possible, and discuss data collection procedures and analytic techniques vis-à-vis their pertinence to the research objectives and to logistical circumstances. On the other hand, it is essential that the statistical consultant be familiar with the substance of psychiatric theory and practice and possess pertinent interpersonal communication skills.

A topic related to the latter is the teaching of statistics for the psychiatrist, a need well emphasized and examined by Everitt. Additionally, it would seem advisable to consider, from a content stand point, discussion of methods for formalizing and quantifying psychiatric concepts as well as of the value and limitations of various data collection strategies. More generally, didactic effectiveness may be enhanced by integrating this teaching of statistics within the context of educational programs on research design and methodology, and by emphasizing practical and thoughtfully supervised exercises with actual research cases.

ADDITIONAL REFERENCES

BAYARRI, M. J. and DEGROOT, M. H. (1986). Bayesian analysis of selection models. Technical Report No. 365, Dept. Statistics, Carnegie Mellon Univ.

BRYK, A. S. and RAUDENBUSH, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychol. Bull.* **101** 147–158.

FLEISS, J. L., WILLIAMS, J. B. W. and DURBO, A. F. (1986). The logistic regression analysis of psychiatric data. *J. Psychiatric Res.* **20** 195–209.

GULBINAT, W. (1983). Mental health problem assessment and information support: Directions of WHO's work. *World Health Stat.* **36** 224–232.

KRAMER M. and ANTHONY, J. (1983). Review of differences in mental health indicators used in national publications: Recommendations for their standardization. *World Health Stat.* **36** 256–338.

MEZZICH, J. E., DOW, J. T., GANGULI, R., MUNETZ, M. R. and ZETTLER-SEGAL, M. (1986). Computerized initial and discharge evaluations. In *Clinical Care and Information Systems in Psychiatry* (J. E. Mezzich, ed.). American Psychiatric Press, Washington, D.C.

MORRIS, C. N. (1983). Parametric empirical Bayes inference: Theory and applications. *J. Amer. Statist. Assoc.* **78** 47–65.

PARCHAPPE, J. B. M. (1839). *Recherches Statistiques sur les Causes de l'Aliénation Mentale.* Imprimerie de D. Briere, Rouen.

ROSENTHAL, R. (1979). The "file drawer problem" and tolerance for null results. *Psychol. Bull.* **86** 638–641.

STRAUSS, J. S., HAFEZ, J., LIEBERMAN, P. and HARDING, C. M. (1985). The course of psychiatric disorder. III. Longitudinal principles. *Amer. J. Psychiatry* **142** 289–296.

TEN HORN, S., GIEL, R., GULBINAT, W. and HENDERSON, J., eds. (1986). *Psychiatric Case Registers.* Elsevier, Amsterdam.

Comment

Joel B. Greenhouse

Professor Everitt argues that statisticians can and will play a major role in the development and advancement of psychiatric research. By developing close working relationships with investigators in psychiatry, I have found that there are several levels of involvement where a statistician can make important contributions. On the most basic level, and because "psychiatry is a relatively young science," this relationship will often begin with helping an investigator to carefully articulate a set of research questions, including discussions of what kind of data are available or need to be collected to answer these questions, or whether the questions can be answered at all. Of course, this includes developing protocols which will answer these questions. This activity could formally fall under the heading of the principles of experimental design, but as Professor Everitt has suggested, experimental design in the broader sense of including how to think about and do research, with a strong emphasis on the principles of the philosophy of science. For the statistician, it is an activity which is often time consuming, and, since it may not yield a tangible product, may be difficult to document professionally. Sometimes, in the early stages of an investigation, it results in a fundable grant proposal with the statistician listed as a consultant or even as a co-investigator. More often than not, it results in a psychiatrist who thinks more clearly about research and recognizes and appreciates the collaborative role of a statistician.

By far, the most exciting aspects of consulting in psychiatry are encountering new problems, unique to psychiatric research, that lead to interesting statistical and methodological research. Professor Everitt discusses two examples of the use of Cox's proportional hazards model to problems in psychiatry. Although no new problems are discussed here, the application of this methodology clearly was superior to the initial analyses considered and contributed to a deeper understanding of the problems. It is interesting to note that although a standard methodological tool in other

Joel B. Greenhouse is Assistant Professor of Statistics, Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213.

areas of biomedical research, applications of methods for the analysis of time-to-response data in psychiatry have been few. This is true as well for other widely used statistical methods such as logistic regression and exploratory data analysis. Perhaps Professor Everitt has some insights as to why this has been the case.

With respect to the use of Cox's model, there are, of course, situations for which the assumptions of the proportional hazards model are not satisfied and alternative models should be considered. An example of such a case in psychiatry occurs in treatment studies for depression where a certain proportion of patients will recover or remit spontaneously if left untreated. To accurately determine the efficacy of a treatment, it is necessary to develop models that take into account the proportion of patients who recover spontaneously. A class of survival models which has been applied in other settings when the proportional hazards assumption is not tenable is described below with modifications to address the problem of spontaneous recovery (Farewell, 1982; Greenhouse and Wolfe, 1984; Chen et al., 1985).

Let T_{ij} be a nonnegative random variable denoting the time from treatment to response, which in this context is recovery, for the ith patient in the jth treatment group, $i = 1, 2, \ldots, N_j$, where $j = 0$ denotes the placebo control group, $j = 1$ denotes the treatment group, and t_{ij} is the observed survival time. Following the usual convention, we call $S(t) = P\{T > t\}$ the survival function. The following model for the survival functions in the control and treatment groups, respectively, is based on a mixture of survival functions and explicitly incorporates the possibility of spontaneous recovery into the model. Letting p_0 denote the fraction of the population who recover or respond spontaneously with $q_0 = (1 - p_0)$, we have

$$S_0(t) = p_0 F_0(t; \theta_0) + q_0,$$

$$S_1(t) = p_0 F_0(t; \theta_0) + q_0 H(t; \theta_1),$$

where $F_0(t; \theta_0)$ is the survival function for time to recovery for patients who respond spontaneously, $H(t; \theta_1)$ is the survival function for time to recovery for patients who do not respond spontaneously, and θ_j denotes a vector of parameters that indexes a family of survival functions.

Furthermore we note that among the $100(1 - p_0)\%$ patients who do not respond spontaneously, a fraction, say p_1, may not respond at all (i.e., are treatment failures), while the remaining fraction $q_1 = (1 - p_1)$ do respond to treatment with some survival function $V(t; \theta_1)$. Therefore, $H(t; \theta_1) = p_1 + q_1 V(t; \theta_1)$ and the overall survival function for the treatment group becomes

$$S_1(t) = p_0 F_0(t; \theta_0) + q_0[p_1 + q_1 V(t; \theta_1)].$$

There are some obvious and potentially interesting methodological issues with respect to using this model; for example, extensions of the model to include covariates. We are currently investigating properties of this model and plan on evaluating its usefulness in the analysis of data from a soon to be completed clinical trial evaluating different therapies for the prevention of recurrent depressions currently being conducted in the Department of Psychiatry at the University of Pittsburgh.

As Professor Everitt has observed, psychiatrists sometimes make errors not only in the applications of statistical methods but also in the applications of the principles of the scientific method. A recurrent problem in psychiatric research, which, in my opinion, perhaps more than any other single factor has limited the advancement of the field, has been the failure to use appropriate control groups in comparative clinical experiments (see also DeGroot and Mezzich, 1985). I will now describe an interesting example from the psychiatric literature.

A primary objective of research in biological psychiatry in the last decade has been the search for biological markers of depression. One line of research has looked at differences in electroencephalographic (EEG) patterns during sleep between depressed patients and normal healthy controls. Both the depressed patients and the healthy controls spend several nights in the hospital where their sleep-EEG is recorded and monitored. A consistent abnormality found in depressed patients, relative to the controls, has been a shortened time of onset of the first rapid eye movement (REM) period, or REM latency. This decreased REM latency has been suggested as a fundamental diagnostic marker for depression (Kupfer and Foster, 1978; Kupfer, 1984). In an elegantly conceived and well-executed study, Mullen, Linsell and Parker (1986) recognized the need to mimic the pattern of sleep disturbances characteristic of major depressive disorders, for example, shorten total time asleep, in a group of normal volunteers and found a significant reduction of REM latencies in this group of normals without a concurrent onset of clinical depression. Having manipulated the control subjects in an attempt to make them as similar to the patients as possible with respect to total sleep time, the authors have used a more appropriate comparison group, and have demonstrated that the diagnostic use of REM latency changes must be confined to depressed subjects without current sleep disturbances. They note, however, that in effect, such a limitation has no practical clinical applicability. The Mullen, Linsell and Parker study, although indicating several logical flaws in the foundations of research into the biological correlates of depression, serves as an example of the importance of clear thinking and carefully done research for the advancement of psychiatry.

Professor Everitt argues convincingly that for psychiatry to advance systematically as a scientific discipline, psychiatric researchers will need to rely more and more on the "scientific approach" and on statistical techniques. His clear and stimulating presentation has performed a great service to the field of psychiatry by introducing a large audience of statisticians to the problems and challenges of statistics in psychiatry.

ADDITIONAL REFERENCES

CHEN, W. C., HILL, B. M., GREENHOUSE, J. B. and FAYOS, J. V. (1985). Bayesian analysis of survival curves for cancer patients following treatment. In *Bayesian Statistics II* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 299–328. North-Holland, Amsterdam.

FAREWELL, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* **38** 1041–1046.

GREENHOUSE, J. B. and WOLFE, R. A. (1984). A competing risks derivation of a mixture model for the analysis of survival data. *Comm. Statist. A—Theory Methods* **13** 3133–3154.

KUPFER, D. J. (1984). Neurophysiological 'markers'—EEG sleep measures. *Psychiatric Res.* **18** 467–475.

KUPFER, D. J. and FOSTER, F. G. (1978). EEG sleep and depression. In *Sleep Disorders: Diagnosis and Treatment* (R. L. Williams and I. Karacon, eds.) 163–209. Wiley, New York.

MULLEN, P. E., LINSELL, C. R. and PARKER, D. (1986). Influence of sleep disruption and calorie restriction on biological markers for depression. *Lancet* 1051–1054.

Comment

Craig D. Turnbull

Everitt notes that psychiatry is a relatively young discipline as compared to other branches of medicine since its theories regarding the etiology and treatment of psychiatric disorders are in their infancy. He reports that psychiatrists have become increasingly aware that to build and to advance their theories requires well designed quantitative studies in combination with the use of appropriate statistical tools in order to properly evaluate the results of such studies.

While I concur with these assessments, I wish to provide additional material regarding the historical background and various efforts to develop a nosology (or classification) of mental disorders. This discussion will contain the major portion of my comments since the nosology of a science serves as the grist (i.e., the diagnosis of a case) which yields the data which psychiatrists wish to analyze. Although I will only provide a sketch of the developments which have led to the current nosology (DSM-III), I wish to note that considerable advances have been made in the fields of psychiatry and psychiatric epidemiology since 1800.

Psychiatry has indeed struggled to attain the "dignity of science" by submitting its observations to measurement and quantification. In fact, Grob (1985) has noted that after 1800, several currents converged to create a type of social inquiry whose methodological distinctiveness was a commitment to quantitative research. Underlying this urge to quantify was the assumption that such a methodology could explain social phenomena. He also reported that early and mid-nineteenth century commentators were preoccupied with the development of elaborate classification systems and an almost obsessive concern with the collection of statistical data. That the field of epidemiology emerged in such an environment was not surprising.

Grob observed that American psychiatrists were among the staunchest proponents regarding the collection of statistical data. The annual reports of mental hospitals as well as the *American Journal of Insanity*, which was first published in 1844, included statistics on the demographic and geographical characteristics of mentally ill patients as well as the results of therapy. In addition, the federal census of 1840 provided some aggregate data on the mentally ill population.

Although nineteenth century psychiatrists were avid data collectors, their approach to statistical data consisted of the following interests:

1. They used statistics to demonstrate high "cure" rates.
2. They thought that the collection of data would help to uncover laws governing health and disease.
3. They used statistics for purposes of policy advocacy.
4. They viewed statistical data as a means to establish the legitimacy of public mental hospitals and to build support among state officials and the public.

Mid-nineteenth century psychiatrists thought there was a direct relationship between the rising incidence of mental illness and the advance of civilization. Such

Craig D. Turnbull is Associate Professor of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill, North Carolina 27514.