

EFFICIENT DESIGNS FOR ANIMAL CARCINOGENICITY EXPERIMENTS

Hongshik Ahn, Wei Zhu and Joonseok Yang*

Department of Applied Mathematics and Statistics
State University of New York at Stony Brook
Stony Brook, NY 11794-3600

Ralph L. Kodell

Division of Biometry and Risk Assessment
National Center for Toxicological Research
Food and Drug Administration
Jefferson, AR 72079

Key Words and Phrases: Monte Carlo; Power; Trend Test; Tumor Incidence Rate.

ABSTRACT

In a typical carcinogenicity study, animals, usually rats or mice, are divided into a control and two to three dose groups of 50 or more by randomization. A chemical is administered at a constant daily dose rate for a major portion of the lifetime of the test animals, for example, two years. In general, such an experiment is expensive and time consuming. In this paper, we propose an efficient design with reduced sample size and/or shortened study duration. An equal number of animals per dose group is considered in this study. A power study of the age-adjusted trend test for the tumor incidence rate for single-sacrifice experiments proposed by Kodell et al. (Drug Information Journal, 1997) is conducted. A Monte Carlo simulation study is performed to compare the performance of the trend test for the standard design and various reduced designs. Based on the Kodell et al. test, the 21-month study duration with sample size 50 per group is recommended as the best reduced design over the traditional 2-year study design with the same sample size.

*Current address: Financial Division, Empire Blue Cross and Blue Shield, New York, NY 10017

1. INTRODUCTION

In general, animal carcinogenicity studies with several dose groups of 50 or more animals conducted for the lifetime of the test animal are expensive and time consuming. In order to reduce the cost of experiments, the effect of shortening study duration on the power of a statistical test is investigated, and the power of the test for reduced sample sizes is examined to find how reasonable the standard size of 50 animals per sex/group is. A Monte Carlo simulation study is conducted to compare the standard design (two-year study duration with 50 animals per dose group) and the reduced designs.

Both interim-sacrifice data and cause-of-death data add expense to rodent bioassays, so that neither type of information is always available in the rodent study. In the absence of both interim sacrifices and cause-of-death information, the tumor incidence rate is not identifiable from bioassay data, unless simplifying assumptions are made. For data with no interim sacrifices and no cause-of-death information, Dinse (1991) and Lindsey and Ryan (1994) proposed parametric statistical tests of the tumor incidence rate for dose-related trend. Dinse's test is based on the assumption of a constant difference between the death rates of animals with and without tumors, while Lindsey and Ryan's test assumes a constant ratio for those death rates. Recently Kodell et al. (1997) proposed a nonparametric age-adjusted trend test for a study with a single terminal sacrifice and no cause-of-death information. They assume the constant proportionality of tumor prevalence for live and dead animals. In this paper, the test for the tumor incidence rate by Kodell et al. is used to conduct the power study of the reduced designs. To achieve the goal of reducing the cost of experiments, the following Monte Carlo simulation studies are conducted: First, power of the test for shorter study duration is evaluated. Second, power of the test is evaluated for smaller sample sizes.

2. AGE-ADJUSTED TREND TEST FOR A SINGLE SACRIFICE EXPERIMENT

Kodell et al. (1997) proposed a nonparametric age-adjusted test for assessing dose-related trend with respect to the tumor incidence rate for a study having only a single terminal sacrifice.

Consider an experiment with g treatment groups, where group i ($i = 1, \dots, g$) is exposed to a dose level d_i of a test substance, with $d_i = 0$ representing a control group. Let n_i tumor-free animals be assigned to the i th treatment group at time $t = 0$. Divide the time scale into

m intervals such that the j th interval is $(t_{j-1}, t_j]$, where $0 = t_0 < t_1 < \dots < t_m$. A sacrifice is assumed to occur only at the end of the last time interval. For group i and interval j , let $N_i^1(j)$ and $N_i^2(j)$ denote the number of natural deaths with and without tumors, respectively, and let $N_i^3(m)$ and $N_i^4(m)$ denote, respectively, the number of sacrifices with and without tumors at the end of the study. Let $A_i(j)$ denote the number of animals alive at the beginning of the j th interval and $N_i(j) = N_i^1(j) + N_i^2(j)$ denotes the total number of natural deaths in the j th interval.

Kodell et al. defined two random variables, T_1 and X_D , denoting, respectively, time to onset of tumor and time to natural death. It is assumed that all deaths and tumors occur only at the ends of intervals, where tumors precede natural deaths and deaths precede sacrifices (Malani and Van Ryzin, 1988). Thus T_1 and X_D are discrete random variables, taking on only the values $j = 1, \dots, m$. The discrete tumor incidence rate for the i th group on the j th interval is defined by

$$\begin{aligned}\lambda_i^T(j) &= Pr(T_1 = j | X_D \geq j, T_1 \geq j) \\ &= 1 - \{[1 - p_i^A(j)][1 - \lambda_i^D(j)] + [1 - p_i^D(j)]\lambda_i^D(j)\} / [1 - p_i^A(j-1)],\end{aligned}$$

where $p_i^A(j) = Pr(T_1 \leq j | X_D > j)$ is the tumor prevalence function for live animals, $p_i^D(j) = Pr(T_1 \leq j | X_D = j)$ is the tumor prevalence function for dead animals, and $\lambda_i^D(j) = Pr(X_D = j | X_D \geq j)$ is the discrete hazard rate for X_D .

The log likelihood function for the i th dose group is

$$\begin{aligned}l_i &= \sum_{j=1}^m \{N_i^1(j) \log p_i^D(j) + N_i^2(j) \log [1 - p_i^D(j)] \\ &\quad + N_i(j) \log \lambda_i^D(j) + [A_i(j) - N_i(j)] \log [1 - \lambda_i^D(j)]\} \\ &\quad + N_i^3(m) \log p_i^A(m) + N_i^4(m) \log [1 - p_i^A(m)] + c,\end{aligned}$$

where c is a constant. Maximization of l_i with respect to $p_i^A(j)$, $p_i^D(j)$ and $\lambda_i^D(j)$ is carried out subject to $\lambda_i^T(j) \geq 0$, for all j .

Because there are no interval sacrifices before the terminal sacrifice, a further assumption is needed to make $p_i^A(j)$ and $\lambda_i^T(j)$ identifiable for all $j < m$. Kodell et al. assumed a constant proportionality of tumor prevalences for live and dead animals by adding

$$p_i^A(j) = p_i^D(j)p_i^A(m)/p_i^D(m), \quad j = 1, \dots, m-1.$$

They performed constrained maximization of l_i using the direct-search algorithm of Box (1965).

In order to test the null hypothesis

$$H_0 : \lambda_1^T(j) = \lambda_2^T(j) = \cdots = \lambda_g^T(j), \quad j = 1, 2, \dots, m,$$

Kodell et al. formulated a log-rank-type statistic with asymptotic standard normal null distribution as

$$Z = \frac{\mathbf{d}'(\mathbf{O} - \mathbf{E})}{\sqrt{\mathbf{d}'\mathbf{V}\mathbf{d}}}, \quad (1)$$

where \mathbf{d} is a vector of length g containing an appropriate dose metric, \mathbf{O} is a vector of length g containing observed frequencies, O_i , of tumors for each dose group, \mathbf{E} is a g -vector of expected frequencies, E_i , and \mathbf{V} is $g \times g$ estimate of the variance of \mathbf{O} , having elements V_{ik} . These quantities are further defined by $O_i = \sum_{j=1}^m y_i(j)$, $E_i = \sum_{j=1}^m R_i(j)[y_{\cdot}(j)/R_{\cdot}(j)]$, $R_i(j) = A_i(j)[1 - \hat{p}_i^A(j-1)]$, $y_i(j) = R_i(j)\hat{\lambda}_i^T(j)$, $R_{\cdot}(j) = \sum_{i=1}^g R_i(j)$, and $y_{\cdot}(j) = \sum_{i=1}^g y_i(j)$. The elements of \mathbf{V} are given in Kodell et al. (1997).

Because it is assumed that there are no scheduled interim sacrifices in the bioassay to be analyzed, there are no pre-defined time intervals for constructing the test statistic. For a two-year rodent bioassay, four intervals are commonly used to compare the age-specific tumor incidence rate. The most natural choice for a four-interval comparison is the intervals of 0-52, 53-78, 79-92, and 93-104 weeks of the National Toxicology Program (NTP; Bailer and Portier, 1988). If, in any dose group, there is no animal dying in a time interval, then that interval is collapsed with the next time interval (if the interval is the last one, it is collapsed with the previous one) for all groups in order to calculate the test statistic.

3. SIMULATION STUDY

A Monte Carlo simulation study was conducted to evaluate the power of the dose-related trend test for various designs. Bioassay designs with four dose groups (0, 1, 2, 4) were considered, with an equal number of animals for each dose group. The simulation study was conducted as follows.

1. Fix the study duration at 2 years, reduce the sample size, and study the change in power of the test. The power of the trend test for the reduced sample size is compared with that for the standard sample size of 50 per dose group.

2. Fix the sample size at 50 per dose group, reduce the study duration, and observe the change in power of the test. The power of the test for reduced study duration is compared with that of the standard two year experiment.

In all cases, the trend test was calculated using dose scalings of 0, 1, 2, 3 and 0, 1, 2, 4. We will now proceed with the discussion of how to find the efficient reduced designs with an acceptable power.

3.1 Standard Design

The design used in Kodell et al. (1997) was simulated in this paper. The design has a terminal sacrifice at the end of the study at 104 weeks, which is the normal term of a chronic study in rodents. To implement the test, the NTP time intervals described in Section 2 were adopted.

It was assumed that three independent random variables determined the observed outcome for each animal completely. These were the time to onset of tumor, T_1 , the time after onset until death from the tumor, T_2 , and the time to death from a competing risk, X_C . Note that $T_1 + T_2 = T_D$, where T_D represents the overall time to death from the tumor of interest. Thus the tumor of interest was present in an animal at death if $T_1 \leq \min\{X_C, X_S\}$, where X_S denotes an animal's scheduled sacrifice time. An animal died from the tumor of interest if $T_D \leq \min\{X_C, X_S\}$, whereas from a competing risk (including sacrifice) otherwise.

Distributions of time to onset and death were of the form used by Portier et al. (1986). The distribution of time to onset of tumor, T_1 , was modeled as

$$P(t) = 1 - \exp(-\theta\delta_1 t^{\delta_2}),$$

where $\theta = e^d \geq 1$, $\delta_1 \geq 0$ and $\delta_2 \geq 0$. The value of δ_2 was set to 3 (Portier et al., 1986). For size evaluations (i.e., no dose effect), θ was set to $e^0 = 1$ and $\delta_1 = -\ln[1 - P(104)]/(104)^3$ was chosen such that the probability of tumor onset by 104 weeks was either .01 (rare tumor), .05 or .10 (common tumor). For power comparisons, θ was chosen such that the probability of tumor onset in the highest dose group ($d = d_H$) by 104 weeks would be x times that in controls ($d = 0$), where $x = 15, 5, 4$ for the background tumor rates .01, .05 and .10, respectively. Thus $d_H = \ln \theta$ for $\theta = -\ln[1 - xP(104)]/[\delta_1(104)^3]$. The low and intermediate doses were chosen as $d_H/4$ and $d_H/2$, respectively.

The survival function for time to death from competing risks, X_C , was taken to be

$$Q(t) = \exp[-\phi(\gamma_1 t + \gamma_2 t^{\gamma_3})],$$

where $\phi \geq 1$, $\gamma_1 \geq 0$, $\gamma_2 \geq 0$ and $\gamma_3 \geq 0$ (Portier et al., 1986). With $\phi = 1$, $\gamma_1 = 10^{-4}$ and $\gamma_2 = 10^{-16}$, γ_3 was chosen to be 7.4477031, so that the probability of survival with respect to competing risk was .9. For no treatment lethality, the probability of competing risk survival in each group was set equal to .5. For high treatment lethality, the probabilities of competing risk survival were .5 (control), .4 (low dose group), .3 (intermediate dose group) and .2 (high dose group). The parameter value of ϕ was calculated as $\phi = \ln(x)/\ln(.9)$, where x is the probability of competing risk survival.

The survival function for time to death from tumor, T_2 , had the same form as that for death from competing risks, and the values of γ_1 , γ_2 and γ_3 were the same. The parameter ϕ was selected to reflect both low tumor lethality (less than 1% of observed tumors were cause of death: $\phi = 1$), intermediate tumor lethality (approximately 35% of observed tumors were the cause of death: $\phi = 2,800$ for $P(104) = .01$, $\phi = 200$ for $P(104) = .05$, $\phi = 175$ for $P(104) = .10$), and high tumor lethality (approximately 38% of observed tumors were the cause of death: $\phi = 10,000$ for $P(104) = .01$ and approximately 90% of observed tumors were the cause of death: $\phi = 10,000$ for $P(104) = .05$ and $\phi = 2,700$ for $P(104) = .10$).

The Type I error rate was evaluated with respect to a nominal significance level of 5%. One thousand simulated data sets were generated for each combination of tumor onset probability at 104 weeks (3), tumor lethality rate (3), and competing risks survival rate (2). The power calculations also used a nominal 5% significance level, for 3 tumor onset probabilities in controls at 104 weeks, 1 tumor onset probability in each dose group per control onset probability, 2 different sets of competing risks survival rates, and 3 tumor lethality rates.

3.2 Reduced Sample Size

In this simulation experiment, the power of the trend test was evaluated using a smaller sample size in each dose group. Sample size was reduced by decrements of 5 animals in each dose group. The study duration was fixed at two years (104 weeks), and sample size decreased to as low as 30 animals per dose group. Tables 1 and 2 show the simulated power. With 1% tumor onset probability, tumor lethality probability cannot reach 90%. Thus we did not include 1% tumor rate with 90% tumor lethality rate in this simulation. Reduction of

Table 1: Simulated power (%) corresponding to nominal 5% significance level for the trend test of Kodell et al. based on a dose scaling of 0, 1, 2, 3, applied to data generated from the bioassay design described in Section 3.

Tumor Onset Prob. ^a	Competing Risks Survival ^b	Tumor Lethal. Prob. ^c	Sample Size at Each Dose Group				
			50	45	40	35	30
.01	.50	$\leq .01$	89.3	87.2	82.0	76.9	72.7
.01	.50	$\geq .35$	89.7	87.4	82.0	77.2	72.9
.05	.50	$\leq .01$	87.3	82.8	79.5	75.4	68.6
.05	.50	$\geq .35$	89.0	84.7	81.1	75.7	69.9
.05	.50	$\geq .9$	87.3	84.9	80.8	75.3	69.7
.10	.50	$\leq .01$	96.0	94.4	89.6	87.8	83.1
.10	.50	$\geq .35$	96.0	95.1	91.3	89.8	84.6
.10	.50	$\geq .9$	95.9	95.5	90.9	90.2	86.0
.01	.5, .4, .3, .2	$\leq .01$	83.9	83.2	77.3	71.2	65.5
.01	.5, .4, .3, .2	$\geq .35$	82.9	81.3	75.1	69.2	64.2
.05	.5, .4, .3, .2	$\leq .01$	85.4	81.3	75.2	71.5	65.9
.05	.5, .4, .3, .2	$\geq .35$	83.0	78.2	72.4	68.7	63.4
.05	.5, .4, .3, .2	$\geq .9$	81.6	76.9	70.7	65.5	59.8
.10	.5, .4, .3, .2	$\leq .01$	93.5	91.6	86.9	84.7	80.0
.10	.5, .4, .3, .2	$\geq .35$	91.0	88.1	84.7	80.9	76.2
.10	.5, .4, .3, .2	$\geq .9$	90.5	87.6	82.7	79.3	74.9

^a Cumulative tumor onset probability in control, at 104 weeks in absence of competing risks.

^b Survival probability with respect to all causes except the tumor of interest.

^c Proportion of observed tumors that actually result in death. Same probability in all dose groups.

the power from sample size 50 to 45 was somewhat small. Although the power was reduced greatly by reducing the sample size from 45 to 40, it was still reasonably high (over 70% for dose scaling 0,1,2,3 and over 73% for dose scaling 0,1,2,4). The power for sample sizes less than 40 was unacceptably low. The power for the higher competing risk survival rate (.5) was higher than that for the lower competing risk survival rate (.5, .4, .3, .2).

3.3 Reduced Study Duration

In this simulation experiment, the power of the test for reduced study duration was evaluated. To compare with the standard two-year bioassay design, 91 week (21 month) and 78 week (18 month) bioassay designs were considered. The sample size was fixed at 50 for each dose group. Time intervals proportional to the NTP time intervals were used to implement the test. If the study duration was N weeks, the intervals were determined as $t_0 = 0$, $t_1 = 52 \times N/104$, $t_2 = 78 \times N/104$, $t_3 = 92 \times N/104$ and $t_4 = N$. Note that the death rate was very low for the animals with a low tumor lethality along with a low tumor

Table 2: Simulated power (%) corresponding to nominal 5% significance level for the trend test of Kodell et al. based on a dose scaling of 0, 1, 2, 4, applied to data generated from the bioassay design described in Section 3.

Tumor Onset Prob.	Competing Risks Survival	Tumor Lethal. Prob.	Sample Size at Each Dose Group				
			50	45	40	35	30
.01	.50	$\leq .01$	90.6	88.7	84.7	78.8	74.2
.01	.50	$\geq .35$	90.9	88.8	84.2	78.4	74.1
.05	.50	$\leq .01$	90.1	85.6	81.4	76.6	72.2
.05	.50	$\geq .35$	90.0	86.2	82.3	76.8	72.8
.05	.50	$\geq .9$	90.1	86.5	82.4	78.0	73.3
.10	.50	$\leq .01$	96.2	95.2	91.3	89.7	85.6
.10	.50	$\geq .35$	96.2	95.9	92.7	91.2	87.2
.10	.50	$\geq .9$	96.9	96.2	92.5	92.1	87.6
.01	.5, .4, .3, .2	$\leq .01$	85.6	86.2	80.6	75.8	70.7
.01	.5, .4, .3, .2	$\geq .35$	84.5	84.1	78.3	74.0	69.7
.05	.5, .4, .3, .2	$\leq .01$	87.7	83.2	77.9	73.0	69.6
.05	.5, .4, .3, .2	$\geq .35$	85.1	80.9	75.7	70.6	67.3
.05	.5, .4, .3, .2	$\geq .9$	83.6	78.3	73.3	68.2	63.5
.10	.5, .4, .3, .2	$\leq .01$	93.6	92.4	88.7	86.0	81.9
.10	.5, .4, .3, .2	$\geq .35$	90.9	89.8	86.2	82.9	79.1
.10	.5, .4, .3, .2	$\geq .9$	91.7	89.2	85.1	81.6	77.6

rate and a high competing risk survival rate, especially for the control group. In this case, it is possible that the first few intervals will have no death at all. Kodell et al. (1997) showed in their simulation that this phenomenon is not unusual. As mentioned in Section 2, they combined such an interval with an adjacent interval across different dose groups to activate the trend test.

Table 3 shows the simulation results of the standard and shortened study durations. Similar to cases with sample sizes of 40 and less in Section 3.2, the power of the trend test was substantially reduced by shortening the study duration. The reduction of the power was relatively small for the cases of highly lethal tumors and low competing risk survival rate. If the tumor is highly lethal and competing risk survival rate is low, then many animals die early and not many animals are left at the last interval. Thus, the last interval of the full two-year study duration does not provide much information. Therefore, a shortened study duration is almost as efficient as the full study duration in this case.

The power for 91 week the study duration was considerably lower than that for the full duration. However, the power was still reasonably high (over 74% for dose scaling 0,1,2,3 and over 76% for dose scaling 0,1,2,4). For the 78 week study duration, the power went down to 61% for some cases of 5% tumor rate for dose scaling 0,1,2,3, which is unacceptably

Table 3: Simulated power (%) corresponding to nominal 5% significance level for trend test of Kodell et al. (1997) applied to data generated from the bioassay design with 50 animals at each of doses 0, 1, 2, 4.

Tumor Onset Prob.	Competing Risks Survival	Tumor Lethal. Prob.	Duration of the Study (Weeks)					
			Dose Scaling 0,1,2,3			Dose Scaling 0,1,2,4		
			104	91	78	104	91	78
.01	.50	$\leq .01$	89.3	81.9	67.8	90.6	83.2	69.6
.01	.50	$\approx .35$	89.7	81.7	68.1	90.9	83.0	69.6
.05	.50	$\leq .01$	87.3	79.1	65.6	90.1	80.8	68.7
.05	.50	$\approx .35$	89.0	79.4	65.5	90.0	81.0	69.0
.05	.50	$\approx .9$	87.3	79.4	65.8	90.1	81.7	68.9
.10	.50	$\leq .01$	96.0	90.9	80.9	96.2	91.9	82.2
.10	.50	$\approx .35$	96.0	91.4	81.1	96.2	92.9	82.5
.10	.50	$\approx .9$	95.9	92.2	81.1	96.9	92.9	83.0
.01	.5, .4, .3, .2	$\leq .01$	83.9	76.7	63.2	85.6	80.2	66.0
.01	.5, .4, .3, .2	$\approx .35$	82.9	76.3	63.2	84.5	79.2	65.6
.05	.5, .4, .3, .2	$\leq .01$	85.4	76.8	62.1	87.7	79.0	65.8
.05	.5, .4, .3, .2	$\approx .35$	83.0	75.0	61.5	85.1	77.6	65.4
.05	.5, .4, .3, .2	$\approx .9$	81.6	74.4	61.4	83.6	77.1	66.0
.10	.5, .4, .3, .2	$\leq .01$	93.5	88.9	77.6	93.6	90.8	79.7
.10	.5, .4, .3, .2	$\approx .35$	91.0	88.2	77.9	90.9	90.2	79.5
.10	.5, .4, .3, .2	$\approx .9$	90.5	87.4	77.9	91.7	89.3	80.0

low.

3.4 Conclusions from the Simulation

Tables 4 and 5 show the probability of a Type I error for the standard design, for the design of two-year study duration with sample size 45, and for the design of 91-week study duration with sample size 50. Since the primary emphasis in carcinogenicity studies is testing for increasing tumor incidence rates with increasing doses, the upper tail probabilities are of most interest. However, the lower tail probabilities in the tables are important for assessing the degree of symmetry of the null distributions of test statistics. For the cases of tumor rate 1%, a tumor does not present for most of the animals. Thus, the value of Z in equation (1) is very low and consequently, the size is much lower than the nominal level. Therefore, we will focus on the size for tumor rates 5% and 10%. Both reduced designs maintained the nominal size reasonably well except for the cases of 1% tumor rate. However, the 2-year study duration with sample size 45 in each group was slightly anticonservative. Sizes were over 6% in several cases. From these and the results from the power study, we recommend the 91-week study duration with sample size 50 in each group. The design with 78-week

Table 4: Simulated size (%) corresponding to nominal 5% significance level for trend test of Kodell et al. based on a dose scaling of 0, 1, 2, 3, applied to data generated from the bioassay design described in Section 3.

Tumor Onset Prob. ^a	Competing Risks Survival	Tumor Lethal. Prob.	50 Animals Per Dose Group, 104 Weeks		50 Animals Per Dose Group, 91 Weeks		45 Animals Per Dose Group, 104 Weeks	
			L ^b	U	L	U	L	U
.01	.50	$\leq .01$	3.2	2.4	1.9	1.5	2.4	2.2
.01	.50	$\approx .35$	3.2	2.4	1.9	1.5	2.4	2.2
.05	.50	$\leq .01$	4.6	5.1	5.3	4.7	4.2	5.0
.05	.50	$\approx .35$	4.8	5.4	5.1	4.8	4.4	5.3
.05	.50	$\approx .9$	4.7	5.2	5.3	4.7	4.4	5.4
.10	.50	$\leq .01$	4.3	4.2	3.9	4.4	4.3	5.7
.10	.50	$\approx .35$	4.8	4.4	3.8	4.8	4.1	6.1
.10	.50	$\approx .9$	4.5	4.7	3.8	5.1	4.4	5.8
.01	.5, .4, .3, .2	$\leq .01$	2.7	2.5	2.1	1.3	2.2	2.0
.01	.5, .4, .3, .2	$\approx .35$	3.3	2.1	2.1	1.3	2.4	1.6
.05	.5, .4, .3, .2	$\leq .01$	4.7	5.6	5.4	4.3	4.3	6.2
.05	.5, .4, .3, .2	$\approx .35$	5.6	4.7	5.3	4.1	4.7	6.1
.05	.5, .4, .3, .2	$\approx .9$	6.0	4.3	5.4	3.8	5.4	4.6
.10	.5, .4, .3, .2	$\leq .01$	3.9	5.4	4.5	4.5	4.0	6.6
.10	.5, .4, .3, .2	$\approx .35$	4.8	4.3	5.3	4.5	5.4	5.7
.10	.5, .4, .3, .2	$\approx .9$	5.7	3.3	5.5	4.0	5.4	4.5

^a Cumulative tumor onset probability at 104 weeks in absence of competing risks. Same probability for all four dose groups.

^bL signifies lower probability for a one-sided level 5% test, U signifies upper tail probability for a one-sided level 5% test.

study duration with sample size 50 also controlled the probability of a Type I error well (not reported in this paper), but we do not recommend this design because it does not have desirable power. The 2-year study duration with sample size 45 in each dose group may be considered as the second best reduced design. In fact, its power is higher than the 91-week study with 50 animals per group. However, its failure to maintain nominal size, especially for dose scaling 0,1,2,4, more than offsets its higher power. The reduced design with 40 animals per group and 104-week duration is not recommended, even though its power is comparable to that of the 91-week design with 50 animals, because of its tendency to be anticonservative (results not shown). From the results of the power study, it is clear that much less power will be obtained from the combination of 91-week study duration and sample size 45 in each group. As mentioned in Kodell et al. (1997), dose scaling of 0, 1, 2, 3 performed better than the 0, 1, 2, 4 scaling in terms of maintaining the balance of the lower and upper tail probabilities of the test statistic.

Table 5: Simulated size (%) corresponding to nominal 5% significance level for trend test of Kodell et al. based on a dose scaling of 0, 1, 2, 4, applied to data generated from the bioassay design described in Section 3.

Tumor Onset Prob.	Competing Risks Survival	Tumor Lethal. Prob.	50 Animals Per Dose Group, 104 Weeks		50 Animals Per Dose Group, 91 Weeks		45 Animals Per Dose Group, 104 Weeks	
			L	U	L	U	L	U
.01	.50	\leq .01	2.1	2.8	1.7	1.5	2.0	2.4
.01	.50	\approx .35	2.7	2.7	1.9	1.5	2.1	2.2
.05	.50	\leq .01	4.4	4.9	4.8	5.8	4.3	6.0
.05	.50	\approx .35	4.5	5.0	5.0	6.0	4.2	5.9
.05	.50	\approx .9	4.6	5.1	4.9	5.7	4.1	6.1
.10	.50	\leq .01	4.4	4.2	3.2	5.2	3.6	6.0
.10	.50	\approx .35	4.6	4.7	3.3	5.4	3.8	6.6
.10	.50	\approx .9	4.1	4.7	3.2	5.3	4.0	6.2
.01	.5, .4, .3, .2	\leq .01	0.8	6.0	0.7	2.5	0.8	5.3
.01	.5, .4, .3, .2	\approx .35	1.4	4.2	0.8	2.3	1.4	3.3
.05	.5, .4, .3, .2	\leq .01	3.4	6.2	3.1	5.6	3.0	6.8
.05	.5, .4, .3, .2	\approx .35	3.9	5.6	3.2	5.6	4.0	6.4
.05	.5, .4, .3, .2	\approx .9	4.4	4.8	3.9	5.4	4.4	5.6
.10	.5, .4, .3, .2	\leq .01	4.0	5.6	3.9	5.1	3.7	6.3
.10	.5, .4, .3, .2	\approx .35	5.3	4.1	4.5	4.9	4.7	5.9
.10	.5, .4, .3, .2	\approx .9	5.5	3.7	4.3	4.7	5.2	4.6

4. DISCUSSION

In order to make carcinogenicity studies more cost effective and efficient, we have examined experimental designs with reduced sample size and shortened study duration. Since most animal studies are not designed with numerous sacrifice times, we have focused on data from experiments with a single terminal sacrifice. In this paper, we examined power and the probability of a Type I error of the dose-related trend test by Kodell et al. (1997) for the reduced designs. The power dropped substantially by reducing the sample size and by shortening the duration of the study. However, some reduced designs had sufficient power. We recommend the 91-week study duration with sample size 50 per group as the best reduced design. This design controlled the probability of a Type I error as well as did the standard design.

Kodell et al. (1997) chose three spontaneous tumor rates, 5%, 15%, and 30%. However, the background tumor rates used in that paper are not representative of true tumor rates of real studies. Kodell et al. (1997) reported that the extreme cases involving high tumor onset probabilities (30%) with high tumor lethality, along with substantial differential mortality from competing risks gave lower probability of a Type I error for the standard design. The 91-week study duration with sample size 50 per group controlled the probability of a Type I error better than the standard design in this case (not reported in this paper), although the designs with reduced sample size failed to achieve this. In most of the carcinogenicity studies for new drugs, the background tumor rates are below 5%. In this paper, we chose the tumor rates to represent tumor rates of real studies.

In this paper we focused on equal numbers of animals for each dose group. The power of the trend test given the reduced designs was compared with that given the standard design. However, the traditional equal allocation scheme may not be an optimal design in some situations, as Begg and Kalish (1984), Chaloner and Larntz (1989), Eastwood (1996) and Zhu and Wong (1997) pointed out. For future research, we will seek some potential unbalanced designs which are likely to increase the power of the test. In this paper, the time intervals considered for implementing the test were proportional to the NTP time intervals. As discussed in Section 3, the intervals proportional to the NTP intervals might not include any death in the first few intervals. To solve this problem, we can implement an alternative intervalization scheme. One way is to keep the lengths of the first one or two intervals, and reduce the lengths of the remaining intervals. We can increase the number of intervals and examine the change in the power. We can also study the optimal number of intervals. It would be worthwhile to derive the optimal duration and sample size for the

trend test analytically. As a future study, we can also examine the change in power for the other trend tests designed for a single terminal sacrifice (Dinse, 1991; Lindsey and Ryan, 1994).

The proposed design with shortened study duration will make carcinogenicity studies more cost effective and efficient. In addition, a more rapid experiment will expedite the marketing of new products, such as drugs. Therefore, using an efficient design with a shortened study duration will result in economic and social benefit.

REFERENCES

- Bailer, A. J. and Portier, C. J. (1988). "Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples," *Biometrics*, **44**, 417-431.
- Begg, C. B. and Kalish L. A. (1984). "Treatment allocation for nonlinear models in clinical trials: The logistic model," *Biometrics*, **40**, 409-420.
- Box, M. J. (1965). "A new method of constrained optimization and a comparison with other methods," *The Computer Journal*, **8**, 42-52.
- Chaloner, K. and Larntz, K. (1989). "Optimal Bayesian design applied to logistic regression experiments," *Journal of Statistical Planning and Inference*, **21**, 191-208.
- Dinse, G. E. (1991). "Constant risk differences in the analysis of animal tumorigenicity data," *Biometrics*, **47**, 681-700.
- Eastwood, B. (1997). "Optimal sample size allocation in clinical trials designed to investigate relative risks," *Statistics in Medicine*. To appear.
- Kodell, R. L., Ahn, H., Pearce, B. A. and Turturro, A. (1997). "An age-adjusted trend test for the tumor incidence rate for single-sacrifice experiments," *Drug Information Journal*, **31**, 471-487.
- Lindsey, J. C. and Ryan, L. M. (1994). "A comparison of continuous- and discrete-time three-state models for rodent tumorigenicity experiments," *Environmental Health Perspectives*, **102** (Suppl. 1), 9-17.
- Malani, H. M. and Van Ryzin, J. (1988). "Comparison of two treatments in animal carcinogenicity experiments," *Journal of the American Statistical Association*, **83**, 1171-1177.
- Portier, C., Hedges, J. and Hoel, D. G. (1986). "Age-specific models of mortality and tumor onset for historical control animals in the National Toxicology Program's carcinogenicity experiments," *Cancer Research*, **46**, 4372-4378.
- Zhu, W. and Wong, W. K. (1997). "Optimum dosage allocation in multiple-objective quantal dose response experiments," Submitted.