timely basis has been the longest challenge for the pharmaceutical industry. As a result, tablets contain a large number of excipients including fillers or diluents, binders or adhesives, disintegrants, lubricants and glidants, colors, flavors and sweeteners; it might also be necessary to add miscellaneous components such as buffers, depending on the application. What constitutes an ideal combination of these ingredients is of great value to the formulators since not only they have to prepare an effective and stable formulation but this must be done at the lowest possible cost. This evaluation is best made by using such statistical techniques as multivariate methods.

Multivariate techniques make use of statistical experimental design, especially designs that deal with optimization, where much effort is spent on obtaining detailed knowledge about the investigated domain which may include the multivariate characterization of the excipients, in terms of both physical and spectral properties, together with principal component analysis (PCA), statistical experimental design in principal properties (PPs), and partial least squares projections to latent structures (PLS) analysis.

Component analysis: An $N \times K$ data matrix consists of N rows and K columns. The samples or objects in the rows are described by measured or calculated variables given in the columns. In a graphical illustration of a data matrix, the objects are a swarm of N points in a coordinate system of K variables. In cases where a number of objects are described by many variables, the variables tend to be correlated to some extent. This is especially true for spectral variables, where a high absorbance at one wavelength is usually accompanied by similar absorbance values at neighboring wavelengths. PCA uses this correlation to describe the variation in the data with a minimum number of orthogonal components. PCA corresponds to the least squares fitting of a straight line ($A = 1$) or an A-dimensional hyperplane to the data in the K-dimensional variable space. Objects are projected onto a subspace of lower dimension and receive new identities, t-values, often referred to as PPs or scores. The variation of the objects is summarized in the ($N \times A$) matrix, which includes a score vector for each component. Score values from two principal components (PCs), together span a mathematical plane, often referred to as a score plot. Objects are projected onto the plane to form a two-dimensional model of the data. This facilitates the detection of groupings, trends, and outliers (deviating objects) in data sets. The process of detecting and diagnosing outliers is important both when fitting and interpreting the model. An outlier may be an object that does not fit very well into the model, that is, one for which the distance to the model in X is too large to be accepted. Examining the residuals of that particular object will reveal the cause of the deviation. An outlier may, alternatively, be an object that lies far away from other objects in the score plot. Since PCA is a least squares technique such an outlier may cause one of the PCs to run through it or very close to it, resulting in a skewed model. Such outliers should be removed upon identification. PCA models can be calculated using the nonlinear iterative partial least squares (NIPALS) algorithm. The first component explains as much as possible of the variance, the second component is orthogonal to the first and explains as much as possible of the residual variance, and so on. The diversity of PCA applications makes it a very powerful tool in many situations. PCA can be used as a means to discover trends, groupings, and outliers in many types of data, to classify objects, as well as to reduce the number of dimensions and descriptive variables. The features of the PCA model of most interest in any particular study will depend on the systems being investigated and the purposes of the study.

MSC and SNV: multiplicative scatter correction (MSC) is a method for linearization and scatter correction of NIR. It is assumed that the factors affecting physical light scattering of a particular wavelength differ from the chemical factors affecting light absorption. Hence, a corrected spectrum should include only chemical information. In order to normalize the scatter level an "ideal" sample, often the average of the data set is used to correct data for each of the samples. The sample spectrum is regressed onto the average in order to calculate the additive offset and the multiplicative constant. MSC should be used carefully, as all of the samples influence the correction terms, so a deviating sample could have adverse effects on the corrections. The standard normal variate transformation (SNV) as a method for removing unwanted variation from NIR spectra. In contrast to MSC, the correction is performed on an individual sample basis, thus eliminating the possible negative effects of a deviating sample. One of the drawbacks of using SNV, as well as MSC, is that potentially interesting information regarding the particle size is lost. In cases where a response matrix exists there are other methods for removing noise from spectra. The concept of orthogonal signal correction (OSC), a method for removing information in spectra that is not related to the response prior to investigation.

Missing data can be handled by NIPALS. As a rule of thumb, in order to use this approach, there should be five times as many observations in any row or column as the number of dimensions (A) being calculated. The missing values should also be randomly distributed.

Ultravariate characterization is the basis for multivariate design. Descriptive variables that are used to characterize the excipients (for example) may be either physical properties or other variables. Usually, a homogenous group of constituents are put in the same group and characterized by the same variables, where the class of excipients commonly used as lubricants are described using literature data on relevant physical properties. By applying PCA to the descriptive data, the important information is extracted in a few PCs. The PCs are often referred to as latent variables or the PPs of the data set. Each excipient is assigned a score value in each PC. Thus, the excipients are compared and related to on a continuous scale of PPs, which are assumed to reflect real differences in excipient properties and greater distances between excipients along the PCs reflect greater differences in behavior.

LII. PHYSICAL PROPERTIES

Physical properties of the excipients influence the properties of the tablet, for example particle size and bulk volume. Determining physical properties of excipients demands a systematic approach and may consume substantial resources. To establish an optimal choice of excipients, screening experiments are conducted to gain knowledge about parameters that influence the measured results. The traditional approaches to experimental design are difficult to implement when choosing factors to use in a screening study investigating more excipients than can possibly be managed in a mixture design. One alternative is to use physical properties as factors,for example viscosity or some measure of particle size, for each class of excipients. Only a limited number of descriptive variables can be used for each excipient class for a manageable number of experiments. Orthogonal factors can