the biological equivalent of exploratory data analysis, or EDA [9]. Each test or assay has an associated activity criterion, that is, a level above which the activity of interest is judged to be present. If the result for a particular test compound meets this criterion, the compound may pass to the next stage. This criterion could be based on statistical significance (e.g., all compounds with observed activities significantly greater than the control at the 5 percent level could be tagged). However, for early screens, such a formal criterion may be too strict, resulting in few compounds begin identified as "active."

A useful indicator of the efficacy of an assay series is the frequency of discovery of truly active compounds. The frequency is related to the probability of discovery and to the degree of risk (hazard to health) associated with an active compound passing a screen undetected. These two factors in turn depend on the distribution of activities in the series of compounds being tested and the chances of rejecting or accepting compounds with given activities at each stage.

Statistical modeling of the assay system may lead to the improvement of the design of the system by reducing the interval between discoveries of active compounds. The objectives behind a screen and considerations of (1) costs for producing compounds and testing and (2) the degree of uncertainly about test performance will determine desired performance characteristics of specific cases. In the most common case of early toxicity screens performed to remove possible problem compounds, preliminary results suggest that it may be beneficial to increase the number of compounds tested, decrease the numbers of animals per group, and increase the range and number of doses. The result will be less information on more structure, but there will be an overall increase in the frequency of discovery of active compounds (assuming that truly active compounds are entering the system at a steady rate).

The methods described here are well suited to analyzing screening data when the interest is truly in detecting the absence of an effect with little chance of false negatives. There are many forms of graphical analysis methods available, including some newer forms that are particularly well suited to multivariate data (the types that are common in more complicated screening test designs). It is intended that these aspects of analysis will be focused on in a later publication.

The design of each assay and the choice of the activity criterion should, therefore, be adjusted, bearing in mind the relative costs of retaining false positives and rejecting false negatives. Decreasing the group sizes in the early assays reduces the chance of obtaining significance at any particular level (such as 5 percent), so that the activity criterion must be relaxed, in a statistical sense, to allow more compounds through. At some stage, however, it becomes too expensive to continue screening many false positives, and the criteria must be tightened accordingly. Where the criteria are set depends on what acceptable noise levels are in a screening system.