DANIEL, Wayne Wendell, 1929-
THE USE OF DEMOGRAPHIC CHARACTER-
ISTICS IN PREDICTING LENGTH OF STAY
IN A STATE MENTAL HOSPITAL.

The University of Oklahoma, Ph.D., 1965
Health Sciences, public health

University Microfilms, Inc., Ann Arbor, Michigan

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THE USE OF DEMOGRAPHIC CHARACTERISTICS IN PREDICTING LENGTH OF STAY IN A STATE MENTAL HOSPITAL

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

WAYNE WENDELL DANIEL

Oklahoma City, Oklahoma

1965

THE USE OF DEMOGRAPHIC CHARACTERISTICS IN PREDICTING

LENGTH OF STAY IN A STATE MENTAL HOSPITAL

APPROVED BY

Edward N. Brandt Jr.

William W. Schottstaedt

John G. Driver

J. Robert Assenzo

M. Clinton Miller III

DISSERTATION COMMITTEE

ACKNOWLEDGMENT

I should like to express my deep appreciation to Dr. Edward N. Brandt, Jr., my senior advisor, for suggesting the methodology employed in this dissertation, and for the subsequent guidance and suggestions he provided. I am also grateful to the other members of my committee, Dr. John C. Brixey, Dr. William W. Schottstaedt, Dr. Joseph R. Assenzo, Dr. M. C. Miller, III, and Dr. Joseph B. Goldsmith for the time and effort they have spent in my behalf.

I should like to express my sincere thanks to Dr. Albert J. Glass, Director of the Oklahoma State Department of Mental Health, for granting me permission to use the data of the four Oklahoma mental hospitals. I am very grateful to Mr. J. Paul Costiloe, Director of Biometrics, Oklahoma State Department of Mental Health, for his many valuable suggestions and help in getting the data ready for analysis.

I am grateful to the State of Georgia for the financial support provided me during my period of study and to Dr. Addison M. Duval, Director, Division of Mental Health, Georgia Department of Public Health; Dr. Irville H. MacKinnon, Superintendent; and Dr. James B. Craig, Assistant Superintendent, Milledgeville State Hospital, Milledgeville, Georgia, for granting me a leave of absence to return to school.

Finally, I extend my sincere thanks to Mrs. Rose Titsworth for typing this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

THE USE OF DEMOGRAPHIC CHARACTERISTICS IN PREDICTING LENGTH OF STAY IN A STATE MENTAL HOSPITAL

CHAPTER I

INTRODUCTION AND SUMMARY OF PREVIOUS WORK

Almost 300,000 patients are admitted annually to state mental hospitals in the United States (1). These patients remain in the hospital for varying lengths of time, from a few days to several years. In 1961, of 333,696 patients resident in state and county mental hospitals in twenty states, 149,838, or forty-five per cent had been under the care of a hospital for ten years or more (2).

Anker (3) has said that neuropsychiatric chronicity has become a major problem in mental health. Lindemann, Fairweather, Stone, Smith, and London (4) share this opinion and characterize the increasing number of chronic neuropsychiatric patients as "...perhaps the foremost problem in the field of mental health..."

Pugh (5), looking at the future number of mental hospital patients in light of the "population explosion", does not find future prospects very encouraging. He points out that the increasing size of the population in the fifteen to twenty-four year age bracket in Massachussetts is resulting in an increase of first admissions to public and private mental hospitals in that age group in Massachussetts. He states that in the past, patients admitted to mental hospitals in the fifteen to thirty-four year age group provided the majority of the very long-term patients; and, since this is the age group of the general population that will experience a dramatic growth in the coming decade, mental hospitals face the possibility of increasing numbers of long-term patients.

Therefore, it would seem that some method for identifying potential long-stay patients at the time of their admission to a mental hospital would be a valuable tool in both hospital administration and patient rehabilitation.

The purpose of the present study is to investigate the use of demographic and other routinely recorded characteristics in predicting the length of hospitalization category that will characterize a person admitted to a state mental hospital. The characteristics used in the study are those usually obtained routinely on public mental hospital patients at the time of their admission, or shortly thereafter.

Johnston and McNeal (6) feel that such early identification of potentially chronic patients is important so that staff energies may be focused upon them and appropriate treatment programs started. This feeling is shared by Anker (3), who hopes that early identification could result in new or more intensive treatment procedures so that this group could have shortened stays.

Several procedures designed to answer the question, "If a person is admitted to a mental hospital how long can he expect to remain there?", are in the literature. According to Staudt and Zubin (7) the beginning of the twentieth century saw a direction of interest toward the evaluation of the outcome of mental disease. They point out that most studies consisted of the follow-up of hospital populations over a period of several years after admission. The objective of such studies was to determine the ultimate disposition of the cases. Staudt and Zubin refer to Bond and Fuller as pioneers in this research, and review their work in considerable detail. In one paper (8) Bond showed the length of hospitalization of 111 female patients admitted to the department for mental and nervous diseases of the Pennsylvania Hospital in 1914. He followed the patients for five years following admission. In four other papers (9, 10, 11, 12) status at follow-up was given in terms of condition, such as recovered, improved, and unimproved, rather than in terms of the number who had been released from the hospital.

Fuller (13) attempted to answer the question, "What becomes of a group of mental patients within given periods of their first admission to a mental hospital system?" He felt that his findings threw light on the likelihood of eventual discharge from, without readmission to, the hospital system. He calculated the percentages of patients in hospital, dead, and discharged at the expiration of given periods after first admission by diagnostic categories and sex.

Israel and johnson (14) presented an attempt to develop a method of analyzing available statistics in order "To dilineate the prognosis in the various types of mental disease with such soundness that the physician, the family and the general public may have an accurate basis on which to develop attitudes toward those mental diseases."

Their method was to account for cohorts of admission by tabulating the percentage released within specified time intervals following admission. All first admissions to the Warren (Pennsylvania) State Hospital during 1940, 1941, and 1942 were compared with first admissions of 1930, 1931, and 1932. The comparison was made by specific diagnostic category to determine whether any change in the rate of discharge had occurred.

The authors demonstrate that the type of statistical follow-up they employ can be used to study the rate at which patients return for further hospital care.

Kramer, Goldstein, Israel, and Johnson (15) in 1954 used a life table method for following cohorts of patients to determine the trend in the probabilities of separation from hospital within specified periods following first admission for patients with specific characteristics. The method they used is that described by Dorn (16) for determining probabilities of withdrawal for a population subjected to multiple causes of decrement. They computed probabilities of release, death, and retention in hospital at various intervals following admission for all first admissions to the Warren State Hospital, Warren, Pennsylvania, between 1936 and 1945.

Since then the application of Kramer's method has been reported by Kramer and others in several articles (17, 18, 19, 20).

Malzberg (21, 22, 23, 24, 25) used a method similar to that of Kramer to determine the outcome of admission cohorts at the end of specified intervals.

A number of methods mostly based on psychological test scores and other data not routinely collected on state mental hospital patients, have been developed in an attempt to predict how long a given person will remain in a mental hospital once he has been admitted. Anker (3) points out that such studies generally have produced results of low predictive value. He cites the studies of Meeker (26) and Lindemann, et al (4) as offering more promising results.

Lindemann, et al, selected twenty-one demographic variables for study. Their subjects, 248 male neuropsychiatric patients admitted to a Veterans Administration hospital, were divided into two groups: (1) those who were hospitalized for 90 days or less, and (2) those who were hospitalized for 91 days or longer. The first group, termed the short-stay group, consisted of 120 patients; and the second group, the long-stay group, was comprised of the remaining 128 patients. The selection of 90 days as a cutting point was based on the experience of Geidt and Schlosser (27), who, the authors state, found a high frequency of discharges during the first 90 days of hospitalization. The authors then examined the two groups with respect to the twenty-one variables and tested the differences between the groups by chi-square.

They found that five variables, diagnosis, degree of incapacity, legal competence, alcoholism, and marital status, sufficiently differentiated the two groups to be included in their index. A score for each item was computed by a method which the authors attribute to Moran, Fairweather, Morton, and McGaughran (28). Moran's method consisted of recording items of information known at time of admission and evaluating the significance of differences between groups by chi-square. Each patient was then assigned a weighted score derived from those items which differentiated the groups beyond the .05 level of significance. Weights were computed by ascertaining the logarithm of every probability value derived from the significant information items. Probability values were defined as the proportion of one group to row totals in the chi-square tables. A patient's score was determined by summing the logarithm values that applied to him.

The index developed by Lindemann, et al, was found to be capable of predicting long and short stay with an overall accuracy of 77.2 per cent.

Meeker (26) through an analysis of the Minnesota Multiphasic Personality Inventory (29), developed a scale for predicting length of hospitalization. He found that 48 MMPI items differentiated between patients with short hospital stays and those with long hospital stays. His scale was developed from these 48 items. Meeker's subjects were all schizophrenics.

Anker (3) developed a scale based on items from the Minnesota Multiphasic Personality Inventory which he felt could discriminate meaningfully for durations of hospitalization up to 18 months.

Johnston and McNeal (6) report on an index which they developed using both demographic and Minnesota Multiphasic Personality Inventory data. The percentage of correct predictions using their scale ranged from 67.9 to 74.4.

Parnell and Skottowe (30) developed a prognostic scale based on demographic and other characteristics of 120 male schizophrenics. They found that six factors, age when seen, age at first mental hospital admission, occupation, somatotype, diagnosis, and particular hospital where treatment was received, showed significant statistical association with length of stay.

The authors computed the proportion of patients in each of the six categories, with the exception of diagnosis, who stayed in the hospital less than 150 days. A score for indicating probability of a short stay was constructed by adding the five proportions for each patient. The scale led to a prediction of short-stay cases that was 69 per cent correct and to a prediction of long-stay cases that was 84 per cent correct.

CHAPTER II

CHARACTERISTICS OF PATIENTS STUDIED

The data for this study consist of information on patients added (hereafter referred to as admissions) to the inpatient population of the four state supported mental hospitals of Oklahoma between January 1, 1955 and December 31, 1959, and who were subsequently released alive before, or were present in the hospital on, July 1, 1964. Length of hospitalization is defined as the total number of days spent in the hospital between first admission date and last release date or July 1, 1964, whichever came first. This allows for a follow-up period of at least four and a half years for each patient finally selected for study.

Admissions to the inpatient population of these hospitals include first admissions, readmissions, and returns to the hospital from convalescent leave or unauthorized absence. Definitions of first admission and readmission are those used by the Model Reporting Area for Mental Hospital Statistics (17). A first admission is defined as an admission with no previous admission to any public or private psychiatric hospital. A readmission is an admission with one or more previous admissions to any public or private psychiatric hospital.

The four state supported mental hospitals in Oklahoma are Central State Hospital, Norman, located in the central region of the state; Eastern State Hospital, Vinita, located in the northeast corner of the state; Western State Hospital, Supply, located in the northwest corner of the state; and Taft State Hospital, Taft, located in the East central part of the state.

Central State Hospital admits patients from that portion of central Oklahoma bounded by the counties of Jefferson, Stephens, Canadian, Logan, Payne, Lincoln, Okfuskee, Hughes, Coal, Atoka, and Bryan and the state of Texas. Eastern State Hospital receives patients from the eastern part of the state which is bounded on the west by Osage, Creek, Okmulgee, McIntosh, Pittsburg, Pushmataha, and Choctaw counties. Western State Hospital admits patients from the western side of the state bounded on the east by the counties of Kay, Noble, Garfield, Kingfisher, Blaine, Caddo, Comanche, and Cotton. At the time of admission of the patients in the present study Taft State Hospital admitted only Negro patients and these came from all counties in the state.

According to the Oklahoma Department of Mental Health Policy Manual (31), patients may be admitted to a mental hospital in Oklahoma on a voluntary basis, by order of a County Court, by a trial court, or by a District or Superior Court order. Provision is also made for the admission of minor drug addicts and mentally ill prisoners.

When a person is admitted or readmitted to one of the four Oklahoma state mental hospitals, certain demographic data relating to this person are collected and punched into an IBM card. In addition to the demographic characteristics, data that will serve to identify individual patients are also collected. These data consist of the person's name, date of birth, and hospital case number. The data collected on each patient and stored on punched cards are shown in Table 1.

TABLE 1

DATA AVAILABLE FOR EACH PATIENT ADMITTED OR READMITTED TO THE STATE SUPPORTED MENTAL HOSPITALS OF OKLAHOMA

1. Name
2. Case Number
3. Hospital to which admitted
4. Date of admission
5. Sex
6. Race
7. Date of birth
8. Age
9. Occupation
10. Marital Status
11. Diagnosis (Mental)
12. County from which admitted
13. Type of commitment (Voluntary, court order, etc.)
14. Criminal charges if any
15. Nativity
16. Education
17. Citizenship (Whether by birth, naturalization, etc.)
18. Duration of Oklahoma residence
19. Economic condition (Whether dependent, marginal, or comfortable)
20. Religion
21. Class of admission (Whether first admission or readmission)
22. Veteran status
23. Percent Indian blood
24. Addictions

The data on these patients are on file in the Division of Biometrics of the Oklahoma State Department of Mental Health. In addition to the data on admissions and readmissions, a punched card is prepared for each patient who is discharged, placed on convalescent leave or unauthorized absence, or who dies. This enables the Biometrics Section to have a complete record of the movement of each patient in and out of the hospital.

In the present study some of the patients admitted during the study period were excluded from the study before any analysis was begun. Patients classified under race as Oriental, Polynesian, and "other and unknown"; and those classified under religion as Hebrew, "other", and "unknown" were excluded because of the relatively small number of patients falling into these categories. Patients classified under diagnosis as "dead on admission", "malingerer", "no mental disorder", "undiagnosed", and "undiagnosed mental disorder" were also excluded for the same reason.

Patients admitted under district court orders were excluded in view of the fact that they are admitted for a maximum period of 90 days for observation. Within the 90 day period, those found to be suffering from a mental disorder are admitted to the hospital through other legal procedures; and those found to be free of mental disorder are released to the courts.

Certain other items of information were not used in the analysis because of the suspected low reliability of the information. These items included nativity, citizenship, duration of Oklahoma residence, economic condition, veteran status and percent of Indian blood. The remaining items were considered for use, but were regrouped as shown in Table 2.

TABLE 2

ITEMS OF INFORMATION CONSIDERED FOR ANALYSIS
AND CATEGORIES INTO WHICH GROUPED

Item	Category
1. Hospital	i) Central ii) Eastern iii) Western iv) Taft
2. Sex	i) Male ii) Female
3. Race	i) White ii) Negro iii) Indian
4. Age	i) Under 25 years ii) 25-44 iii) 45-64 iv) 65 and over and unknown
5. Occupation	i) Professional, semi-professional, managerial and official ii) Clerical and sales, service, agriculture, ranching, fishing, forestry, etc. iii) Degree of skill not stated, miscellaneous iv) No occupation, unknown
6. Marital Status	i) Single ii) Married iii) Divorced, separated iv) Widowed v) Unknown
7. Diagnosis	i) Acute brain disorders, chronic brain disorders ii) Involutional psychotic reaction Affective reactions Paranoid reactions Psychotic reactions without clearly defined structural change other than above

TABLE 2 - CONTINUED

Item	Category
	iii) Psychophysiological, autonomic and visceral disorders Psychoneurotic disorders Personality disorders iv) Mental deficiency v) Schizophrenic reactions
8. Geographic Region	i) Northwest Region ii) West Central Region iii) Southwest Region iv) South Central Region v) Central Southeast vi) Southeast vii) East Central viii) North Central ix) Northeast x) Central
9. Type of Commitment	i) Voluntary, voluntary followed by court certification ii) Physician's certificate, court certification, physician's certificate followed by court certification, emergency, other (Deportees, transfers from penal institutions) iii) Order of detention (Criminal act) Order of detention followed by court certification iv) District Court Order
10. Criminal Charges	i) No ii) Yes
11. Education	i) Illiterate - read, write or both no formal schooling ii) Completed from one through eight grades of school iii) Completed from nine through twelve grades of school, Graduate of a trade or business school iv) Completed from one through four years of college

TABLE 2 - CONTINUED

Item	Category
	v) One or more years of graduate school
12. Religion	i) Protestant ii) Catholic iii) None
13. Number of Previous Admissions	
14. Addictions	i) None, other, unknown ii) Alcohol, narcotics, barbiturates or any combination of the three.

A total of 13,731 patients were in the group finally selected for the study. The number and per cent of the study patients possessing the characteristics considered for analysis are shown in Table 3.

TABLE 3

CHARACTERISTICS OF PATIENTS ADMITTED TO OKLAHOMA STATE MENTAL HOSPITALS, JANUARY 1, 1955-DECEMBER 31, 1959, WHO WERE SELECTED FOR STUDY

Characteristic	Number	Per Cent
Hospital to Which Admitted		
Central	6155	44.8
Eastern	3832	27.9
Western	2326	16.9
Taft	1418	10.3
Number of Previous Admissions		
0	5851	42.6
1	4433	32.3
2	1973	14.4
3	847	6.2
4 or more	627	4.5
Sex		
Male	7374	53.7
Female	6357	46.3
Race		
White	12006	87.4
Negro	1418	10.3
Indian	293	2.1
Not stated	14	.2
Occupation		
Professional, etc.	835	6.1
Clerical, etc.	3866	28.2
Miscellaneous	7414	54.0
None and Unknown	1616	11.8
Diagnosis		
Acute and Chronic Brain Syndrome	2915	21.2
Psychotic Disorders other than Schizophrenia	1504	11.0
Psychoneurotic Disorders, etc.	2783	20.3
Mental Deficiency	729	5.3
Schizophrenia	5550	40.4
Other	250	1.8

TABLE 3 - CONTINUED

Characteristic	Number	Per Cent
Marital Status		
Single	3643	26.5
Married	5769	42.0
Divorced and Separated	2794	20.3
Widowed	1045	7.6
Unknown	395	2.9
Other	85	.6
Age (In years)		
< 25	1909	13.9
25-44	5877	42.8
45-64	4180	30.4
65 and Over	1697	12.4
Not stated	68	.5
Criminal Charges		
No	12815	93.3
Yes	916	6.7
Religion		
Protestant	10590	77.1
Catholic	778	5.7
None	1511	11.0
Other and Unknown	852	6.2
Type of Commitment		
Voluntary	1048	7.6
Physician's Certificate and Court Certification	11243	81.9
Order of Detention	1376	10.0
Not stated	64	.5
Education		
None	701	5.1
Elementary	6154	44.8
High School and Trade or Business School	4807	35.0
College	1299	9.4
Not stated	770	5.6
Region from which Admitted		
Central	4329	31.5
Northwest	829	6.0
West Central	718	5.2

TABLE 3 - CONTINUED

Characteristic	Number	Per Cent
Southwest	1034	7.5
South Central	925	6.7
Central Southeast	771	5.6
Southeast	814	5.9
East Central	907	6.6
North Central	852	6.2
Northeast	2552	18.6
Addictions		
None	11694	85.2
Alcohol	1848	13.5
Other Addictions	184	1.3
Not stated	5	0.0

CHAPTER III

PROPOSED METHOD FOR PREDICTING LENGTH OF HOSPITALIZATION

The mathematical model employed in the present attempt to predict the length of stay in a mental hospital is one based on Bayes' Theorem (32). This theorem states that the conditional probability of an event occurring, given that a set of other events has occurred, can be expressed as a function of the conditional probability of the second events given that the first has occurred. In the present application it is desired to determine the conditional probability, P_{y_i}, of a certain length of hospitalization, y_i, experienced by a patient possessing a set of characteristics x, as a function of the conditional probabilities of the characteristics, given the length of hospitalization categories.

The probability of y_i, P_{y_i}, is the probability of a categorized length of hospitalization, say less than 30 days, among the 13,731 study patients. It is estimated by dividing the number of patients, N_{y_i}, who remain in the hospital the specified length of time, y_i, by 13,731.

The conditional probability of a set of characteristics, x, occurring in a patient whose length of hospitalization is y_i is given by Equation 1

$$P_{x \mid y_i} = \frac{N_{xy_i}}{N_{y_i}} \tag{1}$$

where N_{xy_i} is the number of patients whose length of hospitalization is y_i and who also have the set of characteristics x, and N_{y_i} is the total number of patients whose length of hospitalization is y_i. If the numerator and denominator of the right hand term of Equation 1 are divided by the total number of patients, denoted by $N_{y_1, y_2, \ldots, y_k}$, (k is the total number of length of hospitalization categories) Equation 2 results.

$$P_{x \mid y_i} = \frac{P_{xy_i}}{P_{y_i}} \qquad (2)$$

Similarly, the probability of length of hospitalization y_i occurring in the presence of the set of characteristics x may be written as in Equation 3

$$P_{y_i \mid x} = \frac{P_{xy_i}}{P_x} \qquad (3)$$

where P_x is the probability of the set of characteristics x occurring in any patient. If the length of hospitalization categories are considered mutually exclusive, the probability of the set of characteristics x may be written as in Equation 4.

$$P_x = \sum_{\text{All } i} P_{y_i} P_{x \mid y_i} \qquad (4)$$

From Equations 2, 3, and 4, the probability of length of hospitalization, y_i, given the set of characteristics, x, may be expressed as

$$P_{y_i \mid x} = \frac{P_{y_i} P_{x \mid y_i}}{\sum_{\text{All } i} P_{y_i} P_{x \mid y_i}} \qquad (5)$$

Equation 5 is an expression of Bayes' rule for the probability of causes.

If the individual characteristics which make up the set x are independent, the conditional probability of x given y_i is the product of the probabilities of the individual characteristics making up the set. That is,

$$P_{x \mid y_i} = P_{x_1 \mid y_i} \, P_{x_2 \mid y_i} \cdots P_{x_m \mid y_i} \quad . \tag{6}$$

Using Equation 6, Equation 5 may be rewritten in an expanded form as

$$P_{y_i \mid x_1 x_2, \ldots x_m} = \frac{(P_{y_i})(P_{x_1 \mid y_i} \; P_{x_2 \mid y_i} \ldots P_{x_m \mid y_i})}{\sum_{i=1}^{k} (P_{y_i})(P_{x_1 \mid y_i} \; P_{x_2 \mid y_i} \ldots P_{x_m \mid y_i})} \, , \; i=1, \ldots k. \tag{7}$$

With the use of Equation 7 it is possible to calculate the probability that a patient will remain in a mental hospital a specified length of time, say less than 30 days, given that he has a certain set of characteristics such as male, less than 25 years old, single, white, etc.

There are three assumptions basic to Bayes' Theorem. First, it is necessary that the patient characteristics used in the analysis be independent. Second, it is necessary that one know the true probability of each y_i in the population under study and the conditional probability of each characteristic in each of the length of hospitalization categories. Third, it is necessary that the length of hospitalization categories be mutually exclusive. The latter requirement is met since a patient who is released from the hospital within, say thirty days, cannot have remained in the hospital longer than thirty days. However, in the present study the true probabilities are not known. Instead, proportions based on the sample of 13,731 patients are used as estimates of the true probabilities. The use of estimates, rather than the true probabilities, has not been investigated from a theoretical point of view. On the other hand, other research has demonstrated the utility of Bayes' Theorem.

The studies referred to below indicate that Bayes' Theorem has been used with considerable success in diagnosing physical conditions when the probabilities associated with the symptoms and signs of these conditions are estimated. The present study investigates its usefulness in a somewhat different area. However, the concepts are the same; length of hospitalization plays the role of the disease, and patient characteristics are used instead of signs and symptoms. As will be demonstrated in Chapter V, the characteristics used in the present study are not independent. It was felt that it would be worthwhile to investigate the usefulness of the model in predicting length of hospitalization category, even though this basic assumption was not met. To continue seemed especially warranted since other users of the method have met with considerable success while expressing reservations as to whether their data fulfilled the assumption of independence.

Warner et al (33, 34, 35) have used Bayes' Theorem in the diagnosis of congenital heart disease with the aid of an electronic computer. They used a total of fifty-three symptoms in computing the probabilities associated with each of thirty-five diagnoses. The authors report that their experience with congenital heart diseases indicates that these diseases can be diagnosed with an accuracy equal to that of an experienced specialist in the field. Initially, in their paper the accuracy of the physician was compared with that obtained by the use of Bayes' Theorem in diagnosing congenital heart disease in 43 patients. Later an additional 40 patients were added to the sample, and improvements were made in the statistical information in the data matrix used for computing the probabilities. The authors report a significant improvement in diagnostic accuracy using Bayes' formula with the improved data matrix. Warner, et al, (33) point out that at the time of their study they were unable to evaluate their data matrix for independence of symptoms, because much of the data were gathered from the literature. The authors anticipate the possibility of analyzing each symptom for independence as they accumulate additional data from their own experience.

Nugent, Warner, Dunn, and Tyler (36) have used Bayes' Theorem in the diagnosis of Cushing's syndrome. The probability of Cushing's syndrome was computed in 111 patients in whom this diagnosis was suspected. It was later established that thirty-eight of these patients did in fact have Cushing's syndrome and seventy-three did not. Of the thirty-eight patients in whom the diagnosis was confirmed, the use of Bayes' Theorem led to a "very likely" diagnosis ($p \geqq .99$) in nine cases, a "likely" diagnosis ($.99 > p \geqq .9$) in eight cases and an "unlikely" or "uncertain" diagnosis ($.01 < p < .9$) in twenty-one cases. In none of the thirty-eight cases was a probability of .01 or less obtained. Taking the 73 patients who were found to be free of Cushing's syndrome, the use of Bayes' Theorem led to a "very unlikely" diagnosis ($p \leqq .01$) in forty-five cases, an "unlikely" diagnosis ($.01 < p < .1$) in fourteen cases, and an "uncertain" or "likely" diagnosis ($.1 \leqq p < .9$) in fourteen cases. For none of the 73 patients was a probability of .99 or more calculated.

Combining the nine confirmed cases receiving a probability of .99 or more by the formula, and the forty-five syndrome-free cases receiving a probability of .01 or less, the authors conclude that in fifty-four of the 111 cases no diagnostic errors were made. A total of eleven signs were used in calculating the probability of Cushing's syndrome.

Rinaldo, Scheinok, and Rupe (37) used Bayes' Theorem for diagnosing six gastrointestinal diseases with the aid of eight symptoms. The computer correctly diagnosed 73 per cent of hiatus hernia cases, 69 per cent of duodenal ulcer cases, 27 per cent of gastric ulcer cases, 75 per cent of gallstone cases, 38 per cent of functional cases, and 33 per cent of gastric carcinomas.

Mount and Evans (38) have reported results from a computer simulation for the Bayes process of computer diagnosis. By use of a computer they generated complete probabilistic descriptions of a disease-symptom universe conforming to the hypotheses of statistical independence of symptoms occurring in a given disease and mutual exclusiveness of the diseases. Their next step was to generate a group of "people" randomly according to this real distribution. Each "person" was represented by a disease code and a series of 1's and 0's representing the presence or absence of the symptoms. The authors then attempted to simulate the process which a physician faces in using Bayes' formula. The "people" which they had generated were used to construct a conditional probability array and an initial distribution of probabilities. Following this, another group of "people" were generated in the same manner as the first. For each "person" generated in this last step, a mathematical diagnosis was derived using Bayes' formula. The incidence of the diseases and symptoms obtained from the first group of "people" generated, served as the input for the formula. The disease with the maximum probability by Bayes' formula was the one selected as the diagnosis for each "person". The mathematically most probable diagnosis was computed sequentially for each "person" using new symptoms one at a time. After each new symptom was considered, the number of successes, out of the total diagnoses made, was used to measure the per cent of correct diagnoses. The authors found that increasing the sample size of "people" from which the disease-symptom frequency array was constructed, in general, increased the accuracy of the prediction using Bayes' formula. This seems to be a reasonable result since, as the sample size increases, the proportions used as estimates of the true probabilities become more stable. They also found that there was a steady improvement in the percentage of correct diagnoses as the number of statistically independent symptoms was increased.

CHAPTER IV

THE STUDY OF SOME INTERRELATIONSHIPS AMONG SELECTED CHARACTERISTICS

In order to gain some insight into the interrelationships among the characteristics selected for study a number of statistical analyses were undertaken. To ascertain to what extent the various characteristics were associated with length of hospitalization, chi-square values were computed to test the null hypothesis of no association between length of hospitalization and each characteristic. This was done for three sets of length of hospitalization categories, less than thirty days and thirty days or more; less than ninety days and 90 days or more; and less than 30 days, 30 through 59 days, 60 through 89 days, and 90 days or more. The calculated chi-square values for each characteristic for each set of length of hospitalization categories are displayed in Table 4.

The association of certain characteristics with length of hospitalization, at different "levels" of some of the other characteristics, was investigated. A further breakdown, in some instances by sex and in other instances by race, was also made. The chi-square test for association with length of hospitalization was applied to the resulting contingency tables. The results of this analysis are given in Tables 8 through 26.

In addition to making the overall chi-square test, a method due

TABLE 4

CALCULATED CHI-SQUARE VALUES FOR SELECTED CHARACTERISTICS FOR EACH SET OF LENGTH OF HOSPITALIZATION CATEGORY

Characteristic	Length of Hospitalization Categories (In days)					
	< 30 30 or more		< 90 90 or more		< 30, 30-59, 60-89, 90 or more	
	d.f.	χ^2	d.f.	χ^2	d.f.	χ^2
Hospital to which admitted	3	77.33	3	135.80	9	248.25
Number of previous admissions	7	158.60	7	386.70	21	428.79
Sex	1	281.68	1	172.39	3	352.98
Race	2	7.23	2	21.83	6	62.37
Occupation	3	11.76	3	24.50	9	55.95
Diagnosis	4	756.15	4	1257.42	12	1544.94
Marital Status	4	38.09	4	200.38	12	243.11
Age	3	137.45	3	75.95	9	159.47
Criminal Charges	1	628.08	1	416.30	3	738.07
Religion	3	41.68	3	3.97#	9	53.75
Type of Commitment	2	562.61	2	451.51	6	718.57
Education	4	27.11	4	52.58	12	71.29
Region of state from which admitted	9	125.28	9	150.04	27	249.65
Addictions	2	158.92	2	385.92	6	433.81

#Not significant at .05 level.

to Cochran (39, 40, 41, 42) was used for testing for linear trend in the contingency tables. Cochran's method is applicable to 2 X c tables where the columns fall in a natural order. Scores, z_i, are assigned to the columns, in an effort to convert them to values on a continuous scale. The regression coefficient, b, of the estimate of the column probability $\hat{p}_i$ on z_i is a weighted regression given by Equation 8.

$$b = \frac{\sum_i n_{i.} (\hat{p}_i - \hat{\bar{p}}_i)(z_i - \bar{z}_w)}{\sum_i n_{i.} (z_i - \bar{z}_w)^2} \tag{8}$$

where $\hat{\bar{p}} = \sum_i n_{i2} / \sum_i n_{i.}$; $\bar{z}_w = \sum_i n_{i.} z_i / \sum_i n_{i.}$, the weighted mean of z_i; and $\hat{p}_i$ is given weight $n_{i.} / \hat{\bar{p}} (1 - \hat{\bar{p}})$. The notation is that of Steel and Torrie (41). The χ^2 for regression with one degree of freedom is given by Equation 9.

$$\chi^2_{(1)} = \frac{\left[\sum_i n_{i.} (\hat{p}_i - \hat{\bar{p}})(z_i - z_w)\right]^2}{\hat{\bar{p}}(1 - \hat{\bar{p}}) \sum_i n_{i.} (z_i - z_w)^2} \tag{9}$$

In Table 5 are shown all male patients with college training and advanced degrees broken down by age group and length of hospitalization. Cochran presents a computing formula for $\chi^2_{(1)}$ which, in the notation of Steel and Torrie, is given in Equation 10.

$$\chi^2_{(1)} = \frac{\left[\sum_i n_{i2} z_i - n_{.2} (\Sigma n_{i.} z_i / n_{..})\right]^2}{\left[\sum_i n_{i.} z_i^2 - (\sum_i n_{i.} z_i)^2 / n_{..}\right] (\hat{\bar{p}})(1 - \hat{\bar{p}})} \tag{10}$$

Using Equation 10 and the data of Table 5, the following chi-square value for the regression of $\hat{p}_i$ on z_i was calculated:

TABLE 5

AGE AND LENGTH OF HOSPITALIZATION, MALE PATIENTS WITH COLLEGE TRAINING AND ADVANCED DEGREES

Length of Hospitalization (In Days)	Age Group (In Years)				Total
	< 25	25-44	45-64	65 and Over	
	Score z_i				
	1	2	3	4	
< 30 n_{i1}	10	77	22	7	$116 = n_{.1}$
30^+ n_{i2}	36	285	155	41	$517 = n_{.2}$
Total $n_{i.}$	46	362	177	48	$633 = n_{..}$
$\hat{p}_i$	.78	.79	.88	.85	$\hat{p} = .8167$
$n_{i.}\ z_i$	46	724	531	192	$\Sigma n_{i.}\ z_i = 1493$

$$\chi^2_{(1)} = \frac{[\,36(1) + \ldots + 41(4) - (517)(1493/633)\,]^2}{[\,46(1) + \ldots + 48(16) - (1493)^2/633](.8167)(.1833)}$$

$$= 4.87.$$

The total χ^2 can be partitioned as follows:

Source	Degrees of freedom	χ^2
Regression of $\hat{p}_i$ on z_i	1	4.87
Deviations from Regression	2	2.15
Total	3	7.02

It is seen that although the overall χ^2, with 3 degrees of freedom, is not significant at the .05 level, the χ^2 value due to regression, based on only one degree of freedom, is significant at the .05 level. Partitioning the overall value has increased the sensitivity of the test and it is concluded that there is a significant increase in the proportion of these patients remaining in the hospital 30 days or longer with increasing age. We may conclude further that the increase is linear rather than curvilinear since the departure from linear regression is represented by a chi-square value of only 2.15, based on two degrees of freedom, which is not significant. The results of the tests for linear trend are given in Tables 8 through 26.

In the present study, Cochran's method of testing for linear trend is extended in order to test for quadratic trend in contingency tables. The extension is also applicable to 2 X c tables where the columns fall in a natural order. Scores, z_i, are assigned to the columns as they were in the linear case. In order to test for quadratic trend four columns are needed.

For estimates of the two partial regression coefficients, designated by a and b, the normal equations are given in Equations 11 and 12.

$$a\Sigma n_{i.}\ (z_i - \overline{z}_w)^2 + b\Sigma n_{i.}\ (z_i - \overline{z}_w)\ (z_i^{\ 2} - \overline{z_w^{\ 2}})$$

$$= \Sigma n_{i.}\ (z_i - \overline{z}_w)\ (\hat{p}_i - \overline{\hat{p}}) \tag{11}$$

$$a\Sigma n_{i.}\ (z_i - \overline{z}_w)\ (z_i^{\ 2} - \overline{z_w^{\ 2}}) + b\Sigma n_{i.}\ (z_i^{\ 2} - \overline{z^2})^2$$

$$= \Sigma n_{i.}\ (z_i^{\ 2} - \overline{z_w^{\ 2}})\ (\hat{p}_i - \overline{\hat{p}}) \tag{12}$$

In Equations 11 and 12, $\overline{\hat{p}}$ and $\overline{z_w}$ are the same as in Equation 8 and $\overline{z_w^{\ 2}} = \sum_i n_{i.}\ z_i^{\ 2} / \sum_i n_{i.}$. The weights, $n_{i.}$, are the column totals. The equations are analogous to those for the multiple regression case with the z_i playing the role of the first independent variable and the $z_i^{\ 2}$ taking the place of the second dependent variable.

For the quadratic case, estimates of the two partial regression coefficients, designated by a and b, are computed by Equation 13 and 14.

$$a = \frac{\begin{array}{l}[\Sigma n_{i.}\ (z_i - \overline{z}_w)(\hat{p}_i - \overline{\hat{p}})]\ [\Sigma n_{i.}\ (z_i^{\ 2} - \overline{z_w^{\ 2}})^2] \\ \quad - [\Sigma n_{i.}\ (z_i^{\ 2} - \overline{z_w^{\ 2}})\ (\hat{p}_i - \overline{\hat{p}})]\ [\Sigma n_{i.}\ (z_i - \overline{z}_w)(z_i^{\ 2} - \overline{z_w^{\ 2}})]\end{array}}{[\Sigma n_{i.}\ (z_i - \overline{z}_w)^2]\ [\Sigma n_{i.}\ (z_i^{\ 2} - \overline{z_w^{\ 2}})^2] - [\Sigma n_{i.}(z_i - \overline{z}_w)(z_i^{\ 2} - \overline{z_w^{\ 2}})]^2} \tag{13}$$

$$b = \frac{\begin{array}{l}[\Sigma n_{i.}\ (z_i - \overline{z}_w)^2]\ [\Sigma n_{i.}\ (z_i^{\ 2} - \overline{z_w^{\ 2}})\ (\hat{p}_i - \overline{\hat{p}})] \\ \quad - [\Sigma n_{i.}\ (z_i - \overline{z}_w)(z_i^{\ 2} - \overline{z_i^{\ 2}})]\ [\Sigma n_{i.}\ (z_i - \overline{z}_w)\ (\hat{p}_i - \overline{\hat{p}})]\end{array}}{[\Sigma n_{i.}\ (z_i - \overline{z}_w)^2]\ [\Sigma n_{i.}\ (z_i^{\ 2} - \overline{z_w^{\ 2}})^2] - [\Sigma n_{i.}(z_i - \overline{z}_w)(z_i^2 - \overline{z_w^{\ 2}})]^2} \tag{14}$$

Chi-square for quadratic regression, with two degrees of freedom, is given by Equation 15.

$$\chi^2_{(2)} = [a \, \Sigma \, n_{i.} \, (z_i - \bar{z}_w)(\hat{p}_i - \hat{\bar{p}}) + b \, \Sigma \, n_{i.} \, (z_i^{\,2} - \overline{z_w^{\,2}}) (\hat{p}_i - \hat{\bar{p}}) \,] \,/\, (\hat{\bar{p}}) \, (1 - \hat{\bar{p}}) \tag{15}$$

Table 6 shows the breakdown of all patients having three previous admissions by length of hospitalization and age group. Using Equation 15 and the data from Table 6 the χ^2 value due to quadratic regression may be computed. The χ^2 value for linear regression may be computed also. The total χ^2, with three degrees of freedom, may be partitioned as follows:

Source	Degrees of freedom	χ^2
Quadratic Regression	2	6.12
Linear Regression	1	3.58
Quadratic-Linear	1	2.54
Deviations from Regression	1	1.06
Total	3	7.18

Although the total χ^2 value, with three degrees of freedom, is not significant at the .05 level, the component attributable to quadratic regression is significant, while the component attributable to linear regression is not. It is concluded that there is a significant increase in the proportion of these patients remaining in the hospital 30 days or longer as age increases, but that there is a significant decrease for this group of patients who are 65 years old or older. The results for additional tests for quadratic trend are given in Tables 8 through 26.

A further extension of Cochran's method of testing for trend in contingency tables was made to include the case of multiple regression

TABLE 6

AGE AND LENGTH OF HOSPITALIZATION, ALL PATIENTS HAVING THREE PREVIOUS ADMISSIONS

Length of Hospitalization (In Days)	Age Group (In Years)				
	$\leq$ 25	25-44	45-64	65 and Over	Total
	Score z_i				
	1	2	3	4	
< 30 n_{i1}	6	35	14	6	$61 = n_{.1}$
30^+ n_{i2}	41	371	304	70	$786 = n_{.2}$
Total $n_{i.}$	47	406	318	76	$847 = n_{..}$
$\hat{p}_i$	.87	.91	.96	.92	$\hat{\bar{p}} = .9280$
$n_{i.}\ z_i$	47	812	954	304	$\Sigma n_{i.} z_i = 2117$

with three variables.

Let x_i indicate the i^{th} score assigned to one of the independent variables, y_j the j^{th} score assigned to the other independent variable, n_{ij} the total number of patients in the ij^{th} cell, $\hat{p}_{ij}$ the proportion of the n_{ij} patients staying in the hospital 30 days or more, $\bar{\hat{p}}$ the proportion of all study patients staying in the hospital 30 days or more, $\bar{x}_w$ the weighted mean x score and $\bar{y}_w$ the weighted mean y score. The partial regression coefficients b_1 and b_2 are obtained by solving Equations 16 and 17.

$$b_1 \Sigma x^2 + b_2 \Sigma xy = \Sigma xp \tag{16}$$

$$b_1 \Sigma xy + b_2 \Sigma y^2 = \Sigma yp \tag{17}$$

where $$\Sigma x^2 = \Sigma n_{ij} (x_i - \bar{x}_w)^2$$

$$\Sigma xy = \Sigma n_{ij} (x_i - \bar{x}_w)(y_j - \bar{y}_w)$$

$$\Sigma xp = \Sigma n_{ij} (x_i - \bar{x}_w)(\hat{p}_{ij} - \bar{\hat{p}})$$

$$\Sigma y^2 = \Sigma n_{ij} (y_j - \bar{y}_w)^2$$

and $$\Sigma yp = \Sigma n_{ij} (y_j - \bar{y}_w)(\hat{p}_{ij} - \bar{\hat{p}}) .$$

Table 7 shows the study patients broken down by number of previous admissions by age group and length of hospitalization category. The $\hat{p}_{ij}$'s are shown in parentheses. Table 7 contains all the information necessary for obtaining the normal equations.

Using the data from Table 7 the following quantities are computed.

TABLE 7

STUDY PATIENTS CATEGORIZED BY NUMBER OF PREVIOUS ADMISSIONS, AGE GROUP, AND LENGTH OF HOSPITALIZATION

Age (In Years)	x_i Scores	Length of Hospitalization (In Days)	Number of Previous Admissions y_j Scores					
			0	1	2	3	4 or More	Total
< 25	1	< 30	917	411	136	41	33	1538
		30+	256	91	17	6	1	371
		Total	1173	502	153	47	34	1909
			(.782)	(.819)	(.889)	(.872)	(.971)	
25-44	2	< 30	1855	1746	843	371	250	5065
		30+	373	280	83	35	41	812
		Total	2228	2026	926	406	291	5877
			(.833)	(.862)	(.910)	(.914)	(.859)	
45-64	3	< 30	1308	1301	683	304	212	3808
		30+	189	105	39	14	25	372
		Total	1497	1406	722	318	237	4180
			(.874)	(.925)	(.946)	(.956)	(.895)	
65 and Over	4	< 30	757	442	160	70	62	1491
		30+	151	34	12	6	3	206
		Total	908	476	172	76	65	1697
			(.834)	(.929)	(.930)	(.921)	(.954)	

$\hat{p} = (1538 + \ldots + 1491)/13{,}663 = .871$

$\bar{x}_w = [(1909)(1) + \ldots + (1697)(4)] \;/\; 13{,}663 = 2.415$

$\bar{y}_w = [(5806)(0) + \ldots + (627)(4)] \;/\; 13{,}663 = .981$

$\Sigma\, xp = (1173)(-1.415)(-.089) + \ldots + (65)(1.585)(.083) = 316.763$

$\Sigma\, yp = (1173)(-.981)(-.089) + \ldots + (65)(3.019)(.083) = 451.044$

$\Sigma\, xy = (1173)(-.981)(-1.415) + \ldots + (65)(3.019)(1.585) = 724.973$

$\Sigma\, x^2 = (1173)(-1.145)^2 + \ldots + (65)(1.585)^2 = 8{,}320.057$

$\Sigma\, y^2 = (1173)(.981)^2 + \ldots + (65)(3.019)^2 = 16{,}805.128.$

The normal equations are then

$$8{,}320.057\, b_1 + 724.973\, b_2 = 316.763$$

$$724.973\, b_1 + 16{,}805.128\, b_2 = 451.044,$$

the solution of which gives $b_1 = .036$ and $b_2 = .025$.

The χ^2 attributable to multiple regression (with two degrees of freedom) is given by Equation 18.

$$\chi^2_{(2)} = (b_1\, \Sigma\, xp + b_2\, \Sigma\, yp) \;/ (\hat{\bar{p}})\; (1 - \hat{\bar{p}}) \tag{18}$$

Substituting the above values into Equation 18 gives $\chi^2_{(2)} = 202.5$.

The total chi-square value may be subdivided as follows:

Source	d.f.	χ^2
Multiple Regression	2	202.5
Deviations from Regression	10	669.4
Total	12	871.9

It is seen that, although a significant component of the total chi-square value is accounted for by multiple regression, the deviations from regression component is also highly significant.

A similar analysis was made in which the independent variables were number of previous admissions and education. The results were as follows:

Source	d.f.	χ^2
Multiple Regression	2	124.6
Deviations from Regression	10	142.0
Total	12	266.6

The results of applying the analysis to education by age by length of hospitalization were as follows:

Source	d.f.	χ^2
Multiple Regression	2	196.1
Deviations from Regression	7	1089.9
Total	9	1286.0

Again a significant component of the total χ^2 value was accounted for by multiple regression, but the deviations from regression is also significant.

CHAPTER V

RESULTS

Results of Chi-Square Analysis

Tables 8 through 10 summarize the results of testing for linear and quadratic trend in contingency tables using the methods discussed in the previous chapter. The total chi-square value is listed in the right hand column of the tables. The first two columns contain the chi-square values due to quadratic and linear regression, and the quadratic component minus the linear component and the deviations from regression component are given in the third and fourth columns. Chi-square values significant at the .05 level are indicated by an asterisk.

The comparisons reported in these tables are not all independent. However, since the nature of the interrelationships of the variables studied is of interest it was felt that the analysis as presented here would be more helpful than misleading.

Table 4 shows that all chi-square values are highly significant except for religion versus length of hospitalization in the less than 90 days and 90 days or over group. Tables 8 through 26 reveal that there is a great deal of interdependence among the characteristics. The tests for trends indicate that in many cases the apparent association between a characteristic and length of hospitalization is due to either a linear or quadratic trend in the proportions of patients falling in a particular

TABLE 8

CHI-SQUARE ANALYSIS, MALE AND FEMALE PATIENTS, AGE IN YEARS (<25, 25-44, 45-64, 65 AND OVER) BY LENGTH OF HOSPITALIZATION (<30 DAYS, 30 DAYS AND OVER) AND DIAGNOSIS#

	Quadratic	Linear	Quadratic-Linear	Deviations from Regression	Total
A. Acute or Chronic Brain Syndrome					
Male	.57	.22	.35	3.60	4.17
Female	.14	.00	.14	3.11	3.25
B. Psychotic Disorder (Excluding Schizophrenia)					
Male	13.08*	8.22*	4.86*	.40	13.48*
Female	9.20*	3.41	5.79*	.02	9.22*
C. Psychoneurotic Disorder, etc.					
Male	36.92*	24.65*	12.27*	8.02*	44.94*
Female	2.82	.88	1.94	.54	3.36
D. Mental Deficiency					
Male	1.59	.50	1.09	.72	2.31
Female	2.48	2.44	.04	.02	2.50
E. Schizophrenia					
Male	21.52*	21.37*	.15	2.00	23.52*
Female	5.36	4.94*	.42	.04	5.40

#In Tables 8 through 14, the number of degrees of freedom involved are as follows: quadratic, 2; linear, 1; quadratic-linear, 1; deviations from regression, 1; total, 3.

*Indicates significance at the .05 level.

TABLE 9

CHI-SQUARE ANALYSIS, MALE AND FEMALE PATIENTS, AGE IN YEARS (<25, 25-44, 45-64, 65 AND OVER) BY LENGTH OF HOSPITALIZATION (<30 DAYS, 30 DAYS AND OVER) AND MARITAL STATUS

	Quadratic	Linear	Quadratic-Linear	Deviations from Regression	Total
A. Single					
Male	4.54	2.40	2.14	.52	5.06
Female	52.65*	46.10*	6.55*	.04	52.69*
B. Married					
Male	81.84*	60.50*	21.34*	9.20*	91.04*
Female	24.83*	.75	24.08*	1.82	-- 26.65*
C. Divorced and Separated					
Male	41.66*	36.29*	5.37*	.12	41.78*
Female	28.95*	11.68*	17.27*	7.89*	36.84*
D. Widowed					
Male	.27	.13	.14	1.64	1.91
Female	10.40*	3.99*	6.41*	3.22	13.62*
E. Unknown					
Male	2.14	1.87	.27	.02	2.16
Female	3.03	.40	2.63	.12	3.15

TABLE 10

CHI-SQUARE ANALYSIS, MALE AND FEMALE PATIENTS, AGE IN YEARS (<25, 25-44, 45-64, 65 AND OVER) BY LENGTH OF HOSPITALIZATION (<30 DAYS, 30 DAYS AND OVER) AND NUMBER OF PREVIOUS ADMISSIONS

	Quadratic	Linear	Quadratic-Linear	Deviations from Regression	Total
A. None					
Male	28.37*	22.95*	5.42*	3.31	31.68*
Female	10.44*	.00	10.44*	.62	11.06*
B. One					
Male	47.89*	47.56*	.33	5.07*	52.96*
Female	12.81*	6.20*	6.61*	.83	13.64*
C. Two					
Male	8.17*	8.11*	.06	3.17	11.34*
Female	1.77	.11	1.66	.34	2.11
D. Three					
Male	2.76	2.55	.21	.97	3.73
Female	3.67	1.20	2.47	.27	3.94
E. Four or More					
Male	4.86	2.26	2.60	1.07	5.93
Female	1.77	.01	1.76	.25	2.02

TABLE 11

CHI-SQUARE ANALYSIS, MALE AND FEMALE PATIENTS, AGE IN YEARS (<25, 25-44, 45-64, 65 AND OVER) BY LENGTH OF HOSPITALIZATION (<30 DAYS, 30 DAYS AND OVER) AND OCCUPATION

	Quadratic	Linear	Quadratic-Linear	Deviations from Regression	Total
A. Professional					
Male	8.13*	6.01*	2.12	.01	8.14*
Female	3.06	.00	3.06	.93	3.99
B. Clerical, etc.					
Male	61.68	30.91	30.77	.88	62.56
Female	11.49	5.24	6.25	2.53	14.02
C. Miscellaneous					
Male	42.55*	42.54*	.01*	9.24	51.79*
Female	22.57*	.76*	21.81*	4.44	27.01*
D. Unemployed					
Male	2.15	2.14	.01	.42	2.57
Female	5.52	.45	5.07*	3.26	8.78

TABLE 12

CHI-SQUARE ANALYSIS, MALE AND FEMALE PATIENTS, AGE IN YEARS (<25, 25-44, 45-64, 65 AND OVER) BY LENGTH OF HOSPITALIZATION (<30 DAYS, 30 DAYS AND OVER) AND CRIMINAL CHARGES

	Quadratic	Linear	Quadratic-Linear	Deviations from Regression	Total
A. Without Criminal Charges					
Male	23.86*	22.58*	1.28	14.74*	38.60*
Female	34.91*	2.01	32.90*	2.18	37.09*
B. With Criminal Charges					
Male	16.69*	16.68*	.01	1.81	18.50*
Female	1.21	1.13	.08	.31	1.52

TABLE 13

CHI-SQUARE ANALYSIS, MALE AND FEMALE PATIENTS, AGE IN YEARS (<25, 25-44, 45-64, 65 AND OVER) BY LENGTH OF HOSPITALIZATION (<30 DAYS, 30 DAYS AND OVER) AND EDUCATION

	Quadratic	Linear	Quadratic-Linear	Deviations from Regression	Total
A. No Education					
Male	2.22	.35	1.87	.00	2.22
Female	.17	.16	.01	.59	.76
B. Elementary School					
Male	22.40*	18.79*	3.61	9.35*	31.75*
Female	20.90*	.17	20.73*	1.51	22.41*
C. High School					
Male	52.30*	51.92*	.38	2.31	54.61*
Female	25.23*	5.51*	19.72*	.48	25.71*
D. College Training and Advanced Degrees					
Male	4.90	4.87*	.03	2.12	7.02
Female	1.05	.00	1.05	.02	1.07

TABLE 14

CHI-SQUARE ANALYSIS, MALE AND FEMALE PATIENTS, AGE IN YEARS (<25, 25-44, 45-64, 65 AND OVER) BY LENGTH OF HOSPITALIZATION (<30 DAYS, 30 DAYS AND OVER) AND ADDICTIONS

	Quadratic	Linear	Quadratic-Linear	Deviations from Regression	Total
A. No Addictions					
Male	83.63*	62.17*	21.46*	11.31*	94.94*
Female	31.72*	2.03	29.69*	3.56	35.28*
B. Alcohol					
Male	18.87*	18.87*	.00	.07	18.94*
Female	4.36	.76	3.60	1.47	5.83
C. Other Addictions					
Male	8.81*	6.23*	2.58	2.98	11.79*
Female	4.29	.18	4.11*	1.83	6.12

TABLE 15

CHI-SQUARE ANALYSIS, WHITE AND NEGRO PATIENTS, EDUCATION BY LENGTH OF HOSPITALIZATION (<30 DAYS, 30 DAYS OR MORE) AND SEX#

	Quadratic	Linear	Quadratic-Linear	Deviations from Regression	Total
A. Males					
White	36.00*	30.00*	6.00*	13.53*	49.53*
Negro	5.93	.99	4.94*	.00	5.93
B. Females					
White	6.77*	3.52	3.25	1.24	8.01
Negro	1.45	.41	.04	.59	2.04

#In Tables 15 through 26, the numbers of degrees of freedom involved are as follows: quadratic, 2; linear, 1; quadratic-linear, 1; deviations from regression, 2; total, 4.

##Chi-square values for these categories not calculated because of the small number of patients falling into most of the cells.

TABLE 16

CHI-SQUARE ANALYSIS, WHITE AND NEGRO PATIENTS, EDUCATION BY LENGTH OF HOSPITALIZATION (<30 DAYS, 30 DAYS OR MORE) AND OCCUPATION

	Quadratic	Linear	Quadratic-Linear	Deviations from Regression	Total
A. Professional					
White	2.88	1.14	1.74	5.45	8.33
Negro	##				
B. Clerical, etc.					
White	7.61*	3.60	4.01*	1.33	8.94
Negro	1.54	.19	.35	.00	1.54
C. Miscellaneous					
White	.95	.14	.81	5.30	6.25
Negro	2.12	1.71	.41	.81	2.93
D. Unemployed					
White	5.45	.17	5.28*	7.73*	13.18*
Negro	2.38	.28	.10	1.37	3.75

TABLE 17

CHI-SQUARE ANALYSIS, WHITE AND NEGRO PATIENTS, EDUCATION BY LENGTH OF HOSPITALIZATION (<30 DAYS, 30 DAYS OR MORE) AND DIAGNOSIS

	Quadratic	Linear	Quadratic-Linear	Deviations from Regression	Total
A. Acute or Chronic Brain Syndrome					
White	1.83	.88	.95	3.35	5.18
Negro	2.44	2.11	.33	3.45	5.89
B. Psychoneurotic, etc.					
White	3.44	.72	2.72	.63	4.07
Negro	1.54	.29	.25	.01	1.55
C. Schizophrenia					
White	.55	.21	.34	.06	.61
Negro	.54	.50	.04	1.38	1.92

TABLE 18

CHI-SQUARE ANALYSIS, WHITE AND NEGRO PATIENTS, EDUCATION BY LENGTH OF HOSPITALIZATION (<30 DAYS, 30 DAYS OR MORE) AND MARITAL STATUS

	Quadratic	Linear	Quadratic-Linear	Deviations from Regression	Total
A. Single					
White	10.63*	6.21*	4.42*	2.79	13.42*
Negro	3.31	.01	3.30	.01	3.32
B. Married					
White	.36	.29	.07	11.10*	11.46*
Negro	3.87	1.97	1.90	.00	3.87
C. Divorced and Separated					
White	4.03	.55	3.48	.24	4.27
Negro	2.61	1.42	1.19	.94	3.55
D. Widowed					
White	3.87	1.66	2.21	.08	3.95
Negro	##				

TABLE 19

CHI-SQUARE ANALYSIS, WHITE AND NEGRO PATIENTS, EDUCATION BY LENGTH OF HOSPITALIZATION (<30 DAYS, 30 DAYS OR MORE) AND AGE (IN YEARS)

		Quadratic	Linear	Quadratic-Linear	Deviations from Regression	Total
A.	< 25					
	White	11.22*	3.43	7.79*	.87	12.09*
	Negro	##				
B.	25-44					
	White	1.38	.63	.75	.03	1.41
	Negro	.49	.45	.04	.78	1.27
C.	45-64					
	White	1.34	.42	.92	.36	1.70
	Negro	.63	.29	.34	.96	1.59
D.	65 and Older					
	White	.68	.39	.29	.03	.71
	Negro	##				

TABLE 20

CHI-SQUARE ANALYSIS, WHITE AND NEGRO PATIENTS, EDUCATION BY LENGTH OF HOSPITALIZATION (<30 DAYS, 30 DAYS OR MORE) AND ADDICTIONS

		Quadratic	Linear	Quadratic-Linear	Deviations from Regression	Total
A.	No Addictions					
	White	7.14*	.35	6.79*	11.25*	18.38*
	Negro	2.15	1.91	.24	.92	3.07
B.	Alcohol					
	White	2.35	2.30	.05	.22	2.57
	Negro	1.10	.47	.63	.00	1.10

TABLE 21

CHI-SQUARE ANALYSIS, WHITE AND NEGRO PATIENTS, NUMBER OF PREVIOUS ADMISSIONS BY LENGTH OF HOSPITALIZATION (<30 DAYS, 30 DAYS OR MORE) AND SEX

	Quadratic	Linear	Quadratic-Linear	Deviations from Regression	Total
A. Males					
White	58.88*	36.79*	22.09*	3.02	61.90*
Negro	3.44	2.91	.53	2.02	5.46
B. Females					
White	68.63*	47.47*	21.16*	.95	69.58*
Negro	.96	.04	.92	4.67	5.63

TABLE 22

CHI-SQUARE ANALYSIS, WHITE AND NEGRO PATIENTS, NUMBER OF PREVIOUS ADMISSIONS BY LENGTH OF HOSPITALIZATION (<30 DAYS, 30 DAYS OR MORE) AND OCCUPATION

	Quadratic	Linear	Quadratic-Linear	Deviations from Regression	Total
A. Professional					
White	17.84*	17.56*	.28	.33	18.17*
Negro	##				
B. Clerical, etc.					
White	31.87*	17.75*	14.12*	.30	32.17*
Negro	.36	.32	.04	.47	3.83
C. Miscellaneous					
White	102.49*	73.89*	28.60*	3.49	105.98*
Negro	2.22	.11	2.11	8.05	10.27
D. Unemployed					
White	6.90*	3.98*	2.92	1.16	8.06
Negro	7.37*	6.42*	.95	.44	7.81

TABLE 23

CHI-SQUARE ANALYSIS, WHITE AND NEGRO PATIENTS, NUMBER OF PREVIOUS ADMISSIONS BY LENGTH OF HOSPITALIZATION (<30 DAYS, 30 DAYS OR MORE) AND DIAGNOSIS

		Quadratic	Linear	Quadratic-Linear	Deviations from Regression	Total
A.	Acute or Chronic Brain Syndrome					
	White	9.22*	7.40*	1.82	2.29	11.51*
	Negro	2.31	1.47	.84	1.44	3.75
B.	Psychotic Disorder (Excluding Schizophrenia)					
	White	2.17	1.79	.38	2.19	4.36
	Negro	9.82*	8.19*	1.63	.83	10.65*
C.	Psychoneurotic, etc.					
	White	11.88*	6.71*	5.17*	.47	12.35*
	Negro	##				
D.	Mental Deficiency					
	White	4.07	3.28	.79	1.72	5.79
	Negro	##				
E.	Schizophrenia					
	White	6.79	.08	6.71	6.40*	13.19*
	Negro	3.30	.01	3.29	11.30*	14.60*

TABLE 24

CHI-SQUARE ANALYSIS, WHITE AND NEGRO PATIENTS, NUMBER OF PREVIOUS ADMISSIONS BY LENGTH OF HOSPITALIZATION (<30 DAYS, 30 DAYS OR MORE) AND MARITAL STATUS

		Quadratic	Linear	Quadratic-Linear	Deviations from Regression	Total
A.	Single					
	White	47.15*	13.89*	33.26*	2.01	49.16*
	Negro	3.47	3.41	.06	1.97	5.44
B.	Married					
	White	104.28*	89.35*	14.93*	.44	104.72*
	Negro	2.74	.20	2.54	4.37	7.11
C.	Divorced and Separated					
	White	23.01*	16.82*	6.19*	2.25	25.26*
	Negro	3.34	2.79	.55	.51	3.85

TABLE 25

CHI-SQUARE ANALYSIS, WHITE AND NEGRO PATIENTS, NUMBER OF PREVIOUS ADMISSIONS BY LENGTH OF HOSPITALIZATION (<30 DAYS, 30 DAYS OR MORE) AND EDUCATION

	Quadratic	Linear	Quadratic-Linear	Deviations from Regression	Total
A. Elementary					
White	57.01*	32.35*	24.66*	.32	57.33*
Negro	.41	.40	.01	1.81	2.22
B. High School					
White	93.44*	73.36*	20.08*	1.05	94.49*
Negro	2.97	2.60	.37	2.94	5.91
C. College Training and Advanced Degrees					
White	18.73*	10.98*	7.75*	1.61	20.34*
Negro	##				

TABLE 26

CHI-SQUARE ANALYSIS, WHITE AND NEGRO PATIENTS, NUMBER OF PREVIOUS ADMISSIONS BY LENGTH OF HOSPITALIZATION (<30 DAYS, 30 DAYS OR MORE) AND ADDICTIONS

	Quadratic	Linear	Quadratic-Linear	Deviations from Regression	Total
A. No Addictions					
White	163.19*	114.19*	49.00*	.39	163.58*
Negro	5.55	.00	5.55*	3.95	9.50
B. Alcohol					
White	2.13	2.08	.05	5.68	7.81
Negro	7.78*	6.29*	1.49	1.11	8.89

length of hospitalization category with an increase in the "level" of the characteristic.

Results of Application of Bayes' Theorem

As the previous section indicates there is a marked interdependence among the variables selected for study. However, only one item of data, hospital to which admitted, was excluded from analysis at this stage. As far as the patients of this study are concerned, the hospital to which admitted is highly dependent on race and the region of the state from which admitted. The remaining items, sex, race, occupation, diagnosis, marital status, age, criminal charges, religion, type of commitment, education, number of previous admissions, addictions, and region of state from which admitted were retained for inclusion in the Bayes' formula despite their interdependence.

Equation 7 was first applied to predict length of hospitalization by using less than 30 days and 30 days or more as the two length of hospitalization categories, y_i. The equation was later applied using categories of less than 90 days and 90 days or more, less than 60 days and 60 days or more, and the categories less than 370 days and 370 days or more. Four categories were also used, less than 30 days, 30 through 59 days, 60 through 89 days, and 90 days or longer.

The 30 day cutting point was selected in order to determine how well the method could discriminate between patients who have a relatively short stay and those who stay longer. The 370 day cutting point was used to see how well prediction could be made when both length of stay categories were relatively long. The other cutting points were chosen arbitrarily and serve to indicate the ability of the method to predict intermediate lengths of stay.

The probability, P_{y_i}, of each duration of hospitalization category was computed by dividing the number of study patients remaining in the hospital for the length of time specified by the total number of study patients. The number of patients in each length of stay category and the corresponding probabilities are given in Table 27.

The proportion of patients within each length of hospitalization category possessing each of the study characteristics was computed. These proportions constitute the $P_{x_j \mid y_i}$. A length of hospitalization-characteristic matrix, displaying the P_{y_i} and the $P_{x_j \mid y_i}$, was constructed for each of the five sets of length of hospitalization categories. The matrix for the categories < 30 days and 30 days or more is shown in Table 28. The entries in the table are the $P_{x_j \mid y_i}$ and a decimal is understood to be to the left of each. For example, at the intersection of the column for the characteristic x_{10} and the row for length of hospitalization category 30 days or more, it is seen that the probability of characteristic x_{10} given a length of stay in the hospital of 30 days or more is .27.

For each of the 13,731 patients in the study, a probability for each length of stay category was computed. This was done with an IBM 1401 computer by determining the patient's characteristics from his record which was stored on magnetic tape. The probabilities associated with these characteristics were extracted from the length of hospitalization-characteristic matrix for each length of hospitalization category. These values were then used in Equation 7 to compute the probability of the patient's staying in the hospital less than 30 days and the probability of his staying 30 days or more.

TABLE 27

NUMBER OF PATIENTS AND INCIDENCE, P_{y_i}, FOR EACH LENGTH OF HOSPITALIZATION CATEGORY

Length of Hospitalization Category (In Days)	Number of Patients	Incidence P_{y_i}
< 30	1,773	.129
30+	11,958	.871
< 90	5,550	.404
90+	8,181	.596
<370	10,927	.796
370+	2,804	.204
< 30	1,773	.129
30-59	2,021	.147
60-89	1,756	.128
90+	8,181	.596
< 60	3,794	.276
60 or More	9,937	.724

TABLE 28

LENGTH OF HOSPITALIZATION-CHARACTERISTIC MATRIX

x_j	Characteristic	Length of Hospitalization and Incidence, P_{y_i}	
		< 30 days (y_1) P_{y_1} = .129	30+ days (y_2) P_{y_2} = .871
	Sex		
1	Male	72	51
2	Female	28	49
	Race		
3	White	89	87
4	Non-white	11	13
	Age (In years)		
5	< 25	21	13
6	25-64	67	74
7	65 and over	12	12
	Occupation		
8	Professional	08	06
9	Nonprofessional	92	94
	Marital Status		
10	Single	25	27
11	Married	44	42
12	Other	30	31
	Diagnosis		
13	Acute and Chronic Brain Syndrome	22	21
14	Psychotic (Excluding Schizophrenia)	05	12
15	Psychoneurotic	42	17
16	Mental Deficiency	04	06
17	Schizophrenia	21	43
18	Unknown	06	01
	Type of Commitment		
19	Voluntary	12	07
20	Physician's Certificate	63	85
21	Order of Detention	24	08

TABLE 28 - CONTINUED

x_j	Characteristic	Length of Hospitalization and Incidence, P_{y_i}	
		< 30 days (y_1) P_{y_1} = .129	30+ days (y_2) P_{y_2} = .871
	Criminal Charges		
22	Yes	80	95
23	No	20	05
	Education		
24	None	04	05
25	Elementary and High School	80	79
26	College and Advanced	09	09
27	Unknown	07	07
	Religion		
28	Protestant	71	78
29	Catholic	06	06
30	Other	08	06
31	None	15	10
	Previous Admissions		
32	None	55	41
33	1	29	33
34	2	08	15
35	3 or more	07	10
	Addictions		
36	None	75	87
37	Alcohol	23	12
38	Other	02	01
	Region		
39	Northwest, Morth Central, Northeast	27	31
40	West Central, East Central, Central	52	42
41	Southwest, South Central, Central Southeast, Southeast	22	27

Table 29 gives an example of the characteristics possessed by one of the patients in the study. Also shown in the table are the conditional probabilities associated with these characteristics given the two length of hospitalization categories.

Using Equation 7,

$$P_{y_i | x_1, x_2, \ldots x_m} = \frac{(P_{y_i})(P_{x_1 | y_i} \; P_{x_2 | y_i} \ldots P_{x_m | y_i})}{\sum_{i=1}^{k} (P_{y_i})(P_{x_1 | y_i} \; P_{x_2 | y_i} \ldots P_{x_m | y_i})}, \; i=1, \ldots k \quad (7)$$

and the probabilities of Table 29, Equation 19 yields a probability of .11 that the patient will remain in the hospital less than 30 days.

$$P_{<30 \text{ days} | x_1, x_2 \ldots x_m} = \frac{(.129)(.55)(.72) \ldots (.75)}{(.129)(.55)(.72) \ldots (.75) + (.871)(.41)(.51) \ldots (.87)}$$

$$= .11 \quad (19)$$

Equation 20 yields a probability of .89 that the patient will remain in the hospital 30 days or more.

$$P_{30+ \text{ days} | x_1, x_2, \ldots x_m} = \frac{(.871)(.41)(.51) \ldots (.87)}{(.129)(.55)(.72) \ldots (.75) + (.871)(.41)(.51) \ldots (.87)}$$

$$= .89 \quad (20)$$

On the basis of this information the prediction would be that the patient would remain in the hospital 30 days or more. The patient was actually in the hospital 181 days, so that in this case the prediction was correct.

A procedure like the one outlined above was carried out for each of the 13,731 patients in the study for each group of length of stay categories. In each case the predicted length of stay was compared with the patient's actual length of stay, and a record was kept of the number of

TABLE 29

EXAMPLE OF PATIENT CHARACTERISTICS AND ASSOCIATED CONDITIONAL PROBABILITIES

Patient's Characteristics	Conditional Probability of the Characteristic Given Length of Hospitalization Category	
	< 30 Days	30 Days or More
1. No previous admissions	.55	.41
2. Male	.72	.51
3. White	.89	.87
4. Clerical, etc. Occupation	.92	.94
5. Acute or Chronic Brain Syndrome	.22	.21
6. Married	.44	.42
7. 65 Years Old or Over	.12	.12
8. No Criminal Charges	.80	.95
9. Protestant	.71	.78
10. Commitment by Physician's Certificate or Court Certification	.63	.85
11. Elementary School Education	.80	.79
12. Admitted from West Central Region	.22	.27
13. No Addictions	.75	.87

cases which were correctly and incorrectly predicted in each length of stay category.

CHAPTER VI

DISCUSSION AND CONCLUSIONS

The mathematical model employed to predict category of length of mental hospitalization assumes three conditions. These conditions are 1) that the length of hospitalization categories be independent, 2) that the characteristics be independent, and 3) that the probabilities used in the predicting formula be the true probabilities. The first assumption was met in the present study, but it was found that the characteristics were not independent. Also, true probabilities were not available, but estimates based on a sample of 13,731 were used.

Despite the lack of independence among the characteristics, the model was employed with varying success, depending on the cutting points employed in making the various length of hospitalization categories. Table 30 shows the number of correct and incorrect predictions and the per cent of correct predictions for each length of stay category. When the length of hospitalization categories < 30 days and 30 days or more were used, 11,854, or 86 per cent of the total cases were correctly predicted with 27 per cent and 95 per cent being correctly predicted in the < 30 days group and the 30 days or more group respectively.

The overall percentage of correct predictions dropped to 75 when the less than 60 days, 60 days or more categories were used. The per cent correctly predicted in the less than 60 days group was 40, while 89 per

TABLE 30

NUMBER OF CORRECT AND INCORRECT PREDICTIONS FOR EACH LENGTH OF HOSPITALIZATION CATEGORY

Length of Hospitalization Category (In Days)	No. of Correct Predictions	No. of Incorrect Predictions	Total	Per Cent Correct Predictions	Overall Per Cent Correct Predictions
< 30	476	1,297	1,773	27	86
30 or More	11,378	580	11,958	95	
< 60	1,512	2,282	3,794	40	
60 or More	8,841	1,095	9,936*	89	75
< 90	2,665	2,885	5,550	48	
90 or More	6,734	1,446	8,180*	82	68
< 370	8,172	2,755	10,927	75	
370 or More	973	1,831	2,804	35	67
< 30	607	1,165	1,772*	34	
30-59	241	1,780	2,021	14	
59-89	0	1,756	1,756	0	61
90 or More	7,584	594	8,178*	93	

*These totals do not equal total reported in Table 7 because of "No decision" cases in which a patient was not predicted to fall in either category because of equal probabilities.

cent were correctly predicted in the 60 days or longer category.

When the categories of length of stay, < 90 days and 90 days or longer were used, only 68 per cent of the cases were correctly predicted. The per cent of correct predictions dropped to 67 for the less than 370 days and 370 days or more categories.

The poorest overall prediction ability occurred when four categories of length of stay were used. In this case the overall percentage correctly predicted was 61. For the less than 30 days group 34 per cent were correctly predicted and the percentages for 30-59 days, 59-89 days, and 90 or more days were 14, 0, and 93 respectively.

It is seen that the overall per cent of correct predictions decrease as the division point between the two length of hospitalization categories is raised. This may be explained by the fact that the difference between the probabilities associated with a given characteristic is greater in the case of the less than 30 days, 30 days or more categories. The probabilities associated with a given characteristic tend to be closer together for the other length of hospitalization categories. For example, the probability of being male, given length of hospitalization category, less than 30 days, is .72 and for 30 days or more is .51, a difference of .21. By comparison, the probability of being male given length of hospitalization category less than 90 days and given 90 days or more is .60 and .49 respectively, a difference of .11. The probability of being male given the category less than 370 days drops to .54, and given the category 370 days or more the probability of being male drops to .46. The difference between the two probabilities in this case is only .08. This is another way of saying that the difference between patients remaining in the hospital less than 30 days and patients remaining 30 days or more is greater than the difference between patients remaining in the hospital less than 90 days and those remaining 90 days or more. The difference between patients becomes less pronounced as the cutting point between categories is raised. In general, too, the probabilities associated with the length of hospitalization categories are more discrepant for the less than 30 days, 30 days or more categories than for any other set of categories used.

Table 30 also reveals that the highest accuracy of prediction occurs in those cases with the largest sample sizes. For example, in the $<$ 30 days, 30 days or more categories prediction was made with 95 per cent accuracy for the group with the largest number of patients, 11,958. On the other hand, the group with only 1,773 patients was predicted with an accuracy of only 27 per cent. This feature holds true for the other groups of categories listed in Table 30, and seems to confirm that sample size plays an important part in the accuracy of prediction. The relationship being, of course, that as the sample size from which the estimates of the true probabilities are derived increases, the estimates become more stable.

The results of the present study may be compared with those of Lindemann, et al (4) who were able to predict with an overall accuracy of 77 per cent, Johnston and McNeal (6) who were able to predict accurately up to 74 per cent of their cases, and Parnell and Skottowe (30) who predicted short and long stay cases with 69 and 84 per cent accuracy respectively.

It has been demonstrated that demographic data routinely collected on a group of patients admitted to the state supported mental hospitals of

Oklahoma can be used to predict the length of their hospitalization. Although the mathematical model employed assumes that the characteristics used for predicting are independent, its use, utilizing non-independent characteristics, afforded a means of predicting with an accuracy equal to or better than any method found reported in the literature.

One of the chief advantages of the method used in this study is that data already being collected and stored on punched cards are used. No additional effort is necessary on the part of the hospital in order to procure the information needed for constructing a length of hospitalization-characteristic matrix. Also available without extra effort are the data needed for determining the probabilities associated with each length of hospitalization category for individual patients. Most of the other studies concerned with predicting length of hospitalization have relied on psychological test scores and other data not usually collected on a routine basis (3, 6, 26, 30). A method of predicting length of hospitalization that does not require additional work at the data collection level should be more attractive to hospital officials and their overworked staffs than one requiring additional effort in order to obtain the information required.

However, if it were suspected that some characteristic, not currently being collected, would substantially improve the accuracy of prediction, data on this characteristic could be collected on a sample of patients and its usefulness in the prediction equation actually tested. It is conceivable that many items of information now being collected and considered of dubious value might be eliminated from the data collection schedule and replaced with an equal number of items which would be of great value in predicting a patient's length of hospitalization category.

A second advantage of the method described in this study is the rapidity with which the results of prediction can be made available to hospital administrators or mental health directors. This second advantage is a direct result of the first.

Because of the large number of calculations necessary for arriving at the probabilities associated with each length of hospitalization category for each patient, an electronic computer is necessary for solving Equation 7. If considerably fewer characteristics, say five or less, could be found which predicted with the desired accuracy, Equation 7, could be solved using a desk calculator. This would be practical if, in addition to using a small number of characteristics, only a few patients had to be processed each day.

When an electronic computer is used the length of hospitalization-characteristic matrix, along with program cards telling the machine what calculations to perform, can be stored in the machine's memory unit. The length of hospitalization-characteristic matrix may be updated from time to time as more information on patient characteristics is accumulated. In order to calculate the probabilities associated with each length of hospitalization category for each patient, a card for each patient can be fed into the machine. The program can be so adapted that the cards punched routinely can be used for this purpose. The machine can be instructed to determine the characteristics of each patient from his card, and consult the matrix stored in memory to find the conditional probabilities associated with each one for each length of hospitalization category. The probabilities can then be calculated by the machine and printed out along with the name of the patient. The length of hospitalization category for each patient, receiving the highest probability will be the predicted length of hospitalization for that patient.

If the data processing equipment were located at the hospital, each day's admissions could be processed at the end of the day or on the following day. This would make available at a very early date the predicted length of hospitalization for each patient. In most states data processing equipment, when available, is located in a central office which serves all hospitals in the system. In these cases the feedback of information would probably be a little slower, but a schedule could be developed whereby the information could be relayed back to the hospital in ample time for it to serve its purpose in administration and treatment.

CHAPTER VII

SUMMARY

This study has been concerned with investigating the feasibility of using routinely collected demographic data for predicting length of hospitalization of patients admitted to state mental hospitals. The literature was reviewed, and, although several studies on predicting length of mental hospitalization were found, most of them made use of psychological test scores and other data not usually available on a routine basis.

The study material consisted of 13,731 patients admitted to the state supported mental hospitals of Oklahoma. The study patients were described in considerable detail. The interrelationships of the characteristics of the patients and the associations between the characteristics and length of hospitalization were investigated. The techniques employed in this phase of the study were chi-square analyses, tests for linear and quadratic trend in contingency tables and tests for multiple regression in contingency tables. The test for linear trend is due to Cochran (39, 40, 41, 42). The tests for quadratic and multiple regression were developed as logical extensions of Cochran's method. These analyses revealed that all of the characteristics were associated with length of hospitalization and that there was a great deal of interdependence among the characteristics. The tests for trends indicated that in many cases the apparent association between a characteristic and length of hospitalization was due to a linear or quadratic trend in the proportions of patients falling in a particular length of hospitalization category with an increase in the "level" of the characteristic.

The characteristics selected for study were tested for independence and found to be, in general, highly dependent on each other.

The mathematical model for the method of prediction selected for investigation was Bayes' theorem. This technique makes possible the calculation of the probability of a patient's remaining in a hospital a certain length of time given that he possesses certain demographic characteristics. The uses of Bayes' theorem in the computer diagnosis of physical conditions such as congenital heart disease and Cushing's syndrome were reviewed. It was found that the technique has been used with considerable success in these areas.

When the Bayes' theorem approach was applied using the data of the present study it was found to be capable of predicting length of hospitalization with varying degrees of accuracy depending on the cutting points employed in making the various length of hospitalization categories. The method predicted with an overall accuracy of 86 per cent when the length of hospitalization categories were less than 30 days and 30 days or more. For the categories less than 60 days and 60 days or more, prediction was made with an overall accuracy of 75 per cent. The overall accuracy figures for the categories less than 90 days and 90 days or more and less than 370 days and 370 days or more were 68 and 67 per cent respectively. When four length of hospitalization categories, less than 30 days, 30 through 59, 60 through 89, and 90 days or more, were used the method predicted with an overall accuracy of 61 per cent.

It was pointed out that these figures compare favorably with other prediction studies found in the literature. It was also pointed out that the advantages of using the method of this study are the fact that use is made of data already being collected routinely and the speed with which the results can be obtained as a result of using data available at the time of the patient's admission.

REFERENCES

1. Patients in Mental Institutions 1962, Part II: State and County Mental Hospitals, Public Health Service Publication No. 1143, Part II, Washington, D. C., 1964.

2. Patient Movement Data, State and County Mental Hospitals, 1961, Public Health Service Publication No. 1187, Washington, D. C., 1964.

3. Anker, James M. "Chronicity of Neuropsychiatric Hospitalization: A Predictive Scale", Journal of Consulting Psychology, 25 (1961), pp. 425-432.

4. Lindemann, James E., Fairweather, George W., Stone, Gideon B., Smith, Robert S., and London, Ira T. "The Use of Demographic Characteristics in Predicting Length of Neuropsychiatric Hospital Stay", Journal of Consulting Psychology, 23 (1959), pp. 85-89.

5. Pugh, Thomas F. "The Population Explosion and the Future Number of Mental Hospital Inpatients", New England Journal of Medicine, 271 (1964), p. 672.

6. Johnston, Roy, and McNeal, Benjamin F. "Combined MMPI and Demographic Data in Predicting Length of Neuropsychiatric Hospital Stay", Journal of Consulting Psychology, 28 (1964), pp. 64-70.

7. Staudt, Virginia M., and Zubin, Joseph "A Biometric Evaluation of the Somatotherapies in Schizophrenia", Psychological Bulletin, 54 (1957), pp. 171-196.

8. Bond, Earl D. "A Review of the Five-Year Period Following Admission in One Hundred and Eleven Mental Patients", American Journal of Insanity, 77 (1921), pp. 385-394.

9. ____________ "Results in Two Hundred and Fifty-One Cases Five Years After Admission to a Hospital for Mental Disease", Archives of Neurology and Psychiatry, 6 (1921), pp. 429-439.

10. ____________ "Follow-up Work in Mental and Surgical Cases", The American Journal of Psychiatry, 2 (1923), pp. 445-449.

11. __________ "Underestimation of Good Results in Mental Disease", The Journal of the American Medical Association, 85 (1925), pp. 503-505.

12. __________, and Braceland, Francis J. "Prognosis in Mental Disease", The American Journal of Psychiatry, 94 (1937), pp. 263-274.

13. Fuller, Raymond G. "Expectation of Hospital Life and Outcome for Mental Patients on First Admission", Psychiatric Quarterly, 4 (1930), pp. 295-323.

14. Israel, Robert H., and Johnson, Nelson A. "New Facts on Prognosis in Mental Disease", The American Journal of Psychiatry, 104 (1948), pp. 540-545.

15. Kramer, Morton, Goldstein, Hyman, Israel, Robert H., and Johnson, Nelson A. "A Historical Study of the Disposition of First Admissions to a State Mental Hospital, Experiences of the Warren State Hospital During the Period 1916-1950", Public Health Monograph No. 32, United States Government Printing Office, Washington, D. C., 1955, pp. 1-25.

16. Dorn, Harold F. "Methods of Analysis for Follow-up Studies", Human Biology, 22 (1950), pp. 238-248.

17. Pollock, Earl S., Person, Philip H., Jr., Kramer, Morton, and Goldstein, Hyman "Patterns of Retention, Release, and Death of First Admissions to State Mental Hospitals, The Experience During the First 12 Months of Hospitalization of Patients Admitted to 11 State Hospital Systems in 1954", Public Health Monograph No. 58, United States Government Printing Office, Washington, D. C., 1959, pp. 1-53.

18. Kramer, Morton, Person, Philip H., Jr., Tarjan, George, Morgan, Richard, and Wright, Stanley W. "A Method for Determination of Probabilities of Stay, Release, and Death, for Patients Admitted to a Hospital for the Mentally Deficient: The Experience of Pacific State Hospital During the Period 1948-1952", American Journal of Mental Deficiency, 62 (1957), pp. 481-495.

19. Epstein, Leon J., and Morgan, Richard D. "Trends in Release Rates of Schizophrenic Patients", Comprehensive Psychiatry, 2 (1961), pp. 196-202.

20. Locke, Ben Z. "Outcome of First Hospitalization of Patients With Schizophrenia", Public Health Reports, 77 (1962), pp. 801-805.

21. Malzberg, Benjamin "Rates of Discharge and Rates of Mortality Among First Admissions to the New York Civil State Hospitals", Mental Hygiene, 36 (1952), pp. 104-120.

22. ______________ "Rates of Discharge and Rates of Mortality Among First Admissions to the New York Civil State Hospitals", Mental Hygiene, 36 (1952), pp. 618-638.

23. ______________ "Rates of Discharge and Rates of Mortality Among First Admissions to the New York Civil State Hospitals", Mental Hygiene, 37 (1953), pp. 619-654.

24. ______________ "Cohort Studies of Mental Disease in New York State, 1943 to 1949", Mental Hygiene, 40 (1956), pp. 450-479.

25. ______________ Cohort Studies of Mental Disease in New York State, 1943 - 1949. New York: National Association for Mental Health, 1958, p. 113.

26. Meeker, Frank O. An Exploratory Study of Specific and General Personality Dimensions Related to Length of Hospitalization Among Psychiatric Patients, Unpublished Master's Thesis, University of California, 1958.

27. Geidt, F. H., and Schlosser, J. R. Movement of Patients Through a Neuropsychiatric Hospital, Unpublished Manuscript, Medical Library, Veterans Administration Hospital, Perry Point, Maryland, 1955.

28. Moran, Louis J., Fairweather, George W., Morton, Robert B., and McGaughran, Lawrence S. "The Use of Demographic Characteristics in Predicting Response to Hospitalization for Tuberculosis", Journal of Consulting Psychology, 19 (1955), pp. 65-70.

29. Hathaway, S. R., and McKinley, J. D., The Minnesota Multiphasic Personality Inventory, Minneapolis: University of Minnesota Press, 1943.

30. Parnell, R. W., and Skottowe, Ian "Length of Stay in Mental Hospitals and Some Factors Influencing It", British Medical Journal, 2 (1959), pp. 1296-1300.

31. State of Oklahoma, Department of Mental Health, Policy Manual.

32. Parzen, Emanuel. Modern Probability Theory and Its Applications. New York: John Wiley and Sons, Inc., 1960, p. 119.

33. Warner, Homer R. "Medical Diagnosis Using a Digital Computer", Proceedings of the Third IBM Medical Symposium, 3 (1961), pp. 293-307.

34. ______________, and Stephenson, Robert "A Mathematical Approach to Medical Diagnosis, Application to Congenital Heart Disease", The Journal of the American Medical Association, 177 (1961), pp. 177-183.

35. _______________, Toronto, Alan F., and Veasey, L. George "Experience With Bayes' Theorem for Computer Diagnosis in Congenital Heart Disease", Annals of the New York Academy of Science, 115 (1964), pp. 558-567.

36. Nugent, Charles A., Warner, Homer R., Dunn, John T., and Tyler, Frank H. "Probability Theory in the Diagnosis of Cushing's Syndrome", The Journal of Clinical Endocrinology and Metabolism, 24 (1964), pp. 621-627.

37. Rinaldo, Joseph A., Jr., Scheinok, Perry, and Rupe, C. E. "Symptom Diagnosis. A Mathematical Analysis of Epigastric Pain", Annals of Internal Medicine, 59 (1963), pp. 145-154.

38. Mount, Joseph F., and Evans, John W. "Computer Aided Diagnosis - A Simulation Study", Proceedings of the Fifth IBM Medical Symposium, 5 (1963), pp. 113-127.

39. Cochran, William G. "Some Methods for Strengthening the Common χ^2 Tests", Biometrics, 10 (1954), pp. 417-451.

40. _______, and Cox, Gertrude M. Experimental Designs, Second Edition. New York: John Wiley and Sons, Inc., 1957, p. 105.

41. Steel, Robert G. D., and Torrie, James H. Principles and Procedures of Statistics. New York: McGraw-Hill Book Company, Inc., 1960, pp. 381-383.

42. Hill, A. Bradford Principles of Medical Statistics, Seventh Edition. New York: Oxford University Press, 1961, pp. 176-178.