An Improved Mantel-Bryan Procedure for “Safety” Testing of Carcinogens

Nathan Mantel, Neeti R. Bohidar, Charles C. Brown, Joseph L. Ciminera, and John W. Tukey

Summary

A published method by Mantel and Bryan for calculating “safe” doses of carcinogens is updated by incorporating several improvements. These improvements include more effective procedures for taking into account any spontaneous tumor rate and for combining data at several dose levels. An added feature is that it permits the combining of data from several experiments by postulating that it is only the spontaneous rate that differs between experiments. The improved method is illustrated with data from five hypothetical experiments, using a risk level of $10^{-8}$, a conservative slope of one probit or normal deviate per tenfold dose increase, and a nominal assurance level of 99%. The hypothetical experiments were geared to bring out particular points as, for example, the applicability of the model in the absence of control data. A large variety of issues involved in the determination of “safe” doses are discussed, including questions of experiment design and extrapolation between species. A statistical appendix is provided, laying the framework for the calculating procedure and detailing complications therein.

The “safe” dose approach helps resolve certain dilemmas in questions relating to food additives. A “no-detectable-level” prescription for chemical residues may be dangerous to the public where detection techniques are insufficiently sensitive, but it can become far too restrictive as exquisitely sensitive detection techniques are developed. Only levels in excess of the “safe” dose would require detection. Calculated values for the “safe” dose could be updated and increased as more clear evidence of safety becomes available.

Introduction

A conservative approach for resolving the question of potential carcinogenic effects of chemicals, particularly food additives, was given in 1961 by Mantel and Bryan (3). These authors were concerned with the possible risk attendant on the requirement that an agent have no carcinogenic hazard. The impossibility of demonstrating the hazard to be actually zero could lead to the acceptance of demonstrations of apparent safety as demonstrations of absolute safety; the resulting risk to large human populations could be unconscionably high.

There is a particular dilemma relating to residual levels of additives in food that the Mantel-Bryan approach helps to resolve. It is that if no detectable level of the chemical is permitted in the food, this may, nevertheless, be dangerous for the public if detection procedures are insufficiently sensitive. It can, however, become an unduly restrictive prescription should extremely refined detection procedures subsequently be developed. The Mantel-Bryan approach is directed to identifying “safe” doses for the chemical, and it is only residual levels in excess of these that would require detection. The approach is characterized by the property of allowing higher calculated values for the “safe” dose as more persuasive evidence of safety is brought into play; there is thus encouragement to perform larger and better experiments.

The basic premise of the Mantel-Bryan approach was that under every agent was regarded as a carcinogen. The problem then became one of determining some dosage level for the agent for which one could with assurance assert that the risk was below some arbitrarily low level, e.g., 1 in 100 million or $10^{-8}$. (By the Mantel-Bryan approach the $10^{-8}$ risk is not accepted as suitably low, although it may seem to be virtually zero; the approach is directed to setting the risk at less than $10^{-8}$ but with guarantee, under the assumptions made, that the risk at the calculated “safe” dosage will not exceed $10^{-8}$.) Actual direct testing for determining “safe” doses was not feasible, and the Mantel-Bryan approach depended on the use of several techniques to get at conservative estimates of the “safe” dose. Thus, whatever might be the observed tumor frequency in a group of test animals, the data would be regarded only as establishing some upper limit on the risk, say as at the 99% assurance level, at the test dose used; no tumors among 100 animals would be taken as establishing only that the actual hazard was probably less than 4.5%. The next technique was one of then extrapolating downward by a conservative rule to expected outcomes at yet lower dosages using a carefully chosen shallow slope not derived from the data. This
downward extrapolation Mantel and Bryan exemplified with the use of a probit slope of 1 normal deviate per 10-fold increase in dose. (The reader is referred to the original report and to the references therein as to the technical meaning of such a probit slope. Clarifying information is given also in Appendix A.) A consequence of the use of this extrapolation rule together with the use of the 99% assurance level is, for example, that when the data consist of no tumors among 100 test animals, the $10^{-8}$ “safe” dose is $1/8300$ of the test dose used. Incidentally, use of the probit slope of unity is not intended to convey that the actual response slope is unity or even that it is linear in the probit scale at low dosage levels. The Mantel-Bryan assumption here is that the true dose-response curve does not drop off less rapidly than that corresponding to the unit probit slope, in consequence of which the true $10^{-8}$ dosage cannot be less than that yielded by such shallow slope extrapolation.

This report improves on the Mantel-Bryan report in 3 particulars, the needs for 2 of which were indicated in their report. The initial report provided an unnecessarily conservative procedure for handling the case of spontaneous tumors arising in control animals; such conservatism would be only mild where the control tumor rate is small but could be substantial under circumstances of high spontaneous rates. Mantel and Bryan had illustrated their procedure with the results of testing with a potent carcinogen that elicited tumors so early that spontaneous tumors were not seen; in fact, however, spontaneous tumor rates in long-term testing can be 40% or even higher.

Yet another example of unnecessary conservatism for the original Mantel-Bryan procedure arose in the treatment of data obtained at several dose levels. Mantel and Bryan, however, indicated in their appendix an alternative method, reflecting discussions with J. Cornfield, which was intuitively more satisfying, avoided the conservatism of the “justifiable” combination approach, but required more complex calculations. That alternative method is implemented in this report, in conjunction with the modified handling of control data.

The last new particular of the present report is that it permits the handling of several sets of independent data. Such sets might represent the results of testing on different occasions or the separate results for male and female animals tested on the same occasion. By the original Mantel-Bryan procedure a separate “safe” dose would be calculated for each set of data, leaving open a variety of problems in their interpretation; should some combined “safe” dose be the lower, the higher, or the average of 2 “safe” doses, or might it not sometimes be higher than either? The new approach permits calculating a combined “safe” dose as well as separate “safe” doses on the assumption that the true “safe” dose is really the same in each experiment. A consequence is that the calculated “safe” dose can be updated as the results of new experiments are added to the data set. Unless there is clear evidence of a real difference among the several experiments, the combined “safe” dose is the indicated choice.

Results of application of the improved Mantel-Bryan procedure will now be given, together with certain insights gained therefrom. The technical aspects of the new procedure are deferred to appendices. For discussion of the practical merits of the Mantel-Bryan approach the reader should see the papers of Schneiderman and Mantel (Ref. 6; Footnote 2); in Schneiderman and Mantel, a wide variety of other aspects of the “safe” dose problem are discussed.

Description of the Improved Mantel-Bryan Method

For a single set of data the 1st 2 modifications indicated in the introduction are incorporated simultaneously. This is done by postulating the 3-parameter model described by Finney (2) for handling the problem of natural mortality. The parameters are specifically: an intercept parameter, $a$, which represents the normal deviate corresponding to the induced response rate attributable to drug action by a unit dose (log dose equal to zero); a slope parameter, $b$, which represents the increase in the normal deviate of the response rate per 10-fold increase in the dose (unit increase in common log of dose) and which is not estimated from the data but rather is set at its conservatively placed value, e.g., unity; and a spontaneous response parameter, $c$, which represents the expected response rate for untreated animals. Under the model of independent spontaneous and induced responses, the expected response rate at a dosage with common log, $X$, is given by:

$$P(a,b,c,X) = c + P(a,b,0,X) - cP(a,b,0,X)$$

in which

$$P(a,b,0,X) = \int_{-\infty}^{a+bX} (2\pi)^{-n} \exp (-\frac{1}{2} Y^2) \, dY$$

(Note that this formulation yields $P(a,b,0,X) = [P(a,b,c,X) - c]/(1 - c)$, the standard Abbott (1) formula for correcting for spontaneous response used in Ref. 3. An alternative formulation here might have been that the spontaneous rate represents the response to the load of the test agent or its equivalent already in the environment. The total load for an individual or animal is then the sum of its administered dose and its environmental load. Under the independence model here assumed, however, the environment is substantially free of the agent under test. This does not preclude the possibility that the environment may contain other active agents including agents capable of potentiating, synergizing, or even antagonizing the effects of the test agent. The results of any analysis should thus be taken as relating to a similar environment.)

Implementation of the alternate method suggested in the appendix of Ref. 3 requires that the data at all dose levels, including control data, be fitted to get the joint maximum likelihood estimate of $a$ and $c$, $b$ being already specified to have value unity. Specific interest centers now on a measure, the loglikelihood, of the correspondence between the actual and fitted results; the loglikelihood value can be determined even if in certain instances the $a$ and $c$ values are not estimable. What is next needed is an upper limit on the intercept parameter, $a$. An approach to setting limits on single parameters or single parameter functions is given by
Mantel and Patwary (5). They suggested that the value of such a limit should be that value which, when coupled with maximum likelihood fitting of the remaining parameters (or restricted maximum likelihood fitting subject to the specified value of the parametric function), would lead to a just significant worsening in fit. They further suggest use of an asymptotic result under which twice the reduction in loglikelihood is distributed like \( \chi^2 \) with a single degree of freedom (d.f.). In the present instance, because of 1-sidedness, the 99% upper limit on a is given by that value for which the reduction in the loglikelihood in 5.412/2 = 2.706, where 5.412 corresponds to a cumulative probability of 0.99 in the distribution of the single d.f. \( \chi^2 \). (We note here that, because we are using an asymptotic result, any assurance levels based upon it will be approximate. Experience with loglikelihood calculations, however, gives good reason to expect this approximation to be of quite adequate quality.) Further, although the method used postulates a spontaneous response rate, it can be used even in the absence of control data. This is because under the model used the expected response at every dose level is sensitive to the spontaneous rate parameter, \( c \).

The step just described for obtaining an upper limit on the intercept parameter, \( a \), using all the data is next repeated deleting the data at the highest dose level, to give an alternative upper limit on \( a \). This process of deletion is successively followed until an upper limit \( a \) is determined in which the only non-zero dose level datum used is that at the lowest dose (a reverse process of adding dose levels would be equivalent). Save for a qualification to be given shortly, the overall “safe” dose is obtained by selecting the lowest of the upper limits on \( a \) and sliding back with the conservative slope of unity to the normal deviate corresponding to the desired limiting value on the risk level, e.g., if the limiting risk level is \( 10^{-8} \) with normal deviate \(-5.612\), the smallest of the upper limits on \( a \) is \( a^* \), and the conservative slope used is \( b \), then the desired overall “safe” dose is given by the antilogarithm to the base 10 of \((-5.612 - a^*)/b\).

The rationale for making use of the lowest of the upper limits on the intercept, and thus of the highest of the alternative “safe” doses, is this. Suppose we have determined a “safe” dose using only the data at the several lowest dose levels, then make a recomputation adding the data at the next higher dose level. If the recomputed “safe” dose is lower we can explain it away on the basis that the true slope in the observable region is steeper than the conservatively assumed shallow slope; the added data have consequently raised the level of the fitted line and lowered the estimate of the “safe” dose. We would, on the other hand, accept an increase in the estimated “safe” dose on the basis that the added information relative to safety provided by the higher dose level data has outweighed the tendency for such data to give rise to reduced values for the calculated “safe” dose.

A qualification exists, however, against using such a rule blindly. The “safe” dose estimate may be increased by the use of higher dose data for reasons other than added information. If the response at higher doses has flattened out or even turned down, as for example might result from toxic or lethal drug effects obscuring carcinogenic effects, then estimates of \( a \) and its upper limit (and of “safe” dose) should not be made involving such doses. (Where the response rate at high doses turns down because of competing toxicity, a partial remedy, frequently used, and often essentially is to exclude from numbers at risk those animals dying early in the experiment.)

We come next to the question on the determination of “safe” levels from the data of several experiments. The model we have proposed for this is that the various data sets share a common intercept parameter, \( a \), the common slope parameter, \( b \), is put at a conservatively low value, but each data set is allowed to have its own spontaneous response parameter, say \( c_j \) for Data Set \( j \). When, for example, the data at all dose levels, all experiments, are considered, maximum likelihood estimates are made for \( a \) and the \( c_j \). We next determine that limiting \( a \) value that when coupled with its associated maximum likelihood estimates of the \( c_j \) gives rise to a just significantly worsened reduction in the loglikelihood as measured by a single d.f. likelihood \( \chi^2 \). The process is successively repeated deleting data at dose levels above a decreasing cutoff resulting in a final determination of \( a^* \), the smallest of the upper limits on \( a \), which, except for the qualification just described, is converted into an estimate of the overall “safe” dose. Since the dose levels used may vary from experiment to experiment, each time higher dose level data are deleted (or added when the reverse process is used), such deleted data may relate to only some of the experiments.

In application of these approaches to actual data, we have made overall “safe” dose determinations as well as separate experiment “safe” dose determinations. In some cases, the separate experiment determinations have differed somewhat, the “safe” dose for male mice in 1 example being considerably less than that for female mice. This result was not, however, altogether inconsistent with the possibility of equal true “safe” doses for the 2 sexes. The male data displayed a much higher spontaneous tumor rate, in consequence of which the intercept parameter estimate was subject to greater variability; by the approach taken the calculated “safe” dose is consequently reduced. This kind of issue we discuss in greater detail below.

An Illustration Using Hypothetical Data

The data, representing 5 hypothetical experiments, which we use for illustration are shown in Table 1. We have by preference used hypothetical data in order to bring out certain features of our method that might not have been apparent with real data. For example, in hypothetical Experiments 4 and 5 no control data are shown, in order to illustrate that the method is applicable even in the absence of such data. We have also contrived that the lowest dose appearing in Experiment 4 should appear in no other experiment; yet the result at this single dose level yields a “safe” dose estimate by itself (essentially on the premise of a zero spontaneous rate) and produces an increase in the estimated overall “safe” dose when coupled with the results of the remaining experiments at all lower dose levels.
Table 1
Hypothetical data for 5 experiments used in illustrating the improved Mantel-Bryan procedure for calculating "safe" doses

<table>
<thead>
<tr>
<th>Dietary dose level (ppm)</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
<th>Experiment 3</th>
<th>Experiment 4</th>
<th>Experiment 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>22/50</td>
<td>4/30</td>
<td>10/40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>16/50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>12/40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>15/40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>20/50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>25/50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>33/40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The presumptive dose levels for the various experiments are in parts per million (ppm) in the diet with the calculated "safe" levels given in Table 2 expressed in parts per trillion (ppt). Each of the 5 experiments has its own distinctive set of dose levels in the range of 1 to 100 ppm, but some dose levels are common to more than 1 experiment. The results of individual-experiment and combined analysis are given in Table 2. For each dose level in each experiment, the table shows the upper limit on the intercept parameter when data only at that and lower levels are considered, with data at higher doses excluded. The lowest of the upper limits for each experiment is italicized and its decoded "safe" level value is shown at the bottom of the table in ppt. Similarly, the upper limits on the intercept parameter, all experiments combined, are given in Table 2, Column 7, with the smallest upper limit italicized, then decoded below into a "safe" dose estimate. All calculations are based on a 10^-8 limitation on the risk, a conservative slope of unity, and a nominal assurance level of 99%.

While the individual-experiment "safe" doses range from 31 to 291 ppt, the overall "safe" dose of 331 ppt is greater than any of these. Although this overall "safe" dose is not much higher than that for the individual experiments, it does at least allow us to discount the low "safe" levels in some of the experiments, particularly those displaying high spontaneous response rates. In fact, the combined "safe" dose can sometimes be substantially higher than the component "safe" doses. From Table 2 we can determine what the results would have been had testing at the level of 20 ppm or higher been omitted; the highest individual experiment

Table 2
Results of the application of the improved Mantel-Bryan procedure to the individual-experiment data and to the combined data of Table 1

<table>
<thead>
<tr>
<th>Dietary dose level (ppm)</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
<th>Experiment 3</th>
<th>Experiment 4</th>
<th>Experiment 5</th>
<th>All experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.8374</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-1.2152</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-1.3447</td>
<td>-1.6369</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-1.4922</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>-1.4925*</td>
<td>-1.4129</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>-1.2312</td>
<td></td>
<td>-2.0765</td>
<td>-1.4666</td>
<td>-2.0246</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>-1.4974</td>
<td>-1.4974</td>
<td>-1.4974</td>
<td>-1.4974</td>
<td>-1.4974</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>-1.3457</td>
<td></td>
<td>-1.3457</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>-1.8047</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Risk Level: Calculated "safe" dose corresponding to a* (ppt)

| 10^-8 | 76 | 291 | 77 | 259 | 31 | 331 |

* Numbers in italics, lowest value (a*) in each column.
"safe" level would have corresponded to the lowest limit on $a$ of $-1.8372$ in Experiment 4, but the overall "safe" dose, corresponding to a lowest limit of $a$ of $-2.1044$ would have been $85\%$ greater.

In each column of the table, the $a^*$ value arises at an intermediate dose level, reflecting that our data were consistent with a true slope steeper than unity. We did not wish here to complicate issues by inserting instances in which the "safe" dose increased because the response curve had flattened out or turned down, making the qualification discussed above applicable. In the analysis of actual data we have encountered instances in which such effects influenced the individual-experiment "safe" dose but not the overall "safe" dose.

If we examine the result in Table 2 relative to Experiment 4, we see that it yields a relatively high "safe" dose even when only the data at the single dose level of 5 ppm are considered and that addition of these data to those for lower dose levels in other experiments has probably resulted in somewhat of an increase in the "safe" level, since the upper limit for $a$ has decreased. This has occurred despite the absence of control data for Experiment 4. For Experiment 5, however, the data at the lowest dose level alone yield a poor "safe" level since, in the absence of control data, the method of analysis treats the spontaneous rate as zero when no control data are available. However, as data at higher dose levels are added in Experiment 5, the "safe" level rises sharply, as now there is internal evidence for a substantial spontaneous rate.

Since the principal interest lies in the determination of "safe" doses, we have refrained from presenting here our estimates of the spontaneous rates. There are in fact a variety of such estimates for each experiment, these differing as between separate-experiment and combined analyses and also as the data at various dose levels are included or excluded.

Although we have illustrated here our results with a specified set of values for the risk level, the nominal assurance level, and the conservative slope, our present computer programming procedure is more general. Thus we can specify in advance 1 or more alternative values for these, further specifying which individual experiments or subsets of experiments should be analyzed in addition to the overall analysis.

### Discussion

In the hypothetical illustration the improved Mantel-Bryan procedure has provided an apparently reasonable overall resolution of the separate-experiment "safe" doses. This reasonable resolution occurred not only in the hypothetical case but also in the analysis of data from actual experiments. Where separate-experiment "safe" doses were not too different, the combined "safe" dose was typically substantially higher than the separate values, as would have occurred if the experiments could have been pooled, since less allowance for uncertainty would then have been required. Where "separate" values were distinctly different, the combined "safe" dose was generally heavily weighted towards the largest of the separate "safe" doses, sometimes exceeding it.

We give now a simple example to help explain the behavior of the combined "safe" dose estimation procedure (Table 3). It consists of the data for 5 experiments in which the control spontaneous rate, based on 50 animals, increases progressively from 0 to 80%, but the apparent induced rate [measured according to Abbott's formula by $(P_i - P_o)/(1 - P_i)$] remains constant at 10%. Along with the data for this hypothetical example, we show the individual experiment and the combined experiment "safe" levels, using as before a unit slope, 99% nominal assurance level, and $10^{-8}$ risk level.

We see here that the highest single "safe" dose attaches to Experiment 1, which is the one exhibiting the most clear evidence for, say, a carcinogenic effect. Yet only a low "safe" dose attends the results for Experiment 5, which displays questionable evidence of carcinogenicity. What is happening is that, although for Experiment 1 the data suggest a carcinogenic effect to exist, the upper limit on that effect is probably something under about 20%, say as due to a true zero control rate and a true induced rate of 20%. The upper limit on the tumor rate for Experiment 5 is considerably higher; the data are not too inconsistent with true rates of 70% for controls, 90% for treated, for an induction rate of 67%. The high spontaneous rate in Experiment 5 has led to such an imprecise indication of the induced rate that the resulting data are consistent with the possibility both of a zero induction rate and of a rather high induction rate of interest. The combined "safe" dose is higher than the individual experiment "safe" doses; the data for the last few experiments, while not highly informative in themselves, did add to the total information available and so served to raise the "safe" dose estimate. (Had the results of these 5 hypothetical experiments been blindly combined and treated as the result of a single experiment, showing the outcome of 100 of 250 for controls and 115 of 250 for treated, the calculated "safe" dose would have been only 1144 ppt. This same "safe" dose would apply for Experiment 3 alone were similar results yielded by a 5-times-larger study.)

The lesson here is that we should avoid uninformative experiments in establishing "safe" doses. If the intent were to conceal carcinogenic effects, an experiment that is less informative by virtue of the obscuring effect of a high spontaneous rate would seem ideal; but by the Mantel-Bryan approach, either in its original form or as modified

### Table 3

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Controls $(r/n)$</th>
<th>Treated 100 ppm $(r/n)$</th>
<th>&quot;Safe&quot; dose (ppt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0/50</td>
<td>5/50</td>
<td>1380</td>
</tr>
<tr>
<td>2</td>
<td>10/50</td>
<td>14/50</td>
<td>700</td>
</tr>
<tr>
<td>3</td>
<td>20/50</td>
<td>23/50</td>
<td>400</td>
</tr>
<tr>
<td>4</td>
<td>30/50</td>
<td>32/50</td>
<td>220</td>
</tr>
<tr>
<td>5</td>
<td>40/50</td>
<td>41/50</td>
<td>90</td>
</tr>
<tr>
<td>Combined</td>
<td></td>
<td></td>
<td>1700</td>
</tr>
</tbody>
</table>
here, such uninformative experiments would be penalized by the assignment of a low "safe" dose. For the case of several experiments, some highly informative, others with only minimal information, the improved Mantel-Bryan procedure would be much more influenced by the high "safe" doses of the informative experiment than by the low ones of the uninformative experiment. (However, if a low "safe" dose arises in a relatively informative experiment, it will be given high weight. This could indicate a situation in which there is a true difference in "safe" doses in the several experiments.) Since high spontaneous rates lead to less clear indications of induced tumor occurrence, certain kinds of experiments would be best avoided. These include experiments with animal strains having high spontaneous rates. Also it would be preferable not to count tumors that occur very late in life (but we can be free to observe them). Late-appearing tumors are, in principle, less important; they make for higher spontaneous rates, hence reduced information about induced rates, so leading to smaller values for the calculated "safe" dose. Moreover, the counting of late tumors could lead to misinterpretation of a life-prolonging effect of an agent as a tumor-inducing effect, since longer-lived animals would display more spontaneous tumors. Another desideratum of a "safety"-testing experiment is that it be conducted at several dose levels. If testing is at a single dose level then, whether the response rate is low or even moderate, there is always the possibility that some unrecognized or unconsidered competitive effect, say as from lethal toxicity, could be keeping the rate low.

Given that we have determined the "safe" dose by the improved Mantel-Bryan method, the next question is what we do with it. Strictly, a calculated "safe" dose applies only to the system for which it was determined, and so some extrapolation rule is needed for extension to another system or species, say man. The possibility for such extrapolation has been mentioned in the paper by Schneiderman and Mantel,2 in which they discuss the implication of the surface-area rule for equating dosages between species. That rule can be thought of instead as a two-thirds power rule. If man is 2800 times as massive as the test species, say a mouse, a comparable dosage, expressed in mg, would be 200 (= ca. the two-thirds root of 2800) times greater for man than for mouse. If dosage is expressed in units of dose per unit mass of the test animals, as in mg/kg, the "safe" dosage for the mouse would have to be divided by 14 (= ca. the cube root of 2800) to be applied to man. This as a 1st approximation would be appropriate barring any strain or species peculiarities. Schneiderman and Mantel have proposed, however, that if dosage is expressed in terms of dietary concentration, say ppm in the food (or water), then no adjustment need be made between species, on the assumption that dietary intake is already closely proportional to surface area to which food needs to maintain body temperature would be related. (This issue is discussed further in the paper of Schneiderman and Mantel.)

The "safe" dose estimation procedure coupled with the surface-area extrapolation rule might, however, provide insufficient protection against the possibility that somehow man is more sensitive than mouse even on a surface-area basis. This could require further reduction in setting "safe" doses for humans, as by a "safety factor"; such additional "safety factors" should not be confused with extrapolation factors or with use of the Mantel-Bryan method and will undoubtedly seem to some to be an unnecessary piling on of conservatism.

An interesting aspect of the extrapolation rule is this. To the extent that we are willing to accept any particular rule, we can combine experimental data obtained with different species in the same way as the procedure described herein permits the combination of experiments for a fixed species. This requires only that dosage be expressed in the proper unifying scale for the various species, e.g., ppm in diet or mg/sq m body surface. There are 2 reservations to this, however. The unifying scale is correct only as a 1st approximation and may be seriously wrong in the face of strain or species peculiarities. Further, the experiments in the various species may not be equivalent as, for example, in the sense of having equivalent ages of treatment initiation and/or durations of observations. Where experiments are equivalent, as when the strain or species remains fixed and the experiments performed are essentially identical, the method we describe here can be used to update an analysis to take into account the most current experimental results.

At this point it may be in order to respond to certain recurrent questions relating to the Mantel-Bryan procedure of which we have become aware. These relate to the use of the arbitrary shallow slope of unity, and it is our position that some such arbitrariness cannot be avoided. Alternative suggestions have been to use the slope suggested by the data, but after putting a statistical lower limit on such slope. In an insufficiently precise experiment relative to slope determination, such a lower limit can be extremely shallow (and could even be negative). Suppose the experiment is a precise one so that we determine the central slope with little error; under the Mantel-Bryan approach we cannot extrapolate into the low dose range with any resulting slope, which is presumably a steep slope, since we cannot assume that the slope remains steep. Another tempting suggestion is that we might use the arbitrary shallow slope only to the left of the lowest dose in the experiment but use the sample slope or some limiting value on it to the right. The difficulty is essentially the same as before. We could do a precise experiment in the central range but add an additional group of limited or finite size at a very low dose level, one likely to yield no induced tumors. Even if we extrapolate with a shallow slope to the left of the low dose level, we may have used too steep a slope between the central dose range and the low dose level.

A reverse position is taken by some who assert that our arbitrary shallow slope of unity may be too steep, citing instances in which shallower slopes of only about 0.5 have been obtained. How such shallow slopes can arise artificially is something of which we are too conscious. Thus if we are in a situation with a high or even moderate spontaneous

---

tumor rate, a naive analysis of the data ignoring spontaneity would yield too shallow a slope. Another cause of apparently shallow slopes could be the flattening or turndown of the response curve because of toxicity or any of a number of other reasons, as we have suggested above.

A final caution we should like to give is against a certain misuse of the "safe" dose result. A “safe” level generally could be interpreted as alternatively a “safe” level in drinking water on the basis that water intake and solid food intake are roughly equal. Such “safe” drinking water level should not, however, be interpreted as a “safe” level generally for water supplies or for bodies of fresh water, where biological concentration may play a crucial role. The impact of such a level generally could be much greater than when it is limited to the relatively small amount of water actually drunk; plant and animal life in the water would be the first, but not the only, forms of life affected.

Acknowledgments

Some early programming and extensive calculation work in connection with this research was carried out by James G. Karas, Senior Statistician, Merck Sharp and Dohme Research Laboratories, whose efforts we gratefully acknowledge.

Appendix A

Let $X_i$ represent the i’th distinctly different log dose (to base 10) occurring overall in the several experiments performed. For Experiment j, $r_j$ represents the number responding among $n_j$ test animals, these both being zero when test level i does not actually appear in Experiment j.

We can make the following simplifications from the symbolism used in the text:

$$P_i = P(a,b,0,X_i) = \int_{-\infty}^{\infty} f(a,b,Y_i) \exp \left(-\frac{1}{2} Y_i^2\right) dY$$

$$P_{ij} = P(a,b,X_i,X'_i) = c_j + P_i - c_j P_i$$

adding for convenience the normal density expression.

$$Z_i = Z(a,b,0,X_i) = (2\pi)^{-\frac{1}{2}} \exp \left(-\frac{1}{2} Y_i^2\right)$$

in which $Y_i = a + b X_i$. For the zero dose level, $X_i = -\infty$, $P_i = 0$, $P_{ij} = c_j$.

The loglikelihood for our data is

$$\log L = \text{Constant} + \sum_i \left[ r_j \log P_{ij} + (n_j - r_j) \log (1 - P_{ij})\right]$$

The following 1st derivatives of the loglikelihood obtain:

$$\frac{\partial \log L}{\partial a} = \sum_i Z_i (r_j - n_j P_{ij}) / P_{ij} (1 - P_{ij});$$

$$\frac{\partial \log L}{\partial c_j} = \frac{1}{1 - c_j} \sum_i (r_j - n_j P_{ij}) / P_{ij}.$$

The 2nd derivatives are given by:

$$\frac{\partial^2 \log L}{\partial a^2} = -\sum_i Z_i \left[ \frac{r_j}{P_{ij}^2} + \frac{n_j - r_j}{(1 - P_{ij})^2} (1 - c_j)^2 \right] + \frac{Y_i (r_j - n_j P_{ij})}{Z_i P_{ij} (1 - P_{ij})};$$

$$\frac{\partial^2 \log L}{\partial c_j \partial a} = -\sum_i (1 - P_{ij})^2 \left[ \frac{r_j}{P_{ij}^2} + \frac{n_j - r_j}{(1 - P_{ij})^2} \right];$$

$$\frac{\partial^2 \log L}{\partial c_j} = \frac{-c_j (1 - c_j)}{P_{ij} (1 - P_{ij})};$$

$$\frac{\partial^2 \log L}{\partial \phi_j} = \frac{c_j (1 - c_j) \sum_i (r_j - n_j P_{ij}) - r_j c_j (1 - P_{ij})}{P_{ij}^2};$$

$$\frac{\partial^2 \log L}{\partial \phi_j \partial \phi_k} = 0, j \neq k.$$
Yet another device accompanied our computing strategy, that of iterating for the \( \phi \), effectively the logit transform of the \( c \), rather than the spontaneous rates themselves. This was intended to avoid the difficulty that some iterative step might lead to a \( c \) value outside the permissible 0 to 1 range. For various reasons the standard iterative procedure of seeking the ML estimates by equating the 1st derivatives to zero and solving was not straightforward, as we now detail.

1. The ML estimate of \( a \) could sometimes be \(-\infty\). This would arise when the data failed to indicate an increase in the response rate with dose. Ordinarily, a constant response rate would be indicated by a \( b \) value of zero, while a negative \( b \) would indicate a declining rate. However, our \( b \) value was stipulated in advance at a positive value; with such a stipulation a constant response rate over finite dose levels would result from positive \( c \) values coupled with an \( a \) value of \(-\infty\). Our remedial device was to halt computation when the \( a \) estimate had become strongly negative and to accept it. The loglikelihood for the data would be virtually the same whether \( a \) were at \(-\infty\) or at some other extreme negative value, and it was only the maximum loglikelihood that was needed if an upper limit on \( a \) was being sought. (Another implication here would relate to the setting of a lower limit on \( a \), were it desired. With \( b \) prespecified, the data could be no more inconsistent with an \( a \) of \(-\infty\) than it would be with the model of a nonincreasing response rate. Thus, unlike the case of an upper limit on \( a \), there would always be a critical \( \chi^2 \) value for which the lower limit on \( a \) was \(-\infty\).

2. In iterating on the \( \phi \), a peculiarity occurred when in some experiment or experiments there were no control positives (whether because all controls were negative or because no controls were used). Under this circumstance the derivative of the loglikelihood with respect to the associated \( \phi \) would be zero at \( \phi \) = \(-\infty\), \( c \) = 0, whether or not the likelihood was indeed a maximum there. This could result in the iterative procedure converging at a local minimum with respect to the particular \( c \) involved, a saddle point with respect to the entire parameter set. For \( \phi \), headed towards \(-\infty\), we used the same device as that for \( a \) of stopping at some extreme negative value, but we added the feature of computing loglikelihoods to ensure that the likelihood had increased. If the loglikelihood were actually reduced, the computing routine would be thrown back to an earlier stage, but with some modification, and the iterative search resumed.

Relative to this problem we could visualize that, in the case of a single \( \phi \), convergence would be to the improper minimum of \(-\infty\) if the initial trial value for \( \phi \) were to the left of some critical inflection point. This would suggest that if several \( \phi \)'s were fitted simultaneously, it could be advantageous to start with entering \( \phi \) values which were on the high side, a feature that characterizes our computing routine. Essentially, we are in the single \( \phi \) situation whenever we use trial values of \( a \) in the secant search for the upper limit on \( a \). An increased value for \( a \), however, should lead to a decreased value in the associated estimate of \( \phi \), so that the 1st stage ML estimates of the \( \phi \) would provide high entering estimates for them when the upper limit on \( a \) is being sought. These complexities relative to the spontaneous rate parameters arise only in the case of the absence of positive responding controls. It was only our desire to have our computing program more completely general that has led to emphasis of this feature.

3. A way to avoid the minimum or saddle-point problem relative to the spontaneous rate parameter could be to iterate directly on the \( c \), using other devices to keep the estimate within the 0 to 1 range. We might for example start with a low positive \( c \) value near zero, minimizing the possibility that the iterative procedure will give rise to an estimate beyond unity, being prepared to reinsert a nonnegative \( c \) value should the result of a cycle of iteration be negative. A fresh difficulty arises with this approach. It is that when the true ML estimate of \( c \), in the 0 to 1 range, is indeed zero, the derivative of the loglikelihood with respect to that \( c \) will not be zero at that point. The derivative will instead be zero at some negative value for \( c \), and it is to this negative \( c \) value to which Newton-Raphson iteration will converge as we try to equate the 1st derivatives to zero. For the case in which only a single \( c \) value is involved, as occurs in the secant search procedure, there is no great problem: simply set that \( c \) value at zero, reestimate any other unspecified parameters in the system, and then compute the loglikelihood. Where there are several negative \( c \) estimates involved, because of their interdependence it may not be appropriate to set them all at zero. An approach we can see is this. Suppose that only in a limited number of the experiments do we fail to have positively responding controls; say there are \( w \) such. We then construct \( 2^w \) iterative searches, each reflecting a different specification of which subset, full set, or empty set of the associated \( c \) should be set at zero. The desired ML estimate is given by that solution with the largest loglikelihood for which there are no negative \( c \) estimates.

A number of other minor corrective adjustments attend our computing procedure, as when oscillatory or divergent behavior occurs. While at the present time our computing routine seems to be performing satisfactorily, it should nevertheless be considered as only tentative and subject to further improvement. For the while, the starting values we take in the Newton-Raphson iterative procedure are 0 for \( a \) and \( \log (r_j + 3)/(n_j - r_j + 3) \) for \( r_j \), in which \( r_j \) is the number of responders among \( n_j \) test animals at the lowest \( X_j \) (this would be the control group if one is used) in Experiment \( j \). The computing routine could be simplified a little, but probably with slower convergence resulting, through the use of expected 2nd derivatives, this requiring only the replacement of the \( r_j \) by \( n_iP_{ij} \) in the formulas shown. The inverse of the matrix of expected 2nd derivatives would perhaps be more appropriate for obtaining asymptotic variances.

References

An Improved Mantel-Bryan Procedure for "Safety" Testing of Carcinogens


Cancer Res 1975;35:865-872.

Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/35/4/865.citation

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/35/4/865.citation. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.