The Biggest Public Health Experiment Ever: The 1954 Field Trial of the Salk Poliomyelitis Vaccine

Paul Meier   *University of Chicago*

The largest and, until the 1980's, the most expensive medical experiment in history was carried out in 1954. Well over a million young children participated, and the immediate direct costs were over 5 million midcentury dollars. The experiment was carried out to assess the effectiveness, if any, of the Salk vaccine as a protection against paralysis or death from poliomyelitis. The study was elaborate in many respects, most prominently in the use of placebo controls (children who were inoculated with simple salt solution) assigned at random (that is, by a carefully applied chance process that gave each volunteer an equal probability of getting vaccine or salt solution) and subjected to a double-blind evaluation (that is, an arrangement under which neither the children nor the physicians who evaluated their subsequent state of health knew who had been given the vaccine and who the salt solution).

   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.

3

BACKGROUND

Polio was never a common disease, but it certainly was one of the most frighten-
ing 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 6%
of the deaths in the age group 5 to 9 in the early 1950s, 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 epi-
demic 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 com-
munity 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 determina-
tion to overcome his paralytic handicap and the commitment to the fight against
polio made by Basil O'Connor, his former law partner, enabled a great deal
of attention, effort, and money to be expended on the care 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 caused
by a virus and that three main virus types are involved. Although clinical mani-
festations of polio are rare, it was discovered that the virus itself was not rare,
but common, and that most adults 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 better off hygienically (that
is, 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 countries with the poorest hygiene. The ex-
planation is that because there was plenty of polio virus in the less-favored
populations, almost every infant was exposed to the disease early in life while
still protected by the immunity passed on from the mother. As a result, everyone
had had polio, but under protected circumstances, and, thereby, everyone had
developed 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* (a part of the gamma globulin fraction of the
blood) to the *antigen* (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 for the vaccine a closely related virus, cowpox, which is ordinarily incapable of causing serious disease in humans, but which gives rise to antibodies that also protect against smallpox. (In a very few individuals this vaccine is capable of causing a severe, and occasionally fatal, reaction. The risk is small enough, however, so that before smallpox was conquered we did not hesitate to expose all our schoolchildren 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 virus that has been killed by treatment 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 sufficiently like the live virus to be a satisfactory vaccine.

For 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 that would produce a high level of long-lasting antibodies. With such a vaccine, however, there might be a risk that a vaccine virus so similar to the virulent polio 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 researchers. In the 1930s, 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 had been responsible for cases of paralytic polio soon caused both 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 late 1940s by the development of a tissue culture technique for growing polio virus. Those working with live preparations developed harmless strains from virulent ones by growing them for many generations in suitable tissue culture media. There was, of course, considerable worry lest these strains, when used as a vaccine in humans, might revert to virulence and cause paralysis or death. (It's now clear that the strains developed are 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 infectiousness, but not so long as to destroy its antigenic effect. This was more difficult than expected, 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 story (Meier, 1957), but our concern here is with the evaluation of effectiveness.

EVALUATION OF EFFECTIVENESS

In the early 1950s the Advisory Committee convened by the National Foundation for Infantile Paralysis (NFIP) 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 50% 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 about the average incidence in the United States during the 1950s). 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 this magnitude could fairly easily be attributed to random variation. It would suggest that the vaccine might be effective, but it would not be persuasive. With 100,000 in each group, the expected numbers of polio cases would be 50 and 25, and such a result would be persuasive. In practice, a much larger study was clearly required because it was important to get definitive results as soon as possible, and if there were relatively few cases of polio in the test area, the expected number of cases might be well under 50. It seemed likely, also, for reasons we shall discuss later, that paralytic polio, rather than all polio, would be a better criterion of disease, and only about half the diagnosed cases are classified "paralytic." Thus the relatively low incidence of the disease, and its great variability from 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 use 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 polio 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 out an effective vaccine against the common cold, convincing evidence might be obtained in this way.
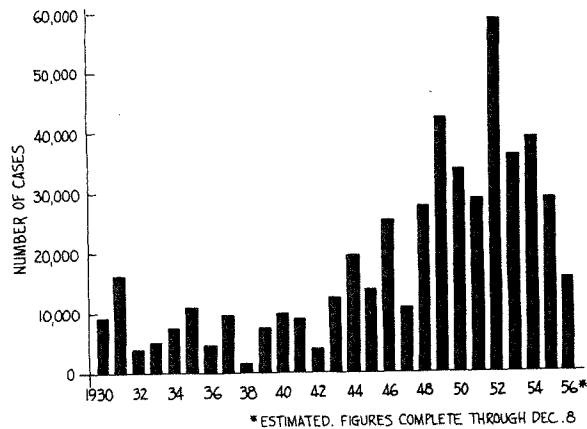
Figure 1 *Poliomyelitis in the U.S., 1930–56. Source: Rutstein (1957)*

In the case of polio—and, indeed, in most cases—so simple an approach would almost surely fail to produce clear-cut evidence. First, and foremost, we must consider how much polio incidence varies from season to season, even without any attempts to modify it. From Figure 1, which shows the annual reported incidence from 1930 through 1955, we see that had a trial been conducted in this way in 1931, the drop in incidence from 1931 to 1932 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—for example, 1952. (On the general problem of drawing inferences from such time series data see the essay by Campbell.) One might suppose that such mistakes 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 polio 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 polio, but the overwhelming majority of cases are less clear-cut. Fever and weakness are common symptoms of many illnesses, including polio, and the distinction between weakness and slight transistory paralysis will be made differently by different observers. Thus the decision to diagnose a case as nonparalytic polio instead of some other disease might well be influenced by a physician's general knowledge or feeling about how widespread polio is in his or her 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

most 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 polio. Later evaluation by specialists cannot, of course, bring into the picture those cases originally diagnosed as something other than polio.

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 polio 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 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 polio incidence does vary with age. Not much variation from grade to grade was expected, however, so it seemed reasonable to assume that the average of first and third grades would provide a good control for the second grade.

Despite the relative attractiveness of this plan and its acceptance by the NFIP advisory committee, serious objections were raised by certain health departments that were expected to participate. In their judgment, the results of such a study were likely to be insufficiently convincing for two important reasons. One is the uncertainty in the diagnostic process mentioned earlier and its liability to influence by the physician's expectations, and the other is the selective effect of using volunteers.

Under the proposed study design, physicians in the study areas would have been aware of the fact that only second-graders were offered vaccine, and in making a diagnosis for any such child, they would naturally and properly have inquired whether the child had been vaccinated. Any tendency to decide a difficult diagnosis in favor of nonpolio when the child was known to have been vaccinated would have resulted in a spurious piece of evidence favoring the vaccine. Whether or not such an effect was really operating would have been almost impossible to judge with assurance, and the results, if favorable, would have been forever clouded by uncertainty.

A less conjectural difficulty lies in the difference between those families who volunteer their children for participation in such a trial and those who do not. Not at all surprisingly, it was later found that those who do volunteer tend to be better educated and, generally, more well-to-do than those who do not participate. There was also evidence that those who agree to participate tend to

be absent from school with a noticeably higher frequency than others. The direction of effect of such selection on the incidence of diagnosed polio is by no means clear before the fact, and this important difference between the treated group and the control group also would have clouded the interpretation of the results.

RANDOMIZATION AND THE PLACEBO CONTROL APPROACH

The position of critics of the NFIP plan was that the issue of vaccine effectiveness was far too important to be studied in a manner that would leave uncertainties in the minds of reasonable observers. No doubt, if the vaccine should appear to have fairly high effectiveness, most public health officials and the general public would accept it, despite the reservations. If, however, the observed control scheme were used, a number of qualified public health scientists would have remained unconvinced, and the value of the vaccine would be uncertain. Therefore, the critics proposed that the study be run as a scientific experiment with the use of appropriate randomizing procedures to assign subjects to treatment or to control and with a maximum effort to eliminate observer bias. This plan follows what we call the *placebo control approach.*

The chief objection to this plan was that parents of schoolchildren 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 polio.

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 because the injections would not be given during the polio season, the risk associated with the placebo injection itself was vanishingly small. Certain health departments took a firm stand: they would participate in the trial only if it were such a well-designed experiment. The consequence was that in approximately half the areas, the randomized placebo control method was used, and in the remaining areas, the alternating-grade observed control method was used.

A major effort was put forth to eliminate any possibility of the placebo control results being contaminated by subtle observer biases. The only firm way to accomplish this was to ensure that neither the subject, nor the parents, nor the diagnostic personnel could know which children had gotten the vaccine until all diagnostic decisions had been made. The method for achieving this result was to prepare placebo material that looked just like the vaccine but was without any antigenic activity, so that the controls might be inoculated and otherwise treated in just the same fashion as were the vaccinated.

Each vial of injection fluid was identified only by a code number so that no one involved in the vaccination or the diagnostic evaluation process could know which children had gotten the vaccine. Because no one knew, no one

could be influenced to diagnose differently for vaccinated cases and for controls. An experiment in which both the subject getting the treatment and the diagnosticians who will evaluate the outcome are kept in ignorance of the treatment given each individual is called a *double-blind* experiment. Experience in clinical research has shown the double-blind experiment to be the only satisfactory way to avoid potentially serious observer bias when the final evaluation is in part a matter of judgment.

For most of us, it is something of a shock to be told that competent and dedicated physicians must be kept in ignorance lest their judgments be colored by knowledge of treatment status. We should keep in mind that it is not deliberate distortion of findings by the physician that concern the medical experimenter. It is rather the extreme difficulty in many cases of making an uncertain decision that, experience has shown, leads the best of investigators to be subtly influenced by information of this kind. For example, in the study of drugs used to relieve postoperative pain, it has been found that it is quite impossible to get an unbiased judgment of the quality of pain relief, even from highly qualified investigators, unless the judge is kept in ignorance of which patients were given the drugs.

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 treatment 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, 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 some 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. All cases of poliolike illness 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 and paralytic, with paralytic divided into nonfatal 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, of which 358 were confirmed as polio, and, among these, 270 were classified as paralytic (including 4 that were fatal). The third and fourth rows show corresponding entries for those who were vaccinated and those who received placebo, respectively. Beside each of these numbers is the corresponding rate. Using the simplest measure—all reported cases—the rate in the vaccinated group is seen to be half that in the control group (compare the boxed rates in Table 1) for the placebo control areas. This difference is greater than could reasonably be ascribed to chance, according to the appropriate probability calculation. The apparent effectiveness of the vaccine is more marked as we move from reported cases to paralytic cases to fatal cases, but the numbers are small and it would be unwise to make too much of the apparent very high effectiveness in protecting against fatal cases. The main point is that the vaccine was a success; it demonstrated sufficient effectiveness in preventing serious polio to warrant its introduction as a standard public health procedure.

Not surprisingly, the observed control area provided results that were, in general, consistent with those found in the placebo control area. The volunteer effect discussed earlier, however, is clearly evident (note that the rates for those not inoculated differ from the rates for controls in both areas). Were the observed control information alone available, considerable doubt would have remained about the proper interpretation of the results.

Although there had been wide differences of opinion about the necessity or desirability of the placebo control design before, there was great satisfaction with the method after the event. The difference between the two groups, although substantial and definite, was not so large as to preclude doubts had there been no placebo controls. Indeed, there were many surprises in the more detailed data. It was known, for example, that some lots of vaccine had greater antigenic power than did others, and it might be supposed that they should have shown a greater protective effect. This was not the case; lots judged inferior

Table 1 Summary of study cases by diagnostic class and vaccination status (rates per 100,000)

Study Group	Study Population	All Reported Cases		Poliomyelitis Cases									
				Total		Paralytic		Non-paralytic		Fatal polio		Not Polio	
		No.	Rate	No.	Rate	No.	Rate	No.	Rate	No.	Rate	No.	Rate
All areas: Total	1,829,916	1,013	55	863	47	685	37	178	10	15	1	150	8
Placebo control areas: Total	749,236	428	57	358	48	270	36	88	12	4	1	70	9
Vaccinated	200,745	82	41	57	28	33	16	24	12	—	—	25	12
Placebo	201,229	162	81	142	71	115	57	27	13	4	2	20	10
Not inoculated*	338,778	182	54	157	46	121	36	36	11	—	—	25	7
Incomplete vaccinations	8,484	2	24	2	24	1	12	1	12	—	—	—	—
Observed control areas: Total	1,080,680	585	54	505	47	415	38	90	8	11	1	80	7
Vaccinated	221,998	76	34	56	25	38	17	18	8	—	—	20	9
Controls†	725,173	439	61	391	54	330	46	61	8	11	2	48	6
Grade 2 not inoculated	123,605	66	53	54	44	43	35	11	9	—	—	12	10
Incomplete vaccinations	9,904	4	40	4	40	4	40	—	—	—	—	—	—

*Includes 8,577 children who received one or two injections of placebo.
†First- and third-grade total population.

Source: Adapted from T. Francis, Jr. (1955), Tables 2 and 3.

in antigenic potency did just as well as those judged superior. Another surprise was the rather high frequency with which apparently typical cases of paralytic polio were not confirmed by laboratory test. Nonetheless, there were no surprises of a character to cast serious doubt on the main conclusion. The favorable reaction of those most expert in research on polio was expressed soon after the results were reported. By carrying out this kind of study before introducing the vaccine, it was noted, we had facts about the Salk vaccine that we still lack about the typhoid vaccine and about the tuberculosis vaccine, after many decades of use.

EPILOGUE

It would be pleasant to report an unblemished record of success for the Salk vaccine following so expert and successful an appraisal of its effectiveness, 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 and it 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 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 prompt disappearance of polio 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 displaced in the United States, but not in Sweden, 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 in 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."