RESEARCH TECHNIQUES IN THE STUDY OF HUMAN BEINGS

William G. Cochran[1]

WHEN Dr. Boudreau invited me to speak tonight, he happened to mention some of the hopes expressed by Dr. Sydenstricker when the Milbank Memorial Fund's division of research was established in 1929. I was struck by one phrase which Dr. Sydenstricker used: he referred to "the possibility of including social data in the domain of scientific research." This phrase set me to trying to sort out my impressions of the quantitative study of human beings as a supposed branch of science. How well is it progressing relative to other branches of science?

Consequently, I would like to present a few of these impressions, with particular reference to the tools of measurement and the general methods of investigation that have been developed.

The claims of the study of social data to be regarded as a branch of science were examined in the 1830's. The occasion was an application made to the British Association for the Advancement of Science to form a section in statistics. In those days, statistical data dealt largely with economic or social matters. The Association appointed a committee to report on the application, and one of its tasks was to consider whether statistics *was* a branch of science. The committee's verdict is interesting. So long as statistics confined itself to the collection, tabulation, and orderly presentation of data, that was science. But if statistics were to concern itself with the interpretation of economic and social data, that would be argumentation, with passions and politics entering, and that was not science. In the picturesque language of the committee, the interpretation of such data could not be allowed as a branch of science, "lest we admit the foul demon of discord into the Eden of philosophy."

The same point of view was maintained a few years later

[1] Professor of Biostatistics, The Johns Hopkins University.

when a statistical society (which became the Royal Statistical Society) was formed in London. The Committee's verdict was in fact embodied in the motto of the new society. This motto consisted of a fat sheaf of wheat, representing the abundant harvest of data that has been collected and tidily arranged. Around the motto was an ornamental ribbon, like the ribbon worn by Miss Atlantic City in the Beauty contests. But in place of the words "Miss Atlantic City" was the Latin motto "Aliis exterendum"—"Let others thrash it out." I am slightly embarrassed that the statisticians should have started their organized career by timidly proclaiming to the world what they will *not* do. The motto is also curious in that the chairman at the early meetings in which statistical organization was discussed was a man well-known to some of you, by the name of Thomas Malthus. It is true that he left much material for others to thrash out, but he did a certain amount of "thrashing out" himself.

Since I shall speak from the viewpoint of the statistician—I can't help it—I must first say a little about statisticians and their relations to scientists. The statistician has long been known as a person who handles data, and the scientist tends to think of seeing a statistician when he has some problem in the analysis of his data. In earlier days, this happened mostly when something had gone wrong with the experiment or survey—or more accurately when the scientist realized that something had gone wrong. As a result, statisticians used to see a sorry collection of the wrecks of research projects.

Now it is a hard fact, which the statistician and the scientist both had to learn, that little could be done to get these wrecks floating again. Usually, some error in the way in which the data were collected made it impossible to draw sound conclusions, manipulate them how you will. This led to two developments. The statistician began to advise scientists to come and see him at the beginning of an investigation—to make him, as it were, an accessory before the fact. Also, the statisticians began to study the process of collecting data in order to learn what procedures and precautions were necessary to ensure that sound conclusions *could* be drawn at the end.

The result is that, at present, the role played by the statistician in the planning of research is often that of verifying that the scientific methodology is sound and sometimes even that of supplying the scientific methodology. Of course, the statistician has other duties. He helps with the arithmetic, tells where the decimal point goes, and he may supply technical formulas from statistical theory, but these are often secondary contributions. Perhaps the statistician's role as a consultant in scientific methods is temporary, because one would expect scientists to perform this function themselves. Indeed, there are signs of a trend in this direction. A few years ago, a conference with a physician about the testing of new drugs on hospital patients seemed to be mainly a matter of trying to wheedle or cajole the physician into taking some precautions that he regarded as a nuisance and as unnecessary. Now he is often found insisting on these precautions himself before the statistician can open his mouth, and the statistician's contribution is to nod his head in agreement at diplomatic intervals. In time, the doctors may decide that they don't need this yes-man.

While it lasts, this role requires close and friendly cooperation between statisticians and research workers. In statistical training centers, something is done to teach young statisticians how to get along with scientists. While in the presence of a number of distinguished scientists, I would like to give a few hints on how to get along with statisticians. In any extensive discussion of a statistical problem, some wag is likely to repeat the old chestnut about the three kinds of lies: "lies, damn lies, and statistics." If you feel an urge to give birth to this witticism, please remember that it is not new, and it was not funny when it was new. The statistician also gets tired of hearing the scientist say "Of course, I am no statistician," in a tone of voice which implies that he is mentioning one of his most sterling virtues. If you *are* no statistician, this fact will probably reveal itself in the course of the conversation, and if you must tell us about it, please do so with an apologetic air. Remember also that the statistician is a poor marriage risk, and may be suffering from marital strains. The reason for this is that the statistician has to cultivate a dislike for imprecise statements, and the person most likely to be making imprecise statements in his vicinity is his unfortunate wife. Statisticians' wives have not, thank goodness, formed an international union, but if they do, the first plank in its platform will be to stop their husbands from being so persnickety.

Tools of Measurement

The consulting statistician has a fascinating opportunity to learn something about the triumphs and the difficulties of research in different branches of science. He begins to wonder why some branches are forging ahead in an exciting way, while others seem to be creeping. Among the numerous factors that influence the rate of progress of scientific research, two of the most important are the tools of measurement that the research worker has at his disposal and the general methods of investigation available to him.

I shall use the phrase "tools of measurement" in a broad sense, to cover both the range of phenomena that can be measured and the precision of the measuring devices. It can be argued that the available tools of measurement are the most important single factor in determining the rate of progress in a field of research. I do not wish to build up this argument, but one example may be quoted from physics. Towards the end of the last century, the laws of physics seemed to have reached a pinnacle. They were of high accuracy, of immense scope, and were pleasing to common sense. Then improvements in measurement were made that enabled very minute bodies as well as very distant bodies to be more accurately studied. Large cracks appeared in the edifice of physical theory, and to the rescue the physicists had to bring in the revolutionary ideas of quantum mechanics and the theory of relativity. To judge from their difficulties in understanding these concepts and in reconciling them with simple common sense ideas, they must have felt at times as if they had brought in the Marx brothers to repair the building.

The importance that is rightly attached to an improvement in measuring technique is illustrated by the recent award of the Nobel prize in medicine to Dr. Enders and his associates. Their contribution was to grow poliomyelitis virus in tissue culture, and other workers helped to perfect the technique. What does this mean to research in the field? In measuring virus concentrations in specimens from suspected polio cases, as many tests can now be made from the kidneys of a discarded monkey as required 600 monkeys previously. The monkey can be dispensed with entirely, by use of the Hela human cancer cells. Experiments that were impossible can now be done in a week. After a development of this type, any area of research can expect to take great strides forward.

A second impression about measurement techniques is that one never knows where the next advance is coming from. Often it does not come from the field of work that desperately needs the advance. The anthropologists, after measuring skulls from every angle with admirable zeal, have to thank the geneticists for blood group methods that for some purposes are much more reliable. The paleontologists were presented with a new and independent method of dating fossils—radioactive carbon—by the physicists. The electron microscope is a godsend to the manufacturers of paint. And so on.

In the study of human beings many of the problems of measurement are formidable. Not only have we to measure fairly concrete attributes like the state of disease in the individual, (which the doctors will assure us is not easy to measure well) but we need to classify and if possible to measure many things that are hard enough to define in the first place, like motives, morale, intentions, feelings of stress. This means a vast undertaking that has had to start from the ground with rude homemade tools. Thus far, for want of anything better in sight, we have obtained our raw data mainly from what the individual tells us. And the recording instrument has usually been another individual.

We are having to learn about the idiosyncrasies of the human being as a reporter. On the whole, he is surprisingly cooperative, and his good nature in taking up his time to talk to us is heartwarming. He is, in fact, a little too friendly, and will sometimes give the kind of answers which he thinks we would like to have. He is anxious to put on a good front: his statement about the amount he paid for his present car is not entirely to be trusted, and his plans for buying all sorts of expensive gadgets in the future are still less so. On the other hand, he can shut up like a clam. At the end of the war, I helped to gather some data from a carefully selected sample of the German civilian population. According to our results, the Nazi party was one of the world's most exclusive clubs. He is loyal to those whom he likes. The English, in their industrial mortality statistics, were puzzled by the fact that the death rate for the drivers of railway engines (i.e. the locomotive engineers) was above the national average, while that for the man with the apparently less healthy job of stoking the coal furnace was well below. The explanation was that father, after a life of service as stoker, was often posthumously promoted by mother to the position of engine driver on the death certificate.

The recording device—the interviewer—is not perfect either. A quotation from Bertrand Russell, although rather overdrawn, illustrates this point. He is writing about studies of learning in animals. "The animals that have been carefully observed have all displayed the national characteristics of the observer. Animals studied by Americans rush about frantically, with an incredible display of hustle and pep, and at last achieve the desired result by chance. Animals observed by Germans sit still and think, and at last evolve the solution out of their inner consciousness. To the plain man, such as the present writer, this situation is discouraging."

As instances of the amount that has to be learned in order to make the best use of human beings as reporters and recorders, the following are some, but by no means all, of the questions that arise in the planning of morbidity surveys in which the data are obtained by interviews in the home. Over what period of time can the subject remember episodes of illness? What types of illness are easily remembered and accurately reported, and what types are poorly reported? What aids to memory are worth while? How well does the housewife remember and report illnesses of other members of the family? To what extent can the reports be used for a diagnostic classification of the illnesses? How much is gained by checking the reports with physicians who have attended the families? How do lay interviewers compare in effectiveness with public health nurses or medical students? How much information can be picked up at a second visit that was missed at the first? Since a substantial amount of experience has been accumulated for morbidity surveys, at least partial answers can be given to these questions. In other words, something is now known about the precision and the limitations of this type of measuring tool, and about good and bad ways of applying it. Research on more difficult concepts like attitudes and sources of motivation will in time have to answer an analogous list of questions about the interviewer-respondent relationship.

Social scientists are attacking vigorously the fascinating problems involved in devising ways of classifying and measuring what might be called, for want of better words, the strengths and directions of opinions, attitudes, and feelings. They are making surprisingly early use of quantitative scales, with an implied continuum in the background, and have shown ingenuity in constructing methods for testing the internal consistency of the scale and for checking how well the scaled results can be reproduced from a second examination of the same group of people. The criticism has been made, with some justification, that these scales may deceive research workers into thinking that they have measured some rather intangible quantity that they are nowhere in sight of measuring. I don't think that the difficulty arises from the use of quantitative scales themselves: the dangers in pushing this process too fast do not seem to me great. It would be well, however, to be cautious and humble in making claims about what we have measured. Until we are very sure of our ground, use of long Greek names for the things measured might be preferable, rather than claiming to have measured, say, the strength of maternal affection.

General Methods of Investigation

Methods of investigation in scientific research can be classified roughly into three types, which may be called chance observations, planned observations, and experiments.

Chance Observations. Something unusual strikes the curiosity of an alert scientist, and off he goes into a chain of speculation and then into action. Many of you have heard Sir Alexander Fleming's account of the beginnings of his discovery of penicillin. He happened to notice an unusual contamination from the air of some plates lying in his laboratory. The contribution of chance observations to progress in science must be very great. Last week I was talking to a productive scientist who had had occasion to review carefully his work during the past fifteen years. He remarked that, to his surprise, all his most important discoveries had arisen in unexpected deviations from his main path of research. None of them would have appeared in that anathema of the modern scientist—the "Statement of work to be done during the next fiscal year."

Planned Observations. Here the scientist knows what he is after—he knows the questions to which he would like answers—and he maps out a plan of observation which he hopes will provide the answers. Some of the current investigations of the relation between smoking and cancer of the lung are of this type. In the British Medical Research Council's study, all the British doctors were asked three years ago to fill out a questionnaire giving their ages and their recent smoking habits. The rest of the study is just a matter of waiting until a reasonable proportion have died, and then examining whether the death rate and the causes of death are related to smoking habits. Doctors present many advantages for this kind of study: they are likely to cooperate, it is relatively easy to find out if they have died, and when they do die there is reason to believe that the cause of death will be more accurately known than for laymen.

Experiments. The word "experiment" has a very broad meaning both in common speech and among scientists. For my present purpose I would like to restrict it to situations in which we are able to *interfere* with nature. In this sense, the essence of an experiment is that we deliberately apply certain chosen procedures for the purpose of measuring their effects. The power of experimentation in speeding up progress in science is tremendous. It has two strong advantages over the observational method. It enables us to select for investigation the factor or factors that will be most informative, whereas with observations, we are restricted to those factors that nature is kind enough to give us the opportunity to observe. The experiment is also the surest method of working out the causal relations that underlie the associations which we observe. With the observational method, the step from correlation to causation is often hazardous and uncertain. For instance, even if several studies in different countries should reach the common conclusion that the death rate from cancer of the lung increases steadily as the amount of smoking increases, the objection will be made (in fact, it has already been made) that this is not a cause and effect relationship, because of the alternative possibility that the kinds of men who smoke heavily are unusually susceptible to cancer of the lung, and would be so even if they did not smoke. Whatever our opinions about the plausibility of this explanation, it is hard to devise an observational study that will clearly support or rebut it. If experimentation were possible, the issue could be cleared up much more easily.

In the study of human beings we are groping our way around among these general methodologies, trying to find which ones pay off best in results. Thus far, observational methods have been used to a large extent, since opportunities for experimentation appear limited.

In particular, we are having to learn how much can be obtained from past data, originally gathered for some other purpose, for example, in connection with the adminstration of a program. Since the data are already there, the method is much speedier than a fresh start would be. In cost, it may mean the difference between $5,000 and $150,000. Although the past data are seldom what we would like to have if we were doing the job anew, yet often there are masses of it, and perhaps it will be possible to select what we need.

Although it is difficult to generalize, experience with past data has been disappointing. It has often given a confused picture from which no clear leads can be drawn, and it has sometimes given leads that turned out to be the wrong ones. The main difficulties appear to be that the definitions used in the data are not rigorous and clear-cut enough for scientific investigations, and that the effects that we wish to study are inextricably tangled up with other effects. Some of my own disappointments with past data remind me of a statement made by Available Jones in the Little Abner comic strip. Available Jones makes his living in part by giving advice. He has two kinds of advice, the 10-cent and the 50-cent kind. Of the 10-cent kind, he says (after some modifications of his spelling): "For 10 cents, I barely listens—in fact I yawns in your face, and the cheap advice you gets will do you more harm than good."

In many human studies, workers are realizing that they must face the long and hard business of planning new observations in order to obtain the 50-cent advice. I do not mean, however, to condemn the use of past data in any outright manner: if a few factors predominate, this should be revealed, and very often, past data are all that we have. Moreover, I owe my first post, in the depression, mainly to the fact that my employers had a large batch of past data which were regarded as a potential mine of information. They hired me to dig it out. I dug furiously, but I doubt whether they received their money's worth. Fortunately, my salary was so low that this moral problem caused me no loss of sleep.

In new studies, we are having to learn how much ground can be covered, that is, how many different questions can be investigated in a single study. At the moment, the lesson seems to be not to be too ambitious. This can be illustrated with respect to one approach to exploratory studies that might be called the method of casting the net widely, if you happen to like it, or the method of shooting blindly in all directions, if you don't happen to like it. Suppose that there is some phenomenon about which not much is known, and we are trying to discover which factors or variables have the most predominant influence on it, or are at least most clearly associated with it. It seems rational to write down all the factors that are likely to have an influence on the phenomenon, include them in the study, and rely on statistical techniques, particularly those of multiple classification or regression, to reveal the most important ones. I know of no one as clever as the social scientist at writing down a ten-page list of factors that might influence any given phenomenon. For the relatively poor results given by this method, the statisticians may be partly to blame, because they may have oversold the power of statistical techniques to unscramble an omelet. If nature mixes things up thoroughly, as she sometimes seems to do, statistical methods will not sort them out very well. Indeed, the more factors that are included in the study, that is, the more painstaking the scientist is, the harder it becomes to disentangle all their effects. Many studies now go to the opposite extreme, concentrating on learning something about a single factor, such as differences between premature and normal children, or between public and slum housing. This means slow progress, and perhaps with more experience some intermediate method will prove rewarding.

Social scientists are having to learn how to observe the same people over a period of years, as in the study of chronic diseases or of the effects of administrative programs. Such studies are expensive and hazardous, because it is difficult to foresee the contingencies that may arise to plague us. For one thing, the human subjects won't stay where they are: off they go to Portland or Honolulu, and if we cannot find means to keep observing them, the group under study dwindles year by year to a remnant consisting of the most settled families. Sometimes it is the scientist who is off to Rangoon or Monte Carlo. My guess would be that we now know how to observe groups for as long as three years, and perhaps for as long as five years: beyond that, there are too few successfully completed studies to be able to say that the technique has been mastered.

In such long-term studies the subjects are sometimes influenced by the fact that they are being studied, in a way that vitiates the purposes of the study. I have heard of farm management studies of poor farmers where the list of questions opened the eyes of some of the farmers to financial opportunities that had never occurred to them. In a few years, these farmers were offering the interviewers jobs. In the British study of smoking and cancer, the Medical Research Council's scientists became alarmed at the number of doctors who replied to the original questionnaire by saying "I have been smoking twenty cigarettes a day, but after reading this questionnaire I have given up smoking for ever."

These long-term studies require, for their direction, a type of scientist who is quite different from the "ivory tower" concept of a scientist. He must be able to assemble a team of workers and to maintain good relations among them: he must obtain the cooperation of various administrators and their agencies, and must handle a considerable amount of paper work. Scientific competence alone does not guarantee success in this type of research: some scientists are too shy, and others too quarrelsome, to meet the requirements.

Social scientists are also having to learn to exercise the kind of ingenuity that is delightful when it comes off. Nature occasionally provides golden opportunities to study some group that will be particularly revealing, as with identical twins who have been reared under different circumstances, or with groups of people who have been long isolated. Ingenuity may also enable us to take the difficult step from correlation to causation. If we have established correlation between two variables A and B in an observational study, we may *think* that A is the cause of B, but nobody saw the murder committed, and the evidence pointing to A as the culprit is only circumstantial. But if by ingenuity we can build up a series of separate pieces of evidence, all pointing to A, it becomes harder and harder to think of an alternative hypothesis that will explain them all away simultaneously. In this connection the social scientist has to use the methods of the detective, the good criminal lawyer, and even the man who is trying to prove that Bacon wrote Shakespeare.

As I have indicated, the use of the more powerful method of experimentation has been small. The obstacles with human subjects are obvious. Yet with persistence and tact, the difficulties can sometimes be overcome, and it may be that experimentation will come to play a more important role than it now does. In medical research on the prevention and cure of disease, some notable successes have been scored by experimentation, and experiments are now being attempted that would, I believe, have been considered impossible a few years ago. The main problem is to secure the tightness of control that is essential for a good experiment, without relaxing the ethical requirement that the welfare of the patient is the paramount consideration.

The trial of the polio vaccine conducted this summer is an example. In some of the states, this trial was made by a method that I would describe as observational, but in others, involving hundreds of thousands of children, the trial was a genuine experiment. The children were divided into two groups at random. Those in one group received, at intervals, three shots of the vaccine. Those in the other group received in the same way three shots of an inert substance that is expected to have no effect. No one in the areas concerned knows which child received vaccine and which control. This, in fact, is known to very few persons, and it will not be revealed until necessary in the final stages of the analysis.

A second example is an international cooperative experiment on drugs for the treatment of leprosy, conducted under the leadership of the Leonard Wood Memorial, that was an organizational masterpiece. The same six drugs were tested at the same time in three different institutions, with uniform methods of measuring and recording the dermatological, neurological and bacteriological progress of the patients, and with random allotments of patients to drugs. One institution was in Japan, one in the Philippines, and one in Pretoria, South Africa. In fact, the chief barrier to progress in this line of research is probably a deficiency in tools of measurement. Since no experimental animal has been found in which leprosy can be studied in the laboratory, it is difficult to obtain clues as to the most promising types of drug to test in the future.

Summary

The quantitative study of human beings, particularly in their social aspects, is a young field. Because of the multitude of critical problems in human relations facing the world today, research workers are trying to obtain helpful answers on practical questions with rather crude tools of measurement and none too powerful methods of investigation.

In hazarding a few suggestions about the use of resources in this area of research, I should make it clear that I have not surveyed the present use of resources in any adequate way. It may be that my suggestions are already being prosecuted as vigorously as seems worth while.

The field needs to devote ample resources to improving its tools of measurement. This is best done by workers who do not have to produce answers to practical questions at the same time. Raymond Pearl used to urge biologists to stop beating their breasts about the difficulty of doing accurate work in biology. If the biologists would devote as much brains, energy and care to refining their measurements as the physicists do, he

claimed that they would obtain as accurate results. Although I think he promised too much, the amount of research that physicists devote to measuring devices as such is impressive, and the returns are equally so. In the social sciences, the work of the psychometricians in the construction of scales is a good beginning. I have heard some hard words about the Rorschach test, but both orthodox and unorthodox methods of measurement should be developed and tried.

Experimentation (in the sense in which I have used it) needs to be exploited as much as possible. The question: "Why can't I do an experiment?" is always worth asking, even if it sounds unrealistic. There may be many opportunities for simple experiments using students as volunteers. A colleague, one of the few men still working on the discouraging task of producing a vaccine for the common cold, finds his volunteer subjects among the convicts.

A balance should be retained between studying what people say they will do and studying what they actually do. Here there is perhaps a contrast between economics and sociology. The economist has kept a close eye on what people do, but has tended to rely on armchair reasoning to uncover the motives for their actions, to the neglect of attempts to study motives independently. The sociologists have been enterprising in tackling the difficult task of studying motives, but they need also to be constantly checking reported motivations against actions.

In addition to scientists engaged in large-scale studies, the field needs a supply of those German animals (in the quotation from Russell) who sit still and think. These might be younger scientists with steady incomes, but with restricted research budgets.

Finally, the field needs to keep strong lines of communication with other branches of science, and particularly with biology, and to recruit some of its research workers from these other branches. For certain research problems, the "interview" method of obtaining data is likely to prove inadequate, and progress may have to await new measuring techniques that are adapted from developments outside of social science. The need for links with biology is obvious: man is biological as well as social; moreover, although biology has access to more powerful and flexible research techniques than social science, many of the problems are the same.

In conclusion, I hope that my comments have not sounded pessimistic. If there is one lesson to be learned from the history of science, it is that the optimists are always right, except that they should have been more optimistic.