An Analysis of the Wilson-Worcester Method for Determining the Median Effective Dose of Pertussis Vaccine*

MARGARET PITTMAN, PH.D., AND J. E. LIEBERMAN, M.S.

Biologics Control Laboratory, National Institute of Health, and Public Health Methods Division, U. S. Public Health Service, Washington, D. C.

IN a study leading to the development of the "intracerebral" mouse protection test, proposed by Norton [1] as a means of evaluating the potency of *Hemophilus pertussis* vaccine, a statistical method was needed for analyzing the results obtained within a single experiment and between different experiments. Therefore a study of the techniques now in use was undertaken in order to select the one that would be best suited for interpretation of such data. The ultimate purpose, however, was the selection of a statistical method which could be recommended for use in the assay of pertussis vaccine.

It was considered essential that the method would provide a statistic which would be in conformity with those criteria that are recognized as being essential for a bioassay analysis. These have been stated or implied by both Gaddum [2] and Bliss,[3] namely: (1) a clear statement of the theoretical formulation of the estimate, (2) a statement of its exact standard error, (3) a minimum sample error of the statistic, and (4) a measure of assurance that the estimated relation between two biologic products at the ED 50 will hold throughout the range of dosage. In addition it was deemed necessary to include three other criteria, namely (1) that different computers using the method would obtain the same statistic, (2) that the statistic would be in agreement with one calculated by any other standard efficient method, and (3) that its computation could be done with ease and speed.

The statement of error is considered of particular importance. In the pertussis vaccine experiments, we are dealing with relatively small samples of animals. Therefore in order to evaluate with any precision, the response due to the vaccine, it is essential that there be some expression of the experimental or sampling errors. This need in biological assays has been realized by many workers; particular attention was directed to it by the British Pharmacopoeia Committee in 1932.[4] Nevertheless there has been deplorable lack of its use.

The methods that have been most widely used in the United States are the Probit method of Gaddum [2] and Bliss [5] and the so-called Reed-Muench method.[6] (It is of interest to note that the latter method, although independently developed, is essentially the same as the Behrens [7] method which had been in use in Europe prior to its publication by Reed and Muench.) The Probit method has proved its value in bioassay and it is generally considered the best to use in exact work, but it has the disad-

* Presented at a Joint Session of the American Statistical Association and the Vital Statistics and Laboratory Sections of the American Public Health Association, at the Seventy-fifth Annual Meeting in Atlantic City, N. J., October 8, 1947.

vantage that the computations are complicated and not too easily comprehended by many workers in problems of bioassay. The simplicity of the Reed-Muench method has contributed to its popularity. It, however, does not provide an exact statement of error and lacks other information that is desirable.

Other methods or modification of the Probit method have been proposed for bioassay problems by a number of workers, among whom are Karber,[8] Litchfield and Fertig,[9] Berkson,[10] Wilson and Worcester,[11] Knudson and Curtis,[12] and Thompson.[13] Each, with the exception of the Wilson-Worcester method, is either not applicable to the set-up of the pertussis test or fails in one or more phases to meet the criteria previously stated.

The Wilson-Worcester method, rather recently proposed, uses the logistic in place of the integrated normal curve for obtaining the probability of an observed response in a sample. Explicit and not very complicated formulae are given for the determination of the ED 50 point and its standard error and also for determining whether the ratio of two assays at the ED 50 point will hold, within the sampling error, throughout the range of dosage.

In this paper we shall attempt to show that the Wilson-Worcester method furnishes a statistic for the analysis of the results of the pertussis test which is in agreement with the one obtained by the Probit method. In addition, the statistics obtained by the Reed-Muench method will be given. This method has been used rather widely in bioassay problems and it was the first that was given consideration in the present problem. It will be shown that the Reed-Muench estimate is not always in complete conformity with the one obtained by the Probit method.

Technics—

Only a brief description of the mouse protection test and the Wilson-Worcester method will be given.

In the animal test, three doses of vaccine varying 4 or 5 fold are used to immunize mice in groups of 15 to 20 each. Their immune response is determined by survival following intracerebral injection of an amount of culture approximately 250 times the median lethal dose.

The ED 50 estimate by the Wilson-Worcester method for a three dose test is obtained by making two combinations of the percentages of survivals, which are used to make entrance into a double entry table provided by these authors.[11b] A factor is extracted and multiplied by the logarithm of the dilution increment. This product is either added to or subtracted from the logarithm of the middle dilution to give the median effective dose. The table also gives the value of that parameter which specifies the slope of the curve at its 50 per cent point, to the end that the experimenter may determine whether the accord between the slopes is sufficient to justify comparison at the ED 50 point alone. The quantities proportional to the standard errors for both the ED 50 and the homogeneity factors are also provided in the table.

Comparison of the median effective doses calculated by the Probit method with those calculated by the Wilson-Worcester method and by the Reed-Muench method—

The results from about 60 tests have been analyzed by the three methods. A comparison of the ED 50 values obtained by the Wilson-Worcester method with those obtained by the Probit method for 49 tests is presented graphically in Figure 1. The two estimated values for a single vaccine are plotted as one point, using the Wilson-Worcester ED 50 as the ordinate and the Probit ED 50 as the abscissa. It is evident that there is very close agree-

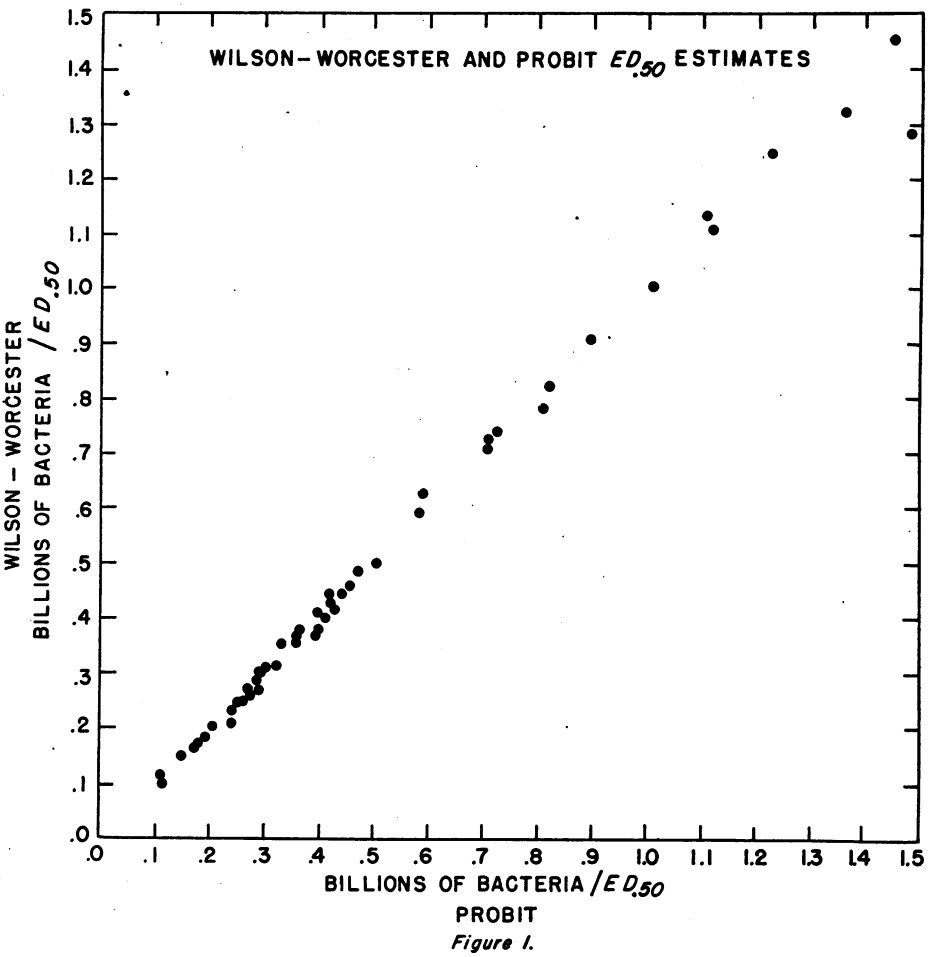

Figure 1.

ment. From the ED 50 obtained by the Wilson-Worcester method, the ED 50 by the Probit method can be predicted with a fair degree of accuracy and vice versa. It may be seen that if a line were fitted to these points, it would pass through the origin. This demonstrates that a predicting equation would not contain an arbitrary constant which would have to be added to the Wilson-Worcester estimate in order to predict the Probit estimate. Therefore the Wilson-Worcester estimates are directly proportional to those of the Probit method and there would be no "bias" term in a predicting equation.

In Figure 2, there is shown a graph which was similarly prepared by plotting the Reed-Muench ED 50 estimate and the Probit ED 50 for the same 49 vaccines. In contrast, however, there is lack of complete agreement which is evidenced by the scattering of the points. In this case, a prediction of the ED 50 value of one method from that of the other method would be less successful. Furthermore, it is doubtful whether one line would be sufficient to fit all points. It appears that a separate line should be fitted to the highest values and another to the next lower values with possibly a third to the lowest values. A line fitted to the upper values would certainly intercept the X-axis.

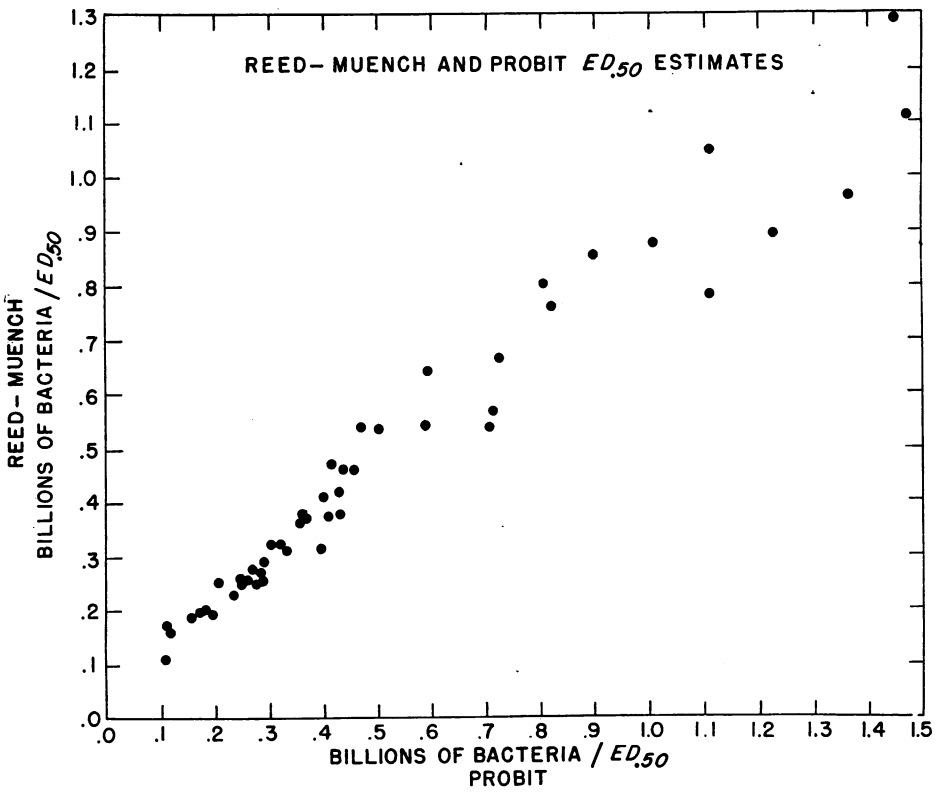

Figure 2.

Thus a predicting equation for these values would contain a constant. In other words, the ED 50 values estimated by the Reed-Muench method may be biased relative to those obtained by the Probit method, with the greater incidence of disagreement being evident when the ED 50 value is not close to the middle dilution. This lack of conformity was to be expected, for it had been shown previously by Irwin,[14] that under certain conditions of spacing and number of doses employed, a bias is introduced in the estimation by the Reed-Muench or similar methods.

Elimination of inconclusive experiments—

Among 58 tests, 49 of which were used in compilation of Figures 1 and 2, there were 8, which on the basis of their standard error established by the Probit method, were interpreted as being inconclusive. The results of these differed from the accepted ones in that either the percentages of survivals at all three dilutions were almost the same, or the percentages, although varying somewhat, did not reach sufficiently close to the 50 per cent point to permit an estimate of the ED 50 within the range of dilutions tested. In the first case a line fitted to the percentages against the doses would be practically parallel to the dose axis and in the second case the ED 50 estimate could be obtained only by extrapolation. The rejection of these tests by the Probit method came only after involved calculations.

Similarly, when the Wilson-Worcester method was applied, the same 8 tests were rejected; one other was also re-

jected. In contrast, however, the rejection came very early in the calculations. The table contains no value for entry if the difference between the proportion survivals at the first and third dilutions is less than 0.30. The authors omitted values below this point because the standard error increases very rapidly in those tests where three dilutions and small samples of animals are employed. With this omission in the table, a test is automatically rejected as being inconclusive at the beginning of the analysis and the computer is spared the time-consuming calculations required by the Probit method before rejection.

In accepting the difference of 0.30 as the critical difference between the highest and lowest response, there are the possibilities of accepting experiments which are inconclusive and rejecting experiments which are conclusive. However, until further information is available, this arbitrary criterion can be accepted with the assurance that both types of errors will be kept to a minimum.

The rejection of a test by the Reed-Muench method, in case of equal numbers of survivals at different dosages must be done subjectively as there is no exact formula for establishing the standard error. Too often a 50 per cent point has been derived by accumulation of survivals and deaths without taking into account that an equal response had been obtained by dilutions of agents that varied many times, although Reed and Muench warned against blind use of the method. In the other case where the 50 per cent point is outside the range of dosage the Reed-Muench method is also not applicable.

An illustration of advantages obtained by comparison over the entire range of dosage—

In one experiment, two vaccines were being compared. They are designated as 3M and 3H in Table 1. By the

TABLE I
NON-HOMOGENEITY OF TWO PERTUSSIS VACCINES

Vaccine		Survivals	$ED_{.50}$		Limits of two standard errors	Homogeneity constant α
No.	Dose	Total	Reed-Muench	Wilson-Worcester		
	(million)		(million)	(million)	(percent)	
3M	1500	$\frac{9}{15}$			35	
	300	$\frac{6}{16}$	582	765	and	0.79
	60	$\frac{2}{15}$			289	
3H	1500	$\frac{12}{15}$			60	1.95 ± 0.54
	300	$\frac{4}{15}$	618	612	and	limits for 2σ
	60	$\frac{0}{15}$			168	3.03 and 0.87

Reed-Muench method the ED 50 estimates were 582 and 618 million bacteria respectively. From these figures alone it might have been interpreted that 3M was slightly better in protective activity than 3H. In contrast, by the Wilson-Worcester method, the respective ED 50 estimates were 765 and 612 million bacteria with limits of two standard errors expressed in percentages, of approximately 35 and 290, and 60 and 168 respectively. It is clear that the standard error of the ED 50 for 3M is rather large and it may be suspected that the factors causing this error are not entirely due to the variability of the test animals. Furthermore, the ratio of the two vaccines at the 50 per cent point was not valid at other points. This lack of parallelism between the action of the two vaccines is shown in the last column of the table. If a comparison is to remain valid over the whole range, the values of a which specifies the slopes of the curves must be near enough alike so that the difference between them may be ignored. In this case the value of a for 3H is within the limits of 3.03 and 0.87 unless a one in twenty chance is operating. But the value of a for 3M, which is 0.79, falls outside these limits. It therefore is indicated that the two vacancies are not homogenous in capacity to induce a like response in mice. One explanation for this difference is that 3M contained some residual toxicity which interfered with induction of immunity by the largest dose.

SUMMARY

The results of a large number of "intracerebral" mouse protection tests of pertussis vaccines have been analyzed by the methods of Wilson and Worcester and of Reed and Muench in comparison with the Probit method. The Wilson-Worcester estimates of the ED 50 were as exact and consistent with the Probit estimates. The formulae of the former are simpler and the calculations less complicated and less time-consuming. Tests that were rejected by the Probit method as being inconclusive were likewise rejected by the Wilson-Worcester method.

The Reed-Muench estimates of the ED 50 were not entirely consistent with the Probit estimates. Furthermore, the calculation of the ED 50 by the Wilson-Worcester method does not take any longer than by the Reed-Muench method. With the expenditure of an additional few minutes, an exact estimate of a standard error and a test for validity of comparisons at points other than the ED 50 can be obtained by the Wilson-Worcester method but not by the Reed-Muench method. With this information the experimenter is better able to assess the meaning and importance of the results relevant to the agent, while making due allowance for the errors caused by disturbing influences.

An example of the informative nature of the Wilson-Worcester statistic in contrast to the comparative lack of information provided by the Reed-Muench statistic has been given.

This study suggests that the Wilson-Worcester method is adequate for use in the assay of the pertussis vaccine.

REFERENCES

1. Norton, John F. Personal Communication, Jan., 1944.

2. Gaddum, J. H. Reports on Biological Standards. *Med. Res. Council, Spec. Rep. Series,* No. 183, 1933.

3. Bliss, C. I. Biometry in the Service of Biological Assay. *Indust. & Engin. Chem.,* 13:84 (Feb.), 1941.

4. Gen. Med. Council. *Brit. Pharm. Comm. Rep. Comm.,* 9:56 and 60, 10:6 and 8, 1936.

5. Bliss, C. I. The Dosage-Mortality Curve. *Ann. Applied Biol.,* 22:134, 1935.

6. Reed, L. J., and Muench, H. A Simple Method of Estimating Fifty Per Cent Endpoints. *Am. J. Hyg.,* 27:493 (May), 1938.

7. Behrens, B. Zur Answertung der Digitalisblätter im Froschversuch. *Arch. F. exper. Path u. Pharmakol.,* 140:237, 1929.

8. Kärber, G. Beitrag zur kollektiven Behandlung pharmakologischer Reihenversuche. *Arch. exper. Path. u. Pharmakol.,* 162:480–483, 1931.

9. Litchfield, J. T., Jr., and Fertig, J. W. On a Graphic Solution of the Dosage-Effect Curve. *Bull. Johns Hopkins Hosp.,* 69:276–86, 1941.

10. Berkson, J. Application of the Logistic Function to Bio-Assay. *J. Am. Statist. A.*, 39:357 (Sept.), 1944.

11a. Wilson, E. B., and Worcester, J. The Determination of LD 50 and Its Sampling Error in Bio-Assay, I. *Proc. Nat. Acad. Sc.*, 29:79 (Feb.), 1943. II. 29:114 (Mar.), 1943.

————. Bio-Assay on a General Curve. *ibid.*, 29:150 (May), 1943.

————. The Determination of LD 50 and Its Sampling Error in Bio-Assay III. *ibid.*, 29:257 (Aug.), 1943.

11b. Worcester, Jane, and Wilson, E. B. A Table Determining LD 50 or the Fifty Per cent Endpoint. *ibid.*, 29:207 (July), 1943.

12. Knudsen, L. F., and Curtis, J. M. The Use of the Angular Transformation in Biological Assays. *J. Am. Statist. A.*, 42:282 (June), 1947.

13. Thompson, W. R. Use of Moving Averages and Interpolation to Estimate Median-Effective Dose. *Bact. Rev.*, 11:115 (June), 1947.

14. Irwin, J. O. Statistical Method Applied to Biological Assays. *Suppl. J. Roy. Statist. Soc.*, 4:1, 1937.

Recommendations for Control of Rabies

Early in 1947 a subcommittee of the New York Academy of Medicine prepared a report, with recommendations, on the whole problem of rabies control. This report is based upon a questionnaire sent to state health departments and Canadian provincial health authorities, and on information secured from various governmental departments such as the U. S. Public Health Service and from other agencies such as the National Research Council and the Rockefeller Foundation which has carried on a nine year research project in Alabama at the request and with the coöperation of that State Health Department.

The report indicates a recent increase in rabies among both man and animals, dogs, cattle, and cats in that order being the chief sufferers. It analyzes the legal situation and control measures in a number of states and reports on recent scientific developments.

Among the recommendations are:

1. Uniform national control measures.
2. Transportation of dogs and other susceptible animals regulated by federal and state coöperation, with vaccination as a prerequisite for entry into a state.
3. Rabies in man or animals should be a reportable disease and the figures published in *Public Health Reports.*
4. Annual licensing and vaccination of dogs; in urban areas licensing should be contingent upon vaccination.
5. Standard quarantine measures should be promulgated by either or both the U. S. Public Health Service and Bureau of Animal Industry.
6. National Research Council recommendations on vaccination should be generally accepted until further knowledge is available.
7. Every unit of local health jurisdiction should have a dog pound.
8. An educational program by health authorities should explain the necessity for control measures and the efficacy of present vaccines.
9. The U. S. Public Health Service should formulate and administer health provisions governing the importation of dogs and other susceptible animals.

The report includes a list of references and appendices showing the types of rabies legislation and the enforcement agencies of the various states.

The full report is on file with the New York Academy of Medicine. An abridged version of it appeared in *Public Health Reports,* Vol. 62, No. 34 (August 22), 1947.