TESTING OF NEW DRUGS

The issue of the *Journal of the American Medical Association* for July 8, 1961 is a Therapeutic Number, the first of an intended annual

edition featuring the work of the A.M.A. Council on Drugs. In the first special article Isaac Starr,[1] the chairman of the Council, relates a classical example of therapeutic error of judgment, the ideas underlying the "experimental method" involving animal experiments and experiments on man, the role of statistics, and the elimination of the personal bias of the observers by means of the paired experiment and more elaborate experimental designs—with discussion of both the advantages and the difficulties and limitations of the latter.

### STATISTICS

Most of the section on statistics, which tells of what statistical analysis can and cannot do, is reproduced below, by permission.

Training in mathematics has never been a necessary or conspicuous part of the doctor's education, and one wonders how well the average practitioner is equipped to interpret the statistical analysis which accompanies the best medical papers of today. The danger is that, not fully understanding the mathematical background, doctors will attribute to such analysis virtues that it does not possess. So it seems wise to point out the difficulties and dangers as well as the advantages inherent in the statistical methods on which we have placed so much reliance in our testing of therapeutic agents in recent years.

The validity of the results secured by statistical methods depends on an assumption. For the results to be trustworthy, the variation in the data subjected to such analysis should obey the law of probability, the law which defines the expected frequency of any result when dice are cast repeatedly. Needless to say, this law does not always accurately describe the variability found in medical data; for example, it would not apply to a situation in which the magnitude of the errors in one direction exceeds that in the other. Often, when we must draw conclusions from medical data, we do not know enough about the inherent errors to make a firm judgment whether they do or do not conform to the law of probability. We usually assume that they do, and proceed to draw our conclusions on that basis. This risk has proved well worth while.

I pause to re-emphasize that statistics can never prove anything. The results of a statistical analysis demonstrate one thing only, namely, the likelihood that chance will explain the phenomena under study. If the odds are large that chance will explain what happens after giving a new drug, the drug is not worthy of further study. If the odds are large that the results cannot be so explained, the drug may, or may not, be worthy of further study. The important fact for doctors to remember is that the value of statistics is negative; it helps us to eliminate from our thoughts many things which doctors in the past thought worthy of consideration in order to concentrate our efforts in directions with greater promise of success.

There is a danger that doctors, unfamiliar with mathematical thought and methods, will expect too much of statistics. A difference between the mortality from pneumonia in patients who received and those who did not receive a new drug may be statistically significant and therefore should not be attributed to chance; however, the cause of the difference may be something other than the drug. To use as illustration a matter keenly debated at this moment, the relation between the increasing frequency of lung cancer and the increasing consumption of tobacco has proved to be significant, and one has the right to wonder whether the one is a factor in the causation of the other. This might indeed be true and, as the statistics indicate, the matter is well worthy of further study. But the yearly increase in frequency of lung cancer is also significantly correlated with the yearly increase in the number of automobiles manufactured in Detroit, with the

[1] STARR, I. The testing of new drugs and other therapeutic agents. J. American Med. Assoc. **177** (1961) 14-22.

number of nylon stockings sold yearly, and, indeed, with everything else in this growing country which is increasing yearly. The demonstration that the relation between two variables cannot be explained by chance is not valid evidence that the one causes the other. Statistics never prove a causal relationship. The judgment that tobacco smoking may be a factor in lung cancer, and that the sale of stockings is not likely to be, is not based on the statistics, but is based on reasoning of another kind. I myself have stopped smoking, but I continue to buy stockings.

Nevertheless, the value of statistics to modern medicine has been very great indeed, and no modern study of therapeutic agents can be considered complete without the criticism involved in such an analysis. This rigorous criticism has caused us to prune out a great body of data which many doctors had thought worthy of attention, to the great benefit of the tree of knowledge. And the rigorous mathematical thought, which has taught us to apply the statistical method to our problem, has led to many advances.

Later, after discussing controlled experiments, he says:

[Such experiments] will permit identification of the drug which has but little beneficial action, when simpler methods would fail. On the other hand, because the results of such an experiment do not demonstrate that a certain drug exerted a "significant" beneficial effect on a series of patients, one has no logical right to conclude that it is worthless, though persons of the reformer type often so conclude. When statistical significance is not attained, the logical conclusion is not condemnation but that judgment should be suspended. Those who insist on waiting for statistical significance before approving should logically wait for statistical significance, in the other direction, before condemning.

CONTROLLED EXPERIMENTATION

The author then discusses the "rough experiment" in which one-half of a group of patients is given a drug and the other half a placebo (or, sometimes, an established drug with which the new one is competing). This experiment may suffice if the effect of the new drug is striking, but unforeseen factors may cause so much variability as to mask the effect of the drug. Better is the experiment in which each of the patients is treated during alternate periods with the drug or the placebo (which one to use first to be determined by the toss of a coin), and calculating the differences. This type of experiment is impossible in many types of studies. In such case matched pairs may be used for the comparison, but again there are difficulties especially in finding enough sufficiently-well matched pairs.

More elaborate experiments are discussed, as the "double blind" one in which neither the patient nor the observer knows whether the drug or the placebo is given. To find enough adequately paired patients requires large numbers of cases, and such experiments are often set up cooperatively in several clinics. Trained teams are needed, and these usually consist largely of nurses and technicians. Such studies, usually designed and presided over by a statistician, have become important in recent years. Unfortunately, they are very expensive, very difficult to organize, and time-consuming—also, it is very boring and unrewarding for keen-minded physicians to do such work with no knowledge of what is going on.

These and other considerations lead the author to suggest that the

individual physician, if trained and unbiased, may still have a place in therapeutic research.

My expectation is that elaborate team research, despite the many advantages, will not supplant research carried out by the single doctor, who, in the full knowledge of what he is doing, gives therapeutic agents to patients he knows intimately and carefully and objectively observes the effects that follow. We all know the enthusiast so biased by his emotions that he confuses his hopes with the truth. But do we not also know the man whose intellectual attainment has put him far above this, the man who has schooled himself to see things exactly as they are? . . . When such a trained observer gave a drug, and the result on his patient . . . was much different from that which his learning and experience had led him to expect, his attention was attracted. From following up such unexpected experiences, the profession has learned much. . . . In contrast, the eleborate team experiment, as at present designed, diverts the attention from the unusual response, and the man at the top who writes the paper may be altogether unaware of it. The great therapeutic discoveries in the past have been made by competent men, not by teams working blindly. I fully expect that this will be the case in the future.