STATISTICS IN MEDICINE
Statist. Med. **18**, 3401–3408 (1999)

A SELECTION OF MANTEL'S CONTRIBUTIONS TO LABORATORY RESEARCH

SAMUEL W. GREENHOUSE*

The Biostatistics Center, George Washington University, 6110 Executive Boulevard, Suite 750, Rockville, MD 20852-3903, U.S.A.

SUMMARY

Nathan Mantel served as a statistician in the National Cancer Institute of the NIH for more than 25 years. During this period he was a consultant to many laboratory investigators, epidemiologists, clinicians and incidentally statisticians. These consultant activities in the design and analysis of data led to many of the research problems that he was engaged in as an independent investigator. Among these were many contributions to quantitative methods in laboratory research. This report discusses six such issues.

INTRODUCTION

The title is somewhat misleading since some of the results discussed are of mathematical interest not directly related to applications in laboratory research but motivated by it. It is necessary to point out that Mantel was involved in many more consultations with laboratory scientists during his more than 25 years at the National Institutes of Health (NIH) than is recounted here. Many of his original and imaginative analytic techniques were published in subject matter journals.

The list of topics to be considered follows: (i) the probit plane; (ii) estimating the standard error of a mean in small samples; (iii) accuracy of blood cell counts in the haemacytometer; (iv) safety testing; (v) litter matched data; (vi) light bulbs.

THE PROBIT PLANE[1]

Dr. Abraham Goldin was a biologist in the National Cancer Institute who had been exploring the effect of certain pharmacologic agents in prolonging the survival time of mice with Leukaemia L-1210. He was using lethal doses of aminopterin which affected the tumour but which was also toxic to the host animal. Previous experiments had shown that the mortality response due to the lethal doses of aminopterin might be reduced by the administration of citrovorum factor (CF). At this point Goldin needed advice on how to design and analyse experiments that would account for the joint administration of doses of CF and aminopterin. Goldin had been experienced in doing classical bioassay, that is, analysing dose response data of per cent mortality against log

* Correspondence to Samuel W. Greenhouse, The Biostatistics Center, George Washington University, 6110 Executive Boulevard, Suite 750, Rockville, MD 20852-3903, U.S.A.

CCC 0277–6715/99/243401–08$17.50

dose of aminopterin. However, he needed advice on the design and analysis of data derived from experiments in which he would administer concomitantly the metabolite and the anti-metabolite. Mantel, working jointly with Cornfield, had just completed developing some methodology on probit analysis and agreed to serve as consultant to Goldin's group.

Mantel first developed a formula for relating the probit (a normal deviate plus five) of mortality to the log of the ratio of the dose of aminopterin to the dose of CF administered plus any endogenous amount in the host (denoted by C). This linear relationship seemed reasonable as a result of the previously observed linear relationship between the median lethal dose (LD50) of aminopterin and the dose of concomitantly administered citrovorum factor (CF). Thus if Y is the probit of response, this equation is

$$Y = a + b \log A/(C + \mathrm{CF})$$

where A is moles of A/kg, CF is moles CF/kg administered simultaneously with A, C is endogenous reserve of protection, expressed as moles CF/kg, b is probit slope constant, expressed as a 10-fold increase in $A/(C + \mathrm{CF})$ and a is probit of per cent mortality when $A = C + \mathrm{CF}$.

Note that for a given per cent mortality, p, the relationship between A and $C + \mathrm{CF}$ is linear with slope, K_p, given by antilog $\{(Y - a)/b\}$, which is consistent with the experimental observation that the LD50 is linear with the simultaneously administered dose of CF. Mantel calls the above equation a modified probit plane. Statistically, it enables one to obtain the effect of a dose of aminopterin on the per cent of animals killed for a given level of protective citrovorum factor. However, more importantly, with the use of this formula Goldin was able to increase the anti-tumour activity of the aminopterin by increasing its dose to levels which were lethal without the administration of CF thereby increasing the survival time of these tumour bearing animals. Goldin made other uses of the probit plane in order to explore the effect of delayed administration of CF on the antileukaemic action of A and also to investigate the specificity of action of A by varying the time of treatment and dosage schedule.

In order to validate the model, an interesting experiment was carried out following a rather complex design. Six levels of CF were chosen (0, 3·96, 7·93, 15·9, 31·7, and 63·4) 10^{-6} moles CF/kg. At four of these levels, five doses of A were administered and at two CF levels six doses of A were given, making a total of 32 $A \times \mathrm{CF}$ combinations. Ten animals were randomly assigned to each treatment combination. Goldin's laboratory carried out the random selection of 320 animals and their random assignment to a treatment combination in about three hours.

RAPID ESTIMATION OF THE STANDARD ERROR OF THE MEAN IN SMALL SAMPLES[2]

Mantel presented this short cut procedure for obtaining the standard error of a sample mean as an aid to laboratory experiments working with small groups of animals or for small samples in general. The method is simple and can be done very quickly. It divides the sample range by the sample size, n. Estimates of the standard error of the mean so obtained have a small bias for normal samples of size up to 15. To show that the procedure can be used more generally, Mantel also gives the bias for small samples taken from a rectangular distribution. As an illustration consider a sample of eight values; 47; 49; 51; 51; 57; 62; 64; 65. The range is 18 yielding an estimate of the standard error of the mean when divided by 8, 2·25. The ordinary procedure based on the computation of the sample deviation (7·18) yields 2·54.

For normal samples, the method can be derived from tables presenting the factors associated with the mean range (for example, table 27 in vol. 1 of *Biometrika Tables For Statisticians*) that multiplies the standard deviation of the population. Thus $E(\text{range})/c_n = \text{SD}$. Divide both sides by $\sqrt{n}$, substitute the sample range, r, for the mean range, and $r/c_n\sqrt{n}$ is an unbiased estimate of the standard error of the mean. Mantel observed that $c_n\sqrt{n}$ is approximately equal to the sample size, n, for small samples up to about 15. A similar argument can be made for those factors that apply to the mean range for samples from the rectangular distribution. Here the factor that is approximately equal to n is $\{(n-1)/(n+1)\}\sqrt{12n}$.

REJECTION PROCEDURES ON THE ACCURACY OF BLOOD COUNTS[3]

A haemacytometer is a chamber consisting of many large squares in which various blood cells are counted. It was customary to sample five cells in sequence. In order to get more uniform results, technicians would accept the fifth count only if it fell between the previous four. If it was outside the four counts, they continued to sample cells until they obtained one that was acceptable. Mantel studied the properties of this procedure. He also quickly surmised that the fifth count should fall between the previous four about 60 per cent of the time. He reasoned that if one considers the cumulative number of counts to be a cumulative probability distribution, it is then a number between 0 and 1. Since the mean range of a sample of n observations from a (0, 1) rectangular distribution is $(n-1)/(n+1)$, for $n = 4$, the mean range is 0·6.

In order to obtain more specific properties of the procedure, he, together with his co-authors, devised an experiment. It was well known that cell counts among squares were distributed as a Poisson variable. Instead of using actual data, he advised that they perform a simulation by sampling counts from a Poisson population constructed from a set of random numbers where the expected mean count in each square was 80. This approximates the average count of erythrocytes. They planned three different counting procedures: T1, five successive counts were taken; T2, four counts were taken – if the fifth fell in the range of the four, it was accepted, if it was outside, another square was taken and accepted regardless where it fell; T3, four counts were taken and the fifth was not accepted until one was obtained which fell within the acceptable range. A total of 100 runs were made recording the counts in five successive squares for each run, that is, T1. If the fifth count fell outside the range of the first four, additional squares were sampled to obtain T2 and T3.

The results were as follows: there was complete agreement in 62 of 100 runs; in the remaining 38 runs, there were 26 where T1 gave a closer estimate of the mean of 80 than T3, and there were 11 where T3 gave a better estimate of the mean than T1. This ratio of 26 to 11 is statistically significant. It also turned out, as one might expect, that the sample variance was increased in T3 and T2 as compared to T1, thus making the estimates of the mean less accurate whenever rejection techniques were employed.

SAFETY TESTING OF CARCINOGENIC AGENTS[4–7]

Mantel probably did some of his best work in a series of papers dealing with this difficult subject. In the earliest paper with Bryan he also did some of his best writing; clear and direct in his exposition and well-reasoned and surprising in his judicious philosophic position that he took with respect to the concept of a 'safe dose'. In addition, in this paper he argues quite effectively for his procedure of getting to this dose, which had been used in the past by the Food and Drug

Administration. He nicely presents the method in a very simple situation and successively extends its use in more general experimental conditions. In this early period, there was a good deal of controversy concerning definitions and procedures related to the subject. Mantel's views held up rather well over the years even to the point of having his definition of a 'virtual safe' dose accepted as policy by the Food and Drug Administration for a period of time.

The problem is the determination of safe dose levels of an agent suspected to be carcinogenic in humans. The usual setting is when the agent is an ingredient added to food, in most instances for good reasons, but in relatively small amounts. Data relevant to the issue usually derive from assays performed on animals at much higher doses than those imbibed by humans. Thus the problem resolves into the hazardous area in statistical inference of extrapolation down to low doses.

In the 1961 paper with Bryan, Mantel points out that it is useless to argue for absolute safety which may have as many problems as any non-zero risk dose. Rather he refers to a conservative 'safe' dose associated with a risk of 1 per 100 million for most ordinary conditions. In certain situations, however, this might be increased to 1 per million, say. He defends the 1/100 million dose by pointing that although this may seem to be very conservative it may not be when one takes account of other potentially carcinogenic factors and their interactions operating in one's environment. Clearly, there is no way in which this dose or 1/10 million or even 1/million, 'both because of practical considerations and statistical variation' can be 'made directly but must be by extrapolation from observed data'. He then presents three essential components for what he states to be a conservative approach: (i) the acceptance of an 'arbitrary' definition of 'virtual safety', taken by him to be 1/100 million – however he points out that 'other definitions of 'virtual safety' may be employed as conditions require'; (ii) the acceptance of a definition of an 'arbitrarily high level of statistical assurance taken by him to be at the 99 per cent level; (iii) the use of a rule of extrapolation by use of an 'arbitrarily' shallow slope. It is at this point that Mantel makes his unique contribution to the subject by proposing an extrapolation slope of one normal deviate or probit per common log (or ten-fold change in dose). Furthermore, Mantel differed from other statisticians with regard to the starting point of the extrapolation. He did not start from the observed rate of tumour occurrence if there was only one dose level or from the lowest rate associated with the lowest dose in a bioassay experiment. He calculated the upper limit on the true rate, P, of tumour incidence for the which the probability of observing as few tumours or less is 1 per cent (that is, the 99 per cent assurance rule). We present the simplest illustration of the method based on one dose. For more general and more realistic examples consult the paper.

Assume that a given dose of an agent yielded no tumours in 100 experimental animals. We seek that value of P, the true tumour rate that could give rise to zero tumours with a probability of 0·01. Solving for P in the equation

$$(1 - P)^{100} = 0{\cdot}01$$

yields $P = 0{\cdot}045$ or 4·5 per cent.

Thus we conclude that the observed result of no tumours in 100 animals is consistent with a true tumour rate of 4·5 per cent. The normal deviate of 0·045 is $-1{\cdot}695$, or a probit value of 3·305. Now the normal deviate of 1/100 million is $-5{\cdot}612$ or a probit value of $-0{\cdot}612$. Thus the upper 99 per cent confidence limit on the observed zero rate is 3·917 (or note the same value in probits) above the safe dose. Hence to have a decreasing slope of one normal deviate per common log (as postulated by rule (iii) it is necessary to decrease the log dose by 3·917 logs to get to the 'safe' level. Since the antilog of 3·917 is about 8300, the 'safe' dose is 1/8300 times the dose tested.

As indicated, their early paper extends the illustrations to a bioassay design and to an experiment of an experimental agent and controls. In an Appendix, Mantel presents an interesting alternative procedure when the experiment involves multiple dosages. He sets up multiple assays obtained from the lowest, another from the lowest two doses, another from the lowest three doses, etc. This gives rise to a series of safe doses from which the largest is selected as the 'safe' dose. In a later paper with Bohidar, Brown, Ciminera and Tukey,[7] Mantel writes on an improved 'Mantel–Bryan' procedure, where extensions are made to account for different sets of assays such as from different laboratories. (An early use of meta-analysis?)

Finally, it should be mentioned that at some times in the past, several agencies, while not establishing Mantel's 1/100 million safe dose criterion as policy, did accept its use. Today a more favoured safe dose is 1/million.

LITTER MATCHING[8,9]

Two papers should be considered together: the first in 1977 by Mantel, Bohidar and Ciminera (MBC)[8] and the second in 1979 by Mantel and Ciminera (MC).[9] The problem was the testing of a chemical compound for carcinogenicity. At the time it was common practice to match an animal assigned to an agent with a control animal from the same litter. In MBC, the data (taken from a larger set) consisted of 50 litters of female rats, one of which was assigned to a treatment and the other two served as controls. Conventional analysis compared the frequency of tumours among the treated animals with that in the controls. For this case, Mantel chose to employ time-to-response methods to analyse the data. The procedure adopted for this analysis was the Mantel–Haenszel approach for time-to-response data in general and survival analysis in particular (see Mantel[10]). Essentially this method sums the events over all the time intervals in which an event occurs and compares the total with the sum of the expectations of each event and computes a chi-square value with one degree of freedom as the square of the difference between the two sums divided by the hypergeometric variance under the null hypothesis.

An analysis was conducted for treated versus control groups ignoring litter membership. The corrected chi-square (1 d.f.) was 7·63. When the analysis took account of inter-litter effects, the corrected chi-square was 5·99, a result which implies a less effective mode of analysis. At this point, Mantel comments that the nature of the data was such that an intra-litter analysis need not, and in this case, obviously did not, take account of all the inter-litter information in the data. The reason for this was that there were a number of litters where a positive response, or a loss to observation, occurred first, thereby leaving remaining animals in the litter without a contrasting litter mate. Such animals were called remnants. As an example, consider the litter where the animal assigned to treatment was lost to observation. Then the two control animals do not have a litter mate that can contribute to the treatment versus control comparison in that litter. At this point Mantel devised three alternative methods for incorporating remnants into respective cohorts of treated and control animals, analysing these remnant tables and combining them with the intra-litter analysis. In each instance, the resulting chi-square was greater than the analysis ignoring the matched litter data. This paper also contains an interesting little section illustrating 'extremely simple' computations necessary to carry out the inter-litter recovery of information.

In the second paper, Mantel produces a unified method where all animals are used in time-to-response data making full use of intra- and inter-litter information. The method makes use of the fact that as the interval of observation gets smaller and smaller so that only one event occurs in the treated or control groups, the Mantel–Haenszel method is equivalent to a ranking

procedure. If one assumes that the events arise from an exponential distribution then a score can be assigned to the rth rank the value of which is the expected value of the rth order statistic from an exponential distribution. These are the Savage scores; for the rth order statistic the score is

$$1/n + 1/(n-1) + 1/(n-2) + \cdots + 1/(n-r+1).$$

Mantel then merges the data from all litters 'and the merged data are used to assign a score to each treated or control animal, such score depending on when or whether the animal developed a tumor'. It then becomes possible to obtain the deviation between the treated score in a litter from its expectation in that litter and also to calculate the finite population variance for that litter. One then computes a 1 d.f. chi-sqaure by squaring the difference between the sum of the observed treatment scores within a litter and the sum of the expected treatment scores within the litter and dividing by the sum of the intra-litter variances. (A similar procedure is done ignoring litters by forming the sum of all observed treatment scores and the sum of all expected scores of treated animals. This expectation will be the same as the previous expectation if the number treated and control animals are the same in each litter).

The assignment of scores to each animal allowed all animals to contribute to the recovery of inter-litter information without the need to make arbitrary decisions on handling the remnant animals. With logrank scores as the measurement, the 1 d.f. degree of freedom chi-square allowing for litter effects was 9·48 and the chi-square ignoring litter effects was 8·34. Thus the score analysis led to some improvement in power.

In comparing the Mantel–Haenszel method in the MBC paper with the logrank score method in the MC paper, Mantel points out: (i) the latter maskes objective use of full and proper recovery of information; (ii) the latter is the method of choice if the litters are of arbitrary and varying size with arbitrary allocation to treatment or controls in which case none of the methods given in MBC is appropriate; (iii) the score method does not yield an estimate of relative risk whereas the MH method does; and finally, (iv) an advantage of the former (MH) is that it permits continual updating as the study progresses.

As usual Mantel inserts many interesting ideas not directly related to the main topics in both papers. In particular, he gives a short discussion as to why the MH procedure is known as 'logrank'.

LIGHT BULB STATISTICS[11, 12]

There are two papers on light bulbs, one, in 1966, is joint with Pasternack[11] (MP), the second appeared in 1969[12] (M). The results obtained in these two papers are too numerous to review so that this report will merely describe the general contents. Mantel makes very clear all results are known; what he presents are ingenious, heuristic ways of obtaining them.

The basic concept which gives rise to all the derivations is an idealized light bulb that does not decay while in use but is subject to a constant probability of failure. The distribution of the time to failure for this bulb, assuming the expected lifetime is one, is the standard exponential or waiting time distribution

$$f(t) = \exp(-t),\ 0 < t < \infty\,.$$

If we let $t = \chi^2/2$, this distribution becomes the chi-square density with 2 degrees of freedom.

In the first paper he derives the following distributions: sums or differences of products of pairs of independent standard normal deviates; the difference of two chi-squares, each with even

degrees of freedom, or what amounts to the same thing, the difference between two total waiting times; the ratio of two independent chi-squares, each with even degrees of freedom, or, the ratio of two total waiting times; distributions associated with a simple exponential growth process. In the second paper, he considers two geometric probability problems posed by Gnedenko, namely, (i) to find the probability that three line segments formed by three points chosen at random on the interval $(0, a)$ can form a triangle, where the lengths are taken as the distance between 0 and each of the three points; (ii) to find the probability that a triangle can be formed from three segments of a rod broken at two points chosen at random. Mantel solves the second problem first making use of the light bulb concept and immediately generalizes to the formation of an n-sided polygon from n successive individual segments. He then proceeds to solve a generalization of the first problem.

I am tempted to present three specific results in detail: the sums of products of pairs of independent standard normal deviates; the simple exponential growth process, and a problem in geometric probability posed by Gnedenko. However, discussing all three issues would make this report overly long so that I only present the second in order to impart the flavour of his reasoning.

We start with a single 'idealized' organism that has a constant probability of dividing into two. The time to division then follows the standard waiting time distribution. Now assume the two daughter organisms at time of birth are independently subject to the same division process which continues through future generations. The question raised is what is the distribution of the time, T, required for the population to attain $n + 1$ organisms:

$$T = T_1 + T_2 + T_3 + \cdots + T_n$$

where T_i is the time needed to go from i to $i + 1$ organisms. The above equation for T is the same but in reverse order of the linear combination of n light bulbs burning simultaneously where T now represents the time required for all bulbs to be extinguished. Each T_i was shown before to be a 2 d.f. chi-square. Furthermore, whether we consider the failures of light bulbs or the growth of organisms, T is the nth order statistic for n independent waiting times. It is then easy to find the density and distribution functions by using the usual order statistics formulae. Thus for T representing the time for the reduction of n bulbs to 0 or the time for the increase of 1 to $n + 1$ organisms, we have

$$f(T) = n\mathrm{e}^{-T}(1 - \mathrm{e}^{-T})^{n-1}$$

$$F(T) = (1 - \mathrm{e}^{-T})^{n}.$$

The section then goes on to derive the distribution of population size, say n, starting with n_0 organisms, and also gives the probability of getting exactly n organisms at time T.

Mantel subsequently published a short note[13] in the *American Statistician* deriving the LaPlace distribution of

$$Y = X_1 X_2 + X_3 X_4$$

by characteristic functions.

Clearly, although the light bulb papers did not arise from any specific laboratory problem, the procedures used to obtain these results could prove useful and are of interest in their own regard.

Note: There are instances where the wording of statements in this report is the same as in the original Mantel paper.

REFERENCES

1. Goldin, A., Mantel, N., Venditte, J. M. and Greenhouse, S. W. 'An analysis of dose-response for animals treated with Aminopterin and Citrovorum Factor', *Journal of the National Cancer Institute*, **13**, 1463–1471 (1953).
2. Mantel, N. 'Rapid estimation of standard errors of means for small samples', *American Statistician*, **5**, 26–27 (1951).
3. Schneiderman, M. A., Mantel, N. and Brecher, G. 'The effect of rejection procedures on the accuracy of blood counts', *American Journal of Clinical Pathology*, **21**, 973–978 (1951).
4. Mantel, N. and Bryan, W. R. 'Safety testing of carcinogenic agents', *Journal of the National Cancer Institute*, **27**, 455–470 (1961).
5. Mantel, N. and Schneiderman, M. A. 'Estimating "safe" levels, a hazardous undertaking', *Cancer Research*, **35**, 1379–1386 (1975).
6. Mantel, N. 'Limited usefulness of mathematical models for assessing the carcinogenic risk of minute doses', *Archives of Toxicology*, Suppl. **3**, 305–310 (1980).
7. Mantel, N., Bohidar, N. R., Brown, C. C., Ciminera, J. L. and Tukey, J. W. 'An improved Mantel–Bryan procedure for "safety" testing of carcinogens', *Cancer Research*, **35**, 865–872 (1975).
8. Mantel, N., Bohidar, N. R. and Ciminera, J. L. 'Mantel–Haenszel analyses of litter-matched time-to-response data, with modifications for recovery of interlitter information', *Cancer Research*, **37**, 3863–3868 (1977).
9. Mantel, N. and Ciminera, J. L. 'Use of logrank scores in the analysis of litter-matched data on time to tumor appearance', *Cancer Research*, **39**, 4308–4315 (1979).
10. Mantel, N. 'Evaluation of survival data and two new rank order statistics arising in its consideration', *Cancer Chemothrapy Reports*, **50**, 163–170 (1966).
11. Mantel, N. and Pasternak, B. S. 'Light bulb statistics', *Journal of the American Statistical Association*, **61**, 633–639 (1966).
12. Mantel, N. 'More light bulb statistics', *American Statistician*, **23**, 21–23 (1969).
13. Mantel, N. 'A characteristic function exercise', *American Statistician*, **27**, 31 (1973).

