Journal of Pharmaceutical & Biomedical Analysis
Vol. 9, No. 8, pp. 659–662, 1991
Printed in Great Britain

0731–7085/91 $3.00 + 0.00

The selection of laboratory tests for organ dysfunctions using multivariate statistics*

H.E. SOLBERG

Department of Clinical Chemistry, Rikshospitalet, N-0027 Oslo 1, Norway

Abstract: There is a need for selecting good combinations of laboratory tests for various medical decision situations. Multivariate statistical methods are better tools for this type of selection than multiple univariate comparisons of candidate tests. Linear discriminant analysis is a suitable method. The best selection strategies are stepwise selection and evaluation of all possible subsets. Guiding criteria in the selection process may be measures of statistical separation between the clinical groups or empirical probabilities of correct reallocation. The latter criterion has the advantage that an optimum subset size may be identified.

Keywords: *Laboratory tests; medical decision-making; multivariate statistics; variable selection.*

Introduction

Laboratory tests play a major role in medical decisions concerning clinical diagnosis and follow-up, evaluation of biochemical and physiological functions, estimation of prognosis, and selection of treatment. The problem is to select the most suitable laboratory test or panel of tests from the vast repertoire of a modern laboratory. The aim of this paper is to review guidelines for the use of statistical methods for the identification of optimum combinations of laboratory tests.

Laboratory tests may be included in packages (1) because they provide indispensable data for the interpretation of biochemical or physiological states or (2) because they contribute to reliable allocation of patients in clinical categories. The following considerations concern the second of these two problem domains.

Examples will be drawn from the author's previous studies on optimum combinations of laboratory tests for thyroid and liver disorders [1]. A more comprehensive review of the theory and use of multivariate statistical methods, focusing on discriminant analysis, has been published [2].

Clinical Data

To identify an optimum combination of laboratory tests for a specified type of medical decision, one needs at least a training data set to use for the selection of the test panel. In addition, it is often necessary to have an evaluation data set for the performance testing of the selected test combination since the use of the training data for this purpose may result in a too optimistic picture of the efficiency.

The training data set

This should consist of a sufficient number of individuals correctly classified into two or more relevant clinical groups, e.g. a group of healthy controls and one or several groups of patients having the diseases under study. A set of candidate laboratory tests should be performed on each of these persons. To avoid circular reasoning, the results of the laboratory test to be evaluated should preferably not be used for the initial categorization of the individuals.

The evaluation data set

This may contain the same type of data, i.e. other individuals belonging to the same clinical categories with results of the laboratory tests selected during the training phase of the study. It is, however, an advantage to include other clinical groups as well in the evaluation data set. The benefits of this extension are caused by two factors. (1) Allocation rules are usually only valid for clinical groups that match those included in the training data set. It is therefore

*Presented at the Symposium on "Chemometrics in Pharmaceutical and Biomedical Analysis", November 1990, Stockholm, Sweden.

of interest to test the performance of these rules on individuals belonging to related or similar clinical groups. (2) The training data set often represents an ideal situation with individuals without disturbing conditions. Therefore, a more realistic picture of the real-life performance is obtained by including "diagnostic noise", i.e. individuals that have other diseases or conditions in addition to those studied [3].

The Insufficiency of the Multiple Univariate Approach

Given a suitable training data set, it might be tempting to identify a panel of laboratory tests on the basis of their ability distinguish between clinical classes when evaluated separately, as has often been done in the past. This is the *multiple univariate approach*. There are many possible criteria for univariate ranking of the laboratory tests [1], for example, Student's t test (two clinical groups), F-ratio (analysis of variance, two or several groups), or correlations between the test results and a property variable (with a different value assigned to each clinical group). The correlation method is a natural choice if the property variable is of an ordinal nature as is the case when the clinical groups belong to a functional scale, e.g. 1 = hypothyroid, 2 = euthyroid, and 3 = thyrotoxic.

An example from the Oslo Liver Study [4] may show the inappropriateness of the multiple univariate approach. When Student's t test was used for ranking laboratory tests to distinguish between chronic active hepatitis and primary biliary cirrhosis, the following four serum and plasma tests gave the highest t-values: Ceruloplasmin ($t = 6.5$), Normotest (a coagulation parameter, $t = 6.2$), phospholipids ($t = 4.8$), and alkaline phosphatase ($t = 4.3$). All these tests were, however, intercorrelated ($0.52 < r < 0.81$). In fact, all four tests are measures of biliary obstruction (at least partially).

The Student's t test example shows the main problem with the multiple univariate approach. It cannot exclude inclusion of laboratory tests that carry essentially the same type of information, i.e. it selects *redundant* tests. This statement probably holds true for any multiple univariate method, as was confirmed in a study where several univariate criteria were tested [1].

The Multivariate Approach

It is more appropriate to identify the best laboratory test panels by multivariate methods that disentangle the intercorrelations between the variables. By this approach the tests are allowed to carry only the discriminatory information that is unique to each of them.

Furthermore, multivariate allocation rules are more efficient than decisions based on single test results evaluated separately. In the liver disease study referred to above [4], a linear discriminant function based on the four best laboratory tests (ceruloplasmin, IgA, haptoglobin and ALAT), identified as described below, correctly reallocated 93% of the patients with chronic active hepatitis and primary biliary cirrhosis, while each test alone only succeeded in 4–27% of the cases.

The cause of this phenomenon is that considering each observation (the set of an individual's laboratory results) as a point in multidimensional space increases diagnostic sensitivity and specificity because the observations for different clinical groups tend to cluster in different regions of this space. Figure 1 shows an example with two tests and two groups. The ellipses represents the projections of two bivariate distributions on a plane through both axes. The univariate distributions for the two groups (see the projections on planes along each of the two axes) overlap considerably, in contrast to the projections on the ideal plane along the line B–B.

Linear discriminant analysis is, in fact, a multivariate statistical method that identifies the optimum separating planes through the multidimensional space. In the bivariate example shown in Fig. 1, a single plane is located along the line A–A. In other words, the two bivariate distributions are transformed to the two univariate distributions shown on the plane along the perpendicular line B–B. With more than two groups, a corresponding number of discriminant functions establish a set of separating planes.

Multivariate Methods

Classical linear discriminant analysis (the Fisher's type) [2] is only one of several multivariate methods that can be applied for identifying optimum combinations of laboratory tests [1].

There is no simple answer to the question of

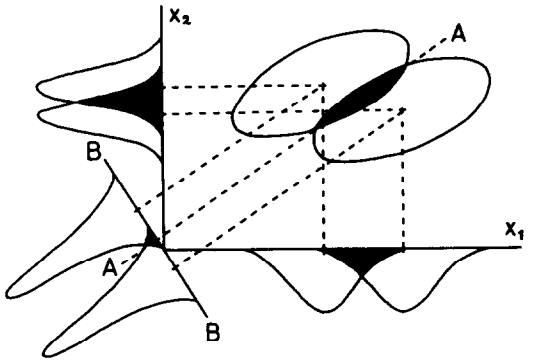

Figure 1
The separation of two bivariate distributions (ellipses) and their projections on planes along the variable axes and along line B–B. From H.E. Solberg, *Scand. J. Clin. Lab. Invest.* **35**, 705–712 (1975), with the editor's permission.

which method should be chosen when several are available. The choice depends partially on the data structure. Most multivariate methods, in particular those belonging to the class of parametric methods, make assumptions concerning distribution types and, in many cases, presuppose equality of the variance–covariance matrices in the groups. The availability of computer programs that support variable selection may also determine the choice.

An incomplete list of alternative methods follows.

(1) Quadratic discriminant analysis, which caters for differences in the variance–covariance matrices [2].

(2) Other variants of discriminant analysis, including non-parametric types, which may be more appropriate when the underlying data do not conform to multivariate Gaussian distributions [1–3].

(3) Pattern recognition methods that are based on partial or total disentangling of the intercorrelation between the variables [1, 3]. The SIMCA-method may also be placed in this category.

(4) Multiple linear regression, particularly when the clinical groups belong to an ordinal scale, e.g. functional classes like 1 = "low", 2 = "normal", and 3 = "high".

The author has tested and applied several methods [1], many of which produced acceptable results, although the classical linear discriminant analysis was preferred for most studies. This method presupposes multivariate Gaussian distributions with equal variance–covariance matrices. This assumption is often not fulfilled with medical data. The method seems, however, to be reasonably robust against violations of the theoretical assumptions [2]. The remainder of this paper discusses the use of linear discriminant analysis for the selection of optimum combinations of laboratory tests although many topics are relevant for other methods as well.

Selection of Tests by Discriminant Analysis

In addition to the possible effects of violated assumptions, the following topics must be considered when using linear discriminant analysis for the identification of an optimum subset of laboratory tests for a specified clinical purpose: (1) selection strategy; (2) selection criterion; (3) subset size; and (4) final evaluation.

Selection strategy

Most computer programs for linear discriminant analysis offer several methods for variable selection (in our case: selection of laboratory tests) [2]. By forward selection one identifies the best single variable and then adds at each stage the variable, among the remaining candidates, which adds most discriminating power in combination with those already included. By backwards elimination the process starts with the whole set of candidate variables followed by sequential elimination of the variables that contributes least to the discrimination. Both of these methods have drawbacks [2] that makes stepwise selection a better alternative. The latter method combines the two best features of the other two. It is a modified forward selection strategy: variables that become redundant as new variables are added to the subset are expelled (they are allowed to re-enter at later stages). This method usually selects an acceptable combination of laboratory tests although it may in some cases not identify the absolutely best subset. Another drawback: this strategy only identifies one subset among all the subsets that might be almost as good as the one selected. The user is, accordingly, not given any possibility to choose, such as is provided when testing all possible subsets (of all sizes or of a specified size). Then a ranked list of, say, the top ten best combinations can be produced. This method may, however, be prohibitive in terms of computing time. For example, the number of possible subsets of size four among 28

candidate variables is 20,475! But it is possible to reduce the search space by applying a branch and bound algorithm [2].

Selection criterion

When searching for good subsets, one needs a selection criterion [2]. Two classes of criteria exist. Most computer programs for variable selection by discriminant analysis apply measures of statistical separation between groups such as *F*-ratio or Mahalanobis' squared distance (or its multi-group analogue). There is, however, often no constant relation between statistical separation of groups and the empirical probability of correct allocation. Therefore, some programs are based on the frequency of correct reallocation of the cases in the training data set for each of the variable subsets evaluated. The best estimate of reallocation probability seems to be obtained by the iterative leave-one-out or split-sample techniques [2].

Subset size

One also needs a stopping rule to identify a variable subset of optimum size. Computer programs that employ separation statistics as criteria for the stepwise method (see above), usually stop the process when the increase in separation is below a threshold value. This method, however, often includes more variables in the subset than desirable. Figure 2 gives an example. As can be seen, the minimum reallocation error is obtained with a subset of three laboratory tests, but the statistical separation continues to rise as more variables are included. This phenomenon is a good argument in favour of using reallocation rates as the guiding criterion in the selection process (see above). In this case the optimum size is accordingly three. More tests only cause redundancy.

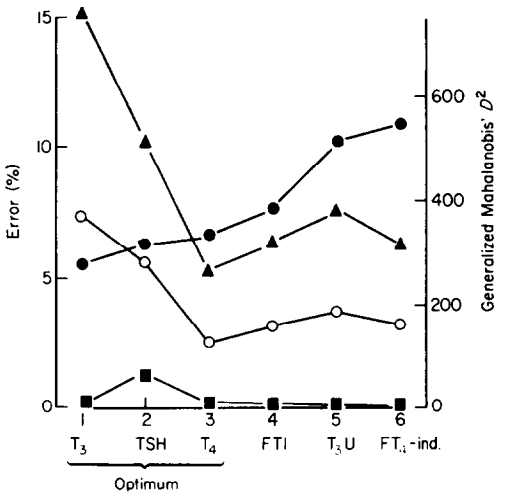

Figure 2
Stepwise selection by linear discriminant analysis of laboratory tests for the separation between thyrotoxicosis and euthyroidism. Filled circles: Mahalanobis' squared distance; open circles: total reallocation error (%); squares: reallocation error of euthyroid cases; triangles: reallocation error of thyrotoxic cases. From K. Rootwelt and H.E. Solberg, *Scand. J. Clin. Lab. Invest.* **38**, 477–485 (1978), with the editor's permission.

Final evaluation

The final evaluation of identified good subsets of laboratory tests has two aspects. First, a laboratory investigator might, for reasons of economy or the information content of the tests in addition to that conveying separation, wish not to select the absolutely best test combination identified. Therefore a ranked list of good test subsets is valuable. Second, as mentioned above, one usually needs an evaluation data set to test the real-life performance of the laboratory test panel.

Final Remark

The statistical methods for variable selection used for the identification of good laboratory test combinations for clinical purposes may also provide algorithms for computer-aided medical decision. The methods are, however, valuable tools even though such decision support is not used in clinical practice.

References

[1] H.E. Solberg, S. Skrede and K. Rootwelt, *Clin. Lab. Med.* **2**, 735–750 (1982).
[2] H.E. Solberg, in *Handbook of Clinical Chemistry*, Volume III (M. Werner, Ed.), pp. 215–245. CRC Press, Boca Raton, Florida (1985).
[3] H.E. Solberg and K. Rootwelt, in *Advanced Interpretation of Clinical Laboratory Data* (C. Heusghem, A. Albert and E.S. Benson, Eds), pp. 65–77. Marcel Dekker, New York (1982).
[4] S. Skrede, H.E. Solberg, S. Ritland, J.P. Blomhoff, E. Schrumpf, K. Elgjo and E. Gjone, *Clin. Chem.* **28**, 1177–1181 (1982).

[Received for review 26 November 1990; revised manuscript received 3 April 1991]