THE MOUSE INFECTIVITY TITRATION OF INFLUENZA VIRUS*

By MAX A. LAUFFER, Ph.D., and GAIL LORENZ MILLER, Ph.D.

(From the Department of Animal and Plant Pathology of The Rockefeller Institute for Medical Research, Princeton, New Jersey)

(Received for publication, November 3, 1943)

In order to evaluate the statistical significance of an experimental result, it is necessary to have some quantitative estimate of the variability due wholly to errors of measurement. Even though determinations of influenza virus infectivity have long been made using the mouse as a test animal, no detailed analysis of the errors involved has heretofore been carried out. In this paper the results of experiments designed to determine the variation encountered in the measurement of mouse infectivity titers of influenza virus preparations are presented and discussed in detail.

Method of Determining Mouse Infectivity

Serial tenfold dilutions of the virus under study were prepared with sterilized 0.1 M phosphate buffer at pH 7. Six successive dilutions covering the range in which the end point was expected to occur were chosen for inoculation into mice. Doses of 1/20 cc. of each of the dilutions were introduced intranasally into five mice under light ether anesthesia. The mice were marked and kept in galvanized iron cages accommodating 15 each. In each experiment, which consisted of a series of titrations, the positions of the groups of five mice within the cages were mixed according to a previously designed code, in order to help to eliminate bias in the evaluation of the results at autopsy. The cages were searched daily for dead mice, and whenever possible autopsies were performed to ascertain the cause of death. At the end of 10 days, the surviving mice were sacrificed by heavy ether anesthesia and were autopsied. Lung pairs with no pulmonary consolidation were recorded as $-$, those with slight or doubtful consolidation were recorded as $\pm$, and those with 1/4, 1/2, and 3/4 consolidation, respectively, were recorded as $+$, $++$, and $+++$. Surviving mice with more than 3/4 lung consolidation were rarely found.

Fifty per cent end points were calculated by the method of Reed and Muench (1). Three different criteria were utilized: (*a*) the occurrence of death due to influenza within 10 days, (*b*) the development of pulmonary consolidation within 10 days, and (*c*) a weighted composite taking into account the occurrence of death and the extent of lung consolidation. In the case of (*a*), a positive group score for each dilution was obtained by dividing the number of mice dying by the number inoculated. In

*The work described in this paper was done under a contract, recommended by the Committee on Medical Research, between the Office of Scientific Research and Development and The Rockefeller Institute for Medical Research.

the case of (*b*), a similar positive group score was obtained for each dilution by dividing the sum of the mice dying plus those showing pulmonary involvement with a score of + or more by the number inoculated. Single mice showing scores of ± were considered negative, but when two mice in one group showed scores of ±, one was considered positive and the other negative. In the case of (*c*), mice which died of influenza within 10 days were each given a score of 4, and those which showed lung involvement upon 10-day autopsy of +++, ++, +, ±, and −, respectively, were given scores of 3, 2, 1, 0.5, and 0. The positive group score assigned to each dilution was then obtained by adding the scores of the individual mice in the test group and by dividing by the number of mice. This method is similar in principle to the "50 per cent maximum score end point" described by Horsfall (2). In all three cases, mice which died of causes other than influenza were eliminated from consideration in the computation of the positive group scores. The assigned positive group scores were employed as the bases for the calculations of the 50 per cent end points.

Reproducibility of Infectivity Titrations

In order to determine the error associated with the mouse titration of influenza virus, five tests were carried out in each of which five replicate titrations were made on a single batch of egg-adapted PR8 virus isolated by centrifugation from the allantoic fluid of infected chick embryos (3). Different preparations of virus were used for each test. The concentration of protein in each preparation was estimated roughly and dilutions were prepared on this basis. Separate dilutions were prepared for each replica within a given test. The titrations were carried out exactly as described in the preceding section. Three-week-old mice from the colony of the Department of Animal and Plant Pathology of the Institute were used throughout. The results are presented in Table I.

The end points computed on the bases of the three criteria described previously are shown in the last three columns on the right of Table I. A statistical study of the variation of the end points was carried out as follows: The variance, V, which is equal to the square of the standard deviation, and which is defined for small sample statistics as the sum of the squares of the deviations of the individual variates from the mean divided by one less than the number of variates, was calculated for each type of end point for each test. The results are listed in Table II. In computing these statistics for Test 1, Replica 2 was discarded, because the deviation of its end point from the rest is far too great to be attributed to random errors.[1] From the average variance for each particular type of end point, the standard deviation of the distribution of

[1] It can be seen from the data of Table I that the weighted end point for this replica differs by 1.45 units from the average of the rest of the test. As is shown later in the text, the most probable value of the standard deviation for this type of end point is 0.260 units. Thus, the discarded value differs by 5.5 standard deviation units from the mean of the other values in the test. From a normal frequency distribution table, one can find that there is less than one chance in ten million that a deviation of this order of magnitude would occur due to random errors.

TABLE I

Results of Replicate Titrations of Influenza Virus in Mice

Test No.	Replica No.	Concentration of inoculum in gm. of protein per cc.						50 per cent end points (negative exponents of 10)		
		10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-10}	10^{-11}	Weighted	Deaths	Lesions
1	1	3D,*1±,1M‡	3D,2+	2D,1+,1−,1M	3+,2−	5−	5−	7.7	7.3	8.7
	2	3+,1−,1M	1D,1+,1±,2−	1++,2+,1±,1−	1+,1±,3−	1+,4−	1+,4−	6.0	—	—
	3	5M	2+,2±,1−	1++,1+,1±,2−	3D,2+	1D,1+,3−	5−	7.5	6.9	9.2
	4	5D	3++,1+,1M	1D,2+,2±	1+,1±,3−	1+,4−	5−	7.5	6.6	8.6
	5	4D,1M	1D,4+	1+,3±,1−	5−	5−	5−	7.1	6.7	7.8
2	1	5D	4D,1±	2++,1+,1±,1M	3±,2−	5−	5−	7.7	7.4	8.4
	2	5D	3D,1++,1−	1++,1+,2±,1−	2±,3−	2+,3−	5−	7.5	7.2	8.5
	3	4D,1+	5D	1D,2+,2±	3+,1±,1−	1±,4−	1±,4−	7.8	7.5	9.0
	4	4D,1−	5D	2++,1+,2±	2±,3−	1±,4−	1±,4−	7.7	7.4	8.4
	5	5D	5D	1D,1++,3+	1D,3+,1±	2±,3−	5−	8.3	7.8	9.5
3	1	5D	4D,1M	5D	2+,2±,1−	1+,2±,2−	5−	8.6	8.5	9.5
	2	5D	4D,1+	3D,2+	1D,1++,1+,2±	1+,3±,1−	5−	8.6	8.2	9.6
	3	4D,1+	4D,1M	1D,2++,2+	2+,2±,1−	1+,1±,3−	2±,3−	8.1	7.5	9.5
	4	5D	5D	2D,3++	2+,3±	5−	5−	8.4	7.8	9.2
	5	5D	5D	3D,2+	2±,3−	2+,1±,2−	2+,1±,2−	8.5	8.2	9.3
4	1	5D	5D	1+,3±,1−	5−	2±,3−	5−	7.6	7.5	8.0
	2	5D	5D	2+,2±,1−	5−	5−	5−	7.6	7.5	8.2
	3	5D	4D,1+	1+,2±,2−	1±,4−	1+,4−	5−	7.4	7.4	8.0
	4	5D	5D	2+,3±	4±,1−	5−	5−	7.7	7.5	8.5
	5	5D	3D,1++,1+	2+,1±,2−	1D,1+,3±	5−	5−	7.3	7.3	8.5
5	1§	4D,1+	3D,1++,1−	1D,1+++,1++,2+	2+,1±,2−	3±,2−	5−	6.9§	6.2	7.8
	2	5D	4D,1+++	2D,1++,1+,1±	1+,1±,1−,2M	2+,3−	5−	7.5	6.9	8.1
	3	4D,1++	5D	5+	2±,3−	5−	5−	7.0	6.8	7.7
	4	5D	2D,1+++,1+,1M	2+,1±,2−	1+,4−	2±,3−	5−	7.1	6.6	7.8
	5	3D,1++,1−	5D	1++,2+,1±,1M	5−	5−	5−	7.3	7.0	8.0

* D = died of influenza. ‡ M = missing or dead of another cause. § In Test 5, Replicas 1 to 5 were applied at 10, 6.3, 4.0, 2.5, and 1.6, respectively, times the concentrations indicated. All of the end points were corrected for the differences in concentration.

individual end points was calculated. The standard deviations are listed in the last column of Table II. These standard deviations are in reality measures of the reproducibilities of the end point determinations. It is clear, therefore, from the above data, that the death and the weighted end points are more reproducible than the lesion end point.

Horsfall (2), using a method of titration similar to that described above, carried out a series of ten titrations on a single preparation of PR8 virus. He calculated 50 per cent mortality end points comparable to our death end points and 50 per cent maximum score end points similar to but not quite identical

TABLE II

Variances and Standard Deviations of End Points Obtained in Titrations of Influenza Virus in Mice

Test No.	Type end point	ΣX^2	$\frac{(\Sigma X)^2}{N}$	Difference	$\frac{\text{Diff.}}{N-1} = V$	σ
1	Weighted	222.200	222.010	0.190	0.063	
2	"	304.560	304.200	0.360	0.090	
3	"	356.340	356.168	0.172	0.043	0.237
4	"	282.860	282.752	0.108	0.027	
5	"	256.560	256.328	0.232	0.058	
1	Deaths	189.350	189.062	0.287	0.096	
2	"	278.450	278.258	0.192	0.048	
3	"	323.820	323.208	0.612	0.153	0.283
4	"	276.800	276.768	0.032	0.008	
5	"	224.850	224.450	0.400	0.100	
1	Lesions	295.130	294.122	1.007	0.336	
2	"	384.620	383.688	0.932	0.233	
3	"	443.790	443.682	0.108	0.027	0.356
4	"	339.740	339.488	0.252	0.063	
5	"	310.580	310.472	0.108	0.027	

with our weighted end points. From the data of Table I of his paper, we have calculated standard deviations of 0.262 and 0.302 log units for his mortality and maximum score end points. It is thus evident that the reproducibility which he obtained is in close agreement with that which we obtained. In his case, the mortality end point seemed to be more reproducible than the maximum score end point, whereas in our case the opposite was true. It is probable that neither difference is significant.

The reproducibility of an experimental method is a measure of the accuracy only if all errors are random—that is, only if there are no systematic errors. A possible systematic error could arise from choosing too wide an interval in the concentrations applied to the test animals. Thus, if a dilution interval

considerably greater than the interval between a region of practically complete response and one of practically complete lack of response were selected, the end point would appear to be midway between the two dilutions chosen, in spite of the fact that it really might be much nearer one extreme than the other. If a series of titrations were carried out on a given solution always at the same dilutions, this error would not be detected. In most virus titrations the concentration interval between the regions of virtually certain response and virtually certain lack of response is about $2\frac{1}{2}$ common logarithmic units. One would therefore not expect an interval of one logarithmic unit, such as was used in this investigation, to be great enough to cause a systematic error. It appeared worthwhile, nevertheless, to subject the possibility to an experimental test. In Test 5 summarized in Table I, each of the five replicate titrations was carried out on a different series of decimal dilutions. The series differed from each other by 0.2 logarithmic units. All of the end points were calculated on the basis of the concentrations assumed for the most concentrated series. If there were a systematic error due to the dilution interval being too great, one should expect the three variances for Test 5 to be greater than those for the other tests. It can be seen in Table II, however, that the variances for the three types of end points for this test do not differ appreciably from the mean variances of the other tests. Hence, these results bore out the expectation that there should not be any very considerable systematic error due to the magnitude of the dilution interval. On this basis, it seems reasonable, therefore, to regard the reproducibility of the mouse titration of PR8 virus as a measure of its accuracy.

Significance of Differences between End Points

These results may be used to establish a basis for objectively deciding whether or not the differences between two end points obtained in virus titrations represent real differences between the activities of the virus samples. To solve this problem, one must first have an estimate of the distribution of end points that would be obtained if an infinite number of titrations were carried out on a single virus preparation. Since the reliability of that estimate increases as the number of tests increases, we have pooled Horsfall's data (2) with our own for the cases of the death and the weighted end points. The standard deviations from the combined data were calculated to be 0.277 and 0.260, respectively. The standard deviation for the lesion end point was computed from our data alone and, as may be seen in Table II, has a value of 0.356. These values, which were computed by the methods of small sample statistics, are reasonable estimates of the most probable values of the standard deviations for infinite supplies, and they will be used for further deductions. Nevertheless, it must be recognized that the true statistics could possibly, though not probably, be considerably different.

From the above estimates one may next calculate the distribution of the differences between successively determined end points on a single sample. This problem is similar to the determination of the distribution of the differences between sample means. Thus it can be shown that the standard deviation of the distribution of differences between successively determined end points is equal to $\sqrt{2}$ times the standard deviation of the distribution of end points. For example, if weighted end points are calculated for pairs of titrations of a given virus sample, one should expect the distribution of differences between the members of the pairs to have a standard deviation of $\sqrt{2} \times 0.260 = 0.368$ logarithmic units. One would therefore expect to encounter differences of 0.368 units or more between members of the pairs about one time in three and differences of $2 \times 0.368 = 0.736$ or more about one time in 20. Conversely, if titrations on two virus samples not known to be identical give a difference of 0.736 units, one could conclude that the chances were

TABLE III

End Point Difference Required for Various Levels of Probability of Significance

Probability	End point difference required		
	Weighted end point	Death end point	Lesion end point
0.90	0.61	0.65	0.83
0.95	0.73	0.77	0.99
0.99	0.95	1.01	1.31
0.999	1.21	1.29	1.66

about 19 out of 20 that the two samples were not identical. We have here, therefore, a reasonably objective means of determining the significance of a difference between two end points obtained in the titration of the PR8 strain of influenza virus with mice in the manner described in this publication. In Table III are listed the differences between end point pairs for the three types of end points required to insure various levels of probability that the differences are significant. Since it seems reasonable to assume that the variability in the end points is principally related to the technique of inoculation and to the natural variation of the resistance of the mice, it may be expected that the variability encountered in titrating other strains of influenza virus should be of a comparable order of magnitude. At least until such data become available, these data may serve as a guide to the interpretation of all mouse titrations of influenza virus.

SUMMARY

A study has been made to establish the statistical significance of results obtained in mouse infectivity titrations of influenza virus. Five titrations,

each composed of five replicas, were carried out and 50 per cent end points were calculated for each titration. Three criteria for evaluating the end points were employed, namely, the presence or absence of pulmonary lesions, the occurrence of death, and a weighted composite taking into account both the extent of lung consolidation and the occurrence of death. Standard deviations of the distribution of end points obtained by each method were computed, and from these data levels of probabilities for significance in the differences between end points were determined. It was found that the chances are 19 out of 20 that differences of 0.99, 0.77, and 0.73 logarithmic units, respectively, for the lesion, the death, and the weighted end points are significant.

BIBLIOGRAPHY

1. Reed, L. J., and Muench, H., *Am. J. Hyg.*, 1938, **27,** 493.
2. Horsfall, F. L., Jr., *J. Exp. Med.*, 1939, **70,** 209.
3. Stanley, W. M., *J. Exp. Med.*, **1944, 79,** in press.