1992-01

Modeling and statistical analysis of Medaka bioassay data

Gaver, Donald P.

Naval Postgraduate School

(2)

AD_____

MIPR NO: 91MM1598

TITLE: MODELING AND STATISTICAL ANALYSIS OF BIOASSAY DATA

PRINCIPAL INVESTIGATOR: Donald P. Gaver

CONTRACTING ORGANIZATION: Department of the Navy
Naval Postgraduate School
Monterey, CA 93940

REPORT DATE: January 1, 1992

TYPE OF REPORT: Final Report

PREPARED FOR: U.S. ARMY MEDICAL RESEARCH AND DEVELOPMENT COMMAND
Fort Detrick, Frederick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
distribution unlimited

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

20030220 21

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1 January 1992	Final Report (8/15/91 - 12/31/91)

4. TITLE AND SUBTITLE
Modeling and Statistical Analysis of Bioassay Data

5. FUNDING NUMBERS
MIPR No.
91MM1598

62720A
3E162720A835.00.009
WUDA 335951

6. AUTHOR(S)
Donald P. Gaver

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of the Navy
Naval Postgraduate School
Monterey, CA 93940

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Development Command
Fort Detrick
Frederick, Maryland 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

14. SUBJECT TERMS
Bioassays, Analysis, Modeling, RA III, MIPR

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

PROGRESS REPORT ON

"MODELING AND STATISTICAL ANALYSIS OF BIOASSAY DATA"

Donald P. Gaver
Professor of Operations Research
Naval Postgraduate School
Monterey, CA 93943

1. BACKGROUND

The objectives of the above project were formulated in discussion with Mr. Henry Gardner of U.S. Army Medical R&D Command, Ft. Detrick, MD. The project purpose and workscope was stated in a MIPR, effective Aug. 05, 1991, as follows:

> The Department of Operations Research, Naval Postgraduate School, Monterey, California, shall conduct quantitative analysis of environmental bioassay data to be provided by the U.S. Army Biomedical Research and Development Laboratory (USABRDL). Individual cancer bioassays shall be evaluated as well as all field test data. Meta-analytical methods for assessing integrated biological assessment data shall also be accomplished.

2. APPROACHES TAKEN, AND PROGRESS

a) Data Availability

Detailed data, concerning pathologist's assessments of physical changes in (medaka) fish organs possibly resulting from toxin dosage, were sent to the investigator on Dec. 19, 1991; they arrived about a week later. Summarized data were available somewhat earlier, i.e. at the end of November. The form of the data clearly influences the type of statistical analysis and modeling, so our effort was directed towards formulating appropriate model types and

developing techniques for applying or fitting these to data. Actual applications, i.e. fits and error assessments, plus comments about the apparent sensitivity of results to the experimental/test designs (numbers of subjects/fish in groups, intervals between sacrifice, *etc.*) have been made in Appendix C but require extension. They can be enhanced during a subsequent contract phase, which will be proposed.

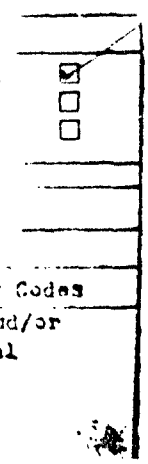
b) Model Development

The objective of the project, to date, has been to focus on biological issues believed to be important in converting toxin dosage to pre-cancerous and cancerous cells, and to translate these into appropriate quantitative mathematical terms. In particular, models that stem from the **clonal expansion** mechanisms identified by Moolgavkar and co-workers, cf. sample references (1979), (1983), have been studied, generalized (to account for possible variability or susceptibility between individual fish), and adapted for fitting to actual data. For an example see Appendix C, where numerical results, including uncertainty analyses, are presented.

c) Some Details Concerning Model Conception

It appears to be widely believed that pre-cancerous conditions in an organ (the liver) occur as a result of cell clonal expansion, followed by a promotion (to tumor) event. Specific models for this has been proposed and developed by Moolgavkar and co-workers. More recent work is by C. J. Portier and co-workers. References appear later.

The basic mechanism is treated as random or probabilistic: an **initiating event**, e.g., caused by contact with toxin, affects a cell within an organ in accordance with a simple Poisson process with rate parameter λ. That is, the



chance of an uninitiated cell being initiated in time interval $(t, t+h)$ is approximately λh. *If* a cell is initiated during exposure time it clones itself into other cells at rate β; the original cells and its clones die randomly at rate δ. All cells in the organ perform thus independently, according to the model. Depending upon the values of β and δ (birth and death rates respectively) a colony of initiated cells (pre-cancerous, presumably) either tends to grow exponentially, or to die off to zero (also exponentially fast). The fates of colonies characterized by the same values of birth rate and death rate may actually be entirely different, as befits experience with variability characteristic of the real biological world. This behavior is *roughly* analogous to that of the flipping of the same coin: on one occasion 10 flips may well result in an excess of 5 Heads (7 Heads and 3 Tails), analogous to more births (Heads) than deaths (Tails); on another sequence of 10 flips with the same coin the result may be exactly reversed (7 Tails, 3 Heads). Processes analogous to coin flipping or dice rolling can describe much, but possibly not all interesting biological variability pertinent to risk analysis. Other options are suggested later.

The values of β and δ describe clone colony properties in a precise probabilistic manner *if* the model is correct. It is only certainly approximate, but may still provide a useful tool for quantifying risk of tumor formation. The second step in the malignant cell development process is postulated to be *promotion*. A model for this is that at rate μ, i.e. with probability μh in time $(t, t+h)$ a promotion event occurs that affects one of the clone colony members in proportion to the current size of the colony; such events are assumed to occur in accordance with a Poisson process with rate proportional to instantaneous clone population size. At the instant that the *first* such

promotion event occurs the clone colony (if one exists, i.e. has been initiated) will be said to have developed a tumor, at least in informal layman's terms. Note that all original cells in an organ are assumed to be independently exposed to initiation and, thereafter, to promotion. Therefore all organ cells and subsequent clones, if any, must survive from initiation to the end of the observation period without being promoted in order for the organ to survive throughout.

The probabilistic mechanism described has been used to obtain a formula for the survival probability for an organ for any observation time t. See Appendix A for the formula and its derivation. Similar formulas have been derived also by Moolgavkar and others. Our formula provides the basis for statistically estimating from pathology data, (combinations of) the parameters: λ, the initiation rate; μ, the promotion rate; and β and δ, the clonal birth and death rates. Such estimates can, in turn, be used to estimate the probability of cell, and organ, survival for any time period. We also supply an alternative model, based on diffusion theory (use of mathematical Brownian motion) that adds certain intuitive features. Appendix B contains a discussion of maximum likelihood estimation from data so as to specify parameters of a preliminary model. Appendix C uses the preliminary statistical model of Appendix B to describe a particular data set. Further work is required to obtain additional statistical models and procedures to analyze other experimental data.

(d) Extensions to the Model: Extra-variation of Parameters

The above model, and the consequences thereof in the form of a survival probability function, are appealing since they have some plausible biologic al basis and represent a form of variability from organ to organ outcomes

(malignancy evident, or not) that is anticipated. On the other hand, *variability between organs in different subjects* (e.g. fish) *is not well represented.* Different, but superficially identical, biological entities, be they fish, rats, or humans, can be expected to have some differences; specifically these may be between the effective parameters λ, μ β and δ. *If* the above are estimated from data without recognizing the possibility of extra-variation, biased results will be obtained. See Harris (1990) for biological explanations of inter-organ (subject) variability.

There are two simple and preliminary ways of dealing with the above problem. One is by attempting to "explain" parameter variation by representing it as a function of some causal variable, such as the age, sex, weight, *etc.*, of the host subject. The technique could be some form of regression analysis; methods of McCullagh and Nelder (1983) suggest themselves. A description of a preliminary computational procedure to estimate model parameters is described in Appendix B. This procedure is used to estimate model parameters for a particular data set in Appendix C. A second approach is to assume that the variability between individual host organs can be represented by treating some or all of the parameters as random variables with their own distributions. A typical survival function is then obtained by *mixing*: the parametric survival function of Appendix A is "simply" randomized according to the (joint) distribution of the parameters. In principle it is desirable to recognize both sources of variability between individuals, adjusting for known sources of variation by a regression technique where possible, but recognizing the "unexplainable" variation by use of a mixing

technique. The latter has been carried out to a limited extent, see Gaver and Jacobs (1992).

(e) Recommendation

Further work, both theoretical and applied, is required to put the above ideas for characterizing and explaining extra variability into practice. It is proposed that this work, plus effort to characterize the errors or uncertainties of survival probabilities, be carried out in a follow-on to the current project.

3. CONTACT WITH RELATED RESEARCHERS AND INSTITUTIONS

We have established information-transfer and possible collaborative relations with several other establishments having related objectives. These are

(a) The National Institute of Environmental Health Sciences (NIEHS), Research Triangle, NOCAR

Specific contacts have been made with David *Hoel*, Joseph *Hasemian*, and Chris *Portier*. They have forwarded papers. Hoel has invited me to spend some time there, which I plan to do in the spring of FY92.

(b) RAND Corp., Santa Monica, CA,

A group interested in environmental epidemiology. Among them are Naihua *Duan* and Sandy *Geschwind*.

(c) EPA, Washington, DC

A Group interested in risk analysis. Dr. Herman *Gibb*.

(d) Naval Medical Research Institute, Toxicology Detachment, Wright-Patterson AFB, Dayton, OH.

Particularly Dr. Robert *Carpenter*.

We also explore contacts with Dr. Alice *Whittemore*, Professor of Epidemiology and Biostatistics at Stanford Univ., and with Prof. Nicholas *Jewell* of UC Berkeley, and elsewhere in academia.

APPENDIX A. TWO-STAGE MODELS

In this appendix we present two models for the distribution of time until a normal cell becomes promoted. The first model is a birth-and-death model. The second model is a diffusion model.

1. A BIRTH-AND-DEATH MODEL FOR THE TIME UNTIL A NORMAL CELL BECOMES MALIGNANT (IS PROMOTED)

We first develop an expression for the distribution of time until an initiated cell or one of its descendants becomes malignant.

Assume that there is one initiated cell at time 0. Such cells divide at an exponential rate β, and die at an exponential rate δ. Any initiated cell turns malignant at an exponential rate μ; i.e. μ is the promotion rate.

(a) Time to Promotion of an Initiated Cell

Let T be the random time at which some initiated cell or its descendent turns malignant; note that T may actually be infinite if the population of initiated cell and its descendents dies out. Put

$$z(t) = P\{T > t\}.$$

Then simple probability arguments give the equation

$$z(t) = e^{-(\beta+\delta+\mu)t} + \int_0^t (\beta+\delta+\mu)e^{-(\beta+\delta+\mu)s}\left[\frac{\delta}{\beta+\delta+\mu} + \frac{\beta}{\beta+\delta+\mu}z(t-s)^2\right]ds$$

$$= e^{-(\beta+\delta+\mu)t} + \frac{\delta}{\beta+\delta+\mu}\left[1-e^{-(\beta+\delta+\mu)t}\right] + \beta\int_0^t e^{-(\beta+\delta+\mu)s}z(t-s)^2 ds$$

$$= e^{-(\beta+\delta+\mu)t} + \frac{\delta}{\beta+\delta+\mu}\left[1-e^{-(\beta+\delta+\mu)t}\right] + \beta e^{-(\beta+\delta+\mu)t}\int_0^t e^{(\beta+\delta+\mu)y}z(y)^2 dy.$$

(A.1)

Differentiating with respect to t and simplifying gives

$$\frac{d}{dt}z(t) \equiv -(\beta+\delta+\mu)\left[z(t)-\frac{\delta}{\beta+\delta+\mu}\right]+\beta e^{-(\beta+\delta+\mu)t}e^{(\beta+\delta+\mu)t}z(t)^2$$

$$=\beta z(t)^2-(\beta+\delta+\mu)z(t)+\delta. \qquad \text{(A.2)}$$

Hence $z(t)$ satisfies a Riccati equation with initial condition

$$z(0) \equiv 1 \qquad \text{(A.3)}$$

The solution to (A.2) with initial condition (A.3) is

$$z(t)=\frac{\rho_1(1-\rho_2)-\rho_2(1-\rho_1)e^{\beta(\rho_1-\rho_2)t}}{1-\rho_2-(1-\rho_1)e^{\beta(\rho_1-\rho_2)t}} \qquad \text{(A.4)}$$

where $\rho_{1,2}$ are the solutions to the quadratic equation

$$x^2-\left(1+\frac{\delta}{\beta}+\frac{\mu}{\beta}\right)x+\frac{\delta}{\beta}=0; \qquad \text{(4.5)}$$

$$\rho_{1,2}=\frac{1}{2}\left[\left(1+\frac{\delta}{\beta}+\frac{\mu}{\beta}\right)\pm\left[\left(1+\frac{\delta}{\beta}+\frac{\mu}{\beta}\right)^2-4\frac{\delta}{\beta}\right]^{\frac{1}{2}}\right]. \qquad \text{(A.6)}$$

Since $\left[\left(1+\frac{\delta}{\beta}+\frac{\mu}{\beta}\right)^2-4\frac{\delta}{\beta}\right]^{1/2} \le \left(1+\frac{\delta}{\beta}+\frac{\mu}{\beta}\right)$, both ρ_1 and ρ_2 are positive. Further $\rho_2 \le 1$ and

$$\rho_1-\rho_2=\left[\left(1+\frac{\delta}{\beta}+\frac{\mu}{\beta}\right)^2-4\frac{\delta}{\beta}\right]^{\frac{1}{2}}>0.$$

Hence,

$$\lim_{t\to\infty} z(t) = \rho_2.$$

If the death rate $\delta = 0$, then $\rho_2 = 0$ and $\lim_{t\to\infty} P(T > t) = 0$; if $\delta \neq 0$, then there is no death of initiated cells and thus an initiated cell will transition to a malignant cell in a finite time with probability 1. If $\delta > 0$, then the initiating cells can die, thus preventing a transition to malignancy and hence $\lim_{t\to\infty} P(T > t) = \rho_2 > 0$.

(b) A Model for the Time until a Normal Cell becomes Malignant (is promoted)

Assume that each normal cell is initiated at an exponential rate λ_0. Let N be the total number of normal cells in an organ. Let S denote the first time a normal cell transitions to a malignant cell.

$$P\{S \geq t\} = \left[e^{-\lambda_0 t} + \int_0^t \lambda_0 e^{-\lambda_0 s} z(t-s)ds \right]^N \tag{A.7}$$

where z is a given in (A.4). Assume λ_0 is small and put $\lambda = \lambda_0 N$, a constant. Then

$$P\{S > t\} \approx \exp\left\{ N \ln\left[1 - \frac{\lambda}{N} t + \frac{\lambda}{N} \int_0^t z(s)ds \right] \right\} \tag{A.8}$$

$$\approx \exp\left\{ -\lambda t + \lambda \int_0^t z(s)ds \right\} \tag{A.9}$$

$$= \exp\left\{ \lambda(\rho_1 - 1)t - \lambda \frac{1}{\beta} \ln\left[\frac{1 - \rho_2 + (\rho_1 - 1)e^{\beta(\rho_1 - \rho_2)t}}{\rho_1 - \rho_2} \right] \right\}. \tag{A.10}$$

2. A DIFFUSION APPROXIMATION TO THE TIME TO PROMOTION

An alternative approach to constructing a model for time to promotion, i.e. appearance of a tumor, is *via diffusion approximation*; see Karlin and Taylor (1981), Chap. 15. The approach can be mathematically justified as an asymptotic limit of a sequence of discrete-state processes by use of weak convergence theory; see Kopp-Schneider, Portier and Rippman (1991) who have used this methodology also and discussed the weak convergence issues. Here our motivation is to model biological phenomena, i.e. variability, in a convenient and flexible manner, so questions of mathematical rigor can be temporarily suppressed.

Consider an organ that at time 0 contains x cells that have just been initiated. Think of x as a *real* fixed positive number, not insisting that it be an integer. Now suppose promotion is a Poisson process of rate μ, applying independently to each of the x cells. This implies that no promotion occurs in time $(t, t+h)$ with probability $e^{-\mu h} = 1-\mu h + o(h)$ as $h \to 0$. The diffusion approximation to the growth/death of the entire initiated cell population, of size $I(0) = x$ initially, changes to

$$I(h) = x + e_x$$

at time h (h small after initiation), with

e_x Normally distributed with mean $v(x)h + o(h)$, and variance $\sigma^2(x)h + o(h)$

Example. A diffusion approximation to the previous birth-death model would naturally equate the mean and variance of change in a small time interval of length h to $v(x)h$ and $\sigma^2(x)h$ respectively; these turn out to be

$$v(x)h = x(\beta-\delta)h \qquad \text{(A.11)}$$

and

$$\sigma^2(x)h = x(\beta+\delta)h \qquad \text{(A.12)}$$

and the birth-death model's increment distribution (net births over deaths) by a normal variable with the matched moments.

The parameters $v(x)$ and $\sigma^2(x)$ are known as, respectively,

$v(x)$: **infinitesimal mean, or drift;**

$\sigma^2(x)$: **infinitesimal variance, or diffusion coefficient.**

The $v(x)$ parameter describes the tendency of the process to move deterministically in a given direction during a time of length h; the $\sigma^2(x)$ parameter provides a scale for the random variability around the mean.

Put $z(t; x)$ for the survival probability, until tumor appearance, given that x cells are initially initiated:

$$z(t; x) = P\{T_x > t\} = P\{T > t \mid I(0) = x\}.$$

The assumptions made initially then suggest that z satisfies the following equation if $h \ll 1$:

$$z(t;x) = \int_{-\infty}^{\infty} e^{-\mu x h} \frac{e^{-\frac{1}{2}(y-v(x)h)^2 / \sigma^2(x)h}}{\sqrt{2\pi}\sqrt{\sigma^2(x)h}} z(t-h, x+y). \qquad \text{(A.13)}$$

Now as h is small then y will also assume small values with appreciable probability so replace $z(t-h, x+y)$ by its Taylor series expansion

$$z(t-h, x+y) = z(t,x) - \frac{h\partial z}{\partial t} + y\frac{\partial z}{\partial x} + \frac{y^2}{2}\frac{\partial^2 z}{\partial x^2}.$$

Then the integral can be performed and

$$z(t;x) = (1-\mu\times h)\left(z(t;x) - h\frac{\partial z}{\partial t} + v(x)h\frac{\partial z}{\partial x} + \frac{\sigma^2(x)}{2}h\frac{\partial^2 z}{\partial x^2} + o(h)\right).$$

Now collect up terms of order h, cancel them and get the following partial differential equation for z:

$$\frac{\partial z}{\partial t} = -\mu xz + v(x)\frac{\partial z}{\partial x} + \frac{\sigma^2(x)}{2}\frac{\partial^2 z}{\partial x^2} \qquad \text{(A.14)}$$

as $h \to 0$.

Special Case. The diffusion approximation appropriate for modeling an exponentially growing (or dying) population puts

$$v(x) = x \cdot v,\ \sigma^2(x) = x \cdot \sigma^2. \qquad \text{(A.15)}$$

Now by analogy with the previous model argue that in order for the initiated colony to survive *all* must survive, so the form of the solution to (A.14) with coefficients of the form (A.15) should be

$$z(t;x) = e^{x\varphi(t)}.$$

Substituting this into (A.14), we get

$$x\frac{d\varphi}{dt}e^{x\varphi(t)} = -\mu x e^{x\varphi} + xv\varphi e^{x\varphi} + \frac{x\sigma^2}{2}\varphi^2 e^{x\varphi}. \qquad \text{(A.16)}$$

Cancellation of $xe^{x\varphi}$ yields a Ricatti equation once again:

$$\frac{d\varphi}{dt} = -\mu + v\varphi + \frac{\sigma^2}{2}\varphi^2 \qquad \text{(A.17)}$$

with $\varphi(0) = 1$.

Once again an explicit solution is available:

$$\varphi(t) = \frac{\xi_1(1-\xi_2) - \xi_2(1-\xi_1)\exp\left\{\left(\sqrt{\upsilon^2 + 2\mu\sigma^2}\right)t\right\}}{(1-\xi_2) - (1-\xi_1)\exp\left\{\left(\sqrt{\upsilon^2 + 2\mu\sigma^2}\right)t\right\}} \qquad \text{(A.18)}$$

where

$$\xi_{1,2} = -\frac{\upsilon}{\sigma^2} \pm \frac{1}{\sigma^2}\sqrt{\upsilon^2 + 2\mu\sigma^2}. \qquad \text{(A.19)}$$

Thus, if at time 0 there is one initiated cell, then the probability that this cell or one of its descendants is not promoted by time t is

$$P\{T > t\} = \exp\{\varphi(t)\} \qquad \text{(A.20)}$$

where $\varphi(t)$ is given by (A.18). This expression is to be compared to the birth-and-death model expression (A 4). Note that as $t \to \infty$, $P\{T > t\} \to \exp\{\xi_2\}$, which is always positive. A further argument similar to that in (A.7) is needed to obtain the distribution of time until a normal cell is initiated and promoted. Further discussion of the relationship between these models will be given elsewhere.

APPENDIX B. MODEL FITTING METHODS AND QUANTIFYING BIOASSAY SURVIVAL

1. PRELIMINARY STATISTICAL MODELS AND METHODS FOR ANALYZING BIOASSAY DATA

Suppose N organisms (for example fish) are used in an experiment. Groups of these organisms may be exposed to different treatments. Let $\mathbf{T}_i$ be the random time until organism i develops a particular symptom, e.g., cystic degeneration. Let $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{ip})$ be covariates which (possibly) influence $\mathbf{T}_i$; the $\mathbf{X}_i$ could be levels of substances having possible toxic effects to which the organisms are exposed. Let $G(t; \mathbf{x}_i) = P\{\mathbf{T}_i \le t \mid \mathbf{X}_i = \mathbf{x}_i\}$. We will assume that the organisms develop symptoms independently of each other. In this initial model, the symptom is either present or not.

Suppose that n_k organisms are sacrificed at time t_k with $t_1 < t_2 < \ldots < t_K$. We will label the organisms so that organisms 1 through n_1 are sacrificed at time t_1; organisms $n_1+1, \ldots, n_1+n_2$ are sacrificed at time t_2; etc. Let $s_i = 1$ if organism i exhibits the symptom when it is examined. Under the assumption of independence, the likelihood function is

$$L = \prod_{k=1}^{K} \prod_{i=1}^{n_k} s_{n_{k-1}+i} G\left(t_k; \mathbf{x}_{n_{k-1}+i}\right) + \left(1 - s_{n_{k-1}+i}\right) \bar{G}\left(t_k; \mathbf{x}_{n_{k-1}+i}\right) \qquad \text{(B.1)}$$

where $n_0 = 0$ and $\bar{G}(t; \mathbf{x}) = 1 - G(t; \mathbf{x})$. The likelihood functions form the basis for estimation of parameters in the distributions that model survival times, i.e. G.

Example (Simple Binomial Model). If there are no covariates, then (B.1) becomes

$$L = \prod_{k=1}^{K} \binom{n_k}{f_k} G(t_k)^{f_k} \bar{G}(t_k)^{n_k - f_k} \tag{B.2}$$

where f_k is the number of the n_k organisms exhibiting the symptom.

A procedure to estimate the parameters of the distribution G for the simple binomial model is as follows.

2. MAXIMUM LIKELIHOOD ESTIMATION IN THE SIMPLE BINOMIAL MODEL

(a) Likelihood and Parameter Estimation Formulas

Assume the distribution of the time to appearance of a symptom, G, is a function of the parameters α_j, $j = 1, \ldots, J$. In this section we discuss maximum likelihood estimation of α_j for the simple binomial model. Presumably the n_k subjects examined at time t_k, $k = 1, 2, \ldots, K$ have all been subjected to a common dosage of a potential toxin. The purpose of the present analysis is to predict survival probabilities as they depend on such dosage. The log-likelihood function for the simple binomial model is

$$l = \sum_{k=1}^{K} \ln\binom{n_k}{f_k} + f_k \ln G(t_k;\alpha) + (n_k - f_k)\ln \bar{G}(t_k;\alpha) \tag{B.3}$$

where $\alpha = (\alpha_1, \ldots, \alpha_J)$. Differentiating, we obtain

$$\frac{\partial}{\partial \alpha_j} l = \sum_{k=1}^{K} \frac{f_k}{G(t_k;\alpha)} \frac{\partial}{\partial \alpha_j} G(t_k;\alpha) + \frac{(n_k - f_k)}{\bar{G}(t_k;\alpha)} \left[-\frac{\partial}{\partial \alpha_j} G(t_k;\alpha) \right]$$

$$= \sum_{k=1}^{K} \frac{f_k[1-G(t_k;\alpha)]-(n_k-f_k)G(t_k;\alpha)}{G(t_k;\alpha)\bar{G}(t_k;\alpha)}\left[\frac{\partial}{\partial\alpha_j}G(t_k;\alpha)\right]$$

$$= \sum_{k=1}^{K}\left[\frac{f_k-n_kG(t_k;\alpha)}{G(t_k;\alpha)\bar{G}(t_k;\alpha)}\right]\frac{\partial}{\partial\alpha_j}G(t_k;\alpha). \qquad \text{(B.4)}$$

Since $E[f_k] = n_kG(t_k;\alpha)$

$$E\left[\frac{\partial^2}{\partial\alpha_j\partial\alpha_m}l\right] = -\sum_{k=1}^{K} n_k \frac{\frac{\partial}{\partial\alpha_j}G(t_k;\alpha)\frac{\partial}{\partial\alpha_m}G(t_k;\alpha)}{G(t_k;\alpha)\bar{G}(t_k;\alpha)}. \qquad \text{(B.5)}$$

Thus a Newton procedure for finding the maximum likelihood estimates of $\{\alpha_j; j = 1, \ldots, J\}$ would iteratively solve the system of linear equations

$$0 = \frac{\partial}{\partial\alpha_j}l(\alpha^0) + \sum_{m=1}^{P} E\left[\frac{\partial^2}{\partial\alpha_j\partial\alpha_m}l\right](\alpha_l - \alpha_l^0) \qquad \text{(B.6)}$$

where $\alpha^0 = \{\alpha_1^0, \ldots, \alpha_J^0\}$. Such iterative procedures can be programmed for a digital computer, and the resulting parameter values can be used to compute predictions for survival probabilities, or risk, as the latter depend upon the parameters of such models as described in Appendix A. An illustration is given in Appendix C.

APPENDIX C. A PRELIMINARY ANALYSIS OF DATA

We present and discuss a specific data analysis and uncertainty assessment utilizing the methodology described above. Note that this analysis is less comprehensive than that potentially possible.

The data analyzed are taken from the U.S. Army Biomedical Research and Development Laboratory Study N: Utilization of Fish to Evaluate Carcinogenic Potential of Tricholoethylene Contamined Groundwater—Pathology Report, Vol. 1 of 3. The report presents results of a "histopathologic examination of tissues from fish of the species Oryzias latipes (Japanese medaka) which was performed to evaluate the carcinogenic potential of trichloroethylene (TCE) in groundwater. Fish exposed to groundwater with various concentrations of TCE were either pre-treated or not pre-treated with 10 mg/l diethylnitrosamine (DEN) in water for 48 hours at 17 days post hatch. Exposures were begun in the biomonitoring trailer on the sixth day after treatment with DEN." After three months of exposure 25 fish in each treatment group were sacrificed (interim sacrifice). The remaining fish in each of the treatment regimes were divided into two groups "one group of fish, designated the chronic fish, continued to be exposed to various concentrations of TCE in the water for an additional three months." The chronic fish were sacrificed at 6 months into the study. It is the data on these fish that we will consider.

The data we consider are the occurrence or nonoccurrence of cystic degeneration (CD) in the chronic fish in each treatment group. The data are presented in Table 1.

TABLE 1.

				# fish with symptom/# fish killed Sacrifice Time	
Group	DEN	no DEN	%TCE	3 months	6 months
1		X	0	6/25	4/15
3		X	25	4/25	5/13
5		X	50	2/25	4/14
7		X	100	3/25	3/14
2	X		0	11/25	6/12
4	X		25	4/25	8/13
6	X		50	6/25	5/12
8	X		100	7/25	3/8

The statistical model used is the simple binomial model of Appendix B with exponential distribution

$$G(t) = 1-\exp\{-e^{\gamma}t\}.$$

The parameter $-\gamma$ is the natural logarithm of mean time to exhibit symptoms. The Newton procedure described in Appendix B was used to estimate the parameter γ numerically for the data in each group. Each group contains two data points, one for three months and one for six months. The maximum likelihood estimates $\hat{\gamma}$ appear in Table 2.

TABLE 2. MAXIMUM LIKELIHOOD ESTIMATES OF MEAN TIME TO EXHIBIT CD

Group	DEN	no DEN	%TCE	$-\hat{\gamma}$	$e^{-\hat{\gamma}}$ = Estimated Mean Time to CD
1		X	0	2.66	14.3
3		X	25	2.68	14.5
5		X	50	3.32	23.8
7		X	100	3.28	24.4
2	X		0	1.85	6.4
4	X		25	2.31	10.0
6	X		50	2.40	11.0
8	X		100	2.32	10.2

The last column of Table 2 shows that the estimates of the mean time to CD for those groups in which the fish were pretreated with DEN are always considerably smaller than those for fish that were not pretreated. The estimates are based on use of a very simple model, the exponential. Note, however, that the two-stage clonal expansion model of survival probability described in Appendix A has approximate exponential behavior in the right tail, so choice of the exponential for data analysis is not inconsistent with that theory.

There has been no opportunity to assess possible between-fish variability in susceptibility to incur different possible toxin effects. This is for the future.

VARIABILITY OR UNCERTAINTY ASSESSMENT

In order to assess uncertainty in the estimate of γ and the mean time to CD appearance the technique of bootstrapping was used, cf. Efron and Tibshirani [1986]. This is a re-sampling methodology that is of wide and useful application. For each group treatment, experimental data was simulated using the estimated model using the γ-estimate in Table 2; that is, for each bootstrap replication for each group of fish, two independent binomial random variables were simulated; one with 25 trials and probability of success $1-\exp\{-3e^{-\gamma}\}$, and the other having probability of success $1-\exp\{-6e^{-\gamma}\}$ with the number of trials equal to the number of fish in the chronic population sacrificed at six months. Any fish that have presumably died (or vanished) were simply ignored; note that these *could* have been removed because they very early reached a dose-related fatal endpoint, and hence the treatment of these is a potential source of bias. These simulated data were then used to obtain one bootstrap estimate of γ. For each treatment group 500 independent bootstrap estimates of γ were obtained.

Figure 1 presents **boxplots** of the bootstrap estimates of $-\gamma$ for each treatment group; recall that $-\gamma$ corresponds to the logarithm of the mean time to symptom development.

The following is taken from the documentation of GRAFSTAT, a developmental product of IBM which the Naval Postgraduate School is using under a test agreement with IBM. "The box portion of the plot extends from the lower quartile of the sample to the upper quartile. (The lower quartile is the point for which one quarter of the sample lies below and three quarters above. The upper quartile is analogous.) The line across the center of the box marks the median. The circle in the box represents the mean.

The distance from the lower to the upper quartile is called the **interquartile distance**, and it will be represented by Q. The points at the ends of the two lines (called **whiskers**) are the smallest and largest points, respectively, within $1.5Q$ of the quantiles. The points beyond the whiskers are outlying values."

The boxplots for groups 2, 4, 6 and 8 (those groups pretreated with DEN) generally lie below those for groups 1, 3, 5, and 7 respectively (those groups not pretreated with DEN). Thus treatment with DEN tends to shorten the mean time to symptom development, even accounting for sampling errors in the estimates. The bootstrapping results tend to strengthen inference concerning the effect of DEN treatment, and they provide perspective on the sensitivity of the experimental procedures, e.g., how well dosage effects can be distinguished. Comparison of the width of the boxes of groups 2, 4, 6, and 8 with those of 1, 3, 5, and 7 suggests that pretreatment with DEN tends to decrease the amount of variability of the γ-*estimates* (as well as reducing those estimates). The variability of the estimate of γ precludes any strong conclusions concerning the effect of the different dosages of TCE on corresponding mean times to symptom

development. The design of the experiment needs to be modified to increase the experiment's sensitivity if higher sensitivity is required.

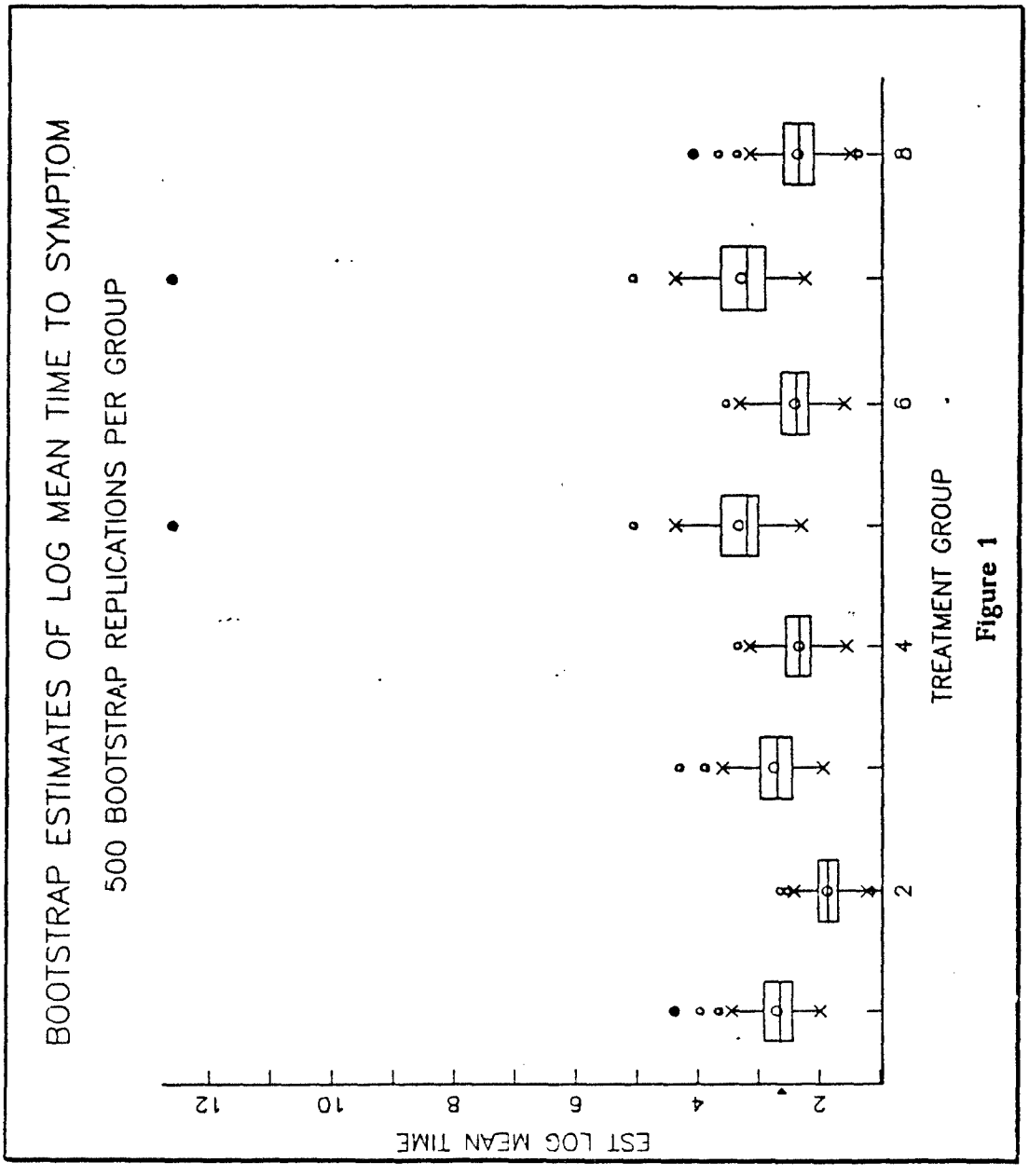


Figure 1

BIBLIOGRAPHY AND REFERENCES

Efron, B. and Tibshirani, R., "Bootstrap Methods for Standard Errors, Confidence Intervals, and other Measures of Statistical Accuracy," *Statistical Science*, Vol. 1, (1986), pp. 54-77.

Gaver, D. P. and Jacobs, P. A., *Susceptibility and Exposure Variation in the Two-stage Clonal Expansion Model*, in progress.

Harris, C. C., "Interindividual Variation in Human Chemical Carcinogenesis: Implications for Risk Assessment," *Scientific Issues in Quantitative Cancer Risk Assessment* (Abbreviate SIQCRA) ed. by S. H. Moolgavkar. Birkhauser. Boston, 1990. pp. 235-251.

Karlin, S., and H. M. Tayler, *A Second Course in Stochastic Processes*, Academic Press, New York, 1981.

Kopp-Schneider, A., C. J. Portier and F. Rippman, "The Application of a Multistage Model that Incorporates DNA Damage and Repair to the Analysis of Initiation/Promotion Experiments," *Math. Biosciences*, Vol. 105, 1991, pp. 139-166.

McCullagh, P., and J. A. Nelder, *Generalized Linear Models*, Chapman and Hall, New York, 1983.

Moolgavkar, S. H., and D. J. Venzon, "Two-event Models for Carcinogenesis: Incidence Curves for Childhood and Adult Tumors," *Math. Biosciences*, Vol. 47, 1979, pp. 55-77.

Moolgavkar, S. H., "Model for Human Carcinogenesis: Action of Environmental Agents," *Environmental Health Perspectives*, Vol. 50, 1973, pp. 285-291.

Moolgavkar, S. H., N. E. Day and R. G. Stevens, "Two-stage Model for Carcinogenesis: Epidemiology of Breast Cancer in Females," *JNCI*, Vol. 65, 1983, pp. 559-569.

Portier, C. J., "Two-stage Models of Tumor Incidence for Historical Control Animals in the National Toxicology Program's Carcinogenicity Experiment," *J. of Toxicology and Environmental Health*, Vol. 27, 1989, pp. 21-45.

Portier, C. J., "Utilizing Biologically Based Models to Estimate Carcinogenic Risk," *SIQCRA*, pp. 252-266.