THE RELATIVE SIGNIFICANCE OF GRADED IMMUNIZING AND CHALLENGE DOSES IN MEASURING THE POTENCY OF VACCINES

A STUDY OF MOUSE PROTECTION BY TYPHOID VACCINE

BY H. C. BATSON, PH.D.

(*From the Department of Biologic Products, Army Medical Department Research and Graduate School, Army Medical Center, Washington, D. C.*)

Many of the difficulties encountered in measuring, quantitatively, the degree of resistance (immune response) developed in an individual animal or in groups of animals in response to an antigenic stimulus have been presented and discussed at considerable length by Wilson and Miles (1). While the nature of these difficulties has not been clearly understood, their existence has been common knowledge to laboratory workers faced with the practical problem of evaluating the potency of an antigenic substance such as typhoid vaccine. In an attempt to develop an assay procedure of greater reliability, a comprehensive study of the sources of variation in potency tests of typhoid vaccine has been undertaken by the Department of Biologic Products, Army Medical Department Research and Graduate School. This report of the relative significance of graded immunizing and challenge doses in mouse-protective potency tests is the first of a series of reports of the results of the study.

Until relatively recently the potency of typhoid vaccine, *in lieu* of a more meaningful method, was expressed in terms of its ability to produce agglutinins in rabbits following parenteral administration of the product. The validity of results obtained in this manner has been questioned repeatedly. The agglutinin response as measured was primarily of the anti-H type, and various workers (2, 3) early reported that there was essentially no correlation between the ability of a typhoid vaccine to protect experimental animals and its content of flagellar antigen.

As a consequence of the above, recent emphasis has been placed on the development of assay methods based directly on the ability of typhoid vaccines to protect animals against lethal doses of the test organism, *Salmonella typhosa*. Such a test, using mucin to enhance the invasiveness of challenge cultures (4) and mice as test animals, was developed by Siler and his coworkers (5) as an aid to the selection of strains of superior immunogenic potency for vaccine production. Shortly afterward, the National Institute of Health (6) adopted as an official potency test for typhoid vaccine an active-immunity mouse-protection test. More recently, Griffitts (7) has proposed a similar procedure which he

considered minimized certain sources of experimental variation, and Luippold (8) has proposed a procedure which incorporates an expression of the protective potency in terms of an absolute unit, the T.I.U. (Typhoid Immunogenic Unit).

The above mentioned mouse-protection tests developed for estimating the potency of typhoid vaccine have been based on the survival (or mortality) rates in groups of mice given either, (a) graded immunizing doses and a constant challenge dose or, (b) a constant immunizing dose and graded challenge doses. In either case the objective has been to obtain a graded degree of response (survival or death) in successive groups of mice which would permit the calculation of a numerical expression of protective potency. Of the aforementioned procedures, that proposed by Luippold (8) was based on graded immunizing doses and a constant challenge dose, while the others (5–7) were based upon a constant immunizing dose and graded challenge doses. In each case the selection of the factor to be varied, the immunizing dose or the challenge dose, apparently was decided arbitrarily.

The lack of agreement as to which factor should be varied and which should be held constant is not peculiar to the assay of typhoid vaccines. Comparable differences of opinion also exist in relation to the assay of cholera and pertussis vaccines. In most instances factual information regarding the accuracy and reproducibility of the results obtained by the two methods which would permit a logical selection of either method has been lacking. However, Kendrick et al. (9) recently reported that the use of graded immunizing doses and a constant challenge dose in the assay of pertussis vaccine yields results superior to those obtained by the reverse procedure.

It was the objective of the study reported here to determine, experimentally, the relative effect of graded immunizing and challenge doses on the degree of gradation in response in successive groups of actively immunized mice and to determine, if possible, which variable should logically be made the basis of an assay procedure.

Materials and Methods

Standard Typhoid Vaccine.—The vaccine used in this study was a lyophilized monovalent typhoid vaccine prepared from a suspension of *S. typhosa*, strain 58,[1] produced by the method routinely used by the Army Medical Department Research and Graduate School (10). Immediately before use the dried product was rehydrated with sterile distilled water and appropriate dilutions were prepared with sterile physiological saline so that the desired quantities were contained in 0.5 ml. volumes.

Test Mice.—White Swiss mice (Bagg strain) weighing 14 to 16 gm. were assigned to jars in groups of five at random except for sex. The assignment of groups of mice to blocks of the experiments, the location of the jars in the room, and the order of injection of mice by groups were all decided by random selection. The temperature of the mouse room was maintained at 81° ± 1°F.

---

[1] For a full description of the strains of *S. typhosa* used in production of vaccine and for challenge of mice *cf.* Siler *et al.* (5).

Challenge Suspensions.—Challenge suspensions containing approximately the desired number of viable organisms were prepared by serial dilutions from veal infusion broth cultures of *S. typhosa*, strain 63, incubated at 37°C. for 16 hours. Final dilutions were made in 5 per cent mucin, so that the desired challenge doses were contained in 0.5 ml. volumes. Immediately before use the challenge suspensions were diluted in sterile physiological saline and nine 1 ml. aliquots of suitable dilutions were plated to provide a check on the actual content of viable organisms. Challenge suspensions were all injected within 2 hours of preparation.

Immunization and Challenge Procedure.—Vaccine dilutions and challenge suspensions were both injected intraperitoneally, the interval between injections being 6 days. The period of observation following challenge was 72 hours.

Statistical Analysis.—The nature of the experiments reported here suggested that the analysis of variance (11) would be an appropriate method for testing the significance of the sources of variation studied. By this procedure it is possible to estimate the portions of the total variation in the data contributed by each known source and also the residual variation or experimental error.

Snedecor (12) has pointed out that percentages based on small numbers of individuals are not suitable for this type of analysis unless some transformation of the data is employed. Statistical analyses of the data using either the probit (13) or the angle (14) transformation have yielded essentially the same results. Since in many instances in this study dosage-mortality (or dosage-survival) curves have been calculated by the Bliss (15) method employing probits, this transformation has been adopted throughout.[2] In those instances where 0 or 100 per cent of the group died, probit values have been assigned arbitrarily as though one-fourth individual had died or survived as suggested by Bartlett (16).

In certain instances it has been considered desirable to express the standard errors of ED_{50} values in terms of per cent of the ED_{50}. In these cases the standard errors have been approximated by the method proposed by Miller and Tainter (17).

Selection of Immunizing- and Challenge-Dose Levels

It was deemed essential to this investigation that the range of the progressively increased immunizing and challenge doses employed should be such as to insure that statistically significant different degrees of response (mortality or survival) would be elicited.

Even a large difference in response in groups treated differently does not insure that the difference will be significant since marked differences in response in groups treated alike also are commonly observed. The proper evaluation of the significance of differences in response in groups treated differently is based on the ratio of this difference (between groups) to the difference in response in groups treated alike (within groups).

Precision in the estimation of differences between and within groups can be increased by making multiple observations. Thus it was considered that experimental efficiency would be gained by making multiple observations at each of a limited number of immunizing-dose, challenge-dose levels rather than by making fewer observations at each of a larger number of levels. This necessitated

[2] Tables for the transformation of percentages to probits have been provided by Bliss (15) and by Finney (19). Weiss (21) has prepared a chart which permits a direct transformation to probits without first calculating percentages.

careful selection of the dosage levels to be employed. Accordingly, preliminary experiments were conducted to determine the approximate fold increments of dosage required to insure significantly different responses with each variable.

Significantly Different Challenge Dose Levels.—

Twelve groups of five mice each, six groups of males and six of females, were immunized with 0.01 ml. of standard typhoid vaccine made up to 0.5 ml. with physiological saline. Six days later all mice were challenged with *S. typhosa*, strain 63, suspended in 5 per cent mucin. Challenge doses of approximately 100, 500, 1000, 5000, 10,000, and 50,000 viable organisms were given to each of one group of males and one group of females. The virulence of the challenge culture used was checked by injecting four groups of five normal mice each with the challenge suspension diluted in 5 per cent mucin so as to contain approximately 10 organisms per 0.5 ml. dose.

TABLE I a

Mortality of Mice Immunized with 0.01 ml. Vaccine and Challenged with Graded Doses of S. typhosa, Strain 63

(Deaths/Total)

Sex	Challenge dose (Viable organisms in 0.5 ml. of 5 per cent mucin)						Totals
	100	500	1000	5000	10,000	50,000	
Female........	2/5	1/5	3/5	3/5	4/5	4/5	17/30
Male..........	0/5	2/5	3/5	4/5	4/5	5/5	18/30
Totals.......	2/10	3/10	6/10	7/10	8/10	9/10	35/60

Non-Immunized Controls (Challenge Dose, 10 Organisms)
Deaths/Total

Female........	4/5	2/5
Male..........	2/5	2/5

TABLE I b

Probit Values Corresponding to Mortality Data Presented in Table I a

Sex	Challenge dose (Viable organisms)						Totals
	100	500	1000	5000	10,000	50,000	
Female........	4.75	4.16	5.25	5.25	5.84	5.84	31.09
Male..........	3.36*	4.75	5.25	5.84	5.84	6.64*	31.68
Totals.......	8.11	8.91	10.50	11.09	11.68	12.48	62.77
Means.......	4.06	4.46	5.25	5.54	5.84	6.24	5.23

* Assigned values, *cf.* Bartlett (16).

The deaths which occurred in each group within 72 hours and empirical probit

values corresponding to the percentage of deaths in each group are presented in Tables I a and I b.

The general significance of the differences in response to the different challenge dose levels was determined by means of analysis of variance as given in Table II.

As shown in Table II, the differences in response of groups of immunized mice to challenge doses ranging from 100 to 50,000 organisms were not sufficiently great to reach the 5 per cent level of significance, or, in other words, results such as these could have been expected to occur by chance alone more often than once in twenty trials.[3] It is difficult to accept that the incidences of death pro-

TABLE II

Analysis of Variance of Preliminary Graded Challenge-Dose Constant Immunizing-Dose Probits-Mortality Data (Table I b)

Source of variation	Degrees of freedom	Mean square	Variance ratio (F)
Between doses (CD)...............	5	1.3892	4.33‡
Between sexes (S)..................	1	0.0290	—
Error..............................	5	0.3210	
Total..............................	11		

‡ Not significant at 5 per cent level.

gressively increased from $\frac{2}{10}$ to $\frac{9}{10}$ are not significant. Actually the lack of significance is more apparent than real. The correlation of probits-mortality with log-challenge dose was highly significant ($r = 0.88$; $p < 1$ per cent) but the data were too meager and the challenge doses selected were insufficiently different in size for the demonstration of significance by analysis of variance. Also, the error mean square was relatively large and, therefore, a larger mean square for differences between mean group responses to the graded challenge doses would be required for a significant value of F, the variance ratio.

The nature of the error term in the statistical analysis of this experiment merits further consideration. This estimate of error is not true experimental error; *i.e.*, it is not an estimate of differences in response of comparable groups treated alike. Rather, it is an estimate of the discrepance in response of the two sexes to the various challenge doses. Such discrepance is commonly referred to

[3] In this and the succeeding tables presenting analysis of variance, tests of significance (variance ratios) are the ratios of the mean squares for the factors of interest to the appropriate error mean squares. Mean squares are obtained by dividing the sums of squares of deviations from the grand mean by the corresponding degrees of freedom. Tables of variance ratios required for various levels of significance such as 5 per cent, 1 per cent, and 0.1 per cent are readily available (12, 22). These levels of significance correspond to probabilities of chance occurrence of 0.05, 0.01, and 0.001, respectively.

as interaction and the latter term is employed throughout this report. Since both sexes of mice were used in this preliminary experiment, this interaction is the proper error term for testing the significance of the differences in response to the graded challenge doses.

The objective of this preliminary experiment was to estimate the least difference between mean group mortalities which would be significant and, in turn, to estimate the increases in challenge doses required to produce these differences in mortality. The least significant difference between mean group mortalities can be estimated by means of the equation

$$LS_{\bar{d}} = t_{0.05} \sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \tag{1}[4]$$

where: $LS_{\bar{d}}$ is the least significant difference between mean group mortalities; $t_{0.05}$ is the 5 per cent significance level value of "t" (11) with degrees of freedom equal to those of the error mean square, σ^2; and n_1 and n_2 are the number of observations (groups of mice) upon which are based the two group means being compared.

Applying this formula to the data obtained, the estimated least significant difference between means expressed in probits is

$$LS_{\bar{d}} = 2.57 \sqrt{0.3210} = 1.45 \text{ probits} \tag{2}$$

Thus if a particular challenge dose caused a mortality corresponding to probits 5 (50 per cent), the least significantly different probits-mortality would be 1.45 probits greater or less than 5. These significantly different probits-mortalities of 6.45 and 3.55 correspond to mortalities of 92.5 per cent and 7.5 per cent, respectively.

The differences in mortality required for significance, as calculated above, are so great that they obviously would be difficult of close achievement in practical experimentation since they approach the absolute limits of the range of possible results, 0 and 100 per cent.

The magnitude of differences required for significance appears even more dramatic when expressed in terms of the challenge doses estimated to effect

---

[4] This equation is derived from the equation used for testing the significance of differences between group means (12)

$$t = \frac{\bar{d}}{\sigma_{\bar{d}}}$$

where $\sigma^2_{\bar{d}}$, the variance of the difference of the means, is equal to the sum of the variances of the means

$$\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

these mortality rates. The dosage-mortality equation calculated from the combined data in Table I b by the method proposed by Bliss (15) is

$$\hat{Y} = 3.0580 + 0.6625\, X \tag{3}$$

where $\hat{Y}$ is the expected mortality in probits and X is the logarithm of the challenge dose. The challenge doses required to effect mortality rates of 50 per cent and rates significantly greater or less than 50 per cent can be estimated by Equations 2 and 3 to be as follows:—

Mortality	Estimated dose (organisms)
per cent	
7.5	6
50.0	854
92.5	132,000

The errors inherent in the above method of estimation are quite large and the predicted challenge doses are of value primarily in depicting the relative magnitude of such doses required to effect significantly different mortality rates.

It was foreseen that significant differences in mortality could be expected with a narrower range of challenge doses by testing these concomitantly in multiple groups of mice. Thus if each challenge dose was tested in four groups of mice immunized with 0.01 ml. of vaccine and the estimate of error remained the same as above (Table II), the calculated least significant difference between means would be 0.73 probits[5] and significantly different mortalities could be expected with consecutive challenge doses of approximately 63, 850, and 10,000 organisms. Thus under the latter conditions a twelve- to thirteenfold increase between successive challenge doses should be sufficient to insure significantly different mortality rates in immunized mice.

Significantly Different Immunizing Dose Levels.—

Determination of the approximate progressive increases in graded immunizing doses required to establish significantly different degrees of resistance of groups of mice to a constant challenge dose was approached in a manner similar to that described above for graded challenge doses.

Ten groups of five mice each, five groups of females and five groups of males, were immunized with progressively graded quantities of standard vaccine. Immunizing doses of 0.005, 0.01, 0.025, 0.05, and 0.1 ml., in each case made up to a constant volume of 0.5 ml. with physiological saline, were each given to one group of females and one group of males. Six days later all mice were challenged with approximately 1000 viable *S. typhosa*, strain 63, suspended in 0.5 ml. of 5 per cent mucin.

---

[5] The values of n_1 and n_2 in Equation 1 would each be 4 and the least significant difference between group means would be calculated as follows:

$$\mathrm{LS}_{\bar{d}} = t_{0.05} \sqrt{\sigma^2(\tfrac{1}{4} + \tfrac{1}{4})} = 0.73 \text{ probits}$$

The number of survivors in each group after 72 hours and empirical probit values corresponding to the percentage of survival in each group are presented in Tables III a and III b.

The analysis of variance of the probits-survival data is presented in Table IV.

TABLE III a

Survival of Mice Given Graded Doses of Vaccine and Challenged with 1000 S. typhosa, Strain 63, in 5 per cent Mucin

(Survival/Total)

Sex	Vaccine dose (ml. in 0.5 ml.)					Totals
	0.005	0.01	0.025	0.05	0.1	
Female.................	1/5	3/5	4/5	5/5	5/5	18/25
Male...................	1/5	1/5	3/5	4/5	5/5	14/25
Totals...............	2/10	4/10	7/10	9/10	10/10	32/50

Non-Immunized Controls (Challenge Dose, 10 Organisms)
Deaths/Total

Female........	2/5	1/5
Male..........	4/5	2/5

TABLE III b

Probit Values Corresponding to Survival Data Presented in Table III a

Sex	Vaccine dose (ml.)					Totals
	0.005	0.01	0.025	0.05	0.1	
Female.................	4.16	5.25	5.84	6.64*	6.64*	28.53
Male...................	4.16	4.16	5.25	5.84	6.64*	26.05
Totals...............	8.32	9.41	11.09	12.48	13.28	54.58
Means...............	4.16	4.70	5.55	6.24	6.64	5.46

* Assigned values, cf. Bartlett (16).

The differences in mean survival rates in the groups of mice given progressively increased doses of vaccine were highly significant, the odds against chance occurrence of these results being greater than 100 to 1. This high degree of significance was due partly to the actual differences in survival and, also, to the relatively small mean square for error. As was pointed out in the discussion of Table II, this again is not a true measure of error but is an interaction, in this case the interaction between immunizing dose and sex. The actual difference in

survival of the two sexes, although considerable, proved not to be statistically significant and as this difference was relatively consistent the interaction mean square was small.

The least significant difference between group means was calculated to be 0.96 probit. By inspection of the group means in Table III *b* it can be seen that the differences between the means of the groups given 0.005 and 0.025 ml. of vaccine and, also of those given 0.01 and 0.05 ml., exceed 0.96 probit. The difference in immunizing doses given these pairs of groups was fivefold in each case.

A better prediction of the relative size of immunizing doses required to effect significantly different survival rates can be made by means of the equation of the dosage-survival curve in the manner previously described in predicting significantly different challenge doses. For calculation of the dosage-survival

TABLE IV

Analysis of Variance of Preliminary Graded Immunizing-Dose Constant Challenge-Dose Probits-Survival Data (Table III b)

Source of variation	Degrees of freedom	Mean square	Variance ratio (*F*)
Vaccine dose (VD).................	4	2.1340	18.04**
Sex (S)............................	1	0.6151	5.19‡
Error.............................	4	0.1183	
Total...........................	9		

** Exceeds 1 per cent level of significance.

‡ Not significant at 5 per cent level.

curve the vaccine doses were first multiplied by 1000 to obviate the use of negative logarithms and the equation was calculated by the method of Bliss. The final prediction equation is as follows:—

$$\hat{Y} = 2.543 + 2.210\,X \qquad\qquad (4)$$

where $\hat{Y}$ is the expected survival in probits and X is the log of 1000 times the vaccine dose estimated to elicit this degree of protection. By the use of this equation the approximate doses of vaccine predicted to effect 50 per cent survival and survival rates significantly less or greater than 50 per cent are as follows:—

Survival	Estimated dose
per cent	*ml.*
17	0.005
50	0.013
83	0.035

Thus it was predicted that significantly different survival rates could be obtained using approximately threefold increments in successive doses of vaccine even though only two groups of mice, one group of males and one of females, were used at each immunizing dose level. This is in sharp contrast to what could be expected using graded challenge doses and a constant immunizing dose.

Direct Comparison of the Effect of Graded Immunizing and Challenge Doses

It was realized that in any biological experiment, the individuals or groups observed may be subject to variation from many unrecognized sources. To insure a valid comparison of the effects of graded immunizing doses and challenge doses, a factorial experiment was conducted which permitted the calculation of the effect of the two variables from the same groups of mice.

Since it was estimated from the preliminary experiments that approximately twelve- to thirteenfold progressive increases in challenge doses would be required for significantly different results at one immunizing dose level, it was expected that even greater fold increments might be required when applied at more than one immunizing dose level. Accordingly it was decided to use challenge doses increased by twentyfold increments. The doses selected were 50, 1000, and 20,000 organisms. In a similar manner it was decided to use immunizing doses progressively increased by fourfold increments. The doses of vaccine selected were 0.003125, 0.0125, and 0.05 ml. Finally, in order to obtain a high precision in estimating error, each possible combination of vaccine dose and challenge dose was tested in six groups of ten mice each,.three groups of females and three groups of males.

The experiment was conducted in three parts, at each time one group each of females and males being tested at each challenge-dose, vaccine-dose level. The results of the three experiments were found not to differ significantly and the combined results, in terms of actual survivors and probits-survival, are presented in Tables V a and V b.

The probits-survival data in Table V b were analyzed by analysis of variance.[6] The analysis is presented in Table VI.

The complete analysis of variance presented in Table VI confirms, in general, the conclusions drawn from the preliminary experiments. The error term used for testing the significance of the interactions is the true error term; $i.e.$, it is based on the differences in survival of comparable groups treated alike.

The only interaction found to be significant was that between the linear effect of challenge dose and sex. This indicates a lack of consistency in survival of the two sexes at the different challenge dose levels. This is readily confirmed by an inspection of the original data in Table V a, where it is shown that more males than females survived the challenge dose of 50 organisms, survivals were essen-

---

[6] At the suggestion of Dr. A. E. Brandt (23), the original data in terms of actual survivors (Table V a), were analyzed independently by the factorial χ^2 method (11, 24). The results were essentially identical with those obtained by analysis of variance of the data transformed to probits.

TABLE V *a*

Survival of Mice Given Graded Immunizing Doses of Standard Typhoid Vaccine and Graded Challenge Doses of S. typhosa, Strain 63, in 5 per cent Mucin

(Survivors/Total)

Challenge dose	Sex	Vaccine dose (ml. in 0.5 ml.)									Totals	
		0.003125			0.0125			0.05				
50	F	2/10	3/10	1/10	7/10	5/10	6/10	10/10	10/10	9/10	53/90	111/180
	M	4/10	3/10	4/10	5/10	8/10	7/10	8/10	10/10	9/10	58/90	
1000	F	1/10	1/10	2/10	2/10	4/10	7/10	7/10	7/10	9/10	40/90	81/180
	M	1/10	1/10	1/10	5/10	6/10	4/10	8/10	8/10	7/10	41/90	
20,000	F	2/10	0/10	2/10	4/10	5/10	3/10	7/10	9/10	7/10	39/90	64/180
	M	1/10	0/10	1/10	2/10	3/10	1/10	6/10	6/10	5/10	25/90	
Totals.......		30/180			84/180			142/180			256/540	

Non-Immunized Controls (Challenge Dose, 10 organisms)

Deaths/Total

F..........	4/5	5/5	3/5	2/5	2/5	4/5
M..........	3/5	1/5	4/5	1/5	2/5	2/5

TABLE V *b*

Probit Values Corresponding to Survival Data Presented in Table V a

Challenge dose	Sex	Vaccine dose (ml.)									Totals	Means
		0.003125			0.0125			0.05				
50	F	4.16	4.48	3.72	5.52	5.00	5.25	6.96*	6.96*	6.28	48.33	97.75	5.43
	M	4.75	4.48	4.75	5.00	5.84	5.52	5.84	6.96*	6.28	49.42		
1000	F	3.72	3.72	4.16	4.16	4.75	5.52	5.52	5.52	6.28	43.35	86.71	4.82
	M	3.72	3.72	3.72	5.00	5.25	4.75	5.84	5.84	5.52	43.36		
20,000	F	4.16	3.03*	4.16	4.75	5.00	4.48	5.52	6.28	5.52	42.90	81.23	4.51
	M	3.72	3.03*	3.72	4.16	4.48	3.72	5.25	5.25	5.00	38.33		
Totals...............		70.92			88.15			106.62			265.69	
Means...............		3.94			4.90			5.92				4.92

* Assigned values, *cf.* Bartlett (16).

tially equal at the challenge dose level of 1000 organisms, but more females than males survived the challenge dose of 20,000 organisms. Aside from any bio-

logical implications, this observation is of interest in that it makes difficult the proper interpretation of the effect of the graded challenge doses. If the mean square of differences between groups treated alike is used as the error term, the linear effect of the graded challenge doses is highly significant. However, since the interaction $CD_L \times S$ proved to be significant, we must, on the basis of available evidence, assume that such an interaction actually exists and thus this interaction mean square becomes the proper error term for testing the significance of the linear effect of the graded challenge doses. The practical

TABLE VI

Analysis of Variance of Graded Immunizing-Dose Graded Challenge-Dose Probits-Survival Data (Table V b)

Source of variation	Degrees of freedom	Mean square	Variance ratio (F)
Main effects			
Vaccine dose, linear (VD_L)........	1	35.4025	234.29***
Vaccine dose, quadratic (VD_Q).....	1	0.0142	
Challenge dose, linear (CD_L).......	1	7.5809	8.52‡
Challenge dose, quadratic (CD_Q) ...	1	0.2862	
Sex (S).........................	1	0.2230	
First order interactions			
$CD_L \times S$......................	1	0.8899	5.89*
All others.....................	7	None significant	
Higher order interactions	4	None significant	
Error	36	0.1511	
Total........................	53		

*** Exceeds 0.1 per cent level of significance.
‡ Not significant at 5 per cent level in respect to proper error term $(CD_L \times S)$.
* Exceeds 5 per cent level of significance.

interpretation of the above is that it is impossible, from these data, to draw general conclusions regarding the significance of the effect of graded challenge doses when the responses of both sexes of mice are considered.

The test of significance of the linear effect of graded immunizing doses was less difficult since none of the interactions involving vaccine doses was significant.

Although the differences in response to doses of vaccine progressively increased by fourfold increments were of far greater significance than were those to challenge doses increased by twentyfold increments, it is difficult to express the relative effectiveness of the two variables in absolute terms. This is es-

pecially true since the degree of the effect of graded challenge doses was found to be dependent upon the sex of the mice.

The effectiveness of the two individual variables employed concomitantly is best expressed by means of a multiple regression equation (18, 19) relating probits-survival to log-challenge dose and log-immunizing dose. Since the effect of graded challenge doses was found dependent upon sex, separate multiple
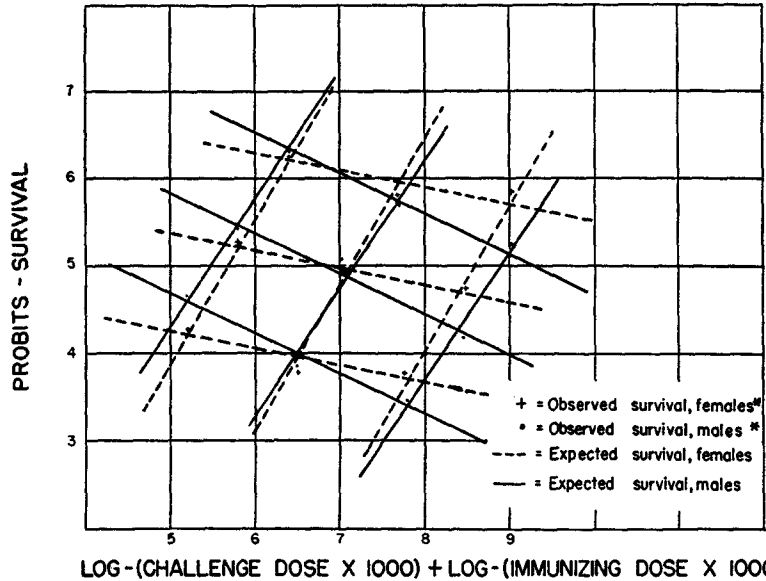

LOG-(CHALLENGE DOSE X 1000) + LOG-(IMMUNIZING DOSE X 1000)

FIG. 1. Probit-planes representing the relationship between expected probits-survival, log-(challenge dose × 1000), and log-(immunizing dose × 1000). Positive slope lines represent effect of increased immunizing doses at each challenge dose level. Negative slope lines represent effect of increased challenge doses at each immunizing dose level.

* Combined data, Table V a.

regression equations for each sex were derived from the probits-survival data presented in Table V b. The two equations

$$\text{Females: } \hat{Y} = 4.2855 - 0.1925\ X_1 + 1.6554\ X_2 \tag{5}$$

$$\text{Males: } \hat{Y} = 6.0331 - 0.4602\ X_1 + 1.4548\ X_2 \tag{6}$$

where $\hat{Y}$ = expected probits-survival, X_1 = log-(challenge dose × 1000), and X_2 = log-(immunizing dose × 1000), were derived using weighting coefficients and corrected probits. These equations represent probit-planes as presented graphically in Fig. 1.

The lack of parallelism evident upon visual inspection of the two probit-planes was confirmed by the appropriate χ^2 test (19). The departure from parallelism was found to exceed the 1 per cent level of significance.

The coefficients of regression of probits-response on log-challenge dose for females and males, -0.1925 ± 0.0845 and -0.4602 ± 0.0836, respectively, were found to differ significantly as was expected from the significant interaction $CD_L \times S$ (Table VI). The coefficients of regression of probits-response on log-immunizing dose for females and males, 1.4548 ± 0.1940 and 1.6554 ± 0.1886, respectively, did not differ significantly. The difference between the log-challenge dose and log-immunizing dose coefficients exceeded the 0.1 per cent level of significance.

It is obvious from a comparison of the respective regression coefficients that graded immunizing doses are more effective than graded challenge doses in effecting a dispersion of group responses. In Equation 5 the absolute value of the coefficient of X_2 is 8.6 times that of X_1. Thus a change of one unit in X_2 effects a change in $\hat{Y}$ equal to that effected by a change in X_1 of 8.6 units. As a practical illustration, by substitution of appropriate values of log-immunizing doses in Equation 5 it can be estimated that, with females, doubling the immunizing dose produces a difference in response equal to that produced by a 388-fold increase in the challenge dose. With males (Equation 6) the difference is not as striking. The coefficient of X_2 is 3.16 times that of X_1. By substitution of appropriate values in Equation 6 it is estimated that doubling the immunizing dose produces a difference in response equal to that produced by an 8.9-fold increase in challenge dose. Since estimation of these relative effects is based on the ratio of the regression coefficients, both of which are subject to considerable error, precision of the ratio is low.

The Dosage-Response Relationship

In the practical assay of typhoid vaccine by mouse protection tests a numerical expression of potency is desirable. This usually is expressed as an ED_{50}[7] value; the dosage estimated to effect response in 50 per cent of the group being tested. In practice it is assumed (or established) that the relationship between dosage and response is linear, and the dosage expected to effect a 50 per cent response is estimated by calculation or interpolation.

It has been established (15, 19) that the dosage-mortality relationship is essentially linear when percentage mortality is transformed to probits and doses are transformed to logarithms. Although standard errors of ED_{50} values are dependent primarily upon the slope of the dosage-response regression line (20), these estimated values would be subject to further error it there was a systematic departure from linearity of the dosage-response relationship.

In the analysis of variance presented in Table VI there was no evidence of any systematic departure from linearity in the relationship of probits-response to log-challenge dose and log-immunizing dose. The degree of departure from linearity with each variable is represented by the estimates of the quadratic

---

[7] The term LD_{50} frequently is employed when dealing with dosage-mortality data. In this presentation, the term ED_{50} is used throughout.

effects of challenge doses and immunizing doses, respectively. While the quadratic effect of challenge doses was greater than was that of immunizing doses, neither was significant.

The practical importance of the greater significance of graded immunizing doses than of graded challenge doses as the basis of estimating potency is more readily apparent from a consideration of the standard errors of ED_{50} values estimated from these data. As Bliss (20) has pointed out, these are standard errors of log-ED_{50} values and are properly used only in the calculation of confidence limits which will bracket the true value in a designated proportion of repeated trials. The lower and upper confidence limits differ unequally from the
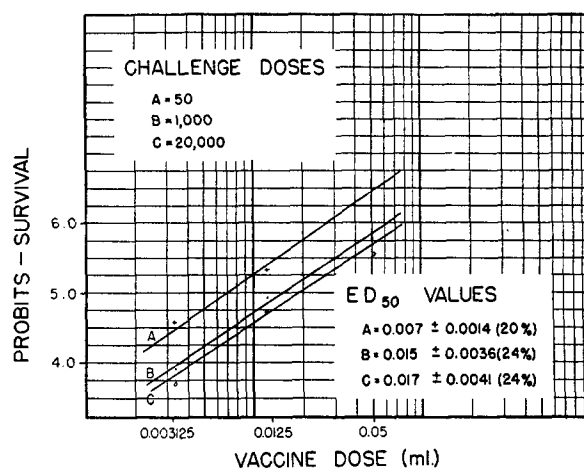
FIG. 2. Immunizing dose-survival relationships at three challenge dose levels.

estimated ED_{50} value, and thus it is difficult to compare directly the accuracy of ED_{50} values when estimated under different experimental conditions.

Miller and Tainter (17) have proposed an approximation to the standard error of an ED_{50} value which is essentially an average standard error. This average standard error can be expressed as a percentage of the ED_{50} and thus permits a direct comparison of the relative accuracy of ED_{50} values estimated under different experimental conditions.

Immunizing dose-survival curves calculated separately for each challenge dose level, ED_{50} values, and average standard errors expressed as percentages of ED_{50} values, are presented in Fig. 2.

The parallelism of the regression lines in Fig. 2 indicates that the immunizing dose-survival relationship qualitatively is the same at each of the challenge dose levels tested. Furthermore, it is apparent that ED_{50} values can be estimated from the immunizing dose-survival curve with comparable accuracy at any of the challenge dose levels. This is of importance in the routine potency testing

of typhoid vaccine, since any toxic effects due to large doses of either vaccine or challenge organisms can be avoided to a considerable degree without sacrificing precision of the assay procedure by using small immunizing doses and a correspondingly low challenge dose. Furthermore, it would appear possible to assay quite small quantities of antigenic material such as frequently is desirable in determining the distribution of such substances in fractionation and purification studies.

Challenge dose-mortality curves calculated separately for each immunizing dose level, ED_{50} values, and average standard errors expressed as percentage of the ED_{50} values, where applicable, are presented in Fig. 3.
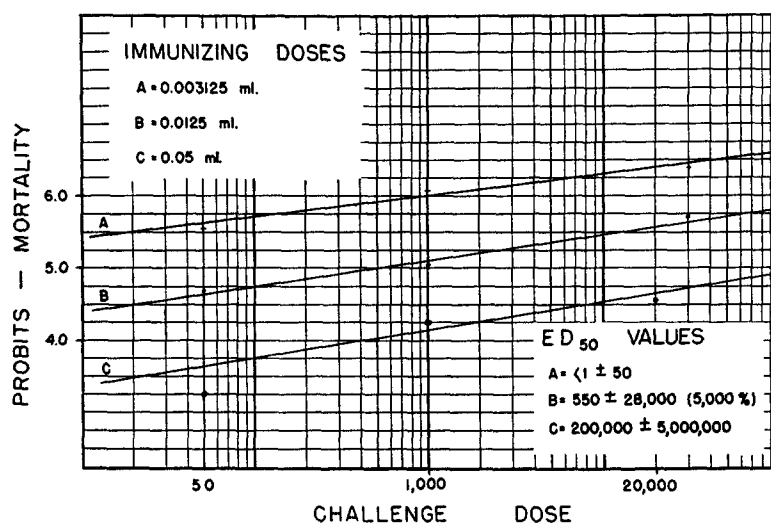

FIG. 3. Challenge dose-mortality relationships at three immunizing dose levels.

Only the challenge dose-mortality curve at the immunizing dose level of 0.0125 ml. is strictly suitable for estimation of an ED_{50} value, since ED_{50} values at the other immunizing dose levels can be estimated only by extrapolation. It is obvious, however, that the average standard errors of ED_{50} values estimated from challenge dose-mortality curves are so large that such values are relatively meaningless. Also, it appears that the average standard errors expressed as percentages of ED_{50} values may differ markedly at the different immunizing dose levels.

DISCUSSION

The data presented in this report reveal the marked variation in response observed in groups of mice even when they are subjected to the same experimental conditions of immunization aand challenge. The existence of this variation, or

experimental error, makes difficult the determination of the significance of qualitative and quantitative differences in response to different factors or different levels of the same factor. Even marked differences in response between groups treated differently, such as in groups given different doses, fail to be statistically significant when there is a comparable degree of variation in the response of groups treated alike.

Statistical significance in experiments such as those reported here is not attained merely by reduction of experimental error through careful preparation of materials, selection of homogeneous test animals, and application of meticulous techniques. The source of a major portion of the variation is unknown and, accordingly, is not subject to direct control. Thus it becomes essential to insure that such variation is of random nature and, further, to employ valid methods for estimating this random variation independently of the effect of any of the experimental factors being studied. Failure to make these provisions inevitably results in the introduction of bias and the final conclusions may be subject to serious error.

In these experiments, random distribution of variation of unknown origin was provided for, at least in part, by the methods of selecting and grouping of mice, location of groups in the test-animal room, and by following a random order of injection. In the main, except for minor restrictions such as segregation of the sexes, there was equal opportunity for any especially resistant or susceptible mouse being assigned to any of the test groups. If any environmental or psychological effect resulted from the particular location of a group of mice in the test-animal room, all groups had equal opportunity of being assigned the location where exposure to the factor existed. Likewise, if there was any tendency for the challenge suspensions to gain or lose virulence during the injection period, it was a matter of chance which groups were favored or adversely affected.

Provision of randomness does not necessarily insure that experimental error (variation within groups) will be decreased. In fact, it even may be greater than that obtained following a non-random procedure since the latter may result in the establishment of intragroup correlation. However, the random procedure does provide a basis for obtaining unbiased estimates of experimental error and thus the performance of significance tests is justifiable.

The true significance of the effect of the sex of mice on their response to either graded immunizing doses or challenge doses is difficult to determine from the data obtained in this study. It was found in the preliminary experiment with graded challenge doses that specifically immunized mice of both sexes were essentially equally susceptible to S. typhosa, although possibly the females were slightly more susceptible to the smallest challenge dose (Table I a). It has been observed repeatedly in other experiments that in unprotected mice, females are slightly more susceptible than males. This tendency can be noted by comparing

the deaths of unprotected mice of both sexes to a challenge dose of approximately 10 organisms as shown by the controls presented in Tables I a, III a, and V a. When these controls are combined it is seen that a total of 29 out of 50 females died as compared with 23 out of 50 males. This difference is so small and the obvious potential error in administering challenge doses of this size is so great, that no significance could be attached to the difference if it were not observed repeatedly.

In the preliminary experiment using graded immunizing doses and a constant challenge dose (Table III a) there was a definite indication that females responded better to immunization than did males. This difference was not statistically significant, primarily as a result of the small number of groups employed, but the tendency is evident.

In the direct comparison of graded immunizing and challenge doses (Table V a), neither the main effect of sex nor the interaction between sex and immunizing dose was significant, while the interaction between sex and challenge dose was significant at the 5 per cent level (Table VI). The explanation of these findings may be indicated from the results of the preliminary experiments as discussed above. The superior resistance of males to the challenge dose of 50 organisms may be due to a lower native susceptibility. The superior resistance of females to the challenge dose of 20,000 organisms may be the result of an actual greater response to immunization. This reversal of resistance at the different challenge dose levels would account for the significant interaction. If this explanation is correct, the significant interaction resulted primarily because of the particular range of immunizing and challenge doses employed and might not be observed if larger doses were used. Additional preliminary experiments have been conducted using larger immunizing and challenge doses and, while the data so far obtained are insufficient for adequate evaluation, a highly significant interaction between sex and challenge dose has not been observed. Further investigation on a more extensive scale of the existence of this interaction is contemplated and will be made the basis of a subsequent report. The mere fact that an interaction is statistically significant is no guarantee that it has any particular biological significance. As used here the term interaction is synonymous with discrepance, and statistical significance may result merely from inconsistent responses to the various levels of one factor at all levels of the second factor. The error of estimation of probable response at any challenge dose level is exceedingly large (Fig. 3), and it is entirely possible that the discrepance observed is of an anomalous nature. Regardless of the true nature of the interaction, it appears justifiable to conclude that the use of mixed sexes should be avoided; certainly whenever graded challenge doses are used to measure the degree of immune response.

The superiority of graded immunizing doses over graded challenge doses as a basis for measuring the degree of response of mice to active immunization

with typhoid vaccine has been discussed in detail previously. This superiority resulted primarily from (a) greater homogeneity in the response of groups of mice given the same immunizing dose than in groups given the same challenge dose, (b) a greater slope of the dosage-response curve and corresponding smaller standard errors of estimated ED_{50} values, and (c) an effect apparently less dependent on the sex of mice employed.

At each immunizing dose level, the response (survival) of mice was surprisingly independent of the size of challenge dose used. This suggests that effective resistance in the individual animal is the result of attainment of some threshold level of immunity and is not primarily a relative state. This observation further suggests that the greater protective effectiveness of increased immunizing doses is the result of stimulation of a greater proportion of the group to respond sufficiently to attain this threshold level.

The use of average approximate standard errors of ED_{50} values estimated by the Miller-Tainter method (17) is somewhat misleading. This method was selected because the errors could be expressed as percentages of the corresponding ED_{50} values and thus permitted a rough comparison of the accuracy of estimation of ED_{50} values under different experimental conditions. As Bliss (20) has pointed out these are standard errors of log-ED_{50} values and are properly used only in the estimation of confidence limits which differ unequally from the ED_{50} values they bracket. Confidence limits based on the average standard errors approximated by the Miller-Tainter method can be calculated as $ED_{50} \pm 1.96 \times S.E._{ED_{50}}$ and differ considerably from the confidence limits calculated by the Bliss method. For example; the confidence limits of the immunizing dose ED_{50} for the combined sexes challenged with 1000 organisms (Table V a) calculated by the Miller-Tainter method are 0.0083 ml. and 0.0217 ml. These limits calculated by the Bliss method are 0.0098 ml. and 0.0266 ml. In either case the breadth of these limits clearly emphasizes the potential inaccuracy of estimated ED_{50} values even when such estimates are based on data more extensive than commonly are available in the assay of immunogenic substances.

Although directly comparable data are not available, strongly suggestive evidence exists that the use of graded immunizing doses would be equally superior to graded challenge doses in the assay of other antigenic substances. Thus, Sokhey (25), in an extensive study of the factors affecting the potency of plague vaccine, found that by employing an assay procedure based on graded immunizing doses and a constant challenge dose, it was possible to detect differences in the potency of products prepared from strains of different sources or subjected to different processing conditions. As mentioned previously, Kendrick et al. (9) reported that assays of pertussis vaccine were most reliable and reproducible when the graded immunizing doses method was employed.

A similar advantage of the use of graded immunizing doses methods has been observed in passive immunity experiments. Smith (26) found marked superi-

ority of this method in determining the mouse protective potency of anti-pneumococcal sera. Batson (27), in a direct comparison of the effects of graded serum doses and challenge doses in the titration of the mouse protective potencies of typhoid antisera, obtained similar results. In fact, when two serum solutions were prepared by dilution so that one contained twice as much original serum as the other, the experimentally determined relative potency of the two was 2.96 by the graded serum dose method and 27.5 by the graded challenge dose method.

Finally, there is evidence that the superiority of the graded immunizing doses method holds for the biological assay of bacterial toxoids as well. Greenberg *et al.* (28) obtained quite reproducible results in the assay of tetanus toxoids by such a procedure and reported they consistently were able to demonstrate differences in potency of toxoids from different manufacturers.

SUMMARY

A study of the effect of graded immunizing doses of typhoid vaccine and graded challenge doses of *S. typhosa* in 5 per cent mucin on the degree of gradation of response (survival or death) elicited in successive groups of mice is reported.

In the range of doses employed the effect of graded immunizing doses was markedly greater than was the effect of graded challenge doses. Statistically the difference exceeded the 0.1 per cent level of significance.

It was concluded that the use of graded immunizing doses was preferable to the use of graded challenge doses as a basis for the immunological assay of typhoid vaccine, since with the former there was obtained (*a*) a greater significance of differences in response of groups given graded doses even with smaller fold increments in successive doses, (*b*) a greater slope of the dosage-response curve which permitted estimation of ED_{50} values with smaller standard errors, and (*c*) an effect on response apparently less dependent on the sex of mice used.

BIBLIOGRAPHY

1. Wilson, G. S., and Miles, A. A., Topley and Wilson's Principles of Bacteriology and Immunity, Baltimore, The Williams & Wilkins Co., 3rd edition, 1946, 2.
2. Arkwright, J. A., *J. Path. and Bact.*, 1927, **30**, 345.
3. Perry, H. M., Findlay, H. T., and Bensted, H. J., *J. Roy. Army Med. Corps*, 1934, **62**, 161.
4. Rake, G., *Proc. Soc. Exp. Biol. and Med.*, 1935, **32**, 1523.
5. Siler, J. F., *et al.*, Immunization to Typhoid Fever, Baltimore, Johns Hopkins Press, 1941.
6. National Institute of Health, Minimum Requirements: Typhoid Vaccine, Bethesda, 1942.
7. Griffitts, J. J., *Pub. Health Rep.*, *U. S. P. H. S.*, 1944, **59**, 1515.

8. Luippold, G. F., *Am. J. Pub. Health*, 1945, **35,** 153.
9. Kendrick, P. L., Eldering, G., Dixon, M. K., and Misner, J., *Am. J. Pub. Health*, 1947, **37,** 803.
10. Holt, R. L., and Hitchens, A. P., *Pub. Health Rep., U. S. P. H. S.*, 1937, **52,** 829.
11. Fisher, R. A., Statistical Methods for Research Workers, Edinburgh, Oliver and Boyd, 10th edition, 1946.
12. Snedecor, G. W., Statistical Methods, Ames, Iowa State College Press, 4th edition, 1946.
13. Bliss, C. I., *Science*, 1934, **79,** 38.
14. Bliss, C. I., Plant Protection, Leningrad, 1937, No. 12, as quoted by Snedecor (12).
15. Bliss, C. I., *Quart. J. Pharm. and Pharmacol.*, 1938, **11,** 192.
16. Bartlett, M. S., *Supp. J. Roy. Statist. Soc.*, 1937, **4,** 137.
17. Miller, L. C., and Tainter, M. L., *Proc. Soc. Exp. Biol. and Med.*, 1944, **57,** 261.
18. Bliss, C. I., *Ann. Entomal. Soc. America*, 1940, **33,** 721.
19. Finney, D. J., Probit Analysis, Cambridge University Press, 1947.
20. Bliss, C. I., *Biometrics Bull.*, 1945, **1,** 57.
21. Weiss, E. S., *Am. J. Pub. Health*, 1948, **38,** 22.
22. Brownlee, K. A., Industrial Experimentation, Brooklyn, Chemical Publishing Co., Inc., 1947.
23. Brandt, A. E., personal communication.
24. Mather, K., Statistical Analysis in Biology, New York, Interscience Publishers, Inc., 2nd edition, 1947.
25. Sokhey, S. S., *Rep. Haffkine Inst. 1932–1935*, Bombay, 1936.
26. Smith, W., *J. Path. and Bact.*, 1932, **35,** 509.
27. Batson, H. C., *Ann. New York Acad. Sc.*, in press.
28. Greenberg, L., Morrell, C. A., and Gibbard, J., *J. Immunol.*, 1943, **46,** 333.