How much of each lot of vaccine must be tested to provide reasonable assurance of its safety? This paper considers some of the statistical issues associated with this question and, in the process, develops the concept of the consistency of a production process.

Some Statistical Aspects of Safety Testing the Salk Poliomyelitis Vaccine

By JEROME CORNFIELD, MAX HALPERIN, Ph.D., and FELIX MOORE

MANY circumstances can influence the safety of any lot of vaccine. They may be conveniently considered as falling into two distinct classes. The first consists of all those circumstances affecting the ability to produce a safe vaccine, the second those affecting the ability to detect an unsafe vaccine. Problems involving the second class of circumstances we shall refer to as problems of safety testing. In practice the safety of the vaccines released for general use will depend on the successful solution of both sets of problems, and it would be hazardous to place reliance for safety exclusively on either one of the two. Nevertheless, in formulating criteria for safety testing, it is useful to inquire into the amount of testing required to assure a high level of safety without making any assumptions about the safety of the production process. This is equivalent to asking whether it is possible to assure a high level of safety by testing alone, even under the most unfavorable production circumstances that one can envisage. While we shall not be able to answer this question definitively, it will be because of the lack of key biological information and not because the problem is analytically insoluble or even necessarily that the solution, given the key information, would require impracticably large amounts of testing.

The authors are all with the National Institutes of Health, Public Health Service. Mr. Cornfield is assistant chief of the Biometrics Branch of the Division of Research Services. Dr. Halperin is chief of the Biometrics Office, Division of Biologics Standards, and Mr. Moore is chief of the Biometrics Research Section of the National Heart Institute. *They presented this paper at the 115th Annual Meeting of the American Statistical Association, held in New York City December 28, 1955.*

Vaccine Preparation

We begin by reviewing briefly those aspects of vaccine preparation and the minimum requirements for safety testing (*1*) which are pertinent to the subsequent discussion. There are three immunologically distinct forms of poliomyelitis virus. An attack by type 1 virus will confer immunity against further infection by that type but not necessarily against infection by type 2 or type 3. Since the vaccine must provide protection against all three types, it must contain antigens for each. A vaccine prepared from a single virus type is referred to as a single-strain vaccine, while a trivalent

vaccine, the form in which the vaccine is actually used, is a combination of equal amounts of three single-strain vaccines.

Each batch of single-strain vaccine is prepared from a virus pool obtained by propagation on cultures of monkey kidney tissue. The pool is filtered and then tested for infectivity. If sufficiently infective for tissue culture, it is ready for the next step, the preparation of vaccine. A pool is considered sufficiently infective if 0.5 cc. is capable of infecting tissue culture after at least a one-millionfold dilution. The amount by which a preparation must be diluted before it loses infectivity is referred to as its titer. There is, in fact, no single dilution point at which infectivity turns abruptly to noninfectivity, and in practice the titer used is that dilution estimated to result in infectivity for 50 percent of the inoculated tissue culture tubes. The amount of virus present in an inoculum capable of infecting 50 percent of the tubes is referred to as one tissue culture infectious dose ($TCID_{50}$).

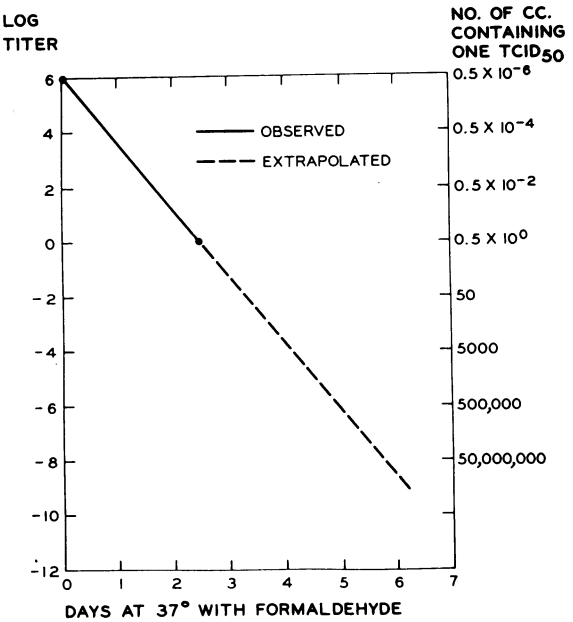
In practice it is more convenient to work with log titers than with titers. We thus say that a virus pool is ready for the next step when its log titer is at least 6. In the next step the pool is exposed to formaldehyde at a temperature of 37° C. and heated for 6 days or more. The preparation loses infectivity continuously but still retains the ability to stimulate antibody production. At intervals during this inactivation process samples are taken and the titer of 0.5 cc. determined. At some point during the process, usually 2 to 3 days, the titer has dropped from at least 1 million to unity, that is, after 2 to 3 days the 0.5 cc. of the preparation, if diluted at all, will infect less than half the tubes into which it is inoculated. Unless one uses larger inoculums or concentrates the preparation, the log titer cannot be easily determined for any period after this time. The curve relating log titer to time heated is referred to as an inactivation curve.

Subsequent to theoretically complete inactivation, tissue culture safety tests are performed. The current tissue culture safety test requires two independent tests of 500 cc. each for each single-strain vaccine, the first test 6 to 9 days after the initiation of inactivation, the second 3 days after the first. In addition, 1,500 cc. of each trivalent vaccine must be tested. The test batch is passed if it produces no tissue changes indicative of the presence of live virus and in addition passes a monkey safety test. This latter test requires that each filling of the final trivalent lot must be tested on at least 5 monkeys, a minimum of 20 being used for each lot. Each monkey receives 2.5 cc. of vaccine. The lot is passed if histological and other studies on the test monkeys "leave no doubt that poliomyelitis infection did not occur"(*1*).

Shape of Inactivation Curve

To point up the difficulties that can arise when exclusive reliance is placed upon the safety of the production process rather than on the adequacy of the safety test, we consider Salk's original concept of factors affecting safety. As elaborated in several publications, the chief guarantee of the safety of the final vaccine was not felt to be the result of a monkey or tissue culture safety test but rather the nature of the inactivation process itself. Thus, it was observed that if log titer was plotted against time of exposure over the observable period, that is, the first 2 or 3 days, that log titer was a linear function of time. Figure 1, which has been adapted from one of the discussions of

Figure 1. Theoretical relationship between log titer and inactivation time.

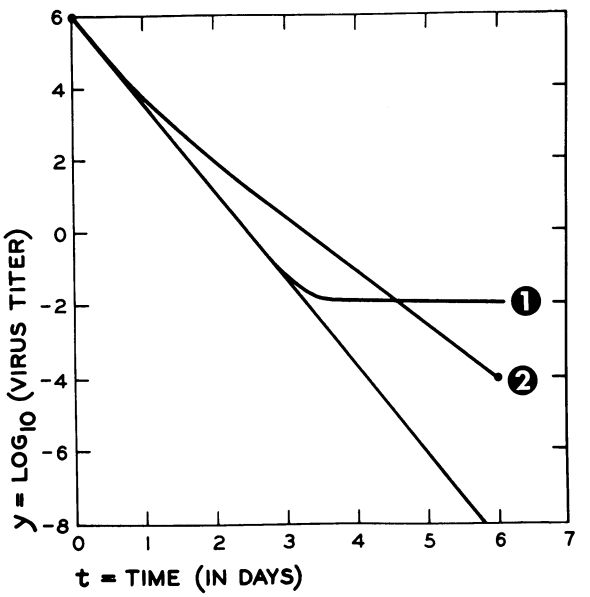


this point subsequently published by Salk (*2*), illustrates this point with a theoretical inactivation curve in which log titer is assumed to drop from an initial value of 6 to 0 in 2½ days. Salk has written (*2*) that the linear nature of this relation "makes possible the prediction, rather precisely, of the time required to render each preparation free of living virus." Thus, in the initial virus pool one tissue culture infectious dose could be found in as small amount as one two-millionths of a cubic centimeter. By 2½ days one would find one $TCID_{50}$ only in every half cubic centimeter, and, by simple extrapolation, by 6 days only one in every 125 million cc. Again quoting from one of Salk's publications (*3*) "[if] the reaction is allowed to proceed for a total period equal to three times the interval required for interception of the base line . . . the margin of safety which guarantees absolute safety has been assured."

The assumption that log titer was a linear function of inactivation time was not an entirely empirical one, simply suggested by inspection of data, but a relationship often found in theoretical chemistry. If the inactivation process is thought of as a chemical reaction analagous to a situation wherein one molecule of virus combines with one molecule of formaldehyde, the latter being present to excess and consequently not limiting the speed of the reaction, then the relationship between log titer and time would indeed be linear, if the system was homogeneous (*4*). Whatever the merits of this formulation from the point of view of inactivation kinetics under controlled laboratory conditions, the occurrence of lots of vaccines containing live virus, even after 9 days of inactivation (*5*), indicated that extrapolation of the inactivation curve was no substitute for safety testing.

It is not hard to see why the extension of results possibly applicable in a homogeneous system to a potentially heterogeneous system might cause difficulty. Thus, figure 2 compares the inactivation curves obtained for two different hypothetical heterogeneous systems with the linear inactivation curve of the preceding figure. In the first curve we have assumed a heterogeneous system with two groups of virus particles, each being inactivated at different rates. One group is assumed to have an initial titer of 10^{6} and a loss of log titer of 2.4 per

Figure 2. Theoretical virus inactivation curves for three different models of the inactivation process.



day, the other an initial titer of 10^{-2}, but it is assumed to have no loss in activity as the reaction proceeds. Such a situation could occur if the bulk of the virus particles were being inactivated as in a monomolecular reaction, but a small fraction, 1/1,000,000, were protected from the action of the formaldehyde by tissue particles. It will be noted that in such a situation the observed inactivation curve would be virtually indistinguishable on any basis from a linear inactivation curve for the first 2½ days, but that thereafter it would level out quite rapidly and, no matter how protracted the time of inactivation, would remain at a level of one $TCID_{50}$ for every 50 cc., a highly infectious level. The shape of the inactivation curve for the observable period would in such a situation provide no guide to the subsequent course of the reaction.

In the second curve we have also assumed two groups of particles being inactivated at different rates. The first group is assumed to have an initial titer of $10^{5.97}$ and to undergo a loss in log titer of 2.4 per day. The second group is assumed to have an initial titer of $10^{4.81}$ and to undergo a daily loss in log titer of 1.47. Some curvature in the observable period will be noted. After 2 days of inactivation this curve is 0.8 of a log titer above the first curve, but it does not

level off as rapidly, eventually crosses it, and at 6 days has one $TCID_{50}$ in every 5,000 cc. Looking only at the observable period, one might place confidence in the eventual noninfectivity of the hypothetical vaccine producing curve 1 and might have serious doubts about the one producing curve 2. Nevertheless, after 6 days of treatment the second vaccine would have only 1/100 the concentration of infectious particles of the first.

These are, of course, only highly oversimplified models of what might happen. The real question is what does happen. To answer this we consider five successive lots of vaccine produced by a single manufacturer. These lots were composed of 104 independently produced and tested single-strain components. Of these, 12 had failed their initial tissue culture test after inactivation was presumably complete and the remainder had passed. We have taken all 12 of the positive components and a haphazardly selected sample of 17 of the 92 negatives. Least squares parabolas have been fitted to each of the 29 inactivation curves. The following tabulation shows for each of the curves the value of the quadratic component at 50 hours for positive and negative lots.

Negative lots		*Positive lots*	
1. 22	0. 66	1. 25	1. 86
1. 82	0. 10	0. 38	0. 22
1. 41	0. 59	0. 85	1. 16
0. 87	0. 63	−1. 59	1. 34
1. 84	0. 28	0. 94	0. 47
−1. 25	0. 57	0. 21	0. 85
0. 07	1. 16		
2. 00	0. 91	Average	0. 66
−0. 11			
Average	0. 75		

NOTE: The quadratic component is the value of ct^2, with $t=50$, where log titer$=a+bt+ct^2$ and t is time in hours.

It will be noted that the value of the quadratic component is less than zero in only three cases, two for negative lots and one for a positive lot. In all other cases the quadratic term is positive, indicating that the best fitting parabola curves up and away from the linear component and that a linear extrapolation will underestimate log titer. The average value of the component at 50 hours, 0.75, involves about the same departure from linearity at that point as does curve 2 of figure 2. The antilog of this value, 5.6, indicates that at 50 hours a difference of more than fivefold in estimated titer had already developed between the best fitting parabola and its linear component.

The fundamental point, however, is that no difference is apparent in the value of the quadratic component for positive and negative lots. Theoretical considerations and actual experience both lead to the same conclusion therefore—that the shape of the inactivation curve up to a certain point provides no necessary indication of its shape thereafter. One can also draw the more general conclusion that no matter how safe a production process is believed to be, common prudence requires safety testing procedures which have high probability of detecting the presence of live virus particles, if by some unforeseen chance the production process permits this to happen.

Size of Test Batch

In many problems of industrial sampling inspection, a decision as to how much to sample is reached by minimizing the monetary loss arising from a combination of testing cost and the costs arising from erroneously rejecting good lots or accepting bad lots of a product (with due regard for the a priori probability that a lot will be bad) (*6*). But in the present problem the loss arising from erroneously accepting an infectious lot is entirely incommensurable with the cost of testing or with the cost of erroneously rejecting good lots. This suggests that the methods of industrial sampling inspection cannot be applied to the present problem without some modification if they can be applied at all.

We may approach a solution by considering first of all an idealized suspension of virus particulates of which we may assume that (*5*):

1. The particulates are randomly and independently dispersed throughout the suspension.

Second, we consider an idealized test system for which we may assume that:

2. One particulate is an effective dose and when introduced into the test system will invariably make its presence known by eliciting some characteristic response.

Since our immediate interest is the logical structure of the problem of safety testing, we

defer to a subsequent section a discussion of the correspondence, if any, between (*a*) the idealized viral suspension and an actual vaccine containing residual live virus and (*b*) the idealized test system and the tissue culture and monkey tests actually used. In this and the following section the "lots" referred to are assumed to have the characteristics of this idealized suspension.

The first assumption is sufficient to assure that if a test volume of v cubic centimeters is taken from a suspension containing m particulates per cubic centimeter (infection level m) the probability that the sample will contain exactly x particulates is given by the general term of the Poisson distribution, namely:

$$e^{-mv}\frac{(mv)^x}{x!} \qquad [1]$$

This is true when the volume of the suspension (V) is large relative to the volume of test sample (v), as we shall assume in what follows. When this assumption cannot be made (*7*), the required probability is the general term of the binomial distribution, namely:

$$\binom{mV}{x}\left(\frac{v}{V}\right)^x\left(1-\frac{v}{V}\right)^{mV-x} \qquad [2]$$

The second assumption says that the probability of detecting growth in the test system is identical with the probability that the test volume contains one or more particulates, namely:

$$\sum_{x=1}^{\infty} e^{-mv}\frac{(mv)^x}{x!}=1-e^{-mv} \qquad [3]$$

Thus, in testing v cc. from a suspension at infection level m the probability of an erroneously negative test is e^{-mv}. By varying v and m it is possible to explore numerically the probabilities of erroneously accepting suspensions at different infection levels and using different sample sizes. For example, if one tests 500 cc. from a suspension containing 5 particulates per 1,000 cc. the probability of a negative result is 0.08, since

$$e^{-(.005)(500)}=.082$$

Thus, 8 percent of all suspensions at this infection level would pass a test using 500 cc.

The minimum requirements imply that 1,500 cc. of each single strain vaccine will be tested on tissue culture and at least 50 cc. more on monkeys. If assumptions 1 and 2, held for both tissue culture and monkey tests, the additional safety assured by the 50 cc. could for the purposes of this calculation be disregarded. In that case the probability of passing a single-strain pool at infection level 5 per 1,000 cc. would be

$$e^{-(.005)(1000)}\times(\text{probability of a negative in the 1,500 cc. in the trivalent pool})$$

If the infection level for the trivalent pool is also assumed to be .005, this gives a final probability of a false negative of $e^{-(.005)(2500)}$.

Thus, the probability of accepting single-strain pools containing 5 virus particulates per 1,000 cc. would be less than 1/100,000 if assumptions 1 and 2 were satisfied. This is the probability given in the White Paper for passing a single-strain vaccine produced at infection level 5 per 1,000 cc. (*5a*).

Consistency

There are a number of questions that can be raised about this formulation. We note first that it appraises a lot solely on the basis of the evidence furnished by that lot and makes no use of prior information on the consistency or inconsistency with which negative lots have been produced in the past. In practice the Public Health Service's Technical Committee on Poliomyelitis Vaccine, which must approve each lot before it is released, has "been influenced as much by the plant record for consistency of performance as by the negative results of tests on the individual lots considered" (*8*). But there has been no precise criterion of what is meant by consistency. This is the question to which we now turn.

We start by borrowing a concept from the literature of quality control and consider the average outgoing quality of lots passing the new safety test. More precisely we ask: What is the probability that a cubic centimeter of a suspension passing the safety test will contain some specified number of particulates, say one or more? Making the same two assumptions as were made earlier we find that no answer to this

question is possible because we do not know the infection level at which any given vaccine is produced. Suppose, for example, that all lots being produced by a manufacturer contain exactly one particulate per liter. Then no matter what the safety test, so long as any lots at all are passed, the outgoing lots will also contain one particulate per liter. In such a case, of course, a considerable proportion of batches submitted would fail the safety test, and it is unlikely that anyone, producer or tester, would feel any great confidence in the safety of the batches that passed. This example suggests that if one wishes to control the probability that an outgoing cubic centimeter contains live virus, one must consider not only the lot being tested but also the past history of testing, that is, the consistency with which safe lots have been produced. It also suggests a general way of proceeding.

Subject to the assumptions made earlier let us initially consider a manufacturer producing a single-strain pool at constant infection level m. Denote the probability that a cubic centimeter contains one or more particulates infectious for the test system by P. Then

$$P=1-e^{-m} \qquad [4]$$

We shall henceforth refer to P as outgoing quality. Denote the probability that a batch produced at this level of infection will pass when v cc. are tested by Y. Then

$$Y = e^{-mv} \qquad [5]$$

and

$$P = 1 - Y^{1/v} \qquad [6]$$

For this simplest situation we thus have a relation between the probability that an outgoing cubic centimeter contains one or more particulates, P, and the proportion of lots, which pass, Y, for constant test level, v. P is a quantity that we wish to keep below some minimum level; the amount tested, v, is subject to our control; and Y, the proportion of lots passed, can be estimated from past experience. As it stands the model is too simple to be realistic, but solely in the interests of understanding its implications let us explore it numerically. Suppose we set P at some low level, say 5/100,000 and consider v=4,500, that is, we consider the entire testing process to consist of a single test of 4,500 cc. of the final trivalent pool. We then have

$$5\times10^{-5}=1-Y^{1/4500}$$

Solving, we obtain Y=80.0.

That is to say, if a manufacturer is producing lots at a constant but unknown level of contamination, and if 4,500 cc. of each batch are tested and 80.0 percent pass, then, given the assumptions previously made, it follows that out of every 100,000 cc. released, 5 would be expected to contain one or more live virus particulates.

If now under this model we wish to assure that the probability of live particulates in a cubic centimeter of passed vaccine never exceeds 5/100,000, we pass a lot if, and only if: (a) the lot under consideration passes a safety test involving 4,500 cc.; and (b) at least 80.0 percent of previously tested lots have passed.

In practice we should, of course, wish to safeguard ourselves against a number of contingencies, perhaps the most important of which is that the level of m fluctuates from time to time. In that case one might wish to use only recent production information in estimating the value of Y for a producer. Suppose, for example, one looked at only the last 10 lots produced. If the probability of a negative were in fact constant and equal to .8, then the probability of passing all 10 is .11, which is rather high and suggests that a run of 10 negatives is not too improbable even for a Y less than .8. The probability of passing 20 out of 20 when Y=.8 is, however, .012, while the probability of failing 1 out of 20 is .058. One might thus regard 20 negatives out of 20 as evidence at approximately the .99 level of confidence that Y was at least equal to .8 and at least 1 positive out of 20 as evidence at this level that Y might be below .8. An amended procedure for providing that the probability of live particulates in a cubic centimeter from passed lots does not exceed 5/100,000 would be to pass a lot if, and only if, it formed part of a run of 20 negative lots. More generally if we denote by n the size of the run of negative lots required, we have

$$n = \frac{\log \alpha}{v \log (1 - P)}$$

where $(1-\alpha)$ is the level of confidence.

Size of negative run (n) required to insure given confidence ($1-\alpha$) that average outgoing infectivity per cubic centimeter is less than P (for selected values of α, P, and test volume, v).

$(1-\alpha)$	$P=1\times10^{-5}$	$P=5\times10^{-5}$	$P=25\times10^{-5}$	$P=50\times10^{-5}$
	$v=1{,}000$			
.95	303	60	12	6
.99	465	92	18	9
.999	698	138	28	14
.9999	930	185	37	18
.99999	1, 162	231	46	23
	$v=5{,}000$			
.95	61	12	2	1
.99	93	18	4	2
.999	140	28	6	3
.9999	186	37	7	4
.99999	233	46	9	5
	$v=10{,}000$			
.95	30	6	1	1
.99	47	9	2	1
.999	70	14	3	1
.9999	93	18	4	2
.99999	116	23	5	2

We show in the table above the values of n for various levels of confidence, sample volumes, and levels of outgoing quality.

The rule derived is in a general way consistent with recent practice in accepting and rejecting lots. In the 1954 field trials, however, its use would have led to the rejection of the two lots whose production was preceded and followed by lots which tested positive. The lots which tested positive were discarded, but the two lots sandwiched in between were used in the field trials, a practice inconsistent with any consistency rule. (A rereading of the slides which led to calling these lots positive has, in the light of the accumulated experience, prompted rediagnosis. All four lots are now believed to have tested negative, according to a personal communication from David Bodian. This finding, of course, should not be construed as justification of the 1954 practice but rather as an explanation of why it did not lead to difficulties.)

Most of the modifications of this model which are required to make it more realistic are straightforward, but involved, and we shall not discuss them. There is one modification of possible interest that we mention here, however. We have justified the rule of estimating Y only from the 20 previously produced lots by considering the possibility that the level, m, fluctuates from lot to lot. The relation between P and Y was obtained on the assumption that it is constant, however. Does the relation between P and Y continue to hold when m is no longer assumed constant from lot to lot?

A perfectly general answer can be supplied. No matter how m fluctuates one can show

$$\overline{P}\leq 1-\overline{Y}^{1/v} \qquad [7]$$

where $\overline{P}$ and $\overline{Y}$ are averages of equations 4 and 5 over the appropriate distributions of m (see inset). The equality between P and Y pre-

Derivation of Equation 7

Define P and Y by equations 4 and 5; let m be the level of infection of a lot, and let v be the cubic centimeters tested of each lot. If now we suppose that m has an arbitrary distribution $F(m)$, we define $\overline{Y}$ by

$$\overline{Y}=\int_0^1 e^{-mv}dF(m) \qquad [8]$$

and $\overline{P}$ by

$$\overline{P}=\frac{\int_0^1(1-e^{-m})e^{-mv}dF(m)}{\int_0^1 e^{-mv}dF(m)} \qquad [9]$$
$$=1-\frac{\int_0^1 e^{-m(v+1)}dF(m)}{\overline{Y}}$$

From Liapounoff's inequality

$$\int_0^1 e^{-m(v+1)}dF(m)\geq\left[\int_0^1 e^{-mv}dF(m)\right]^{\frac{v+1}{v}} \qquad [10]$$

Substituting from equation 10 in equation 9, equation 7 follows. A somewhat more general result has been obtained by Paul Meier in a study as yet unpublished.

viously used now turns out to be an inequality and, fortunately, in the direction to make it useful. Thus, if 80.0 percent of a long series of lots, 4,500 cc. of each of which are tested, test negative, then the probability that a cubic centimeter of passed material contains one or more particulates is equal to or less than 5/100,000. The assumption that m is constant is consequently the least favorable one for the safety tester, and the procedure suggested is one which protects him against the least favorable a priori distribution of m.

What are the factors that will influence the level at which P, the level of outgoing quality, is set? This is not a statistical question and consequently not one to which we can give an answer. It is, nevertheless, a question to which an answer is required, and it is worth indicating briefly some of the issues involved. First of all, P cannot be set at zero. That is to say, no amount of consistency testing can assure the complete absence of infectivity. The most that can be done is to keep P, the proportion of infected cubic centimeters, below some preassigned level. In selecting a numerical value for P, one must be guided by the consequences of the choice. The first major difficulty is that the consequence of introducing one virus particle into a human host is unknown. That is to say, if P is set at some value say 5/100,000, and the production process is such that five 1-cc. inoculums in every 100,000 released do in fact contain one virus particulate, we are unable to say whether any of the five children receiving the infected inoculums would contract the disease. In the next section we shall consider more closely the relation between exposure to live virus particulates and the subsequent development of disease. Here we shall simply make the overly conservative assumption that all children exposed to one or more live virus particulates invariably develop the disease—that in setting a value of P we are also setting the incidence rate for poliomyelitis.

Now, the average annual incidence of paralytic poliomyelitis is very low. In the average epidemic year of 1954 it was about 50 per 100,000 persons in the age group 5–9 in areas covered by the field trial of the vaccine (*9*). The Francis report indicates that the 1954 field trial vaccines reduced this rate by at least 50 percent, and epidemiological analysis of the 1955 experience leads to much the same conclusion (*10*). Thus, a value of P set at, say 50 percent of 50/100,000 would not be safe enough since such a safety testing procedure could barely assure that passed vaccines would not cause more cases than they prevented.

The Cutter experience is illuminating. Approximately 400,000 children were inoculated with vaccines from 17 different filling lots produced by Cutter Laboratories. There were 61 cases of paralytic poliomyelitis among these children within 50 days of vaccination and an additional 97 among family or community contacts within 65 days. This amounts to an overall paralytic rate of 40 per 100,000 persons, most of which can be attributed to the vaccine. If we insist on setting P at 50 percent of 50/100,000, then the level selected is approximately equal to the average Cutter level of infection. Undeniably this is not safe enough and P must be set well below 25/100,000. How far below 25/100,000 is suggested by the fact that Cutter vaccines were withdrawn from use not after 61 cases but after the first 6 cases out of 400,000 vaccinations.

One might of course argue that it is inappropriate to consider the annual incidence of poliomyelitis and that, in fact, a more appropriate magnitude is the lifetime probability of contracting the disease. This probability is a good deal higher, about 800 per 100,000 persons by age 24 according to a study of children of native white parents in 28 cities (*11*). Since this is 10 to 15 times the annual incidence, one might incline to a value of P well above 25/100,000. There are several problems raised by this issue, however. First of all, if the risk of infection by the vaccine is to be balanced against the lifetime probability of developing the disease, then it is necessary that the vaccine confer lifetime immunity. Whether this is in fact the case is not now known and presumably will not be known for some time (see Salk (*12*), however). Second, and perhaps more fundamental, it is doubtful whether any community would (or should) tolerate safety standards that will permit the release of vaccines that raised the incidence of paralytic poliomyelitis in that year

on the grounds that the increase would be more than counterbalanced by decreases in subsequent years.

It is even more difficult to indicate what factors should be considered in selecting a level of confidence. From a practical point of view, however, the value selected is not as crucial as that of P. Thus, increasing the level of confidence from .99 to .99999 increases the required length of negative runs by about two- and one-half-fold, whereas decreasing P from 5 to 1 per 100,000 increases it by fivefold. In general, the choice requires a compromise between the desire for a high degree of confidence and practical limitations on possible sizes of n.

It is important to realize that, after one has determined a value for outgoing quality, say x/100,000, and a confidence level, $1-\alpha$, it does not necessarily follow that in proportion α of the negative runs x cc. in every 100,000 will in fact contain live virus. What will actually happen depends on the safety of the production process. If the production process is safe, such levels will not occur. The logical structure of safety testing, in short, necessitates fixing a maximally tolerable level of outgoing infectivity, but this level need not necessarily ever be realized.

On the Assumptions Used

The key assumptions of the preceding section are that: (*a*) a vaccine can be considered as an idealized suspension of randomly and independently dispersed particulates, (*b*) the tissue culture and monkey tests used can be considered as an idealized test system capable of invariably detecting the presence of a single virus particulate, and (*c*) a child may be considered maximally sensitive and invariably capable of developing poliomyelitis, even when exposed to a single virus particle.

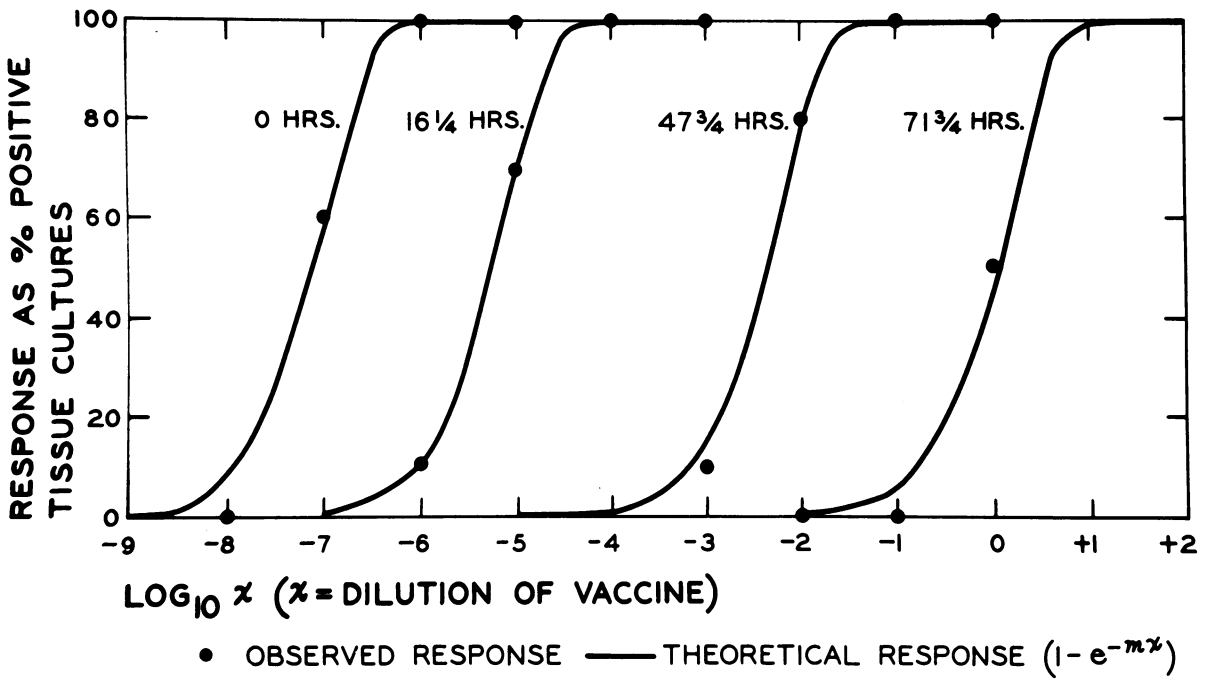
Several bits of evidence suggest that the last assumption is incorrect by several orders of magnitude. First of all, most persons have developed an immunity to the disease by the time they reach adult age even though not more than 1 percent have ever had clinically manifest poliomyelitis. Associated with this is an increased level of neutralizing antibodies (*13*), which suggests that most adults were at some time in life infected by poliomyelitis virus without ever having developed the disease. More direct evidence on this point is provided by a longitudinal study of familial infection with poliomyelitis virus by Fox and associates (*14*). During a 3-year study of 156 households, they noted the development of 240 cases of infection with poliomyelitis virus as indicated either by the recovery of virus from stools or elevated serum antibody levels or both. There was not a single instance of paralytic disease in any of the 240 cases. Extrapolating these results to the community of which the households were considered to be a sample (for which community the incidence of paralytic poliomyelitis was known) they concluded that 1 paralytic case develops for every 710 cases of infection with poliomyelitis virus by the oral route.

Further suggestive evidence on this point has been brought to our attention by Nathanson and Hall. There were 105 cases of poliomyelitis among family contacts of Cutter vaccinated individuals. These may be presumed to have been infected by vaccinated family members. Only 1 of these 105 family members developed a case of the disease. Thus, on this premise, of 105 individuals sufficiently infected to pass the disease on to others 104 did not develop clinical poliomyelitis. The assumption that a child is maximally sensitive is thus a very conservative one. Testing procedures based upon this assumption will possess a considerable margin of safety, at least on this score.

Turning to the first two assumptions, we note that they imply that if one tests v cc. of a vaccine at infection level m, the probability of a positive result is from equation 3, $1-e^{-mv}$.

This relationship between the probability of a positive result and amount tested, often referred to as the one-particle curve, is a well-known relationship in virology. It has been tested on a variety of plant and animal viruses and usually, although not invariably, found to apply (*15*). In principle, its applicability to the present problem could be tested by varying v in a vaccine preparation known to contain incompletely inactivated live virus. In practice, v can be varied only by diluting the vaccine, and the amount of live virus present in the vaccines tested has not been sufficient to give positive responses after dilution. To investigate

Figure 3. Tissue culture response during inactivation.



the question, we have consequently turned to data lying behind the inactivation curves previously discussed. The log titers on these curves are obtained by testing four successive tenfold dilutions. A volume of 0.5 cc. of the diluted suspension is introduced into each of 10 tissue culture bottles at each dilution and the presence or absence of viral growth noted. We show in figure 3 the results of one such run at four different inactivation times. The description of the relation between proportion of positive bottles and dose provided by the one-particle curve appears satisfactory, although a more searching examination would be possible if the spacing between dose levels were not so wide. A very large number of such comparisons is in fact possible, one for each single-strain lot of vaccine produced by each manufacturer. We have not investigated more than a fraction of them, but, for most of those that we have, the agreement between observation and hypothesis shown in the figure is by no means unusual.

Such agreement would appear to validate both the assumptions of an ideal suspension and of an ideal test system. This is too hasty a conclusion, however. Aside from the fact that no data are given for inactivation times beyond the third day, the assumption that one particulate will invariably initiate growth is sufficient but not necessary for the derivation of the one-particle curve. Thus, if we substitute for assumption 2 (p. 1048) the less limiting assumption:

> 2′. The probability that a virus particulate will initiate growth is constant for all tissue culture bottles and equal to p, and this probability does not depend on whether other particulates present have or have not initiated growth,

we also obtain the one-particle curve. Thus, the probability that a test batch of v cc. will contain z particles and that none of these will result in growth is

$$e^{-mv}\frac{(mv)^z}{z!}(1-p)^z \qquad [11]$$

The probability of no growth is simply the sum of such terms over all values of z, and this sum is simply

$$e^{-pmv} \qquad [12]$$

We can estimate the product pm from data such as that given in figure 3, but not p and m separately. In consequence the agreement between observed and theoretical in figure 3 provides no evidence on the numerical value of p and

hence on the choice between assumptions 2 and 2′.

In the literature of virology the parameter m is usually referred to as the number of elementary bodies and the parameter pm as the number of infectious units. The ratio of infectious units to elementary bodies has been determined for a number of virus-host systems. Two lines of evidence suggest that for the poliomyelitis virus its value in tissue culture may be well below unity and that assumption 2′ rather than 2 is the appropriate one. The first is provided by electron microscope photographs of purified poliomyelitis virus preparations. From these it has been estimated that $p=1/30$ to $1/60$ (*16*). The second is provided by the results of intraspinal inoculation with infectious vaccine of monkeys rendered especially sensitive by pretreatment with cortisone. In a number of such experiments the dose required to infect the monkeys has been only a fraction of the dose required for tissue culture. Neither piece of evidence can at present be considered much more than suggestive. The results of the monkey experiments are as yet unpublished and require confirmation, while uncertainty as to the viability of the particles seen in the electron micrograph enjoins caution in interpretation. In the words of Dulbecco and Vogt (*17*) the relation between infectious units and elementary bodies "is still an open problem of highest interest." The mere fact that it is an open question does suggest, however, that a logical structure which is dependent upon the validity of assumption 2 may not be firmly grounded.

The first assumption, that a vaccine may be considered as an idealized suspension of randomly and independently dispersed particulates has also been questioned. Thus, Veldee has argued (*18*) that in the original virus suspension a significant proportion of virus particles are known to be imbedded in gelatinous protein material which cannot be removed by the finest filters. He suggests that formaldehyde may harden this material so that the imbedded virus particles cannot attach to the tissue cell in tissue culture. Growth of the virus in tissue culture is thereby prevented. Once the vaccine has been injected into a living animal, he goes on to suggest, enzymes present in the animal, but not in tissue culture, may free the virus particle of its coating, after which growth may take place. No evidence that would either support or contradict this hypothesis is known to us.

If in the light of this discussion we reexamine the preceding section, it becomes apparent that the three assumptions listed at the beginning of the present section are sufficient but not necessary. The necessary assumptions are less limiting, namely, (*a*) the test system used is at least as sensitive to the presence of live virus as the human subject and (*b*) the probability of a negative response in tissue culture is a decreasing exponential function of test volume. The second assumption is supported by results of the type summarized in figure 3. The first is the crucial one and unfortunately the one about which only indirect evidence is ever likely to be available.

To cast some light on its possible validity, we consider the only evidence now available—the results of a cooperative study, undertaken immediately after the Cutter incident, of 16 of the 17 filling lots of vaccine produced by Cutter Laboratories. All 16 lots were tested in tissue culture, the total amount tested being somewhat less than 6 liters, or considerably less than is called for by present minimum requirements. Nine of the sixteen lots were also tested in 391 normal monkeys and 10 of the 16 in 178 intraspinally inoculated monkeys, which had been pretreated with cortisone. Of the 16 lots tested 6 were associated with an excess incidence of poliomyelitis (*9*). All 6 of these lots gave positive results, and, in addition, 2 of the lots that were not epidemiologically implicated also tested positive, 1 by tissue culture and 1 by cortisone treated monkey. These tests, which are not as extensive as those now called for by the minimum requirements, thus indicate that at least so far as the Cutter lots of the spring of 1955 are concerned the test system now used is as sensitive as and probably more sensitive than the human host. It would, of course, be desirable to have a good deal more information than can be extracted from that experience, but the results as far as they go are in the direction of validating the present testing program. Between the Cutter incident (and after the adoption of the new minimum requirements on May 26) and January 1, 1956, 37,500,000 cc. were released for public use without, so far as is

known, being related to the development of any further cases. This is consistent with the conclusion suggested by the post-Cutter cooperative tests although it is of course hopelessly confounded with simultaneous improvements in the safety of the production process.

Summary

The safety of a vaccine will depend both upon the basic safety of the production process and the ability of safety testing procedures to detect an unsafe vaccine, if one is produced. Exclusive reliance upon the safety of a production process with a good past record, without the second line of defense provided by an adequate safety test, may be hazardous. Salk's early discussions of the nature of the inactivation process are reviewed, and his conclusion that safety was assured by the predictable nature of the inactivation process is critically appraised.

The major statistical problem in safety testing involves a decision as to how much testing is required. The procedures used in industrial quality control to solve this problem are not applicable to vaccine testing because the costs of erroneously rejecting good lots and of erroneously accepting bad ones are entirely incommensurable. An answer is derived instead by making certain statistical assumptions as to the dispersion of live virus in the vaccine, the sensitivity of the test system, and the viral concentration which is infective for man. The question of consistency testing is considered and a general theory derived for deciding how many successive lots of vaccine testing negative are required before a producer can be said to be a consistent producer of safe vaccine. Some of the questions that must be answered to apply this theory are considered. The assumptions on which this theory is based are critically analyzed.

REFERENCES

(*1*) U. S. Public Health Service: Regulations. Biologic products—Additional standards: Poliomyelitis vaccine. 21 Fed. Reg. 4922 (1956).

(*2*) Salk, J. E.: Vaccination against paralytic poliomyelitis, performance and prospects. Am. J. Pub. Health 45: 575–596 (1955).

(*3*) Salk, J. E., Krech, V., Youngner, J. S., Bennet, B. L., Lewis, L. J., and Bazeley, P. L.: Formaldehyde treatment and safety testing of experimental poliomyelitis vaccine. Am. J. Pub. Health 44: 563–570 (1954).

(*4*) Clark, W. M.: Topics in physical chemistry. Baltimore, Williams and Wilkins Co., 1948, ch. 12.

(*5*) U. S. Public Health Service: Technical report on Salk poliomyelitis vaccine. Washington, D. C., 1955; (*a*) p. 69. Mimeographed.

(*6*) Barnard, G. A.: Sampling inspections and statistical decisions. J. Roy. Stat. Soc. 16 (series B): 151–174 (1954).

(*7*) Prigge, R., Gunther, O., Bonin, O., Eissner, G., Hallevorden, T., and Spaar, W.; Probleme der Staatlichen prufung von Poliomyelitis—Impstoffen. Deutsche med. Wchnschr. 1956. In press.

(*8*) U. S. Public Health Service Technical Committee on Poliomyelitis Vaccine: Interim report. Washington, D. C., 1955. Mimeographed.

(*9*) Francis, T., Korns, R. F., Voight, R. B., Boisen, M., Hemphill, F. M., Napier, J. A., and Tolchinsky, E.: An evaluation of the 1954 poliomyelitis vaccine trials. Am. J. Pub. Health 45 (pt. 2): 1–63, May 1955.

(*10*) Langmuir, A. D., Nathanson, N., and Hall, W. J.: The surveillance of poliomyelitis in the United States in 1955. Am. J. Pub. Health 45: 75–88, January 1956.

(*11*) Collins, S. D.: The incidence of poliomyelitis and its crippling effects, as recorded in family surveys. Pub. Health Rep. 61: 327–355, Mar. 8, 1946.

(*12*) New York Times, May 2, 1956, p. 1.

(*13*) Melnick, L., Paul, J. R., and Walton, M.: Serologic epidemiology of poliomyelitis. Am. J. Pub. Health 45: 429–437, April 1955.

(*14*) Fox, J. P., Gelfand, H. M., Le Blanc, D. R., and Conwell, D. P.: A continuing study of the acquisition of natural immunity to poliomyelitis in representative Louisiana households. Am. J. Pub. Health 46: 283–294, March 1956.

(*15*) Lauffer, M. A., and Price, W. C.: Infection by viruses. Arch. Biochem. 8: 449–468, December 1945.

(*16*) Fogh, J., and Schwerdt, C. E.: Physical particle per plaque ratio observed for human poliomyelitis viruses. Federation Proc. 15: 253–254 (1956).

(*17*) Dulbecco, R., and Vogt, M.: Biological properties of poliomyelitis viruses as studied by the plaque technique. *In* Biology of poliomyelitis. Ann. New York Acad. Sc. 61: 790–800 (1955).

(*18*) Veldee, M. V.: Letter to the editor. New England J. Med. 253: 483–484, Sept. 15, 1955.