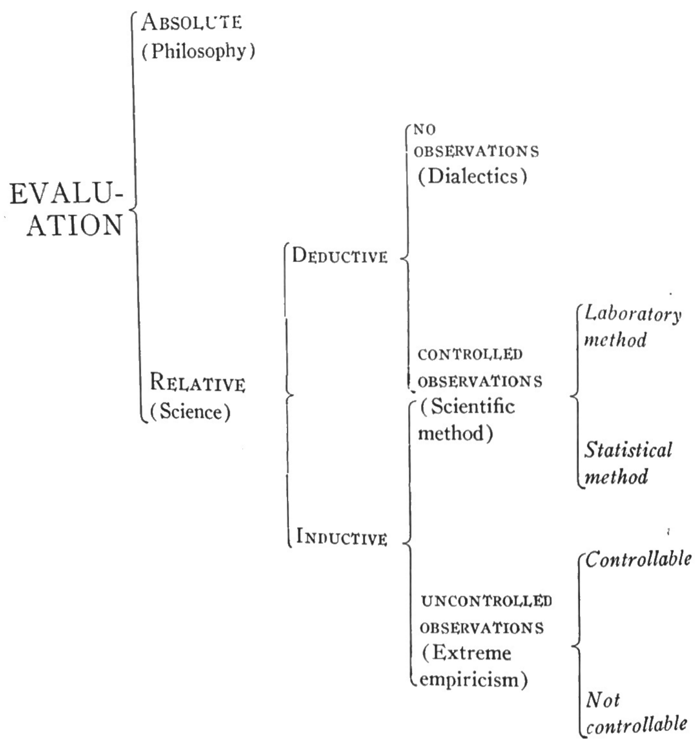
EVALUATION OF PUBLIC HEALTH PROCEDURES[1]

ELBRIDGE SIBLEY

STATISTICIAN, TENNESSEE DEPARTMENT OF PUBLIC HEALTH

INTRODUCTION

The title of this paper may possibly seem to imply an attempt to say that public health work is a valuable thing, in the sense in which we say that human life is valuable. No such interpretation is intended. *Evaluation* is a dangerous word in the hands of a scientist, and it is necessary, before proceeding to use it, to make clear what kind of evaluation is meant. I mean, specifically, determination of the relative values of various expedients as means for reducing the amount of sickness in a population.



[1]Read before the Tennessee Academy of Science at the Nashville meeting, November 27, 1931.

The accompanying outline is not offered as a new solution to all problems of logic and epistemology; its purpose is simply to indicate in rather crude language a possible classification of ways of thinking about a certain class of phenomena, namely the health of large numbers of people and things done to improve it.

As the diagram indicates, we can obviously seek to evaluate a thing either in absolute terms or in relative terms. If we attempt the former, we become philosophers. Science, as righteously outraged metaphysicians often cry out, is incapable of determining the ultimate goodness or badness of anything. If we remain scientists, as we are expected to do while in this Academy, we are confined to setting a relative value on one thing in terms of some other thing. Concretely, the present paper deals with methods of evaluating certain public health activities in terms of illness and death. That illness and death are evil and their absence good, is a proposition for others to prove.

There are various processes of reasoning by which conclusions as to the relative values of such things may be reached, with varying degrees of certainty. We may conclude inductively from the observation of actual cases, that a certain public health measure is or is not effective in limiting the amount of sickness and death. An example of extreme uncontrolled empiricism consists in observing that there are fewer and fewer deaths from year to year from a certain cause and concluding without further investigation that certain activities of the public health department have therefore been effective. Stated in such a bald fashion, the unreliability of such reasoning is obvious; yet there are actual problems in which the number of interacting factors is so great that for the present at least we have to be content with it, *faute de mieux*.

At the other extreme of the diagram is the item "Deductive evaluation with no observations." Such dialectics do not strictly belong in a classification of scientific methods, but have been included because they are sometimes passed off in the guise of scientific reasoning. As H. G. Wells has observed,

> . . . When we have a name we are predisposed—and sometimes it is a very vicious predisposition—to imagine forthwith something answering to the name . . . If I say *wodget* or *crump*, you find yourself passing over the fact that these are nothings, . . . and trying to think what sort of a thing a *wodget* or a *crump* may be. You find yourself insensibly, by subtle associations of sound and ideas, giving these blank terms attributes.

To sum up, then, a sound appraisal of the value of public health procedure, whether the procedure be very simple or very complex, must be neither purely rhetorical nor purely empirical (if I may put these two terms in opposition). Technicians, and those who deal in "applied" science of any kind, rebel almost automatically at the mention of dialectics. In its extreme form, especially in the "inexact" sciences, this result sometimes leads to the habit of thinking that a sufficient quantity of data will solve any problem.

But enough of this discussion of what we are not going to do. As everyone knows, the advancement and application of science involve both inductive and deductive reasoning, and verification of hypotheses by actual observations. Observations, to be scientific, must be controlled, and there are two methods of controlling them which may be conveniently but not accurately called the laboratory method and the statistical method. The former involves artificially limiting the number of elements in a problem; the latter, measuring the behavior of certain elements which are more or less entangled with factors irrelevant to the problem at hand. It is the latter method to which we must most often resort in attempts to measure social phenomena of which public health is one aspect.

There are two times when it is important for a public health worker to evaluate his procedures: one is before a procedure has been carried out, and the other after it has been carried out. The latter case, being more transparent, I shall pass over very briefly; I shall give more time to describing a particular instance of the former, which I hope may be of interest.

Evaluation ex Post Facto

A fallacy to be avoided in measuring the results of causes which have already operated is that of unmitigated empiricism, or perhaps I should say wishful empiricism. When we have spent a good deal of time and money on efforts intended to produce a certain result, and when we see that the hoped for state of affairs has come to pass, there is sometimes an overwhelming temptation to "prove" that our efforts have been fruitful by simply setting down in parallel columns the efforts made and the subsequent changes in conditions toward which the efforts were directed. Such a proof, of which innumerable published examples might be cited, is either wholly convincing or not at all convincing, depending upon whether or not we very much wish to believe it and allow our logic to be dominated by our wishes.

It would, I fear, be gratuitous to try to rehearse here a large number of the methods which it is necessary to follow in measuring objectively the results of social forces or activities as in public health work. The basic principles are those of science in general, but their corollaries are different in almost every actual case. If we are concerned only with demonstrating to the public that its money may well be spent on measures against typhoid fever, comparatively crude data may be used; on the other hand, if we want to determine which of several methods is most economically effective in controlling that disease, our observations must be more rigidly controlled, and our methods of inference more complicated.

In the attempt of public health workers to substitute the accuracy and effectiveness of a high-powered rifle for the more diffuse but less penetrating bombardment of a shotgun, there has come an in-

creased emphasis on carefully controlled projects whose primary objective is not the immediate improvement of conditions in a particular place but the accumulation of knowledge about the effectiveness of specific procedures. It is not always humanly—or humanely—possible to carry out such projects in a strictly scientific manner. Dr. Martin Arrowsmith, in Sinclair Lewis' story, went to a plague-stricken tropical island and was faced by the dilemma of advancing scientific knowledge by controlled experiments, or doing all he could to relieve the actual suffering which he saw. It is hard to blame him for having reached toward the latter horn and lost his grip on the former. Fortunately, however, it is frequently possible to collect

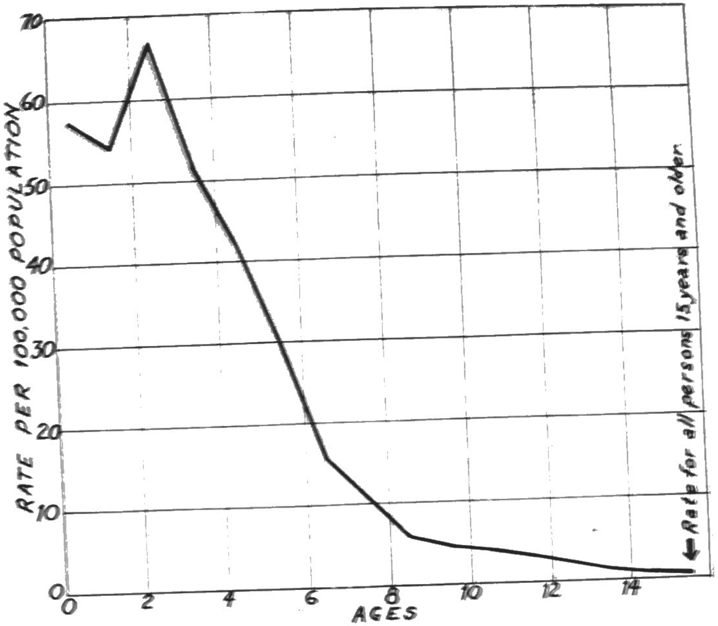


Fig. 1. The Annual Diphtheria Death Rate by Ages for the Rural Part of Tennessee (1923-1927).

valid controlled data without hardening one's heart toward the victims of some catastrophe. The essential requirement is that the procedure in question be applied either in a place where no other variable forces are impinging on the situation, or else in a place where the effects of such variables can be measured or estimated with sufficient accuracy to make possible their "elimination" from final conclusions. This is the logical procedure followed when the value of a special kind of nursing program, for example, is appraised in terms of the experience of its clientele contrasted with the experience of members of the same community who have not received such service. Statistical criteria based on the laws of probability have to be applied in order to draw usable conclusions from experiments, such as that just mentioned, in which only one variable is artificially controlled, but others are known to vary.

Evaluation A Priori

Statistical operations are sometimes interesting but seldom useful unless they yield generalizations which may be expected to hold good in the future under similar circumstances. The statistical appraisal of a nursing program which has already been carried out is useful because it helps in deciding whether that program should continue, and whether similar ones are worth undertaking. Obviously, such an appraisal would be far more useful if it could be made before instead of after carrying out a project. If this were always possible,

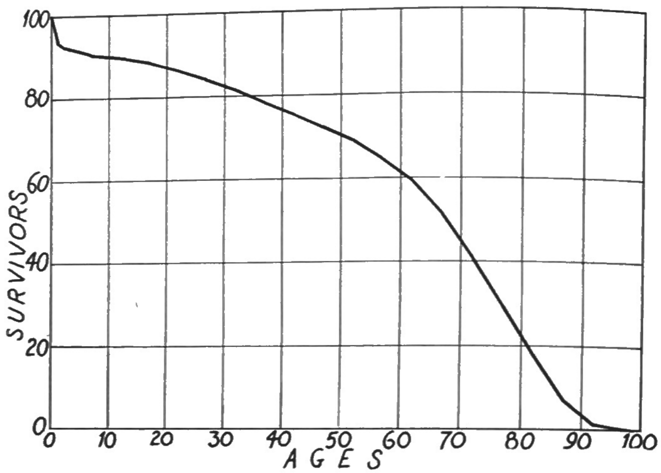


Fig. 2. The Survivors at Given Ages out of One Hundred Born Alive in the White Population of Tennessee (1919-1920). Data from the United States Abridged Life Tables, 1919-1920.

a great deal of time and money which is now spent on experimental work, would be saved for other purposes. Sometimes it is possible. I should like to give a concrete illustration of how empirical data can be manipulated and made to give us an *a priori* estimate of the results of a certain project of public health work.

Diphtheria is a disease for which medical scientists have found specific preventive medicines, namely toxin-antitoxin mixture and toxoid. Biologically, these are different products, but the argument which is to follow is not immediately concerned with these differences. Most individuals can be prevented from having diphtheria by inoculation with a series of doses of toxin-antitoxin. Hence, if we could inoculate everyone, the disease would be practically abolished. But, our entire budget is not big enough to pay the cost, and there

are other diseases to control on which we must spend a part of our money. Therefore, the immediate question is, What is the most economical way, within the limits of our resources, to reduce the prevalence of diphtheria?

There have been divergences in practice even more than in theory with respect to the age at which large numbers of children should be immunized by public health agencies. As to the individual child, there is no difficulty. The earlier in life immunization is performed, the greater the saving of risk; further, there is general agreement on the part of clinicians that there is no objection to administering the doses as early as the sixth month of age. All of this would be admitted by any health officer. But, he might add, it is very much more expensive to canvass a hundred homes in order to immunize a hun-

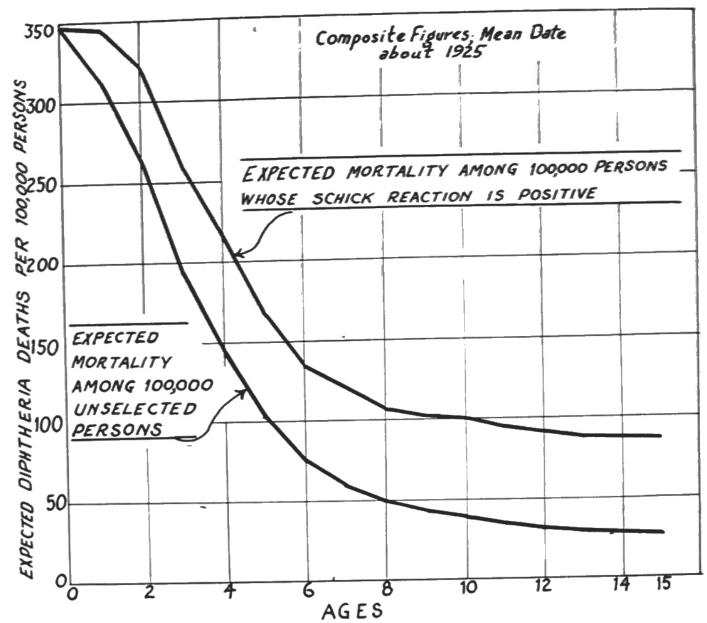


Fig. 3. The Numbers Out of 100,000 Survivors at Given Ages Expected to Die Eventually from Diphtheria in Rural Tennessee.

dred babies, than it is to visit a single schoolhouse and immunize the same number of children; our time and money are limited, so we shall devote our efforts to immunizing the largest possible number of individuals.

The empirical way to solve this problem would of course be to spend equal sums of time and money on the immunization of several separate groups of children of different ages, and watch the results.

The procedure would be expensive and slow. Some *a priori* reasoning will give us good information without the expense and delay.

The necessary data are (1) the incidence of diphtheria by ages, (2) the percentage of persons surviving from age to age (*i. e.*, expectation of life), (3) the efficiency of the immunizing agent—in this particular case, toxin-antitoxin mixture. The first two items are shown in figures 1 and 2.

Our problem is to combine these data into an expression for the life- or health-saving value of inoculation of a person at a given age. First the proposition must be stated in the negative: Out of 100,000 children who have survived to age x, how many can be expected (under existing conditions) ultimately to die from diphtheria?

The answer involves some mathematics which it will not be profitable to reproduce here; the lower curve in figure 4 shows the results. Under prevailing conditions, we can expect that out of 100,000 new-born infants about 350 will ultimately die of diphtheria. Out of 100,000 children who have reached their fifth birthday, 100 will die of diphtheria after various intervals of time. When a child reaches his ninth birthday, the chances that he will die of diphtheria at some later time are only 50 out of 100,000. Thus if we artificially immunize Billy Jones on his fifth birthday so that he can never contract diphtheria, the chances are 100 out of 100,000 or 1/1,000th that we shall thereby save his life. Put in other words, if we immunize Billy Jones and 999 of his playmates, all five years old, we shall probably have forestalled one death from diphtheria.

In practice it has been found, however, that a series of three doses of toxin-antitoxin will produce immunity to diphtheria in about 70 out of 100 persons. (This is determined by the Schick test, which consists in observing the reaction to a minute quantity of diphtheria toxin injected into the skin). Therefore we must revise our prediction about Billy Jones and say, for practical purposes that if we inoculate Billy and 1,428 others of the same age, we shall probably produce immunity in 1,000 of them, and shall thus forestall one death from diphtheria.

The foregoing can be summarized thus (referring to the lower curve in figure 3): Immunization of an infant under a year of age, is roughly twice as valuable as immunization of a three-year-old, four times as valuable as immunization of a five-year-old, and seven times as valuable as immunization of an eight-year-old. Combining these figures with data on the cost of traveling about and reaching a given number of children of different ages, it will be possible to determine the most productive way to spend a given sum on diphtheria prevention.

So far, it has been assumed in the discussion that toxin-antitoxin would be administered to groups of children irrespective of whether or not they were already naturally immune to diphtheria. Another plan is, first to test the immunity of an individual by the Schick method, and then to inoculate him only if he is found susceptible or

"Schick-positive." Practice varies in this respect; it is desirable to find out objectively the relative economies of the two methods.

By Schick-testing a large number of children who have not been artificially immunized, it is possible to derive a curve representing the development of natural immunity to diphtheria with increasing age. Such a curve based on observations in a rural county in Middle Tennessee is shown in figure 4. It is assumed that for practical purposes natural immunity once acquired endures permanently.

Now, if Schick testing is done before giving inoculations, only a certain portion of the population at a given age will need to be inoculated. At less than one year of age this portion will be nearly 100 per cent; at three years, 80 per cent; at six years, 60 per cent, at ten years, 40 per cent. Since indications are that a person once

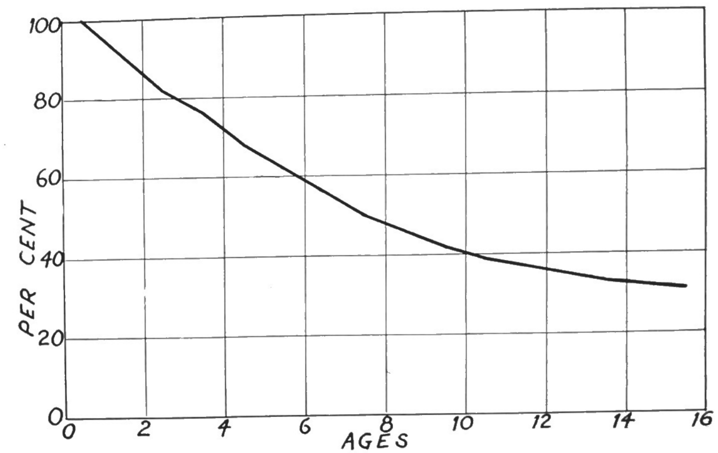


Fig. 4. The Per Cent of Children That Are Schick-Positive, by Ages in a Rural County in Tennessee (1924-1928).

Schick-negative will never become positive, it is clear that just as much illness and death can be prevented by immunizing only those members of the population who give a positive Schick reaction, as can be prevented by inoculating the entire population. Concretely, given a population of 1,000 children, only 50 per cent of whom are susceptible to diphtheria: we can completely eliminate diphtheria either by inoculating the entire 1,000, or equally well by immunizing only the 500 who are actually susceptible. Similarly, half of the deaths could theoretically be prevented either by inoculating 500 unselected children, or by immunizing 250 children taken from the susceptible half of the population. In other words, twice as much theoretical good would be done by inoculating a known susceptible child as by inoculating an unselected child.

Applying the percentages of non-immune children at different ages as divisors to the figures of expected mortality from diphtheria, we may obtain another curve (the upper curve in figure 4), which shows the expected number of deaths from diphtheria which will ultimately occur among 100,000 children who have survived to age x without yet having acquired immunity to diphtheria. In negative terms, the curve shows the potential saving of life to be effected by a given number of immunizations performed on Schick-positive children at given ages.

Comparing the two curves in figure 4, it is clear that the differential between the values of inoculation to an individual selected at random (lower curve) and to an individual selected by the Schick test increases with increasing age. For administrative purposes, the saving of expense can be converted into terms of dollars and cents saved, and then compared with the additional cost of the preliminary Schick test. When the latter data are compiled, it may perhaps be demonstrated that it is unprofitable to give the preliminary test to infants and pre-school children, but that a saving of resources for other work could be effected by testing school children before inoculation. These last statements, it must be reiterated, are not conclusions but guesses, as the last link in the chain of evidence has not yet been formed.

An Actual Example

It may be of interest to comment upon an application of the first part of the scheme of evaluation which has just been described, to some work which was actually done in Tennessee. About 20,000 children in a certain group of counties were inoculated with toxin-antitoxin within a period of two years. Comparison of the diphtheria mortality for the following years in these counties with the death rate in counties where no wholesale immunization had been done, did not show the results which an uninformed optimist might have dreamed. Actually, the diphtheria rate rose in the counties where immunization had been carried on, and fell in the other group. Should it then be concluded that the effort to immunize 20,000 children was wasted?

Applying the age-specific annual death rates from diphtheria in rural Tennessee (Fig. 1) to the numbers of children at different ages among the 20,000 inoculated, it is found that 4 or 5 deaths would be expected to occur during one year among this group of 20,000 children, and that ultimately about 20 of them would die from diphtheria. The number of deaths thus prevented is not large enough to have a perceptible influence upon the diphtheria death rate of the counties involved, particularly since the incidence of diphtheria varies greatly from year to year from a number of causes which are not well understood. Hence it would be equally fallacious to say that the work was useless because it was followed by an increased death rate from diphtheria, or to say, if the death rate had

happened to fall subsequent to the immunization work, that the latter was the cause of the decline.

Another observation may be added. A majority of the group of 20,000 were school children; if it had been feasible to select a younger group, twice as many deaths might have been prevented by the same number of inoculations.

Conclusion

The example of diphtheria immunization has been chosen for presentation, not because it is necessarily the leading problem in public health work, but because it offers a clear-cut example of the application of deductive reasoning where a purely empirical approach is impractical. We have a lot of other problems, some of them much bigger, in which we are more or less groping about for methods which will enable us to evaluate our preventive measures.