Application of step-wise discriminant analysis and Bayesian classification procedure in determining prognosis of acute myocardial infarction

K.S. BAY, PH D; S.J.K. LEE, MD, FRCP[C]; D.P. FLATHMAN, PH D; J.W. ROLL, B SC; W. PIERCY, RN

A retrospective study was carried out to assess the feasibility of computer-assisted prognostication by discriminant analysis and the Bayesian classification procedure based on clinical information collected on patients with acute myocardial infarction. The overall accuracy was 94.2% in predicting hospital death but the prediction of late death after discharge was less accurate. It was found that not all of the 44 variables used for analysis were necessary to reach the same level of predictive accuracy — 16 to 20 variables would result in almost the identical prediction. The Bayesian classification procedure was applied to estimate probabilities of individual patients belonging to the different prognostic categories.

Une étude rétrospective a été réalisée afin d'évaluer la possibilité d'établir un système de pronostic par ordinateur s'appuyant sur une analyse différentielle et la méthode Bayesienne de classification de l'information clinique obtenue de patients atteints d'infarctus du myocarde. La précision globale concernant les prévisions de décès à l'hôpital a été de 94.2% mais on a été moins précis en ce qui a trait à la prédiction du décès tardif après le congé de l'hôpital. On a trouvé que les 44 variables utilisées n'étaient pas toutes nécessaires pour atteindre ce niveau de précision: 16 à 20 variables suffiraient à obtenir une prédiction identique. La méthode de classification Bayesienne a été appliquée dans le but d'établir les probabilités que les patients appartiennent à chacune des différentes catégories de pronostic.

From the division of health services administration and the department of medicine, University of Alberta

Reprint requests to: Dr. S.J.K. Lee, 6-125 Clinical Sciences Building, University of Alberta, Edmonton, Alta. T6G 2G3

This paper reports the results of a retrospective study undertaken to investigate the feasibility of computer-assisted prognostication following acute myocardial infarction. The basis for prediction of outcome was a set of "predictor" variables selected from all clinical information collected routinely in the coronary care unit (CCU).

The early prognosis following acute myocardial infarction depends on many variables, including age, sex, previous history and laboratory findings, according to several studies,[1-6] but these studies were carried out before the era of the modern CCU, where patients are monitored continuously, at least for a few days. It was postulated that incorporation of such detailed information with computer technology could yield a more accurate prognosis. Another objective of this study was to examine the accuracy of late prognosis, which was not done in the previous studies.

Estimation of risk of death from acute myocardial infarction is not new. Peel and colleagues[7] proposed a coronary prognosis index that has been widely used. Hughes and associates,[5] applying a linear discriminant analysis, established a prognosis rating for 22 variables. In these and other investigations[8,9] the early prognosis was studied in patients who were not monitored continuously in a CCU. Recently, prognostic values of hemodynamic measurements in myocardial infarction were discussed by Marx and Yu.[10] To the best of our knowledge, no study has been done to determine late prognosis by discriminant analysis, nor has the Bayesian classification procedure been used to study the prognosis of myocardial infarction.

Methods

The population studied consisted of 410 patients treated for proven acute myocardial infarction in the CCU of the University of Alberta Hospital between January 1969 and January 1971. Late follow-up to determine survival was obtained in January 1973 by corresponding with the patients or their physicians or both.

All patient charts were reviewed and information was recovered on 96 variables of clinical data (including history related to cardiovascular disease and physical findings) and laboratory data collected during the stay in the CCU. After initial data editing and evaluation 30 records were found to be unusable owing to incompleteness of information; information on all variables used in the analysis had to be complete for each subject. Excluding cases with incomplete information may have introduced bias since this procedure tended to exclude cases in which the patient died shortly after admission; thus, the hospital mortality was changed from 16.1% to 13.7% after the 30 cases were excluded.

Among the 96 variables 62 were assumed to have some prediction potential, while the remaining variables were related to patient identification, date of hospital admission and outcome (Appendix A). After preliminary analysis some of the original variables were combined, creating a new set of variables (Appendix B). Thus, the number of predictors was reduced to 44.

The 380 usable records were grouped into the following classifications based on outcome at follow-up:

A. Hospital death (died in CCU or ward).

B. Late death (died after discharge from hospital).

C. Survived (alive at follow-up).

D. Unknown (discharged alive from hospital but fate could not be determined at follow-up).

With the 44 predictor variables and outcome information, the outcome following acute myocardial infarction for each patient was predicted by discriminant analysis and the Bayesian classification procedure; these methods are described in Appendix C.

Three analyses of the data were performed with different prognosis groups:

- Analysis I. Two groups: hospital death and alive at discharge (A and B + C + D).
- Analysis II. Three groups: hospital death, late death and alive at follow-up (A, B and C).
- Analysis III. Two groups: late death and alive at follow-up (B and C).

Of the 410 patients in the original study population, data for 380, 339 and 287 were included in analyses I,

II and III, respectively. Characteristics of the patients are summarized in Table I.

Results

Lists of discriminating predictor variables are given in Tables II to IV; shown are the order of entry to the discriminant functions from the most significant to the least significant and the improvement in prediction with successive addition of predictor variables. The F-value (or equivalent t^2 for analyses I and III) associated with the entry of each variable is also reported to provide a measure of the contribution of each variable. For those familiar with multivariate analysis, Wilks' lambda statistics are also included, together with the prediction rate at each stage.

For example, for the prediction of hospital death or being discharged alive (Table II), ventricular fibrillation, recurrent pain in the CCU, peak lactic dehydrogenase (LDH) value, age at admission and heart failure in the CCU were found to be the most discriminating variables. These five variables were also the most discriminating for analysis II, although in a different order (Table III). However, for analysis III (Table IV), which was concerned with late death, diabetes and ventricular tachycardia were found to be more important than ventricular fibrillation and peak LDH value.

Results of the step-wise discriminant analysis and Bayesian classification procedure are summarized in Tables V to VII. The rates of correct prediction achieved with all 44 variables were 94.2%, 82.9% and 86.8% for analyses I, II and III, respectively. The prediction rate for late death was not as accurate as that for hospital death. For example, in analysis II (Table VI), 40 of 52 hospital deaths (77%) were correctly predicted, while in analysis III (Table VII) 29 of 56 late deaths (52%) were correctly predicted.

Discussion

Hughes and colleagues[5] discussed the controversy over factors affecting death from acute myocardial infarction. Their study showed age, leukocyte count, temperature, systolic blood pressure, conduction defects, pulmonary infarction, congestive heart failure and shock to be closely related to mortality. The Framingham study[22] also found sex, serum cholesterol concentration, smoking history and vital capacity to be related to mortality. Weinblatt, Shapiro and Frank[23] added angina and diabetes to the list. A report from the World Health Organization[24] noted a strong relation between previous transient cerebral ischemia and death from myocardial infarction.

Unlike those studies our study also included variables collected during continuous observation in the CCU, such as ventricular fibrillation, recurrent chest pain, peak LDH, lowest heart rate and bundle branch block. However, the subsequent classification analysis indicated that a set of about 20 variables was optimum; the addition of further variables did not improve the prediction accuracy appreciably.

The patient information was used

Table I—Characteristics of groups of patients with myocardial infarction (MI)

Variable	No. of patients			
		Analysis group		
	Total population (n = 410)	I (n = 380)	II (n = 339)	III (n = 287)
Outcome (cases)				
Hospital death	66	52	52	—
Late death	60	56	56	56
Survived	243	231	231	231
Unknown	41	41	—	—
Sex				
Male	338	320	285	246
Female	72	60	54	41
Admission for MI				
First	385	355	315	270
Second	23	23	22	15
Third	2	2	2	2
Mean age (yr)				
Total population	60.4	60.1	60.3	59.0
By outcome				
Hospital death	67.8	67.8	67.8	—
Late death	63.8	63.8	63.8	63.8
Survived	57.8	57.8	57.8	57.8
Unknown	58.6	58.6	—	—
By sex				
Male	58.9	58.7	58.9	57.7
Female	67.6	67.8	67.6	66.5

Table II—Results of step-wise discriminant analysis I*

Step no.	Variable entered		F-value (t^2)	Wilks' lambda	Rate of correct prediction (%)
	No.	Description†			
01	58	Ventricular fibrillation	78.76	0.837	88.4
02	11	Recurrent pain in CCU	49.44	0.732	88.4
03	03	Peak LDH value	27.54	0.682	88.4
04	01	Age at admission	28.19	0.634	89.7
05	13	Heart failure in CCU	11.66	0.615	90.0
06	06	Lowest heart rate	7.90	0.602	91.1
07	27	IVCD	7.36	0.590	91.1
08	12	Extension of MI	5.871	0.581	91.3
09	72	RBBB & LPHB	4.183	0.575	91.1
10	63	Acute MI	3.583	0.569	91.6
11	19	Acute anterior MI	2.982	0.565	91.2
12	09	Low systolic BP	2.309	0.561	91.6
13	15	Hypertension (history)	2.971	0.557	92.1
14	50	Second-degree AV block (Wenckebach)	1.963	0.554	92.6
15	74	Third-degree heart block	1.931	0.551	92.4
16	70	RBBB	1.804	0.548	93.2
17	57	Ventricular tachycardia	1.604	0.546	93.2
18	08	High systolic BP	1.248	0.544	93.4
19	68	LAHB	1.199	0.542	93.4
20	17	Previous angina	1.155	0.540	92.6
21	18	Previous MI	2.638	0.536	93.4
22	69	LPHB	1.119	0.535	93.4
23	07	Peak temperature	0.900	0.533	93.4
24	59	Wandering pacemaker	1.037	0.532	93.7

*Data for two outcome groups were studied: hospital death (A) and alive at discharge (B + C + D).
†CCU = coronary care unit; LDH = lactic dehydrogenase; IVCD = intraventricular conduction defect; RBBB = right bundle branch block; LPHB = left posterior hemiblock; BP= blood pressure; AV = atrioventricular; LAHB = left anterior hemiblock.

Table III—Results of step-wise discriminant analysis II*

Step no.	Variable entered No.	Variable entered Description†	F-value	Wilks' lambda	Rate of correct prediction (%)
01	58	Ventricular fibrillation	45.24	0.788	72.6
02	11	Recurrent pain in CCU	25.55	0.684	72.6
03	01	Age at admission	22.61	0.602	73.7
04	03	Peak LDH value	15.44	0.551	74.9
05	13	Heart failure in CCU	10.69	0.518	76.1
06	12	Extension of MI	4.31	0.505	76.1
07	57	Ventricular tachycardia	4.27	0.492	75.2
08	16	Diabetes	3.60	0.481	76.7
09	27	IVCD	3.00	0.473	77.6
10	06	Lowest heart rate	2.42	0.466	78.5
11	19	Acute anterior MI	2.15	0.460	78.5
12	63	Acute MI	3.54	0.450	79.1
13	72	RBBB & LPHB	2.39	0.443	79.4
14	73	Atrial fibrillation or flutter	2.26	0.437	80.2
15	15	Hypertension (history)	2.20	0.431	79.1
16	70	RBBB	1.91	0.426	79.9
17	18	Previous MI	1.65	0.422	81.4
18	69	LPHB	1.65	0.418	81.7
19	09	Low systolic BP	1.50	0.414	82.0
20	08	High systolic BP	1.67	0.409	81.7
21	07	Peak temperature	1.51	0.405	81.7
22	17	Previous angina	1.55	0.402	82.0
23	51	Second-degree AV block (type II)	1.30	0.398	82.0
24	49	First-degree AV block	1.29	0.395	81.7
25	67	LBBB	1.50	0.381	81.7
26	71	RBBB & LAHB	1.42	0.388	82.0
27	66	Acute MI, site unknown	1.26	0.385	82.0
28	55	PVCs (> 5/min)	1.17	0.382	82.0
29	74	Third-degree heart block	1.00	0.379	82.0
30	14	Cardiac arrest before admission	0.863	0.377	82.9

*Data for three outcome groups were studied: hospital death (A), late death (B) and alive at follow-up (C).
†LBBB = left bundle branch block; PVC = premature ventricular contraction; and see footnote to Table II.

Table IV—Results of step-wise discriminant analysis III*

Step no.	Variable entered No.	Variable entered Description†	F-value (t^2)	Wilks' lambda	Rate of correct prediction (%)
01	13	Heart failure in CCU	17.19	0.943	80.5
02	01	Age at admission	9.22	0.914	80.5
03	16	Diabetes	7.30	0.891	79.8
04	57	Ventricular tachycardia	5.46	0.874	81.9
05	11	Recurrent pain in CCU	4.97	0.858	81.9
06	68	LAHB	3.91	0.847	81.2
07	19	Acute anterior MI	4.14	0.834	83.3
08	67	LBBB	3.08	0.825	83.6
09	64	Old MI	3.83	0.815	82.9
10	58	Ventricular fibrillation	3.58	0.805	82.6
11	51	Second-degree AV block (type II)	3.36	0.795	82.6
12	49	First-degree AV block	4.47	0.782	83.3
13	66	Acute MI, site unknown	2.70	0.775	83.6
14	74	Third-degree heart block	3.09	0.766	82.6
15	71	RBBB & LAHB	2.42	0.759	83.6
16	50	Second-degree AV block (Wenckebach)	2.35	0.753	83.6
17	73	Atrial fibrillation or flutter	2.17	0.747	83.6
18	70	RBBB	1.83	0.742	85.0
19	09	Low systolic BP	1.47	0.738	84.7
20	08	High systolic BP	2.25	0.731	84.3
21	03	Peak LDH	1.99	0.726	85.4
22	05	Peak heart rate	1.74	0.721	84.7
23	15	Hypertension (history)	1.20	0.718	86.1
24	63	Acute MI	1.51	0.714	86.4
25	59	Wandering pacemaker	0.69	0.712	86.4
26	07	Peak temperature	0.54	0.710	86.4
27	14	Cardiac arrest before admission	0.53	0.709	86.4
28	06	Lowest heart rate	0.35	0.709	85.7

*Data for two outcome groups were studied: late death (B) and alive at follow-up (C).
†See footnotes to Tables II and III.

twice in this study: first, in deriving discriminant functions and prior probabilities, each patient being a member of the study population; and, second, in the classification procedure, each patient being a subject to be classified. Therefore, it may be argued that the accuracy rate for independent subjects would not be as high as the results might suggest. However, the contribution of any individual patient's record to discriminant analysis within a data set of 380 records should be negligible and should not affect the results appreciably.

An example of application of the Bayesian classification method to analysis II is shown in Table VIII. The average probability of belonging to each of the three categories A, B and C was estimated as 0.15, 0.17 and 0.68, respectively, based on the mortality for 339 patients. Patient X's medical record indicated that his chances of having such medical information if he belonged to group A, B or C were 0.029, 0.916 and 0.443, respectively. These are the so-called conditional probabilities, or "likelihoods", obtained using distributional assumptions such as the chi-square distribution. From these two types of probabilities the posterior probability, or the probability of belonging to each group after we know the patient's condition, was obtained. Patient X's posterior probabilities were 0.009, 0.338 and 0.653, respectively; one may conclude that his chance for late survival was high and one may classify him into group C. In fact, this patient survived more than 1 year after discharge.

Another patient, Y, resembled group B patients, with a conditional probability of 0.293 and a posterior probability of 0.460 of dying within 1 year. Thus, one may classify patient Y into the late-death group (group B). In fact, patient Y died in hospital (group A); this was a prediction error. However, this patient's combined posterior probability of belonging to group A or B was very high (0.843); hence, Y was at high risk of dying.

One possible application of the results of this study is to store the discriminant functions and other statistics in an interactive computer program that could provide instant posterior probabilities for new patients to aid physicians in their decisions about length of stay in the CCU or the hospital. For instance, the physician might be able to discharge patients with a lower posterior probability of hospital or late death sooner from the CCU or the hospital, while keeping those at high risk longer.

Another application of Bayesian classification is in identifying the patient

Table V—Actual v. predicted outcome, analysis I

Predicted outcome	Actual outcome: Hospital death A	Alive at discharge (B + C + D)	Total (A + B + C + D)
Hospital death	38	8	46
Alive at discharge	14	320	334
Total	52	328	380

Rate of correct prediction: $\frac{38 + 320}{380} = 0.942$

Sensitivity: $38/52 = 0.731$

Specificity: $320/328 = 0.976$

Table VI—Actual v. predicted outcome, analysis II

Predicted outcome	Actual outcome: Hospital death (A)	Late death (B)	Alive at follow-up (C)	Total (A + B + C)	Fate unknown (D)
Hospital death	40	5	2	47	6
Late death	3	21	9	33	3
Alive at follow-up	9	30	220	259	32
Total	52	56	231	339	41

Rate of correct prediction: $\frac{40 + 21 + 220}{52 + 56 + 231} = \frac{281}{339} = 0.829$

Table VII—Actual v. predicted outcome, analysis III

Predicted outcome	Actual outcome: Late death (B)	Alive at follow-up (C)	Total (B + C)
Late death	29	11	40
Alive at follow-up	27	220	247
Total	56	231	287

Rate of correct prediction: $\frac{29 + 220}{287} = 0.868$

Sensitivity: $29/56 = 0.518$

Specificity: $220/231 = 0.952$

Table VIII—Example of Bayesian classification

Subjects	Hospital death (A)	Death within 1 year (B)	Alive at 1 year (C)	Total
Target population				
No. of patients	52	56	231	339
Prior probability	0.15	0.17	0.68	1.00
Patient X				
Probability				
Conditional	0.029	0.916	0.443	
Posterior	0.009	0.338	0.653	1.00
Classification			X	
Actual fate			X	
Patient Y				
Probability				
Conditional	0.277	0.293	0.025	
Posterior	0.383	0.460	0.157	1.00
Classification		X		
Actual fate	X			

with a high probability of late death, thus enabling the physician to introduce preventive measures after the patient is discharged from hospital.

The major shortcoming of this study is that this analysis was based on retrospective studies, which may have biased the data collection. Thus, the true accuracy of prediction of outcome needs to be tested by a prospective study. Nevertheless, the results of this study support the contention that the accuracy of prediction can be improved by stepwise discriminant analysis and computer-assisted analysis of patient information, including data collected in the CCU. It is possible that incorporation of this classification method into clinical practice may result in improved patient care by reducing mortality and increasing cost-effectiveness, but this hypothesis needs to be tested by properly designed clinical trials. The methods used in this study are not specific for myocardial infarction but can be applied to parallel clinical situations — for example, predicting death among patients with angina, based on clinical and coronary angiographic information.

This work was supported in part by a grant-in-aid of the Alberta Heart Foundation and under national health research and development project no. 609-1041-22 of Health and Welfare Canada.

References

1. Rosenbaum FF, Levine SA: Prognostic value of various clinical and electrocardiographic features of acute myocardial infarction; immediate prognosis. *Arch Intern Med* 68: 913, 1941
2. Honey GE, Truelove SC: Prognostic factors in myocardial infarction. *Lancet* 1: 1155, 1957
3. Beard OW, Hipp HR, Rabins M, et al: Initial myocardial infarction among 503 veterans: five-year survival. *Am J Med* 28: 871, 1960
4. Pell S, D'Alonzo CA: Immediate mortality and five-year survival of employed men with a first myocardial infarction. *N Engl J Med* 270: 915, 1964
5. Hughes WL, Ralbfleisch JM, Brandt EN jr, et al: Myocardial infarction prognosis by discriminant analysis. *Arch Intern Med* 111: 120, 1963
6. Takkunen JH, Tuomaala R, Waris E, et al: Factors in influencing survival after myocardial infarction. *Ann Clin Res* 1: 246, 1969
7. Peel AAF, Semple T, Wand I, et al: A coronary prognostic index for grading the severity of infarction. *Br Heart J* 24: 745, 1962
8. Lemlich A, Covo G, Ziffer H: Multivariant analysis of prognostic factors in myocardial infarction, in *Rochester Conference on Data Acquisition and Processing in Biology and Medicine,* vol 3, New York, Pergamon, 1963, p 65
9. Norris RM, Brandt PWT, Caughey DE, et al: A new coronary prognostic index. *Lancet* 1: 274, 1969
10. Marx JG, Yu PN: Prognosis factors in acute myocardial infarction, in *Myocardial Infarction,* Corday E, Swan HJC (eds), Baltimore, Williams & Wilkins, 1973, pp 195-204
11. Rao CR: *Advanced Statistical Methods in Biometric Research,* New York, Wiley, 1952, pp 273-338
12. Idem: *Linear Statistical Inference and its Application,* New York, Wiley, 1965, pp 487-94
13. Cooley WW, Lohnes PR: *Multivariate Data Analysis,* Toronto, Wiley, 1971, pp 243-86
14. Bay KS, Flathman DP: A patient classification system: an application of discriminant analysis and Bayesian classification procedures to cardiac patients (1973; unpublished; available on request)
15. Draper NR, Smith H: *Applied Regression Analysis,* New York, Wiley, 1966, pp 171-72

Appendix A—Original predictor variables

Var. no.	Description	Var. no.	Description	Var. no.	Description
01	Age at admission	22	Old lateral MI	43	RBBB & LPHB, recent
02	Peak serum glutamic oxaloacetic transaminase value	23	Acute inferior MI	44	RBBB & LPHB, transient
03	Peak lactic dehydrogenase value	24	Old inferior MI	45	RBBB & LPHB, old
04	Leukocyte count	25	Acute high lateral MI	46	Atrial fibrillation
05	Peak heart rate	26	Old high lateral MI	47	Atrial flutter
06	Lowest heart rate	27	Intraventricular conduction defect	48	Atrioventricular (AV) nodal rhythm
07	Peak temperature	28	Left bundle branch block (LBBB), recent	49	First-degree AV block (type II)
08	High systolic blood pressure (BP)	29	LBBB, transient	50	Second-degree AV block (Wenckebach)
09	Low systolic BP	30	LBBB, old	51	Second-degree AV block (type II)
10	Sex	31	Left anterior hemiblock (LAHB), recent	52	Third-degree heart block (junctional)
11	Recurrent pain in coronary care unit (CCU)	32	LAHB, transient	53	Third-degree heart block (idioventricular)
12	Extension of myocardial infarction (MI)	33	LAHB, old	54	Accelerated idioventricular rhythm
13	Heart failure in CCU	34	Left posterior hemiblock (LPHB), recent	55	Premature ventricular contractions (PVCs) (> 5/min)
14	Cardiac arrest before admission	35	LPHB, transient	56	Ventricular bigeminy (PVCs)
15	Hypertension (history)	36	LPHB, old	57	Ventricular tachycardia
16	Diabetes (history)	37	Right bundle branch block (RBBB), recent	58	Ventricular fibrillation
17	Previous angina	38	RBBB, transient	59	Wandering pacemaker
18	Previous MI	39	RBBB, old	60	Artificial pacing
19	Acute anterior MI (transmural or subendocardial)	40	RBBB & LAHB, recent	61	Premature atrial contractions
20	Old anterior MI	41	RBBB & LAHB, transient	62	Paroxysmal atrial tachycardia
21	Acute lateral MI	42	RBBB & LAHB, old		

Appendix B—Created predictor variables

Var. no.	Original variables	Description	Coding procedure
63	19, 21, 23, 25	Acute MI	Total no. of yes's (0 - 4)
64	20, 22, 24, 26	Old MI	Total no. of yes's (1 - 4)
65	21, 25	Acute lateral MI	(1) At least one yes (0) Otherwise
66	19, 21, 23, 25	Acute MI, site unknown	(1) Site unknown (0) Otherwise
67	28 - 30	LBBB	(1) At least one yes (0) Otherwise
68	31 - 33	LAHB	(1) At least one yes (0) Otherwise
69	34 - 36	LPHB	(1) At least 1 yes (0) Otherwise
70	37 - 39	RBBB	(1) At least 1 yes (0) Otherwise
71	40 - 42	RBBB & LAHB	(1) At least 1 yes (0) Otherwise
72	43 - 45	RBBB & LPHB	(1) At least 1 yes (0) Otherwise
73	46, 47	Atrial fibrillation or flutter	(1) At least 1 yes (0) Otherwise
74	52, 53	Third-degree heart block	(1) At least 1 yes (0) Otherwise
75	46 - 62	Arrhythmias	Total no. of yes's (0 - 17)

16. Bay KS: Application of multivariate analysis of variance to educational and psychological research: hypothesis testing by MANOVA programs. *Multivariate Behav Res* 6: 325, 1971
17. Dixon WJ (ed): *BMD: Biomedical Computer Programs,* Berkeley, U of Cal Pr, 1973, pp 233-53
18. Nie NJ, Hull CH, Jenkins JG, et al: *SPSS: Statistical Package for Social Sciences,* 2nd ed, New York, McGraw, 1975, pp 9-10, 435-67
19. McCabe GP jr: Computations for variable selection in discriminant analysis. *Technometrics* 12: 103, 1975
20. Fisz M: *Probability Theory and Mathematical Statistics,* New York, Wiley, 1963, p 23
21. McNeil BJ, Keeler E, Avelstein JA: Primer on certain elements of medical decision making. *N Engl J Med* 293: 211, 1975
22. Kannel WB, Dawber TR, Cohen ME, et al: Vascular disease of the brain — epidemiologic aspects: the Framingham study. *Am J Public Health* 55: 1355, 1965
23. Weinblatt E, Shapiro S, Frank CW: Prognosis of women with newly diagnosed coronary heart disease — a comparison with course of disease among men. *Am J Public Health* 63: 577, 1973
24. World Health Organization: Cerebrovascular disease: prevention, treatment and rehabilitation. *Tech Rep Ser* no 469: 13, 1971

Appendix C — Discriminant analysis and Bayesian classification

In this paper a step-wise discriminant analysis and the Bayesian classification procedure were used to establish a set of predictor variables and to test the accuracy of prediction. The mathematical aspects of this type of analysis were presented by Rao,[11,12] and application to behavioural science problems was discussed by Cooley and Lohnes.[13] Hughes and colleagues[5] and Bay and Flathman[14] discussed in detail the use of discriminant analysis as a tool for clinical prognosis.

Discriminant analysis is a multivariate statistical technique that can aid in distinguishing two or more groups of study subjects. For this study the subjects were myocardial infarction patients and the groups were defined by categories of outcome. For prediction purposes a set of variables for characteristics of the patients may be compared with the outcomes. Discriminant analysis attempts to do this by estimating one or more sets of linear coefficients or weights applicable to each of the predictor variables. Thus, discriminant analysis transforms the raw predictor variables into a smaller number of discriminant functions that maximize separation among the groups. The maximum number of functions that can be derived is usually one less than the number of groups. For the present study there were three categories of patients by group definition; therefore the number of discriminant functions was two.

One major problem in the application of discriminant analysis is determination of what variables should be used for the prediction of outcome. It is desirable to identify the most discriminating among many clinical variables, for some are clearly redundant.

The so-called "step-wise" analysis method has been used widely in regression analysis.[15] Although discriminant analysis uses the general multivariate linear model[16] and is similar to regression analysis, methods of selecting variables are not extensively discussed in the literature. Currently, Biomedical Computer Programs (BMD)[17] and the Statistical Package for Social Sciences[18] are programs that include step-wise discriminant analysis routines. Recently, McCabe[19] proposed an algorithm for all possible subsets of any given size and compared results with the BMD procedure. Although this new algorithm seems to be a great improvement, it requires excessive computer storage and time and appears impractical for studies with more than 20 variables. The present study used a step-wise discriminant analysis program named BMD07M, which is part of the BMD package.[17]

After a set of variables is determined that provides satisfactory discrimination for patients with known group membership, a set of classification functions can be derived that will permit the classification of new patients with unknown outcome. For this purpose the so-called Bayesian theorem or rule may be applied. Mathematical aspects of Bayes' rule may be found in most statistics books,[20] and clinical applications of Bayes' rule have also been discussed.[21]

In its simplest form Bayes' rule may be written as follows:

$$P(H_i|z) = \pi_i P(z|H_i) / \sum_{i=1}^{k} \pi_i$$

$$P(z|H_i) \quad i = 1, 2, \ldots k$$

where $\{\pi_i\}$ are the so-called prior probabilities, $P(z|H_i)$ are the conditional probabilities of obtaining discriminant scores of z or greater if the subject belongs to the ith group, and $P(H_i|z)$ are the so-called posterior probabilities — the probabilities of belonging to each group when the information given by z is taken into account. An example application of this formula is given in the main part of this paper. ■

Long-term therapy of essential tremor with propranolol

T.J. MURRAY, MD

In a double-blind crossover study 12 patients with essential tremor were treated with propranolol and a placebo; 8 improved with propranolol and 3 with the placebo; the degree of improvement with propranolol was greater. In a similar study with diazepam 5 of 12 improved with diazepam and 4 of 12 with the placebo; the degree of improvement was less than that achieved with propranolol.

Response in 21 patients to treatment with propranolol for 2 to 4 years was excellent in 4, good in 4 and fair in 10; the condition of 1 was unchanged and that of 2, worse. Excellent response was maintained for as long as 4 years, but response tended to deteriorate with time if initially it was less than excellent. Response decreased with increasing age. No patient 60 years of age or older had an excellent response, and the four with an excellent response were under age 55, three being under age 35; all four had had their tremor less than 12 years.

Patients with essential tremor should be given a 3-month trial of propranolol at 120 mg/d; if no significant response is seen the dose should be decreased, then the drug discontinued.

Dans une étude à double insu avec chassé-croisé 12 patients souffrant de tremblement essentiel ont reçu du propranolol ou un placebo; 8 se sont améliorés sous propranolol et 3 sous placebo; on a constaté un taux d'amélioration supérieur avec le propranolol. Dans une étude identique, 5 patients sur 12 se sont améliorés avec le diazépam et 4 sur 12 avec le placebo; l'amélioration a été moins marquée qu'avec le propranolol.

Chez 21 patients ayant reçu du propranolol pour des périodes de 2 à 4 ans la réponse a été excellente dans 4, bonne dans 4 et moyenne dans 10; l'état d'un patient est demeuré inchangé, alors qu'il s'est aggravé chez 2 d'entre-eux. Une excellente réponse a pu être maintenue jusqu'à 4 ans, mais une réponse initiale moins qu'excellente avait tendance à se détériorer avec le temps. La réponse a diminué en fonction de l'âge. Aucun patient âgé de 60 ans ou plus n'a eu une excellente réponse, alors que les quatre patients bénéficiant d'une excellente réponse étaient âgés de moins de 55 ans, dont trois de moins de 35 ans; tous quatre souffraient de tremblement depuis moins de 12 ans.

Les patients atteints de tremblement essentiel devraient recevoir un traitement d'essai de 3 mois au propranolol à la dose de 120 mg par jour; si aucune réponse significative ne peut être observée, la dose devrait être diminuée, puis le traitement interrompu.

In 1971 Winkler and Young[1,2] observed that the β-adrenergic blocking agent propranolol was beneficial to patients with essential tremor. In 1972 I reported the findings of a preliminary study of 12 patients treated with propranolol for essential tremor; excellent results were obtained in 6 and lesser improvement in 6.[3] There are now several further reports on propranolol therapy for essential tremor,[1-14] most confirming its value. However, in general, these are anecdotal or represent short-term, often uncontrolled studies. In this paper I report the results of a double-blind study with propranolol, a separate double-blind study using diazepam for comparison, and a long-term follow-up of 21 patients treated with propranolol for 2 to 4 years.

Methods

From 1971 to 1975, 41 patients with essential tremor were assessed for possible inclusion in a study of long-term propranolol therapy. Three were eliminated because of asthma or borderline cardiac output, three because they did not wish to take a medication for a long period and six because of serious concomitant disease. Five patients were given therapy but not studied further because of distance or noncooperation. The remaining 24 patients entered short-term double-blind studies and 21 were followed up while taking propranolol for 2 to 4 years.

Initially a double-blind crossover study was completed for 12 patients, each receiving 120 mg of propranolol daily for 6 weeks and a placebo for 6 weeks. Assessments included examination and evaluation of the tremor by a five-grade system, the patient's subjective evaluation of change, and study of the patient's handwriting and ability to draw an Archimedes spiral. The effect on the tremor when the patient was angry, under stress or in other circumstances of heightened emotion was also evaluated. However, only the evaluation of the tremor proved to be a true measure of response to therapy; hence the complex scoring system was abandoned and only the five-grade evaluation of tremor retained.

Because the value of propranolol could be attributed to its effect on anx-

Presented in part at the Canadian Congress of Neurological Sciences, Saskatoon, June 22, 1974

Reprint requests to: Dr. T.J. Murray, Chief of medicine, Camp Hill Hospital, Halifax, NS B3H 3G2