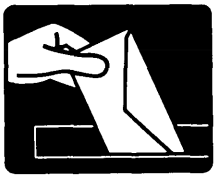
Letters to the Editor



In Defense of Statistical Significance with Multivariate Analysis

Skyler's critique of psychological studies in diabetes in a recent editorial in DIABETES CARE[1] confronts many of the deficiencies in clinical diabetes research head on. His implied criticism of multivariate analyses, however, requires a response that hopefully may clarify some of the misunderstanding in the field concerning its use. He states that "a popular design of many studies is to apply a battery of instruments to a group of subjects, determine outcomes by multivariate analyses, and look for statistical significance. . . . The battery approach with multivariate analysis strikes me as a shotgun design, which is conceptually fuzzy, rather than a carefully planned, focused study." Let's separate out the facts. A shotgun approach is never to be admired in research. A good design is always needed. But the use of multivariate analysis should not be included as a criterion of a conceptually fuzzy study. We are all aware of the GIGO principle in computer analyses of any kind (multivariate or univariate), that is "Garbage-In, Garbage-Out." It is true that thought must be given to the selection of the combination of variables that theoretically "go together" and thus might predict outcome or group differences, and that only these sets of variables belong together in a multivariate analysis. In fact, contrary to what Skyler implies, the inclusion of extraneous variables in a data set will only weaken the chance of finding statistical significance at a multivariate level. Perhaps a brief defense of the rationale behind multivariate analysis is in order.

For example, if we have several outcome measures of the effects of treatment or several different dimensions of psychological functioning, we have multivariate data. This is almost always the case in clinical research. Statistical methods that consider each variable or test by itself are inadequate. Such methods fall within the domain of univariate statistics. We should use multivariate statistics for multivariate data. Almost all multivariate statistical programs give you univariate statistics as well. Table 1 shows some of the common univariate statistics and their multivariate analogues.

Two types of errors can occur if we use univariate methods when the data are multivariate. First, if we compute a lot of separate univariate tests of significance, we will find a few that reach the 5% or 1% level purely by chance. We may make the mistake of attributing meaning to these differences when actually they are nonsignificant. Thus, the multivariate level of significance protects against such chance observations. Second, it may be that no one variable by itself is significant, but that a combination or pattern of variables is. It commonly happens in medical research that a *syndrome* is more informative than each *symptom* by itself. If data from a clinical study have this characteristic, we may mistakenly conclude from univariate tests that nothing is significant when actually there are statistically significant patterns.

Both of these types of errors may be avoided by using multivariate statistics. The overall test of significance embraces all variables in combination. A good example might be that five patient history variables together might predict a response to a drug when no one of the variables alone would be significantly related to drug response.

A simple example may illustrate the point still further. If 100 diabetic individuals were asked two questions in an

TABLE 1
Univariate and multivariate methods

Univariate methods
t test
Analysis of variance
Analysis of covariance
Product-moment correlation
Multivariate analogues
Discriminant function
Multivariate analysis of variance (MANOVA)
Multivariate analysis of covariance
Canonical correlation

attempt to examine the relationship between a health belief model and compliance, the following could occur.

Questions	Answers	Univariate difference
1. Do you see your diabetes as a minor problem?	No (50%) Yes (50%)	No
2. Do you believe treatment will prevent complications?	No (50%) Yes (50%)	No
Multivariate difference		Yes

If 50% answered yes and 50% answered no to each of the questions, there would be no univariate difference between responses to each question. However, if the individuals who answered no to question 1 were the same persons who answered yes to question 2, then there would be a significant difference at a multivariate level, indicating that there was a pattern of response in the group. Such a combination of responses might be related significantly to compliance.

Another common mistake is using less than the most sensitive techniques for analyzing the data we collect. Whenever we have measures on patients before and after treatment, it makes sense to analyze all the data, taking into account the individual differences that existed before treatment. The most obvious way to allow for such differences is to calculate a difference score. If a patient weighed 350 pounds before taking a drug and 300 pounds after, we could say that he lost 50 pounds. If another patient weighed 150 pounds before and 100 pounds after, we would likewise say that he lost 50 pounds. However, this simple and obvious procedure of subtracting initial weight from final weight may not be as innocuous as it seems. It has the effect of lumping all people who lost 50 pounds together for analysis as if they were the same. But, a heavy person who loses 50 pounds may be medically different from a light person who loses 50 pounds. Simple difference scores are much more complicated mathematically than they seem. It has been shown that they have many undesirable statistical properties and may lead to false conclusions because they place people together who are really different.

As an example, in a carefully planned drug study, the authors of an article reported that the effects of several drugs on schizophrenic patients were ineffective when they used difference scores. In fact, they concluded: "There can be little doubt that insofar as drugs are concerned, this paper reports only negative findings. No significant differences were found with an individual drug or combination of drugs over a period of 16 weeks. This was true whether the apparent drug effects were compared with each other, with the initial ratings, or with the predrug phase."[2] Since the authors had published their scores for each patient before and after treatment, Clyde et al.[2] were able to demonstrate the results when the correct statistical technique (multivariate analysis of covariance) was applied to the data. This procedure makes allowances for differences in initial ratings but on a sounder basis than simply subtracting initial from final ratings and analyzing the difference scores. Contrary to the authors' report, the analysis of covariance showed that one of the main effects of the drugs was significant at the 3% level, in that one of the drugs actually made the patients worse.

Statisticians have many valuable techniques for analyzing interrelationships in multivariate data. Almost all of these methods have been programmed for computers. Sophisticated statistical techniques can never take the place of a sound research design that delineates ahead of time the variables and groups required to test a hypothesis. This, I would suggest, is what Skyler had in mind regarding fuzzy studies. Statistical fishing expeditions and manipulations post-hoc have little to do with the specific type of statistic used. However, no longer need an investigator plan and carry out a good study only to miss the significant findings because he employs crude statistical methods that are 30 years out of date.

MARGARET W. LINN

From the Veterans Administration Medical Center, University of Miami School of Medicine, Miami, Florida 33136.

REFERENCES

[1] Skyler, J. S.: Psychological issues in diabetes. Diabetes Care 4: 656–57, 1981.

[2] Clyde, D. J., Cramer, E., and Sherin, R. J.: Multivariate Statistical Programs. Coral Gables, Fl., University of Miami, 1966.

Properties of Dextrostix

We have recently carried out a study designed to investigate several properties of Ames Dextrostix (Ames, Elkhart, Indiana) used for home blood glucose monitoring. The Dextrostix were read using an Ames Glucometer.

First, we wished to confirm Sönksen's statement that halving the reaction time of blood with the Dextrostix and doubling the reading gives an accurate value for high blood glucose concentrations.[1] We measured glucose in 33 blood samples using paired Dextrostix, one in contact with blood for 60 s (recommended procedure), the other for 30 s. The 30-s values, when doubled, were virtually identical to the 60-s values (14.6 ± 4.9 mmol/L versus 14.8 ± 4.6 mmol/L, respectively; mean ± SD). We showed previously, using the recommended procedure, that readings with Ames Dextrostix were unreliable when blood glucose was ≥15 mmol/L.[2] Shortening the reaction time therefore extends the effective range of the Dextrostix and should overcome the problem of inaccuracy at high blood glucose concentrations.