

Web-based Simulator for Sample Size and Power Estimation in Animal Carcinogenicity Studies

Hojin Moon¹, J. Jack Lee¹, Hongshik Ahn² and Rumiana G. Nikolova¹

¹Department of Biostatistics
The University of Texas M.D. Anderson Cancer Center
1515 Holcombe Boulevard - 447, Houston, TX 77030-4009
e-mail: hojin@mdanderson.org

²Department of Applied Mathematics and Statistics
State University of New York, Stony Brook, NY 11794-3600

Abstract

A Web-based statistical tool for sample size and power estimation in animal carcinogenicity studies is presented in this paper. It can be used to provide a design with sufficient power for detecting a dose-related trend in the occurrence of a tumor of interest when competing risks are present. The tumors of interest typically are occult tumors for which the time to tumor onset is not directly observable. It is applicable to rodent tumorigenicity assays that have either a single terminal sacrifice or multiple (interval) sacrifices. The design is achieved by varying sample size per group, number of sacrifices, number of sacrificed animals at each interval, if any, and scheduled time points for sacrifice. Monte Carlo simulation is carried out in this tool to simulate experiments of rodent bioassays because no closed-form solution is available. It takes design parameters for sample size and power estimation as inputs through the World Wide Web. The core program is written in C and executed in the background. It communicates with the Web front end via a Component Object Model interface passing an Extensible Markup Language string. The proposed statistical tool is illustrated with an animal study in lung cancer prevention research.

Key Words: Competing risks; Experimental design; Monte Carlo simulation; Multiple/Single Sacrifice.

1 Introduction

Due to easy accessibility, the World Wide Web (WWW) provides an excellent environment for deploying statistical tools. One major advantage to the users is that statistical design and analysis can be performed at any place where an Internet connection is available. It does not require that a user acquire, download, or purchase a statistical software to run a program. The technology for the WWW interface has been developed and applied to various statistical applications (to name a few: Banfield, 1999; Firth, 2000; West and Ogden, 1997).

A Web-based sample size and power estimator using Monte Carlo simulation is proposed in this paper. This statistical tool is designed for animal carcinogenicity/tumorigenicity studies on occult tumors, and for rodent bioassays with a single terminal sacrifice or multiple (interval) sacrifices. The estimator takes inputs from the WWW through an Extensible Markup Language (XML) interface and simulates the specified design in the background. The design input parameters include (a) number of dose groups, (b) type of experiments with the number of interval sacrifices, (c) dose metric in each group, (d) sample size per group, (e) time points for sacrifice, (f) number of sacrificed animals at each interval, (g) tumor onset probability in the control group, (h) shape of tumor onset distribution, (i) hazard ratio between each dose group and the control group, (j) competing risks survival rate in each group, (k) parameter to determine lethality rate in the control group, (l) significance level of a one- or two-sided test, and (m) number of simulation runs. This tool sends results back to the user via e-mail. The results include summary statistics, such as the simulated average tumor onset probability in each group, the simulated average competing risks survival rate per group, the simulated average tumor lethality rate in each group, along with the design considerations, and the estimated power. In this study, the core simulation program is written in C.

Animal carcinogenicity bioassays are routinely used to evaluate the carcinogenic potential of chemical substances to which humans are exposed. In a typical animal carcinogenicity study on occult tumors, each animal is assumed to begin with a tumor free state. Mice or rats are species commonly used. They are randomized into a control group (typically, animals that are exposed to a control agent or observed without any exposure) or into 2 to 3 test groups that receive specified levels of exposure until they either die or are sacrificed. In an experiment with multiple sacrifices,

sacrificed animals are preassigned to a specific dose level and sacrifice time at the beginning of the experiment. In a single terminal sacrifice, all surviving animals are sacrificed and subjected to necropsy at the end of the experiment, which is typically a period of 104 weeks (2 years). During the study, age at death and the information on the presence or absence of the tumor of interest are collected for each animal. The primary goal of the experiment is to assess a dose-related trend of test agent exposures in the incidence of the tumor of interest. The tumor of interest can be any occult tumor for which the time to tumor onset is not directly observable. Our software can seek a reduced design (78 - 104 weeks) with an acceptable power. The proposed statistical tool can also be used to seek an optimal design by choosing the design with the maximum power of the trend test for a given total sample size. This tool will help researchers conduct more efficient and cost effective experiments.

The logrank test of Mantel and Haenszel (1959) may be used for comparing hazards of death from rapidly lethal tumors. To compare the prevalence of nonlethal tumors, the prevalence test proposed by Hoel and Walburg (1972) may be used for incidental tumors. However, the data obtained from a carcinogenicity experiment generally contain a combination of fatal and incidental tumors. Peto et al. (1980) suggested combining the fatal and incidental tests for comparing tumor onset distributions. The procedure proposed by Peto et al., has been called the cause-of-death test or the Peto test.

The development of the presented statistical tool is motivated by a series of animal studies at M. D. Anderson Cancer Center that explores the mechanisms underlying the chemopreventive effects of test agents. The first step is to establish the carcinogenic potential of the tobacco carcinogen NNK, a byproduct of tobacco smoke, in retinoic acid receptor- β (RAR- β) transgenic mice. The Peto test, recommended by the International Agency for Research on Cancer (IARC), is used to compare the tumor incidence rate among groups in the presence of potential confounders. It is a widely used statistical test to determine a dose-related trend for a test agent in the occurrence of occult tumors.

The purpose of this paper is to present a statistical tool for Web-based sample size and power estimation using the Peto test statistic to provide sufficient power for detecting a dose-related trend in the occurrence of a tumor of interest. This package simulates rodent bioassay experiments with either a single sacrifice or multiple sacrifices. A comprehensive list of design parameters can be

specified by the users through the WWW. The underlying models are described in Section 2. A detailed description of the use of the proposed estimator is demonstrated in Section 3. A design of a carcinogenesis experiment for lung cancer prevention research is illustrated in Section 4 as an example. Concluding remarks and suggested future study directions are described in the last section.

2 Model and Test for a Sample Size and Power Estimator

Consider a carcinogenicity/tumorigenicity experiment with the control group and $G-1$ dose groups of animals. Suppose that N_i animals are randomly assigned to the i -th group, and they are followed over time for the development of irreversible and occult tumors. The animals in the i -th group receive a dose level of ℓ_i of a test agent. We assume that all animals come from the same population and have no tumor on day zero of the experiment. The time scale is divided into J intervals such that the j -th interval is given by $I_j = (t_{j-1}, t_j]$, $j = 1, \dots, J$. Note that $t_0 = 0$ and t_j denotes the sacrifice time point for $j = 1, \dots, J$. For an experiment with a single sacrifice, t_J denotes the terminal sacrifice time point.

It is assumed that three independent random variables completely determine the observed outcome for each animal. The random variables are 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 competing risks, T_C . Also let $T_1 + T_2 = T_D$, where T_D represents the overall time to death from the tumor of interest. Thus the tumor of interest is present in an animal at the time of death if $T_1 \leq \min\{T_C, T_S\}$, where T_S denotes a scheduled time to sacrifice of an animal. When $T_D \leq \min\{T_C, T_S\}$, an animal dies from the tumor of interest. Otherwise, it dies from competing risks or sacrifice.

2.1 Distribution of the Random Variables

- Time to tumor onset (T_1)

Let $S_i(t)$ be a survival function of the i -th group with respect to a random variable T_1 representing time to onset of the tumor of interest. Assume $S_i(t)$ follows a Weibull distribution:

$$S_i(t) = \exp\left[-\theta_i \delta_1 (t/t_{\max})^{\delta_2}\right], \quad (1)$$

where $\delta_1 \geq 0$, $\delta_2 \geq 0$, and t_{\max} represents the duration of the study or the time for a terminal sacrifice. The hazard ratio θ_i between the i -th dose group and the control group ($i = 1$) is typically greater than or equal to 1 ($\theta_i = 1$ for $i = 1$ and $\theta_i > 1$ for $i > 1$) for $i = 1, \dots, G$. The scale parameter δ_1 can be calculated by specifying the tumor onset probability $1 - S_1(t_{\max})$ at the end of the study in the control group. With $\theta_1 = 1$ and a given shape parameter δ_2 , $\delta_1 = -\log S_1(t_{\max})$. In this estimator, the value of the parameter δ_2 of the Weibull distribution ranges between 1.0 and 6.0 in order to reflect a wide variety of tumor onset distributions. When there are no competing risks, the tumor onset probability at the end of the study in each dose group is determined by the hazard ratio and the baseline tumor onset probability in the control group by the end of the study.

• **Time to death from competing risks (T_C)**

The survival function for time to death from competing risks, T_C , is taken to be

$$Q_i(t) = \exp[-\phi_i(\gamma_1 t + \gamma_2 t^{\gamma_3})], \quad (2)$$

where $\phi_i \geq 1$, $\gamma_1 \geq 0$, $\gamma_2 \geq 0$ and $\gamma_3 \geq 0$ (Portier et al., 1986). With $\phi_1 = 1$ in the control group ($i = 1$), γ_3 can be calculated as

$$\log[-\{\log Q_1(t_{\max}) + \gamma_1 t_{\max}\} / \gamma_2] / \log t_{\max}$$

under the constraint that $Q_1(t_{\max}) < \exp(-\gamma_1 t_{\max})$, where $Q_1(t_{\max})$ is the probability of survival with respect to competing risks in the control group at the end of the study. The values of γ_1 and γ_2 are chosen as 10^{-4} and 10^{-16} , respectively. These values are close to the ones fitted to the historical control data such as Fisher 344 rats and B6C3F₁ mice in Portier et al. (1986). These parameter values can be also found in many other settings (Kodell and Ahn, 1996, 1997; Kodell et al., 1997; Ahn, Zhu and Yang, 1998; Ahn et al., 2002). The competing risks survival rate can be determined according to tumor types and historical data showing the survival rates of mice and rats. The value of ϕ_i can be calculated after specifying the competing risks survival rate $Q_i(t_{\max})$ in (2) such that $\phi_i = \log[Q_i(t_{\max})] / \log[Q_1(t_{\max})]$.

• **Time to tumor death (T_2)**

For simplicity, the survival distribution for tumor-induced mortality, T_2 , is taken to have the same form as that for death from competing risks

$$F_i(t) = \exp[-\psi_i(\gamma_1 t + \gamma_2 t^{\gamma_3})], \quad (3)$$

and the values of γ_1 , γ_2 and γ_3 remain the same as in (2). These types of models using a modified Weibull distribution can be found in other literature (Kodell, Chen and Moore, 1994; Ahn and Kodell, 1995; Kodell and Ahn, 1997). The parameter ψ_1 is selected by the user to reflect various tumor lethality from low tumor lethality (where approximately 5% of observed tumors are the cause of death) to high tumor lethality (where approximately 60% or higher of observed tumors are the cause of death) in the control group. We also assume that once the tumor is developed, the distribution of time to death (T_2) is the same in all dose groups. Therefore, $\psi_i = \phi_1$ for all i . The lethality parameter indicates how the presence of a tumor of interest affects survival. A highly lethal tumor could lead to death shortly after its onset. Less restrictive choices for the distribution of T_2 can also be considered.

2.2 Construction of the Peto Test

The data are generated according to the distributions of T_1 , T_2 and T_C for each animal. They are collected at the j -th interval according to the following five events: animals died from the tumor of interest (d_j), animals died from competing risks while having the tumor of interest (a_{1j}), animals died without tumor (b_{1j}), animals sacrificed with tumor (a_{2j}), and animals sacrificed without tumor (b_{2j}). An animal dies from the tumor of interest in the j -th interval if $T_D \in I_j$ and $T_C > T_D$. An animal dies from other causes with the tumor of interest in the j -th interval if $T_D > T_C$, $T_1 \leq T_C$, and $T_C \in I_j$. On the other hand, an animal dies without the tumor of interest in the j -th interval if $T_C \in I_j$ and $T_1 > T_C$. A sacrificed animal has the tumor of interest at the time of sacrifice (t_j) if $T_1 \leq t_j$, $T_D > t_j$, and $T_C > t_j$. It does not have the tumor at the time of sacrifice when $T_1 > t_j$ and $T_C > t_j$. These data are applied to the Peto test to estimate sample size and power.

First, consider the animals that did not have the specific tumor before death or tumor-bearing animals for which that tumor was not the cause of death. Let $n_{ij} = a_{1ij} + a_{2ij} + b_{1ij} + b_{2ij}$ be the number of animals in group i dying during interval I_j from causes unrelated to the presence of the tumor of interest, and let $y_{ij} = a_{1ij} + a_{2ij}$ be the number of these animals in which the tumor was observed in the incidental context, for $i = 1, \dots, G$ and $j = 1, \dots, J$. For each interval I_j , the tumor prevalence data may be summarized in a $2 \times G$ table, as in Table 1. All tumors found

Table 1: Tumor prevalence data for incidental tumors in interval I_j

	Dose group				Total
	1	2	...	G	
# with tumors	y_{1j}	y_{2j}	...	y_{Gj}	$y_{.j}$
# without tumors	$n_{1j} - y_{1j}$	$n_{2j} - y_{2j}$...	$n_{Gj} - y_{Gj}$	$n_{.j} - y_{.j}$
# deaths	n_{1j}	n_{2j}	...	n_{Gj}	$n_{.j}$

in sacrificed animals are classified as incidental. The intervals defined by the pre-assigned NTP intervals (Bailer and Portier, 1988) are recommended to implement the incidental part of the Peto test.

The expected number of tumors in the i -th group for the j -th interval is $E_{ij} = y_{.j}K_{ij}$, where $K_{ij} = n_{ij}/n_{.j}$. Thus, the observed and expected numbers of tumors in the i -th group over the entire experiment are $O_i = \sum_{j=1}^J y_{ij}$ and $E_i = \sum_{j=1}^J E_{ij}$, respectively, for $i = 1, \dots, G$. Define

$$D_i = O_i - E_i = \sum_{j=1}^J (y_{ij} - E_{ij})$$

and

$$V_{ri} = \sum_{j=1}^J \kappa_j K_{rj} (\delta_{ri} - K_{ij}),$$

where $\kappa_j = y_{.j}(n_{.j} - y_{.j})/(n_{.j} - 1)$, and δ_{ri} is defined as 1 if $r = i$ and 0 otherwise. Let $\mathbf{D}_a = (D_1, \dots, D_G)^T$, and \mathbf{V}_a be the $G \times G$ matrix with (r, i) entry V_{ri} .

Second, consider the animals that died with a tumor of interest. The method used for the fatal tumors is similar to that used for the incidental tumors. Table 2 is a contingency table for tumor mortality data in the j -th interval. Data-determined intervals defined by the actual death time of an animal were used for the fatal tumor analysis. Let m_{ij} be the number of animals in group i surviving at the beginning of the interval, and $x_{ij} = d_{ij}$ be the number of these animals dying from the tumor of interest in that interval. A vector \mathbf{D}_b that has the differences of observed and expected values using the data in Table 2 is calculated in the same way as for the incidental tumors, and the corresponding covariance matrix \mathbf{V}_b is computed.

The analysis of data on occult tumors using contexts of observation is based on the vector $\mathbf{D} = \mathbf{D}_a + \mathbf{D}_b$, with covariance matrix $\mathbf{V} = \mathbf{V}_a + \mathbf{V}_b$. Then a dose-related trend test can be considered by using

$$Z_R = \mathbf{l}^T \mathbf{D} / \sqrt{\mathbf{l}^T \mathbf{V} \mathbf{l}},$$

Table 2: Tumor mortality data for interval I_j

	Dose group				Total
	1	2	...	G	
# fatal tumor deaths in I_j	x_{1j}	x_{2j}	...	x_{Gj}	$x_{.j}$
	$m_{1j} - x_{1j}$	$m_{2j} - x_{2j}$...	$m_{Gj} - x_{Gj}$	$m_{.j} - x_{.j}$
# surviving in I_j	m_{1j}	m_{2j}	...	m_{Gj}	$m_{.j}$

where $\mathbf{l} = (\ell_1, \dots, \ell_G)^T$, and ℓ_i stands for the dose metric for the i -th group with $0 = \ell_1 < \ell_2 < \dots < \ell_G$. Under the null hypothesis, Z_R is asymptotically distributed as a standard normal.

3 Usage of Sample Size and Power Estimator

The proposed estimator for sample size and power takes input parameters from a series of pages in the Web site (<http://biostatistics.mdanderson.org/ACSS>). The title page in Figure 1 provides a general description of the proposed estimator. By clicking the *Continue* button, it moves to a user log-in and registration page, as shown in Figure 2. A new user is required to register to obtain a user name and a password.

Users may save or retrieve their work by entering or selecting a file name. For a test run, the session may remain as “default”, as shown in Figure 3. A default session name automatically generated by concatenating the user name, date and time is provided, and it shall be used as the “subject” of an e-mail for delivering the output. The session name may be changed as the user wishes. Figure 4 shows a page of the detailed input parameters in an experimental design. It starts with requesting three input parameters (a) the number of dose groups, (b) whether the experiment uses multiple sacrifices or a single terminal sacrifice, and (c) an integer seed for the random number generator. The number of dose (or treatment) groups commonly considered are 2 to 4 groups, including the control group. An experiment with multiple sacrifices is simulated to perform sacrifices at specified interim time points, as well as a terminal sacrifice at the end of the study. All the remaining live animals are assumed to be sacrificed at the end of the experiment. The number of scheduled sacrifices, including the terminal sacrifice, is typically either 3 or 4 in a two-year study. In addition, a seed for the random number generator can be chosen by the user as any positive integer for the Monte Carlo simulation.

Figure 5 shows other detailed input parameters, which are (a) dose metric in each group, (b)

The University of Texas MD Anderson Cancer Center

Sample Size and Power Estimation for Animal Carcinogenicity Studies

Name: Web-based simulator for sample size and power estimation in animal carcinogenicity studies

Objective: A Web-based statistical tool for sample size and power estimation in animal carcinogenicity studies is developed. Given the design parameters, the program computes statistical power for detecting a dose-related trend in the occurrence of a tumor of interest in the presence of competing risks. By varying the sample size the program can be used to determine the required sample size for achieving sufficient power. It is designed for rodent tumorigenicity assays that have either a single terminal sacrifice or multiple sacrifices. It is applicable for studies on occult tumors for which the time to tumor onset is not directly observable.

Authors: Hojin Moon, J. Jack Lee, Hongshik Ahn and Rumiana Nikolova

Input Parameters:

- Number of dose groups
- Type of experiment with number of sacrifices
- Dose metric in each group
- Sample size per group
- Sacrifice time points
- Number of sacrificed animals at each interval
- Tumor onset probability in the control group
- Distribution of tumor onset
- Hazard ratio between each dose group and a control group
- Competing risks survival rate in each group
- Lethality parameter to determine lethality rate in the control
- Level of a one- or two-sided test
- Number of simulation runs

Output:

- Average tumor rate per dose group
- Average competing risks survival rate per dose group
- Average lethality per dose group
- Average death rate with information on tumors and sacrifices per dose group
- Simulated power

Reference: "**Web-based Simulator for Sample Size and Power Estimation in Animal Carcinogenicity Studies**"
Hojin Moon, J. Jack Lee, Hongshik Ahn and Rumiana Nikolova submitted to JSS (2002)

Continue

Figure 1: A general description of the proposed estimator

The University of Texas MD Anderson Cancer Center
Sample Size and Power Estimation for Animal Carcinogenicity Studies

User name
Password
New user? [Start here](#)
Do you want to change your password? [Click here](#)

Figure 2: A user log-in page

Input page 1 of 3

The University of Texas MD Anderson Cancer Center
Sample Size and Power Estimation for Animal Carcinogenicity Studies

Load data from file:
Session name:

Figure 3: The first input page

The University of Texas MD Anderson Cancer Center

Sample Size and Power Estimation for Animal Carcinogenicity Studies

- Number of groups (a control and dose/treatment groups):
- Experiment:
 - multiple sacrifices : Number of scheduled sacrifices (including a terminal sacrifice):
 - single sacrifice
- Positive integer seed for the random number generator:

Figure 4: Input parameters on the sample size and power estimation

sample size per group, (c) sacrifice time points, (d) number of sacrificed animals at the end of each time interval, (e) tumor onset probability in the control group, (f) shape parameter of the Weibull distribution for time to tumor onset, (g) hazard ratio between each dose group and the control group, (h) competing risks survival rate per group, (i) tumor lethality rate (low, intermediate, and high), (j) value of the lethality parameter to determine a lethality rate in the control group, (k) significance level of one- or two-sided test, and (l) number of simulation runs. Typically, animal carcinogenicity studies on occult tumors are conducted in a duration of 104 weeks. This statistical tool seeks an efficient reduced design (78 - 104 weeks) with an acceptable power. The default values are shown in Figures 3 - 5 for a typical two-year bioassay experiment that has multiple sacrifices. For a design with 4 groups, the dose levels may be set as a relative dose metric of 0, 1, 2 and 3 or 0, 1, 2 and 4 with respect to the control and 3 dose groups. Alternatively, the actual dose levels may be used in a design. In a typical two-year animal carcinogenicity study on occult tumors, 50 or more animals are considered in a group. However, a different number of animals per group may be used in a design. One needs to specify the time points of sacrifices in weeks. For a two-year study, NTP intervals are, for example, 0 - 52, 53 - 78, 79 - 92, and 93 - 104 weeks. For analyzing single-sacrifice data with the Peto test (described in Section 2), NTP intervals were used for the

incidental tumors. The NTP intervals for a two-year study are given as a default. The number of sacrificed animals is decided at the beginning of a study. Either the same number of sacrificed animals or a different number of animals can be specified in each group and/or each interval.

The probability of tumor onset by the end of the study in the control group needs to be specified in a design to determine the time to tumor onset. A tumor onset probability of each dose group, therefore, is determined by a hazard ratio and a tumor onset probability in the control group. An underlying distribution of time to tumor onset is assumed to be a Weibull distribution (Kodell and Ahn, 1997). The shape parameter in the Weibull tumor onset distribution ranges between 1.0 (exponential distribution) and 6.0 in order to reflect a wide variety of tumor onset distributions. The hazard ratio is a ratio of hazard rates between a dose group and the control group. A user needs to specify $G - 1$ hazard ratios, where G is the number of groups, in the estimation of sample size to have a desirable degree of power to detect a specified hazard ratio. Intercurrent mortality refers to all deaths not related to tumors of interest. The causes of intercurrent mortality are referred to as competing risks. Since a specific type of tumor may occur in the presence of competing risks, intercurrent mortality rates (competing risks survival/death rates) are considered in our design. They may be the same or different across dose groups.

The lethality parameter needs to be set to govern the tumor lethality rate in the control group. Three tumor lethality rates (low, intermediate and high) are considered in this paper. Tumors of low lethality have rates below 5%. Tumors of intermediate lethality are those with rates between 35% and 40%. The rates 60% and above are considered as highly lethal. To facilitate the choice of a tumor lethality parameter, at least one reference table, depending on study duration, tumor onset probability and tumor onset distribution, will be shown in a separate window by clicking the *Help* button that is illustrated in Figure 5. The table entries are obtained from a database that stores the information of a wide range of lethality rates by the duration of the study, tumor rate and tumor onset distribution. The value of a lethality parameter, then, can be chosen from the reference tables. A reference table is shown in Figure 6 as an example.

The significance level of the test is selected from 1%, 5% or 10%. The choice of a one- or two-sided test is also needed to estimate power. The number of data sets (i.e., simulation runs) needs to be decided for the Monte Carlo simulation. In this example, 5000 simulation runs are given as a default value. The e-mail address of the user is required to receive simulated results via e-mail.

The University of Texas MD Anderson Cancer Center

Sample Size and Power Estimation for Animal Carcinogenicity Studies

- Dose metric in each group in the order of control, dose1, dose2, etc.
(Typically, 0 for the control group: group 0)

group 0	group 1	group 2	group 3
0	1	2	4

- Number of animals in each group:

group 0	group 1	group 2	group 3
50	50	50	50

- Sacrifice time points (in WEEKS):
(Note: Please use integer values.)

Interim Sacrifices

t ₁	t ₂	t ₃	
52	78	92	weeks.

Terminal sacrifice at end of the study:

104	weeks.
-----	---------------

- Number of INTERIM sacrificed animals in each group (row) at each interim sacrifice time point (column). Exclude the number of terminal sacrifice animals.

group	Number of animals per sacrifice		
	t ₁	t ₂	t ₃
0	6	6	6
1	6	6	6
2	6	6	6
3	6	6	6

- Tumor onset probability by the end of the study in the control group:

0.33

Figure 5: More input parameters on the sample size and power estimation

- Tumor onset distribution: Weibull with a shape parameter
Note: Choose a parameter between 1.0 (exponential distribution) and 6.0

- Hazard ratio of each dose group versus the control group:

group 1	<input type="text" value="2"/>
group 2	<input type="text" value="2.5"/>
group 3	<input type="text" value="3"/>

- Competing Risks Survival Rate (CRSR) for each group (low -> high dose group).

group 0	group 1	group 2	group 3
<input type="text" value="0.7"/>	<input type="text" value="0.7"/>	<input type="text" value="0.7"/>	<input type="text" value="0.7"/>

- Please select the lethality rate:

- Type a value for the **tumor lethality parameter** to decide tumor lethality rate in the control group [Help](#):

- Level of the test(0.01 for 1%, 0.05 for 5%, 0.10 for 10%):

- Enter 1 for one-sided test or 2 for two-sided test :

- Number of simulation runs (recommended 5000):

- The output will be sent to the following e-mail address:

Save input data to file:

Figure 5: More input parameters on the sample size and power estimation (continued)

Helpful guidelines for selecting the tumor lethality parameter

104 weeks of study (Weibull 3)		
Tumor rate	Lethality rate	Tumor Lethality Parameter
0.3	high	1500
0.35	high	1400

Figure 6: Helpful guidelines for selecting the tumor lethality parameter

Input Parameters

Parameter	Value
Fixed Seed	3000
Number of groups	4
Number of sacrifices	4
Duration of the study (in weeks)	104
Tumor onset probability	0.33
Weibull tumor onset distribution with shape parameter	3
Tumor lethality	high
Tumor lethality parameter	1450
Level of test	0.05
One-sided or two-sided test	1
Number of simulation runs	5000

Hazard Ratios

Hazard Ratio
2
2.5
3

Competing Risks Survival Rates (CRSR)

Group	Number of animals	CRSR	Dose Metric
0	50	0.7	0
1	50	0.7	1
2	50	0.7	2
3	50	0.7	4

Figure 7: Input parameters considered

Interim Sacrifices

Number of Interim Sacrifices			
group	Number of animals per sacrifice		
	52	78	92
0	6	6	6
1	6	6	6
2	6	6	6
3	6	6	6

Figure 7: Input parameters considered (continued)

Before submitting input parameters, all design parameters are displayed once again on the Web page for verification (See Figure 7). Once all the input parameters entered are confirmed, they are submitted to the compiled C program for simulation runs.

When the simulation runs are completed, an output file is sent to the e-mail address specified by the user. A simulation with 5000 runs typically takes about 5 minutes. The output file contains design parameters, average tumor rate, average competing risks survival rate, average lethality rate for each group, and average death rate with information on tumors and sacrifices per dose group. At the end, the power to detect a dose-related trend is shown. Figure 8 is an image of an output file as an example.

4 Example: Testing the Carcinogenic Potential of NNK in a Prevention Study for Lung Cancer

The development of this statistical tool was motivated by a recent lung cancer prevention study developed at M. D. Anderson Cancer Center. One particular goal was to evaluate the carcinogenic potential of the tobacco carcinogen NNK, a byproduct of tobacco smoke, on the incidence of lung adenomas and adenocarcinomas (occurring either spontaneously or following exposure to the tobacco carcinogen NNK) in transgenic mice that express antisense retinoic acid receptor- β (RAR-

To: hmoon@mail.mdanderson.org
cc: (bcc: Hojin Moon/MDACC)

Subject: hojin::3/8/2002::14:42:41

The University of Texas MD Anderson Cancer Center
Sample Size and Power Estimation for Animal Carcinogenicity Studies

*** Input Parameters ***

Selected Seed = 3000
Number of Groups = 4
Dose metric of each group:
0.00 1.00 2.00 4.00
Number of animals in each group
50 50 50 50
Number of sacrifices including a terminal sacrifice = 4
Sacrifice time points in weeks:
52 78 92 104
Study duration = 104 weeks
Number of INTERIM sacrificed animals in each interval:
6 6 6
6 6 6
6 6 6
6 6 6
Background tumor onset probability at the end of the study = 0.33
Tumor onset distribution assumed: Weibull with a shape parameter 3.00
Hazard ratio(s) of dose vs. control group
2.00 2.50 3.00
Competing Risks Survival Rate (CRSR) for each group:
0.70 0.70 0.70 0.70
Tumor lethality parameter entered = 1450.00
Level of the test = 0.05
One-sided or two-sided test = 1 sided test
Number of simulation runs = 5000

Figure 8: Contents of an output file

```

*** Simulation Results ***

dose group 0:
average tumor rate = 0.3298
average competing risks survival rate = 0.6982
average lethality = 0.8483

sacrifice time d      a1      b1      a2      b2
52             0.0332  0.0000  0.0068  0.0017  0.1133
78             0.0763  0.0009  0.0292  0.0037  0.0970
92             0.0533  0.0027  0.0548  0.0047  0.0783
104            0.0411  0.0050  0.0767  0.0226  0.2989

dose group 1:
average tumor rate = 0.5520
average competing risks survival rate = 0.6998
average lethality = 0.8719

sacrifice time d      a1      b1      a2      b2
52             0.0664  0.0001  0.0064  0.0034  0.1077
78             0.1411  0.0016  0.0258  0.0064  0.0819
92             0.0892  0.0042  0.0431  0.0071  0.0597
104            0.0624  0.0075  0.0531  0.0317  0.2014

dose group 2:
average tumor rate = 0.6341
average competing risks survival rate = 0.6993
average lethality = 0.8811

sacrifice time d      a1      b1      a2      b2
52             0.0829  0.0001  0.0062  0.0041  0.1054
78             0.1683  0.0019  0.0244  0.0079  0.0747
92             0.1027  0.0049  0.0392  0.0079  0.0514
104            0.0677  0.0083  0.0444  0.0336  0.1640

dose group 3:
average tumor rate = 0.6982
average competing risks survival rate = 0.7007
average lethality = 0.8909

sacrifice time d      a1      b1      a2      b2
52             0.0974  0.0002  0.0063  0.0047  0.1026
78             0.1932  0.0022  0.0232  0.0087  0.0689
92             0.1124  0.0052  0.0345  0.0084  0.0455
104            0.0706  0.0083  0.0381  0.0343  0.1350

Positive Trend (Power): 0.9386

```

Figure 8: Contents of an output file (continued)

β). A Web-based sample size and power estimator was applied to provide sufficient power to detect a dose-related trend of the occurrence of lung adenomas and adenocarcinomas. The estimation of sample size and power was conducted by the Peto test (see Section 2).

The power computation is based on the following assumptions. One hundred antisense RAR- β hemizygous (+/0) mice and one hundred antisense RAR- β homozygous (+/+) mice will be obtained for the experiments. For each type of mice, fifty mice will be randomized into a group that either receives or does not receive exposure to NNK. To test the carcinogenic potential of NNK, a dose metric of 0 or 1 is used for the control group or exposed group, respectively. Serial sacrifices are scheduled at weeks 39, 52, 65, and 78 (at the end of the study). The time to lung cancer development is assumed to follow a Weibull distribution with the shape parameter 3. It is expected that 55% of the antisense RAR- β hemizygous mice will develop a lung cancer by 78 weeks. On the other hand, 86% of the antisense RAR- β homozygous mice are expected to develop a lung cancer by 78 weeks. In this experimental design, 85% competing risks survival rate is considered. The lung adenomas and adenocarcinomas are assumed to be highly lethal. Five thousand simulation trials are run.

The statistical power (in %) under the one-sided 5% nominal significance level is listed in Table 3. A hazard ratio between the treatment group and the control group in each type of mice is chosen as 2.0, 2.5 and 3.0. Three different designs are considered. One is to estimate power with 6 mice at each serial sacrifice (39, 52 and 65 weeks) in a total of 50 mice per group. Another setting is to calculate power with 6 sacrifices at 52 weeks out of a total of 55 in the control group and with 3 sacrifices at 52 weeks out of a total of 45 in the dose group. The other configuration is to estimate power with a total of 30 mice per group and 3 mice per serial sacrifice at 39, 52 and 65 weeks. The same competing risks survival rate (85% for the control and a dose group) and different competing risks survival rates (85% for the control, 50% for a dose group) are considered in this example. Under the null hypothesis of no treatment effect on the tumor of interest, it may still be reasonable to assume different competing risk rates among different dose groups. For example, even though the carcinogen may have no effect on the development of lung cancer, it may increase competing risks, such as the development of a liver tumor or bladder tumor, etc, and, consequently, results in lower competing risks survival rates in the dose groups. A lung cancer is typically considered as highly lethal in humans. The median survival time is less than a year for stage III and less than

6 months for stage IV non-small cell lung cancer (Ginsberg et al., 1993). Lethality parameters of 1500 and 800 are selected in hemizygous and homozygous mice, respectively, to reflect highly lethal tumors. Expected lethality corresponding to these parameters is about 80% in this example.

With 50 hemizygous mice in each group and the same competing risks survival rate between groups, 79.6% power was achieved to detect a hazard ratio of 2. However, the power decreased slightly to 76.3% with different competing risks survival rates. On the other hand, at least 80% power was achieved with 50 homozygous mice in each group. With 30 mice in each group of hemizygous and homozygous mice, we had less than 80% power to detect a hazard ratio of 2 in the presence of both the same and different intercurrent mortalities. A better design, in this example, was achieved with the control group of 6 mice for an interim sacrifice at 52 weeks out of 55 mice in total, and with a treatment group of 3 mice for an interim sacrifice at 52 weeks out of 45 mice in total. In this design, 80% power was achieved to detect a hazard ratio of 2 or higher among antisense RAR- β hemizygous and homozygous mice with the same and different competing risks survival rates.

5 Discussion

We have developed, based on our knowledge, the first Web-based sample size and power simulator for animal carcinogenicity studies to detect a dose-related trend in tumor incidence following exposure to a putative carcinogen. It is applicable for studies on occult tumors for which the time to tumor onset is not directly observable. It was designed for rodent bioassays that have either multiple sacrifices or a single terminal sacrifice.

The Peto test (Peto et al., 1980) is used to compare the incidence rate of occult tumors among groups in the presence of potential confounders. It requires data with cause-of-death information determined by pathologists. Monte Carlo simulation is introduced in this tool to simulate experiments of rodent bioassays because no closed-form solution is available.

This package could be used to construct a design with a sufficient power to detect specified hazard ratios by varying the sample size, number of sacrifices, time points for sacrifice, and number of sacrificed animals at each sacrifice, if any, under the given design considerations. As an example, the application of this tool was illustrated by an animal experiment for lung cancer prevention

Table 3: Power calculation for animal carcinogenicity studies of NNK in Antisense RAR- β mice

Type of mice	CRSR ^a	Hazard ratio	(6,50) ^b	(6,55,3,45) ^c	(3,30) ^d
Antisense RAR- β hemizygous (+/0) mice	same ^e	2.0	79.6	84.7	60.8
		2.5	95.9	97.5	84.4
		3.0	99.4	99.7	94.1
	different ^f	2.0	76.3	80.7	56.9
		2.5	94.6	96.1	80.7
		3.0	99.1	99.3	92.6
Antisense RAR- β homozygous (+/+) mice	same	2.0	85.6	87.3	65.8
		2.5	97.1	97.9	84.6
		3.0	99.4	99.6	93.0
	different	2.0	83.4	84.2	63.9
		2.5	96.6	96.8	83.3
		3.0	99.4	99.3	92.5

^aCompeting Risks Survival Rate

^b6 mice at each serial sacrifice (39, 52, 65 weeks) and a total of 50 mice per group

^c6 mice at an interim sacrifice (52 weeks) and a total of 55 mice (control group);

3 mice at an interim sacrifice (52 weeks) and a total of 45 mice (dose group)

^d3 mice at each serial sacrifice (39, 52, 65 weeks) and a total of 30 mice per group

^e85% CRSR for the control and a dose group

^f85% CRSR for the control and 50% CRSR for a dose group

research. An advantage of the proposed Web-based sample size and power estimator is wider accessibility to the user, provided that an Internet connection and a Web browser are available (Microsoft Internet Explorer 5.5 and above, or Netscape 4.76 and above). In addition, this statistical tool provides a user-friendly environment so that the user can search for an optimal design.

Figure 9 shows a flowchart for the communication between the Web front end and the core program. Input data are supplied to the proposed tool via a user-interface implemented with Web pages. When the user presses the *Submit* button on the last Web page, the input data are sent to the Web server machine. There it is packaged into XML format and sent to a queue for temporary storage until the time at which it can be scheduled for processing by the core calculation program. Every few minutes a *Queue Reader* process running on a separate compute-server machine scans the queue for waiting input data. When new data are found, it is sent (still in XML format) to the core calculation program. This returns the results to the *Queue Reader*. The *Queue Reader* then sends the results to the user via e-mail. This asynchronous “queued” architecture is chosen so that multiple jobs requiring calculation times ranging from a few seconds to several hours could be accommodated. All communication with the queue is done through a custom software component implemented with Microsoft’s Component Object Model (COM) technology. This allows the actual queue (a relational database) to be abstracted from the remainder of the system, so the queue can be changed/upgraded without requiring additional maintenance on the rest of the system.

Through a Monte Carlo simulation study, the proposed tool can seek an efficient design. For a given sample size and study duration, an optimal design can be obtained using our tool by choosing the design with the maximum power of the trend test. This tool can help investigators conduct more efficient and cost effective experiments.

Acknowledgements

This work was supported in part by the Tobacco Settlement Funds approved by the Texas State Legislature (Hojin Moon and J. Jack Lee), and NIH grants CA91844 (J. Jack Lee) and CA77289 (Hongshik Ahn). Authors thank Clift Norris for programming the COM object and LeeAnn Chastain for editorial assistance.

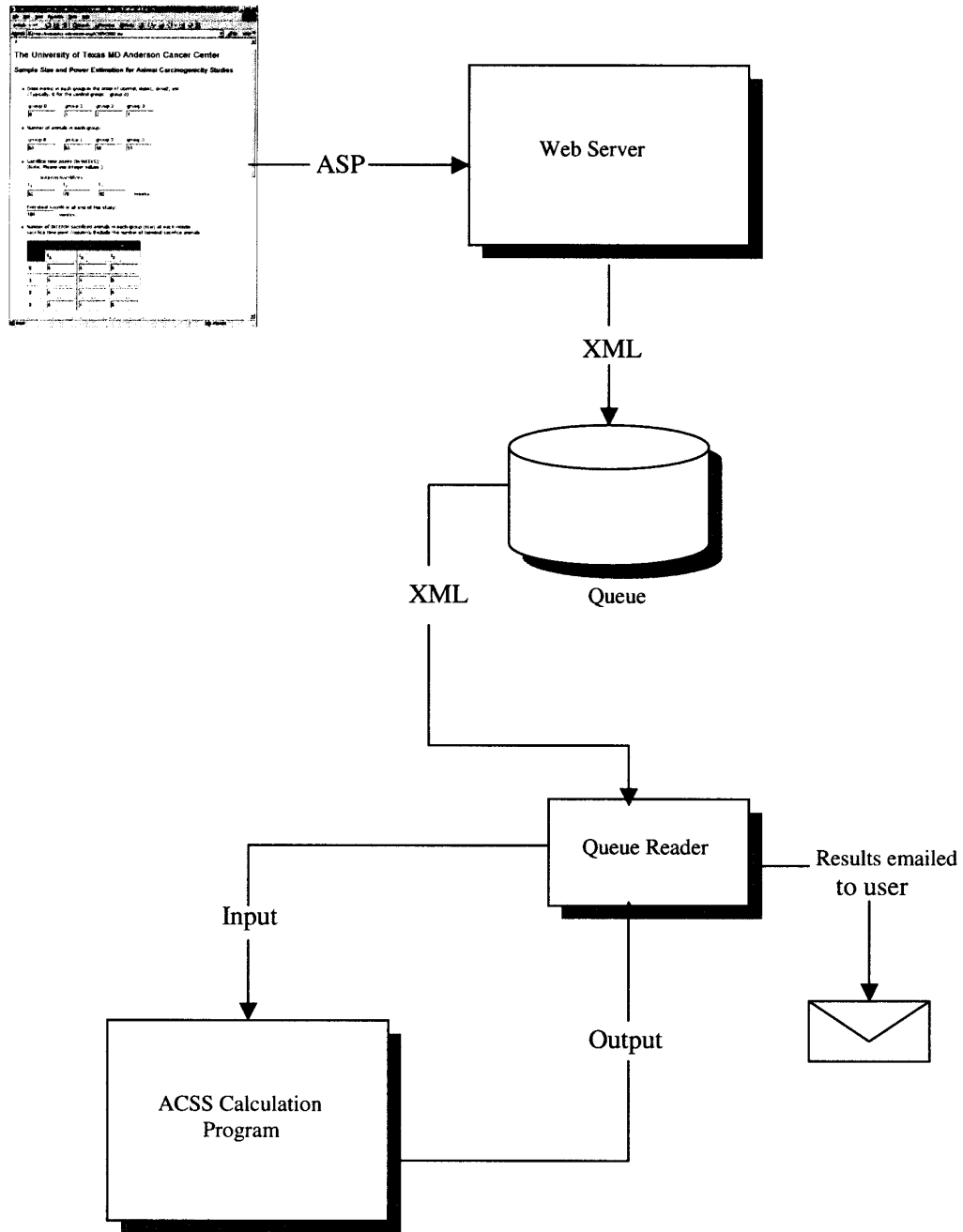


Figure 9: A flowchart for the communication between the front end and the core program

References

- Ahn, H. and Kodell, R. L. (1995), Estimation and Testing of Tumor Incidence Rates in Experiments Lacking Cause-of-Death Data. *Biometrical Journal*, **37**, 745-763.
- Ahn, H., Moon, H., Kim, S. and Kodell, R. L. (2002), A Newton-based approach for attributing tumor lethality in animal carcinogenicity studies. *Computational Statistics and Data Analysis*, **38**, 263-283.
- Ahn, H., Zhu, W. and Yang J. (1998), Efficient designs for animal carcinogenicity experiments. *Communications in Statistics - Theory and Methods*, **27**, 1275-1287.
- 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.
- Banfield J. (1999), Rweb: Web-based Statistical Analysis. *Journal of Statistical Software*, **4.1**, URL: <http://www.stat.ucla.edu/journals/jss/>.
- Firth, D. (2000), Quasi-variances in Xlisp-Stat and on the web. *Journal of Statistical Software*, **5.4**, 1-13. URL: <http://www.stat.ucla.edu/journals/jss/>.
- Ginsberg, R. J., Kris, M. G. and Armstrong, J. G. (1993), Non-small cell lung cancer. In: *Cancer Principles and Practice of Oncology*, DeVita, V., Hellman, S. and Rosenberg, S. A. (Eds.), 673-722. J. B. Lippincott Co., Philadelphia.
- Hoel, D. and Walburg, H. E. (1972), Statistical analysis of survival experiments. *Journal of National Cancer Institute*, **49**, 361-372.
- Kodell, R. L. and Ahn, H. (1996), Nonparametric trend test for the cumulative tumor incidence rate. *Communications in Statistics - Theory and Methods*, **25**, 1677-1692.
- Kodell, R. L. and Ahn, H. (1997), An age-adjusted trend test for the tumor incidence rate. *Biometrics*, **53**, 1467-1474.
- Kodell, R. L. and Ahn, H., Pearce, B. A., and Turturro, A. (1997), Age-adjusted trend test for the tumor incidence rate. *Drug Information Journal*, **31**, 471-487.
- Kodell, R. L., Chen, J. J., and Moore, G. E., (1994), Comparing distributions of time to onset of disease in animal tumorigenicity experiments. *Communications in Statistics - Theory and Methods*, **23**, 959-980.
- Mantel, N. and Haenszel, W. (1959), Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of National Cancer Institute*, **22**, 719-728.
- Peto, R., Pike, M. C., Day, N. E., Gray, R. G., Lee, P. N., Parish, S., Peto, J., Richards, S., and Wahrendorf, J. (1980), Guidelines for simple, sensitive significance tests for carcinogenic effects in long-term animal experiments. Annex to: *Long-term and Short-term Screening Assays for Carcinogens: a Critical Appraisal*. IARC Monographs, Supplement 2, 311-426. International Agency for Research on Cancer: Lyon, France.

- 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.
- West, R. W. and Ogden, R. T. (1997), Statistical Analysis with Webstat, a Java applet for the World Wide Web. *Journal of Statistical Software*, **2.3**,
URL: <http://www.stat.ucla.edu/journals/jss/>.