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.

ACKNOWLEDGMENT

This work was supported by Grant DMS 86-02282 from the National Science Foundation.

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 reliabilities 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

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