

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  $\kappa$  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-

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

The  $\kappa$  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,  $\kappa$  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.*

methods of procedure are specified in advance. Everitt correctly points out that the pursuit of statistical significance is unfortunate and damaging, but to recommend informality as an alternative is to invite a return to the not-so-long-ago days when psychiatric research had the deserved reputation for producing junk.

Unlike Everitt, I would support the statistician who couldn't or wouldn't help the psychiatrist with a 500-item questionnaire that had been administered to 100 depressed patients. The statistician, if he or she had several years of experience working in the mental disorders, probably knew better than the psychiatrist, who may have been new to psychiatric research, that there wasn't much left to learn about the dimensions underlying depression, that hundreds of factor analyses of rating scales applied to depressives had already been performed, and that virtually nothing of value would be gained by the performance of yet another such factor analysis. Knowledge in psychiatry, and the psychiatrist's career in research, would both have been better served by the specification and testing of hypotheses, perhaps by a *confirmatory* factor analysis (Everitt and Dunn, 1983).

The opinion implicit in the preceding paragraph is that a statistician who's had extensive experience

in a medical or scientific specialty may sometimes have as much or even more knowledge than a person formally trained in that specialty. Does Everitt subscribe to such heresy? How would he recommend a statistician to act if there were a serious disagreement on substantive matters between the statistician and the subject matter "expert?"

#### ADDITIONAL REFERENCES

- AMERICAN PSYCHIATRIC ASSOCIATION COMMITTEE ON NOMENCLATURE AND STATISTICS. (1980). *Diagnostic and Statistical Manual of Mental Disorders*, 3rd ed. American Psychiatric Association, Washington.
- COHEN, J. (1960). A coefficient of agreement for nominal scales. *Ed. Psychol. Meas.* **20** 37-46.
- EVERITT, B. S. and DUNN, G. (1983). *Advanced Methods of Data Exploration and Modelling*. Heinemann, London.
- FLEISS, J. L., COHEN, J. and EVERITT, B. S. (1969). Large sample standard errors of kappa and weighted kappa. *Psychol. Bull.* **72** 323-327.
- FLEISS, J. L. and ZUBIN, J. (1969). On the methods and theory of clustering. *Multivariate Behav. Res.* **4** 235-250.
- KORAN, L. M. (1975). The reliability of clinical methods, data and judgments. *New England J. Med.* **293** 642-646 and 695-701.
- SCHMIDT, H. O. and FONDA, C. P. (1956). The reliability of psychiatric diagnosis. *J. Abnorm. Soc. Psychol.* **52** 262-267.
- SPITZER, R. L., COHEN, J., FLEISS, J. L. and ENDICOTT, J. (1967). Quantification of agreement in psychiatric diagnosis. *Arch. Gen. Psychiat.* **17** 83-87.

## Comment: The Biometric Approach to Psychiatry

Joseph Zubin

Everitt points out that psychiatry for the last several decades has been trying to emerge from its phenomenological descriptive cocoon into a more objective science. Galton was not alone in demanding measurement and numbers as a sine qua non for attaining "the dignity of a science." Thorndike is quoted as saying that whatever exists, exists in some amount and therefore could be eventually subjected to measurement and counting. Lord Kelvin is quoted as saying that one cannot understand a phenomenon

until it is subjected to measurement. Both Emil Kraepelin and Karl Jaspers were appreciative of the importance of objective data and their evaluation. Kraepelin (1896) indicated his interest in measurement in the following statement:

"As soon as our methodology has sufficiently proved itself through experience with healthy individuals, it would be possible to approach the actual ultimate goal of these efforts, the investigation of the sick personality, especially of the inborn pathological disposition. . . . We, therefore, have first of all to investigate whether it is possible by means of psychological tests to determine individual deviations, which cannot be recognized by ordinary observation. If that succeeds, we would be in the position, through the quantitative determinations at our disposal, to establish the

---

*Joseph Zubin is Research Career Scientist and Coordinator for Research and Development, Veterans Administration Medical Center, and Distinguished Research Professor of Psychiatry, University of Pittsburgh Medical School. His mailing address is Biometrics Research, Veteran's Administration Medical Center, Highland Drive, Pittsburgh, Pennsylvania 15206.*