Comment

Samuel W. Greenhouse

Mr. Everitt has been closely associated with psychiatric research in his role as a statistician in the Psychiatric Institute. His discussion clearly reflects his experiences and insights in dealing with psychiatrists who reluctantly have to learn statistical methods in order to analyze their data or with those psychiatrists who, having learned statistics, overemphasize significance tests, are "obsessed" with .05 and .01, and at times misuse statistical procedures. Everitt points out that increasingly, more advanced techniques such as log-linear models, logistic regression, and survival analysis (Cox proportional hazard models), among others are being applied in psychiatric research. In connection with survival analysis models, he presents two examples wherein he persuaded the investigators to analyze all their data, both censored as well as uncensored, utilizing a proportional hazards model. (In the second example, where the dependent variable is duration of hospitalization and observations of patients still in the hospital are considered as censored, it is difficult to see how one can assume the censoring mechanism to be independent of the outcome variable: time to hospital discharge.) Of interest are the topics the author seems to favor: greater use of exploratory data analysis, introduction and use of concepts such as support, lack of support, weak support, and use of statistical evidence to adjust degrees of belief. The author also presents an interesting approach to teaching statistics to psychiatrists.

As we think about these issues, we wonder what is there about them that warrants a special article on statistics and psychiatry. Everitt is not alone in having written on the subject. In fact there have been a number of such papers published in the past twenty years among which are DeGroot and Mezzich (1985), Garside and Roth (1978), Hand (1985), Laska, Seigel and Meisner (1985), Moran (1969), and Pocock (1980). All of these papers discuss some aspects of statistical methodology as applied to psychiatric research. Although some authors imply something different about the kind of procedures needed in psychiatry only DeGroot and Mezzich directly consider this issue. They ask: "In particular why is psychiatric statistics not just a branch of biostatistics or psychometrics, the

Samuel W. Greenhouse is Professor of Statistics, Department of Statistics/Computer and Information Systems, George Washington University, Washington, D. C. 20052.

areas of statistics that deal with biomedical and behavioral issues, respectively?" They then respond: "A glib but reasonably accurate answer to this question is that biostatistical methods are typically inadequate for handling the behavioral components of problems in psychiatric statistics, and the psychometric methods are typically inadequate for handling the biomedical components." DeGroot and Mezzich then note that most of the psychiatric literature uses little methodology beyond classical methods and that this is a propitious time for the application of more modern techniques, "suitably modified where required" and for the development of new "theory and methodology to keep pace with and spur new developments in psychiatric research." This theme of the need for more modern techniques occurs in other papers. Indeed, as Everitt points out, some of these, log linear models, logistic regression, survival analysis, have already been instituted.

Although the reader will not find in Everitt's paper what makes statistics in psychiatric research different from the statistics applied in other areas, neither will he find it in the other articles. It is true that DeGroot and Mezzich declare it is different and discuss some of the differences, but they do not indicate why it is or should be different other than the classical procedures used are inadequate and presumably the newer techniques will somehow be adequate.

As one who was closely associated with research in psychiatry, psychology, and the social sciences from 1954 to 1974, I am acutely aware of the force of these arguments. There is no question that psychiatric statistical methods should be strengthened. There exist a number of data collection methods that are indeed peculiar to psychiatry and to the entire rubric of social and behavioral sciences dealing with the assessment of personality, attitude, and behavior. But, except possibly for the major issue of classification and diagnosis, the problems in analysis and inference are not too much different in psychiatry from those occurring in any of the other chronic diseases such as cancer, heart diseases, arthritis, etc. The problems of overuse of .05 and .01 significance testing and the need for newer techniques especially in categorical repeated measurements occur in these other areas also. Even some elements of the classification problem occur currently in heart disease research, as witness the uncertainties and ambiguities connected with evaluating cardiac arteriograms. We do not, however, see

statisticians writing on "statistics and cancer" or "statistics and cardiovascular disease."

My own view is that whatever is unique in the subject of statistics and psychiatry lies as much in the nature of psychiatry and psychiatric reseach as it does in the need for different and more adequate statistical procedures. With .Everitt's opening sentences, it appeared as if finally a statistician closely associated with psychiatric research was going to address the appropriate issues. He declares: "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." This introduction raises an extremely important question, namely, how successful has psychiatry been in its struggles of the last 40 to 50 years in attaining the dignity of a science. A related question asks whether these struggles have reflected changes in measurement in order that the application of numerical methods can prove useful for analysis and prediction. It is a truism, requiring no extensive argumentation, that the answers to these questions relate to, and to a large extent determine, the kinds of statistical procedures that are to be used in psychiatric research.

For example, 30 years ago it seemed that every major clinical investigation of mental diseases or of personality disorders found it necessary to devise their own anxiety rating scale or their own personality function scale or their own mental status rating scale. There were clinical psychiatrists involved in the early studies of the effects of chlorpromazine and reserpine who claimed they did not need any measured scales to assess the status, or the degree of improvement, of their schizophrenia patients. Of interest would be an account of the progress made to date in getting all research investigators to use a common standardized test for each potential outcome variable.

Clearly, the extent to which different measuring instruments are used to determine the status of the same trait or behavior impedes the climb of psychiatry to the status of a science. (It must be noted that there are a small number of psychiatrists who are not dis-turbed if psychiatry is not considered a science, who in fact do not even consider psychiatry a branch of medicine.)

Another area worthy of discussion is the nature of the measurements themselves. For example, what is the meaning of a significant difference between a mean of 2 and a mean of 4 on a seven-point rating scale? What is the meaning of a correlation coefficient based on a seven-point scale? This leads me to the following observation: Probably the major distinguishing characteristic between the research in psychiatry and the research in other disciplines is that almost all of the structure of clinical (not necessarily laboratory) psychiatric research is based on a correlational logic, whereas most of the research in other disciplines is based on a regression logic. Correlations are needed when a discipline is still finding its way—trying to identify and define homogeneous sets of conditions either through similar variables (factor analysis) or similar individuals (cluster analysis). To quote Cornfield (1959): "The degree of articulation of a field is measured by the extent to which the phenomena with which the field is concerned are potentially capable of being explained and predicted in terms of a small number of fundamental concepts and constants." Thirty or forty years ago, the field of psychiatry could well have been described as a field of "low articulation." The interesting question is whether any progress has been made in discovering fundamental principles or indeed whether any such principles and unifying concepts and constants exist in psychiatry. Is every individual's psychotic or neurotic disorder unique to that individual with no comparable features in others with the same disorders? This view is held by some statisticians and psychologists who maintain that research should be centered on the intensive study of the individual. The reason for this contention, I assume, is that the disorder or disease is manifested by the occurrence of abnormal behavioral or cognitive symptoms which are peculiar to the individual and therefore a therapy to be effective must alter that individual's symptoms and not someone else's. (Is this possibly the reason why many therapists object to evaluating psychotherapy with conventional statistical group designs?)

On the other hand, in cancer research or in coronary artery disease research, decreasing tumor size or reducing coronary blockage is a common objective for all individuals with these diseases regardless of differences in detailed manifestations of the disease. A majority of statisticians would find the views in the preceding paragraph detrimental to the process by which we gain new knowledge. I for one agree with this position but have often also wondered whether studying an individual intensively can lead to any less

knowledge than obtaining a 50 order correlation matrix where the highest correlation may be .45, all those greater than .25 are significant at .05, and many of the 50 variables are rating scales with dimensions ranging from two-point to ten-point scales.

In summary, if "statistics and psychiatry" requires special attention over and above the application of statistical methods in other biomedical disciplines, it is not enough to write merely on the statistical side. In order to obtain a more complete view of the issues which do contribute to making "statistics and psychiatry" different, we should also consider problems on the psychiatric side—problems in concepts and problems in measurement. In my view, such a discussion would be most useful if it were made not by a psychiatrist, but by a statistician like Everitt who is aware of these matters because

he has been associated with psychiatric research for a long period of time.

ADDITIONAL REFERENCES

CORNFIELD, J. (1959). Principles of research. *Amer. J. Ment. Defic.* **64** 240-252.

LASKA, E., SEIGEL, C. and MEISNER, M. (1985). Applications of statistics to psychiatry. In *Comprehensive Textbook of Psychiatry IV* (H. I. Kaplan and B. J. Sadock, eds.) **1**, 4th ed. 327-336. Williams & Wilkins, Baltimore.

MORAN, P. A. P. (1969). Statistical methods in psychiatric research (with discussion). *J. Roy. Statist. Soc. Ser. A* **132** 484-524.

POCOCK, S. J. (1980). The role of statistics in medical research. *British J. Psychiatry* **137** 188-190.

Comment

Joseph L. Fleiss

In the early 1920's, Joseph Zubin and a few fellow graduate students undertook a study of 4-, 5-, 6- and 7-year-old children to put to the test Sigmund Freud's Oedipus hypothesis. Data were collected and analyzed, and the statistical results seemed to confirm the master's theories. It was Joseph Zubin's task to prepare the tables, charts, and summary statistics and to send them to Freud. "Ganz amerikanisch" was his disparaging reply, implying that only in America was the need felt to test what was obvious.

Freud might have added "und britisch," because the realization of the need to put psychiatric theories to the test has been a tradition in Britain as well. This paper testifies to the vigor of that tradition. Everitt has provided several examples of the impact made by statistics on psychiatry. Examples exist of the reverse, of the influence that psychiatry has had on statistics.

The long-standing concern that researchers in the mental disorders have had with the unreliability of psychiatric diagnosis (Schmidt and Fonda, 1956) probably provided the major impetus to statistical research on the κ coefficient of chance-corrected agreement (Cohen, 1960; Spitzer et al., 1967; Fleiss, Cohen and Everitt, 1969). On the basis of changes in the value of this statistic, the American Psychiatric Association's Committee on Nomenclature and Statistics (1980) could validly demonstrate that the reli-

Joseph L. Fleiss is Professor and Head, Division of Biostatistics, Columbia University, 600 West 168th Street, New York, New York 10032.

abilities of many important psychiatric diagnostic categories had improved over the preceding 20 years.

The κ coefficient is defined as the ratio $(p_o - p_e)/(1 - p_e)$, where p_o is the observed proportion of cases on whom two diagnosticians agree and p_e is the estimated proportion of agreement expected if the diagnosticians were assigning diagnostic categories randomly. Although originally applied almost exclusively to psychiatric classifications, κ has proven useful in the study of the reproducibility of diagnoses in other medical specialties (Koran, 1975).

Dissatisfaction with psychiatric nomenclature provided an important impetus to research in another area of statistics, cluster analysis (Fleiss and Zubin, 1969; Everitt, 1980). I share Everitt's perception that the reciprocal impact of cluster analysis on psychiatry has been weak. One might even say that the impact has been nil. It is my impression that neither of the current editions of the two diagnostic classification systems most in use in the world today, the American Psychiatric Association's Diagnostic and Statistical Manual and the World Health Organization's International Classification of Diseases, benefitted from the results of cluster analyses or of exercises in numerical taxonomy. I don't know why this is so. Does Everitt have any opinions?

I worry about Everitt's advice to psychiatrists and journal editors to move "away from...tests (of hypotheses) to the more informal methods of exploratory data analysis." We're talking about research, after all, and one of the hallmarks of good research is that one's