks, the percent that will take one or two months to recover, the percent which will have effects that linger throughout their lifetime, the percent which will be likely to have slightly shorter lives, and the percent which are likely to die within a short period. While such figures are mostly unpublished, physicians appear to be relatively willing to generating the data. And while such a task involves substantial "guestimating", the degree of guessing must be less than that necessary to include the influence of pain or debility associated with each division of patients. For example, Torrence, who estimates social utility levels for five states of health, finds that physicians assign widely different utility values (0% to 99% of perfect health) to a single condition (hospital dialysis). (Torrence, G. W., et al., 1973, p. 158).

Despite the expected superiority of days lost over a utility index and in order to test the hypothesis that doctors behave so as to minimize the patient's general discomfort rather than to minimize the patient's days lost, a measure of disutility ("discomfort") associated with each disease was generated and included in the simulation. Because the large number of data points required by this model could be generated more quickly, the discomfort index was estimated directly, rather than using the standard gamble techniques suggested by Torrence, et al. The unsuccessful results of the diagnosis simulation using discomfort in the place of days lost are reported below.

The various measures of health status provide a second alternative days lost. Most measures of health status are designed for groups and are simply inappropriate. Moreover even the studies which assign a health status to individual patients appear to have designed their indices for an examination of a group's health levels[5] and are therefore too aggregate to be adequate for this paper. Although designing a health status index with sufficient sensitivity is judged beyond the scope of this paper, days lost might be considered as a first approximation.

B. Utility Function

Variables which provide physician's utility in diagnosis include: 1) being able to select effective treatment for the patient, E; 2) being certain that a patient does (or does not) have a serious disease, S; 3) being certain about the patient's condition, C. Other determinants which are less likely to appear important and are therefore excluded from the diagnostic simulations of this model, but which might be important in clinical practice, include selecting tests which: 1) provide certainty that a patient does (or does not) have an intrinsically interesting disease; 2) directly increase the physician's income; 3) determine the effectiveness of some treatments for a specific illness; and 4) provide etiological or pathological information which does not affect either treatment decisions or the doctors hypothesis.

In the diagnostic process, the value of the first variable--being able to select effective treatment--follows from the physician's desire to provide quality care for the patient. Providing effective care is important to the physician both because of a desire to fulfill his Hippocratic oath and because he gets certain satisfaction from increasing the utility of his patient. The second variable--certainty about serious conditions--has value to the MD as a means of avoiding unpleasant repercussions of incorrectly informing a patient and/or his family of the presence of a serious disease. The third variable--general certainty about the patient's condition--is included as an example of a variable which cannot be directly associated with the delivery of quality medical care. Nevertheless, general certainty may be sought by recently graduated doctors for educational purposes, or by specialists who wish to maintain their role as a source of diagnostic information to the referring primary physician.

1. Effective Treatment, E.

Where the doctor's hypothesis is expressed as subjective probabilities that the patient has each disease, and where it is possible to determine the expected increase in normal days for each treatment and disease, the expected increase in normal days for each treatment may be calculated.

(1) $$E_j = \sum_i P(d_i) \cdot E_{ij}$$

where $P(d_i)$ indicates the physician's subjective probability that the patient has condition i, and

E_{ij} indicates the expected number of days treatment j will save (or lose) if the patient has condition i.

After some amount of diagnosis, doctors might select the treatment associated with the largest expected increase in normal days, i.e.,

$E = \max(E_1, E_2, \ldots, E_j, \ldots, E_m)$. In the context of diagnostic decisions, a physician may select a particular test because he expects it to have the greatest impact on his ability to select the most effective treatment.

2. General Certainty, C

The index of general certainty is based on the subjective probabilities of the doctor's hypothesis. To compare the degrees of certainty indicated by various combinations of probabilities, the following index of a doctor's general certainty is proposed:

$$(2) \qquad C = \frac{n}{n-1}\left[\sum_{i=1}^{n} (P(d_i) - (1/n))^2\right]$$

where $P(d_i)$ is the physicians's judgment of the probability that the patient has the i^{th} disease or condition, and

n is the total number of possible diseases or conditions.

When the doctor is absolutely certain that a patient has a particular disease, C = 1. When the doctor has no clue to which of the n possible diseases the patient might have, C = 0. For the intermediate example indicated by the vector (.4, .3, .2, .1), C = .07. As part of the diagnostic process, some doctors may select a test because they believe it will have the greatest impact on their certainty about the patient's condition.

3. Certainty About Serious Diseases, S.

The index of certainty about serious diseases is similar to the index of general certainty. But since some diseases are more serious than others, certainty about the serious diseases may be more important than certainty about inconsequential diseases. In accordance with equation 2, the level of certainty for each disease is defined as the squared difference between the doctor's judgment about the likelihood that a patient has the disease and the probability that the patient would have the disease if the doctor were totally uncertain, 1/n. Since the expected loss of a patient's normal days indicates serious diseases, an index can be constructed by weighting the certainty of each disease by the expected days lost:

$$(3) \qquad S = \sum_i \left[L_i (P(d_i) - \frac{1}{n})^2 \right]$$

where L_i is the expected loss of normal days from disease i if untreated,

$P(d_i)$ is the physician's judgment of the probability that the patient has disease i, and

n is the total number of possible diseases or conditions.

Some physicians may select a diagnostic test because they expect it will give them the most information about certainty of serious diseases.

To summarize the model, physicians are assumed to make diagnostic decisions by maximizing their utility achieved from a series of variables. Total utility is constrained by the time they have available for all their patients. This maximization process may be expressed in the familiar Langrangian form:

$$(4) \qquad U(E,C,S) - \lambda \left\{ \sum_k \left[\sum_\alpha (t_\alpha A_{\alpha k}) + \sum_\ell (t_\ell B_{\ell k}) \right] - T \right\}$$

where E indicates effective treatment,

C indicates general certainty,

S indicates certainty about serious diseases (which could cause large loss of normal days),

λ is the Lagrangian multiplier and the marginal utility of time,

t_α is the physician's time necessary to perform diagnostic procedure α,

$A_{\alpha k}$ is a dummy variable with a value one if the physician performs the diagnostic procedure α on patient k and zero otherwise,

t_ℓ is the physician's time necessary to perform treatment,

$B_{j\ell}$ is a dummy variable with the value one if the physician performs treatment ℓ on patient k and zero otherwise, and

T is the total amount of time per day a physician devotes to all his patients.

C. Expected Information From Diagnostic Tests

This section demonstrates that each diagnostic procedure may be associated with an expected increase in each of the utility variables. Therefore physician preferences for diagnostic procedures can be translated into preferences for additional ability to select the most effective treatment, to be certain about serious diseases, or to be certain over all diseases.

The information which must be elicited from the participating physician includes his: 1) initial hypothesis, $P(d_i)$; 2) normal days lost for each disease if it is left untreated, L_i; 3) normal days lost for each disease and treatment; 4) the conditional probabilities of test outcomes for patients known to have each disease; and 5) the time each test takes the physician. In this model, the physician's subjective estimates of all this information is appropriate for two reasons. First, the model examines the physician's utility function. Thus, for example, the doctor's perceptions about treatment effectiveness are more appropriate than the truth. Second, since some of the combinations are ridiculous, much of the treatment effectiveness data are unobtainable from morbidity and mortality statistics. For example, physicians do not put patients with minor disc problems on renal dialysis machines.

With information about the incidence of test results for patients known to have certain diseases, the probabilities that a patient has a particular disease given a specific test result may be calculated according to the Bayes formula:

$$(5) \qquad P(d_i \mid \alpha = \alpha_j) = \frac{P(\alpha = \alpha_j \mid d_i) \cdot P(d_i)}{P(\alpha = \alpha_j)}$$

where $P(d_i \mid \alpha = \alpha_j)$ is the "posterior" probability that a patient has disease i given the test α has result α_j,

$P(\alpha = \alpha_j \mid d_i)$ is the probability that test α will have the result α_j if the patient is known to have disease i,

$P(d_i)$ is the "prior" probability or the physician's subjective judgment that the patient has disease i, and

$$P(\alpha = \alpha_j) = \sum_i \left\{ P(\alpha = \alpha_j \mid d_i) \cdot P(d_i) \right\} .$$

Substituting $P(d_i \mid \alpha = \alpha_j)$ for $P(d_i)$ in equations 1, 2 and 3, the values of each utility variable conditional upon test α having result α_j may be calculated. For example, the index of general certainty conditional upon $\alpha = \alpha_j$, $C \mid \alpha = \alpha_j$, may be calculated as

$$(6) \qquad C \mid \alpha = \alpha_j = \frac{n}{n-1} \sum_i \left[(P(d_i \mid \alpha = \alpha_j) - \frac{1}{n})^2 \right].$$

Further, since $P(\alpha = \alpha_j)$ is known, it is possible to calculate the expected value of the variables E, C, S for each test. For example, the expected certainty associated with test α, $E(C \mid \alpha)$, is

$$(7) \qquad E(C \mid \alpha) = \sum_j \left[P(\alpha = \alpha_j) \cdot C \mid \alpha = \alpha_j \right]$$

$$\sum_j \left\{ P(\alpha = \alpha_j) \cdot \frac{n}{n-1} \sum_i \left[(P(d_i \mid \alpha = \alpha_j) - \frac{1}{n})^2 \right] \right\}$$

After making such calculations for E, C and S for all tests, it will be clear that some tests are associated with large expected increases in the physician's ability to choose effective treatment while others may be associated with large expected increases in the physician's certainty that the patient has a serious disease. Physicians who prefer tests associated with large increases in one index may be said to consider that index to be more important than the others. If his preferences are well-behaved, it may be possible to sketch the physician's utility function on a map of additional units of E, C and S per unit of physician time.

IV. The Diagnosis of Backache and Dysuria

This section considers the application of the model to the diagnostic process of Dr. Donald L. Crombie.[6] The simulated patient is assumed to be a 45 year old male who expects to live to 70 and who has a backache and at least a vague dysuria. Discussion of Dr. Crombie's preferences in the first stage of the diagnosis--the selection of the first diagnostic procedure--is followed by an evaluation of the simulation.

A. Selection of the First Diagnostic Process

Judging from his own practice and assuming no information other than knowledge of the pain and dysuria, Dr. Crombie interpreted his initial hypothesis about the patient's condition in the required probabilistic form, Table 1. Using that list of probabilities and the estimated days saved given each disease and treatment, the expected days saved by each treatment was calculated according to equation 1, Table 2. The negative numbers suggest that the net effect of many incorrect treatments is a loss of normal days. If the physician was forced to make a decision at this initial stage, the model indicates that he would almost be indifferent between Furadantin or Furadantin with restricted ambulation, both of which would be expected to add 26 days to

Table 1

Dr. Crombie's Initial Hypothesis

Patient: Male; 45 years of age

Presenting Symptom: Backache and Dysuria

Disease Probability (percent)	Diseases
44.00	Distorted Disc with spontaneous recovery
.40	Distorted Disc requiring operative interference, hip spica and/or fusion
7.00	Torn muscle and/or ligament with no bone injury
.40	Torn muscle and/or ligament with wedge fracture
.40	Torn muscle and/or ligament with fracture of a transverse process
39.94	Fibrositis
1.00	Spondylitis (acute exacerbation)
1.00	Spondylitis (chronic)
.40	Secondary carcinoma of vertebra (only secondary manifestation and/or pain warrants local radiotherapy)
.40	Secondary carcinoma of vertebra (analgesics)
.10	Myelomatosis
.10	Other bone disease
.60	Pyelitis (no stone)
.01	Pyelo-nephritis with renal failure (no stone)
.04	Pyelo-nephritis without renal failure (no stone)
.07	Pyelocystitis secondary to primary urethritis
1.00	Small discrete renal calculus without infection
.60	Small discrete renal calculus with infection
.20	Large "stag-horn" calculus without infection
.01	Large "stag-horn" calculus with infection and renal failure
.06	Large "stag-horn" calculus with infection and without renal failure

Table 1 (cont'd.)

Disease Probability (percent)	Diseases
.13	Hydronephrosis secondary to calculus
.50	Hydronephrosis secondary to previous chronic pyelitis with renal failure
.50	Hydronephrosis secondary to previous chronic pyelitis without renal failure
.01	Congenital repairable by surgery
.01	Congenital not repairable by surgery with renal failure
.02	Congenital not repairable by surgery without renal failure
.70	Trauma with kidney viable
.10	Trauma with kidney nonviable
.02	Renal infarct with renal failure
.08	Renal infarct without renal failure
.02	Thrombo-phlebitis with renal failure
.08	Thrombo-phlebitis without renal failure
.02	Arterial thrombosis with renal failure
.08	Arterial thrombosis without renal failure

Table 2

Expected Days Saved for Each Treatment
Given the Initial Hypothesis

Expected Days Saved	Treatment
-5.82	Active ambulation
-2.77	Restricted ambulation (including bedboards as appropriate)
-2.70	Simple analgesics (such as aspirin)
-4.00	Simple analgesics with active ambulation
-2.17	Simple analgesics with restricted ambulation
-2.70	Intermediate analgesics (such as diconal)
1.60	Intermediate analgesics with restricted ambulation
-393.47	Powerful analgesics (such as morphine) with restricted ambulation
26.43	Furadantin or other appropriate antibiotic
26.47	Furadantin with restricted ambulation
-.76	Surgical removal of stone
17.26	Surgical removal of stone with appropriate antibiotic
-323.65	Surgical removal of kidney
-317.94	Surgical removal of kidney with appropriate antibiotic
-662.77	Renal dialysis
-1017.89	Renal transplant
-108.50	Hip spica
-130.14	Fusion
-715.03	Radiotherapy with intermediate analgesics
-804.71	Radiotherapy with powerful analgesics
-28.46	Nephropexy or equivalent plastic repair to kidney pelvis

the patient's life. The additional E, C and S per unit of physician's time (Table 3) are calculated as the difference between the expected E, C and S from each test (the expected C from test α is calculated by equation 7) and the initial E, C and S (equations 1, 2, and 3) all divided by the physician's time ($t_{\alpha k}$).

Analysis of the results of the clinical simulations reported in Table 3 indicates that additional units of certainty, C, were totally unimportant to Dr. Crombie. Tests with high values of dC/t were not included in any list of preferences. Choices between tests with essentially identical values for dE/t and dS/t were not affected by differences in dC/t. Thus the following analysis is limited to preferences for dS/t and dE/t.

At the first stage, several tests gave almost identical expected information about dE/t and dS/t. To avoid problems which might be associated with this bunching, Dr. Crombie was first asked which of the bunched tests he would prefer, Tables 4A-4C. Eliminating the less preferred of each of these groups, Dr. Crombie was then asked to rank his preference for those tests with substantial expected information, Table 4D. The results of this ranking suggest the indifference map sketched in Figure 1.

At this initial stage, and for the arbitrary units employed for dE/t and dS/t, Dr. Crombie reveals a preference for information which will allow him to select more effective treatments. Dr. Crombie's utility would be increased more rapidly by tests with substantial expected information about days saved from more appropriate treatments than from certainty about days lost from serious diseases.

B. Evaluation

This section evaluates the strength of the evidence which confirms the hypothesis that physicians have preferences for the various kinds of information expected from diagnostic procedures. First, the reader is undoubtedly

Table 3

Additional E, C and S Per Second of Physician's Time

dE/sec	dC/sec	dS/sec	Diagnostic Procedure
.0	.06	.7	1) Is he limping as he walks in?
1.3	.02	.6	2) Did he have difficulty sitting down?
8.4	.01	2.1	3) Is he obviously ill, shocked and in severe pain?
.0	.05	.3	4) Onset: sudden or gradual?
.3	.02	.2	5) Onset: on bending or lifting?
2.0	.00	.4	6) Pain into groin or testicle?
.0	.00	.0	7) Other severe trauma or blow to area?
.0	.01	.1	8) Presence of sciatic pain?
.0	.02	.2	9) Pain down leg or coughing?
.4	.00	.2	10) Dysuria at onset of micturition?
2.4	.00	2.0	11) Dysuria during micturition?
2.4	.00	2.2	12) Dysuria at end of micturition?
1.5	.00	.9	13) Past history of attacks of severe renal pain?
.0	.01	.1	14) Quality of Pain? (continuous or spasmodic)
.0	.00	.1	15) Headache?
2.0	.00	1.3	16) Vomiting?
1.3	.00	.8	17) Pyrexia?
1.5	.00	.2	18) Local renal tenderness in back?
.0	.00	.1	19) Limitation of movements of back?
.0	.05	.3	20) Straight leg raising? (restricted or normal)
.0	.02	.1	21) Straight X-Ray of disc
2.9	.00	1.0	22) Straight X-Ray of kidney stone

dE/sec	dC/sec	dS/sec		Diagnostic Procedure
.0	.00	.7	23)	Straight X-Ray of back for myelomatosis
.0	.00	.0	24)	Myelography
.2	.00	.1	25)	Intravenous pyelogram (stone, hydronephrosis, both, none)
.2	.00	.1	26)	Retrograde pyelogram (stone, hydronephrosis, both, none)
1.8	.00	2.5	27)	Midstream Urine (MSU) bacteriology and routine sensitivity
2.9	.00	.9	28)	MSU red blood cells present
1.9	.00	2.5	29)	MSU white blood cells present
.0	.00	.1	30)	Electrophoresis
.0	.00	.7	31)	Alkaline Phosphatase
.2	.00	5.5	32)	Straight X-Ray for secondary carcinoma
.2	.00	.2	33)	Straight X-Ray for Spondylitis
.0	.00	.0	34)	Straight X-Ray for wedge fracture
1.0	.00	1.6	35)	Straight X-Ray for hydronephrosis
.0	.00	.0	36)	Straight X-Ray for fracture of transverse process

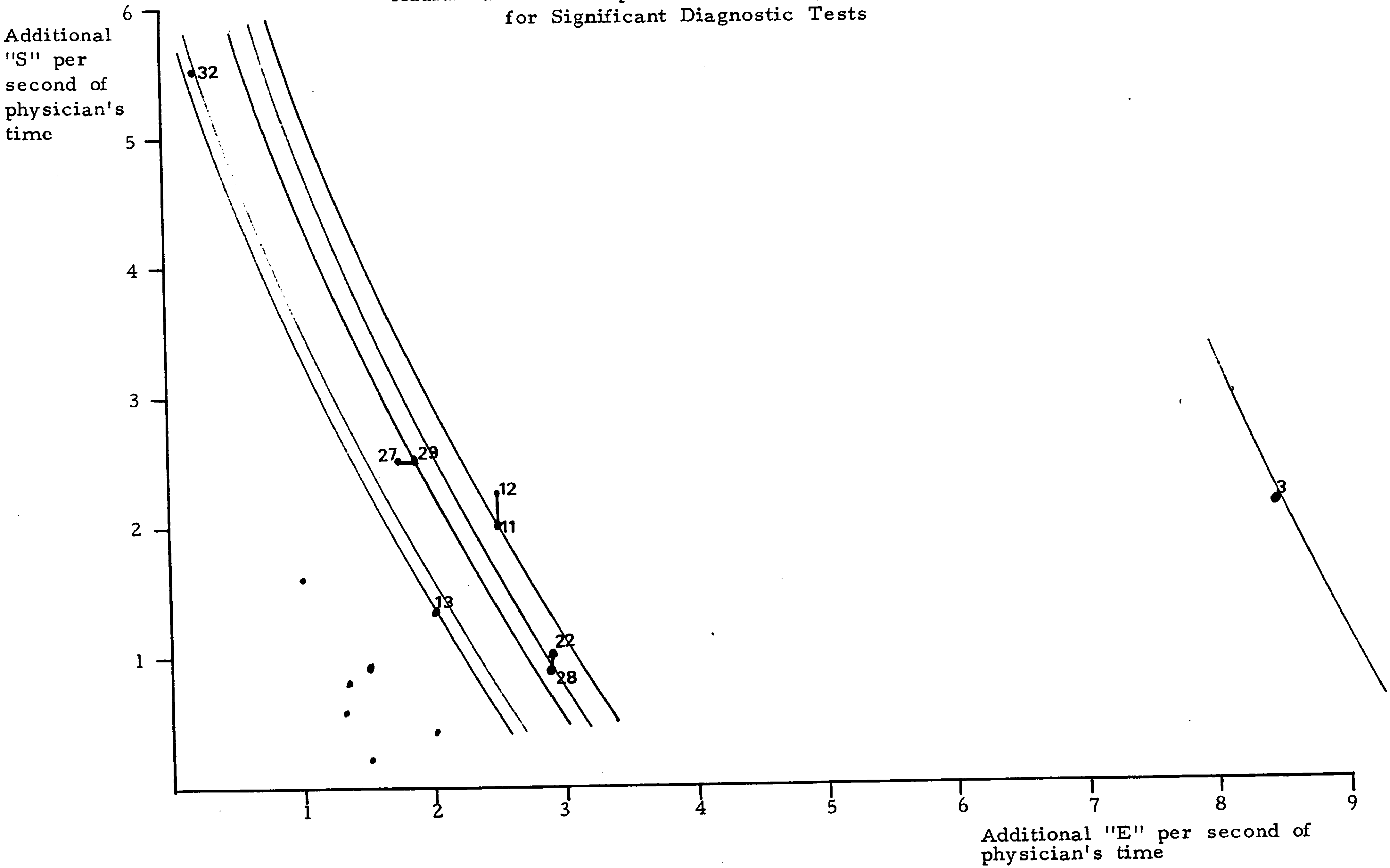
TABLE 4

Diagnostic Preferences

	Diagnostic Process Alternatives	Rank
A.	11	1
	12	2
B.	27	2
	29	1
C.	22	2
	28	1
D.	3	1
	11	2
	13	6
	28	3
	29	4
	32	5

FIGURE 1

Additional E and S per Second of Physicians' Time
for Significant Diagnostic Tests



aware of the limitations of the technique used to elicit Dr. Crombie's preferences. However, when Dr. Crombie was asked to rank his preferences for all diagnostic tests he included: 1) tests which had very low values of expected information (such as whether back movements were limited); or 2) which took a lot of his time relative to the information expected (such as whether the patient had any renal tenderness). Thus only in the situation where Dr. Crombie's alternatives were limited was it possible to demonstrate his preferences.

Although the model is conceptually applicable to any stage of the diagnostic process, the data requirements for a general application to the physicians preferences for the second test are great. In particular, data for all $P(\beta_k|d_i \cap \alpha_j)$--the probabilities of each possible second test outcome (β_k) given each disease (d_i) and each possible first test outcome (α_j)--would be required. Attempts to apply the model to the selection of the second diagnostic procedure assuming the independence indicated by $P(\beta_k|d_i) = P(\beta_k|d_i \cap \alpha_j)$ proved unsuccessful.

Despite these qualifications, the existence of preferences for E, C and S in the diagnostic process appears confirmed (or the null hypothesis rejected) by the utility map derived from Dr. Crombie's priorities among diagnostic procedures. According to his preferences, Dr. Crombie selected the questions about obvious illness (procedure 3), the timing of dysuria (procedure 11), and the mid-stream urine (MSU) sample (procedures 28 and 29) before an X-ray for a secondary carcinoma (procedure 32). When procedures 3, 11, 28 and 29 are compared with 32, Dr. Crombie demonstrates a preference for the additional expected information of 3, 11, 28 and 29 which would allow him to select more effective treatments rather than for the expected information about serious diseases available from 32. A physician whose <u>only</u> concern is to establish whether or not a patient has a serious disease would select the X-ray for the secondary carcinoma (procedure 32), the MSU white blood cells and bacteriology (procedures 27 and 29) and the timing of the dysuria (procedure 12) before Dr. Crombie's most preferred, "Is the patient obviously ill?".

V. Summary

The model of physicians' diagnostic processes presented in this paper establishes the existence of a physician's preferences for alternative forms of diagnostic information. This section considers the importance of the results for clinical behavior and considers policy implications. First, the model is an operational interpretation of the first three of Elstein's four criteria--probability, seriousness, treatability, and novelty--which he suggests (p. 89) physicians use to rank alternative diagnostic hypotheses. Probability, seriousness and treatability correspond, respectively, to the initial hypothesis ($P(d_i)$), the index of seriousness (S), and the index of effective treatment (E) of this model.

Second, preferences are likely to determine, at least in part, each physician's diagnostic routine and his procedures once his routine has been completed. While there is substantial agreement about the necessity of certain diagnostic procedures in every physician's routine, controversy remains about the appropriateness of other tests, especially where there may be some delay or difficulty obtaining the results. The physician's judgment about whether the expected test information is likely to be important may well determine whether he includes the test in his normal routine and which test he selects after he has completed or departed from the routine.

Third, since diagnostic procedures are associated with different kinds of expected information, since physicians are shown to have preferences for these alternative informations, and since there is no reason to believe that their preferences are the same, scholars should expect different diagnostic procedures. Moreover, the search for a standard (or optimal) diagnostic pathway, even for well-defined symptom and disease group, depends upon the definition of the standard (or optimal) preferences. A general agreement among physicians about the relative importance of providing effective treatment versus establishing certainty about the presence of serious diseases, or even the degree of certainty required before treatment decision, seems highly unlikely.

Diagnostic computer programs represent the attempt to increase the efficiency and accuracy of diagnostic procedures. Their efforts to be efficient, that is to reach their conclusions with as little time and cost as possible must be based on: 1) a programmer's evaluation of the relative importance of various kinds of information such as the E, C and S defined in this paper; and 2) the programmer's judgement about which of the constraints--the doctor's time, the patient's time, the pain or discomfort of the patient, the life-threatening risk of the patient, or the dollar cost--is binding. Thus it is unlikely that the diagnostic pathway will be efficient for any physician other than the programmer.

Despite their lack of general efficiency, it is recognized that computer program diagnoses are likely to be at least as accurate as those of senior clinicians. The computer's abilities: 1) to process large bodies of data; 2) to distinguish those tests which do have positive expected information content, and 3) to calculate accurately the implications of test results are unsurpassed by the human mind. Still, the implications of this paper are that a physician should not be surprised if he disagrees with the diagnostic process of the machines. Indeed, if preferences are recognized as important in the diagnostic process, and if computer-aided diagnosis is to become popular among physicians, programmers may well be called upon to custom design their diagnostic routines to fit the preferences of each clinician.

FOOTNOTES

*I am deeply indebted to Dr. Donald Crombie, M.D., for his guidance during the model development and sympathetic and patient cooperation during the data generation. Other valuable comments have been received from Ian McWhinney, MD.; Martin Bass, M.D.; Lionel Reese, M.D.; Bill Comanor, Ph.D.; Kevin Burley, Ph.D.; Peter Chinloy, Ph.D.; Dave Scheffman, Ph.D.; Mike Johnson, Ph.D.; and Katherine Benesch, MPH. An earlier version of this paper was delivered to the Econometric Society on December 28, 1974. The research was funded in part by a National Health Grant from Health and Welfare Canada.

[1]Pathways refer to the ordered procedures, such as the parts of a physical examination or laboratory tests, which are selected.

[2]See Hull for a study which finds that even physicians in the same practice exhibit "great variation between methods of examination" (p. 257).

[3]The Profession Standards Review Organizations have been established under the revisions of the Social Security Act (Public Law No. 92-603) "in order to promote the effective, efficient, and economical delivery of health care services of proper quality for which payment may be made (in whole or in part) under this Act..." Sec. 1151. In short, in anticipation of a national health insurance program, the United States Government is making every effort to determine appropriate levels of care. These levels may be used to limit the expected growth of health expenditures when the national health insurance program is initiated.

[4]See Horrocks, et al., and Dombal, et al., for reports of successful diagnostic computer programs.

[5]For example, see Berdit, M., and Williamson, J., who measure changes in health of patients in a therapeutic program by evaluating the status of each patient on a scale from 1 to 6.

[6]Dr. Donald L. Crombie, O.B.E., F.R.C.G.P., and Director of the General Practice Research Unit of the Royal College of General Practitioners, has numerous publications in the areas of diagnostic process, medical care and medical education. His interests in analyses of the diagnostic process, his willingness to devote considerable time from his schedule, and his patience with the problems caused by my lack of medical training made him an ideal source of data for the simulation. I am deeply indebted to him.

References

M. Berdit and J. Williamson. "Function Limitation Scale for Measuring Health Outcomes," in Health Status Indexes, R. Berg (ed.), Chicago 1973.

J. T. Chu. "Some Decision Making Methods Applicable to the Medical Sciences," Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume IV, Berkeley 1967, 925-34.

D. L. Crombie. "Diagnostic Process," Journal of the College of General Practitioners, Nov. 1963, 6, 579-89.

de Dombal, et al. "Computer-aided Diagnosis of Acute Abdominal Pain," British Medical Journal, April 1972, 9-13.

A. S. Elstein, et al. "Methods and Theory in the Study of Medical Inquiry," Journal of Medical Education, Feb. 1972, 47, 85-92.

M. Grossman. The Demand for Health: A Theoretical and Empirical Investigation, N.B.E.R. Occasional Paper No. 119, New York 1972.

D. Gustafson, et al. "Initial Evaluation of a Subjective Bayesian Diagnostic System," Health Services Research, Fall 1971, 6, 205-13.

J. Horrocks, et al. "Computer-aided Diagnosis: Description of an Adaptable System, and Operational Experience with 2,034 Cases," British Medical Journal, April 1972, 5-9.

J. G. R. Howie, et al. "Respiratory Illness and Antibiotic Use in General Practice," Journal of the Royal College of General Practitioners, Nov. 1971, 21, 657-63.

F. M. Hull. "Diagnostic Pathways in General Practice," Journal of the Royal College of General Practitioners, April 1972, 22, 241-58.

B. Kleinmuntz. "The Processing of Clinical Information by Man and Machine," in B. Kleinmuntz (ed.), Formal Representation of Human Judgment, New York 1968.

M. L. Lee. "A Conspicuous Production Theory of Hospital Behavior," Southern Economic Journal, July 1971, 38, 48-59.

L. B. Lusted. "Logical Analysis in Medical Diagnosis," Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume IV, Berkeley 1967, 903-23.

__________. Introduction to Medical Decision Making, Springfield 1968.

I. R. McWhinney. "Problem-Solving and Decision-Making in Primary Medical Practice," Albert Wander Lecture delivered to the Royal Society of Medicine, mimeo 1972.

J. Newhouse. "Toward a Theory of Nonprofit Institutions: An Economic Model of a Hospital," American Economic Review, March 1970, 60, 64-74.

M. Pauly and M. Redisch. "The Not-For-Profit Hospital as a Physician's Cooperative," American Economic Review, March 1973, 63, 87-89.

G. Torrance, et al. "A Utility Maximization Model for Evaluation of Health Care Programs," Health Services Research, Summer 1972, 7, 118-33.

________________. "Utility Maximization Model for Program Evaluation: A Demonstration Application," in Health Status Indexes, R. Berg (ed.), Chicago 1973.