"Safety" Testing of Carcinogenic Agents[1]

NATHAN MANTEL *and* W. RAY BRYAN, *Biometry Branch and Laboratory of Viral Oncology, National Cancer Institute,*[2] *Bethesda, Maryland*

SUMMARY

The problem of determining what dose levels of an agent are safe, *e.g.*, noncarcinogenic, cannot be resolved unless one first defines some level of permissible risk, no matter how small, rather than insisting on absolute safety. Both because of practical considerations and statistical variation, the determination of low-risk dose levels, for example 1/100 million, cannot be made directly but must be by extrapolation from observed data. A conservative approach for doing so is given. In addition to an arbitrary definition of "virtual safety," it is necessary to define an arbitrarily high statistical assurance level and a rule for extrapolation by use of an arbitrarily shallow slope. Illustrative data by Bryan and Shimkin (J. Nat. Cancer Inst. 3: 503–531, 1943) on the carcinogenic action of methylcholanthrene yield a "safe," 1/100 million, dose of 9×10^{-8} mg per mouse when a statistical assurance level of 99 percent and a conservative probit slope of 1 normal deviate per log for extrapolation are used. The principles given are of general applicability in other safety-testing problems, the point of emphasis being that since direct observation cannot be made that the risk at some dose level is clearly low, indirect conservative procedures for the determination of low risk levels must be made. The arbitrary risks and definitions for so doing may change with circumstances. The procedure does not require specification of an experimental protocol; the "safe" dose is determined on the basis of whatever data are available. Minimum protocols may, however, be desirable, since greater amounts of data will ordinarily permit specifying large "safe" levels.—J. Nat. Cancer Inst. 27: 455–470, 1961.

ALTHOUGH IT is not definitely known whether all chemical compounds that induce cancer in experimental animals will also cause cancer in man, it has been fairly well established that, with the possible exception of arsenical compounds, every chemical or physical agent known to produce cancer in man will likewise do so in one or more species of lower animals (*1*).

Of necessity, the potential deleterious effects of chemical compounds must be tested in laboratory animals. The reaction of a particular species of animal does not constitute proof that humans will react similarly, but

[1] Received for publication March 29, 1961.

[2] National Institutes of Health, Public Health Service, U.S. Department of Health, Education, and Welfare.

597373—61——20

the only recourse the investigator has in implementing test programs for control of the human environment with respect to injurious chemical or physical agents is to proceed under the assumption that any substance harmful to animals is potentially harmful to man. Also in carrying out control tests in animals one must proceed, initially, as if protection of the animal population were the actual problem. When reliable estimates of the doses tolerated (with specified probability) are achieved for the animal population, the problem then becomes one of judgment, based on accumulative experience, in transferring the implications of the results to man. Results of testing with a variety of animals could suggest how extrapolation could be made properly to mammalian species of higher order or larger size.

Many chemical compounds may be harmful at certain concentrations, though beneficial at others. The control of "toxic" or "harmful" substances therefore does not imply the necessity of their complete or absolute elimination, which in some cases would be either impossible or economically infeasible, but their reduction to concentrations that can be tolerated by essentially all individuals of the population at risk.

In rapidly acting toxic substances that are either quickly eliminated from the body or are readily transformed to less harmful compounds through metabolic processes, the estimation of tolerated levels is not too difficult. With a reasonable allowance for an extra margin of safety introduced in the form of an arbitrary "safety factor," the results obtained in laboratory animals can be successfully projected to humans. For the most part, modern pharmacology is based on just such usage of laboratory animals.

There are other compounds, however, for which the results obtained in laboratory animals cannot be so confidently projected to man. These substances are not readily excreted or metabolized, and because of their weak solubility in aqueous solution, may remain in the body or on body surfaces for very long periods. Other substances may be metabolized to some extent, but because of their selective affinity for cells of certain types they may accumulate selectively within these cells to yield harmfully high concentrations if supplied on a continuing basis. Most chemical compounds that cause cancer, *e.g.*, polycyclic aromatic hydrocarbons, naphthylamines, azo dyes, and some steroids, have properties falling into one or the other of these categories.

Still other compounds have an immediate initial effect, which is not considered harmful, and are rapidly eliminated or metabolized; yet, during their brief sojourn in the body they produce some critical intracellular damage or change, which does not manifest itself until later. Urethan, which interferes with nucleic acid metabolism, is a classical example of a compound of this type. Injected parenterally, urethan acts immediately to induce transient anesthesia, but single anesthetic doses may cause lung tumors in mice many months later. Fortunately, this drug was never approved for use as an anesthetic for man.

Biometric methodology has been highly developed for the study of

acute toxicity and rapidly developing biological reactions of other types (*see 2–4* for review). Some progress has been made in the development of methods for the analysis of responses that are greatly prolonged in time, such as cancer (*see 5* for review); however, the problems associated with the estimation of tolerated levels of carcinogenic agents and other chronically acting compounds have not yet received adequate study.

The purpose of this communication is to discuss the major problems in the development of methods for estimating limits of tolerance, or "safety levels," of carcinogenic compounds and to describe certain biometric procedures, based on available data, that are applicable to presently known carcinogenic compounds and experimental animal systems. Factors which one must consider in transferring the inferences derived from results in animals to man are also discussed. The suggested procedures are not restricted to carcinogenic agents, but are applicable, in principle, to compounds which cause various other types of harmful reactions or disease.

SOME PROBLEMS IN PLANNING AND ANALYSIS OF SAFETY STUDIES

In a safety-testing program both the design of the experimental protocols and the method of analysis or interpretation of resulting data must be determined. The experimental design cannot be properly determined apart from the plan for analysis of the data.

Certain issues in a safety-testing program must be resolved in advance. These relate to what we mean by safety and what kind of feasible results we are willing to accept as proof of safety. Settling of these issues, if necessary on some arbitrary but conservative basis, may permit answers to problems that would otherwise be insoluble. These problems are:

1) *How safe is safe?* Absolute safety can never be unquestionably demonstrated experimentally. Rather, experimental results can be used only to establish limits on the risk involved. With the specification of some level of risk, no matter how small, the possibility of determining whether or not that risk is exceeded opens. We may, for example, assume that a risk of 1/100 million is so low as to constitute "virtual safety." Other arbitrary definitions of "virtual safety" may be employed as conditions require.

Incidentally, an inflexible requirement for absolute safety may lead to acceptance of high levels of hazard. The impossibility of really demonstrating absolute safety leads to the acceptance, as a satisfactory demonstration, that no hazard was observed in an experimental protocol of moderately large size, 100 or even 1,000 animals. Such evidence, however, only provides assurance, at the 99 percent probability level, that the true risk is under 4.5 percent in the 100 animals or 0.46 percent in the 1,000 animals.

2) *What constitutes proof of safety?* In principle, one could use an experimental protocol sufficiently large to demonstrate that "virtual

safety" obtained. For this purpose it must be realized that an observed outcome of no tumors among 100 million treated mice does not necessarily demonstrate clearly that treatment was either absolutely or even virtually safe. This outcome could arise with a probability of 1 percent, even if the risk involved were as high as 4.6/100 million. It would in fact require a total of some 460 million tumor-free mice to demonstrate at the 99 percent assurance level that "virtual safety" obtained. Similarly, tumor-free results for 10,000 mice would only indicate that the risk was less than 1/2,200 and it would require tumor-free results in a total of some 450 mice to establish with high probability that the risk was under 1 percent. Studies of feasible size can be used to establish directly only risks of the order of 1/100 or higher. Data from such studies can be used to ascertain the treatment level consonant with a prescribed risk or to establish limits on the risk for a particular treatment level. The determination of "safe" levels can be made only by indirect methods extrapolating from the data obtained in a feasible study.

The use of extremely large studies to establish safety may well be self-defeating. The almost certain occurrence of unusual syndromes in one or more of a large number of test animals, albeit these may have arisen spontaneously, will require admitting the possibility that they may be attributable to drug treatment.

3) *How can protocol data be extrapolated safely?* Since it is only feasible to use experimental protocols for the direct determination of relatively high-risk dose levels, *e.g.*, 1 percent, which makes extrapolation methods necessary, one must consider that such extrapolation methods might yield misleading results. Procedures exist which permit extrapolating the results obtained at a number of test-agent levels to determine the dose level corresponding to any desired degree of risk and to establish, with a high level of assurance, a minimum bound on this dose level. These methods, however, are based on the assumption that the relationship observed between tumor occurrence and dose at the levels tested will continue to apply in the regions to which extrapolation is being made. The validity of such an assumption cannot be tested and, if it is false, may lead to a serious overestimate of the "safe" level. Such overestimation would arise if the relationship of response to dose is less pronounced at the dose levels to which extrapolation is made than at the levels at which tests were performed. Another related source of difficulty is that tests are performed on relatively pure inbred strains of laboratory animals. Characteristically, such pure strains will show steep dose-response relationships, while the heterogeneous population to which it is intended to apply the results of testing may exhibit a shallow relationship.

To avoid the risk of overestimation of "safe" levels which may result from extrapolating with too steep a slope, it is suggested here that a conservative result may be obtained by extrapolation with an arbitrarily low slope from the data at hand. For example, quantal-response data, that is, all-or-none response, frequently exhibit a somewhat linear relationship when plotted on probability paper with a normal or probit

scaling for the percent responding and a logarithmic scaling on dose. According to the kind of system being investigated the dose-response slopes observed may vary widely. With systemic poisons, rather steep dose-response slopes are generally obtained, reflecting the narrow logarithmic range between the lowest dose levels at which any toxic deaths occur and the levels which are lethal for all animals—such response slopes are on the order of 10 to 50 or more probits per common logarithmic or tenfold dose increase (the technical definition of this slope is not given here; it permits one to perform any necessary extrapolations). The all-or-none response also arises in the study of the therapeutic effects of antibiotics. The response slopes in these instances are generally shallower, on the order of 3 probits per common logarithm. Lower slopes on the order of 2 do arise in virus-assay work, but in these instances it may be that use of a different response curve, the single-hit or one-particle curve, would be more appropriate. From such experience it would appear that the use for purposes of extrapolation of a slope as low as one probit per common logarithm is likely to be conservative. While slopes in the regions to which we wish to extrapolate cannot be established, the suggested slope of one is rather low compared with that ordinarily obtained in the observable region.

The indicated low slope is a key feature in the method to be suggested as conservative for the establishment of "safe" levels. For this reason it may be well to make clear just how weak or strong are the assumptions being made in its use. In fact the only assumption being made for the procedure to be conservative is that whatever form the true response curve may take over the region of extrapolation, the average slope is not less than the assumed one. There is no requirement that the true response curve be linear or even that the true slope should nowhere be less than assumed. The use of the indicated conservative slope is, of course, arbitrary. Other values may be specified and other scales for extrapolation may be employed.

Once answers to the three questions are provided, a defined level of "virtual safety," a prescribed level of statistical assurance, and a conservative rule for extrapolation, it becomes possible to determine, from protocol data, "safe" dose levels and to undertake the planning of any necessary experimental protocols. This is considered in the succeeding sections.

ANALYSIS OF RESULTS AT A SINGLE DOSE LEVEL

In what follows we will take the defined level of "virtual safety" to be 1/100 million, the statistical assurance level to be 99 percent. Extrapolation will be on the basis of 1 normal deviate or probit per common log or tenfold change in dose.

To illustrate how definitive "safe" levels are obtained, consider that a prescribed dose of an agent has elicited no tumors in a group of 100 ex-

perimental animals. While the observed rate of tumor occurrence is 0 percent, we will take, as an upper limit on the true rate, that risk for which the probability for occurrence of as few as zero tumors is 1 percent (100% less the assurance level of 99%). This is given by the solution for P to the equation

$$(1 - P)^{100} = 0.01.$$

Solving, we have

$$100 \log (1 - P) = \log 0.01 = -2; \log (1 - P) = -0.02 = 9.98\text{–}10;$$
$$1 - P = 0.955; P = 0.045 \text{ or } 4.5 \text{ percent.}$$

We now know that the observed outcome of no tumors in 100 animals is consistent with the possibility that the true risk was, in fact, 4.5 percent. From tables of the normal probability function (*6*) we determine that the normal deviate, Y, such that the integral from $-\infty$ to Y equals 0.045 is -1.695, for a probit value of $3.305 = 5 - 1.695$. However, the normal deviate, Y_o, corresponding to a risk of 1/100 million can similarly be determined as -5.612, the probit being -0.612. The upper limit on the risk for the dose employed is $3.917 = -1.695 - (-5.612)$ normal deviates above the desired safe risk, and, at a slope of one normal deviate per common log, it is necessary to reduce the log dose by 3.917 logs to attain a "safe" level. The antilog of 3.917 being about 8,300, it is determined that the "safe" dose is 1/8,300 times that which had been tested.

Table 1 shows the preceding results together with those for several other hypothetical experiment sizes in which no tumors were observed to occur in a group of treated mice. For each such group the table shows the greatest risk consistent, at the 99 percent assurance level, with the observed outcome. Also shown are the corresponding conservative estimates, with the slope of one normal deviate per tenfold dose increase, of the "safe" (1/100 million) dose level expressed as a fraction of the dosage tested. The larger the experimental group among which no tumors occurred, the greater is the value determined as the "safe" dose.

TABLE 1.—Illustration of "safe" doses determined when no risk was observed in single groups

Number of mice with tumors/ No. tested	0/10	0/50	0/100	0/500	0/1,000
Upper limit on tumor risk at level employed, 99 percent assurance (percent)	37	8. 8	4. 5	0. 92	0. 45
Estimated "safe" dose (1/100 million) dose employed = 1	1/190,000	1/18,000	1/8,300	1/1,800	1/1,000

Study of this table indicates that a control system can be established without the need for specifying a design protocol, though there might be some merit in specifying a minimum size. For, whether the amount of evidence adduced to show that an agent is safe is great or small, it can be properly weighted to determine conservative safe limits. If the promulgator of a drug wishes to have high tolerances established for his

compound, it would be worth while for him to produce results of experiments with a large number of animals, which would show that the agent is not especially dangerous at certain dose levels. Where the lack of danger is demonstrated with a small number of animals or as a result of testing at rather low doses, the dose levels determined as "safe" will be lower. A control system can be constructed about the possibility for interpreting the data submitted, with no specification needed as to how much data should be obtained.

THE CASE OF SOME OBSERVED RISK

The method indicated for determining "safe" levels is not restricted to the case in which no observable danger was noted. That case was used for illustration because of the simpler mathematical solution involved. In general it will be that an experiment testing n animals will yield r unfavorable results (tumors). The upper limit on risk at the 99 percent assurance level is then the solution for P to

$$\sum_{i=0}^{r} {}_nC_i P^i (1 - P)^{n-i} = 0.01$$

or

$$\sum_{i=r+1}^{n} {}_nC_i P^i (1 - P)^{n-i} = 0.99$$

The solution for P is a value such that the chance of observing as few or fewer than r tumors is 1 percent. At values for P in excess of the solution, the chance for such an outcome is less than 1 percent.

The preceding equation can be solved approximately by reference to tables of cumulative binomial probabilities. Such tables, as for examples those of the Ordnance Corps (7), show values of quantities such as

$$\sum_{i=r+1}^{n} {}_nC_i P^i (1 - P)^{n-i}.$$

Let us consider as a hypothetical outcome that, of $n = 100$ mice, $r = 10$ have developed tumors, for an observed rate of 10 percent. Referring to these tables we see that for $P = 0.19$

$$\sum_{i=11}^{100} {}_nC_i P^i (1 - P)^{n-i} = 0.9891$$

and that for $P = 0.20$

$$\sum_{i=11}^{100} {}_nC_i P^i (1 - P)^{n-i} = 0.9943.$$

We may take the solution for P as approximately 0.192. The calculating procedure follows as before. The normal deviate corresponding to a 19.2 percent probability is -0.871 and, at the slope assumed, it will require a reduction in log dose of $4.741 = -0.871-(-5.612)$ to obtain a risk of 1/100 million. The "safe" dose is then determined as 1/55,000 of the dose which had been employed, 55,000 being the antilog of 4.741.

USE OF CONTROL DATA

In the methodology shown, it was assumed that the response of interest, appearance of tumors, did not occur spontaneously. In general, however, it will be desirable to use controls to check this and to allow data obtained for such controls to modify the determination of "safe" dose made. However, with the method already shown, failure to use controls or to take control data into account will result in more conservative determinations of the "safe" dose. If, in fact, spontaneous rates are rather low, they will have little effect on the determination made. For this reason we may adopt a procedure which is somewhat more conservative than necessary for taking control data into account. This we can do by taking, as before, the upper limit for the risk in the treated group as the solution for P_t to

$$\sum_{i=0}^{r_t} n_t C_i P_t^i (1 - P_t)^{n_t - i} = 0.01$$

while the lower limit on the control group risk is the solution for P_c to

$$\sum_{i=r_c}^{n_c} n_c C_i P_c^i (1 - P_c)^{n_c - i} = 0.01$$

where r_t of n_t treated animals and r_c of n_c control animals, respectively, showed positive response. These equations can be solved through the use of binomial tables.

At this point Abbott's formula (*8*) can be used to obtain a modified value for the treated-group risk, this being computed as

$$P_t' = (P_t - P_c)/ (1-P_c).$$

The computation follows as before, the normal deviate being obtained corresponding to P_t'.[3] Since P_t' cannot exceed P_t, the use of control data cannot result in decreased values for the calculated "safe" dose.

ANALYSIS OF RESULTS OBTAINED AT SEVERAL DOSE LEVELS

When an agent is tested, it is sometimes desirable to do so over a number, perhaps even a wide range, of dose levels. In such instance, all the available data should be considered in the determination of "safe" levels.

[3] The fastidious statistician may object to the moderately conservative procedure described here for determining P'_t. A more rigorous procedure would require setting limits on the ratio of binomial parameters. P_t' would be given by reducing by unity the upper limit on the ratio of control-to-treatment nonresponse probabilities.

Parametric Procedures

As already indicated, it may be unwise to extrapolate the data with the observed response slope. Procedures for taking into account the statistical variation of the fitted slope will not suffice to make such extrapolation methods conservative. To see this, one need only consider the use of quite large study sizes. In this case, statistical variation will be negligible with the result that extrapolation to low-risk levels will be substantially with the slope obtaining in the observable range. (However, with small study sizes or studies, the designs of which are inefficient for estimation of the slope, taking account of the statistical variation of the slope determination may result in extremely conservative estimates of "safe" levels. The lower confidence limit on the "safe" dose in such instances may be less than that which would be obtained when an arbitrarily low slope value is used for extrapolation, as suggested.)

An alternative device could be to employ parametric procedures, for example, fitting the maximum likelihood probit line to the data (*9*), to determine the lower confidence limit on the dose corresponding to some moderate percentage risk, *e.g.*, 1 percent, to which risk-level extrapolation with the observed slope is considered reliable. Then, using an arbitrarily conservative value for the slope, and anchoring at the lower limit on the 1 percent dosage, one could extrapolate to the desired "safe" level.

While the procedure just indicated is straightforward, its employment is based on the validity of the parametric function employed. Quite useful results have derived from such parametric assumptions in bioassay work. But for bioassay purposes it is not essential that a parametric function be exactly appropriate; it is sufficient that the function assumed do a reasonably good job of graduation. (Even somewhat inappropriate curve forms can yield reasonably good relative potency estimates as long as the preparations being compared are tested over the same regions of response.) The use of an invalid parametric function can lead to inappropriate estimates of dose levels corresponding even to moderate risks.

In many situations the estimates may be only moderately inappropriate, in others quite serious. With parametric procedures, the use of rather large experimental groups in one region of the response curve can be reflected in narrow-range confidence intervals for dose levels corresponding to risks in other ranges. This can produce a false sense of security in one's estimate of the moderate-risk dose level when the parametric model is violated.

Accordingly, while agreeing that parametric procedures may be useful, we will consider the possibility for extending the method described for the single dose-level case without the need for assuming any particular model. The only assumption is that the arbitrary low slope assumed is conservative and it is, of course, implicit that the response curve is monotone. In instances in which it can be recognized that use of a parametric procedure is not misleading, workers may prefer this procedure rather than the more generally appropriate nonparametric procedures described in the next section.

Nonparametric Procedures

In the preceding section it was suggested that the use of parametric methods, while straightforward, could lead to nonconservative results. There are no simple fixed rules for conservative estimates of the "safe" dose when several dose levels are employed and one is unwilling to make assumptions about the dose-response curve in the region of observation. How estimates can be made in these circumstances can best be demonstrated by illustration.[4]

We will begin with some simple ideas. Suppose investigators at two laboratories independently test an agent at a level of 100 mg/kg. At the first, with 500 mice tested, no tumors are observed, and with the methods described previously the "safe" dose is estimated as 100 mg/kg/1,800 = 0.056 mg/kg (*cf.* table 1). A somewhat lower "safe" dose of 0.012 mg/kg is obtained at the second laboratory, based on the observation of no tumors among only 100 mice. It can readily be recognized here that it would be inappropriate to reject the high "safe" level of the first laboratory just because of the low estimate obtained at the second laboratory. The two sets of data are consistent and in fact confirm each other. If any modification is to be made, it should be to consider that, with results combined, no tumors have occurred among 600 mice which would lead to a safe dose of about 0.065 mg/kg.

Suppose that at still a third laboratory, tests are made at a dose of 50 mg/kg and, with no tumors occurring among 500 mice, the calculated "safe" dose at that laboratory is 0.028 mg/kg. Here again we can see that the "safe" dose obtained at the first laboratory should not be modified downward just because a consistent result at the third laboratory yielded a lower "safe" dose. If anything, it should be considered that the 500 mice not responding at the higher dose at laboratory 1 would not have responded at the lower dose employed at laboratory 3. With these 500 mice treated as nonresponders at the low dose, there is then, including those at laboratory 3, a total of 1,000 mice not responding at the low dose. (Laboratory 2 results are being ignored for this illustration.) This yields as a calculated "safe" dose 50 mg/kg/1,000 = 0.050 mg/kg. In the present case the "safe" dose based on the combined calculation is less than that for the data of laboratory 1 alone of 0.056 mg/kg and so the higher figure is retained. Had the combined calculation led to a higher "safe" dose it would have been correct to take that as the estimate.

The point of these illustrations is that, when the data obtained from a series of doses are consistent with each other, it is appropriate to take as the calculated "safe" level the highest one pertaining to the results at any one dose. Even a higher "safe" level may be taken when it can be obtained through a justifiable combination of the results at the various doses used. What is meant here by "justifiable" combinations can be seen from the following example in which hypothetical results at four dose levels, low, middle, and high, are considered.

[4] An alternative method to the one about to be described is given as an appendix.

Dose in size order	Observed results Number of tumors/ number of mice	"Justifiable" combined results Combined number of tumors/combined number of mice			
1	0/100	(0/100);[5]	0/200;	1/300;	5/400
2	0/100	0/100;	1/200;	5/300	
3	1/100	1/100;	5/200		
4	4/100	4/100			

[5] Parentheses indicate that this result need not be considered, as the next must yield a higher value for the "safe" dose.

In the absence of inversions in the data, we can determine the "justifiable" combined results at a dose by adding to the results at that dose, in succession, the results at still higher doses. A calculated "safe" dose can be determined for each dose used and for each of the various "justifiable" combined results corresponding to each dose. The over-all "safe" dose would be the highest of the various determinations.

Where data show an inversion the procedure is altered. Consider a simple example:

Dose in size order	Observed results: Number of tumors/ number of mice	"Justifiable" combined results: Combined No. of tumors/combined No. of mice
1	1/100	(1/100);1/200
2	0/100	[(0/100);1/100]

At the first dose level it is clear that the calculated "safe" dose would be larger if based on the combined results for both levels than if based on the results observed at this level alone; accordingly, as noted in one instance in the preceding example, the result at the lower level alone is shown in parentheses. At the higher dose level an inversion occurs; there is a lower incidence of tumors even though the dose level is higher. In view of the inversion, one would be less willing to accept as "safe" the calculated value obtained on the basis of results for this dose level alone. The two alternative results shown in brackets at this dose level are the results at this dose level and the contradictory results at the lower dose level. The significance of the use of brackets here is that the calculated "safe" value is now to be taken as the lesser of the values suggested by the alternative results. In the present instance, the result at the lower dose 1/100 would yield the lower "safe" dose and so the alternative result is shown in parentheses. (It will not always be necessarily true, when an inversion occurs, that the retained result will correspond to the higher tumor incidence at the lower dose level.) In the present example the calculated "safe" dose will be that corresponding to the first dose with combined result 1/200 or that corresponding to the second dose with retained result 1/100, whichever is the greater.

In practice the application of the methods just indicated is much simpler than the explanation would suggest. Ordinarily only one or perhaps two of the combined results at a dose level will need to be considered.

The results at some dose levels may immediately permit us to drop them from consideration. After only a limited amount of experience it should be possible so do this rather rapidly. The calculations are actually simpler than those for the maximum likelihood probit method. [In fact, the confidence limit procedures ordinarily employed in connection with the probit method are not fully satisfactory. A more appropriate method is described by Mantel and Patwary (*10*), but it could require a somewhat extravagant level of computational effort.] And, while for completeness, we have indicated the need for considering the possibility of inversions, this will ordinarily not pose a problem.

An Illustrative Example

An example from the literature shows how the procedure just discussed can be applied. The data, from Bryan and Shimkin (*11*), are the results obtained after a single injection of methylcholanthrene into mice, 12 different dose levels being used in the study. The reader may refer to the original article for details.

No peculiarities arise in this example. There are no inversions. At the four lowest levels no tumors occurred and the appropriate combined result is readily recognized in these instances. At the middle four levels it can be recognized that there is no point in combining results and, finally, the four highest levels can be disregarded as these all yielded 100 percent tumor occurrence.

The procedure is illustrated in table 2. The first three columns show, respectively, the dose, log dose, and the observed result. Column 4 shows each combined result considered, and there should be a separate line for each such result. In the present instance only one combined result required to be considered at each dose. For each such result, column 5 shows the calculated maximum risk at the 99 percent assurance level. These were obtained from binomial tables or calculated directly for the case of no tumors occurring. The normal deviate corresponding to the maximum risk is obtained from tables of the normal distribution and is shown in column 6. Finally, column 7 shows the calculated "safe" (1/100 million) log dose. The maximum for this, 2.962–10, appears in the second line, and the over-all calculated "safe" dose is 9×10^{-8} mg per mouse.

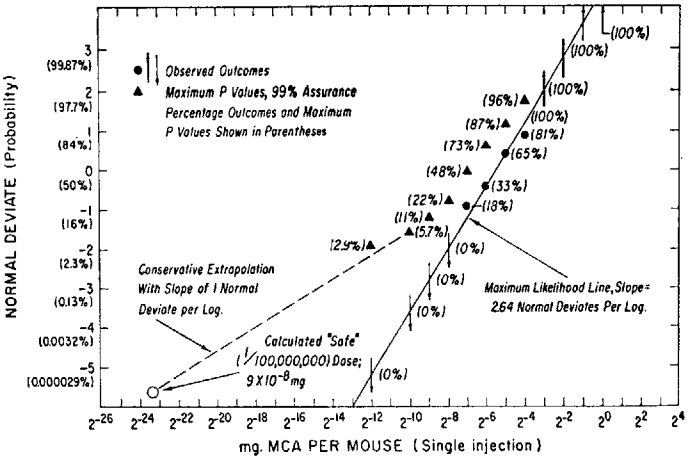
One might remark that this "safe" dose is so low as to make impractical any use of it which may result in its ingestion by humans. But we are dealing here with a rather potent carcinogen and if any compounds are to be assigned tolerated levels which are virtually zero, this is one of them.

Text-figure 1 shows graphically the results and analysis of the experiment just considered. Normal deviates are shown on the vertical scale, while on the horizontal scale the dose employed is shown as negative descending powers of 2. The points shown represent the outcomes at each dose level; 0 and 100 percent outcomes are shown by arrow. The solid line on the figure is the maximum likelihood probit line fitted to the

TABLE 2.—Illustration of methodology for determining the "safe" dose from results at several dose levels; data from Bryan and Shimkin (*11*)

Dose mg/mouse	Log dose	Result: No. of tumors / No. of mice	Combined result: No. of tumors / No. of mice	Maximum *P* value 99% assurance	Corresponding normal deviate	Calculated "safe" (1/100 million) log dose (2) − (6) −5.612
(1)	(2)	(3)	(4)	(5)	(6)	(7)
0. 000244	6. 388–10	0/79	0/158	0. 0288	−1. 899	2. 675–10
0. 000975	6. 990–10	0/41	0/79	0. 0566	−1. 584	*2. 962–10*
0. 00195	7. 291–10	0/19	0/38	0. 1141	−1. 205	2. 884–10
0. 0039	7. 592–10	0/19	0/19	0. 2152	−0. 789	2. 769–10
0. 0078	7. 893–10	3/17	3/17	0. 480	−0. 050	2. 331–10
0. 0156	8. 194–10	6/18	6/18	0. 729	+0. 610	1. 972–10
0. 0312	8. 495–10	13/20	13/20	0. 871	+1. 131	1. 752–10
0. 0625	8. 796–10	17/21	17/21	0. 958	+1. 728	1. 456–10
0. 125	9. 097–10	21/21	—	—	—	—
0. 25	9. 398–10	21/21	—	—	—	—
0. 50	9. 699–10	21/21	—	—	—	—
1. 0	10. 000–10	20/20	—	—	—	—

data. Above the first 8 data points the triangles shown correspond to the maximum *P* values of table 1. Extrapolation, with the slope of one normal deviate per common log to the over-all calculated "safe" value, is indicated by a broken line. All triangles, other than the one from which



TEXT-FIGURE 1.—Estimation of the "safe" dose from test results with a carcinogen methylcholanthrene, at several dose levels. At each test level both the observed percentage response and an upper limit, 99 percent assurance, based on combined data are shown. *Solid line* is the maximum likelihood probit line fitted to the data. The "safe" level of 9×10^{-8} mg per mouse is in this instance estimated by extrapolation with the conservative slope of 1 normal deviate per log from the upper limit on *P* at the second dose level.

extrapolation was made, should fall above the line. The difference in slope between the solid and the broken line may be noted.

RELATION TO OTHER SYSTEMS

While at various points in the preceding section the possibility for arbitrary selection of the assurance level, the level of safety desired, the slope value used for extrapolation, and even the extrapolation curve (one could use some scale other than normal deviates or probits along which a risk probability may be defined) have been emphasized, this should not be taken to mean that the methods suggested are completely general. Rather, in each case the nature of the risk situation should be considered.

Some principles do carry over, however. For example, one may be interested in trying to extinguish a bacterial or viral population by exposure to increasing temperatures or by increasingly long exposure to bactericidal or viricidal conditions. Determining the appropriate temperature or duration of exposure through the use of a conservatively shallow slope may be appropriate in such cases.

Consideration of the "single-particle" or "one-hit" problem (*12*) gives rise to a quite different answer than that developed in the preceding section. In this problem it is considered that a single particle can cause infection or death. If particles are distributed at random in material being inoculated, so that if there is an average of m particles per inoculation, the probability that a particular inoculation will contain none and so be safe is e^{-m}; the probability that it will not be safe is then $1-e^{-m}$. The problem is to determine from test data what reduction in the inoculum is necessary to ensure safety.

One can use directly the methodology already given to set maximum values on the risk as a result of the outcome of testing. This in turn establishes a maximum value for m as $-\log_e$(1-max. P value), the necessary reduction to any desired risk level following directly.

In this instance the inoculum corresponding to a risk of 1/100 million is approximately one millionth that corresponding to a risk of 1 percent. This contrasts sharply with the ratio of about 1/2,000 when extrapolation is made between these two risk levels with a slope of one probit per common log. While it has been shown that in the central region the "one-particle" curve mimics the probit curve with a slope of 2 probits per common log (*13*), the contrast would suggest that the comparatively steep slope of 2 noted virtually disappears in the lower tail. What further suggests itself is that the "single-particle" curve provides a most conservative rule for extrapolation. Where there is any suspicion that this curve may apply, the procedure described in the text should not be used.

One might also visualize a two-hit or two-stage process, the low-level probabilities for each stage being approximately proportional to the dose used. A reduction in dose by a factor of 1,000 should reduce the joint probability by a factor of 1 million. In this case the 1/100 million risk dosage would be 1/1,000th that of the 1 percent risk dose.

APPENDIX

Alternative Method

An alternative method to that described for handling data at several dose levels may be more appealing to the biometrically oriented reader. It was evolved in discussions with a reviewer (J. Cornfield) and is objective and intuitively satisfying.

In this method the data are considered to consist of a series of assays: one comprising results at the lowest dose level only; another comprising results at the two lowest levels; still another in which the results are given for those at the three lowest levels, etc. (Levels where there is interference, as for example lethality interfering with a carcinogenic response, should not be included.) For each assay, with appropriate probit procedures determine the maximum likelihood, 1/100 million dose level and its lower limit, subject to the restriction that the probit slope is known to be unity. Of all the lower limits on the 1/100 million dose level determined for the series of assays, the largest is selected as the "safe" dose.

A drawback to this procedure could be that standard methodology does not correctly permit the determination of lower limits on the 1/100 million level. Such methodology goes awry if, for example, there are no positive responses at the three lowest dose levels. A principle described by Mantel and Patwary (*10*) permits a solution. Calculate the chi-square goodness-of-fit, χ_1^2, (or log likelihood for those so inclined) for the maximum likelihood solution. (This chi square would be taken as zero if there are no positive responses at any of the dose levels considered; the maximum likelihood 1/100 million level is actually infinite. A zero chi square obtains also when only results at the lowest dose level are considered.) Consider alternative trial values for the 1/100 million dose level less than the maximum likelihood estimate. Each such trial value, in conjunction with the assumed slope of unity, provides an estimate of the response rate at each dose level of the assay. There is thus, in turn, a chi-square value, χ_2^2, for departures of the observed responses from the estimated responses based on the trial 1/100 million dose level. χ_2^2 will exceed χ_1^2 and for some trial value $\chi_2^2 - \chi_1^2$ will equal 5.412, the 98 percent upper limit on a chi square with one degree of freedom (a 98% limit on chi square corresponds to a 99% on the "safe" level).

At this point we introduce a modification which, for simplicity, has just been glossed over. Instead of computing χ_1^2 based on weights implied by the maximum likelihood solution, one should compute the quantity χ_1^{2*} based on departures from the maximum likelihood fit, the weightings being derived from the fit to the trial "safe" dose.

To illustrate: Suppose that at the i'th dose level, r_i of n_i animals respond; that the maximum likelihood estimate of the response rate is $\hat{P}_i$ and the estimate corresponding to the trial value considered is $\hat{P}_i^*$. Then

$$\chi_2^2 = \sum_i (r_i - n_i\hat{P}_i^*)^2/n_i\hat{P}_i^*(1 - \hat{P}_i^*)$$

and

$$\chi_1^{2*} = \sum_i (r_i - n_i\hat{P}_i)^2/n_i\hat{P}_i^* (1 - \hat{P}_i^*).$$

The trial value for which $\chi_2^2 - \chi_1^{2*}$ equals 5.412 is the lower limit on the safe dose for the assay considered. With k dose levels, there will be k sets of assay data and k lower limits, the maximum of these limits being the one selected. (The likelihood ratio may be taken as an alternative criterion for setting limits; *see 10.*)

REFERENCES

(*1*) MIDER, G. B.: The role of certain chemical and physical agents in the causation of cancers. Hearings before the Committee on Interstate and Foreign Commerce, 86th Congress, 2d Session, on H.R. 7624 and S. 2197, January 26, 1960. Washington, U.S. Govt. Print. Office, 1960, pp. 45–60.

(*2*) BLISS, C. I., and CATTELL, M.: Biological assay. Ann. Rev. Physiol. 5: 479–539, 1943.

(*3*) FINNEY, D. J.: Statistical Methods in Biological Assay. New York, Hafner Publishing Co., 1952.

(*4*) BURN, J. H., FINNEY, D. J., and GOODWIN, L. G.: Biological Standardization, 2d ed. London, Oxford University Press, 1950.

(*5*) BRYAN, W. R.: Quantitative biological experimentation in the virus and cancer fields. J. Nat. Cancer Inst. 22: 129–159, 1959.

(*6*) FEDERAL WORKS AGENCY, WORK PROJECTS ADMINISTRATION FOR THE CITY OF NEW YORK: Tables of probability functions 2: 1942.

(*7*) ORDNANCE CORPS PAMPHLET ORDP 20-1: Tables of the cumulative binomial probabilities. Washington, Ordnance Corps, September, 1952.

(*8*) ABBOTT, W. S.: A method of computing the effectiveness of an insecticide. J. Econ. Ent. 18: 265–267, 1925.

(*9*) CORNFIELD, J., and MANTEL, N.: Some new aspects of the application of maximum likelihood to the calculation of the dosage response curve. J. Am. Stat. Assoc. 45: 181–210, 1950.

(*10*) MANTEL, N., and PATWARY, K. M.: Interval estimation of single parametric functions. Proc. 32d Session of the International Statistical Institute, Tokyo, Japan, 1960. In press.

(*11*) BRYAN, W. R., and SHIMKIN, M. B.: Quantitative analysis of dose-response data obtained with three carcinogenic hydrocarbons in strain C3H male mice. J. Nat. Cancer Inst. 3: 503–531, 1943.

(*12*) LEA, D. E.: Actions of Radiations on Living Cells. Cambridge, Cambridge Univ. Press, 1946.

(*13*) MEYNELL, G. G.: Inherently low precision of infectivity titrations using a quantal response. Biometrics 13: 149–163, 1957.