Statistical Science
1987, Vol. 2, No. 2, 107–134

Statistics in Psychiatry

B. S. Everitt

Abstract. Statistics is a most important basic science for psychiatry. More and more often psychiatric researchers resort to sophisticated and powerful statistical techniques to help them unravel the complexities of their data. In this paper, the relationship between the two disciplines and their members is discussed in the context of the use of statistics in psychiatry in general.

Key words and phrases: Cox regression, history, psychiatric statistics, statistical consultancy, teaching.

1. INTRODUCTION

A widely quoted remark of Galton is that until the phenomena of any branch of knowledge have been submitted to measurement and number, it cannot assume the dignity of a science. Psychiatry has for the last 40 to 50 years struggled to attain "the dignity of a science" by submitting its observations to measurement and quantification, and psychiatrists have become increasingly aware that for their discipline to progress requires a strict scientific approach. Allied to this has been a growing appreciation of the need for some type of *statistical* evaluation of the data collected. A consequence of this change in approach and attitude is reflected in the increase in statistical content of most psychiatric journals, so that currently the majority of published papers contain at least some statistical analyses; Table 1 taken from DeGroot and Mezzich (1985) shows the usage of various statistical techniques in several well known psychiatric journals. (The question of whether or not these techniques are always used appropriately will be taken up later in the paper.)

Clearly psychiatrists need to learn at least a little about the methods of statistics simply to be able to evaluate the psychiatric literature critically; in addition, however, those engaged in research may need a fairly firm grasp of a number of techniques such as t tests, χ^2 tests, and the analysis of variance. For many psychiatrists, the need to acquire such knowledge will not be a particularly appealing way of spending their time; statisticians should not find this surprising—after all how can a description of the appropriate way to do a t test compare with a consultant psychiatrist's anecdotes about a patient who can constantly hear a Scottish piper playing or the man who keeps mistaking his wife for a hat?

Nevertheless, psychiatry has become more dependent on statistical techniques both simple and sophisticated in the evaluation of its data, and psychiatrists ignore such a development at their peril. In this paper I shall consider the use of statistical methods in psychiatry, and the difficult problem of how statistics should be taught to psychiatrists. To begin, however, I shall briefly consider the historical background of the use of statistics in psychiatry.

2. HISTORICAL BACKGROUND

In the nineteenth century and earlier, the use of statistics in psychiatry was largely restricted to simple descriptive measures, and it is only in the second half of the twentieth century that the use of inferential and other more complex statistical methods has become widespread. But the use of even simple descriptive statistics was important and their presentation often led to changes in policy if not to changes in attitude to the problems of lunacy. For example, Table 2, taken from Scull (1979), shows the number of people officially identified as insane and the rate of insanity per 10,000 people in England and Wales at various times during the nineteenth century. The increase in lunacy as suggested by these figures became one of the main weapons in reformers' arguments for new legislation to deal with the insane, since they indicated that insanity was now a serious social problem, a view endorsed by the following from the 1844 Report of The Metropolitan Commissioners on Lunacy.

> Lunatics have unfortunately become so numerous throughout the whole kingdom, that the proper construction and cost of asylums for their use has ceased to be a subject which affects a few counties only, and has become a matter of national interest and importance.

B. S. Everitt is Reader in Statistics in Behavioural Science, and Head of the Biometrics Unit, Institute of Psychiatry, University of London, Denmark Hill, London, SE5 8AF, United Kingdom.

TABLE 1
American Journal of Psychiatry (AJP), British Journal of Psychiatry (BJP), and Archives of General Psychiatry (AGP) during 1980 by categories of statistical usage

Categories of statistical usage	AJR (1980) 339 papers	BJP (1980) 148 papers	AGP (1980) 110 papers
1. Expository, literature review, etc.	18 (5.3%)	6 (4.1%)	4 (3.6%)
2. No statistical data: case reports, etc.	115 (33.0%)	12 (8.1%)	2 (1.9%)
3. Descriptive statistics only: tables, graphs, means, variances	65 (19.2%)	14 (9.5%)	11 (10.0%)
4. ! and *t* tests, Fisher exact test: 1 or 2 samples, contingency tables	95 (28.0%)	75 (50.7%)	66 (60.0%)
5. Product-moment correlations, rank correlations	42 (12.4%)	22 (14.9%)	30 (27.3%)
6. Analysis of variance, *F* tests: 1-, 2-, and higher-way	25 (7.4%)	22 (14.9%)	32 (19.1%)
7. Nonparametric rank methods (other than rank correlations)	9 (2.7%)	11 (11.5%)	10 (9.1%)
8. Measures of association and agreement (other than correlation)	10 (2.9%)	13 (8.8%)	9 (8.2%)
9. Regression analysis: simple, multiple polynomial stepwise	6 (1.8%)	9 (6.1%)	10 (9.1%)
10. Discriminant and factor analyses	4 (1.2%)	6 (4.1%)	7 (6.4%)
11. Estimation: maximum likelihood, interval estimation, etc.	0 (0.0%)	3 (2.0%)	2 (1.8%)
12. Cluster analysis, classification	1 (0.3%)	0 (0.0%)	1 (0.9%)
13. Life tables, life testing, survival analysis	1 (0.3%)	0 (0.0%)	2 (1.8%)
14. Time-series analysis, spectral analysis	2 (0.6%)	0 (0.0%)	1 (0.9%)
15. Classical experimental design: Latin squares, hierarchical models	0 (0.0%)	3 (2.0%)	1 (0.9%)
16. Bayesian methods	0 (0.0%)	0 (0.0%)	1 (0.9%)

From DeGroot and Mezzich (1985).

Scull points out, however, that the achievement of reform (the construction of asylums and the employment of doctors to affect the cure of the insane) did *not* bring about a halt or even a diminution in the rapid upward spiral of cases of lunacy. Between 1844 and 1860, when the population as a whole grew by just over 20%, the number of lunatics all but doubled; and the growth in the number of the insane continued to far outstrip the rate of increase of the general population for the rest of the century. Scull discusses the various "official" explanations of the increase, one of which was that a large number of cases previously unreported had only recently been brought under observation because the methods of gathering statistics on insanity had previously been slipshod and inadequate; the apparent increase was therefore considered to be largely a statistical artefact. Others, however, preferred different explanations for the increase, assuming it to be real and attributable to stresses attendant upon life in a higher "mechanical" civilization. (Not so very different from explanations of increases in neuroses and depression in the second half of the twentieth century!)

Examination of early issues of the *Journal of Mental Science* (the forerunner of today's *British Journal of Psychiatry*) show an early reference to "statistical proof" in Matt's (1913) investigation of the heritability of insanity; in this case, however the "proof" derived simply from the presentation of a set of data. Inferential statistics in the form of t tests and χ^2 tests entered the psychiatric research scene only at a later date.

An indication of the psychiatrists attitude to mathematical and statistical topics in the early part of the twentieth century may perhaps be gleaned from Edward Mapother's comments when reviewing Spearman's book *The Abilities of Man: Their Nature and Measurement* in the *Journal of Mental Science* in 1928:

> "Doubtless most readers of this JOURNAL, like the reviewer, will be content to take for granted the mathematics involved."

Nevertheless statistical techniques began to appear in the JOURNAL around this time. Cameron (1933) in a study of perseveration used the correlation coefficient and a test to assess its significance. A report by the Royal Medico-Psychological Association Committee on Mental Deficiency on the incidence of neuropathic conditions in the relatives of normal persons published in the JOURNAL in 1937 used a t test to examine the difference in average family size for two groups of families, those who included a weak-minded person and those not including such a person. (The t test was explained in detail in an appendix to the paper.) The same paper also contains an application of a one-way analysis of variance. Masserman and Carmichael (1938) in an investigation of diagnosis and prognosis in psychiatry use Pearson's contingency coefficient to assess the relationship between diagno-

TABLE 2
Total population, total number officially as insane, and rate of insanity per 10,000 people in England and Wales

1 Jan.	Population	No. officially identified as insane[a]	Rate per 10,000	Source of data on number insane
1807	9,960,000	2,248	2.26	House of Commons 1807
1819	11,106,000	6,000	5.40	Burrows 1820
1828	13,106,000	8,000	6.10	Halliday 1828
1829	13,370,000	16,500	12.34	Halliday 1829
1836	14,900,000	13,667	9.18	Parliamentary Return 1836
1844	16,480,000	20,893	12.6	Metropolitan Commissioners on Lunacy
1850			Not available	
1855	18,786,914	30,993[b]	16.49	Commissioners on Lunacy
1860	19,902,713	38,058	19.12	Annual Reports
1865	21,145,151	45,950	21.73	
1870	22,501,316	54,713	24.31	
1875	23,944,459	63,793	26.64	
1880	25,480,161	71,191	27.94	
1885	27,499,041	79,704	28.98	
1890	29,407,649	86,067	29.26	

From Scull (1979).
[a] Includes lunatics in asylums, but also those in workhouses, at large in community, etc.
[b] The Commissioners found 20,493 luncatics in asylums of all types in 1855; lacking a complete enumeration of all lunatics not so confined, they estimated that these amounted to some 10,500 persons. (Commissioners on Lunacy Annual Report 1855, Volume 9, page 39.)

sis and prognosis; the same year saw an increase in the use of statistical techniques in general, following perhaps the publication in 1937 of Bradford Hill's classic text *Principles of Medical Statistics.*

Nowadays the majority of papers in the *British Journal of Psychiatry* and in similar psychiatric journals contain the results of applying one or another statistical technique. Littered among the t tests, χ^2 tests, and F tests, it is not uncommon to find the results from a stepwise regression, or from fitting a log linear model or from applying discriminant function analysis. By taking advantage of the powerful statistical packages now available, psychiatrists are clearly using more sophisticated statistical tools in their research. Whether these are always used appropriately and the possible problems that can arise are discussed in the following two sections.

3. STATISTICAL METHODS AND PSYCHIATRIC RESEARCH

Even a cursory examination of current psychiatric journals will indicate the extent to which psychiatric researchers have taken on board a variety of statistical tools, so much so that it is difficult nowadays to find an article not liberally scattered with r's, t's, χ^2, and $p <$ or *, **, *** (a nomenclature regarded by Sprent (1970) as more suitable for a hotel guidebook than a serious scientific paper!). There is some evidence to suggest that such techniques are not always used wisely. For example, White (1979) considered 12 issues of the *British Journal of Psychiatry* from July 1977 to June 1978. Of the 168 papers published in this period, 139 contained numerical results and 103 contained some form of statistical analysis. In 47 papers there were major statistical errors, these being defined as errors which could potentially affect at least one conclusion. The errors ranged from the failure to give any measure of dispersion when summarizing data to failure to use a *correlated t* test for matched data.

Table 1 indicates that the most commonly used statistical techniques are clearly the simple significance tests such as the t test for the equality of two means or the χ^2 test for the equality of two proportions; in addition, correlation coefficients are often given, together with the results of a test that the population value is zero. The use of such tests has become so routine that psychiatric researchers now rarely stop to ask whether they are appropriate in their particular study; indeed, they probably feel that the presence of such tests is essential to make their study and the subsequent papers scientifically respectable. The criticisms of significance tests frequently encountered in the statistical and psychological literature (see for example, Rozeboom, 1960) often seem to have by-passed or simply have been ignored by psychiatrists. The problem is that researchers in the area continue to use significance tests irrespective of the type of sample, type of research problem, or type of research design. In addition to such important technical errors, fundamental errors in the philosophy of science are frequently involved in the indiscriminate use of the tests. (It is not uncommon to come across claims that a "hypothesis has been shown to be false" or that a "hypothesis has been proved to be true.") Many psychiatrists (like psychologists before them) still seem convinced of the "sacred" nature of 0.05 and cling to the language of "acceptance" and "rejection" even when the empirical phenomena under investigation are continuous in nature. As Skipper, Guenther and Nass (1967) comment:

> "The current obsession with 0.05, it would seem, has the consequence of differentiating significant research findings from those best forgotten, published studies from unpublished ones, and renewal of grants from termination. It would not be difficult to document the joy experienced by a social scientist when his F ratio or t-value yields significance at 0.05 nor his horror when the table reads "only" 0.10 or 0.06. One comes to internalize the difference between 0.05 and 0.06 as "right" versus "wrong," "creditable" vs "embarrassing," "success" vs "failure.""

It would clearly be an advantage if psychiatrists could be persuaded to use terms such as "support," "lack of support," "weak support," "strong support," etc. when discussing results, since in science, adjustment of degree of belief based on the strength of evidence rather than firm decision is the appropriate response to any set of observations. Trainee psychiatrists might with advantage be referred to Lyttleton's simple "bead-on-a-wire" model for the correct scientific attitude to adopt in relation to each and every hypothesis of interest (see Lyttleton, 1977). Psychiatrists (*and* the editors of psychiatric journals) might be well advised to reappraise their attitude to the significance test. Current thinking in statistics tends to be away from such tests to the more informal methods of exploratory data analysis advocated by Tukey and his colleagues (see Tukey, 1977). Many of these techniques are of a graphical nature, in line with the opinion expressed by Chambers et al. (1983) that "There is no single statistical tool that is as powerful as a well-chosen graph."

Although the above is, in many ways, an argument in favor of a more simplistic (but more thoughtful) approach to data analysis, particularly the avoidance of the *automatic* use of significance tests, there are situations where an *increase* in the sophistication of the statistical tools employed would be appropriate and in some cases necessary to avoid misleading conclusions. Two examples illustrating this point will be discussed in the next section.

One area where psychiatrists have used a variety of relatively sophisticated statistical techniques is that of the *classification* of the mentally ill. Despite reservations by many psychiatrists about the worth of any such classification, it remains an area of great interest and one in which researchers have employed a variety of statistical techniques including *factor analysis, discriminant function analysis*, and *cluster analysis* in an effort to refine or even redefine current diagnostic labels. Of particular interest has been the lively and occasionally heated debate as to whether depressed patients are of two types (usually labeled *endogeneous* and *neurotic*) or simply form a continuum. Investigations by Kiloh and Garside (1963), Carney, Roth and Garside (1965), Sandifer, Wilson and Green (1966), and Pilowsky, Levine and Boulton (1969) support the former view, while those of McConaghy, Joffe and Murphy (1967), Kendell (1960), and Kendell and Gourlay (1970) indicate that the latter is more likely. Much of the argument concerns the appropriate use of various statistical techniques, particularly factor analysis and cluster analysis, and Eysenck (1970) and Garside and Roth (1978) have attempted to clarify the issues involved. Eysenck points to the confusion in many of the studies in this area caused by the failure to separate arguments for a one- or two-factor model for depressive symptoms, from those concerned with a dimensional versus categorical view of psychiatric illness. He fails to mention, however, the possibility of arriving at misleading results from a factor analysis when the data consist of distinct groups of patients; such a possibility arises because, in such cases, the correlation matrix of the data as a whole will not necessarily reflect those of each group. Consequently, it might be argued, that in these studies, an examination of the data by some form of cluster analysis is a logical prior step to the application of factor analysis. Unfortunately, the last two decades have seen a vast proliferation of clustering techniques without a parallel increase in understanding of their properties or their problems; consequently, practical applications of the methods reported in the literature in general, and the psychiatric literature in particular, remain less than satisfactory (see Everitt (1980) for a detailed account of the area).

Recent years have seen an increase in the use of statistical techniques such as logistic regression, log linear models, and structural equation models in psychiatric research. For example, Dunn (1981) uses a linear logistic model to investigate patterns of psychotropic drug consumption; Tennant and Bebbington (1978) use log linear models to assess a "vulnerability" model of depression originally proposed by Brown and Harris (1978) and Fergusson and Horwood (1984) use structural equation modeling to examine the relationship between life events and depression in women.

Such sophisticated and powerful techniques can now be used routinely thanks to the availability of statistical packages such as SPSS and SAS. The "user friendly" manuals of such packages means that they can be used by those who are relatively untutored in statistics. Unfortunately, they can as easily be *misused*, a danger which has been commented on several times previously, see for example, Hooke (1980) and Hand and Everitt (1987).

Psychiatrists may complain that they have had to resort to using such packages themselves simply because the statistican they consulted was not willing to undertake the analyses required. After all, a statistician who can recommend only looking at a few histograms or scatterplots and who keeps mumbling something about what are the hypotheses of interest, is really of very little use when what one clearly needs is a factor analysis of the 500-item questionnaire you have just finished administering to 100 severely depressed patients!

Faced with such an attitude it is easy to see why statisticians occasionally feel like retreating to their ivory towers, or suggest producing manuals for statistical packages that are anything but "user friendly." (A number of statisticians do in fact appear to have achieved this goal!) Such a reaction, while understandable, is likely to be counterproductive and statisticians will improve the situation only by showing even greater patience and effort to communicate than they have in the past. An obvious way to begin is to improve the statistics teaching received by psychiatrists and a number of suggestions are made in Section 5. Before this, however, I shall discuss two psychiatric examples which were originally analyzed using relatively simple statistical methods, but for which a more sophisticated approach seemed more appropriate (at least in the opinion of the author!).

4. TWO EXAMPLES OF THE USE OF COX'S REGRESSION IN PSYCHIATRY

An increasingly widely used technique in applied statistics is an adaptation of multiple regression for use with censored survival data first suggested by Cox (1972). Essentially this method seeks to assess the relationship between the distribution of survival time and a number of explanatory variables such as age, sex, treatment group, etc., by modeling the log transformed *hazard function* as a linear function of the explanatory variables. A relatively simple description of the method is given in Allison (1984). Here I would like to briefly describe two applications of the technique to psychiatric data.

4.1 Clinical Trial of Bromocriptine

This trial involved a comparison of a "high" with a "low" dose regimen of bromocriptine in the treatment of Parkinson's disease. During the first 6 months of the trial, patients randomly allocated to one or the other of the two treatment groups were given increasing amounts of the drug according to some prescribed schedule and rated by a neurologist on a clinical rating scale designed to assess their disability. Incremental dosing continued until the patient had achieved a 33% improvement in his or her clinical rating scale from the average of two pretreatment assessments. More details of the trial are given in the first report of the U.K. Bromocriptine Research Group (1987).

As with many such studies, the involvement of a statistician (in this case the author) only began after the trial had been running for a considerable time, and after a great deal of data had already been collected. Consequently, my reservations about some aspects of the trial's design and about the reliability and validity of the measuring instrument had to be placed firmly in the background so that I could concentrate on analyzing the data actually available. One of the questions of interest was whether, at the end of 6-month treatment, the average time required to reach the improvement criterion differed in the two groups. Initially this has been assessed by a simple t test with the results given in Table 3.

Such an analysis however fails to account for the different "drop out" rates in the two groups, or for differences in the proportion of patients who, at 6 months, had failed to achieve the 33% improvement criterion. The relevant figures are shown in Table 4.

Consequently a reanalysis of the data was undertaken using a Cox's regression model (see Cox, 1972), using treatment group as one of a number of explanatory variables of interest, and counting drop outs due to severe side effects and nonresponders as "censored" observations for which we know only that the time needed to reach the improvement criterion is longer than the time at which they dropped out or greater than 6 months (for nonresponders.) The results of this reanalysis are summarized in Table 5. Now we find a significant difference between the two groups.

In order to find an estimate of the average time to reach the improvement criterion for each group, it was

TABLE 3

Comparison of the time in weeks needed to achieve a 33% improvement in clinical rating score in the slow and fast treatment regimes considering only patients who reached this criterion in the 6 months of the trial

Group	No. of patients	Mean	SD	t test value	d.f.	p
Slow	37	11.64	7.31	1.00	62	0.3
Fast	28	10.00	5.43			

TABLE 4

Comparison of treatment groups at end of 6 months of the trial

	Slow	Fast
No. of patients	80	54
Drop outs due to severe side effects	21	19
Drop outs due to noneffects	10	2
No. not reaching improvement criterion	12	5

TABLE 5
Cox's regression for time taken to reach improvement criterion

Explanatory variables considered
1. Age
2. Sex
3. Treatment group
4. Years since contracting disease
5. Initial clinical rating score

A backward elimination procedure leaves only *treatment group* with $\chi^2 = 4.84 \quad p = 0.0279$

TABLE 6
Estimates of mean time to reach improvement criterion allowing for censored observations

Group	Mean	SD	Approximate 95% CI
Slow	28.7	4.7	19.4–37.9
Fast	22.3	4.2	14.0–30.6

assumed that time to improvement was exponentially distributed (an examination of histograms, etc. showed that this was not a totally unrealistic assumption); the maximum likelihood estimate of the mean time to improvement is then given by

$$\hat{\mu} = \frac{1}{r}\left[\sum_{i=1}^{r} t_i + \sum_{i=1}^{n-r} t_i^+\right] \tag{1}$$

where n is the total number of patients, r is the number of patients who reach the improvement criterion within 6 months, and $t_1, t_2, \ldots, t_r$ are their times to improvement; $t_1^+, \ldots, t_{n-r}^+$ are the censoring times of the $n - r$ patients who failed to reach the improvement criterion within 6 months. The results are shown in Table 6. These estimates are likely to be far more realistic than the simple averages of uncensored observations given in Table 3.

4.2 Length of Stay of Patients in Broadmoor

The second example I would like to consider involves an investigation of the factors associated with the length of stay of mentally ill patients committed, by the courts for a variety of offenses, to Broadmoor Hospital in the United Kingdom. Information about such patients collected on admission includes social background, criminal record, and diagnosis, in addition to variables such as age, sex, and country of birth. In this study the researchers considered male patients admitted in the years 1972–1974 for whom at least 8 hours had elapsed by the time the release data were collected in 1982.

The analyses reported in the original description of the study began by categorizing length of stay into short (4 years or less), intermediate (between 4 and 8 years), and long (over 8 years). These three groups were then compared with respect to the various explanatory variables to see on which they differed. Such comparisons were made by way of simple χ^2 tests taking each explanatory variable in turn. Those patients with a Mental Health Act classification of psychopathic disorder were considered separately from other patients. The results indicated that the three length of stay groups did not differ on variables such as age, early family history, social class, or previous psychiatric hospitalization. Differences, however, were found with respect to type of offense for which the men had been admitted and in terms of motive for the crime. In particular for the psychopathic disorder patients, the long stay group was made up largely of men who had committed offenses of violence against the person, sexual or otherwise, while those men whose offenses did not physically injure other people accounted for the majority of the short stay group. In addition, for the psychopaths, the investigators found that where a sexual motive had been attributed to the offense, the patient was very unlikely to be in the short stay group, and where the psychiatrist at admission had recorded the offense as being "apparently motiveless," the patient was almost always among those in the long stay group. In the case of the remaining patients, only whether or not the offense was deemed to have a sexual motive related to length of stay.

The investigators duly produced a paper describing their study and their results and submitted it to a well known psychiatric journal. My involvement began when I received the paper to referee, and suggested an analysis using actual length of stay as the dependent variable in a Cox's proportional hazards model with observations involving patients still in detention at the time of the study being considered as censored. Fortunately, the researchers involved occupied rooms almost opposite mine and so it was not difficult to obtain their data and carry out my suggested analysis! (Not all explanatory variables considered in the original analysis could be used in my reanalysis, however, making a direct comparison of results somewhat difficult.)

The explanatory variables considered in the model are described in Table 7. Type of offense was recorded as a series of dummy variables. The regression coefficients for the psychopathic disorder patients are shown in Table 8. The results of a backward elimination procedure are shown in Table 9. For the psychopaths then we see that the results are in partial agreement with those detailed earlier except that type of offense is not included in the model. The most likely explanation for this difference is that the variable sexual motive acts as a surrogate for the important components of type of offense. (In fact the sexual motive variable appears to contain a lot of the information contained in other variables since *not* including it as an explanatory variable leads to a model for the psychopathic patients which includes arson versus other types of offense, age, and age at first admission as significant predictors.)

The results of similar analyses for the remaining patients are shown in Tables 10 and 11. Here again sexual motive is found to be the most important explanatory variable, although age at admission is almost significant at the 5% level. (Not, of course, that we regard 0.05 as sacred!) (See Section 3.)

A second analysis was performed in which the diagnosis of a patient, psychopath or not, was coded as a dummy variable and included as an explanatory variable. In addition two "interaction" variables were included, sexual motive × group and motiveless × group. The results of a backward elimination procedure are shown in Table 12. The first interaction variable is not included in the final model indicating that sexual motive is an equally important predictor

TABLE 7
Variables involved in length of stay investigation

Dependent variable
  Length of stay in days
Explanatory variables
  1. Type of offense
    Homicide
    Sexual offenses
    Wounding, assault
    Arson
    Other
  2. Country of birth (COB)
    0 = British
    1 = Other
  3. Age in years (AGE)
  4. Age at first court appearance (AGEFA)
  5. Number of previous court appearances (NOPCT)
  6. Number of previous admissions to psychiatric hospitals (NOPPH)
  7. Age at first admission (AGEFAD)
  8. Sexual motive (SM)
    0 = yes
    1 = no
  9. Motiveless (MOTL)
    0 = apparently motiveless
    1 = some motive for crime
Type of offense was coded in terms of the following four dummy variables
  Homicide
    (TO1 = 1, TO2 = 0, TO3 = 0, TO4 = 0)
  Sexual offenses
    (TO1 = 0, TO2 = 1, TO3 = 0, TO4 = 0)
  Wounding, assault
    (TO1 = 0, TO2 = 0, TO3 = 1, TO4 = 0)
  Arson
    (TO1 = 0, TO2 = 0, TO3 = 0, TO4 = 1)
  Other
    (TO1 = 0, TO2 = 0, TO3 = 0, TO4 = 0)

TABLE 8
Regression coefficients in Cox's proportional hazards model for psychopathic patients

Explanatory variable	Regression coefficient	Standard error
TO1	−0.818	0.727
TO2	−0.561	0.653
TO3	0.396	0.728
TO4	0.514	0.656
COB	0.429	0.774
AGE	−0.095	0.054
AGEFA	0.049	0.028
NOPCT	−0.024	0.060
NOPPH	−0.041	0.076
AGEFAD	0.024	0.020
SM	1.881	0.595
MOTL	1.904	0.570

Information is based on 68 observations and 42 uncensored observations.

TABLE 9
Backward elimination analysis for psychopathic patients

Explanatory	Regression	Standard error	χ^2
SM	1.562	0.355	19.31
MOTL	1.607	0.493	10.60

Test for all variables not in the model: $\chi^2 = 15.65$ with 10 d.f. and $p = 0.11$.

TABLE 10
Regression coefficients in Cox's proportional hazards model for nonpsychopathic patients

Explanatory variables	Regression coefficient	Standard error
TO1	−1.396	0.553
TO2	−0.998	0.477
TO3	−8.573	24.074
TO4	−1.542	0.593
COB	0.751	0.312
AGE	0.028	0.013
AGEFA	−0.008	0.012
NOPCT	0.002	0.028
NOPPH	0.004	0.034
AGEFAD	−0.023	0.009
SM	0.720	0.548
MOTL	−0.155	0.298

Information is based on 79 observations and 61 uncensored observations.

TABLE 11
Backward elimination analysis for nonpsychopathic patients

Explanatory variable	Regression coefficient	Standard error	χ^2
SM	1.072	0.521	4.23
AGEFAD	−0.019	0.009	3.84

Test for all variables not in the model: $\chi^2 = 14.20$ with 10 d.f. and $p = 0.16$.

TABLE 12
Results of a backward elimination analysis on length of stay data with group membership and interaction variables included

Explanatory variable	Regression coefficient	Standard error	χ^2
SM	1.312	0.297	19.50
Group × MOTL	0.517	0.224	5.32

Information is based on 147 observations and 103 uncensored observations. Test for all variables not in the model: $\chi^2 = 18.39$ with 12 d.f. and $p = 0.10$.

for both groups. The second interaction variable is included and indicates that among psychopathic disorder patients, attributing a motive to their offense is generally associated with a shorter stay. This analysis, of course, does little more than confirm the results found from the analyses of the separate groups; it was, however, useful in demonstrating to the researchers involved how such models can accommodate information on group membership.

The results of the original analysis (described in Dell, Robertson and Parker, 1986) and those obtained from Cox's regression are not strikingly different in this case, although I feel that the latter arise from a more appropriate approach and that in other situations the differences could be more important. In practical terms, the results simply indicate that the discharge decision makers tend to use predominantly the motive of a patient's crime when making their decision as to whether to release the patient or not.

5. TEACHING STATISTICS TO PSYCHIATRISTS

The content of previous sections of this paper clearly indicate that at least one statistics course should be part of every psychiatrist's training. For reasons hinted at earlier, this is likely to be regarded by the psychiatrist more as a necessary evil than an eagerly awaited treat! For the statistician unlucky enough to have drawn the short straw, teaching such a course is unlikely to be a great source of enjoyment or satisfaction!

Certainly if the traditional option of 10 to 20 lectures on elementary statistics is chosen, then the chances are that it will not be a very satisfactory or stimulating experience for either the psychiatrist or the statistician. An alternative approach which has much to recommended it is to begin with four to five lectures on basics such as the differences between descriptive and inferential statistics and between experimental and observational studies, and an introduction to the concept of hypothesis testing and the scientific method. These lectures would be followed by several sessions in which the psychiatrists would tackle simple problems using an interactive package such as MINITAB (with which, in my experience, they can become relatively familiar within half an hour of "hands on" study). These sessions would, of course, need careful supervision. Following this part of the course, the psychiatrists (depending on the numbers) could be divided into small groups and asked to design a small study for which they would then collect data and analyze as they considered appropriate. Such projects could involve small-scale experiments, the use of hospital case records, etc. On completion of the projects they could be discussed both with the psychiatrists in other groups and with the statistican teaching the course.

Such an approach may require more time and effort than conventional service courses in statistics (both from lecturer and student), but it is likely to provide a far firmer grounding in basic statistics. The course could be further supplemented with illustrations of the misuse of statistics in the psychiatric literature (such illustrative papers should not be too hard to find!) and with caveats about the dangers of an unthinking reliance on statistical packages.

6. WORKING AS A STATISTICAL CONSULTANT IN A PSYCHIATRIC INSTITUTE

Working as a statistical consultant in any environment can be a challenging, exhausting, and occasionally frustrating task. In a psychiatric institute such as the Institute of Psychiatry, all the usual problems may be magnified because the majority of one's clients are medically trained; such training often tends to instill attitudes and expectations which are not particularly helpful to the formation of a good working relationship. Psychiatrists may assume that medically trained researchers such as themselves are ex officio, "top of the tree" in the inevitable hierarchies present in any such organization; they may further assume that the beings that they perceive to be lower in the hierarchy, such as statisticians, should be available at any time to answer any questions they may have. Such assumptions can irritate even the most humble and self-effacing statistician! To illustrate the point I shall ignore the advice generally given early in one's academic career "never resort to anecedotes in making an argument," and recount a story from my own experience as a statistical consultant:

> Late one Friday afternoon (after 2:30 p.m.), I was working in my room struggling with a tricky optimization problem, namely how to position my chair and an open drawer of my desk to achieve the most comfortable position in which to rest while scanning *The Annals of Statistics.* A knock on my door disturbed my intense concentration (this is *not* a euphemism for it woke me up!), and into the room entered a person

whom I had never met before, but who proceeded to inform me that he had a "small problem" with which he needed some statistical help. Ever eager to please I asked him to explain his problem, and was almost at once bombarded with words and phrases such as "visual cortex," "arrays of implanted electrodes," "phosphenes," "ghosts," etc., etc. I eventually managed to stop the flow and reminded my client that I was a statistician not a physiologist and that he would need to explain his problems using language I could understand. Clearly this came as a surprise, as if the fact that a phosphene is a spot of light seen by a blind person with an array of electrodes implanted onto the visual cortex, when these electrodes are stimulated by radio waves across the skull, was self evident! Nevertheless, I persevered and eventually began to understand what appeared to be a very interesting problem, and we began to discuss possible approaches to its solution. All of these involved a considerable amount of work on my part and I estimated that it would take me at least 3 months before I would have any answers. "Oh that's no use" was the reply, "I need the results in the next 2 weeks so that I can finish writing up my M.D. thesis!" (I later discovered that the data had taken 5 years to collect and, yes, the visit to me was the first to a statistician!)

Lest readers should think this typical, I should hasten to add that many of my clients (medically trained or otherwise) are, of course, a pleasure to work with, although it is in the nature of things that one's memories concern those that are not. The relationship between the statistician and the client clearly depends on the personalities and preconceptions of each party; it also, however, depends on a number of other factors. How much statistics does the client know? Is he prepared to accept advice or is he confident he knows the answers and is just seeking confirmation? Is the consultant a freelance statistician or is he a junior member of a University Department of which the client is the Head? Will the consultation lead to a genuine collaboration or will the statistician merely be acknowledged in some subsequent paper? (And does he want this anyway, if the client has ignored his advice?) Perhaps most important of all is the question of whether the statistician is perceived as a scientist in his own right. This sets the tone of the relationship. (Many other issues concerning statistical consultancy in general are taken up in Hand and Everitt, 1987).

Despite the rumors, the statistician has not been largely replaced by the ubiquitous statistical package. Improvements in communication which can only be advantageous for both psychiatrist and statistician are ultimately only achievable by each being aware of and sympathetic toward the problems of the other.

7. SUMMARY

As a scientific discipline, psychiatry is still relatively young. Compared to other branches of medicine, theories of the etiology and treatment of psychiatric disorders are in their infancy. Psychiatrists are increasingly aware that to build and then advance such theories requires well designed quantitative studies in combination with the use of the appropriate statistical tools for the evaluation of results. In addition, the increased use of advanced technology in psychiatry as seen in the shape of CAT and PET scanners is likely to lead to an even greater need for statistical expertise (see, for example, Vardi, Shepp and Kaufman, 1985). Hand (1985) summarizes the point nicely:

> It is clear that statistics serves a major role in modern psychiatry, and that awareness and understanding of statistical concepts is of increasing importance to all psychiatrists, but especially those who wish to advance their field by undertaking research themselves.

The statistician and the psychiatrist clearly have much to offer one another, but patience and understanding in good measure will be needed on both sides to ensure a lasting and fruitful relationship.

REFERENCES

ALLISON, P. D. (1984). Event History Analysis: Regression for Longitudinal Event Data. Sage Publications, Beverley Hills, Calif.

BROWN, G. W. and HARRIS, T. (1978). *Social Origins of Depression. A Study of Psychiatric Disorder in Women.* Tavistock, London.

CAMERON, D. E. (1933). Studies in perseveration. *J. Ment. Sci.* **79** 735–745.

CARNEY, M. W. P., ROTH, M. and GARSIDE, R. F. (1965). The diagnosis of depressive syndromes and the prediction of ECT response. *British J. Psychiatry* **111** 659–674.

CHAMBERS, J. M., CLEVELAND, W. S., KLEINER, B. and TUKEY, P. A. (1983). *Graphical Methods for Data Analysis.* Wadsworth, Belmont, Calif.

COX, D. R. (1972). Regression models and life tables. *J. Roy. Statist. Soc. Ser. B.* **34** 187–220.

DEGROOT, M. H. and MEZZICH, J. E. (1985). Psychiatric statistics. In *A Celebration of Statistics, The ISI Centenary Volume.* (A. C. Atkinson and S. E. Fienberg, eds.) 145–165. Springer, New York.

DELL, S., ROBERTSON, G. and PARKER, E. (1986). Detention in Broadmoor: factors in length of stay. To appear in *Psychol. Med.*

DUNN, G. (1981). The role of linear models in psychiatric epidemiology. *Psychol. Med.* **11** 179–184.

EVERITT, B. S. (1980) *Cluster Analysis*, 2nd ed. Gower, London.
EYSENCK, H. J. (1970). The classification of depressive illness. *British J. Psychiatry* **117** 241–250.
FERGUSSON, D. M. and HORWOOD, K. J. (1984). Life events and depression in women: a structural equation model. *Psychol. Med.* **14** 881–889.
GARSIDE, R. F. and ROTH, M. (1978). Multivariate statistical methods and problems of classification in psychiatry. *British J. Psychiatry* **133** 53–67.
HAND, D. J. (1985). The role of statistics in psychiatry. *Psychol. Med.* **15** 471–476.
HAND, D. J. and EVERITT, B. S. (1987). *The Statistical Consultant in Action*. Cambridge Univ. Press, Cambridge.
HOOKE, R. (1980). Getting people to use statistics properly. *Amer. Statist.* **34** 39–42.
KENDELL, R. E. (1969). The continuum model of depressive illness. *Proc. Roy. Soc. Med.* **62** 335–339.
KENDELL, R. E. and GOURLAY, J. (1970). The clinical distinction between psychotic and neurotic disorders. *British J. Psychiatry* **117** 257–266.
KILOH, L. G. and GARSIDE, R. F. (1963). The independence of neurotic depression and endogenous depression. *British J. Psychiatry* **109** 451–463.
LEE, E. T. (1980). *Statistical Methods for Survival Data Analysis*. Lifetime Learning Publications, Belmont, Calif.
LYTTLETON, R. A. (1977). The nature of knowledge. In *The Encyclopedia of Ignorance* (R. Duncan and M. Weston-Smith, eds.). Pergamon, Oxford.
MASSERMAN, J. H. and CARMICHAEL, H. T. (1938). Diagnosis and prognosis in psychiatry. *J. Ment. Sci.* **84** 893–946.
MATT, F. W. (1913). The neuropathic inheritance. *J. Ment. Sci.* **59** 222–263.
MCCONAGHY, N., JOFFE, A. D. and MURPHY, B. (1967). The independence of neurotic and endogenous depression. *British J. Psychiatry* **113** 479–484.
PILOWSKY, I., LEVINE, S. and BOULTON, D. M. (1969). The classification of depression by numerical taxonomy. *British J. Psychiatry* **115** 937–945.
ROZEBOOM, W. W. (1960). The fallacy of the null hypothesis significance test. *Psychol. Bull.* **57** 416–428.
SANDIFER, M. G., WILSON, I. C. and GREEN, L. (1966). The two-type thesis of depressive disorders. *Amer. J. Psychiatry* **123** 93–97.
SCULL, A. T. (1979). *Museums of Madness*. Allen Lane, London.
SKIPPER, J. K., GUENTHER, A. L. and NASS, G. (1967). The sacredness of 0.05: A note concerning the uses of statistical levels of significance in Social Science. *Amer. Sociologist* **1** 16–18.
SPRENT, P. (1970). Some problems of statistical consultancy (with discussion). *J. Roy. Statist. Soc. Ser. A* **133** 139–165.
TENNANT, C. and BEBBINGTON, P. (1978). The social causation of depression. A critique of Brown and his colleagues. *Psychol. Med.* **8** 565–575.
TUKEY, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, Mass.
U. K. BROMOCRIPTINE RESEARCH GROUP (1987). Bromocriptine (Parlodel), in the treatment of Parkinson's disease: a double blind study of long term efficacy using slow and rapid introductory regimes. Analysis of Phase 1. To appear in *Lancet*.
VARDI, Y., SHEPP, L. A. and KAUFMAN, L. (1985). A statistical model for positron emission tomography (with discussion). *J. Amer. Statist. Assoc.* **80** 8–37.
WHITE, S. J. (1979). Statistical errors in papers in the British Journal of Psychiatry. *British J. Psychiatry* **135** 336–342.

Comment

Donald Guthrie

Psychiatry may be unique among the medical disciplines in the breadth of its scientific collaboration. Research in psychiatry includes components from the biological, medical, behavioral, physical, and social sciences. Their mixture provides an ideal working environment for a statistician with interests in diverse application experiences. Brian Everitt is one of the best known and highly respected statisticians with application interests in psychiatry. His collaboration with Michael Rutter, for example, has led to significant extensions of understanding in child psychiatry.

Everitt is to be commended for providing enlightening and entertaining insight into the role of statisticians in psychiatric research. He has accurately described some of the more rewarding aspects (collaboration on interesting scientific problems) and some of the more frustrating aspects (unwillingness to seek and accept statistical advice). I have had similar experience to Everitt's in reading and contributing to the psychiatric literature. First, I find that there tends to be an obsession with p values and other mechanistic approaches to data interpretation. Second, I share the concern over more-or-less blind use of packaged programs by naive users. It is, I believe, unfair to accuse the psychiatrists I know of making these errors, but the correct blame may lie with those who provide support which should be supplied by statisticians. Most psychiatrists are quite eager to seek and accept expert opinion from statisticians.

Everitt has illustrated applications of Cox regression in his two examples. These examples are illustrative but by no means exhaustive. Statisticians are useful in virtually all aspects of psychiatric research. Let me consider a few additional examples, all of which involve substantial contributions by and from statisticians.

Donald Guthrie is Professor of Psychiatry and Biostatistics, University of California, 760 Westwood Plaza, Los Angeles, California 90024.