Statistical modeling and analyses of a base-specific *Salmonella* mutagenicity assay

Walter W. Piegorsch [a,*], Susan J. Simmons [a], Barry H. Margolin [b], Errol Zeiger [c],
Xavier M. Gidrol [d], Pauline Gee [d]

[a] *Department of Statistics, University of South Carolina, 216 LeConte College, Columbia, SC 29208, USA*
[b] *Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599, USA*
[c] *National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA*
[d] *Xenometrix, Inc., Boulder, CO 80301, USA*

Abstract

Statistical features of a base-specific *Salmonella* mutagenicity assay are considered in detail, following up on a previous report comparing responses of base-specific *Salmonella* (Ames II™) strains with those of traditional tester strains. In addition to using different *Salmonella* strains, the new procedure also differs in that it is performed as a microwell fluctuation test, as opposed to the standard plate or preincubation test. This report describes the statistical modeling of data obtained from the use of these new strains in the microwell test procedure. We emphasize how to assess any significant interactions between replicate cultures and exposure doses, and how to identify a significant increase in the mutagenic response to a series of concentrations of a test substance. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Ames II strains; *Salmonella typhimurium*; Complementary log–log link function; Culture-by-dose interaction; Fluctuation test; Generalized linear model; *His⁻* mutant tester strains; Many-to-one testing; Multiple comparisons with a control; Statistical methods

1. Introduction

The *Salmonella*/microsome reversion assay has been used extensively in genetic toxicology testing [1–4]. The procedure employs bacterial tester strains that identify the reversions of missense and small frameshift mutations in the *his* operon. Despite the widespread acceptance of this test, new *Salmonella* tester strains are constantly being developed and studied. In a previous report [5], we discussed the use and validity of a series of six new *his⁻* mutant strains (TA7001, TA7002, TA7003, TA7004, TA7005 and TA7006), each of which was designed to revert to *his* independence by unique base-pair substitutions [6]. The TA700X series of tester strains has been designated ''Ames II™'' (Xenometrix, Boulder, CO, USA).

In order to help automate the data collection process, and to allow the assay to be adapted to high-throughput, robot-controlled procedures, a modified fluctuation protocol [7,8] has been developed

* Corresponding author. Tel.: +1-803-777-7800; fax: +1-803-777-4048.
E-mail address: piegorsc@stat.sc.edu (W.W. Piegorsch).

for use with the TA700X tester strains (AMAX™: Ames II Mutagenicity Assays by Xenometrix). In our previous report on the AMAX™ procedure, the performance of these strains was compared with the results obtained using the traditional *Salmonella* tester strains in a preincubation procedure [4,5]. Thirty coded chemicals (five of which were duplicates with different code numbers) were tested in the individual strains TA7001–TA7006 to identify mutagens that produced base-pair substitutions, in a mixture of these six strains, and in the traditional strains TA98 and TA1537 to detect frameshift mutagens. All testings were done using a modified liquid fluctuation test procedure as designated in the AMAX™ protocol. The results were compared to results obtained previously using a preincubation protocol with strains TA98, TA100, TA1537 (or TA97) and TA1535 [5].

One important component of the validation effort in this previous study was proper recognition and adjustment for the various sources of statistical variability in the AMAX™ data. However, the statistical analysis of those data was conducted at an introductory level only. The current manuscript presents more detailed statistical aspects of the AMAX™ assay, based on the data from the previous validation study. The issues discussed include the aspects of the sampling protocol, the possible interactions between cultures and exposures to mutagens, and the identification of exposure-related effects using a form of generalized linear statistical model. Recommendations are made for statistical analysis of microwell fluctuation test (e.g., AMAX™) data from these *Salmonella* strains.

2. Methods

2.1. AMAX™ protocol and experiments

Eight individual *Salmonella typhimurium* strains and a mixture of the base-specific strains were employed in our previous report, where the strains and their genotypes are described [5]. Each of the six base-specific strains (TA7001–TA7006) carries a target missense mutation in the histidine operon, which reverts to prototrophy by base-substitution events unique to each strain. Strains TA7001,

TA7002 and TA7003 detect base substitutions at A:T base pairs, while TA7004, TA7005 and TA7006 detect base changes at G:C base pairs.

The TA700X strains are not designed to identify mutagens that induce only frameshift mutations. Thus, routine tests of unknown chemicals must also use one or two of the traditional strains that are designed to detect frameshift mutations, viz. TA98 and TA1537. For the purposes of the previous study [5], these two strains were used in the AMAX™ procedure along with the TA700X strains.

All 30 coded chemicals were studied without exogenous metabolic activation. Those that appeared to be negative in the initial test (18 out of 30) using a subjective, non-statistical assessment were also tested in the presence of S9 (4.5% rat liver S9 fraction), in order to allow for possible metabolic activation. Protocols used to generate the data analyzed herein were detailed in our previous report [5]. Data were collected for a positive control, five doses of each test chemical (including a zero-dose control), and for a solvent control. Each culture was treated independently with the test chemical in a total exposure volume (Exposure Medium) of 0.5 ml per dose, in triplicate. Following exposure, 2.5 ml of selective medium (Indicator Media) was added and 50 μl aliquots were dispensed into 48 wells of 384-well microtiter plates, either manually or using a ML 2200 pipeting station (Hamilton, Reno, NV). Each experiment was performed using three independent cultures by one of two teams of experimenters on the same day and/or up to 3 weeks later. Each culture was inoculated from separate frozen vials, although each vial of cells used for the study was from the same production lot. Each chemical was tested up to 5 mg, the maximal tolerated dose, or the limits of the solubility, whichever was reached first.

A concern noted previously [5] with this assay system is that under otherwise-homogeneous preparation, a few cultures may exhibit an extreme response, called *jackpot mutations*. Jackpots result from reversion events that occur early in the growth of the overnight culture, such that the wild-type revertant population expands during the overnight growth and pre-exists in the culture before exposure to the test agent. High spontaneous counts are usually attributed to jackpot mutations, and may obscure any increase in reversion events caused by the test

agent. The frequencies of jackpots for the different strains correspond to each strain's individual inherent genetic instability. For example, 6% of cultures of TA1537 exhibited high spontaneous positive wells, while TA7001 and TA7006 cultures did not exhibit any jackpot mutations.

Where jackpot mutations were observed to obscure the results, the experiments were repeated. The final data set may have included experiments where jackpot mutations did not affect the overall results and were therefore included in the analysis. Statistically, the presence, or even the possibility of jackpots associated with individual cultures must be viewed as a source of additional variability in this assay, and some adjustment or correction must be made to account for it in the analysis (see below).

2.2. Statistical models

The basic experimental design of the AMAX™ assay for any given chemical (in the presence or absence of S9) consists of an exposure regimen involving an untreated control, a solvent control, a positive control, and four increasing doses of the chemical. The doses are indexed via $k = 0, 1, \ldots, 4$, where $k = 0$ indicates the untreated (zero-dose) control. Each dose was tested in triplicate in aliquots of each independent culture, and each experiment was performed three times, thus using three different overnight cultures of the *Salmonella* strains. This design yielded three experiments, each consisting of a series of solvent controls and chemicals tested in triplicate.

Statistically, the experimental response is a set of dichotomous outcomes in 48 wells of a 384-well plate. Each well is scored for growth of his^+ reverent bacteria (growth is indicated by a positive yellow well, vs. an otherwise purple well), and we record 0 for purple/negative growth, or 1 for yellow/positive growth.

For the ith plate ($i = 1,2,3$) in the jth culture ($j = 1,2,3$) at the kth dose level, we denote by $X_{ijk}/48$ the proportion of mutagenic wells observed for a given combination of chemical/strain/S9. In this design, cultures are *crossed* with dose levels, representing a form of *two-way design* [9]. Because

each culture appears at each dose level an equal number of times, the two-way design is *balanced*.

The standard statistical model for X_{ijk} is the binomial distribution [10]. Data across plates within any culture/dose combination are assumed homogeneous; we pool these values into a single summary proportion, say $Y_{jk}/144 = \{X_{1jk} + X_{2jk} + X_{3jk}\}/144$, and write $Y_{jk} \sim \text{Binomial}(144, \pi_{jk})$, where π_{jk} is the (unknown) probability of mutation in the jth culture at the kth dose level.

To adjust for the effect of jackpots and to analyze more generally the mutant proportions for any chemical/strain/S9 combination, we chose to take advantage of the balanced feature of the treatment design through a form of analysis of variance (ANOVA) appropriate for binomial proportions. Specifically, we applied a generalization of the common ANOVA model, known as the Generalized Linear Model, or GLiM [11]. A GLiM can involve two specialized components: (i) a statistical model for the data other than the normal (in our case, based on the observed binomial proportions), and (ii) a function that links the unknown mutation probability π_{jk} to the features of the treatment design. For component (i), we incorporated the binomial assumption on Y_{jk}; for component (ii), we recognized an experimental feature that induces a specialized form of link function. Specifically, we assumed that the random number of mutations per well, say U_{ijk}, is described by a Poisson distribution with unknown, positive mutation rate $\lambda_{ijk} > 0$. Of course, U_{ijk} is unobservable; all that is recorded is whether *any* mutations occurred in a given well. Thus, the per-well Poisson variate U_{ijk} is truncated to the dichotomous observation X_{ijk}, which equals 1 if any mutations occurred in that well, and 0 otherwise. Denote the probability that X_{ijk} equals 1 by ϕ_{ijk}. Then under this truncated Poisson model, ϕ_{ijk} is given by

$$\phi_{ijk} = P\left[X_{ijk} = 1\right] = P\left[U_{ijk} \geq 1\right]$$

$$= 1 - P\left[U_{ijk} = 0\right] = 1 - \exp\{-\lambda_{ijk}\},$$

the latter equality following from the basic form of the Poisson probability mass function [10]. Assuming that triplicate wells are homogeneous, and thus, that no per-plate effects are present, we may drop the i subscript in λ.

The usual ANOVA formulation for the mean response under a two-way design is a linear combination of the unknown effect parameters:

$$\mu + \gamma_j + \delta_k + \psi_{jk}, \tag{1}$$

where for our setting, μ is an overall effect parameter, γ_j is a term for the culture effect, δ_k is a term for the dose effect, and ψ_{jk} represents a possible *interaction* between culture and dose. In Eq. (1), γ_j represents a ''blocking'' term that accounts for any jackpot-related culture-to-culture variability.

In most ANOVA settings, the linear expression in Eq. (1) is set equal to the mean response and analyzed accordingly. Here, however, equating (1) to the mean mutation rate λ_{jk} fails to account for the constraint that λ must be positive. To overcome this, we can model λ_{jk} itself as an exponential form:

$$\lambda_{jk} = \exp\{\mu + \gamma_j + \delta_k + \psi_{jk}\}. \tag{2}$$

Under Eq. (2), λ_{jk} is guaranteed to be positive for any realization of the linear effect parameters in Eq. (1).

Collecting all of these model components together, the resulting GLiM may be written as $Y_{jk} \sim$ Binomial$(144, \pi_{jk})$, where $\pi_{jk} = 1 - \exp\{-\exp[\mu + \gamma_j + \delta_k + \psi_{jk}]\}$ and Y_{jk} is the sum over the three replicate plates of the mutant wells at the jth culture ($j = 1,2,3$) and the kth dose ($k = 0,1,2,3,4$) for any chemical compound being tested. Inverting this model for π gives:

$$\log\{-\log(1 - \pi_{jk})\} = \mu + \gamma_j + \delta_k + \psi_{jk}, \tag{3}$$

which is known as a complementary log–log GLiM [10,11].

2.3. Statistical analyses

Under the complementary log–log GLiM in Eq. (3), we can assess whether there is an effect due to the dose after correcting for possible culture-to-culture variability. As is well known, however, it is inappropriate to test for any main effects due to individual factors, such as dose, in the presence of a significant interaction. Indeed, P-values for testing the main dose effect possess no sensible interpretation if given in the presence of a significant interaction (see Ref. [9]). Thus, before assessing the dose-related effects for any chemical/strain/S9

combination under study, we first must test the null hypothesis of no interaction. This translates to H_o: $\psi_{jk} = 0$ for each j,k, vs. an alternative hypothesis that $\psi_{jk} \neq 0$ for some combination of j and k. H_o may be assessed via a likelihood ratio test, which is similar in form to the usual F-test for the interaction in a block design/ANOVA. Under our design, the likelihood ratio statistic, G_ψ^2, for the culture $\times$ dose interaction is distributed as χ^2 with $(5 - 1)(3 - 1) = 8$ degrees of freedom (df). Departure from H_o is indicated if the P-value $P = P[\chi^2(8) \geq G_\psi^2]$ drops below a pre-assigned α-level.

The likelihood ratio computations must be performed on a computer. We employ the SAS® computer package via its GLiM procedure Proc Genmod [12]. [Sample SAS code for fitting the complementary log–log model under Eq. (3) is given in Fig. 1. To identify the likelihood ratio statistic in Proc Genmod, invoke the Type 1 option in the Model statement, being sure to order the model components with culture first, dose second, and culture ∗ dose last.]

If a particular chemical/strain/S9 combination of interest tests negative for culture $\times$ dose interaction, we can move directly to testing the main effect due to dose. This translates to the null hypothesis H_o: $\delta_0 = \delta_1 = \cdots = \delta_4$. The alternative hypothesis, H_a, is that some departure from pure equality exists among the δ_ks. The SAS output again provides a likelihood ratio statistic, G_δ^2, that is referenced to a χ^2 distribution with $5 - 1 = 4$ df; the corresponding P-value is $P = P[\chi^2(4) \geq G_\delta^2]$. When P drops below a pre-assigned α-level, there is a departure from H_o, and hence, some dose effect is indicated.

If significant, the dose effect can take on many forms. Of interest in a mutagenicity testing setting is the set of one-sided departures from the control, H_{ak}: $\delta_k > \delta_0$ ($k = 1,\ldots,4$). If any such alternative hypothesis is significant at the kth dose, it indicates a significant, dose-related mutagenic effect. Notice that we can also write H_{ak} as H_{ak}: $\delta_k - \delta_0 > 0$.

To test against these one-sided alternatives, SAS' Proc Genmod is particularly useful. First, refit the model after removing the culture $\times$ dose interaction. Then, under the Proc Genmod output for Analysis of Parameter Estimates, find the point estimates of the dose effect parameters. Information in these point estimates is employed in testing against the alternatives H_{ak}: $\delta_k - \delta_0 > 0$.

```
*Sample SAS code to fit complementary log-log GLiM;
*Includes test for culture*dose interaction;

 data work1;
 infile 'data_file_name_here' ;
     *Ensure that the control level is coded as last;
 input  dose culture y n;

 proc genmod;
  class culture dose;
  model y/n = culture dose culture*dose / dist=b link=cll type1;

* if interaction is insignificant, re-fit with interaction term removed;
  proc genmod;
   class culture dose;
   model y/n = culture dose / dist=b link=cll ;
  run;
```

Fig. 1. Sample SAS Proc Genmod complementary log–log GLiM code for fitting a two-way model with interaction terms.

Users are cautioned to proceed carefully, however, since the outputs from the Proc Genmod analysis do not produce point estimates of each δ_k. As is common with two-factor models such as Eq. (1), there are certain identifiability constraints required to complete the fit [9]; SAS defaults to a reference-cell constraint, where the last δ-parameter is set equal to zero. As a result, the reported SAS ''parameter estimates'' turn out to be estimates of the *differences* $\delta_k - \delta_4$, $k = 0, 1, \ldots, 4$. Admittedly, this SAS artifact can add confusion to the analysis, but it is a necessary consequence of the two-factor model being employed in Eq. (1).

For testing against H_{ak}, this differencing can nonetheless be manipulated to our advantage. The quantities we wish to study are differences from the control, so if in the SAS input code, we enter and code the dose levels such that the control level is *last* — say, discard the $k = 0$ subscript and *relabel the control as $k = 5$* — the corresponding ''parameter estimates'' will relate to the differences $\delta_1 - \delta_5$, $\delta_2 - \delta_5, \ldots, \delta_4 - \delta_5$. These are the precise differences we wish to estimate.

Under this relabeled scheme, denote the SAS estimates as $d_k - d_5$. The SAS output also supplies standard errors, $\mathrm{se}[d_k - d_5]$, from which a Wald statistic for testing against H_{ak} is calculated as $W_k = (d_k - d_5)/\mathrm{se}[d_k - d_5]$. This is referenced in large

samples to a standard normal distribution, with corresponding one-sided P-value $P_k = 1 - \Phi(W_k)$. [The function $\Phi(z)$ is the cumulative distribution function of the standard normal.] However, further caution is advised here: the SAS output also reports P-values under the heading Pr > Chi, but these are actually *two*-sided P-values. To convert them to the one-sided values we desire, use the following rule: (i) if the output ''parameter estimate'' $d_k - d_5$ is zero or positive, divide SAS' output P-value by two to find P_k, or (ii) if the ''parameter estimate'' $d_k - d_5$ is negative, divide SAS' output P-value by two *and subtract this from* 1.0 to find P_k.

It is important to recognize that in most cases, this analysis of the dose-effect will be performed at all non-zero dose levels. Thus, e.g., under our $k = 5$ construction, there are four separate significance tests being performed for the dose effect. Each is a comparison of a specific dose level against the control level; hence this is often called a *multiple comparison with the control*, or a ''MCC.'' (Some authors also call this a *many-to-one* analysis.) Due to the multiple comparisons being performed, however, there will be an inflation in the false positive error rate for testing the dose effect. One possible MCC adjustment to account for error inflation that operates well with binomial GLiMs is a simple Bonferroni correction [13]: this amounts to multiplying the raw

```
*Sample SAS code to fit complementary log-log GLiM;
*Assumes culture*dose interaction is significant;

 data work1;
 infile 'data_file_name_here' ;
    *Ensure that the control level is coded as last;
 input  dose culture y n;

 proc sort; by culture;

 proc genmod;
 by culture;
  class dose;
  model y/n = dose / dist=b link=cll type1;
 run;
```

Fig. 2. Sample SAS Proc Genmod complementary log–log GLiM code for fitting a one-way model of only dose. Model assumes a significant culture × dose interaction and consequently stratifies the analysis over levels of culture.

P-value by the number of individual comparisons being made. In our case, the MCC-adjusted *P*-value at each of the four dose levels is $P_k^* = 4\{1 - \Phi(W_k)\}$. Reject in favor of the one-sided alternative H_{ak} if P_k^* drops below α.

2.4. Statistical analysis under significant culture × dose interaction

When a particular chemical/strain/S9 combination tests positive for culture × dose interaction, the MCC analysis described above must be modified, since we cannot make interpretable inferences about the main dose effect in the presence of a significant interaction. In this case, we are forced to assess the possible dose effects at a simpler level. Specifically, we test for a dose-related increase by assessing the dose effect at each level of culture — under our design this is at each $j = 1,2,3$. In effect, we *stratify* the dose analysis over the levels of culture.

The computations for this stratified analysis are no more complex than those for testing the main effects; sample SAS code for this is given in Fig. 2. The resulting output contains a dose analysis at every level of the culture indicator. In each case, conduct the analysis in the same manner as above, i.e., read the parameter differences $d_k - d_5$ from the Analysis of Parameter Estimates output, calculate the correct one-sided P_k-values, adjust the P_ks for multiplicity via a Bonferroni correction, etc. If any of the

three per-culture tests of dose effect indicates a significant increase over the corresponding control response, we judge the dose effect to be significant.

3. Results: example with cumene hydroperoxide and nitrofurantoin

To illustrate the complementary log–log analysis, we selected two different chemicals from our previous study [5]: cumene hydroperoxide (CASRN 80-15-9) and nitrofurantoin (CASRN 67-20-9). The first represents a case of no culture × dose interaction so that main-effect testing is warranted; the second illustrates a case where a significant culture × dose interaction requires a stratified analysis.

3.1. Example 1: cumene hydroperoxide

Cumene hydroperoxide is a chemical intermediately used to synthesize organic peroxides for the

Table 1
Proportions of positive wells in Ames II™ Strain TA7006 after exposure to cumene hydroperoxide. Case: no S9 activation

Replicate culture	Dose index (doses (in μg/ml))				
	$k = 5$ (0)	$k = 1$ (1)	$k = 2$ (5)	$k = 3$ (10)	$k = 4$ (25)
$j = 1$	3/144	1/144	6/144	21/144	4/144
$j = 2$	4/144	3/144	15/144	17/144	3/144
$j = 3$	1/144	3/144	11/144	13/144	2/144

Table 2
Results from complementary log–log analysis of data from Ames II™ Strain TA7006 after exposure to cumene hydroperoxide. Case: no S9 activation (data from Table 1)

Dose level, k	MCC comparison	Estimated difference	SAS two-sided P-value[a]	Upper one-sided P-value	Bonferroni adjusted P_k^*-value
1	$\delta_1-\delta_5$	-0.1353	0.7938	0.6031	N.S.[b]
2	$\delta_2-\delta_5$	1.4156	0.0003	0.0002	0.0007
3	$\delta_3-\delta_5$	1.9054	5.43×10^{-7}	2.72×10^{-7}	1.09×10^{-6}
4	$\delta_4-\delta_5$	0.1185	0.8073	0.4037	N.S.

[a] SAS typically limits its output P-values to only four digits of accuracy. For values of $P < 0.0001$, we have supplied more accurate values based on direct computations.

[b] N.S. = Not significant.

manufacture of plastic resins and polymerization catalysts. Table 1 contains proportion response data from the original study in Ames II™ Strain TA7006 with no S9 activation. Notice the coding of the control dose as $k = 5$.

Applying the SAS code in Fig. 1 to these data results in the following likelihood ratio test for the culture $\times$ dose interaction: $G_\psi^2 = 8.3326$, with P-value $P = 0.4017$. At $\alpha = 0.05$, this is insignificant, so we continue with this analysis and move to the tests of the dose main-effect. The overall likelihood ratio statistic for the dose effect is $G_\delta^2 = 71.163$, with a P-value of $P < 0.0001$. Although significant, this omnibus P-value provides no guidance regarding which dose levels deviate significantly from the control, and also judges departures below the control level equal in importance to those that exceed the control level. For a more pertinent analysis, the one-sided MCC analysis described above is required.

The SAS output for the main-effect MCC analysis gives the results in Table 2. From the Bonferroni-adjusted P^*-values, we see that a significant increase in mutagenic response over the control is observed at the middle two dose levels ($P < 0.001$ in both cases). This represents definitive evidence of mutagenicity for this chemical in Ames II™ Strain TA7006.

3.2. Example 2: nitrofurantoin

The pharmaceutical product nitrofurantoin is a potent germicide employed to treat urinary tract infections. Proportion response data from the original study in Ames II™ Strain TA7004 with no S9 activation are given in Table 3. Again, notice the coding of the control dose as $k = 5$.

Applying the SAS code in Fig. 1 to these data results in the following likelihood ratio test for culture $\times$ dose interaction: $G_\psi^2 = 47.3442$, with a P-value of $P < 0.0001$. At $\alpha = 0.05$, this is significant, so to analyze the dose effect, we must turn to a culture-stratified analysis. Applying the SAS code in Fig. 2 yields the results given in Table 4. (Notice that the Bonferroni-adjusted P^*-values are the raw one-sided values multiplied now by 12. This is because there are $4 \times 3 = 12$ different MCC comparisons being performed for this data set.) From the Bonferroni-adjusted P^*-values, we observe significant increases in mutagenic response at the two middle doses for all the three cultures, along with significant increases at high dose in the first culture and at the low dose in the last culture. We view this pattern of consistent increases across cultures as indicative of a significant mutagenic effect, rather

Table 3
Proportions of positive wells in Ames II™ Strain TA7004 after exposure to nitrofurantoin. Case: no S9 activation

Replicate culture	Dose index (doses (in μg/ml))				
	$k = 5$ (0)	$k = 1$ (0.1)	$k = 2$ (0.5)	$k = 3$ (1.0)	$k = 4$ (5.0)
$j = 1$	5/144	6/144	44/144	76/144	36/144
$j = 2$	6/144	18/144	50/144	70/144	5/144
$j = 3$	2/144	18/144	53/144	92/144	13/144

Table 4
Results from a complementary log–log analysis, stratified by level of culture, of data from Ames II™ Strain TA7004 after exposure to nitrofurantoin. Case: no S9 activation (data from Table 3)

Dose level, k	MCC comparison	Estimated difference	SAS two-sided P-value[a]	Upper one-sided P-value	Bonferroni adjusted P_k^*-value
Culture: j = 1					
1	$\delta_1-\delta_5$	0.1859	0.7588	0.3794	N.S.[b]
2	$\delta_2-\delta_5$	2.3339	7.71×10^{-7}	3.85×10^{-7}	4.62×10^{-6}
3	$\delta_3-\delta_5$	3.0555	3.90×10^{-11}	1.95×10^{-11}	2.23×10^{-10}
4	$\delta_4-\delta_5$	2.0969	5.63×10^{-6}	1.13×10^{-5}	0.0002
Culture: j = 2					
1	$\delta_1-\delta_5$	1.1434	0.0153	0.0076	0.0918
2	$\delta_2-\delta_5$	2.3047	9.82×10^{-8}	4.91×10^{-8}	5.89×10^{-7}
3	$\delta_3-\delta_5$	2.7500	1.08×10^{-10}	5.40×10^{-11}	6.48×10^{-10}
4	$\delta_4-\delta_5$	−0.1859	0.7588	0.6206	N.S.
Culture: j = 3					
1	$\delta_1-\delta_5$	2.2563	0.0025	0.0013	0.0150
2	$\delta_2-\delta_5$	3.4909	1.27×10^{-6}	6.35×10^{-7}	7.62×10^{-6}
3	$\delta_3-\delta_5$	4.2881	2.05×10^{-9}	1.02×10^{-9}	1.23×10^{-8}
4	$\delta_4-\delta_5$	1.9118	0.0118	0.0059	0.0708

[a] SAS typically limits its output P-values to only four digits of accuracy. For values of $P < 0.0001$, we have supplied more accurate values based on direct computations.

[b] N.S. = Not significant.

than as a series of random jackpot increases. This represents definitive evidence of mutagenicity for this chemical.

4. Discussion

We find the complementary log–log GLiM to be a useful model under which to test interaction and MCC dose effects when presented with data in the form of proportions (i.e., number of positive wells divided by total number of wells) from the AMAX™ protocol. Facilitated by the use of simple SAS programming code, the statistical methods can be easily implemented. Using the SAS output, multiplicity-adjusted P^*-values are straightforward to calculate and can yield proper inferences on the ability of a chemical or environmental agent to induce mutagenesis in these Ames II™ strains. The examples displayed above were chosen to be representative of the larger body of experimental results reported from our previous, larger study of the AMAX™ protocol [5]. Both illustrate the methodology and show qualitative agreement with our corresponding previous results.

It is interesting to note that in both examples, the dose response is non-monotone, i.e., in both Tables 1 and 3, there is a consistent increasing-then-decreasing trend in the response as dose increases. This is not uncommon with Ames test data [14], and we were not surprised by the phenomenon here. The MCC methods we apply to detect the increases above the background response are designed to be unaffected by such downturns. (Although this is at the cost of some sensitivity to detect a monotone-increasing trend over dose, if one did exist. If desired, a trend test that accounts for the downturns may be useful, such as that suggested in Ref. [15] or Ref. [16]. The issue of testing non-monotone trend specifically with the proportion data is problematic, however [17,18], and is open for further statistical research.)

One additional problem for further statistical study concerns the small-sample properties of the Bonferroni-adjusted MCC inferences that form the core of our dose analysis. Previous research [13] has suggested that the Bonferroni correction exhibits acceptable false positive error properties for binomial-based GLiMs under a simple one-way model (say, with

only a single factor such as Dose). The correction is generally conservative in that it protects against false-positive errors too strenuously, but as the sample size increases, this conservative nature tends to lessen somewhat. Whether this performance carries over to the two-way setting with interaction, as studied herein, is unclear. Clearly, more research is required in this area.

References

[1] D.M. Maron, B.N. Ames, Revised methods for the *Salmonella* mutagenicity test, Mutat. Res. 113 (1983) 173–215.

[2] K. Mortelmans, S. Haworth, T. Lawlor, W. Speck, B. Tainer, E. Zeiger, *Salmonella* mutagenicity tests: II. Results from the testing of 270 chemicals, Environ. Mutagen. 8 (Suppl. 7) (1986) 1–119.

[3] E. Zeiger, B. Anderson, S. Haworth, T. Lawlor, K. Mortelmans, W. Speck, *Salmonella* mutagenicity tests: III. Results from the testing of 255 chemicals, Environ. Mutagen. 9 (Suppl. 9) (1987) 1–109.

[4] E. Zeiger, B. Anderson, S. Haworth, T. Lawlor, *Salmonella* mutagenicity tests: IV. Results from the testing of 300 chemicals, Environ. Mutagen. 11 (Suppl. 12) (1988) 1–158.

[5] P. Gee, C.H. Sommers, A.S. Melick, X.M. Gidrol, M.D. Todd, R.B. Burris, M.N. Nelson, R.C. Klemm, E. Zeiger, Comparison of responses of base-specific *Salmonella* tester strains with the traditional strains for identifying mutagens: the results of a validation study, Mutat. Res. 412 (1998) 115–130.

[6] P. Gee, D.M. Maron, B.N. Ames, Detection and classification of mutagens: a set of base-specific *Salmonella* tester strains, Proc. Natl. Acad. Sci. U.S.A. 91 (1994) 11606–11610.

[7] M.H.L. Green, W.J. Muriel, B.A. Bridges, Use of a simplified fluctuation test to detect low levels of mutagens, Mutat. Res. 38 (1976) 33–42.

[8] D.G. Gatehouse, G.F. Delow, The development of a ''microtitre®'' fluctuation test for the development of indirect mutagens, and its use in the evaluation of mixed enzyme induction of the liver, Mutat. Res. 60 (1979) 239–252.

[9] J. Neter, M.H. Kutner, C.J. Nachtsheim, W. Wasserman, Applied Linear Statistical Models, 4th edn., R.D. Irwin, Chicago, IL, 1996.

[10] W.W. Piegorsch, A.J. Bailer, Statistics for Environmental Biology and Toxicology, Chapman & Hall/CRC Press, Boca Raton, FL, 1997.

[11] P. McCullagh, J.A. Nelder, Generalized Linear Models, 2nd edn., Chapman & Hall, London, 1989.

[12] SAS Institute, SAS/Stat® Software: Changes and Enhancements through Release 6.12, SAS Institute, Cary, NC, 1989.

[13] W.W. Piegorsch, One-sided significance tests for generalized linear models under dichotomous response, Biometrics 46 (1990) 309–316.

[14] B.S. Kim, B.H. Margolin, Statistical methods for the Ames *Salmonella* assay: a review, Mutat. Res. 436 (1999) 113–122.

[15] D.G. Simpson, B.H. Margolin, Recursive nonparametric testing for dose–response relationships subject to downturns at high doses, Biometrika 73 (1986) 589–596.

[16] D.H. Lim, D.A. Wolfe, Nonparametric tests for comparing umbrella pattern treatment effects with a control in a randomized block design, Biometrics 53 (1997) 410–418.

[17] N.-Z. Shi, Testing for umbrella order restrictions on multinomial parameters, Sankhya Ser. B 51 (1989) 13–23.

[18] A. Cohen, H.B. Sackrowitz, Inadmissibility of some tests for order-restricted alternatives, Stat. Probability Lett. 24 (1995) 153–156.