Multivariate Statistics and Meaning in the Study of Psychopathology

by William T. Carpenter, Jr., John J. Bartko, and M. St. Claire Evans

At Issue

Abstract

A recent article dealing with Schneiderian first-rank symptoms illustrates a frequent occurrence: the use of sophisticated statistical data analysis without clear concepts of either psychopathology or statistics. The purpose of this article is to present some guidelines for the interpretation of multivariate statistics and to increase sensitivity to the importance of integrating underlying clinical concepts with statistical data analysis.

Computer assisted, sophisticated data analysis techniques are presently available to most workers at modest expense. A current trend in investigations is to establish a data base which includes a wide range of systematically recorded, codable variables relevant to the description and study of psychopathology. Elsewhere we have noted an untoward trend in the literature whereby commonly used methods are reported without documenting within study reliability or defining the clinical data crucial to clinical understanding (Carpenter, Heinrichs, and Hanlon 1981). More recently we have noted that powerful statistical approaches are often used to scrutinize large data sets without a compelling presentation of the conceptual basis upon which questions of psychopathology are addressed. Such circumstances are not simply wasteful, but create unnecessary complications for the field.

A recent article by Bland and Orn (1980) will illustrate what we believe is now frequent in the literature: the use of sophisticated statistical data analysis without clear concepts of psychopathology. We select this publication since it permits critique on conceptual grounds, and we can introduce actual data into the discussion. It is, of course, more than incidental that the subject matter is an old favorite: Schneider's (1959) first-rank symptoms (FRSs). Bland and Orn have published many excellent reports, including an investigation of FRSs (1979). Their 1980 report is used to illustrate a common problem. Our purpose here is to increase sensitivity to the importance of integrating underlying clinical concepts with statistical data analysis.

Based on a stepwise regression analysis, Bland and Orn (1980) reported that FRSs (scored present or absent) relate to outcome, explaining from 17 to 26 percent (albeit without statistical significance values) of the variance on four outcome measures. Some FRSs had positive regression coefficients, others had negative regression coefficients, and symptoms differed in their predictive value for the four outcome measures. Bland and Orn concluded that the assessment of only the presence, absence, or number of FRSs obscures the relationship of these symptoms to outcome.

Concept and Hypotheses Leading to Investigation

Our work (Carpenter et al. 1973; Strauss and Carpenter 1974; Hawk, Carpenter, and Strauss 1974), a previous study by Bland and Orn (1979), and investigations by others into the prognostic significance of FRSs (Mellor 1970; Taylor 1972; Abrams and Taylor 1973; World Health Organization 1974; Silverstein 1978; Chandrasena and Rodrigo 1979; Kendell, Brockington, and Leff 1979) have been based on the concept that these symptoms identify nuclear...
schizophrenia. This results from Schneider's (1959) emphasis on the primacy of these symptoms in diagnosis, on the presumption that nuclear schizophrenia has a graver prognosis than schizophrenia-like syndromes, and on Langfeldt's (1969) presumed discrimination between poor prognosis schizophrenia and good prognosis schizophreniform psychosis on cross-sectional symptomatology including some FRSs. Hence, earlier investigators hypothesized or assumed that diagnosis had prognostic significance and used FRSs to define the diagnostic entity. One might postulate that FRSs are prognostic indicators independent of their diagnostic utility, but presentation of a new point of view should include a conceptual rationale that clarifies why the prediction was made and its presumed direction. Since it is not intuitively obvious that specific FRSs should bode well for some components of outcome and poorly for others, elaboration of the concept should reveal the basis for this reasoning. Such elaboration would explain the purpose of the study and enable the reader to judge whether results confirm expected relationships.

Rather than approximate the above format, Bland and Orn (1980) note that FRSs have not been tested as prognostic variables devoid of all assumptions as to which are related to outcome, the direction of that relationship, and with which component of outcome they are associated. However, a number of these symptoms failed to have prognostic significance as individual symptoms in an earlier study (Carpenter et al. 1978) in which a large number of discrete signs and symptoms were examined in a hypothesis-generating framework. Nevertheless, Bland and Orn (1980) contend that an FRS may predict one aspect of course while failing to predict another and predicting in the opposite direction on a third dimension of course. Also, they speculate that each FRS may be a very weak predictor but that the cumulative weight of the FRSs may be significant. If either or both of these possibilities is correct, previous studies could reflect a false "negative" concerning the prognostic importance of FRSs. In most clinical studies of psychopathology, failure to confirm hypotheses does not provide definitive rejection. Of the many possible (sometimes fanciful) explanations of failure to confirm, it is usually prudent to pursue only those which seem both plausible and important. In this instance, the merit of a renewed statistical assault on the problem would have been more convincing had a conceptual or theoretical framework explained the failure of previous studies and predicted the findings reported. Instead of establishing this enabling framework, Bland and Orn (1980) used multiple regression techniques devoid of any a priori judgment as to the direction of predictive power of each FRS, what combination of FRSs might be predictive, and what components of outcome might be predicted.

Statistical Support of Investigations

The availability of computer processing statistical packages and today's emphasis on the numerical aspects of clinical observation appear to have encouraged an abundance of fishing expeditions and the reporting of statistical errors (Everitt 1975; Garside and Roth 1978; White 1979; Pocock 1980; Paykel 1981). Nevertheless, there is a valid role for computer-assisted scrutiny of data sets relatively free of presuppositions concerning the association between variables. In some circumstances extensive artifact and bias may have skewed the literature and clinical observation, and a fresh examination of associations may generate new concepts and hypotheses.

In other instances, data reduction may be required to enable an intuitive appreciation of its structure. Whether analyses were conducted for reasons such as these or for little reason beyond the capability to apply multivariate statistics, the results require the utmost scrutiny and skepticism by the investigators. Too often the average reader will be unable to assess the data analyses critically, and simple rules of thumb are not generally known for multivariate techniques, as they are for commonplace univariate techniques (e.g., by chance one expects 1 in 20 independent t tests or chi-squares to reach the p < .05 level). Relevant to the present discussion, if a number of variables in any population are to be examined for their cumulative prognostic significance, it is a statistical fact that some combination of these variables will achieve a degree of predictive power. The investigator can ascertain the strength of the prediction in his data and clarify the extent to which it is greater than that achieved by some chance distribution of scores on these variables.

Next, the investigators should ask the possible meaning of any findings. If findings are contrary to expectations, self-contradictory, or both, the suspicion of chance associations will be increased. For example, if employment record and social functioning have a positive correlation in the course of schizophrenia (as indeed they do, although the correlation is modest), then a variable which predicts good outcome in employment but poor outcome in social function-
ing must be considered a candidate for having a chance rather than a valid predictive relationship to either. It is possible, of course, that such seemingly contradictory relationships are valid and meaningful, but it is here that the role of theory, concept, and clinical observation is critical.

Bland and Orn (1980) report six to eight cumulative predictor variables in stepwise multiple regression accounting for 17 to 26 percent of the variance. They do not report statistical significance (F or p values). In fact, none of their $R^2$'s (percent variance accounted for) reach the classic $S$ percent significance level (see table 1a), while most $R^2$'s hover near their expected values (long range average). When we see that a particular Schneiderian symptom predicts poor outcome for one dimension and good outcome for another, and when we see that the vast majority of variable additions in a multiple regression add insignificantly to prediction, a compelling explanatory theory is required to balance skepticism.

Replication Multiple Regression

How, then, can one determine whether such results are valid? It is not sufficient to say that replication in a fresh sample can determine the issue, for such studies are often difficult, time consuming, and take years to appear in the literature. Since the findings may be variable additions in a multiple regression add insignificantly to prediction, a compelling explanatory theory is required to balance skepticism.

<table>
<thead>
<tr>
<th>Outcome measure</th>
<th>No. of FRSs*</th>
<th>$R^2$</th>
<th>$ER^2$</th>
<th>5% $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic productivity</td>
<td>7</td>
<td>.263</td>
<td>.167</td>
<td>.314</td>
</tr>
<tr>
<td>Social adjustment</td>
<td>6</td>
<td>.174</td>
<td>.143</td>
<td>.282</td>
</tr>
<tr>
<td>Psychiatric condition</td>
<td>7</td>
<td>.183</td>
<td>.167</td>
<td>.314</td>
</tr>
<tr>
<td>Combined outcome</td>
<td>8</td>
<td>.203</td>
<td>.190</td>
<td>.344</td>
</tr>
</tbody>
</table>

Table 1b. FRSs as predictors of outcome: Multiple regression random number study ($n = 43$)—10 replications

<table>
<thead>
<tr>
<th>Outcome measure</th>
<th>No. of FRSs</th>
<th>Average $R^2$</th>
<th>Median $R^2$</th>
<th>$ER^2$</th>
<th>5% $R^2$</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined outcome</td>
<td>7</td>
<td>.168</td>
<td>.166</td>
<td>.167</td>
<td>.314</td>
<td>.05-</td>
</tr>
</tbody>
</table>

* Taken from Bland and Orn (1980). Other entries provided by the present authors.

Note.—See Technical Appendix for definitions of statistical terms.
their predictive relationship with comparable dimensions evaluated at 11 years.1

The 11-year period is closely comparable to the 14-year long-term follow-up period of Bland and Orn.

Multiple regressions yielded similar predictive power (n = 43, seven predictors) as that reported by Bland and Orn (table 1c). In table 1c none of the $R^2$s (percent variance accounted for) reached the 5 percent level of statistical significance, except for IPSS economic productive outcome. Most of the $R^2$s hover around their expected values (column ER; this figure can be thought of as a long run average). The $R^2$ value (for a particular number of FRS predictors) required for 5 percent statistical significance is given throughout table 1.

For comparative purposes in table 1c, we allowed the stepwise multiple regression program to run its course. However, following standardly accepted statistical practice, we would have stopped at step 1 of the stepwise procedure for two of the outcome measures (economic productivity where for one predictor $R^2 = 0.32$ and combined outcome score where one predictor had an $R^2 = 0.11$; the 5 percent value with one predictor is $R^2 = 0.09$). Social adjustment and psychiatric condition were nonsignificant at step 1.

**Table 1c. FRSs as predictors of outcome: Stepwise regression analysis (IPSS data; n = 43)**

<table>
<thead>
<tr>
<th>Outcome measure</th>
<th>No. of FRSs</th>
<th>$R^2$</th>
<th>ER</th>
<th>5% $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic productivity</td>
<td>6</td>
<td>.351</td>
<td>.143</td>
<td>.282</td>
</tr>
<tr>
<td>Social adjustment</td>
<td>7</td>
<td>.108</td>
<td>.167</td>
<td>.314</td>
</tr>
<tr>
<td>Psychiatric condition</td>
<td>7</td>
<td>.135</td>
<td>.167</td>
<td>.314</td>
</tr>
<tr>
<td>Combined outcome</td>
<td>7</td>
<td>.168</td>
<td>.167</td>
<td>.314</td>
</tr>
</tbody>
</table>

**Discussion**

Increasing methodologic rigor, availability of research instruments and statistical computer packages, and an increased emphasis on psychometric evaluation with clinical research data have radically changed the face of psychiatric research, mostly for the good.

In the above discussion, however, we note a common but unnecessary hazard associated with recent developments in clinical research, and caution investigators and their readers against poorly conceived applications of powerful statistical techniques in multivariate research. If chance findings abound, then the perspective gained by theoretical or conceptually derived research and hypothesis testing is obscured. This can be reduced by a sharp critical eye wherein the investigative group attempts to ascertain meaning, not mere statistical significance. The rationale and hypotheses underlying a study should be sufficiently compelling to enable one to understand the possible meaning and importance of results, and it must be appreciated that relationships of note require clinical and conceptual meaningfulness as well as statistical significance.

We concur with White (1979) and others in urging investigators to collaborate with statisticians and journals to make more extensive use of statistical refereeing.

**References**


Carpenter, W.T., Jr.; Bartko, J.J.; Strauss, J.S.; and Muleh, S. Are there pathognomonic symptoms in schizophrenia? An empirical investigation of Kurt Schneider's first rank...


Acknowledgment

The IPSS was a nine-nation collaborative effort conducted under the auspices of the World Health Organization. Although the data reported here were collected at the Washington Center, many contributed to the development and implementation of methods. Volume 1 (World Health Organization 1973) and Volume 2 (World Health Organization 1974) of the IPSS reports list participating centers and collaborators.

The Authors

William T. Carpenter, Jr., M.D., is Director, and M. St. Clare Evans, was Research Associate, Maryland Psychiatric Research Center, Catonsville, MD. John J. Bartko, Ph.D., is Mathematical Statistician, Division of Biometry and Epidemiology, National Institute of Mental Health, Bethesda, MD.
Technical Appendix

1. \( ER^2 \) is read as the expected value of \( R^2 \). It can be thought of as the long-range average value of \( R^2 \) given a particular number of predictors and number of subjects. It is provided as a multivariate (normal) statistic benchmark. Note from the tables that the \( 5\% R^2 \) values are greater than the \( ER^2 \) values. The formula for computing \( ER^2 \) (number of predictors divided by the sample size less one; \( p/N-1 \)) allows the investigator to determine various \( ER^2 \)’s before conducting a computer data run.

2. \( 5\% R^2 \) is the “table value” of \( R^2 \) required for the classic 5 percent level of statistical significance. This value varies as the number of predictors changes and as the number of subjects changes. At the 5 percent level, computed values of \( R^2 \) must be greater than the \( 5\% R^2 \) value for classic 5 percent statistical significance.

3. In table 1b there were 10 random number multiple regression runs. The columns labeled AVG \( R^2 \), MEDIAN \( R^2 \), and \( R^2 \) RANGE are values obtained from the sum, midpoint, and upper and lower values, respectively, of these 10 computed values of \( R^2 \).

An Invitation to Readers

Providing a forum for a lively exchange of ideas ranks high among the Schizophrenia Bulletin’s objectives. In the section At Issue, readers are asked to comment on specific controversial subjects that merit wide discussion. But remarks need not be confined to the issues we have identified. At Issue is open to any schizophrenia-related topic that needs airing. It is a place for readers to discuss articles that appear in the Bulletin or elsewhere in the professional literature, to report informally on experiences in the clinic, laboratory, or community, and to share ideas—including those that might seem to be radical notions. We welcome all comments.—The Editors.

Send your remarks to:

At Issue
Center for Studies of Schizophrenia
National Institute of Mental Health
Alcohol, Drug Abuse, and Mental Health Administration
5600 Fishers Lane, Rm. 10-95
Rockville, MD 20857