Why was such elaboration necessary? Did it really result in more or better knowledge than could have been obtained from much simpler studies? These are the questions on which this discussion is focused.

BACKGROUND

Polio was never a common disease, but it certainly was one of the most frightening and, in many ways, one of the most inexplicable in its behavior. It struck hardest at young children, and, although it was responsible for only about 8% of the deaths in the age group 3 to 9 in the early fifties, it left many helpless cripples, including some who could survive only in a respirator. It appeared in epidemic waves, leading to summer seasons in which some communities felt compelled to close swimming pools and restrict public gatherings as cases increased markedly from week to week; other communities, escaping an epidemic one year, waited in trepidation for the year in which their turn would come. Rightly or not, this combination of selective attack upon the most helpless age group and the inexplicable vagaries of its epidemic behavior, led to far greater concern about polio as a cause of death than other causes, such as auto accidents, which are more frequent and, in some ways, more amenable to community control.

The determination to mount a major research effort to eradicate polio arose in no small part from the involvement of President Franklin D. Roosevelt, who was struck down by polio when a successful young politician. His determination to overcome his paralytic handicap and the commitment to the fight against polio made by Basil O'Connor, his favorite law partner, enabled a great deal of attention, effort, and money to be expended on the cause and rehabilitation of polio victims and—in the end, more importantly—on research into the causes and prevention of the disease.

During the course of this research, it was discovered that polio is carried by a virus and that three main virus types are involved. Although clinical manifestations of polio are rare, it was discovered that the virus itself was not rare, but common, and that most adult individuals had experienced a polio infection sometime in their lives without ever being aware of it.

This finding helped to explain the otherwise peculiar circumstance that polio epidemics seemed to hit hardest those who were lower off hygienically (i.e., those who had the best nutrition, most favorable housing conditions, and were otherwise apparently most favorably situated). Indeed, the disease seemed to be virtually unknown in those communities with the poorest hygiene. The explanation is that because there was plenty of polio virus in the broad-based population, almost every infant was exposed to the disease early in life while he was still protected by the immunity passed on from his mother. As a result, everyone had had polio, but under protected circumstances, and thereby, everyone had developed his own immunity.
As with many other virus diseases, an individual who has been infected by polio and recovered is usually immune to another attack (at least by a virus strain of the same type). The reason for this is that the body, in fighting the infection, develops antibodies, which are a part of the gamma globulin fraction of the blood, to the antigen, which is the protein part of the polio virus. These antibodies remain in the bloodstream for years, and even when their level declines so far as to be scarcely measurable, there are usually enough of them to prevent a serious attack from the same virus.

Smallpox and influenza illustrate two different approaches to the preparation of an effective vaccine. For smallpox, which has long been controlled by a vaccine, we use the vaccine a closely related virus, cowpox, which is ordinarily incapable of causing serious disease in man, but which gives rise to antibodies that also protect against smallpox. (In a very few instances this individual is capable of overcoming a severe and occasionally fatal reaction. The risk is small enough, however, so that we do not hesitate to expose all our school children to it in order to protect them from smallpox.) In the case of influenza, however, instead of a closely related live virus, the vaccine is a solution of the influenza virus itself, prepared with a live virus that has been killed by treatment for a time with formaldehyde. Provided that the treatment is not too prolonged, the dead virus still has enough antigenic activity to produce the required antibodies so that, although it can no longer infect, it is, in this case, sufficiently like the live virus to be a satisfactory vaccine.

In the case of polio, both of these methods were explored. A live-virus vaccine would have the advantage of reproducing in the vaccinated individual and, hopefully, giving rise to a strong reaction which would produce a high level of long-lasting antibodies. With such a vaccine, however, there might be a risk that a virus similar to the vaccine virus could mutate into a virulent form and itself be the cause of paralytic or fatal disease. A killed-virus vaccine should be safe because it presumably could not infect, but it might fail to give rise to an adequate antibody response. These and other problems stood in the way of the rapid development of a successful vaccine. Some unfortunate prior experience also contributed to the cautious approach of the vaccine developers. Attempts had been made to develop vaccines against polio; two of these were actually in use for a time. Evidence that at least one of these vaccines, in fact, had been responsible for cases of paralytic polio soon caused them to be promptly withdrawn from use. This experience was very much in the minds of polio researchers, and they had no wish to risk a repetition.

Research to develop both live and killed vaccines was stimulated in the polio epidemic. Those working with live preparations developed harmless strains from virulent ones by growing them for many generations in suitable tissue cultures. There was, of course, considerable worry lest these strains, when used as a vaccine in man, might revert to virulence and cause paralysis or death. (By 1952 it seemed clear that the strains developed were indeed safe—a live-virus preparation taken orally is the vaccine presently in widespread use throughout the world.) Those working with killed preparations, notably Jonas Salk, had the problem of treating the virus (with formaldehyde) sufficiently to eliminate its infectiveness, but not so long as to destroy its antigenic effect. This was more difficult than, at first, had appeared to be the case, and some early lots of the vaccine proved to contain live virus capable of causing paralysis and death. There are statistical issues in the safety study (Miezer, 1957), but our concern here is with the evaluation of effectiveness.

EVALUATION OF EFFECTIVENESS

In the early fifties the Advisory Committee convened by the National Foundation for Infantile Paralysis (NPIP) decided that the killed-virus vaccine developed by Jonas Salk at the University of Pittsburgh had been shown to be both safe and capable of inducing high levels of the antibody in children on whom it had been tested. This made the vaccine a promising candidate for general use, but it remained to prove that the vaccine actually would prevent polio in exposed individuals. It would be unjustified to release such a vaccine for general use without convincing proof of its effectiveness, so it was determined that a large-scale "field trial" should be undertaken.

That the trial had to be carried out on a very large scale is clear. For suppose we wanted the trial to be convincing if indeed the vaccine were 90% effective (for various reasons, 100% effectiveness could not be expected). Assume that, during the trial, the rate of occurrence of polio would be about 50 per 100,000 (which was the average incidence in the United States during the fifties). With 40,000 in the control group and 40,000 in the vaccinated group, we would find about 20 control cases and about 10 vaccinated cases, and a difference of researchers. In the thirties, of paralytic polio soon caused them to be promptly withdrawn from use. This experience was very much in the minds of polio researchers, and they had no wish to risk a repetition. Research to develop both live and killed vaccines was stimulated in the polio epidemic. Those working with live preparations developed harmless strains from virulent ones by growing them for many generations in suitable tissue cultures. There was, of course, considerable worry lest these strains, when used as a vaccine in man, might revert to virulence and cause paralysis or death. (By 1952 it seemed clear that the strains developed were indeed safe—a live-virus preparation taken orally is the vaccine presently in widespread use throughout the world.) Those working with killed preparations, notably Jonas Salk, had the problem of treating the virus (with formaldehyde) sufficiently to eliminate its infectiveness, but not so long as to destroy its antigenic effect. This was more difficult than, at first, had appeared to be the case, and some early lots of the vaccine proved to contain live virus capable of causing paralysis and death. There are statistical issues in the safety study (Miezer, 1957), but our concern here is with the evaluation of effectiveness.

EVALUATION OF EFFECTIVENESS

In the early fifties the Advisory Committee convened by the National Foundation for Infantile Paralysis (NPIP) decided that the killed-virus vaccine developed by Jonas Salk at the University of Pittsburgh had been shown to be both safe and capable of inducing high levels of the antibody in children on whom it had been tested. This made the vaccine a promising candidate for general use, but it remained to prove that the vaccine actually would prevent polio in exposed individuals. It would be unjustified to release such a vaccine for general use without convincing proof of its effectiveness, so it was determined that a large-scale "field trial" should be undertaken.

That the trial had to be carried out on a very large scale is clear. For suppose we wanted the trial to be convincing if indeed the vaccine were 90% effective (for various reasons, 100% effectiveness could not be expected). Assume that, during the trial, the rate of occurrence of polio would be about 50 per 100,000 (which was the average incidence in the United States during the fifties). With 40,000 in the control group and 40,000 in the vaccinated group, we would find about 20 control cases and about 10 vaccinated cases, and a difference of researchers. In the thirties, of paralytic polio soon caused them to be promptly withdrawn from use. This experience was very much in the minds of polio researchers, and they had no wish to risk a repetition. Research to develop both live and killed vaccines was stimulated in the polio epidemic. Those working with live preparations developed harmless strains from virulent ones by growing them for many generations in suitable tissue cultures. There was, of course, considerable worry lest these strains, when used as a vaccine in man, might revert to virulence and cause paralysis or death. (By 1952 it seemed clear that the strains developed were indeed safe—a live-virus preparation taken orally is the vaccine presently in widespread use throughout the world.) Those working with killed preparations, notably Jonas Salk, had the problem of treating the virus (with formaldehyde) sufficiently to eliminate its infectiveness, but not so long as to destroy its antigenic effect. This was more difficult than, at first, had appeared to be the case, and some early lots of the vaccine proved to contain live virus capable of causing paralysis and death. There are statistical issues in the safety study (Miezer, 1957), but our concern here is with the evaluation of effectiveness.

EVALUATION OF EFFECTIVENESS

In the early fifties the Advisory Committee convened by the National Foundation for Infantile Paralysis (NPIP) decided that the killed-virus vaccine developed by Jonas Salk at the University of Pittsburgh had been shown to be both safe and capable of inducing high levels of the antibody in children on whom it had been tested. This made the vaccine a promising candidate for general use, but it remained to prove that the vaccine actually would prevent polio in exposed individuals. It would be unjustified to release such a vaccine for general use without convincing proof of its effectiveness, so it was determined that a large-scale "field trial" should be undertaken.

That the trial had to be carried out on a very large scale is clear. For suppose we wanted the trial to be convincing if indeed the vaccine were 90% effective (for various reasons, 100% effectiveness could not be expected). Assume that, during the trial, the rate of occurrence of polio would be about 50 per 100,000 (which was the average incidence in the United States during the fifties). With 40,000 in the control group and 40,000 in the vaccinated group, we would find about 20 control cases and about 10 vaccinated cases, and a difference of researchers. In the thirties, of paralytic polio soon caused them to be promptly withdrawn from use. This experience was very much in the minds of polio researchers, and they had no wish to risk a repetition. Research to develop both live and killed vaccines was stimulated in the polio epidemic. Those working with live preparations developed harmless strains from virulent ones by growing them for many generations in suitable tissue cultures. There was, of course, considerable worry lest these strains, when used as a vaccine in man, might revert to virulence and cause paralysis or death. (By 1952 it seemed clear that the strains developed were indeed safe—a live-virus preparation taken orally is the vaccine presently in widespread use throughout the world.) Those working with killed preparations, notably Jonas Salk, had the problem of treating the virus (with formaldehyde) sufficiently to eliminate its infectiveness, but not so long as to destroy its antigenic effect. This was more difficult than, at first, had appeared to be the case, and some early lots of the vaccine proved to contain live virus capable of causing paralysis and death. There are statistical issues in the safety study (Miezer, 1957), but our concern here is with the evaluation of effectiveness.
place to place and time to time, required that the trial involve a huge number of subjects—as it turned out, over a million.

THE VITAL STATISTICS APPROACH

Many modern therapies and vaccines, including some of the most effective ones, such as smallpox vaccine, were introduced because preliminary studies suggested their value. Large-scale trials subsequently provided clear evidence of efficacy. A natural and simple approach to the evaluation of the Salk vaccine would have been to distribute it as widely as possible, through the schools, to see whether the rate of reported polo was appreciably less than usual during the subsequent season. Alternatively, distribution might be limited to one or a few areas because limitations of supply would preclude effective coverage of the entire country. There is even a fairly good chance that were one to try an extensive vaccine campaign against the common cold or against measles, convincing evidence might be obtained in this way.

In the case of polo—and, indeed, in most cases—as simple an approach as this would virtually fail to produce clear-cut evidence. First, and foremost, we must consider how much polo incidence varies from season to season, even without any attempts to modify it. From Figure 1, which shows the annual reported incidence from 1980 through 1955, we see that had a trial been conducted in this way in 1951, the drop in incidence from 1951 to 1952 would have been strongly suggestive of a highly effective vaccine because the incidence dropped to less than a third of its previous level. Similar misinterpretations would have been made in 1935, 1937, and other years—most recently in 1952. (On the general problem of drawing inferences from such time series data see the essay by Campbell.) One might suppose that such missteps could be avoided by using the vaccine in one area, say, New York State, and comparing the rate of incidence there with that of an unvaccinated area, say, Illinois. Unfortunately, an epidemic of polo might well occur in Chicago—as it did in 1956—during a season in which New York had a very low incidence.

Another problem, more subtle, but equally burdensome, relates to the vagaries of diagnosis and reporting. There is no difficulty, of course, in diagnosing the classic respirator case of polo, but the overwhelming majority of cases are less clearcut. Hence, the decision to diagnose a case as paralytic polo instead of some other disease may well be influenced by the physician's general knowledge or feeling about how widespread polo is in his community at the time.

These difficulties can be mitigated to some extent by setting down very precise criteria for diagnosis, but it is virtually impossible to obviate them completely when, as would be the case after the widespread introduction of a new vaccine, there is a marked shift in what the physician expects to find. This is even more especially true when the initial diagnosis must be made by family physicians who cannot easily be indoctrinated in the use of a special set of criteria, as is the case with polo. Later evaluations by specialists cannot, of course, bring into the picture those cases originally diagnosed as something other than polo.

THE OBSERVED CONTROL APPROACH

The difficulties of the vital statistics approach were recognized by all concerned, and the initial study plan, although not judged entirely satisfactory, got around many of the problems by introducing a control group similar in characteristics to the vaccinated group. More specifically, the idea was to offer vaccination to all children in the second grade of participating schools and to follow the polo experience not only in these children, but in the first- and third-grade children as well. Thus the vaccinated second-graders would constitute the treated group, and the first- and third-graders would constitute the control group. This plan follows what we call the observed control approach.

It is clear that this plan avoids many of the difficulties that we listed above. The three grades all would be drawn from the same geographic location so that an epidemic affecting the second grade in a given school would certainly affect the first and third grades as well. Of course, all subjects would be observed concurrently in time. The grades, naturally, would be different ages, and polo incidence does vary with age. Not much variation
The chief objection to this plan was that parents of school children could not reasonably be expected to permit their children to participate in an experiment in which they might be getting only an ineffective salt solution instead of a probably helpful vaccine. It was argued further that the injection of placebo might not be ethically sound, since a placebo injection carries a small risk, especially if the child unknowingly is already infected with polo.

The proponents of the placebo control approach maintained that, if properly approached, parents would consent to their children’s participation in such an experiment, and they judged that the injections would not be given during the polo season, the risk associated with the placebo injection itself was vanishingly small. Certain health departments took a firm stand: they would participate in such an experiment, and they judged that the injections would not be given during the polo season, the risk associated with the placebo injection itself was vanishingly small. Certain health departments took a firm stand: they would participate in such an experiment, and they judged that the injections would not be given during the polo season, the risk associated with the placebo injection itself was vanishingly small.
The second major feature of the experimental method was the assignment of subjects to treatments by a careful randomization procedure. As we observed earlier, the chance of coming down with a diagnosed case of polio varies with a great many factors including age, socioeconomic status, and the like. If we were to make a deliberate effort to match up the treatments and control groups as closely as possible, we should have to take care to balance these and many other factors, and, even so, we might miss some important ones. Therefore, perhaps surprisingly, we leave the balancing to a carefully applied equivalent of coin tossing: we arrange that each individual has an equal chance of getting vaccine or placebo, but we eliminate our own judgment entirely from the individual decision and leave the matter to chance.

The gain from doing this is twofold. First, a chance mechanism usually will do a good job of evening out all the variables—those we didn’t recognize in advance, as well as those we did recognize. Second, if we use a chance mechanism in assigning treatments, we may be confident about the use of the theory of chance, that is to say, probability theory, to judge the results. We can then calculate the probability that so large a difference as that observed could reasonably be due solely to the way in which subjects were assigned to treatments, or whether, on the contrary, it is really an effect due to a true difference in treatments.

To be sure, there are situations in which a skilled experimenter can balance the groups more effectively than a random-selection procedure typically would. When new factors may have a large effect on the outcome of an experiment, it may be desirable, or even necessary, to use a more complex experimental design that takes account of these factors. However, if we intend to use probability theory to guide us in our judgment about the results, we can be confident about the accuracy of our conclusions only if we have used randomization at some appropriate level in the experimental design.

The final determinations of diagnosed polio proceeded along the following lines. First, all cases of poliomyelitis reported by local physicians were subjected to special examination, and a report of history, symptoms, and laboratory findings was made. A special diagnostic group then evaluated each case and classified it as nonpolio, doubtful polio, or definite polio. The last group was subdivided into nonparalytic, paralytic, and fatal polio. Only after this process was complete was the code broken and identification made for each case as to whether vaccine or placebo had been administered.

### RESULTS OF THE TRIAL

The main results are shown in Table 1, which shows the size of the study populations, the number of cases classified as polio, and the disease rates, that is, the number of cases per 100,000 population. For example, the second line shows that in the placebo control area there were 428 reported cases...
but it is more realistic to recognize that such success is but one step in the continuing development of public health science. The Salk vaccine, although a notable triumph in the battle against disease, was relatively crude and, in many ways, not a wholly satisfactory product that was soon replaced with better ones.

The report of the field trial was followed by widespread release of the vaccine for general use, and it was discovered very quickly that a few of these lots actually had caused serious cases of polio. Distribution of the vaccine was then halted while the process was reevaluated. Distribution was reinitiated a few months later, but the momentum of acceptance had been broken and the precept disappearance of the disease that researchers hoped for did not come about. Meanwhile, research on a more highly purified killed-virus vaccine and on several live-virus vaccines progressed, and within a few years the Salk vaccine was dispensed by live-virus vaccines.

The long-range historical test of the Salk vaccine, in consequence, has never been carried out. We do not know with certainty whether or not that vaccine could have accomplished the relatively complete elimination of polio that has now been achieved. Nonetheless, this does not diminish the importance of its role is providing the first heartening success in the attack on this disease, a role to which careful and statistically informed experimental design contributed greatly.

PROBLEMS

1. Using Figure 1 as an example, explain why a control group is needed in experiments where the effectiveness of a drug or vaccine is to be determined.

2. Explain the need for control groups by criticizing the following statement: "A study on the benefits of vitamin C showed that 90% of the people suffering from a cold who take vitamin C get over their cold within a week."

3. Explain the difference between the observed control approach and the placebo control approach. Which one would you prefer, and why?

4. Why is it important to have a "double-blind" experiment?

5. If "double-blind" experiments provide the only satisfactory way to avoid observer bias, why aren't they used all the time?

6. If only volunteers are used in an experiment, instead of a random sample of individuals, will the results of the experiment be of any value? What can you say about the results?

7. Why did the polio epidemics seem to hit hardest those who were better off hygienically?
8. Why was a large-scale field trial needed to get convincing evidence of the Salk vaccine effectiveness?

9. Refer to Figure 1. In which year did the highest polio incidence occur? the lowest? the largest increase? the smallest increase? Give the approximate values of these incidences and increases.

10. Refer to Figure 1. Comment on the use of the number of cases. Can you suggest a different indicator of the spread of poliomyelitis in the U.S. during 1930-56. When are the two indicators equivalent? (Hint: refer to Table 1.)

REFERENCES


SAFETY OF ANESTHETICS

Lincoln E. Moses Stanford University
Frederick Mosteller Harvard University

In 1958, American hospitals began using a new anesthetic called halothane. It soon became widely accepted for its many desirable properties. Unlike some of the more commonly used anesthetics, it could not catch fire, so fire and explosive hazards did not have to be a concern during surgical operations. Patients found it less disagreeable and recovered from anesthesia more quickly and with less severe side effects. Extensive laboratory research and trials on animals and humans in surgery had encouraged belief in its safety. So there were good reasons for halothane to come rapidly into widespread use. By 1962, surgeons used halothane in half of their operations.

After a few years, however, halothane came under suspicion as accounts appeared in the medical literature of some strikingly unusual—but strikingly similar—deaths of patients who had recently had this anesthetic. A few patients recovering from surgery suddenly took a turn for the worse, ran fevers,