MICROBIOLOGICAL REVIEWS, Mar. 1978, p. 237-249
0146-0749/78/0042-0237$02.00/0

Vol. 42, No. 1
Printed in U.S.A.

Evaluation of Survival in Challenge Experiments

F. D. K. LIDDELL

Department of Epidemiology and Health, McGill University, Montreal, Quebec, Canada H3A 2B4

INTRODUCTION

To make the best use of experimental material, good design is essential; the results are then worth the best form of statistical analysis, i.e., that which allows the maximum amount of information to be derived from the data (7).

Principles of Design and Analysis

The principles of experimental design, first codified by R. A. Fisher (4, 5) in the 1920s and 1930s, apply in all biological fields. An excellent simple (non-mathematical) account of them is provided by D. R. Cox in his *Planning of Experiments* (3), on which I rely heavily for the following discussion. There are many technical accounts, among which the handbook (2) of W. G. Cochran and G. M. Cox is comprehensive and of high reputation. The terminology derives from agricultural research, wherein Fisher developed the principles, and technical meanings have often to be distinguished from those in common usage within different disciplines.

An experiment consists in applying one *treatment* to each *experimental unit* and making one or more *observations* on each unit, the assignment of treatments to units being under the control of the experimenter. A *treatment* (originally how a plot of land was treated, by sowing various crops on it as well as by application of various fertilizers, etc., to it) is here used for a *specific combination of factors of interest*, such as strain or infectious agent or vaccine. In a challenge experiment, the experimental unit has usually been one group of animals; the observations are the records of the interval, within a fixed period of time, in which each animal died (or whether it survived). The requirements for a good experiment are: absence of systematic error; precision and the calculation of uncertainty; wide range of validity of the results; and simplicity.

Experimental units receiving one treatment should show only *random* differences from units receiving any other treatment (including the control) and should be allowed to respond independently of one another. The absence of systematic error can only be ensured by a *randomization* procedure. Colloquially, random means haphazard, lacking aim or obvious pattern; technically (as in the present context) it means that differences are equally frequent in the units and hence are not systematic. Randomization gives to every experimental unit the same chance of being allocated to a particular relevant treatment; for example, the units of one strain are rearranged into random order and allocated to the combinations of vaccination or no vaccine and dose of inoculum. Thus uncontrolled variation between units is converted into random variation between treatments, and this by definition is without systematic error. There is a great deal of evidence from many fields including bacteriology that any other method of allocation can be subject to serious bias. So can the lack of randomization of the sequence in which treatments are dealt with.

The precision of any challenge experiment depends upon: the intrinsic variability of the experimental material; the design; and the number and size of the units (groups of animals). Within the basic design, there should be sufficient animals for cogent conclusions to be drawn, but not too many because of the waste implied. To widen the range of validity, we should, in designing the experiment, examine as wide a

range of conditions as possible without decreasing the precision of the experiment. It will remain important to recognize eplicitly the restrictions imposed on the generalizability of the conclusions from any particular experiment.

An important class of experiment is where the treatments (as defined above) consist of all combinations of different *levels* of a set of factors; they are called *complete factorial experiments*. For example, Robson and Vas (10) used a 9 × 7 × 2 design: there were 9 strains of mice × 7 doses of inoculum × 2 levels of the other factor, whether vaccinated or not; the number of treatments was 9 × 7 × 2 = 126, and one unit (of 10 animals) was allocated to each treatment. Factorial experiments have the following advantages over the classical one-factor-at-a-time approach: they give greater precision for estimating overall factor effects; they allow the range of validity of the conclusions to be extended simply; and, perhaps most important, they enable *interactions* between factors to be explored. In the example cited above (10), the design allowed study of how the improvement of survival due to vaccination varied from strain to strain and how it was related to dose of inoculum.

For other objectives (the examination of dose-response relations is one), other experimental designs are appropriate. However, most good designs have in common the need for *analysis of variance*, also due to R. A. Fisher (4), and considered by Sokal and Rohlf (13) to be "indispensable to any modern biologist" and a technique which "provides an insight into the nature of variation of natural events, into Nature in short.... " This technique, which can be used in many diverse situations, is described in most statistical texts, in particular the 1969 *Biometry* (13). Often now called ANOVA, it is a flexible means of *partitioning* the variation between experimental units into elements. One element is used to estimate the residual standard deviation (RSD); each of the other elements is associated with a factor or interaction (in a factorial experiment) or with a linear, quadratic, or higher-order component of a relationship. The exact form of partitioning is determined by the experimental design, but in each design all elements are constrained to be *orthogonal*, i.e., independent of each other, and each is allocated specific degrees of freedom. Indeed, the degrees of freedom of the variation between units (one less than the number of units) are subject to an exactly corresponding partition.

The RSD is a measure of the variation between observations on those units which receive the same treatment. Even where only one unit has been allocated to each treatment, it is usually possible to estimate the RSD. However, if the observations on each animal in an experimental unit can be utilized, the RSD can be calculated from them.

To exploit to full advantage the merits of good experimental design and of ANOVA, it is necessary to obtain a *measure* of survival for at least each experimental unit, and it then becomes possible to estimate the effects and interactions of the various factors, etc., with statements of standard errors. Other forms of analysis, including all nonparametric methods, are less efficient, usually much less so, while the multiple use of *t*-statistics can only be justified as a partitioning of ANOVA, with very careful attention to orthogonality.

Censoring of Survival Time

To estimate treatment differences, we have to average the measures of survival, and they must be analyzed on an appropriate scale (3). When all the animals die before the end of the period of observation, it is fairly straightforward to use the average time to death. However, if the period is not to be excessively long, a substantial proportion of animals will survive. For example, in the experiment of Robson and Vas (10) (Fig. 1), the period was 28 days, and in seven of the 126 groups all mice were still alive after 28 days, whereas in a further 37 units at least one mouse survived the 4 weeks. In other words, for 44 experimental units (35%), information about the time of death was "*censored*"; precise information on the time of death was obtained for only 78% of the 1,260 animals.

The Problem Arising from Censoring

Censoring creates a major problem in the determination of average survival and hence in any comparison of group survival. In particular, the arithmetic mean of survival times becomes indeterminate, but this has not stopped the quotation of "mean survival" calculated on the basis that those animals which were alive at the end of the period of observation died on a certain later day. Which day has not always been made clear; even where it has, the inevitably arbitrary choice of that day can lead to seriously misleading comparisons. The median of survival times also remains indeterminate where at least half the animals in any group survived—as in 30 of the 126 groups (24%) of Fig. 1—and even where it is determinate, the median has well-known disadvantages. Several attempts to force probit analysis on the data have been quite erroneous: if the experiments illustrated in Fig. 1 are thought of as 18 bioassays, probit analysis could have been used for estimating ED_{50} values (here, the dose of inoculum effective in killing just half the animals in 28 days) in only three (17%). This

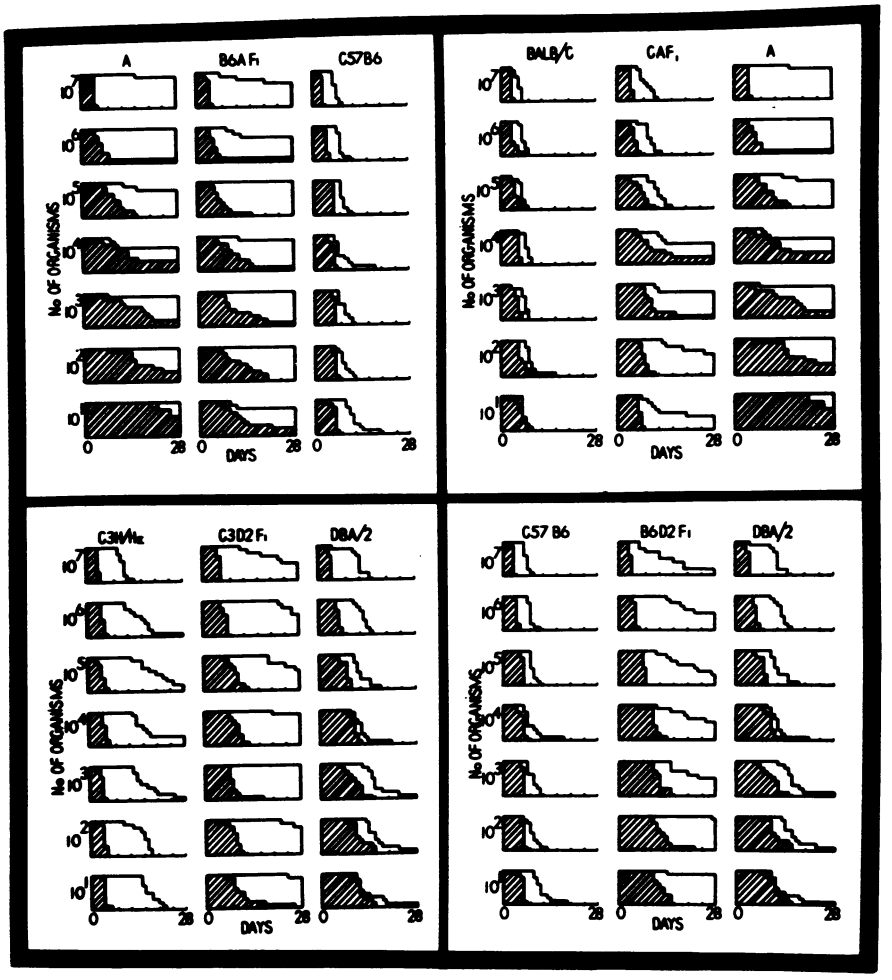

FIG. 1. *Results of Robson and Vas (10): survival of inbred and F_1 hybrid mice infected intraperitoneally with Salmonella typhimurium (Keller). Control mice are represented by shaded blocks; mice vaccinated with phenol-killed S. typhimurium are represented by open blocks. Each block represents 10 mice. (Reproduced with publisher's permission from reference 10, © 1972 by the University of Chicago.)*

is because the proportion of animals surviving 28 days was never greater than 50% for all doses of inoculum in 13 assays and never less than 50% for all doses in two. Many authors, recognizing this difficulty, have ignored the duration of survival. However, as Boyd (1, p. 717) points out: "If we treat the response ... as all-or-none we are throwing away some of the information which the experiment could provide."

Transformation as Solution

Three serious essays toward solving the problem caused by censoring are known to the author and are reviewed below. Each relies upon a *transformation* of survival time, a well-established statistical practice (13) when the transformation is functional; i.e., the transform is determined as a mathematical function of the recorded time. The first transformation reviewed was functional, the others not.

Smith and Westgarth (12) discussed the objections to what they called, in 1957, the time-honored methods for survey neutralization tests. In particular, mortality ratios (those animals dying within an appropriate number of days of observation as a proportion of all those observed) give only a few discrete points on the scale of mortality, which means that wide safety limits must be set and therefore that many repeat tests are required. Further, they noted the loss of information contained in the length of time between inoculation and death of each animal. These authors cited the finding of Gard (6) that "there is an approximately rectilinear relation between log dose of virus administered to a group of mice and the mean of the recipro-

cals of the number of days each mouse survives," and that similar relations had been found by various workers in toxicology, pharmacology, and virology. Smith and Westgarth showed that the relation was truly sigmoid, but claimed that over the region of interest it could be regarded as linear. However, in their view, the main benefit of this transformation was that it had been shown (6) that it tended to equalize the variance of responses over the whole range of doses.

The index used by these authors (12), studying neutralization tests in mice, was $Y = (1{,}000)\Sigma_1^k\{t_i\}^{-1}$, where k was the number of animals and t_i was the number of days survived by the ith animal in the group. This index is simply the group mean value of the transformation $(1/t)$, i.e., the reciprocal transformation, multiplied by 1,000. These authors defined Y only for a restricted range of t; animals that died before the 4th day were, following convention, disregarded on the grounds that the death was not relevant to the experiment; survivors beyond the observational period of 21 days were regarded as living as long again, and dying on the 42nd day. Thus, the possible values of $1{,}000/t$ are: 24, 48, 50, 53, ... (in steps increasing from 3 units per day to 18 units per day) ..., 143 (died on day 7), 167 (day 6), and 200 (day 5). Had earlier deaths been treated in the same way, they would have been scored 250, 333, 500, or 1,000 (days 4, 3, 2, or 1). The arbitrary nature of the effects of the rules relating to survival and to early death are obvious. The same transformation with the same rules was used again in 1962 by Smith et al. (11); it is evaluated against others in a later section of this paper.

An early contribution (8) is quoted in reference 1 (p. 719). Ipsen had devised a "function of death time of guinea pigs giving linear relation to logarithm of dose of diphtheria toxin"; to do so, he had "taken advantage of knowledge accumulated from many experiments with toxin and any particular species of animals." The "function" has properties broadly similar to those of the reciprocal transformation, but does not have a simple mathematical description. Further, it is by definition specific to the combination of toxin and species.

Also cited in this reference (p. 717) is a later essay by Ipsen and co-workers (9), who graded results on the immunizing potency of tetanus toxoid as follows: death within 2 days, 0; death within 3 and 4 days, 2; death within 5 and 7 days, 3; survival on 7th day with tetanus, 4; survival on 7th day without tetanus, 6. This is equivalent to a transformation of survival time into a step function. We must therefore treat Boyd's statement that "this made it unnecessary to make use of any transformation ... " only in relation to the needs for the next stage of analysis. (These few pages [717–722] of reference 1 contain many errors that could be traps for the unwary. Corrections have been incorporated in the present text.) However, there are strong theoretical grounds for believing that a continuous function would have been preferable, whereas the complication that survival to the end of the observational period was scored differently according to whether tetanus was present or absent renders the actual grading of times even more arbitrary.

In the remainder of this paper, we show how *functional* transformation can provide a generalizable solution to the problem created by censoring. The next section presents criteria for the evaluation of transformations and a choice for study. Then these are examined, using the data of Table 1, and the "best" is applied in the following section to some of the other material already introduced. Table 1 contains the data of a 4×2 factorial experiment: 40 mice of strain C were subdivided into experimental units of 10, each of which was challenged by one of four agents; and 40 mice of strain A were similarly treated. Records were maintained daily for 28 days, and the data are presented in complementary forms.

CRITERIA FOR AND CHOICE OF TRANSFORMATIONS

Criterion (1). To be generally acceptable, a transformation $f(t)$ must be a strictly monotonic function of t, the number of days (or other time units) after challenge, it must take a fixed finite value for $t = 0$, and it must approach an asymptote when $t \to \infty$. Without loss of generality, we take $f(0) = 0$ and let $f(t) \to 1$ as $t \to \infty$.

Criterion (2). The transformation should help improve the homogeneity of the standard deviations of $f(t)$ for the various experimental units.

Criterion (3). When t reaches the full period of observation (T), whether one uses $f(T)$ or $f(\infty)$ should have little effect on the mean of $f(t)$ for the group or any comparison between groups.

Criterion (4). It should be possible to perform the inverse transformation in order to make point and/or interval estimates of central tendency of survival on the original time scale.

Criterion (5). "Signal/noise ratios" should be high: it should be possible to distinguish between groups despite the differences between animals within groups.

Two other considerations are as follows. First, death recorded at day t has to be assumed to have occurred at some point in the 24 h from the end of day $t - 1$ to the end of day t. One could take as the measure of response $y(t) =$

TABLE 1. *Typical survival data*

(a) Numbers surviving to end of stated day after injection

Day →	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Strain C																												
P[a]	10	10	6	5	0																							
Q	10	10	10	10	3	0																						
R	10	10	10	10	10	8	2	0																				
S	10	10	10	10	6	5	5	5	4	4	3	2	1	0														
Strain A																												
P	10	10	10	10	10	8	6	2	0																			
Q	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
R	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
S	10	10	10	10	10	10	10	9	9	9	8	7	7	7	6	5	5	5	5	5	5	5	5	5	5	5	5	5

(b) Numbers dying during stated day after injection

Day →	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	Later
Strain C																													
P	—	—	4	1	5																								
Q	—	—	—	—	7	3																							
R	—	—	—	—	—	2	6	2																					
S	—	—	—	—	4	1	—	—	1	—	1	1	1	1															
Strain A																													
P	—	—	—	—	—	2	2	4	2																				
Q	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	10
R	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	10
S	—	—	—	—	—	—	—	1	—	—	1	1	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	5

[a] P, Q, R, S are codes for agent injected.

$f(t - \frac{1}{2})$, but it seems more appropriate to use $z(t) = \frac{1}{2}\{f(t) + f(t-1)\}$; another possibility is discussed later. Second, some convention is required for handling survival beyond the period of observation. It could be assumed that all animals died at one fixed time, say immediately after the period of survival, or after a further fixed period (cf. the use by Smith and Westgarth [12] of an equally long further period), or after an infinite period. Not only is any assumption of this nature arbitrary, but it precludes any estimate of variance when all animals survive: thus criteria (3) and (4) could not be met. Alternatively, it can be assumed that the animals die at varying times, and here two approaches seem feasible. In the first, values of $f(t)$ equally spaced in the interval $[f(T), 1]$ could be allocated to the survivors; in the second, $f(T)$ would be allocated to half the survivors and 1 to the other half. The first approach is attractive, but the second has been adopted because it maximizes the variance of $f(t)$, i.e., indicates the greatest uncertainty about the points of death of survivors.

Negative-exponential transformations, of the form $f(t) = 1 - \theta^{t/T}$, where θ is a constant, clearly meet criterion (1). So that criterion (2) can be met, $f(T) = 1 - \theta$ has to be fairly close to 1, i.e., θ fairly close to zero. We have taken four values of θ (0.01, 0.05, 0.1, and 0.2); for evaluation on the data of Table 1 we have considered the single relevant value of T, 28 days.

The reciprocal transformation does not meet criterion (1), because $1/t \to \infty$ as $t \to 0$. This in turn means that $z(t)$, defined as above, is indeterminate for $t = 1$. However, in view of earlier use, it appeared desirable to consider, if possible, some of the family of inverse-power transformations. To this end, criterion (1) can be relaxed provided the measure of response is $y(t) = f(t - \frac{1}{2})$, ensuring that $y(1) = f(\frac{1}{2})$ remains finite, without violation of the monotonic nature of the transform. This allows the use of transform $1 - \phi/t$, where ϕ is constant. The choice of ϕ affects $f(\frac{1}{2})$ and $f(T)$ and so might be expected, at first sight, to affect also criteria (2), (3), etc. However, in any comparison of group means in terms of their standard errors, ϕ appears in both numerator and denominator, and so its value is immaterial. The choice made was $\phi = \frac{1}{2}$; the measure of response during the first day became zero. Another set of possibilities is of form $1 - \{\psi/(t - \frac{1}{2})\}$, where ψ is a constant, the choice of which is again unimportant. With $\psi = 1$, the transform is proportional to that of Smith and Westgarth (12), except that they used $f(t)$, not $y(t)$; again response during the first day leads to a measure = 0. Because these were both rather "strong" transformations, a weaker version, based on the inverse square root of t, was also studied.

The seven selected transformations of survival time are summarized in Table 2 and illustrated in Fig. 2. In the next section they are applied to the eight groups of animals of Table 1. Mean

TABLE 2. *Transformations studied*

(a) Negative exponential: $f(t) = 1 - \theta^{t/T}$, $T = 28$ days

Transformation	θ	$f(T)$	Response measure
E	0.2	0.8	$z(t) = \frac{1}{2}\{f(t) - f(t-1)\}$
F	0.1	0.9	
G	0.05	0.95	
H	0.01	0.99	

(b) Inverse power

Transformation	Form of $f(t)$	$f(T)$	Response measure
J	$1 - (t + \frac{1}{2})^{-1/2}$	0.8127	$y(t) = 1 - t^{-1/2}$
K	$1 - (t + \frac{1}{2})^{-1}$	0.9649	$y(t) = 1 - t^{-1}$
L	$1 - 1/(2t)$	0.9821	$y(t) = 1 - 1/(2t-1) = (2t-2)/(2t-1)$

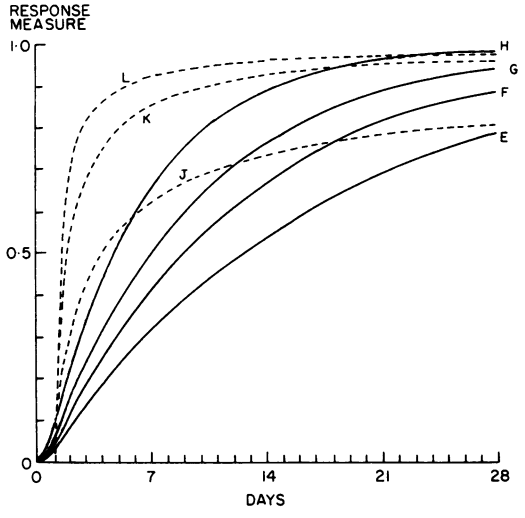

FIG. 2. *Seven transformations studied; see Table 2. The full curves are of the response measure z(t) for transformations E through H. The broken curves are of the response measure y(t) for transformations J through L.*

survival, in terms of the various transforms, together with standard deviations (SDs), are in Table 3.

EVALUATION OF SELECTED TRANSFORMATIONS

As explained in the previous section, the negative-exponential transformations (E, F, G, H) all meet criterion (1), and those based on inverse powers (J, K, L) meet a relaxed version of it.

For a group in which all animals survived (e.g., AQ or AR), criterion (2) suggests that the standard deviation should be within the range of SDs observed in other series; only transformations G and H failed in this respect. The ratio of the largest SD to the smallest lay between 4.74 and 7.36 for transformations (in sequence) K, L, G, J, and F; the ratio was rather higher (11.45) for E and very high (30.98) for H. Of course, it remains important that when the deaths in any group are observed over a wide interval of time, the SD should reflect this: two groups with similar means may yet be correctly distinguished by SD. As an example, groups CS and AP had similar means on all seven transformations, but in each case the SD for CS was about three times that for AP. In this material, the variation in survival clearly depended on the agent, the impression from Table 1(b) being that Q showed the least, there was slightly more for R and again for P, but substantially the greatest variation with S, all for both animal strains. From Table 3, it would seem that the impression is confirmed by the SDs of transformation F, somewhat distorted by those of G and H, rather more so by E and J, and denied by K and L.

Criterion (3) can be examined for group AS: the mean of each transform is given in Table 4, together with the standard error of the mean (i.e., $SD/\sqrt{10}$). Also given are the lowest and highest estimates of each mean, obtained by taking the five survivors as dying immediately after the period of observation, i.e., scoring them all $f(T)$, or as surviving indefinitely, i.e., scoring them all $f(\infty) = 1$. The final line of Table 4 shows the difference between the highest and "best" estimates of each mean in terms of its standard error. The largest such difference was 1.16, and even for this transformation (J) the criterion would seem to be met adequately; it is best met by H. Where all animals survive to at least T, the convention for handling them by scoring $f(T)$ for half and 1 for the rest leads to mean survival $1 - \frac{1}{2}f(T)$, with standard error $\frac{1}{2}f(T)/(n-1)^{1/2}$, where n is the number of animals in the group. Thus, all transformations perform identically for groups such as AQ.

It is also evident that any transform which meets criterion (1) can provide a point estimate, τ, of central tendency on the original time scale. Table 5 provides the expression for reconverting from each transformation. Interval estimates can also be obtained, by applying the same formulae to the confidence limits previously calculated (by appeal to the central limit theorem and the Student t-distribution) for the mean on the transformed scale. Table 6 gives the values of τ with 90% confidence limits. For groups CP, CQ, CR, CS, and AP, in which all animals died, usually within a short interval, there was a consistent tendency for smaller τ from use of the higher-lettered transformations, but it was very slight except for group CS. The widths of the confidence intervals also depended only slightly on the particular transformation. Thus, for groups in which all animals die, there seems little to favor one transformation over any other. In group AS, the median survival was at least 22

TABLE 3. *Means and standard deviations of survival*[a]

Group[b]	Transformation						
	E	F	G	H	J	K	L
CP							
Mean	0.1854	0.2533	0.3158	0.4382	0.4954	0.7471	0.8502
SD	0.0468	0.0614	0.0734	0.0926	0.0647	0.0663	0.0442
CQ							
Mean	0.2405	0.3251	0.4000	0.5431	0.5645	0.8100	0.8950
SD	0.0208	0.0264	0.0303	0.0351	0.0188	0.0161	0.0098
CR							
Mean	0.3110	0.4128	0.4993	0.6536	0.6208	0.8559	0.9223
SD	0.0264	0.0322	0.0357	0.0380	0.0183	0.0140	0.0081
CS							
Mean	0.3556	0.4602	0.5448	0.6864	0.6335	0.8600	0.9238
SD	0.1324	0.1552	0.1660	0.1642	0.0795	0.0580	0.0335
AP							
Mean	0.3337	0.4398	0.5287	0.6834	0.6347	0.8659	0.9280
SD	0.0415	0.0501	0.0550	0.0572	0.0271	0.0202	0.0117
AQ and AR							
Mean	0.9	0.95	0.975	0.995	0.9064	0.9825	0.9911
SD	0.1054	0.0527	0.0264	0.0053	0.0987	0.0185	0.0094
AS							
Mean	0.6941	0.7816	0.8405	0.9182	0.8080	0.9484	0.9731
SD	0.2382	0.1942	0.1586	0.0990	0.1280	0.0417	0.0221
Range of means	0.7146	0.6967	0.6592	0.5568	0.4110	0.2408	0.1409
Mean of SDs	0.0896	0.0781	0.0715	0.0621	0.0667	0.0317	0.0185
$SD_{max} + SD_{min}$	11.5	7.4	6.8	31.0	7.0	4.7	5.5
(Range of means) + (mean of SDs)	8.0	9.0	9.2	9.0	6.2	7.6	7.6

[a] Data of Table 1 after the transformations of Table 2.
[b] The first letter of the group indicates the strain; the second indicates the agent.

TABLE 4. *Mean survival, with standard error and limits*[a]

Determination[b]	Transformation						
	E	F	G	H	J	K	L
BEM	0.6941	0.7816	0.8405	0.9182	0.8080	0.9484	0.9731
SEM	0.0753	0.0614	0.0502	0.0313	0.0405	0.0132	0.0070
90% confidence limits							
Lower	0.5560	0.6690	0.7486	0.8608	0.7338	0.9242	0.9489
Upper	0.8322	0.8942	0.9324	0.9756	0.8822	0.9726	0.9659
LEM	0.6441	0.7566	0.8280	0.9157	0.7612	0.9396	0.9686
HEM	0.7441	0.8066	0.8530	0.9207	0.8548	0.9572	0.9776
(HEM − BEM) + SEM	0.66	0.41	0.25	0.08	1.16	0.67	0.64

[a] Group AS after transformations of Table 2.
[b] BEM, Best estimate of mean (i.e., taking half the survivors as dying immediately and half as surviving indefinitely); SEM, standard error of the mean; LEM, lowest estimate of the mean; HEM, highest estimate of the mean.

days and only transformation J yielded τ as large as this, whereas G and particularly H led to unacceptably low τ and narrow interval. Where all animals have survived (groups AQ and AR), a point estimate of central tendency may be considered meaningless; the lower confidence limit should be somewhere near $T = 28$ days and the interval fairly wide. None of the transformations considered is ideal in these respects; perhaps F might be thought of as providing the best compromise.

The range (largest minus smallest) of means is quoted at the foot of Table 3, together with the arithmetic mean of the SDs. The ranges decrease steadily across the table, and at first sight this might appear powerful contraindication for the higher-lettered transformations. However, the mean SD also decreases in similar fashion, and the two phenomena have to be weighed together. A first evaluation is available in the ratio (range of means)/(mean SD), which is roughly proportional to the signal/noise ratio,

TABLE 5. *Reconversion to original time scale*[a]

(a) Negative-exponential transformations, including E, F, G, and H

$$t = \{T/\log \theta\} \cdot \log(1 - z)$$

(b) Inverse-power transformations: J, K, and L

J	$t = (1 - y)^{-2} - \frac{1}{2}$
K	$t = (1 - y)^{-1} - \frac{1}{2}$
L	$t = \frac{1}{2}(1 - y)^{-1}$

[a] To find the value of t corresponding to a transformed variable z (or y), use the appropriate expression above.

which should be high. The ratio takes values around 9 for transformations F, G, and H and between 6 and 8 for the rest.

Signal/noise ratios can best be evaluated by ANOVA, and Table 7(a) gives the F-ratios for assessing the differences between the eight groups (with 7 df) as the signal against the noise, i.e., variation within series (with 72 df). These are all enormous but, for what it is worth, largest for transformations F and G. Study of Table 1 suggests that the differences between treatments depended on the strain; for strain C, agents Q

TABLE 6. *Point and interval estimates of central tendency of survival*[a]

Group	Transformation						
	E	F	G	H	J	K	L
CP	3.57 (3.00, 4.16)[b]	3.55 (2.99, 4.15)	3.55 (2.98, 4.15)	3.51 (2.95, 4.12)	3.43 (2.90, 4.08)	3.37 (2.87, 4.05)	3.34 (2.85, 4.03)
CQ	4.79 (4.51, 5.06)	4.78 (4.51, 5.06)	4.77 (4.50, 5.05)	4.76 (4.50, 5.04)	4.77 (4.52, 5.05)	4.76 (4.52, 5.04)	4.76 (4.52, 5.03)
CR	6.48 (6.10, 6.87)	6.47 (6.09, 6.87)	6.47 (6.09, 6.86)	6.45 (6.07, 6.85)	6.45 (6.08, 6.86)	6.44 (6.07, 6.85)	6.44 (6.07, 6.85)
CS	7.65 (5.69, 9.85)	7.50 (5.62, 9.71)	7.36 (5.56, 9.58)	7.05 (5.44, 9.25)	6.94 (5.37, 9.24)	6.64 (5.26, 8.90)	6.56 (5.23, 8.81)
AP	7.06 (6.45, 7.70)	7.05 (6.43, 7.69)	7.03 (6.42, 7.69)	6.99 (6.39, 7.67)	6.99 (6.39, 7.68)	6.96 (6.36, 7.67)	6.94 (6.35, 7.67)
AQ and AR	40.06 (31.76, 56.49)	36.43 (30.63, 47.91)	34.48 (30.02, 43.30)	32.21 (29.32, 37.96)	113.5 (43.4, 752.9)	56.5 (34.9, 146.0)	56 (34.8, 144.0)
AS	20.61 (14.13, 31.05)	18.50 (13.45, 27.31)	17.16 (12.90, 25.19)	15.22 (11.89, 22.57)	26.63 (13.61, 71.56)	18.88 (12.70, 35.96)	18.59 (12.59, 35.49)

[a] Data of Table 1, reconverted to original time scale, after manipulation when under the transformations of Table 2.
[b] Figures in parentheses are lower and upper 90% confidence limits.

TABLE 7. *Signal/noise ratios (see text)*[a]

(a) Between all 8 groups, compared with within-groups
Ratios of mean squares: degrees of freedom 7 and 72

Transformation						
E	F	G	H	J	K	L
68.4	84.8	83.6	68.0	42.9	53.3	48.1

(b) Strain × treatment interaction, compared with within-groups
Ratios of mean squares: degrees of freedom 3 and 72

Transformation						
E	F	G	H	J	K	L
21.8	22.0	17.1	8.1	8.3	4.6	3.9

(c) Strains and treatments, compared with interaction
Ratios of mean squares:
(i) Strains: degrees of freedom 1 and 3
(ii) Treatments: degrees of freedom 3 and 3

	Transformation						
	E	F	G	H	J	K	L
(i)	13.6	17.4	22.3	38.8	24.8	55.1	56.3
(ii)	1.8	2.3	3.0	5.6	3.6	9.2	9.9

[a] Data of Table 1 after the transformations of Table 2.

and R led to all animals dying fairly close to the central time and none after 8 days, whereas the strain A animals on these treatments all survived 28 days. It is therefore desirable that this interaction should be clearly recognized. The appropriate ratios are in Table 7(b): those for transformations K and L were rather small, and, although those for H and J would have been judged significant in any test, E, F, and G yielded substantially the largest. The differences between strains and the consistency of differences between treatments have to be assessed against the interaction [see part (c) of Table 7]. The F-statistics for all transformations point clearly to a big difference between strains. For E, F, and G, the difference between treatments did not greatly transcend the very large interaction term; for K and L, the rather low interaction term led to assessment of fairly consistent between-treatment differences; H and J were intermediate.

To summarize, it would appear that all the transformations considered have certain strengths and weaknesses, and the choice between them is not likely to be critical. The only one that is known to have been used previously is K, which performed similarly to L, with both perhaps a little less satisfactory than J. However, none of these inverse-power transformations appeared more acceptable than certain negative-exponential forms, and of these F is judged the most suitable for these data. It is, of course, essential that like be compared with like; therefore when comparisons are to be made between experiments, the identical transformation must be used throughout. However, in any specific field of inquiry, the period of observation (T) may be quite different from the 4 weeks of the data of Table 1. The transformation that is applied in the next section is, then, $f(t) = 1 - (0.1)^{t/T}$.

APPLICATIONS OF NEGATIVE-EXPONENTIAL TRANSFORMATIONS

Three sets of data discussed in the introduction are reexamined here. Details of the rest were not available.

Complete Factorial Experiment

In Robson and Vas's 9 × 7 × 2 factorial experiment (10), the mice were of nine strains, five inbred, here called A, B, C, D, and E, and four hybrids, AC, AD, BE, and BC, where the first letter indicates the strain of the males and the second indicates that of the females. All were infected by intraperitoneal injection of 0.1 ml of *Salmonella typhimurium* (virulent), the dose varying from 1×10^7 colony-forming units/ml, in serial 10-fold dilutions, to 1×10^1 colony-forming units/ml, all 10 animals in an experimental unit receiving the same dose. Half the groups had been vaccinated previously with *S. typhimurium* phenol vaccine. Deaths were scored daily at a standard time for 28 days. The original authors were reduced to presenting their findings in a diagram (Fig. 1) with a descriptive commentary. However, after negative-exponential transformation ($\theta = 0.1$; $T = 28$), ANOVA became possible (Table 8). Although SDs were

TABLE 8. *Analysis of variance to study vaccination in relation to strain and dose of inoculum*[a]

Source[b]	Degrees of freedom	Mean squares	Degrees of freedom	Mean squares
Main effects:				
V	1	24.2903		
S	8	2.3682		
D	6	0.9717		
Interactions:				
V × S	8	0.5123		
V × D	6	0.1509		
S × D	48	0.0402		
V × S × D	48	0.0398	96	0.0400
Between units	125	(0.4633)		
Within units	1,134	0.0100		
Total	1,259	(0.0550)		
		Ratios of mean squares		
V × D interaction	(0.1509)/(0.0400) = 3.8		df: 6 & 96	$P = 0.0020$
V × S interaction	(0.5123)/(0.0400) = 12.8		8 & 96	2.2×10^{-12}
S main effect	(2.3682)/(0.5123) = 4.6		8 & 8	0.0222
D main effect	(0.9717)/(0.1509) = 6.4		6 & 6	0.0197

[a] Data of Robson and Vas (10) after negative-exponential transformation with $\theta = 0.1$, $T = 28$.
[b] V, Vaccination; S, between strains; D, between doses.

unequal for different strains, interpretation was unaffected whether it was based on the RSD averaged over all strains or on an estimate from the strain with the largest degree of variation.

The strains × doses interaction was closely similar to the highest-order interaction ($S \times D \times V$), and so they were pooled to give a basis for testing the other interactions. Both were found to be important, as indicated by the P-values of 0.0020 and 2.2×10^{-12} at the foot of Table 8. Thus the effect of vaccination (last column of Table 9) depended on dose, the increase in survival, on the scale of the transform, being greatest for the largest dose and least for the smallest. Care is required in interpretation on the original time scale (see Fig. 3). The variation in the effect of vaccination with strain was very great, but no clear pattern emerged (see the last column of Table 10, which has been arranged with each hybrid strain between its parent strains). It is reasonable to take the main effects of dose and strain as transcending the varying effects of vaccination; this is indicated by P-values of 0.0222 and 0.0197 (foot of Table 8) and illustrated by the "overall means" of Tables 9 and 10. Indeed, the differences between doses can be further partitioned, by use of orthogonal polynomials (12), to explore the dose-response relationships; the sum of squares (mean square × degrees of freedom: $0.9717 \times 6 = 5.8304$) has three elements, linear (4.9729) and quadratic (0.8373), each using 1 df, and remainder (0.0202), with 4 df. The first term leads to a ratio $(4.9728)/(0.1509) = 32.95$, with $P = 0.0012$; ratios for the other terms are not particularly high. Similarly, the linear element of the relation between the effect of vaccination and dose (last column of Table 9) leads to a ratio $(0.7326)/(0.0400) = 18.32$, with $P = 0.0052$. Corresponding partitions of the strains effect might be possible, but would require a very careful allocation of the 8 df, preserving orthogonality yet making the desired comparisons.

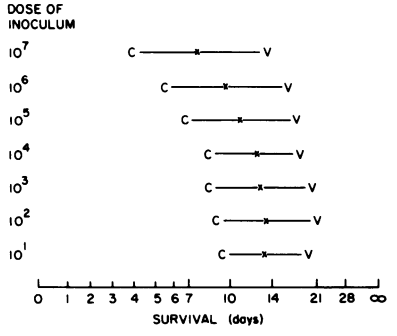

FIG. 3. *Results of Robson and Vas (10). Interaction between vaccination and dose. The ends of the horizontal lines indicate mean survival of (V) vaccinated and (C) control animals, averaged over species; the crosses represent the overall means.*

TABLE 9. *Survival in relation to dose of inoculum and vaccination*[a]

Dose of inoculum	Overall mean	Mean over all strains		Difference[b]
		Vaccinated	Control	
10^7	0.47	0.65	0.29	0.36
10^6	0.55	0.72	0.38	0.34
10^5	0.59	0.74	0.45	0.29
10^4	0.64	0.75	0.52	0.23
10^3	0.65	0.78	0.52	0.26
10^2	0.67	0.80	0.54	0.25
10^1	0.66	0.77	0.56	0.21

[a] Data of Robson and Vas (10) after negative-exponential transformation with $\theta = 0.1$, $T = 28$.
[b] Estimate of effect of vaccination.

TABLE 10. *Survival in relation to strain and vaccination*[a]

Strain[b]	Overall mean	Mean over all doses		Difference[c]
		Vaccinated	Control	
D	0.39	0.44	0.35	0.08
AD	0.60	0.72	0.47	0.25
A	0.77	0.92	0.63	0.29
AC	0.72	0.90	0.53	0.37
C	0.45	0.53	0.37	0.16
BC	0.69	0.85	0.52	0.33
B	0.60	0.68	0.51	0.17
BE	0.71	0.92	0.50	0.42
E	0.53	0.75	0.31	0.43

[a] Data of Robson and Vas (10) after negative-exponential transformation with $\theta = 0.1$, $T = 28$.
[b] Inbred strains are identified by a single letter; hybrid strains are indicated by two letters, the first identifying the males, the second the females.
[c] Estimate of effect of vaccination.

Thus, ANOVA has revealed and quantified survival patterns that, without transformation, could only be discussed in qualitative terms. The same data (10) were analyzed a second time as though all the experiments had terminated at 21 days (but with the same transformation). Virtually no difference could be found in *relative* effects; in other words, strains differed greatly, interaction of the forms discussed above existed for most strains, and vaccination was protective, the degree of protection being very similarly assessed, strain by strain, after 21-day censoring as in the real 28-day data. Both 21-day censored data and original data were also transformed by taking $\theta = 0.1$, $T = 21$; findings were again very similar.

Inverse Prediction of MLD

The objective of Ipsen's 1941 experiment (8) was to estimate the minimum lethal dose (MLD)

of diphtheria toxin in guinea pigs, that is, according to Boyd (1, p. 720), to find "the dose which decreases the survival time of the average guinea pig to 4 days." We transformed the data by using the negative exponential in which the unit of time was taken as 12 h, the length of observation as 144 h ($T = 12$), because Ipsen's function is not defined beyond 144 h, and $\theta = 0.1$; we also used Ipsen's own function. As dose decreased from 0.01 to 0.004 ml, the negative-exponential transform of survival increased from 0.487 to 0.889 (i.e., through 0.785, the 96-h value) and the relationship with log dose was essentially linear. Over the same range of doses, the values of Ipsen's function decreased from 0.353 to 0.035, but the relationship with log dose was definitely curved, particularly near the 96-h value of 0.085. The degree of linearity is indicated by the "coefficient of determination" (12), which was 0.96 and 0.84, respectively. (For higher and lower doses, both relations were sigmoid.)

Because of the high degree of linearity after negative-exponential transformation, the estimate obtained by inverse prediction (13) can be considered the most reliable; it was 0.00535 ml. The corresponding estimate from a line fitted to Ipsen's function yielded an estimate of 0.00511; but, from a fit only in the region of 0.085 and taking the curvature into account, MLD = 0.00523 ml. (Boyd [1, p. 720] incorrectly assumed not only linearity with Ipsen's function but also unit slope. He then used a rather indirect method of estimation that, correctly employed, gives MLD = 0.00552 ml. Boyd's own calculations contained several errors, and it was fortuitous that his estimate, 0.00555 ml, was not farther from the truth.)

Bimodal Distribution of Survival Times

Because the 1953 material of Ipsen et al. (9) differentiated survivors at the end of a week ($T = 7$) according to the presence or absence of tetanus, we treated these survivors as follows: half the mice with tetanus were taken as dying by $t = 8$ days and the other half by $t = 10$; half those without tetanus were assumed to die by $t = 11$ and the rest to survive indefinitely, so that $f(t) \rightarrow 1$. In a second system, all survivors were treated alike in the usual way, i.e., half dying immediately, half surviving indefinitely. Similarly, we used the original scores and also those scores, but treating all survivors alike (scored 4). Survival scores in these four systems were subjected to ANOVA and, for each system, a linear model of the relation between log dose and transformed survival was found adequate. Not unexpectedly, little difference was found between the two negative-exponential systems; indeed, another choice of θ and T (0.1 and 14 days) also led to very similar findings. In the 1953 scoring system, the tetanus adjustment improved the degree of linearity slightly. However, that system did not meet criterion (1); the step function, with or without the tetanus adjustment, also failed to meet criteria (3) and (4).

The objective was stated (1, p. 718) to be the estimation, by inverse prediction, of the dose for a score of 3, corresponding to death "within 5 and 7 days." In the cited assay of tetanus toxoid, the distribution of survival times was markedly biomodal: 21 of the total of 48 mice died within 4 days, while 24 survived 7 days; thus only 3 (or 6.25%) died in the interval associated with the critical score. In such circumstances, the actual time of survival provides virtually no additional information, and transformation is of no assistance in achieving the objective; it has probably given a spurious sense of simplicity to the form of the relation discussed in the previous paragraph.

DISCUSSION

The objective of this paper has been to provide a solution to the problem arising in challenge experiments where observations are censored but a measure of the central tendency of survival is required. The negative-exponential transformation, $1 - \theta^{t/T}$, has been adjudged better than any of several transformations of the inverse-power family, including the reciprocal, which has been used previously. It has been shown to be effective in a considerable range of circumstances and to perform at least as well as certain empiric transformations. The principle of transformation is natural enough to the practicing statistician, but is not necessarily the only solution. For example, P. H. A. Sneath (personal communication) has suggested that it might be worthwhile to consider describing the survival curve as a mixture of short- and long-term mortality rates, bearing in mind the bimodal nature of many of the distributions of survival in this type of experiment. That solution is not yet readily available, whereas transformation is. And it is important to appreciate that we are not proposing just an unnecessary complication. This can be seen from Fig. 4, which portrays the 126 group means from the data of Robson and Vas (10) after classification by the proportion of animals surviving through the period of observation. This proportion is simple and therefore attractive, but undoubtedly conceals a great deal of information available in the mean (and SD) of the transformed survival times.

An attractive feature with functional transformations is that they allow a statement of the central tendency of any group's survival to be made on the original time scale. Table 10 gives

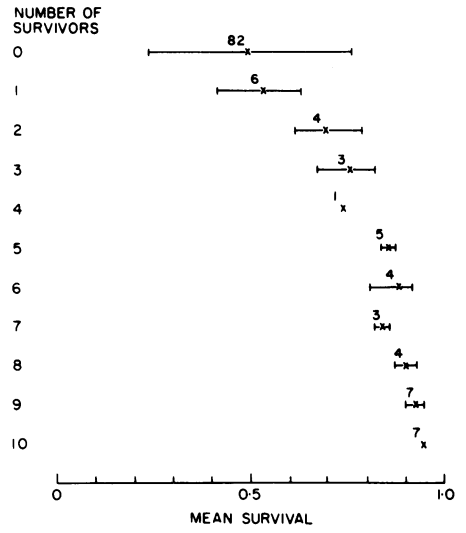

FIG. 4. *Mean survival in relation to the proportion of survivors. The 126 group means, after negative-exponential transformation with $\theta = 0.1$, $T = 28$, of the data of Robson and Vas (10) classified by the number of animals, out of 10, surviving to 28 days. The ends of the bars indicate the lowest and highest means; the crosses represent the mean of the means. The figure above each bar gives the number of groups.*

examples that are of special interest because animals of only two strains all died within 28 days, so that "mean survival" could not have been calculated. A further advantage is that interval estimates can also be made on the original time scale (see examples in Table 6).

The device for handling survivors appears to work satisfactorily. However, that concerned with assumed points of death, although appropriate for estimation of means, could be improved for estimating SDs (without affecting means). Instead of giving a death in the interval $(t - 1, t)$ a single score $[z(t)$, as defined above], it should be "split," counting one-half scored $f(t - 1)$ and one-half scored $f(t)$. Where all deaths in a group occur during a single day, splitting would be required to avoid a zero estimate of SD and hence a spurious idea of precision. If deaths occur over only a few consecutive days, splitting would be preferable, but otherwise is unnecessary. This device used on the data of Table 1 improves slightly the case in favor of transformation F.

In none of the situations examined has the choice of parameters (θ and T) proved critical, although there is some evidence in favor of letting θ be 0.01 when T is taken as the actual period of observation. In two cases, the period of observation was artificially changed, and the effects of this on the transformation itself need to be understood. In fact, two choices of parameters, θ_1, T_1 and θ_2, T_2, yield the same transformation when $T_1 \log \theta_1 = T_2 \log \theta_2$. Thus our analyses of the 28-day data of Robson and Vas (10) can be considered as having been with the appropriate value of T, i.e., 28, but two values of θ (0.1 and 0.0464), and those of their data truncated at 21 days as again with the appropriate T ($= 21$) but $\theta = 0.1$ and 0.1778.

A computer program has been written (in Fortran IV) to perform all the calculations for one or more groups. It allows specification of θ and T and can be used either from a terminal or in batch mode. Its running cost is low, and it stores output and leading input data to permit statistical manipulation of the means. The output includes point and interval estimates, with variable degrees of confidence, on the original time scale. However, it still bases calculation on $z(t)$ as distinct from splitting.

Even without a computer program, the calculations are quite manageable on a desk or pocket calculator, although tedious, as the author discovered in carrying out most of those used here. The first step is to prepare a table of the values of $z(t)$ for all t in the interval $[0, T]$; this takes time and has to be accurate; at least four decimal places are recommended. For survivors to T, half should be scored unity and the other half $f(T)$, not $z(T)$. The rule for an odd number of survivors, say $2s + 1$, is to score unity for $s + 1$ of them and $f(T)$ for the remaining s.

CONCLUSION

As has been appreciated for many years (1, 11, 12) the length of time each animal in a challenge experiment survives is valuable information. To make full use of it, some transformation of survival time is required. Empiric transformations (8, 9) may be adequate in the specific circumstances for which they are devised, but cannot be recommended since they cannot be generalized. The reciprocal transformation has been used (6, 11, 12), but fails to meet certain important criteria that are met by negative-exponential transformations; and these last perform as well as the empiric transformations even in their specific situations.

The use of the negative-exponential transformation is thus recommended. The choice of parameters (θ, T) is not critical, but users have to bear in mind the need to compare like with like. Thus, in any series of comparisons, the same transformation must be used; and, in any publication or exchange of information, the values of θ and T that have been used must be quoted.

The use of this transformation allows the application of the powerful ANOVA (4, 13) to data from well-designed challenge experiments. It

also enables the development of designs to extend the range of validity of the conclusions drawn from them (3). Experimental units have to be alike in all factors other than those deliberately made different between treatments. They might be made homogeneous in, say, body weight; but then the conclusions could apply only to animals of this weight. Alternatively, the units might be constrained to have closely similar distributions of body weight; it would then require only a minor change in design, without increasing the total number of animals in the experiments, to permit weight to be included as a factor, and so for the effect of this factor—and all the interactions of other factors with it—to be evaluated.

The Fortran program can be made available, for a nominal charge, from Stephen I. Vas, Professor of Medical Microbiology in the University of Toronto and Microbiologist-in-Chief, The Toronto Western Hospital, 399 Bathurst Street, Toronto, Ontario, Canada M5T 2S8. One last comment: this paper has not sought to enter the long-standing controversy over the relative merits of prolongation of life within the period of observation and survival beyond that period, nor to examine the relevance of animal experiments to human vaccination.

ACKNOWLEDGMENTS

I am grateful to Stephen I. Vas for bringing this interesting problem to my attention when he was Professor and Chairman of the Department of Microbiology and Immunology at McGill University and for his encouragement once the basic idea had been propounded. The program was written by Godfrey Iloabachie and formed part of the practical work for his M.Sc. (Applied) in the School of Computer Science at McGill. Funding for Mr. Iloabachie's work on this project and for his computing time was provided by Professor Vas from a grant to him by the Medical Research Council of Canada.

LITERATURE CITED

1. **Boyd, W. C.** 1966. Fundamentals of immunology, 4th ed., p. 717–722. Interscience, New York.
2. **Cochran, W. G., and G. M. Cox.** 1967. Experimental designs, 2nd ed. Wiley, New York.
3. **Cox, D. R.** 1958. Planning of experiments. Wiley, New York and London.
4. **Fisher, R. A.** 1925. Statistical methods for research workers. Oliver and Boyd, Edinburgh.
5. **Fisher, R. A.** 1935. The design of experiments. Oliver and Boyd, Edinburgh.
6. **Gard, S.** 1940. Encephalomyelitis of mice. II. A method for the measurement of virus activity. J. Exp. Med. **72**:69–77.
7. **Hill, A. B.** 1961. Principles of medical statistics, 7th ed., p. 1. The Lancet, London.
8. **Ipsen, J.** 1941. Contributions to the theory of biological standardization. Nyt Nordisk Forlag Arnold Busck, Copenhagen.
9. **Ipsen, J., Jr., J. H. Brown, E. G. Gerwe, L. Greenberg, J. L. Hendry, J. Hohenadel, P. Masucci, C. Newman, and E. M. Taylor.** 1953. Precision of potency assay of alum-precipitated tetanus toxoid in mice. J. Immunol. **70**:171–180.
10. **Robson, H. G. and S. I. Vas.** 1972. Resistance of inbred mice to *Salmonella typhimurium*. J. Infect. Dis. **126**:378–385.
11. **Smith, C. E. G., L. H. Turner, and P. Armitage.** 1962. Yellow fever vaccination in Malaya by subcutaneous injection and multiple puncture. Bull. W.H.O. **27**:717–727.
12. **Smith, C. E. G., and D. R. Westgarth.** 1957. The use of survival time in the analysis of neutralization tests for serum antibody surveys. J. Hyg. **55**:224–238.
13. **Sokal, R. R., and F. J. Rohlf.** 1969. Biometry. Freeman, San Francisco.