they appeared to contribute to the desired classification of the two sample groups.

The problem is most clear in the analysis of 65 variables chosen, objectively, as the most commonly occurring (p. 1222). The investigator selected six of these variables—not by objective means but on the grounds that they worked best—and then made a formal discriminant analysis of those six variables. But in effect the analysis involved 65 variables, *more* than the number of cases. The results are therefore suspect; the relevant statistical question for the believability of the results is not whether the classification success might have been found just by chance for six variables, but more nearly for the best six chosen opportunistically from 65 possibilities.

There is nothing wrong with all this as exploration. (And in other respects Jurs' exploratory analysis was very sophisticated: There was an additional selection criterion that the variables not discriminate among other subdivisions of the data and an "internal validation" showed that the findings were not the result of a few highly deviant subgroups in the sample.) But discriminant functions prove to be notoriously poor in external validation, that is, in classifying samples other than the one on which they were based. A serious case for the validity of the pattern recognition in such an exploratory study can be made only by means of actual prediction in a subsequent study.

These remarks have pertained to prediction, which was the focus of Jurs' article. But much more could be said about weaknesses of the procedure for purposes of explanation. Even if the procedure for identification of variables were legitimate, and prediction were significantly better than chance, little credence could be placed in the meaning of the particular coefficients found for the sample unless all important variables were known to be included in the analysis (or were known to be uncorrelated with the variables already included). The inclusion of a single important missing variable could significantly change all of the values, even the signs, of the coefficients.

ANDREW AHLGREN
College of Education and
Office of Academic Affairs,
University of Minnesota,
Minneapolis, MN 55455

REFERENCES

1. T. R. Stouch and P. C. Jurs, *J. Chem. Inf. Comput. Sci.* **25**, 45 (1985).
2. J. A. Pino, J. E. McMurry, P. C. Jurs, B. K. Lavine, A. M. Harper, *Anal. Chem.* **57**, 295 (1985).

Response: I agree that one must be careful about getting results in discriminant studies that are due to chance and are not as meaningful as they appear to be. In addition to the paper cited by Ahlgren, we have published twice more on this subject (*1*). With respect to the selection of six variables from 65 in the cystic fibrosis work, we selected six chromatographic peaks that possessed several attributes simultaneously, both with respect to the desired separation, but also with respect to the undesired separations. In addition, while I agree that prediction on true unknowns is the most satisfying test of discriminants, such unknowns are not always available in real studies. Finally, it is true that the results obtained in exploratory data analysis are dependent upon the variables used, and one can seldom be sure that all important variables have been included in the analysis in this type of work. We must do the best we can.

PETER C. JURS
Department of Chemistry,
Pennsylvania State University,
University Park, PA 16802

REFERENCES

1. T. R. Stouch and P. C. Jurs, *J. Chem. Inf. Comput. Sci.* **25**, 92 (1985); *Quant. Struct. Act. Relat. Pharmacol. Chem. Biol.* **5**, 57 (1986).