

STATISTICAL METHODS USED IN THE EVALUATION OF THE POLIOMYELITIS VACCINE

This chapter includes some amplification and extension of the material published in the "Statistical Methods Used in Analysis" section of the Appendix in the Summary Report of April, 1955. Its primary purpose is to indicate the statistical methods used in the evaluation. In addition, an alternate method of testing differences and of estimating limits is given.

Methods used in this evaluation are commonly employed in statistical analysis and for that reason only those portions of derivations of equations are given which were judged to be necessary for understanding the symbols and steps used in solving the equations. Numerical examples from the data of the Field Trial show (1) results of computation of limits of effectiveness and (2) tests of significance of differences.

Reference to other standard statistical methods including "Chi-square" tests and "Analysis of Variance" are included because these are used in some sections of the text of this report. However, their derivation, procedures for computation, and the results are presented in many texts, and so only the results of quantities from data of the Field Trial may be made in the appropriate models.

Evaluation of data from scientific investigations involves several basic concepts among which are:

1. Uniform classification of data;
2. Establishment of uniform units of measures;
3. Uniformity of methods of comparison of data within classifications.

Examples of the application of these concepts in the evaluation of data from the Field

Trial included the establishment of criteria for classifications of "vaccinated" and "control" children, and "paralytic cases." The concept of uniform units of measures of the data is exemplified by "rates" per 100,000 persons. Instances of the use of uniformity in methods of comparing various classifications of data are shown in the tests of significance of difference between rates among vaccinated and their respective control groups.

The general statistical approaches for analysis were determined prior to the collection of the data of the Field Trial insofar as was feasible. Specific statistical methods applied to analyses must be selected subsequent to consideration of the design and procedures followed in the collection of the data as well as the general nature of the distributions into which the data are assembled for study. Other sections of this report describe the plans for collection of data of the Field Trial and provide bases for evaluating the extent to which these plans were consummated. Many tables and graphs are presented for the classification of data. The implications with regard to the distributions of selected characteristics within the population studied may be drawn.

Several methods of statistical evaluation were considered and applied to various topics for analysis. Those methods which were used for final evaluation were judged to be most reasonably applicable to the distributions of the data arranged for studying a specific topic and having underlying concepts consonant with the design and procedures by which the data were collected. The design and procedures followed in the observed areas indicate that the tests of significance are not valid (though

STATISTICAL METHODS

they do have a limited usefulness) because of the biases known to be present.

The statistical concepts and specific methods used in the final evaluation are given under the following topics:

1. Rates
2. Estimates of effectiveness
3. Tests for significance and confidence limits
4. Another approach to statistical evaluation
5. Estimates of population characteristics
6. Chi-square
7. Analysis of variance
8. Geometric means

For the convenience of those who may wish to study these subjects in detail, references are listed in the Bibliography as follows: 4, 11, 23, 24, 28, 45, and 51.

RATES

In order to compare incidence of poliomyelitis between various groups, rates have been used throughout the evaluation, that is, the number of cases per 100,000 persons as is customary in work involving populations. This rate may be formulated as $R = \frac{n}{N} 100,000$, where N is the number of persons in the group under consideration and n is the number of cases of the category under consideration (reported, paralytic, Type I virus isolated, etc.) occurring among the N persons in the group.

ESTIMATES OF EFFECTIVENESS

Effectiveness as computed for this evaluation was defined as the percentage of cases attributed to prevention by the vaccine. The percentage estimate of effectiveness which was used in the evaluation may be written;

$$E = 100 \left[1 - \left(\frac{R_v}{R_c} \right) \right] \%$$

where R_v is the case rate per 100,000 persons for the vaccinated group under consideration and R_c is the case rate per 100,000 persons for the corresponding control group.

TESTS FOR SIGNIFICANCE AND CONFIDENCE LIMITS

In consideration of effectiveness of the vaccine, there was the possibility that an observed effectiveness could have occurred by sampling variation rather than being due to the vaccine. Statistical tests of significance were made in order to consider the possibility that the observed differences in rates could have occurred as a result of sampling variation. Assuming the vaccine to be ineffective, if the probability of a difference as great as, or greater than, that observed in the Trial is 0.05 (or less), confidence is developed at the 95 percent level in the decision that the vaccine has had some true effect. If this probability is greater than 0.05, any inference about the preventive powers of the vaccine was considered as unjustified.

The test used in the evaluation was a "one-tailed" test since consideration was given only to the "true effectiveness greater than zero" and not the hypothesis "true effectiveness less than zero." The decision to use this one-tailed test would have been inappropriate had there been any significant evidence from the data of the Field Trial that the disease was contracted as a result of vaccination. There was no such evidence of this nature from the Field Trial data. The procedure of analysis involved first the study of safety and then the study of effectiveness. Hence, "safety" was treated as a separate subject and only when the decision had been reached that the vaccine(s) used in the Field Trial had been safe was any effort made to estimate effectiveness. Computations of "two-tailed" limits establish "expected lower limits" somewhat lower than those computed

STATISTICAL METHODS

from the one-tailed approach. However, the decision with regard to all major segments of the study remains the same for either the one- or two-tailed approach.

Probability Distribution of Cases

The following symbols are used throughout the discussion of statistical methods:

- M = number of vaccinated persons
- N = number of persons in corresponding control group
- m = number of cases among vaccinated persons
- n = number of cases among persons in corresponding control group
- T = total number of persons (T = M + N)
- t = total number of cases (t = m + n)

The test for significance which was used assumes that in the i^{th} school the number of cases in the vaccinated group follows a Poisson distribution with mean a_{iv} . The Poisson distribution has been found to hold widely in biological data for the frequency distribution of events that have a low probability of occurrence. The mean a_{iv} will presumably vary from school to school because of variations in the level of poliomyelitis infection, antibody status of the children, effectiveness of the vaccine lots, and so on. However, it is a property of the Poisson distribution that the sum of a number of independent Poisson variates also follows a Poisson distribution with mean $\sum a_{iv} = A_v$ (say), this being true whether the a_{iv} are equal or not. Hence, the total number of cases in the vaccinated group follows a Poisson distribution. Similarly, this argument holds for the control group, where A_c corresponds to A_v .

Letting $\frac{A_v}{M} = p_v$, and $\frac{A_c}{N} = p_c$,

the true effectiveness is

$$E = 100 \left[1 - \left(\frac{p_v}{p_c} \right) \right] \% = 100 (1 - \lambda) \%$$

$$\text{where } \lambda = \frac{p_v}{p_c}$$

Then the probabilities of "m" cases occurring among the vaccinated group and of "n" cases occurring among the corresponding control group are:

$$\frac{(Mp_v)^m}{m!} e^{-Mp_v} \quad \text{and} \quad \frac{(Np_c)^n}{n!} e^{-Np_c},$$

respectively.

The probability of a total of "t" cases in both groups combined is:

$$\frac{(Mp_v + Np_c)^t}{t!} e^{-(Mp_v + Np_c)}$$

Given a total of "t" cases, the probability of "m" vaccinated cases is:

$$\frac{\left[\frac{(Mp_v)^m}{m!} e^{-Mp_v} \right] \left[\frac{(Np_c)^{t-m}}{(t-m)!} e^{-Np_c} \right]}{\frac{(Mp_v + Np_c)^t}{t!} e^{-(Mp_v + Np_c)}}$$

Cancelling out the exponential terms, this becomes:

$$\frac{t!}{m! (t-m)!} \frac{(Mp_v)^m (Np_c)^{t-m}}{(Mp_v + Np_c)^t}$$

But $p_v = \lambda p_c$, and $\frac{t!}{m! (t-m)!} = \binom{t}{m}$.

Thus we obtain $\binom{t}{m} \frac{(M\lambda p_c)^m (Np_c)^{t-m}}{(M\lambda p_c + Np_c)^t}$,

and cancelling out p_c , we have

$$\binom{t}{m} \frac{M\lambda^m N^{t-m}}{(M\lambda + N)^t}$$

STATISTICAL METHODS

Finally, the probability of "m" vaccinated cases, given a total of "t" cases, may be expressed as:

$$\binom{t}{m} \left(\frac{M\lambda}{M\lambda + N} \right)^m \left(\frac{N}{M\lambda + N} \right)^{t-m}$$

This is the binomial distribution with probability parameter

$$\frac{M\lambda}{M\lambda + N}$$

Tests of Significance

The probability expression derived above is appropriate for testing the hypothesis $E = 0$ against the alternative $E > 0$. When $E = 0$, $\lambda = 1$. If the vaccine were ineffective, ($E = 0$), the proportion of cases that had been vaccinated, (m/t), would be equal to the proportion of persons that had been vaccinated (M/T). If the vaccine had some effect, it follows that

$$\frac{m}{t} < \frac{M}{T}$$

The test is made by evaluating the expression:

$$B = \sum_{k=0}^m \binom{t}{k} \left(\frac{M}{T} \right)^k \left(\frac{N}{T} \right)^{t-k}$$

B is the probability of getting "m" or fewer vaccinated cases out of a total of "t" cases. In the evaluation of the vaccine, if $B \leq 0.05$, the effectiveness was considered to be significantly larger than zero at the (100B) percent level of significance. If $B > 0.05$, the effectiveness of the vaccine was not considered to be significantly different from zero.

Although the evaluation of B seems to be a difficult task, there are tables available which make this a very simple procedure. In the evaluation of the poliomyelitis vaccine, the tables used were the Tables of the Cumulative Binomial Probabilities,³¹ which permitted rapid evaluation of "B" for values of "t" as large as 150. In terms of these tables,

$$B = 1 - P(m + 1, t, \frac{M}{T}) \text{ when } \frac{M}{T} \leq 0.50,$$

$$\text{or } B = P(n, t, \frac{N}{T}) \text{ when } \frac{M}{T} > 0.50.$$

When "t" is larger than 150, the normal distribution may be used to approximate the binomial distribution. Find

$$z = \frac{t \left(\frac{M}{T} \right) - m - \frac{1}{2}}{\sqrt{t \left(\frac{M}{T} \right) \left(\frac{N}{T} \right)}}$$

and substitute in

$$B' = \frac{1}{\sqrt{2\pi}} \int_z^{\infty} e^{-a^2/2} da.$$

This integral may be evaluated from Tables of the Probability Functions.²⁸ In terms of these tables,

$$B' = \frac{1}{2} \left[1 - \frac{1}{\sqrt{2\pi}} \int_{-z}^z e^{-a^2/2} da \right].$$

Other less extensive tables of the cumulative normal distribution may be used for this purpose.

The decision procedure regarding effectiveness, using B' , is the same as that described above for B.

Confidence Limits

Corresponding to the approach used in the test of significance, a 5 percent lower limit (one-sided) was used in the evaluation of poliomyelitis vaccine. Only a lower limit was computed, hence, referred to as "one-tailed." The upper limit was not considered pertinent to evaluation in this study.

In order to find L, the 5 percent lower limit for the effectiveness, solve the equation

$$\sum_{k=0}^m \binom{t}{k} \left(\frac{M\lambda}{M\lambda + N} \right)^k \left(\frac{N}{M\lambda + N} \right)^{t-k} = 0.05$$

for λ . Then the lower limit is $L = 100(1 - \lambda)\%$.

STATISTICAL METHODS

For values of $t \leq 150$, the binomial tables cited above may be used. Using these tables, find λ such that

$$P(m + 1, t, \frac{M\lambda}{M\lambda + N}) = 0.95,$$

$$\text{or } P(n, t, \frac{N}{M\lambda + N}) = 0.05.$$

Only one of these approaches will lead to λ in the tables, except when

$$\lambda = \frac{N}{M}.$$

When $t > 150$, the normal approximation was used. This was done by solving the equation:

$$\frac{t \left(\frac{M\lambda}{M\lambda + N} \right) - m - \frac{1}{2}}{\sqrt{t \left(\frac{M\lambda}{M\lambda + N} \right) \left(\frac{N}{M\lambda + N} \right)}} = 1.64$$

for λ (two values) and substituting the larger value of λ in the expression $L = 100(1 - \lambda)\%$.

ANOTHER APPROACH

TO STATISTICAL EVALUATION

As a check on the methods used to obtain tests of significance and confidence limits, these operations can be performed by an alternative approach, suggested by Professor William G. Cochran, which depends on the assumption of a Poisson distribution. In the alternative approach, the county is regarded as the basic unit of sampling. In each county there is a paired trial of vaccine versus control. The standard error of the study-wide difference in case rates is estimated from the observed variation in rates from county to county, by formulas commonly used in the analysis of variance and in sample surveys.

Inferences made by this approach apply to the type of population of which the counties in the Field Trial can be regarded as a random sample. It is realized that the counties in the Trial were not a random sample of the coun-

ties in the United States, since the trial counties were specifically selected for certain characteristics. In view of this selection, there is no statistical technic by which inferences made from the data can be extended to the United States as a whole. Any such inferences must be based, in part at least, on judgment. The method follows:

Tests and Confidence Limits When Sizes of Vaccinated and Control Groups are Equal

In a typical county, say the i^{th} county, let " m_i " and " n_i " be the numbers of cases for vaccinated and control groups, respectively, and let " M_i " and " N_i " be the corresponding numbers (in 100,000's) of vaccinated and control children. The estimated rates are:

$$R_v = \frac{\sum m_i}{\sum M_i} \text{ and } R_c = \frac{\sum n_i}{\sum N_i}.$$

Since M_i and N_i are nearly equal, we may write, to a close approximation,

$$R_c - R_v = \frac{\sum (n_i - m_i)}{W} = \frac{\sum d_i}{W}$$

where W is the average of $\sum M_i$ and $\sum N_i$, and $d_i = n_i - m_i$.

An unbiased estimate of the variance of $\sum d_i$ is:

$$s^2 = \frac{a}{a-1} \left[\sum d_i^2 - \frac{(\sum d_i)^2}{a} \right]$$

where a = number of counties. Hence, the standard error of $(R_c - R_v)$ may be taken as

$$SE = \frac{s}{W}.$$

This estimate of standard error may be subject to question because " W " is a random variable. However, for making a test of significance, " W " is regarded as conditionally fixed.

The test of significance is made by finding

STATISTICAL METHODS

$$z = \frac{R_c - R_v}{SE}$$

and comparing with tabulated values of the normal distribution. This involves some approximation since z is not exactly normally distributed.

To find the 5 percent lower limit by this method, first compute:

$$s_m^2 = \frac{1}{a-1} \sum (m_i - \bar{m})^2,$$

$$s_{mn} = \frac{1}{a-1} \sum (m_i - \bar{m})(n_i - \bar{n}),$$

$$\text{and } s_n^2 = \frac{1}{a-1} \sum (n_i - \bar{n})^2;$$

$$\text{where } \bar{m} = \frac{1}{a} \sum m_i, \text{ and } \bar{n} = \frac{1}{a} \sum n_i.$$

Then, with λ such that $E = 100(1 - \lambda)\%$ as before, by the argument for finding confidence limits of a ratio,

$$\frac{\sum m_i - \lambda \sum n_i}{\sqrt{a(s_m^2 - 2\lambda s_{mn} + \lambda^2 s_n^2)}}$$

is approximately a normal variate with mean zero and unit standard deviation. Setting this expression equal to 1.64 and solving for λ (two values), the 5 percent lower limit L is found by substituting the larger value of λ in the expression $L = 100(1 - \lambda)\%$.

Tests and Confidence Limits When Sizes of Vaccinated and Control Groups are Unequal

Here $\sum M_i$ and $\sum N_i$ are not equal and the ratio $\frac{M_i}{N_i}$ varies from county to county. One way to make a test of significance or to de-

termine a lower limit in this situation is to use the approximate formula for the standard error of a ratio as developed for sample survey work.

Using this method the variance of $R_c - R_v$, under the null hypothesis ($E = 0$), is approximated by

$$(SE)^2 = s_v^2 - 2s_{vc} + s_c^2,$$

$$\text{where } s_v^2 = \left(\frac{a}{a-1}\right) \frac{\sum (m_i - R_v M_i)^2}{(\sum M_i)^2},$$

$$s_{vc} = \left(\frac{a}{a-1}\right) \frac{(m_i - R_v M_i)(n_i - R_c N_i)}{(\sum M_i)(\sum N_i)},$$

$$\text{and } s_c^2 = \left(\frac{a}{a-1}\right) \frac{\sum (n_i - R_c N_i)^2}{(\sum N_i)^2}.$$

The test is made by finding

$$z = \frac{R_c - R_v}{SE}$$

and proceeding as in the method for testing described above in the case when counties have equal numbers in the vaccinated and control groups.

For the lower limit the same approximations of variance as in the "Test of Significance" are used. In the non-null case this leads to the normal variate

$$\frac{R_v - \lambda R_c}{\sqrt{s_v^2 - 2\lambda s_{vc} + \lambda^2 s_c^2}}$$

Setting this expression equal to 1.64 and solving this equation for λ (two values) the lower limit L is found by substituting the larger value of λ in the equation $L = 100(1 - \lambda)\%$ as before.

STATISTICAL METHODS

Comparing the Sampling Survey Method and the Poisson Distribution Method

Using data for paralytic cases (Table 5, Chapter I) the following results were obtained:

total of 1,829,916 was drawn for this purpose. This sample was found to estimate tabulated characteristics quite accurately, thereby giving confidence in estimates of untabulated characteristics. The table on page 339 shows vaccination status as determined from the

Status	Study Population	Paralytic Cases	E	Poisson Method		Sampling Method	
				SL	Lower Limit	SL	Lower Limit
Placebo Areas (80)							
Vaccinated	200,745	33	69.92%	<.001	57.71%	<.001	57.72%
Placebo	201,229	110					
Observed Areas (127)							
Vaccinated	221,998	38	62.49%	<.001	49.72%	<.001	49.01%
Placebo	725,173	331					

S. L. - Level of statistical significance.

Source: Table 5, Chapter I.

Both methods lead to similar results. Other experience indicates that the estimates of degree of significance made by the sampling method tend to be somewhat higher than those computed by the Poisson method when significance is high, and lower when the significance is low. However, the 5 percent lower limits were almost equal for both methods in all cases. In view of the similarity of the results, the approach assuming a Poisson distribution of cases was used since, in comparison, the sampling survey method involves lengthy computation.

ESTIMATES OF POPULATION CHARACTERISTICS

Estimates of selected characteristics of the entire study population were required during the course of the evaluation of the Field Trial. Complete tabulations of the population cards for these characteristics were unwarranted and, therefore, a systematic sample of 2,545 basic data cards out of a

tabulation of the total study population and from the sample for placebo and observed study areas.

The percent distributions agree very well and, therefore, estimates from the sample were used (as the denominators) in adjusting age of children for studying "age-temporal-incidence" variation.

CHI-SQUARE

Chi-square was used several times in the course of the evaluation to test significance of differences between observed frequencies and expected frequencies. The chi-square was computed by using

$$\chi^2 = \frac{\sum(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

The method of calculating expected frequencies varies from one application to another and is explained in each instance.

STATISTICAL METHODS

ANALYSIS OF VARIANCE

In the section of the report dealing with laboratory methods, analysis of variance was used to determine between-laboratory and within-laboratory variation and to test for significance of between-laboratory variation. Standard methods as described in several statistical texts were used.^{23, 24, 45}

GEOMETRIC MEANS

In sections of the report dealing with serum antibody titers, it was necessary in many instances to find mean titers. A geometric mean was used for this purpose. Geometric means were found by taking logarithms of the titers and finding the arithmetic means of the logarithms, and the arithmetic means of the logarithms were then converted back to geo-

metric means of titer readings by taking the antilogarithms.

In most laboratory reports the titers were based on 2-fold or 4-fold dilutions of the sera. Therefore, logarithms to the base "2" were used in these situations. Titer readings recorded as logarithms to the base "2" may readily be converted back to titer readings by applying the equation $\log_{10} X = 0.30103 \log_2 X$ and finding the antilogarithms in tables of logarithms to the base "10."

Reported titers of 1:4 and 1:8 were combined for computation of some geometric means in these studies. This combined category was given a value of one logarithmic (base 2) unit above the < 1:4 category and one logarithmic (base 2) unit below the 1:16 category. Consequently, the geometric means presented are lower than they would have been if the 1:4 and 1:8 categories had been entered separately into the computations.

Table 180
VACCINATION STATUS OF STUDY POPULATION
AS DETERMINED FROM THE TOTAL TABULATION
AND FROM THE SAMPLE TABULATION
PLACEBO AND OBSERVED AREAS

Vaccination Status	Study Population		Sample	
	Number	Percent	Number	Percent
Placebo Areas	749,236	(40.94)*	1,039	(40.83)*
Vaccinated	200,745	26.79	270	25.99
Placebo	201,229	26.86	278	26.76
Partial Vaccinates	8,484	1.13	11	1.06
Partial Placebo	8,577	1.14	6	0.58
Not Inoculated	330,201	44.07	474	45.62
Observed Areas	1,080,680	(59.06)*	1,506	(59.17)*
Vaccinated	221,998	20.55	300	19.92
Control	725,173	67.10	995	66.07
Partial Vaccinates	9,904	0.92	14	0.93
2nd Grade Not Inoculated	123,605	11.44	197	13.08
Total	1,829,916		2,545	

* Numbers in parentheses are percentages of grand total; others are percentages of area totals.

