INTERNATIONAL JOURNAL OF LEPROSY

Volume 50, Number 1
Printed in the U.S.A.

Statistical Analysis of Results Obtained by Two Methods for Testing Drug Activity Against *Mycobacterium leprae*[1,2]

Charles C. Shepard[3]

Three methods have been used for measuring the activity of drugs against *Mycobacterium leprae* in mice. The first was the continuous method ([6]): the drug is administered continuously from the day the mice are infected until the end of the experiment. The method achieves its goal—the detection of even minimal amounts of antibacterial activity—and for technical reasons it is still the most convenient method for measuring the sensitivity of individual patient strains. The chief disadvantage of the method is that the results do not allow the discrimination between a merely bacteriostatic drug and one that is bactericidal, perhaps completely so in a few hours. This is an important point because there are many drugs that are at least bacteriostatic for *M. leprae* but only a few that are bactericidal, and it is only those few bactericidal drugs that have been useful in the therapy of human leprosy ([5]). The kinetic method ([4]) was therefore devised to allow the differentiation of bactericidal drugs. In this method the drug is administered for a limited period, frequently 60 days, beginning about the 60th day after infection. The logarithmic phase of the *M. leprae* growth curve is delineated for each group, and the growth delay (during the logarithmic phase) for the treated groups vs. untreated control is determined. It serves its purpose well, too—the differentiation of bactericidal from purely bacteriostatic drugs. It does not differentiate, however, between partial but pure bactericide, on the one hand, and either pure bacteriopause ([7]) or repository activity of a bacteriostatic drug on the other. Consequently, the proportional bactericidal method was devised ([1]). By this method the *M. leprae* is, in effect, titrated in parallel sets of mice, and the drug regimen is started in one set on the day of infection and continued, often for 60 days. After enough time to permit bacterial growth to reach the plateau stage, usually 12 months after inoculation, *M. leprae* are counted in individual mice and the number of *M. leprae* killed by drug is calculated through estimates of the most probable number (MPN) of organisms that would have produced the observed result if the number remaining after treatment were distributed randomly between mice. The method can provide incontrovertible evidence of bactericide and allows comparison of various degrees of bactericide, but it misses purely bacteriostatic or bacteriopausal drugs. Consequently, in the screening of new drugs or frequently in the comparison of a series of analogs, current practice is to use the kinetic method, in order not to miss what might be interesting activity. Conveniently for studies of quantitative-structure-activity relationships (QSAR), the kinetic method arrays the varying degrees of activities from partial bacteriostasis to complete bactericide on a continuous scale. The proportional bactericidal method is used to advantage in the comparison of bactericidal drugs and drug combinations.

A disadvantage of both the kinetic and the proportional bactericidal methods has been that satisfactory approaches for assessing the statistical significance of the results have not been available. The problem was especially prominent in kinetic method

[1] Received for publication on 24 February 1981; accepted for publication in revised form on 22 September 1981.

[2] Presented in part at the 15th Joint U.S.-Japan Leprosy Research Conference, 9–11 October 1980, Kagoshima, Japan.

[3] C. C. Shepard, M.D., Leprosy and Rickettsia Branch, Virology Division, Center for Infectious Diseases, Centers for Disease Control; U.S. Department of Health and Human Services, Atlanta, Georgia 30333, U.S.A.

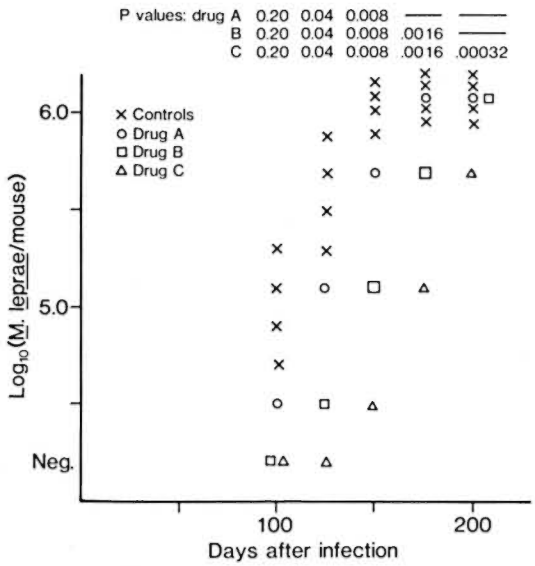

FIG. 1. Theoretical examples to illustrate the basis for the calculation of statistical significance of results obtained by the kinetic method. Neg = negative, i.e., no AFB were encountered during the standard counting procedure.

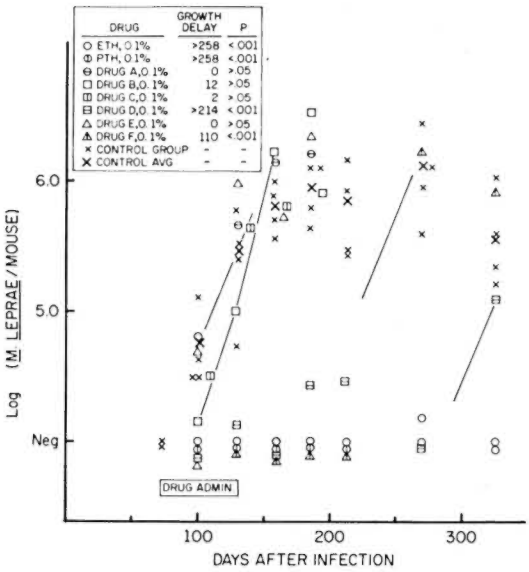

FIG. 2. Actual example of kinetic method applied to a comparison of thioamides administered in the diet. The lines drawn are for the purpose of estimating the growth delay; a generation time of 12.5 days is assumed and the points above $10^{5.5}$ are considered to be more reliable.

results where not enough harvests were performed to delineate the logarithmic phase satisfactorily. Beyond this, the needs for quantitation are greater now that QSAR are playing an important role in the design of new drugs.

Kinetic method. For this method the statistical approach described here is a simple one that requires only a moderate increase in experimental work. The principle is shown in Figure 1. In this example there are four groups of controls and one group for each drug. The probability that the arrangement of values shown at 100 days for either drug A, B, or C would occur by chance alone is $^1/_5$ or 0.20. The probability that this same arrangement would occur at both 100 and 125 days is 0.20×0.20 or 0.04, etc. In practice, we often include 10–12 drug-regimen groups and four control groups per experiment. There is usually a pool of four mice in each harvest. A recent experiment testing a group of drugs is illustrated in Figure 2, with the results tabulated in the upper left. The p values refer to the probabilities that the observed results (with a given drug) would differ from the control group by at least this much by chance alone. To estimate the significance of the difference between two drug groups, the same principle can be used. For only one group for each drug and one harvest from each group at each time interval, however, the p value cannot become less than 0.05 until there have been five comparative counts ($0.5^5 = 0.03$). If the time interval between harvests is 28 days, this would amount to a 140-day difference in growth delays, or longer than the logarithmic phase. When it is important to distinguish between two groups, more counts can be carried out ahead of schedule. For example, three counts from each of the two groups with no overlapping counts between groups could give a p value of 0.05, and two counts from each group at two intervals (with no overlapping between groups at each interval) could give a p value of 0.028. (In these latter cases, the probabilities were estimated with the aid of Fisher's exact test.) It may be worthwhile to increase the number of groups for a standard drug if comparisons with it are important. In the example in Figure 2, the differences are not significant ($p > 0.05$) between drug F and ethionamide (ETH) or prothionamide (PTH). Incidentally, both ETH and PTH

TABLE 1. *Current protocol for the kinetic method.*

1. Include at least four control groups.
2. When comparing with a standard drug, include two or three groups for it.
3. Include at least 30 mice per group.
4. With a "fast" strain in mouse passage, begin harvests (4 mice/pool, 1 pool/group) in the controls at 126 days. Start harvests in the treated groups as soon as all four control groups are positive.
5. Continue at 28-day intervals until 210 days and then continue at 56-day intervals.
6. Stop counts in any group when its count overlaps any control count.
7. If comparison between two particular drugs is important, increase the number of harvests from the two groups in the interval after the counts with the weaker drug have become distinctly positive and before they have reached plateau values.
8. Continue control counts as long as any other groups remain to be counted.

were completely bactericidal in this experiment.

To ensure that the result being analyzed covers the timing of the logarithmic phase of growth and not some interval selected gratuitously to give a desired result, the following rules should suffice. The intervals chosen for estimates of the statistical significance of the result are a) for comparison with the control, from the time when all the control counts have become positive to (but not including) the time when the treated group reaches plateau levels, and b) for comparison between two treated groups, from the time the group receiving the weaker drug has become distinctly positive to (but not including) the time when the group receiving the stronger drug has reached plateau values. That is, during the interval compared, one member of the comparison must clearly be in the logarithmic phase.

This statistical analysis is appropriate for our current protocol, which is shown in Table 1. The minimum activity detectable when this protocol is used is usually a growth delay of about 20 days. This comes about because the spread among the counts in the four control groups is usually less than a factor of $10^{0.7}$ or 0.7 logs to the base 10. A 0.4 log decrease from the mean of the four control groups would thus usually put the treated group outside the range of the controls, and 0.4 logs of growth at a doubling time of 12.5 days takes 17 days.

Proportional bactericidal method. In this method the number of *M. leprae* killed by drug is calculated through estimates of the MPN of organisms that would produce the result. It is assumed that the organisms in the suspensions inoculated are distributed randomly and that one delivered organism will produce an infection. Many other experimental infections, even those in which the agent is not present in clumps and globi, do not fit this model, however, and the situation might be even more complicated in the presence of chemotherapy. When the random (Poisson) distribution of successfully infecting organisms cannot be assumed, the usual approach is to estimate empirically the dose that infects 50% of the animals. The Reed and Muench method is most widely known, but it does not allow estimation of the statistical significance of the result. The Spearman-Kärber method does, however ([3]). This method requires that the titration be carried out over a range from 100% to 0% infections in the control group as well as the treated group. This range can usually be assured in the control group by using a good inoculum and extending the titration from 10^{4} to 10^{-2} acid-fast bacteria (AFB) inoculated per mouse. With treated groups, 100% may be impossible to achieve, but some simplifying assumptions that allow minimal estimates to be made are given below.

Table 2 gives examples of calculations from a comparative titration in CFW and CBA mice of infective *M. leprae* in a skin biopsy specimen. Because the method of calculation is not shown in most statistics textbooks and the Finney reference ([3]) is out of print, the method of calculation is given in detail. F is the number of AFB inoculated per mouse. The dilutions of F must be evenly spaced. Log F is transformed to x (log dose of organisms), so that the logarithm of the smallest dose is equal to 1. One then calculates the proportion of positive mice (p_i) at each dilution and sums the proportions (Sp_i) to use in the formula to calculate m; m is transformed back to log ID_{50}, which is the estimate of the logarithm of the number of AFB necessary to infect 50% of the mice. The variance at each dilution $p_iq_i/(n_i - 1)$ is also summed to use in the formula for Vm, which is the variance of m and also the variance of log ID_{50}. The square root of Vm is the standard deviation.

TABLE 2. *Spearman-Karber calculations of ID_{50}.*[a]

$$m = x_k + \tfrac{1}{2} - dSp_i$$
$$Vm = d^2S(P_iq_i)/(n_i - 1)$$

Mice	Log F	Log dose (x)	No. mice (n_i)	No. pos. (r_i)	r_i/n_i (p_i)	$p_iq_i/(n_i - 1)$
CFW	−1	1	9	0	0.00	.0000
	0	2	10	0	0.00	.0000
	1	3	8	0	0.00	.0000
	2	4	9	4	0.44	.0308
	3	5	10	10	1.00	.0000
				Sp_i	= 1.44	.0308 = $S\frac{p_iq_i}{n_i - 1}$

$Vm = 1^2 \cdot 0.0308 = 0.0308$
$\sqrt{Vm} = \sqrt{0.038} = \pm 0.175$
$m = 5 + \tfrac{1}{2} \cdot 1 - (1 \cdot 144) = 4.06$
Log ID_{50} = 2.06 ± 0.18 AFB

CBA	−1	1	10	0	0.00	.0000
	0	2	10	0	0.00	.0000
	1	3	10	2	0.20	.0178
	2	4	10	3	0.30	.0233
	3	5	7	7	1.00	.0000
				Sp_i	= 1.50	.0411 = $S\frac{p_iq_i}{n_i - 1}$

$Vm = 1^2 \cdot 0.0411 = 0.0411$
$\sqrt{Vm} = \sqrt{0.0411} = \pm 0.203$
$m = 5 + \tfrac{1}{2} \cdot 1 - (1 \cdot 1.50) = 4.00$
Log ID_{50} = 2.00 ± 0.20 AFB

Significance of the difference between CBA and CFW mice:
Pooled Vm = 0.0308 + 0.0411 = 0.0719
$\sqrt{\text{Pooled Vm}} = \sqrt{0.0719} = \pm 0.268$
$$\frac{ID_{50} - ID'_{50}}{\sqrt{\text{Pooled Vm}}} = \frac{2.06 - 2.00}{0.268} = 0.224$$
$p \simeq 0.18$

[a] Symbols are defined as follows:
m is defined by the equation. It is related to log ID_{50} through the transformation of log F to x.
x = log F transformed so that the smallest value is 1.
x_k = the highest value of x used.
d = the difference in values of x.
pi = proportion of positive mice in the ith dilution.
qi = proportion of negative mice in the ith dilution.
Vm = variance of m.
Others are defined in the table.

To calculate the significance of the differences between log ID_{50}s, the assumption is made, on the null hypothesis, that the two groups represent one population. The standard deviation of this difference is the square root of the sum of the two variances ($\sqrt{\text{Pooled Vm}}$). In this example, the difference in the two log ID_{50}s is 2.06 − 2.00 = 0.06, which was 0.224 times the standard deviation of the difference. The p value for such a difference may be looked up in tables of the areas under a normal curve and is about 0.18; i.e., the observed difference is not significant.

The endpoints with *M. leprae* titrations in untreated mice tend to be sharp; with tenfold dilutions one frequently has no more than one dilution with partial results. Under these circumstances the standard error of the ID_{50} is about 0.15 with 10 mice per group and about 0.25 with 5 mice per group, and for statistical significance between two groups, a difference in log ID_{50} of about 0.4 and 0.7, respectively, is needed. This corresponds roughly to a 2.5-fold and a fivefold difference, respectively. The required difference is a little less when the difference is in the expected direction, as it might be in a comparison between control and treated mice, for example.

Colston, Hilson, and Lancaster recently published some results with dapsone and ethionamide treatment ([2]). These are shown in Table 3. Their practice is to express the result as the MPN of viable *M. leprae* that would have produced the result, normal-

TABLE 3. *Results of Colston, Hilson, and Lancaster ([2]).*

Drug regimen[a] (for 60 days)	AFB/Mouse				Col. #6[b] Log (MPN/ 10^4 AFB[c])	Col. #7 Log (ID_{50}/ 10^4 AFB)	Col. #8 Col. #6 − log 0.69
	10^4	10^3	10^2	10^1			
Control	5/5[d]	5/5	5/5	0/5	2.38	2.50 ± 0.2	2.54
ETH cont.	5/5	0/5	0/5	0/5	0.38	0.50 ± 0.20	0.54
ETH, 3×/week	5/5	4/5	1/5	0/5	1.23	1.50 ± 0.28	1.39
ETH, 1×/week	5/5	5/5	3/4	0/5	2.11	2.25 ± 0.25	2.27
DDS cont.	4/4	3/5	1/5	0/5	1.04	1.30 ± 0.32	1.20
DDS cont. + ETH cont.	2/5	0/5	0/5	0/5	<0	$\leq -0.1 \pm \geq 0.24$	
DDS cont. + ETH 3×/week	1/4	1/5	0/5	0/5	<0	$\leq -0.05 \pm \geq 0.32$	
DDS cont. + ETH 1×/week	5/5	4/5	0/5	0/5	1.11	1.30 ± 0.20	1.27

[a] ETH = ethionamide, Cont. = continuously, and DDS = dapsone.
[b] Col. #6 = column number 6.
[c] AFB = acid-fast bacteria.
[d] (No. mice with AFB)/(No. mice counted).

ized to 10^4 AFB. In Table 3, their MPN/10^4 AFB is converted to logarithms in column 6. An infection of 50% of the mice theoretically corresponds to a mean number of infective *M. leprae* of 0.69, so column 8 gives the corresponding corrected figure for comparison to column 7 [the log (ID_{50}/10^4 AFB)]. The agreement is good in these cases.

The calculations of these ID_{50}s illustrate some of the conventions that are necessary (Table 4). In the first example, there were no dilutions that showed partial results. In such a case one assumes, for the purpose of calculating Vm, that one of the dilutions has one less positive than was actually observed. In the other example, 100% positivity was not reached. One accepts, however,

TABLE 4. *Examples of calculations for results in Table 3.*[a]

Log dose	No. mice	No. pos.	p_i	Assume	$p_iq_i/n_i - 1$	Log (AFB/ID_{50})	Log (ID_{50}/10^4 AFB)
Control							
1	5	0	0.00	0.00	0.00		
2	5	5	1.00	0.80	0.04		
3	5	5	1.00	1.00	0.00		
4	5	5	1.00	1.00	0.00		
			$Sp_i = 3.00$		$0.04 = S\frac{p_iq_i}{n_i - 1}$		
$Vm = 1^2 \cdot 0.04 = 0.04$ $\sqrt{Vm} = \sqrt{0.04} = \pm 0.02$ $m = 4 + \frac{1}{2} \cdot 1 - (1 \cdot 3.00) = 1.50$							
						1.50 ± 0.20	2.50 ± 0.20
DDS Cont. + Cont.							
1	5	0	0.00	0.00	0.00		
2	5	0	0.00	0.00	0.00		
3	5	0	0.00	0.00	0.00		
4	5	2	0.40	0.40	0.06		
5				1.00	0.00		
				$Sp_i = 1.40$	$0.06 = S\frac{p_iq_i}{n_i - 1}$		
$Vm = 1^2 \cdot 0.06 = 0.06$ $\sqrt{Vm} = \sqrt{0.06} = \pm \geq 0.24$ $m = \geq 5 + \frac{1}{2} \cdot 1 - (1 \cdot 1.40) = 4.10$							
						$\geq 4.10 \pm \geq 0.24$	$\leq -0.10 \pm \geq 0.24$

[a] Symbols are defined in Tables 2 and 3.

that the next higher concentration could not have given more than a 100% infection rate (or a sharper endpoint) and makes minimal estimates for the ID_{50} and a maximal estimate of its standard deviation.

With these estimates of the ID_{50} and its standard deviation, the following statements can be made about the results of Colston, *et al.* in Table 3: a) All the regimens except ethionamide (ETH) 1× weekly were significantly bactericidal. b) ETH 3× weekly was significantly less active than ETH continuously and ETH 1× weekly was less active than ETH 3× weekly. c) Dapsone (DDS) continously is not significantly less effective than ETH ($p = 0.10$). d) Addition of ETH by any of the three schedules to DDS increased the bactericidal effect. e) As an addition to DDS, ETH 3× weekly appears to be less active than the other two ETH schedules; one cannot say whether there was any difference between the addition of ETH continuously and ETH 3× weekly. To decide about this last point, the experiment could be repeated with shorter drug regimens or lower dosages, so that a 100% infection rate can be reached with the high inoculum (10^4 *M. leprae* per mouse).

SUMMARY

The methods used for the study of antileprosy drugs are briefly reviewed. The two chief methods used for the measurement of activity against *Mycobacterium leprae* are the kinetic method and the proportional bactericidal method. Methods for the statistical analysis of the results are now described for those two methods and their use is illustrated by examples.

RESUMEN

Se revisan los métodos usados para el estudio de las drogas antileprosas. Los dos métodos principales usados para la medición de la actividad contra el *Mycobacterium leprae* son el método cinético y el método bactericida proporcional. También se describen algunos métodos para el análisis estadístico de los resultados obtenidos usando los dos métodos mencionados y su uso se ilustra con ejemplos.

RÉSUMÉ

On passe en revue les méthodes utilisées pour l'étude des médicaments antilépreux. Les deux méthodes principales auxquelles on a recours pour mesurer l'activité contre *Mycobacterium leprae*, sont la méthode cinétique et la méthode bactéricide proportionnelle. Des méthodes sont décrites pour l'analyse statistique des résultats, après utilisation de ces deux types de procédés. Leur emploi est illustré par des exemples.

Acknowledgment. Dr. Ross J. Wood, Statistical Activities, Bureau of Laboratories, Centers for Disease Control, provided very helpful criticisms and suggestions. This work was partly supported by the U.S.-Japan Cooperative Medical Science Program, administered by the National Institute of Allergy and Infectious Diseases (NIAID), by means of an interagency agreement between NIAID and the Centers for Disease Control.

REFERENCES

1. Colston, M. J., Hilson, G. R. F. and Bannerjee, D. K. The "proportional bactericidal test," a method for assessing bactericidal activity of drugs against *Mycobacterium leprae* in mice. Lepr. Rev. **49** (1978) 7–15.
2. Colston, M. J., Hilson, G. R. F. and Lancaster, R. D. Intermittent chemotherapy of experimental leprosy in mice. Amer. J. Trop. Med. Hyg. **29** (1980) 103–108.
3. Finney, D. J. Statistical method in biological assay. New York: Hafner Publishing Co., 1964, pp. 524–530.
4. Shepard, C. C. A kinetic method for the study of the activity of drugs against *Mycobacterium leprae* in mice. Int. J. Lepr. **35** (1967) 429–435.
5. Shepard, C. C. A survey of the drugs with activity against *M. leprae* in mice. Int. J. Lepr. **39** (1971) 340–348.
6. Shepard, C. C. and Chang, Y. T. Effect of several anti-leprosy drugs on the multiplication of human leprosy bacilli in foot-pads of mice. Proc. Soc. Exp. Biol. Med. **109** (1962) 636–638.
7. Videau, D. La pristinamicine et le phénomenène de bacteriopause. Ann. Inst. Post **108** (1965) 602–606.