The Dominant-Lethal Test: Potential Limitations and Statistical Considerations for Safety Evaluation

by Sidney Green* and Janet A. Springer†

Introduction

In June of 1971, two independent laboratories under contract with the Food and Drug Administration were given the responsibility of determining the potential mutagenicity of 24 food additives generally recognized as safe (GRAS) (*1*). This goal was to be accomplished by an examination of these compounds by the host-mediated assay, *in vivo* cytogenetic assay, and the dominant-lethal test.

This communication will focus upon one of the three assays, namely, the dominant-lethal assay. We wish to point up problems encountered in attempting to use the results to evaluate safety and how these may limit the extent to which the assay may be employed in safety evaluations. It is not intended to present the results of the tests of the 24 GRAS compounds by the dominant-lethal assay.

The nature of these problems is twofold: first, the biology of the system, and second, the statistical considerations. The detailed methodology involved for the dominant-lethal test employing mice and rats has been described previously (*2*, *3*). In the present investigations male rats are dosed singly (acute) and mated with groups of untreated virgin females over sequential weekly intervals, usually for 8 weeks. In the subacute test the male rats are dosed on five consecutive days and then mated. Ten males are generally used for each dose level and mated with two virgin females each week. The following three parameters are counted for each female: total corpora lutea, total implants, and dead implants. Corpora lutea minus total implants gives a direct measure of the preimplant losses for each female.

*Division of Toxicology, Food and Drug Administration, Department of Health, Education, and Welfare, Washington, D. C. 20204.

†Division of Mathematics, Food and Drug Administration, Department of Health, Education, and Welfare, Washington D. C. 20204.

Evaluation of Untreated Controls

A large body of control data was available from the three laboratories including the FDA laboratory. Over 260 weekly means were tabulated for control rats from over 30 experiments. These control data were examined for consistency and trends.

There were no significant differences ($p > 0.05$) between the acute and subacute studies in the control values of corpora lutea, total implantations, dead implantations, or preimplantation losses. Within a few experiments, the route of administration had an effect on total implantations for control rats; however, this effect was not consistent from experiment to experiment. In two studies with Holtzman rats, the mean control value over 10 weeks was lower by one im-

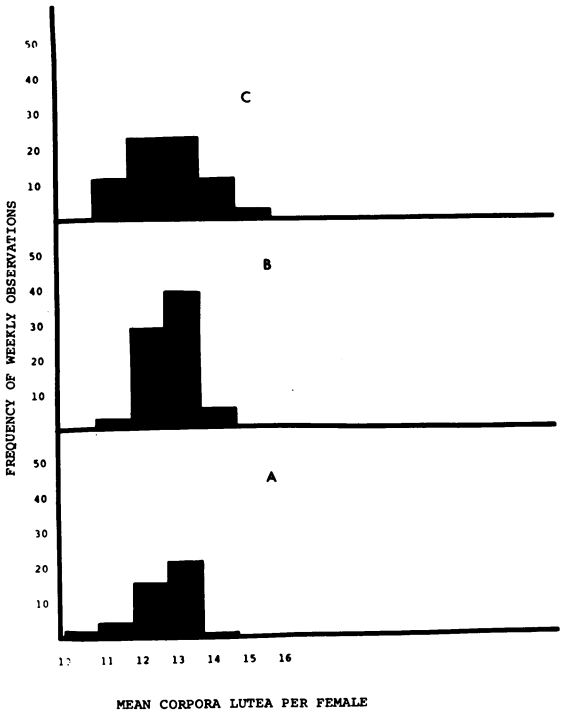

FIGURE 1. Distribution of weekly mean corpora lutea per pregnant female: (*A*) laboratory 1; (*B*) laboratory 2; (*C*) laboratory 3.

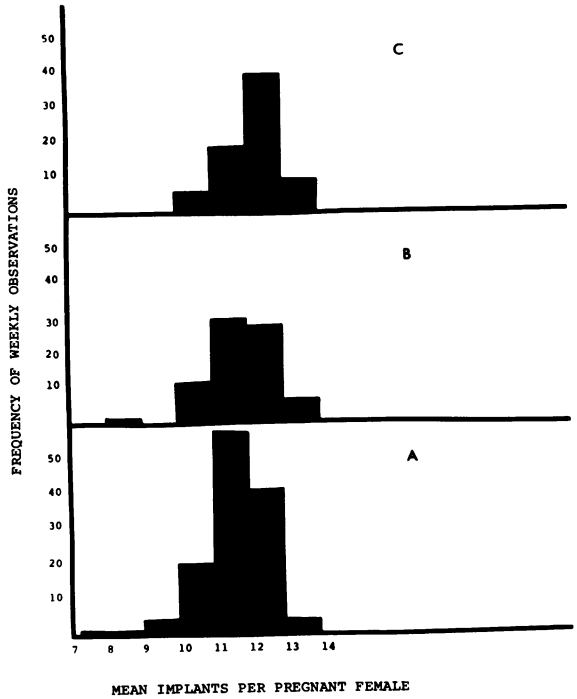

FIGURE 2. Distribution of weekly mean implants per pregnant female: (*A*) laboratory 1; (*B*) laboratory 2; (*C*) laboratory 3.

plantation per female when the control solvent was given intraperitoneally than when it was given by intubation. However, the mean number of dead implantations per female over all weeks was almost identical for the two different routes of administration. Control data were therefore pooled for the different routes of administration and for acute and subacute studies.

Figure 1 shows the distribution of control means for the entire set of control data within each laboratory for corpora lutea per pregnant female, when the mean is the average for one week, one experiment. Figure 2 shows the distribution for mean total implantations within each laboratory. These two parameters are normally distributed for each female. The distribution of control means of these two parameters is fairly comparable for the three laboratories.

Figures 3 and 4 show the distribution of weekly means for preimplantation losses and dead implantations. Late deaths and early deaths were measured for each animal. However, since late deaths were extremely rare, they were included with early deaths and analyzed as dead implantations. Although these two parameters were extremely skewed in individual females, the means are more normally distributed, as would be expected.

The mean control values for both preimplantation losses and dead implantations are higher for laboratory 2 than laboratory 3. This indicates that controls must be run for each laboratory, and although negative control data from one laboratory certainly give an indication of what would be expected in controls from another laboratory, testing of compound effects would necessitate controls within the laboratory doing the test.

In FDA's contracting work, a file of historical control data is being built within each contracting laboratory. Each set of new controls is compared against the historical controls for each week. If the new data are

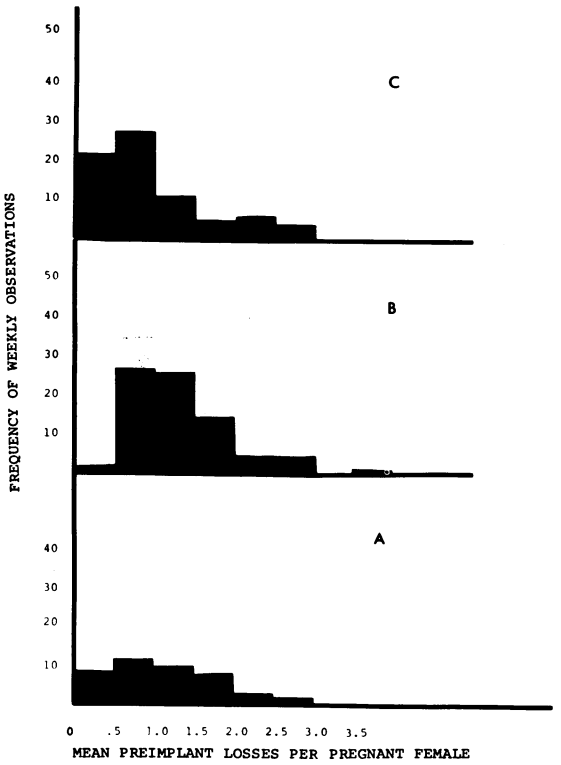

FIGURE 3. Distribution of weekly mean preimplantation losses per pregnant female: (*A*) laboratory 1; (*B*) laboratory 2; (*C*) laboratory 3.

not significantly different ($p>0.05$) from the historical data, the new group is added to the historical file. Each test compound is compared to both the historical control and the concurrent control. Significance of results with concurrent controls is interpreted in light of the significance against historical controls. However, the importance of concurrent controls cannot be overemphasized.

The Necessity For Repeats

In evaluating the results of dominant-lethal testing, one of the areas of major concern is the necessity for replication of tests. Of the 22 GRAS compounds tested and evaluated by contract laboratories to date, only one is a strong presumptive positive for preimplantation and postimplantation loss. The results from eight compounds clearly showed no dominant-lethality and could therefore be regarded as conclusively negative.

The compounds which are considered questionable positives are listed in Table 1. These questionable positives are not based simply upon one dose level in one week showing a significant effect, but rather upon an analysis of variance over all weeks in some instances and in other instances upon clustering of significant effects around one or two weeks.

Thirteen compounds generated enough of a response in one of the parameters measured to preclude a statement regarding their safety as determined by the dominant-lethal assay. Five compounds from laboratory 2 were in this questionable category and eight from laboratory 3. In both laboratories, the parameter with the greatest number of questionables was preimplanation loss as a result of subacute administration. Second in the order of frequency was the measure of postimplantation loss from subacute

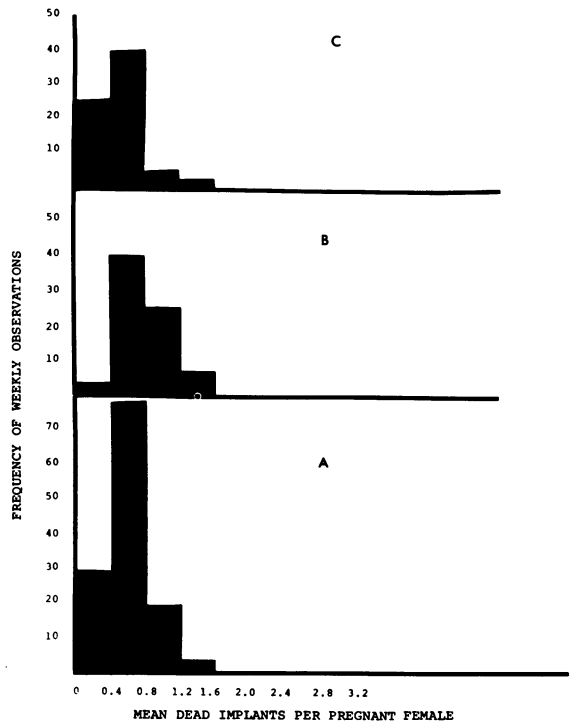

FIGURE 4. Distribution of weekly mean dead implants per pregnant female: (*A*) laboratory 1; (*B*) laboratory 2; (*C*) laboratory 3.

Table 1. Number of GRAS compounds considered "questionable positive" and the response elicited by each.

	No. of compounds examined [a]	GRAS compounds producing preimplantation loss		GRAS compounds producing postimplantation loss	
		Acute	Subacute	Acute	Subacute
Laboratory 2	14 A thru N		A,B,M	C	M,N
Laboratory 3	8 O thru V	P,S,V	O,R,S,T,U	P,Q,V	Q,S,U,V

[a] Letters indicate codes for GRAS compounds: A, FD&C Red No. 2; B, calcium carrageenan; C, gum arabic; M, sodium metabisulfite; N, ional C.P.; O, ammoniated glycerrhizin; P, ammonium saccharin; Q, sodium carrageenan; R, calcium saccharin; S, potassium nitrate; T, potassium nitrite; U, sodium saccharin; V, sodium bisulfite.

treatment. However, the degree of significance for preimplantation loss in most cases was at the 0.01 level, whereas that of the postimplantation loss was at the 0.05.

The investigator faced with making a safety evaluation based upon the configuration of the data has an immediate problem. Due to the lack of consistency, there is a need to repeat the studies in question. Approximately 50% of the compounds tested fell into the questionable category. If one were to extrapolate to the total number of compounds on the GRAS list currently scheduled to be tested for mutagenicity this would represent approximately 45 compounds.

Clearly, there is a need for a more refined method of looking at dominant-lethal effects; one which would not give an inordinate number of false negatives or false positives.

From experience, it appears that few, if any, compounds exert their effect in one week only. Therefore, an analysis of variance over specific spermatogenic stages may be the better method, coupled with weekly analysis. For example in the rat, the first 2 weeks could be analyzed, the following 3, the next 3 and the final 2. These reflect periods in which effects on spermatozoa, spermatids, spermatocytes, and spermatogonia, respectively, are manifest (*4*). This method of analysis could not serve as an adequate protocol for chronic testing because the extended treatment would obliterate the assessment of stage sensitivity. There would be no rationale for the weekly grouping. If the analysis of variance for specific stages is the best method of determining dominant lethality and of limiting the number of questionables, while chronic exposure of animals is the best dosage regimen for safety evaluation, (that is, for substances to which we are chronically exposed), a conflict arises, and the analysis of variance test cannot be employed in what is considered the most statistically powerful and biologically meaningful fashion.

Use of the Analysis of Variance

In an attempt to decrease the incidence of spurious effects at one week only, an analysis of variance was accomplished across all doses and all weeks on 23 compounds. The model used in the analysis was

$$\chi_{ijkl} = \mu + \underset{\text{dose}}{\alpha_i} + \underset{\text{male (dose)}}{\beta_{l(k)}} + \underset{\text{weeks}}{\gamma_j} + \underset{\text{weeks} \times \text{dose}}{\lambda_{ij}} + \underset{\text{male} \times \text{weeks (dose)}}{\omega_{ij(k)}} + \underset{\text{female (dose} \times \text{weeks} \times \text{male)}}{\epsilon_{l(ijk)}}$$

Doses and weeks were considered fixed effects; males were considered random effects.

The data for corpora lutea and total implantations per pregnancy are normally distributed and were not transformed. Preimplantation and post-implantation losses per pregnancy were essentially Poisson-

distributed and were transformed by using the Freeman-Tukey square-root transformation.

The results of the analysis indicate that the week effect is the greatest source of variation within an experiment. This was true of all four parameters. The week variation was greater for preimplantation losses than for dead implantations, but was significant for both. The need for analysis of the data on an individual week basis, or at least on the separate stages of spermatogenesis, is indicated.

The analysis showed that the variability between males is accounted for by the variability between females mated to the same male. The between male component of variance is essentially zero, indicating that analysis on a per female basis is appropriate, even though the males are the treated experimental unit.

The analysis of variance indicated that the group by week interaction was generally not significant for any parameters, that is, the trend across dose levels was not significantly different from week to week, although it was variable and there were inconsistencies.

The analysis of variance was also carried out for each stage of spermatogenesis: weeks 1–2; weeks 3, 4, and 5; weeks 6, 7, and 8. These analyses indicated significant variability between weeks, even within one stage of spermatogenesis. For at least two compounds there was a significant dose-related effect in one stage of spermatogenesis with no corresponding effect in another stage. This further illustrates the need for analysis at each stage of spermatogenesis. It appears that a dominant-lethal effect at only one stage of spermatogenesis cannot be counted as a negative effect.

The mean residual error variance for an experiment, that is, between the two females mated to the same male, was 1.0 for dead implants with a range of 0.34–4.78. These figures are based on pooling the data for all experiments over all dose levels, all weeks. The error variance for individual weeks had a considerably greater range. For preimplantation losses the mean residual error variance was 1.62 with a range of 0.36–5.85. The increase in the mean preimplantation loss for treatment over control that could be detected as significant at the 5% level (assuming 20 females per dose) based on the mean residual variance is 0.9. Across all weeks, the increase in mean preimplantation losses per female that could be detected as significant (based on 160 females per dose) is 0.3. For dead implants, these differences would be 0.7 for an individual week, and 0.2 across all weeks. As was pointed out before, these figures would vary slightly for different laboratories. The strain of rat would also affect the error variance.

Correlations between Parameters

Correlations between corpora lutea and preimplantation losses per female, preimplantation loss and dead implantations, and

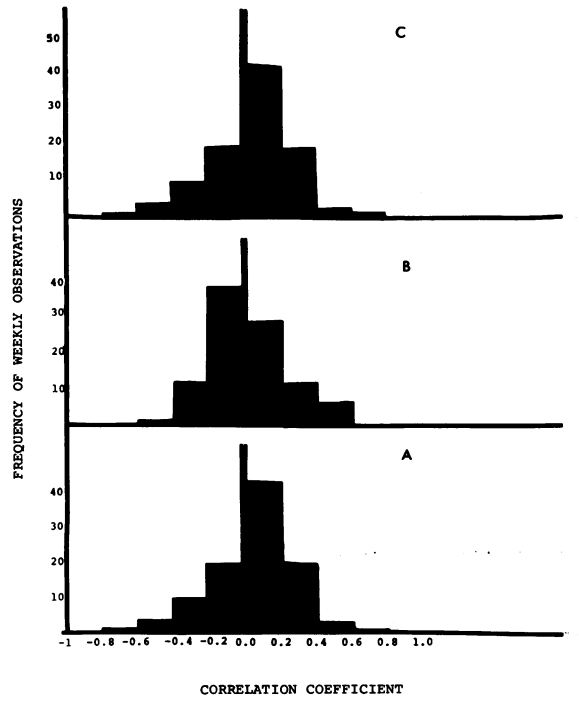

FIGURE 5. Distribution of correlations between the parameters for control rats: (*A*) dead implantations with total implantations; (*B*) preimplantation loss with dead implantations; (*C*) preimplantation loss with corpora lutea.

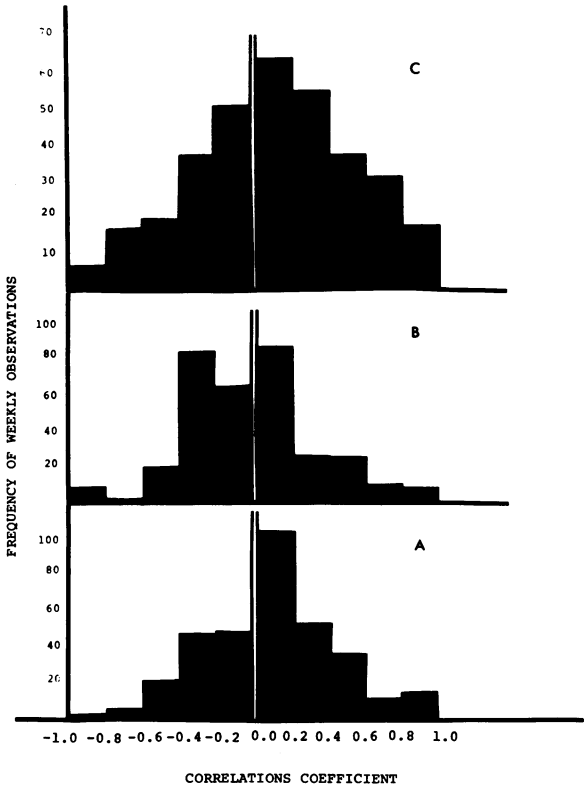

FIGURE 6. Distribution of correlation between parameters for rats (treatment and positive control): (*A*) dead implantation with total implantations; (*B*) dead implantations with preimplantation losses; (*C*) preimplantation loss with corpora lutea.

dead implantations and total implantations per female were calculated for each week of each experiment. A frequency distribution of the correlations is shown in Figures 5 and 6 for rats and Figure 7 for mice. There is a tendency for dead implantations to be positively correlated with total implantations, although this correlation is not a strong one. This may indicate that the commonly used mutagenic index defined as dead implantations/total implantations is a valid index to use in evaluating the dominant-lethal test. This positive correlation would indicate that the dead implantations per female should be adjusted by the total implantations for that female, which is essentially what the use of the index does. These observations are contrary to what has been reported in the work of Epstein et al. (*2*), who reported no correlation between dead implantations and total implantations when using weekly means. The statistician may object to use of indexes on purely aesthetic grounds because of the problems of variance estimates for a ratio, but this does not mean that the index may not be the best interpretation of the data. In any case, the FDA has been analyzing all data on dead implantations on a dead implantations per female basis and on a dead implantations/total implantations per female basis, and the inferences as to significance are comparable.

The preimplantation losses are strongly correlated with corpora lutea count. This would indicate the need for preimplantation losses to be adjusted by corpora lutea count for each female.

Preimplantation losses and dead implantations appear not to be correlated. The higher

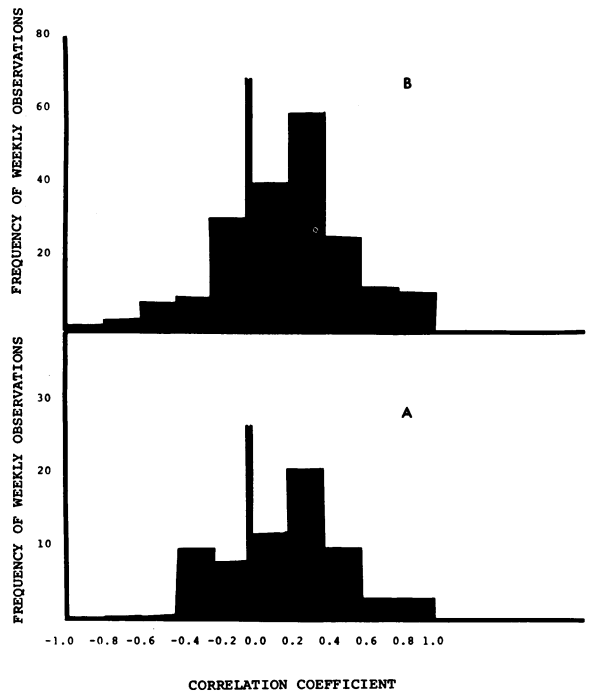

FIGURE 7. Distribution of correlation between dead implantations and total implantations for mice: (*A*) control mice; (*B*) treatment levels and positive control.

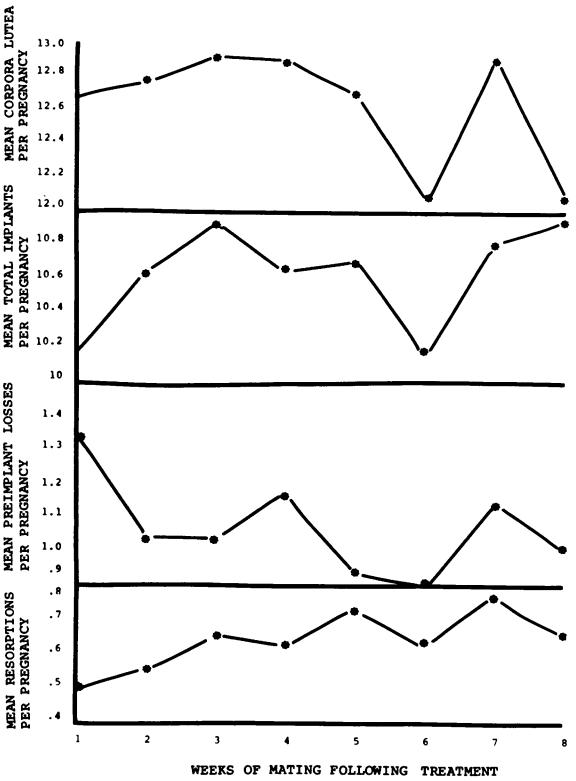

FIGURE 8. Trends in weekly parameter means.

correlations are positive; however, no definite trend is evident. One may expect that the same animal which shows preimplantation losses also shows dead implantations. With a strong positive compound, however, if a heavy preimplantation loss is shown in one female, dead implantations may be few, since there are fewer implantations to be affected.

There was considerable week to week variability in all parameters within an experiment. No consistent trends were apparent, however, from week to week. The control data were combined across all experiments by week to determine if a trend emerged. Figure 8 shows the weekly trends for each parameter. The mean dead implantations per pregnancy shows an interesting cyclic trend. The variability in each point on the chart is the variability between experiments. This variability is much greater than the overall elevation in weekly means, thus this trend is probably not significant.

The mean preimplantation losses show no consistent trend across weeks (Fig. 8). The high value for week 1 may indicate decreased viability of the sperm in the male which is being mated for the first time.

The mean total implantations and the mean corpora lutea show no consistent trends with weeks of mating. Although these data might seem to suggest pooling negative controls over weeks and comparing each dose level to the pooled data, this procedure may not be valid, as will be shown later in this communication.

The Preimplantation Problem

The propensity of the test for showing significant preimplantation loss, especially in the earlier weeks, presents another problem. Table 2 shows the number of questionable compounds exhibiting preimplantation loss in the first two weeks. Four of the eight compounds shown in this table (sodium metabisulfite, ammonium saccharin, potassium nitrite, sodium saccharin) did show a correlation to postimplantation loss at the same week or the week immediately following. This leads us into a comparison of the variability in the preimplantation and the postimplantation parameters among experiments.

The mean resorption rate is consistent from week to week and from experiment to experiment as shown in Figure 3. The preimplantation losses (Fig. 4) are more variable in the controls, particularly from week to week in the same experiment. The latter observation may be due to difficulties in reading corpora lutea. For this reason, the resorptions are a better measure of dominant lethal effects. This observation has been reported before by Epstein et al. (*2*). However, it appears that with many compounds, as shown in Table 2, an effect is seen in preimplantation losses with no corresponding effect on dead implantations. Preimplantation losses may be a more sensitive indicator of toxic activity than dead implants, though more difficult to measure.

Investigators familiar with dominant lethal testing acknowledge that preimplanta-

Table 2. Number of questionable GRAS compounds showing preimplantation loss in the first two weeks

GRAS compounds showing preimplantation loss	
Acute	Subacute
P,R	O,R,T,U A,B,M

[a] Letters indicate codes for GRAS compounds: A, FD&C Red No. 2; B, calcium carrageenan; C, gum arabic; M, sodium metabisulfite; N, ional C.P.; O, ammoniated glycerrhizin; P, ammonium saccharin; Q, sodium carrageenan; R, calcium saccharin; S, potassium nitrate; T, potassium nitrite; U, sodium saccharin; V, sodium bisulfite.

tion loss in earlier weeks when not coupled with postimplantation loss in later weeks or at lower dosage levels is meaningless, because it cannot be correlated with the general basis of the dominant-lethal test; that is, it cannot be correlated with postimplantation losses. This philosophy is manageable for persons interested only in precise genomal effects, but what is the proper posture for those individuals concerned with safety evaluation?

When preimplantation loss occurs without concomitant postimplantation loss, certain questions immediately arise, such as: What are the responsible events? Do they involve true dominant-lethality, or simply aspermia, or oligospermia? The etiology of the preimplantation loss must be investigated, before any decision regarding safety evaluation can be made. The experimental procedures necessary to obtain the answers are very tedious, therefore prolonging the decision-making process.

One possible approach to the preimplantation loss problem is to premate males which should eliminate the questionable results in the first week. There is some evidence from statistical evaluation of males and females in control groups that this problem is due to the males.

Certain results of the analysis of variance may bear on the above problem. For three compounds, there was a significant difference between males at weeks 1 and 2 only, in at least one parameter. The males not having been mated previous to the study, react differently for weeks 1 and 2, particularly week 1, than for the remaining weeks of mating. This may seem to indicate the desirability of premating. However, the significant effects of the test compound seem to be more pronounced in the first two weeks of mating. The virgin male may be more susceptible to the dominant-lethal effect, and premating would mask the effect.

One larger question arises, however, and that is why should virgin males be more sensitive to causing preimplantation losses than experienced breeders. One should therefore be cautious of preimplantation losses in the first week when analysis shows this to be due to male variability. At the same time we cannot totally discount effects in treated groups in the first week, which are highly significant.

The Dose-Response Problem

Another area of concern has to do with the fact that most of the positive effects seem to occur at the lowest dosage level tested. It is generally accepted among pharmacologists that a compound should exhibit a dose response; that is, there is a direct relationship between the dose of a compound and the magnitude or intensity of the response it elicits. This point is especially critical for safety evaluation, as one would expect.

With approximately half of the compounds studied, the most important parameters were not dose-related. When these differences are significant as tested by analysis of variance, they cannot be attributed to individual males or females, because variability between females within a group is taken into account in the analysis of variance. The significance of effects at low or intermediate dose levels compared to controls seems too pronounced in some cases to be ignored.

To illustrate this point, the results obtained for potassium nitrite in the subacute study for preimplantation losses are shown in Table 3. The results at the low dose level are too far above the ranges of all negative control data to be ignored. The experience with dominant-lethal data is not enough to

Table 3. Mean preimplantation losses for the Subacute study on potassium nitrite of laboratory 3.

	Mean preimplantation losses			
Week	Negative control	Low dose	Intermediate dose	High dose
1 and 2	0.53	2.47	1.39	2.12
3, 4, and 5	1.25	2.59	2.24	1.69
6 and 7	0.96	2.29	1.55	1.52

say what models may be applicable, and models assuming a (linear) dose-related response may not be valid (for many of the GRAS compounds) in the dose ranges studied. The classical compounds used in dominant-lethal studies, such as TEM, TEPA, and EMS, all have well-known mechanisms of action which demonstrate definite dose-response relationships (*3, 5*). Compounds having dissimilar mechanisms of action from the alkylating agents may not exhibit as definitive a dose response.

When a compound does not exhibit a direct dose-response relationship, it is generally assumed the compound is not acting pharmacologically. In other words, there are other unknown factors contributing to the response. We must seriously consider the possibility of hormonal effects or of abnormal or disturbed metabolism and pharmacokinetics at the maximum tolerated dosage (MTD) as a possible source of the problem. If this is the case, then testing at lower dosage levels is essential. Therefore, we must modify our protocols to either include additional dosages, or, on the basis of experience, choose levels below the LD_5 as the highest level to be tested.

It may be, however, that the MTD is not causing the above problem. Whatever the cause, this is an area of major concern, for without a dose-response effect it is not certain whether the compound is responsible for the effect. At some later point in time we may question the relevance of a dose response, as now with carcinogens, but the present state of the art dictates that we relate these effects to pharmacological and toxicological principles.

Extrapolation of results is not discussed here. Sufficient models have not been developed, and all relevant problems have not been explored for valid extrapolation of data to lower dose levels. The only investigation in which a model was used employed the probit model for the proportion of females with one or more dead implants in a study on TEPA and METEPA (*5*). The authors stated that their model was not a sufficient fit and there were significant departures from the assumed model.

Expansiveness of Dominant-Lethal Data as a Problem

The inordinate amount of time required to examine adequately results from dominant-lethal testing is another area of concern. Any system promulgated for safety evaluation should have as one of its primary features a minimum amount of information necessary to accomplish its purpose. We realize that a careful and sometimes laborious examination of data must take place in order to draw meaningful conclusions. There should be, however, a development of better methods of evaluation to reduce the expansiveness of data now required for dominant-lethal testing.

In addition to the above consideration, departures from adequate performance on the part of technical personnel increase the probability of obtaining spurious results. Table 4 is an illustration of possible technical difficulties which influence the pattern of control results. Calcium saccharin shows negative control preimplantation losses which are much smaller than historical controls. We might, therefore, expect significance for the test compound which is entirely due to the low negative control. However, here is a case where the preimplantation losses for every level of the test compound and even the positive control are significantly lower than the historical negative control values. Similar results were obtained for the same compound on the subacute study. Therefore, significant differences between concurrent controls and test groups cannot be completely discounted be-

Table 4. Mean preimplantation losses for the acute study on calcium saccharin of laboratory 3.

Week	Mean preimplantation losses					
	Historical control	Negative control	Low dose	Intermediate dose	High dose	Positive control
4	0.6	0.1 [a]	0.3	0.3	0.1	0.8
5	0.6	0.0 [a]	0.1 [a]	0.3	0.1 [a]	0.0 [a]
6	1.0	0.0 [a]	0.3 [a]	0.3 [a]	0.1 [a]	0.1 [a]
7	0.6	0.0 [a]	0.1 [a]	0.2 [a]	0.4	0.3
8	1.0	0.2 [a]	0.2 [a]	0.1 [a]	0.1 [a]	0.1 [a]

[a] Indicates significant decreases ($p<0.05$) from historical control.

cause of unusually low or high control values. In general, mean values for parameters tend to be low at all dose levels when they are low for controls and high when controls are high, because of variability which is due to unexplained causes.

Relationship of Dominant Lethality to Heritable Effects

If the primary focus of our concern in mutagenesis is the threat to future generations, in terms of effects on the gene pool, then the dominant-lethal test has yet another limitation when used for safety evaluation. The test, by definition, measures only death: death in the preimplantation stage and/or in the postimplantation stage. The concern is that there may be ways in which dominant lethality can be induced without producing genetic events with potential transmissibility to live progeny.

It has been pointed out that there is a genetic disease burden in the population (*6*). In order to determine whether agents capable of increasing this burden exist at the present time we need systems and approaches capable of detecting such effects. The dominant-lethal system measures such a narrow spectrum of genetic effects (gross chromosomal aberrations manifested as dead implantations) that it is severely limiting in the type of information one can glean from it.

There have been attempts to examine live progeny for cytogenetic abnormalities in order to expand the amount of information that the system provides. To our knowledge, only one group has been successful with this technique (*7*). There is a need for more information in this area and for additional approaches which measure genetic effects in progeny.

REFERENCES

1. Anonymous. FDA reveals testing priorities for GRAS substances. Food Chem. News **13** (No. 11): 18, (1971).
2. Epstein, S. S., et al. Detection of chemical mutagens by the dominant lethal assay in the mouse, Toxicol. Appl. Pharmacol. **23**: 288 (1972).
3. Bateman, A. J. The induction of dominant lethal mutations in rats and mice with triethylenemelamine (TEM). Genet. Res. Camb. **1**: 381 (1960).
4. Jackson, H. Antifertility Compounds in the Male and Female. Charles C Thomas, Springfield, Ill., 1966, pp. 20–23.
5. Epstein, S. S., et al. Mutagenic and antifertility effects of TEPA and METEPA in mice. Toxicol. Appl. Pharmacol. **17**: 23 (1970).
6. Childs, B. Miller, S. M., and Bearn, A. G. Gene mutation as a cause of human disease. In: Mutagenic Effects of Environmental Contaminants, H. E. Sutton and M. I. Harris, Eds., Academic Press, New York–London, 1972.
7. Datta, P. K., Frigger, H., and Scheiermacher, E. The effect of chemical mutagens on the mitotic chromosomes of the mouse, *in vivo*. In: Chemical Mutagens in Mammals and Man. F. Vogel and G. Rohrborn, Eds., Springer Verlag, Berlin–New York, 1970.