Advances in biotechnology, no matter how impressive, will not be widely adopted unless the benefits and harms are rigorously evaluated in clinical studies. Rigorous evaluation of clinical studies requires familiarity with design and analysis methods developed in the field of biostatistics. Hence, an appreciation of the role of biostatistics is important for researchers in biotechnology. The evaluation of emerging biotechnology works best if there is a partnership between clinical scientists and biostatisticians during the earliest planning stages of the study, rather than seeking statistical input as an afterthought upon completion of the study. A few relevant topics are discussed below.

Evaluating New Interventions

Many aspects of biotechnology involve new medical interventions. The gold standard for evaluating new medical interventions is a randomized trial. In a randomized trial, subjects are randomly assigned to one of two groups: either a new intervention or a control. Randomization is important because it ensures that, on average, the correct answer will be obtained. It is also important because it is the most powerful “tool” available for avoiding systematic bias. Randomized trials have been widely used for decades and there are many books and articles on the topic.
Sometimes, it is not possible to implement a randomized trial because the subjects are unwilling to be randomized into two different types of intervention, such as surgery or chemotherapy. In this situation, an alternative evaluation approach is to perform an observational study in which subjects choose the treatment. The problem is that subjects with risk factors associated with poorer prognosis may preferentially choose one treatment over the other, thus causing incorrect results. The only way to avoid an incorrect interpretation of results is to mathematically adjust for all confounding factors, namely risk factors that are related to both the outcome and the type of subject who would receive an intervention. As observational studies are particularly prone to the effect of unadjusted confounding factors, it is particularly important for the clinical investigator to work with the biostatistician in the design, conduct and analysis of such studies. However, because one can never be sure of identifying all relevant risk factors, there is always some concern about the validity of the results. It is even possible for an unadjusted risk factor to reverse the direction of the outcome. For example, if the risks of outcome differ for men and women and if there is no mathematical adjustment for sex in the analysis, it is possible for an intervention to be “good for men, good for women and bad for people.”

Another approach for evaluating a medical intervention when a randomized trial cannot be implemented is the paired availability design for historical controls. In this design, a new intervention is made available at multiple sites (e.g. hospitals) with a well-defined patient population and little in- or out-migration. The relevant comparison is the change in outcomes among all subjects in time periods before and after the change in availability of the medical intervention divided by the difference in the fraction of patients who receive the intervention before and after the change in availability. This comparison is averaged over the multiple sites. If there are no other changes over time that could affect the outcome, this method can be relied upon to give correct results.

Validating Surrogate Endpoints

A drawback of many randomized trials is the long time period needed to observe the health outcome, such as the occurrence of disease or death (i.e. the true endpoint). Therefore, there is a growing interest in identifying outcomes that might appear sooner. The biostatistical term for an early outcome that is designed to replace the true endpoint is a surrogate endpoint. Many biotechnology projects have, as their goal, the identification of new surrogate outcomes. However, there are two important biostatistical considerations related to their validation.

Firstly, there is a common misconception that if the surrogate endpoint is highly associated with the true endpoint, the surrogate endpoint is valid and it will yield the same inferences about intervention effect as the true endpoint. A better criterion is that a surrogate endpoint is good if the effect of intervention on the surrogate endpoint is a good predictor of the effect of intervention on the true endpoint. The two criteria are not equivalent. For example, a surrogate endpoint may be highly associated with the true
endpoint, but the association could differ for different interventions, leading to a poor surrogate endpoint.

Secondly to validate a surrogate endpoint, one needs data from multiple trials with both surrogate and true endpoints. The reason is that the relationship of surrogate to true endpoint may vary over trials and so it is necessary to capture this variability. A relatively simple approach involves a weighted average of predictions based on surrogate and true endpoints in each previous trial and the surrogate endpoint in the new trial.

Markers for the Early Detection of Cancer

In recent years the goal of many biotechnology studies has been the identification of markers for the early detection of cancer. Special statistical considerations are required in the design and analysis of these studies. A particularly informative study design involves stored specimens obtained from subjects with clinically detected cancer and stored specimens from subjects without cancer. The outcome measures are: 1 the true positive rate (sensitivity), the fraction of subjects with cancer who are positive for the marker or marker combination in the stored specimens; and 2 the false positive rate (one minus specificity), the fraction of subjects without cancer who are positive for the marker or marker combination in the stored specimens. It is important to report both true and false positive rates, as one without the other is a meaningless quantity. Also, for markers for the early detection of cancer, the target value for the false positive rate should be very small, (of the order of 1% or 2%). The reason is that, in the next phase, a promising marker would be tested as a trigger of early intervention in asymptomatic persons where typically only a 1% to 2% rate of unnecessary biopsies would be acceptable.

Classification Using High-throughput Technologies

One area of biotechnology that has grown rapidly in the last five years is high-throughput technology, in which measurements are taken on thousands of markers such as gene expression arrays or proteomics. One important application has been in the realm of classification where the goal is a highly accurate rule for classifying subjects into previously defined categories such as tumor or not, or survived three years or not. Both computer scientists and statisticians have developed algorithms for creating classification rules. An important contribution from statistics is an appreciation of the variability of the results and the need for good experimental design. An important contribution from the clinician scientist is the consideration of biologic plausibility in guiding the development of classification rules. When searching for the best classification rule among thousands of possible markers, there is a high probability of finding a rule that performs well simply by chance and will thus have poor performance in practice. Therefore, it is best to randomly divide the data into separate training and test samples. One should develop the classification rule in the training sample and evaluate its performance in the test sample. To get an even better sense of the variability of the results, one can repeat this
procedure for different randomly selected training and test samples. To avoid misleading results, it is also important to handle specimens the same way from each category used for classification.

**Conclusion**

These examples illustrate the importance of biostatistical thinking and clinical partnership in evaluating biotechnology. Without this partnership, the study product may be either statistically or clinically naive. Because all biotechnology products require rigorous evaluation, familiarity with these concepts and methods is important for biotechnologists.

**References**


