1

Controlled Experiments

Always do right. This will gratify some people, and astonish the rest.
—MARK TWAIN (UNITED STATES, 1835–1910)

1. THE SALK VACCINE FIELD TRIAL

A new drug is introduced. How should an experiment be designed to test its effectiveness? The basic method is *comparison*.[1] The drug is given to subjects in a *treatment group*, but other subjects are used as *controls*—they aren't treated. Then the responses of the two groups are compared. Subjects should be assigned to treatment or control *at random*, and the experiment should be run *double-blind*: neither the subjects nor the doctors who measure the responses should know who was in the treatment group and who was in the control group. These ideas will be developed in the context of an actual field trial.[2]

The first polio epidemic hit the United States in 1916, and during the next forty years polio claimed many hundreds of thousands of victims, especially children. By the 1950s, several vaccines against this disease had been discovered. The one developed by Jonas Salk seemed the most promising. In laboratory trials, it had proved safe and had caused the production of antibodies against polio. By 1954, the Public Health Service and the National Foundation for Infantile Paralysis (NFIP) were ready to try the vaccine in the real world—outside the laboratory.

Suppose the NFIP had just given the vaccine to large numbers of children. If the incidence of polio in 1954 dropped sharply from 1953, that would seem to

prove the effectiveness of the vaccine. However, polio was an epidemic disease whose incidence varied from year to year. In 1952, there were about 60,000 cases; in 1953, there were only half as many. Low incidence in 1954 could have meant that the vaccine was effective—or that 1954 was not an epidemic year.

The only way to find out whether the vaccine worked was to deliberately leave some children unvaccinated, and use them as controls. This raises a troublesome question of medical ethics, because withholding treatment seems cruel. However, even after extensive laboratory testing, it is often unclear whether the benefits of a new drug outweigh the risks.[3] Only a well-controlled experiment can settle this question.

In fact, the NFIP ran a controlled experiment to show the vaccine was effective. The subjects were children in the age groups most vulnerable to polio—grades 1, 2, and 3. The field trial was carried out in selected school districts throughout the country, where the risk of polio was high. Two million children were involved, and half a million were vaccinated. A million were deliberately left unvaccinated, as controls; half a million refused vaccination.

This illustrates the method of comparison. Only the subjects in the treatment group were vaccinated: the controls did not get the vaccine. The responses of the two groups could then be compared to see if the treatment made any difference. In the Salk vaccine field trial, the treatment and control groups were of different sizes, but that did not matter. The investigators compared the rates at which children got polio in the two groups—cases per hundred thousand. Looking at rates instead of absolute numbers adjusts for the difference in the sizes of the groups.

Children could be vaccinated only with their parents' permission. So one possible design—which also seems to solve the ethical problem—was this. The children whose parents consented would go into the treatment group and get the vaccine; the other children would be the controls. However, it was known that higher-income parents would more likely consent to treatment than lower-income parents. This design is biased against the vaccine, because children of higher-income parents are more vulnerable to polio.

That may seem paradoxical at first, because most diseases fall more heavily on the poor. But polio is a disease of hygiene. A child who lives in less hygienic surroundings is more likely to contract a mild case of polio early in childhood, while still protected by antibodies from its mother. After being infected, these children generate their own antibodies, which protect them against more severe infection later. Children who live in more hygienic surroundings do not develop such antibodies.

Comparing volunteers to non-volunteers biases the experiment. The statistical lesson: the treatment and control groups should be as similar as possible, except for the treatment. Then, any difference in response between the two groups is due to the treatment rather than something else. If the two groups differ with respect to some factor other than the treatment, the effect of this other factor might be *confounded* (mixed up) with the effect of treatment. Separating these effects can be difficult, and confounding is a major source of bias.

For the Salk vaccine field trial, several designs were proposed. The NFIP had originally wanted to vaccinate all grade 2 children whose parents would consent,

leaving the children in grades 1 and 3 as controls. And this design was used in many school districts. However, polio is a contagious disease, spreading through contact. So the incidence could have been higher in grade 2 than in grades 1 or 3. This would have biased the study against the vaccine. Or the incidence could have been lower in grade 2, biasing the study in favor of the vaccine. Moreover, children in the treatment group, where parental consent was needed, were likely to have different family backgrounds from those in the control group, where parental consent was not required. With the NFIP design, the treatment group would include too many children from higher-income families. The treatment group would be more vulnerable to polio than the control group. Here was a definite bias against the vaccine.

Many public health experts saw these flaws in the NFIP design, and suggested a different design. The control group had to be chosen from the same population as the treatment group—children whose parents consented to vaccination. Otherwise, the effect of family background would be confounded with the effect of the vaccine. The next problem was assigning the children to treatment or control. Human judgment seems necessary, to make the control group like the treatment group on the relevant variables—family income as well as the children's general health, personality, and social habits.

Experience shows, however, that human judgment often results in substantial bias: it is better to rely on impersonal chance. The Salk vaccine field trial used a chance procedure that was equivalent to tossing a coin for each child, with a 50–50 chance of assignment to the treatment group or the control group. Such a procedure is objective and impartial. The laws of chance guarantee that with enough subjects, the treatment group and the control group will resemble each other very closely with respect to all the important variables, whether or not these have been identified. When an impartial chance procedure is used to assign the subjects to treatment or control, the experiment is said to be *randomized controlled.*[4]

Another basic precaution was the use of a *placebo*: children in the control group were given an injection of salt dissolved in water. During the experiment the subjects did not know whether they were in treatment or in control, so their response was to the vaccine, not the idea of treatment. It may seem unlikely that subjects could be protected from polio just by the strength of an idea. However, hospital patients suffering from severe post-operative pain have been given a "pain killer" which was made of a completely neutral substance: about one-third of the patients experienced prompt relief.[5]

Still another precaution: diagnosticians had to decide whether the children contracted polio during the experiment. Many forms of polio are hard to diagnose, and in borderline cases the diagnosticians could have been affected by knowing whether the child was vaccinated. So the doctors were not told which group the child belonged to. This was *double blinding*: the subjects did not know whether they got the treatment or the placebo, and neither did those who evaluated the responses. This randomized controlled double-blind experiment—which is about the best design there is—was done in many school districts.

How did it all turn out? Table 1 shows the rate of polio cases (per hundred thousand subjects) in the randomized controlled experiment, for the treatment

group and the control group. The rate is much lower for the treatment group, decisive proof of the effectiveness of the Salk vaccine.

Table 1. The results of the Salk vaccine trial of 1954. Size of groups and rate of polio cases per 100,000 in each group. The numbers are rounded.

The randomized controlled double-blind experiment			*The NFIP study*		
	Size	*Rate*		*Size*	*Rate*
Treatment	200,000	28	Grade 2 (vaccine)	225,000	25
Control	200,000	71	Grades 1 and 3 (control)	725,000	54
No consent	350,000	46	Grade 2 (no consent)	125,000	44

Source: Thomas Francis, Jr., "An evaluation of the 1954 poliomyelitis vaccine trials—summary report," *American Journal of Public Health* vol. 45 (1955) pp. 1–63.

Table 1 also shows how the NFIP study was biased against the vaccine. In the randomized controlled experiment, the vaccine cut the polio rate from 71 to 28 per hundred thousand. The reduction in the NFIP study, from 54 to 25 per hundred thousand, is quite a bit less. The main source of the bias was confounding. The NFIP treatment group included only children whose parents consented to vaccination. However, the control group also included children whose parents would not have consented. The control group was not comparable to the treatment group.

The randomized controlled double-blind design reduces bias to a minimum—the main reason for using it whenever possible. But this design also has an important technical advantage. To see why, let us play devil's advocate and assume that the Salk vaccine had no effect. Then the difference between the polio rates for the treatment and control groups is just due to chance. How likely is that?

With the NFIP design, the results are affected by many factors that seem random: which families volunteer, which children are in grade 2, and so on. However, the investigators do not have enough information to figure the chances for the outcomes. They cannot figure the odds against a big difference in polio rates being due to accidental factors. With a randomized controlled experiment, on the other hand, chance enters in a planned and simple way—when the assignment is made to treatment or control.

The devil's-advocate hypothesis says that the vaccine has no effect. On this hypothesis, a few children are fated to contract polio. Assignment to treatment or control has nothing to do with it. Each child has a 50–50 chance to be in treatment or control, just depending on the toss of a coin. Each polio case has a 50–50 chance to turn up in the treatment group or the control group.

Therefore, the number of polio cases in the two groups must be about the same. Any difference is due to the chance variability in coin tossing. Statisticians understand this kind of variability. They can figure the odds against a difference as large as the observed one. The calculation will be done in chapter 27, and the odds are astronomical—a billion to one against.