ORNL/CSD-123
Distribution Category UC-32

MAXIMUM LIKELIHOOD ESTIMATION FOR CYTOGENETIC DOSE-RESPONSE CURVES

ORNL/CSD--123

DE84 001582

E. L. Frome

Mathematics and Statistics Research

R. J. DuFrain

Oak Ridge Associated Universities

Date Published - October 1983

COMPUTER SCIENCES
at
Oak Ridge National Laboratory
Post Office Box Y
Oak Ridge, Tennessee 37830

Research jointly sponsored by the
Office of Health and Environmental Research,
U. S. Department of Energy; and
Oak Ridge Associated University
(contract DE-AC05-76OR00033)

UNION CARBIDE CORPORATION, NUCLEAR DIVISION
operating the
• Oak Ridge Gaseous Diffusion Plant • Oak Ridge National Laboratory
• Oak Ridge Y-12 Plant • Paducah Gaseous Diffusion Plant
under Contract No. W-7405-eng-26
for the
DEPARTMENT OF ENERGY

CONTENTS

LIST OF TABLES

MAXIMUM LIKELIHOOD ESTIMATION FOR CYTOGENETIC DOSE-RESPONSE CURVES

E. L. Frome and R. J. DuFrain

ABSTRACT

In vitro dose-response curves are used to describe the relation between the yield of dicentric chromosome aberrations and radiation dose for human lymphocytes. The dicentric yields follow the Poisson distribution, and the expected yield depends on both the magnitude and the temporal distribution of the dose for low LET radiation. A general dose-response model that describes this relation has been obtained by Kellerer and Rossi using the theory of dual radiation action. The yield of elementary lesions is $\kappa[\gamma d+g(t,\tau)d^2]$, where t is the time and d is dose. The coefficient of the d^2 term is determined by the recovery function and the temporal mode of irradiation. Two special cases of practical interest are split-dose and continuous exposure experiments, and the resulting models are intrinsically nonlinear in the parameters. A general purpose maximum likelihood estimation procedure is described and illustrated with numerical examples from both experimental designs. Poisson regression analysis is used for estimation, hypothesis testing, and regression diagnostics. Results are discussed in the context of exposure assessment procedures for both acute and chronic human radiation exposure.

1. INTRODUCTION

In recent years there has been considerable interest in evaluating the influence of the magnitude and temporal distribution of low linear energy transfer (LET) radiation on biological systems. An extensive review of studies on a wide spectrum of species and experimental systems is given in NCRP Report No. 64--Influence of Dose and its Distribution in Time on Dose-Response Relationships for Low-LET Radiations (1980). The linear-quadratic (LQ) model

$$\lambda(d) = \alpha d + \beta d^2 \qquad (1.1)$$

is used extensively throughout NCRP64 to describe the effect of absorbed dose d on a specific biologic endpoint. The LQ model and its more general form (1.2) are also discussed in the latest report of the Committee on the Biological Effects of Ionizing Radiations of the National Academy of Science (BEIR III, 1980 Chap. 2). It is pointed out that the LQ model is a convenient empirical model for complicated endpoints in complex systems. For "simple" cellular systems the LQ model has been extensively used in the evaluation of radiobiologic data. In the discussion that follows we shall consider studies which focus on specific lesions in the chromosomes of somatic cells as the end point of interest. Most of the early research on the quantitative aspects of the effects of ionizing radiation on specific chromosome aberrations utilized plant cells (see Savage, 1975 for a recent review). Starting in the 1960's and continuing on to the present this line of research has shifted more to the use of animal cells. Most

recent work in human cytogentic dosimetry utilizes cultured peripheral blood lymphocytes to quantitatively assess the effect of low LET radiation on chromosome damage. This approach provides an effective method for the evaluation of one type of radiation damage in man. Numerous studies have demonstrated that chromosome alterations induced in lymphocytes after *in vitro* exposure to low LET radiation are both qualitatively and quantitatively similar to alterations observed after *in vivo* exposure. Thus it is assumed that data obtained with carefully controlled *in vitro* irradiation of human lymphocytes will accurately reflect the effect of dose magnitude and its temporal distribution on exposed persons. This provides a basis for the indirect evaluation of the effect of both acute and chronic human radiation exposure. Cytogenetic methods are currently used to provide dosimetry estimates for radiation accident management (see DuFrain *et al*, 1980, Frome and DuFrain, 1978), and it has been proposed that they be used for the indirect assessment of the long term biologic effects of chronic exposure to radiation and other clastogens in human population--see Evans *et al* (1979), Savage, (1979), Holden, (1982).

In the next section we will describe a maximum likelihood estimation procedure that can be used to estimate the parameters from an *in vitro* experiment. We assume that (i) the dependent variable y (the number of chromosome aberrations) follows the Poisson distribution, and (ii) that a regression function that describes the relation between y and the radiation exposure is specified. The role of the Poisson distribution in describing the dispersion of dicentric chromosome aberrations has been discussed by Edwards, Lloyd, and Purrott

(1979) and Merkle (1981). The index of dispersion can be used as a monitoring test for Poisson variation--see Fisher (1950) and Frome (1982)--and Frome, Kutner, and Beauchamp (1973) have discussed testing for heterogenity of variance and "lack of fit" in a regression context. Two examples are presented to illustrate both linear and nonlinear analysis using both empirically and theoretically derived models. In the first example we present results that were obtained using a "linear models" approach to evaluate the effect of dose and dose-rate on aberration yield. This initial analysis is straightforward and was designed to test the hypothesis that the coefficient of the d^2 term in (1) "depends" on dose-rate. Although this initial analysis is technically correct we were led to reject this approach as being both inappropriate and misleading on biologic grounds. We then propose a more appropriate analysis that utilizes a nonlinear model that is derived from the theory of dual radiation action (DRA) described by Kellerer and Rossi (1972). To emphasize the importance of the DRA theory a second example is presented using data obtained from a dose-fractionation procedure which leads to an appropriate dose response model under the DRA theory.

The DRA theory proposed by Kellerer and Rossi (1972) utilizes concepts from microdosimetry to provide a quantitative characterization of the effect of various temporal distributions of absorbed dose on the production of chromosome aberrations (CAs). It is postulated that elementary lesions are produced at a rate that is proportional to the square of the local energy concentration produced by charged particles in certain "critical sites". The form of the

dose-effect model that is appropriate here (see Kellerer and Rossi, 1972, Section 5.4) is

$$\lambda(d,t) = \kappa[\gamma d+g(t,\tau)d^2], \qquad (1.2)$$

where d denotes dose, t is time, and $\lambda(d,t)$ is the yield of elementary lesions. The parameter κ is a biophysical proportionality constant that reflects the target sensitivity for the biologic system (lymphocyte). The parameter γ depends on the radiation quality and can be related to the specific energy produced in a critical site by a single ionization. The linear term in (1.2) represents the effect due to intratrack interactions and the quadratic term represents the effect of intertrack interaction. The coefficient of the d^2 term is referred to as the 'reduction factor', and assuming an exponential recovery process

$$g(t,\tau) = \int_0^t e^{-s/\tau}\, h(s)ds \ ,$$

where h(t) describes the distribution of the time intervals t between dose increments for a given temporal mode of irradiation. For continuous irradiation of duration t one obtains (see Example 1)

$$g(t,\tau) = \frac{2\tau}{t} - \frac{2\tau^2}{t^2}(1-e^{-t/\tau}) \ . \qquad (1.3)$$

For a dose d given in two fractions separated by time t the reduction factor is (see Example 2)

$$g(t,\tau) = 1-2f(1-f)(1-e^{-t/\tau}) \ , \qquad (1.4)$$

where $f = d_1/d$, and d_1 is the first dose. Substitution of (1.3) and (1.4) into (1.2) gives the appropriate dose-response curve for the

continuous exposure and split dose experiments, respectively. The resulting models are intrinsically nonlinear in parameters and the appropriate statistical analysis is based on the general maximum likelihood estimation procedure described in the next section. Note that as $t \to 0$ in both (1.3) and (1.4) $g(t,\tau) \to 1$, and $\lambda(d) = \kappa[\gamma d + d^2]$ which is equivalent to the LQ model (1.1), for the limiting acute exposure situation. The parameterization in (1.1) has traditionally been used as a matter of computational convenience, and consequently the estimates of α and β can be viewed as 'computational artifacts'. Note that for the continuous exposure, split dose experiments, and acute exposure experiments, the parameters of interest are the same, i.e. κ, γ, and τ. In the acute exposure experiments one assumes that $t \ll \tau$ so that $g(t,\tau) \equiv 1$ for all values of d, and τ cannot be estimated.

2. MAXIMUM LIKELIHOOD ESTIMATION

Let y_i denote the number of dicentric CAs observed at the ith set of experimental conditions, i.e. dose d_i and time t_i for $i = 1, \ldots, n$. The y_i's are assumed to be independent and to follow the Poisson distribution with expectation

$$\mu_i = c_i \lambda(X_i, \underset{\sim}{\beta}) ,$$

where c_i denotes the total number of cells scored (in units of 100 cells) and $\lambda(X_i, \underset{\sim}{\beta})$ denotes the average yield of CAs per hundred cells scored. The regression function $\lambda(X, \underset{\sim}{\beta})$ describes the relation between the expected CA yield, the ith set of predictor variables,

$X_i = (x_{i1}, x_{i2}, \ldots, x_{im})$ and the p-dimensional vector of unknown parameters $\underset{\sim}{\beta}$. The kernel of the log likelihood function of $\underset{\sim}{\beta}$ is

$$L(\underset{\sim}{\beta}) = \sum_{i=1}^{n} \{y_i \log [c_i \lambda(X_i, \underset{\sim}{\beta})] - c_i \lambda(X_i, \underset{\sim}{\beta})\} \qquad (2.1)$$

The maximum likelihood (ML) estimate $\hat{\underset{\sim}{\beta}}$ is a root of the likelihood equations

$$\frac{\partial L(\underset{\sim}{\beta})}{\partial \beta_k} = \sum_{i=1}^{n} \left\{ \frac{\partial \lambda(X_i, \underset{\sim}{\beta})}{\partial \beta_k} \left[\frac{y_i}{\lambda(X_i, \underset{\sim}{\beta})} - c_i \right] \right\}, \quad k = 1, \ldots, p . \qquad (2.2)$$

Since these equations are generally nonlinear with respect to the unknown parameters, the method of scoring is used to develop an iterative procedure to find a root of (2.2). A convenient computational approach is obtained by using iteratively reweighted least squares (IPLS). Let $\bar{y}_i = y_i/c_i$ denote the average CA yield per 100 cells scored, and consider the following weighted sum of squares

$$S(\underset{\sim}{\beta}) = \sum_i w_i [\bar{y}_i - \lambda(X_i, \underset{\sim}{\beta})]^2 , \qquad (2.3)$$

where w_i denotes a weight that is inversely proportional to the variance of $\bar{y}_i$. Since $\lambda(X_i, \underset{\sim}{\beta})$ is, in general, nonlinear in the parameters an iterative procedure is required to obtain an estimate of $\underset{\sim}{\beta}$. On iteration k+1 we replace $\lambda(X_i, \underset{\sim}{\beta})$ with the linear terms in a Taylor series expansion about the current estimate $\underset{\sim}{\beta}^k$

$$\lambda(X_i, \underset{\sim}{\beta}) \approx \lambda(X_i, \underset{\sim}{\beta}^k) + P_i^k \underset{\sim}{\delta}^k , \qquad (2.4)$$

where P_i^k denotes the ith row of the n×p matrix of partial derivatives

$p_{ij} = \partial\lambda(X_i,\underset{\sim}{\beta})/\partial\beta_j$ evaluated at the current estimate $\underset{\sim}{\beta}^k$, and $\underset{\sim}{\delta}^k = (\delta_1^k,\ldots,\delta_p^k)'$. Using (2.4) in (2.3) with the appropriate 'Poisson weights', $w_i = c_i/\lambda(X_i,\underset{\sim}{\beta}^k)$, and the least squares principle we obtain the 'correction vector' $\underset{\sim}{\delta}^k$ by solving following system of p linear equations:

$$\underset{\sim}{A}(\underset{\sim}{\beta}^k)\underset{\sim}{\delta}^k = \underset{\sim}{G}(\underset{\sim}{\beta}^k) , \tag{2.5}$$

where $\underset{\sim}{A}(\underset{\sim}{\beta}^k) = \underset{\sim}{P}(\underset{\sim}{\beta}^k)' \underset{\sim}{W} \underset{\sim}{P}(\underset{\sim}{\beta}^k)$ is the information matrix, $\underset{\sim}{G}(\underset{\sim}{\beta}^k)$ is (2.2) evaluated at the current estimate $\underset{\sim}{\beta}^k$, and $\underset{\sim}{W}=\text{diag}(w_i)$. This leads to the revised estimate $\underset{\sim}{\beta}^{k+1} = \underset{\sim}{\beta}^k + \underset{\sim}{\delta}^k$, and the process continues until some convergence criteria are satisfied.

The ML estimate of $\underset{\sim}{\beta}$, the estimated parameter covariance matrix, and the deviance for each model are obtained using this IRLS algorithm --see Frome, Kutner, and Beauchamp (1973) for a detailed discussion of Poisson regression analysis and for examples of intrinsically non-linear models see Frome and Beauchamp (1968) and Frome (1983). This can be done using any statistical package that supports IRLS and the statistical package GLIM (Baker and Nelder, 1978) is particularly suited for this analysis for generalized linear models. A further advantage of the IRLS approach is that the basic 'building blocks' for regression diagnostics are easily obtained using the IRLS approach. The basic building blocks for Poisson regression diagnostics (see Frome, 1983) are some type of standardized residual and the diagonal terms h_i from the matrix

$$\underset{\sim}{H} = \underset{\sim}{W}^{1/2}\underset{\sim}{P}(\underset{\sim}{P}'\underset{\sim}{W}\underset{\sim}{P})^{-1}\underset{\sim}{P}'\underset{\sim}{W}^{1/2} , \tag{2.6}$$

where all quantities that depend on $\underset{\sim}{\beta}$ are evaluated at the ML estimate

$\hat{\underline{\beta}}$. For generalized linear models $\lambda(X_i\underline{\beta}) = g(\eta_i)$, where $\eta_i = \underline{X}_i\underline{\beta}$, and we obtain

$$\underline{H} = \underline{V}^{1/2}\underline{X}(\underline{X}'\underline{V}\underline{X})^{-1}\underline{X}'\underline{V}^{1/2} \quad ,$$

where $\underline{V}$ is diagonal with $v_i = c_i(\partial g_i/\partial \eta_i)^2/g_i$. For linear models this reduces to

$$\underline{H} = \underline{W}^{1/2}\underline{X}(\underline{X}'\underline{W}\underline{X})^{-1}\underline{X}'\underline{W}^{1/2} \quad ,$$

where the w_i's are the Poisson weights $c_i/\lambda(X_i,\hat{\underline{\beta}})$. Note that $\Sigma_i h_i = p$ and that large values of h_i (say greater than $2p/n$) indicate extreme points in the model space that may have a substantial influence on the fitted model. If u_i denotes a standardized residual for the ith observation, then the variance of u_i is approximately $1-h_i$, and adjusted residuals are given by $u_i/\sqrt{(1-h_i)}$ (see Haberman, Chap. 4, 1974). There are several possible choices for standardized residuals for Poisson data. The most obvious is $u_i = (y_i-\hat{\mu}_i)/\sqrt{\hat{\mu}_i}$, where $\hat{\mu}_i = c_i\lambda(X_i,\hat{\underline{\beta}})$, but this 'chi square type' residual behaves poorly when $\hat{\mu}_i$ is small. An alternative to this is the Freeman-Tukey (FT) residual $u_i=\sqrt{y_i}+\sqrt{(y_i+1)} - \sqrt{(4\hat{\mu}_i+1)}$ (Freeman and Tukey, 1950). The FT residuals appear to be the best choice for routine use in regression diagnostics. Velleman and Hoaglin (Chap. 9, 1981) have noted that when some of the $\hat{\mu}_i$ are small (less than 1) an adjusted degrees of freedom can be obtained by subtracting $\Sigma(1-\hat{\mu}_i)^2$ (where the sum is over values of i where $\hat{\mu}_i<1$) from the usual degrees of freedom n-p for the 'lack

of fit' statistic Σu_i^2. A third choice that would be preferred on theoretical grounds is the 'signed deviance' d_i as defined in (2.7). The signed deviance and various approaches to using the standardized residuals, and h_i's have been given by Pregibon (1981) in the context of logistic models for binomially distributed data.

In order to construct an ANOVA-like table for Poisson regression models we use the deviance $D(\underset{\sim}{y}, \hat{\underset{\sim}{\mu}}) = \Sigma_i d_i^2$ as a measure of residual variation, where

$$d_i = \text{sgn}(y_i - \mu_1)\{2[y_i \log(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)]\}^{1/2}, \qquad (2.7)$$

and $\hat{\mu}_i = C_i \lambda(X_i, \hat{\underset{\sim}{\beta}})$. This measure of residual variation was proposed by Nelder and Wedderburn (1972) and is minus twice the ratio of the log likelihood function of the model defined by $\lambda(X_i, \underset{\sim}{\beta})$ relative to the 'complete' model in which there is one parameter for each value of i. In the analysis that follows we fit a sequence of models and use the deviance as a measure of unexplained variation to construct an ANOVA-like table. The simplest (or minimal) model of interest in this situation is given by $\lambda(X_i, \beta) = \beta x_i$, where x_i is the radiation dose. The maximum likelihood estimate of β is $\hat{\beta} = \Sigma_i y_i / \Sigma_i x_i c_i$, and the deviance for the minimal model is $D[\underset{\sim}{y}, \hat{\underset{\sim}{\mu}}(1)] = -2 \Sigma_i y_i \log (y_i / c_i \hat{\beta} x_i)$. Following the approach described by Efron (1978) for the binomial distribution, we fit an increasing sequence of models for the explanatory vector $\underset{\sim}{\mu}$, say $\underset{\sim}{\mu} H_k$, $H_0 c H_k c \ldots$. The fitted vector for the kth model, say $\hat{\underset{\sim}{\mu}}(k)$, is that value of $\underset{\sim}{\mu} H_k$ that minimizes the deviance, i.e. the ML estimate restricted to H_k. Note that the

decrease in the deviance that is obtained when a less restrictive model is considered is a test statistic for the more restrictive hypothesis. The procedure is illustrated in the next section for a sequence of models that are linear in the parameters,

i.e. $\lambda(X_i, \underline{\beta}) = X_i \underline{\beta}$.

3. EXAMPLES

Example 3.1.1 Continuous Exposure Experiment

The data in Table 1 (Purrott and Reeder, 1976) were obtained from an experiment (using gamma radiation from a caesium-137 source) that was designed to investigate the effect of dose rate on CA yield. According to theoretical predictions from microdosimetry, a quadratic dose-response relation is predicted for low LET radiation, i.e. dicentric frequency is equal to $\alpha d + \beta d^2$, where d is radiation dose. From a biological point of view the two coefficients are thought of as corresponding to two different physical events. The linear term describes the induction of dicentrics by a single ionization or track, and the dose squared term which describes the induction of dicentrics by two different ionizations or tracks. Thus, the two break asymmetic exchange (dicentric) frequency is believed to be the result of these two phenomena, and is described by a second degree polynomial in dose. The validity of the quadratic model is predicated on the assumption that the absorbed dose is delivered to a 'critical site' in a short period of time, i.e. at a high dose rate.

The purpose of the study by Purrott and Reeder was to test the hypothesis that the effect of decreasing the dose-rate would be to decrease the contribution of the dose-squared term, without changing

Table 1

Dicentric CA Yields For CONTINUOUS Exposure Experiment

Dose Rate G/hr	Dose (Grays) 1.0 c	1.0 y	2.5 c	2.5 y	5.0 c	5.0 y
		H		H		
.1	4.78	25	3.28	52	2.10	100
.25	19.07	102	1.85	51	1.38	113
.5	22.58	149	3.42	100	1.60	144
1.0	23.29	160	3.10	100	1.20	106
1.5	12.38	75	2.78	107	.90	111
2.0	14.91	100	2.59	107	1.00	132
2.5	15.18	99	2.49	102	3.13	419
3.0	7.64	50	2.98	110	1.82	225
4.0	13.67	100	2.43	107	1.44	206

NOTE: y = number of dicentrics, c = cells scored (100s)

Source: Purrott and Reeder (1976)

the linear term. Model 4 (see Table 2) corresponds to the most general case in which both the linear and quadratic coefficients are allowed to vary with dose rate, i.e. $\lambda_{jk} = \alpha_j d_k + \beta_j d_k^2$, where j identifies the dose rate group. For each of the models in Table 2 the regression function $\lambda(X, \underset{\sim}{\beta})$ is linear in the parameters, and the procedure described in Section 2 was used to obtain the Poisson ANOVA.

Table 2

Poisson ANOVA Data in Table 1

Regression Model		Number of Parameters	Deviance	df
1	αd_i	1	1075.30	26
2	$\alpha d_i + \beta d_i^2$	2	228.00	25
3	$\alpha d_i + \beta_j d_i^2$	10	21.52	17
4	$\alpha_j d_i + \beta_j d_i^2$	18	11.10	9
5	Complete	27	0.0	0

A test statistic for the hypothesis $\beta_1 = \beta_2 = \ldots = \beta_9$ is obtained using the difference of the deviance $D[\underset{\sim}{y}, \hat{\underset{\sim}{\mu}}(2)] - D[\underset{\sim}{y}, \hat{\underset{\sim}{\mu}}(3)] = 206.48$. This test statistic has an asymptotic chi-squared distribution with 8 degrees of freedom (df), if the more restrictive hypothesis is true. Consequently, we reject the hypothesis that the coefficient of the quadratic term is independent of dose rate. An alternative approach is to test for 'lack of fit' of model 3. The deviance for this model is 21.52 with 17 df indicating that model 3 cannot be rejected.

3.1.2 Ad Hoc Model for Example 1

If the ML estimates of the quadratic coefficients obtained from model 3 are plotted against the log of the dose rate it appears that the $\hat{\beta}_j$s increase linearly with log dose rate, and this can be described by the following regression model

$$\lambda_{jk} = \alpha d_k + [(\theta_1 + \theta_2 \log_{10}(r_j)]d_k^2.$$

The ith row of the model matrix for this ad hoc model is $X_i = (d_i, d_i^2, d_i^2 \log_{10} r_j)$. The ML estimates and estimated standard errors for this model are given in Table 3. The value of the deviance for the model is 29.95 with 24 dF, indicating that this ad hoc model cannot be rejected for these data. This model provides a good description of effect of dose rate on dicentric yield, i.e., the

Table 3

Maximum Likelihood Estimates for Ad Hoc Model
for Dose-Response Curve Data in Table 1

Parameter	Estimate	Standard Deviation
α	2.86	.305
θ_1	3.80	.141
θ_2	2.26	.144

quadratic component increase with the log of dose rate, and the linear component is independent of dose rate.

3.3.3 Dual Radiation Action Model

The ad hoc model described in the previous section can be used as an empirical description of cytogenetic dose response curves for this experiment. The parameters in this model do not have a clear interpretation in terms of the quantitative effects of ionizing radiation. The DRA theory (see the Introduction) leads to the dose

effect model (1.2) and for a continuous exposure experiment the function $g(t,\tau)$--originally proposed by Lea (1955)--is given by (1.3). Using (1.3) in (1.2) we obtain (see Kellerer and Rossi, 1972, Section 5.4)

$$\lambda(X_i,\underline{\beta}) = \kappa\{\gamma d_i + \frac{2\tau}{t_i}\,[1 - \tau\{1 - \exp(-t_i/\tau)\}/t_i]\, d_i^2\}\,, \qquad (3.1)$$

where d is the absorbed dose and t is the duration of exposure at a constant dose rate. The parameters γ, κ, and τ can be related to the radiation quality, target sensitivity, and the recovery process (see the Introduction and Discussion).

The ML estimates of the parameters in (3.1) for the data in Table 1 were obtained using the IRLS procedure described in Section 2. Since the DRA model is nonlinear in the parameters, the partial derivatives of (3.1) with respect to the parameters must be supplied (see the Appendix). The ML estimates and their standard deviations are given in Table 4. The deviance for this model is 28.58 with

Table 4

ML Estimates for the DRA Model for the Data in Table 1.

Parameter	Estimate	Standard Deviation
κ	5.44	.208
γ	.269	.0677
τ	7.40	.857

24 df ($p = .236$) indicating that the DRA model cannot be rejected. The standardized residuals in Table 5(a) are used to identify

outlying observations, and in this example there is one large negative residual. The diagonal terms from the $\underset{\sim}{H}$ matrix (2.7) are given in Table 5(b). There are several large h values (greater than 2p/n=0.22) in column 3, and two of these are in the first two rows, i.e. the highest dose and the lowest exposure rates (see the Discussion).

Table 5

Regression Diagnostics for Data in Table 2 Using the Nonlinear Model (3.1)

(a) Standardized Residuals $u_i = (y_i - \hat{\mu}_i)/\hat{\mu}_i^{1/2}$

0.127	-0.929	1.35
-1.23	0.315	1.19
0.291	-0.627	-1.05
0.383	-0.563	-2.92
-0.927	0.914	-0.140
-0.111	1.48	0.247
-0.423	1.26	0.315
-0.293	0.144	-1.17
0.670	1.88	0.732

(b) Diagonal terms from the $\underset{\sim}{H}$ matrix (p/n=0.111)

0.056	0.164	0.406
0.143	0.038	0.239
0.155	0.036	0.157
0.161	0.035	0.080
0.086	0.037	0.062
0.105	0.038	0.075
0.107	0.039	0.251
0.054	0.049	0.154
0.097	0.043	0.132

3.2.1. Split-dose Experiment

Schmid, Bauchinger, and Mergenthaler (1976) undertook a study to investigate the "time-dependent" quadratic component of the LQ model using a split-dose technique. Two experiments were carried out using 250 KV X-rays for the *in vitro* exposure of human peripheral lymphocytes. The first experiment was carried out to determine the coefficients for the LQ model (see Table 6a). In the second experiment the lymphocytes were irradiated with a dose of 3.4 Grays split into two equal fractions separated by intervals of 50 minutes to 6 hours -- see Table 6b. They assume that the primary damage induced by the first dose fraction decrease at a constant rate, and obtain the follow expression for the "interval dependent" yield

$$\bar{y}_t = \frac{\beta d^2}{2} e^{-t/\tau} \quad .$$

The "interval dependent" yield is taken to be the observed yield at time t for d = 3.6G minus the observed yield at d = 1.7G with t = 0. Using the results from the DRA theory (see Kellerer and Rossi, 1972 Section 5) for a split dose experiment we see that $g(t,\tau)$ is given by (1.4). Using (1.4) in (1.2) with f = 1/2 we obtain

$$\lambda(X_i, \underset{\sim}{\beta}) = \kappa\{\gamma d_i + \frac{1}{2}[1+\exp(-t_i/\tau)]d_i^2\} \quad , \qquad (3.2)$$

where $X_i = (d_i, t_i)$, and $\underset{\sim}{\beta} = (\kappa, \gamma, \tau)'$. Since half the dose is given at t = 0 the coefficient of d^2 can be written $\frac{1}{2}\kappa d^2 + \frac{1}{2}\kappa d^2 e^{-t/\tau}$, i.e. where the second component is the identical to the expression for the interval dependent yield given by Schmid, Bauchinger and Mergenthaler (1976, equation 11). Consequently we can combine the data from Table

Table 6

a) Dicentric Yields For Acute Exposure Experiment (t=0 and c=1)

d-dose (Grays)								
.25	.50	1.0	1.5	2.0	2.5	3.0	3.5	4.0
3	5	9	30	37	54	74	77	128
1	4	12	27	41	57	70	84	123

b) Dicentric Yield for Split-Dose Experiment (d=3.4 Grays)*

Time Interval (t hours)	Cells Analysed (c 100s)	Dicentrics y
0	5	135*
0	6	540
.83	5	417
1.00	5	393
1.17	3	238
1.33	2	150
1.50	3	214
1.67	5	354
1.83	2	141
2.00	4	277
2.50	3	200
3.00	2	122
3.33	2	127
4.00	2	104
5.00	2	107
6.00	2	104

*Dose = 3.4 Grays for all except the first row where d=1.7G.

Source: Schmid, E., et al (1976).

5a and 5b and use (3.2) to obtain ML estimates of κ, γ, and τ as described in Section 2. Table 7 shows the ML estimates of κ and γ obtained using Experiment 1 data only, the ML estimates when experiments 1 and 2 are combined, and the ANOVA for the split dose

Table 7

Results for Split Dose Data in Table 6

a) ML Estimates

	κ	γ	τ
Acute Only	5.49	1.37	-
Acute & Split Dose	6.23	.88	2.15
(St. Deviation)	(.49)	(.28)	.42

b) Poisson ANOVA

Regression Model	df	Deviance
αd	33	162.2
$\alpha d + \beta d^2$	32	115.2
DRA (3.2)	31	18.45
Each (d,t)	9	2.6
Complete	0	0.0

experiment. A 'lack of fit' test for the DRA model is obtained from lines 3 and 4 of Table 7b and the value of the likelihood ratio test statistic is 15.8 with 22 d.f., indicating that the model cannot be rejected.

4. DISCUSSION

The results in Section 3 show how Poisson regression methods can be used in the analysis of cytogenetic dose-response curves. In our original analysis of the continuous exposure experiment data in Table 2 (see Frome and DuFrain, 1978) our objective was to show how to use linear model analysis to test the hypothesis of interest as specified by Purrott and Reeder (1976). In order to simplify the analysis only those data with three doses at each dose rate were included. There were six additional data points at the low dose rates (see Table 8),

Table 8

Additional Data for Continuous Exposure Experiment in Table 1

Dose d	Dose Rate G/hr	Cells Scored C(100s)	Dicentrics y
5.0	.15	2.04	157
2.5	.15	2.25	50
2.5	.05	5.40	100
1.0	.05	14.01	50
1.0	.05	5.74	25
1.0	.019	6.29	25

and these data were also excluded from our latter analysis using the DRA model (see Frome and DuFrain, 1982). This was done partially to ensure comparability with the earlier analysis and partly on

biological grounds since the stability of the unstimulated G_0 lymphocyte maintained in culture for long time intervals can be questioned. Theresults of fitting the ad hoc model, the DRA model, and a third model

$$\lambda(X_i, \underline{\beta}) = \beta_1 d + \beta_2 d^2 + \beta_3 (d^2 \log t)$$

are given in Table 9. When all of the data are included both of the empirical linear models provide better fits for the complete set of

Table 9

Values of the Deviance for Continuous Exposure Study

Regression Model	Table 1 n=27	Table 1+Table 8 n=33
$d+d^2+d^2\log t$	24.54	35.00
$d+d^2+d^2\log r$	29.95	41.96
DRA (eq. 3.1)	28.58	50.37

data. Both of these models can be rejected however, on biological grounds since they do not lead to reasonable results in the limiting situations of interest, i.e. as $t \to 0$ and as $t \to \infty$. Much of the lack of fit for the DRA model comes from the data at the lowest dose rates, and as we noted earlier there are reasons to question these data. The second experiment provides further support for the DRA theory since it provides an appropriate mathematical model for both the split dose and continuous exposure experiments.

It is apparent that both of these studies were motivated by the DRA theory, and consequently we feel that the use of the appropriate model for these and related experiments is of prime importance in furthering our understanding of the effects of the temporal distribution of low LET radiation on the yield of dicentric aberrations. Under similar experimental conditions the results from both continuous exposure and split dose experiment should be comparable for the human lymphocyte data. The parameter γ is related to radiation quality but the values of κ and τ should be the same for normal human lymphocytes. We propose that future research efforts should focus on experiments that are designed to test for lack of fit of model, with particular emphasis on the time-dependent component. It is apparent that a more general form of the model could be obtained, for example, by assuming a more general form for the repair process. The purpose of this paper is to describe the ML estimation, hypothesis testing, and regression diagnostic procedures that can be used for any appropriate dose-response model for CAs that follow the Poisson distribution. We are currently considering the use of resistant regression techniques to reduce the influence of atypical data, so that the DRA model can be fitted to all available human lymphocyte data, from both split dose and continuous exposure experiments without the need for arbitrary decisions concerning potential anomalies.

ACKNOWLEDGMENTS

This research was supported by the Office of Health and Environmental Research, U. S. Department of Energy under contract

W-7405-eng-26 with Union Carbide Corporation, and under contract DE-AC05-76OR00033 with Oak Ridge Associated Universities. The authors wish to thank Drs. J. J. Beauchamp, D. G. Gosslee, L. G. Littlefield, C. C. Lushbaugh, M. D. Morris, and J. Preston for reviewing an earlier version of this manuscript. The secretarial support provided by the MSRD staff is also appreciated, and Donna Poole in particular deserves special recognition for her effort in preparing this report.

APPENDIX

When the function $\lambda(X,\underset{\sim}{\beta})$ is intrinsically nonlinear in the parameters the IRLS procedure can be used to obtain ML estimates of β's (see Frome, Kutner, and Beauchamp, 1973). This requires the partial derivatives of $\lambda(X,\underset{\sim}{\beta})$ with respect to each of the β_is. As an example of the general procedure, consider the model for the split dose experiment--see equations (1.2) and (1.4) --

$$\lambda(d,t) = \kappa\{\gamma d + [1-2f(1-f)(1-e^{-t/\tau})]d^2\} .$$

To obtain ML estimates of the parameters using GLIM we wrote a GLIM macro named FITNL (see Figure 1). The partial derivatives of $\lambda(d,t)$ with respect to κ, γ, and τ are called P1, P2, and P3 in the macro KRSD. Additional macros that are required for this nonlinear model are also listed in Figure 1 and the reader is referred to the GLIM manual (Nelder and Baker, 1977, Chap. 18) for further details.

Identical results can also be obtained using the FORTRAN program PREG (Frome, 1981), the SAS (Goodnight and Sall, 1982) procedure NLIN, or the BMDP (1979) program P3R. Each of these approaches requires the partial derivatives, initial estimates of the parameters, and some convergence criteria. A listing of a GLIM program and detailed computational results for the split dose data in the example can be obtained from the authors. Note that the same GLIM procedures, FITNL in Figure 1, can be easily modified for alternative nonlinear models. Additional examples of nonlinear models are given in Frome and Beauchamp (1968) and Frome (1983). This is done by (i) replacing the macro KRSD with a new macro with the appropriate regression function

```
!      MAXIMUM LIKELIHOOD ESTIMATION FOR POISSON REGRESSION
!                            USING
!      IRLS ALGORITHM FOR NONLINEAR MODEL IN GLIM ( VER 3.12 )
!
$SUBFILE KRSD  !     KRSD.RSS  12 FEB 1983
!
!    MACRO FITNL REQUIRES THE FOLLOWING INPUT DATA:
!     C- NUMBER OF CELLS SCORED
!     L- CA YIELD       T- TIME BETWEEN FRACTIONS
!     D- TOTAL DOSE     F1- FIRST FRACTION
!
!   ON EXIT THE FOLLOWING VECTORS ARE AVAILABLE:
!     LHAT- ESTIMATED CAS PER 100 CELLS SCORED
!     CS-   STANDARDIZED RESIDUAL ( CHI-SQ TYPE )
!     H-    DIAGIONAL TERMS FROM H MATRIX
!
$MAC FITNL  !    FIT NONLINEAR MODEL DEFINED BY MAC KRSD
$DATA 3 B$READ 6 0.9 2.2    !  STARTING VALUES FOR BETA
$CA %K= 12    : %C=0.0001 ! SET CONVERGENCE
 $WEIGHT W  $YVAR Z  $OWN R1 R2 R3 R4 !
$WHILE %K KRSD $DISP ESEXTR %VL $CA H=%VL*W  !
$CA CS = (Y-LHAT*C)/%SQRT(LHAT*C) $
!       DELETE WORK ARRAYS AND MACROS
$DEL %WT %FV %WV %LP %VA %VL %PE  Z DB F !
   P1 P2 P3 W CCHK R1 R2 R3 R4    $END !
!
$MACRO KRSD    !  KELLERER-ROSSI SPLIT DOSE MODEL
!    F IS THE REGRESSION FUNCTION
! P1 P2 P3 ARE THE PARTIAL DERIVATIVES-- SEE EQ 2.4
!    B= (KAPPA,GAMMA,TAU )
$CA F=1-2*F1*(1-F1)*(1-%EXP(-T/B(3)) ) !
$CA P1= B(2)*D + F*D**2 : P2= B(1)*D  !
$CA P3=2*T*B(1)*F1*(1-F1)*%EXP(-T/B(3)) !
        *(  D/B(3) )**2 !
$CA LHAT= B(1) * (  B(2)*D +  F*D**2 ) !
$CA W=C/LHAT : Z=L-LHAT : %LP=Z $SCA 1.0!
$FIT P1+P2+P3-%GM $EXTR %PE  $CA DB=%PE !
$CA B= B + DB $PR %K  '  ESTIMATES=' B  !
  $USE CCHK  $ !  CHECK FOR CONVERGENCE
$END
!     MACROS REQUIRED BY FITNL FOR POISSON REGRESSION
!                MACROS REQUIRED BY OWN
$M R1 $CA %FV=%LP$E $M R2 $CA %DR=1.0 $E $M R3 $CA %VA=1.0$E !
$M R4 $CA %DI= 2*( Y*%LOG(L/LHAT)-C*Z)/W $E !
$MAC CCHK !      CONVERGENCE CHECK FOR FITNL
$CA DB= %IF( %LE(DB,0),-DB,DB)/B !
$CA DB= %IF( %LE(DB,%C),0,1)  $CA %T= %CU(DB)!
$CA %K=%K-1 $CA %K= %IF( %LE(%T,0),0,%K) $E !
$RETURN
```

FIGURE 1

and partial derivatives, and (ii) providing initial estimates of $\hat{\underline{\beta}}$ in FITNL for the data and model being considered. Further note that the computational approach described here can be extended over situations where y is in the regular exponential family for general nonlinear models (see Charnes, Frome, and Yu, 1976). This requires the two modifications just described and appropriate changes in the weight vector $\underline{W}$ and the deviance in macro R4.

REFERENCES

Baker, R. J. and Nelder, J. A. (1978). Generalized Linear Interactive Modelling (GLIM), Release 3, Numerical Algorithms Group, Oxford.

Beir III (1980). The Effects on Populations of Exposure to Low Levels of Ionizing Radiation: 1980. National Academy Press, Washington, D.C.

Charnes, A., Frome, E. L., and Yu, P. L. (1976). The Equivalence of Generalized Least Squares and Maximum Likelihood Estimation in the Exponential Family. Journal of the American Statistical Association 71, 169-172.

DuFrain, R. J., Littlefield, L. G., Joiner, E. E. and Frome, E. L. (1980). In Vitro Human Cytogenetic Dose Response Systems. In: The Medical Basis for Radiation Accident Preparedness (Hubner, K. F. and Fry, S. S., eds.), Elsevier North-Holland, N. Y.

Efron, B. (1978). Regression and ANOVA with Zero-One Data: Measures of Residual Variation. J. Amer. Stat. Assoc. 73, 113-121.

Evans, H. J., Buckton, G. E., Hamilton, G. E., and Carothers, A. (1979). Radiation-Induced Chromosome Aberrations in Nuclear Dockyard Workers, Nature 277, 531-534.

Fisher, R. A. (1950). The Significance of Deviations from Expectation in a Poisson Series, Biometrics 6, 17-24.

Freeman, D. H., Jr. and Tukey, J. W. (1950). Transformations Related to the Angular and the Square Root. Annals of Mathematical Statistics 21, 607-611.

Frome, E. L. (1981). Poisson Regression Analysis. American Statistician 35, 262-263.

Frome, E. L. (1982). Fisher's Exact Variance Test for the Poisson Distribution. Appl. Statistics 31, 67-71.

Frome, E. L. (1983). The Analysis of Rates Using Poisson Regression Models. Biometrics (To appear).

Frome, E. L. and Beauchamp, J. J. (1968). Maximum Likelihood Estimation of Survival Curve Parameters. Biometrics 24, 595-605.

Frome, E. L. and DuFrain, R. J. (1978). The Analysis of Dose-Response Curves for Radiation Induced Chromosome Aberrations. Abstract. Biometrics 34, 731.

Frome, E. L. and DuFrain, R. J. (1978). Estimation of the Amount of Exposure to an Environmental Clastogen Using Human Cytogenetic Dosimetry. In: Proceedings of the Fourth DOE Statistical Symposium (Albuquerque, New Mexico), pp. 169-191.

Frome, E. L. and DuFrain, R. J. (1982). Analysis of Cytogenetic Dose-response Data Using a Model Derived from the Theory of Dual Radiation Action. Biometrics (abstract).

Frome, E. L., Kutner, M. H. and Beauchamp, J. J. (1973). Regression Analysis of Poisson-distributed Data. Journal of the American Statistical Association 68, 935-940.

Haberman, S. J. (1974). The Analysis of Frequency Data. U. of Chicago Press.

Holden, C. (1982). Looking at Genes i the Workplace. Science 217, 336-337.

Kellerer, A. M. and Rossi, H. I. (1972). The Theory of Dual Radiation Action. Curr. Topics Radiat. Res. 8, 85-158.

Lea, D. E. (1955). Action of Radiations on Living Cells. 2nd Edition. Cambridge Press.

Lloyd, D. C., Purrott, R. J., Dolphin, G. W., Bolton, D. and Edwards, A. A. (1975). The Relationship between Chromosome Aberrations and Low LET Radiation Dose to Human Lymphocytes. Int. J. Radiation Biology 28, 75-90.

Merkle, W. (1981). Poisson Goodness-of-fit Tests for Radiation-Induced Chromosome Aberrations. International Journal of Radiation Biology 40, 685-692.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized Linear Models. Journal of the Roy Statistical Society, Series A 135, 370-384.

Pregibon, D. (1981). Logistic Regression Diagnostics. The Annals of Statistics 9, 705-724.

Purrott, R. J. and Reeder, E. (1976). The Effect of Changes in Dose Rate on the Yield of Chromosome Aberrations in Human Lymphocytes Exposed to Gamma Radiation. utation Research 35, 437-444.

Savage, J. R. K. (1975). Radiation-Induced Chromosomal Aberrations in the Plant Tranescantia: Dose-Response Curves, Radiation Biology 15, 87-140.

Savage, J. R. K. (1979). Chromosomal Aberrations at Very Low Radiation Dose Rates, Nature 277, 512-513.

Schmid, E. Bauchinger, M. and Mergenthaler, W. (1976). Analysis of the Time Relationships for the Interaction of X-ray-induced Primary Breaks in the Formation of Dicentric Chromosomes. Int. J. Radiation Biol. 30, 339-346.

Velleman, P. F. and Hoaglin, D. C. (1981). Applications, Basics, and Computing of Exploratory Data Analysis. Duxbury Press, Boston.

ORNL/CSD-123
Distribution Category UC-32

INTERNAL DISTRIBUTION

1. Central Research Library
2. K-25 Plant Library
3. ORNL Patent Office
4. Y-12 Technical Library, Document Reference Section
5. Laboratory Records - RC
6-7. Laboratory Records Department
8. V. A. Alexiades
9. C. K. Bayne
10. J. J. Beauchamp
11. K. O. Bowman
12. J. A. Carpenter
13. H. P. Carter/CSD X-10 Library
14. M. V. Denson
15-19. E. L. Frome
20. R. E. Funderlic
21. G. A. Geist
22. D. G. Gosslee
23. L. J. Gray
24. M. T. Heath
25. T. L. Hebble
26. T. D. Jones
27. E. Leach
28. W. E. Lever
29. L. P. Lewis
30. T. J. Mitchell
31. M. D. Morris
32. D. L. Poole
33. R. J. Preston
34. T. S. Reed
35. D. C. Schlotzhauer
36. R. L. Schmoyer
37. C. A. Serbin
38. K. E. Shultz/ Biometrics Library
39. A. D. Solomon
40. J. S. Trent
41. V. R. R. Uppuluri
42. R. C. Ward
43. D. G. Wilson
44. T. Wright

EXTERNAL DISTRIBUTION

45. Dr. Peter Bloomfield, Department of Statistics, Princeton University, Princeton, New Jersey 08540

46. Dr. Brent Blumenstein, Department of Biostatistics, University of Washington, Seattle, Washington 98195

47. Dr. Norman Breslow, Department of Biostatistics, University of Washington, Seattle, Washington 98195

48. Dr. Elisabeth Cardis, Department of Biostatistics, University of Washington, Seattle, Washington 98195

49. Dr. Keewhan Choi, Centers for Disease Control, EPO, 1600 Clifton Road, NE, Atlanta, Georgia 30333

50-54. Dr. R. J. DuFrain, Medical Division, Oak Ridge Associated Universities, P. O. Box 117, Oak Ridge, Tennessee 37830

55. Dr. L. Edler, DKFZ, Institut für Dokumentation, Information und Statistik, Im Neuenheimer Feld 280, D-6900 Heidelberg 1, GERMANY

56. Dr. Shirley Fry, Medical Division, Oak Ridge Associated Universities, P. O. Box 117, Oak Ridge, Tennessee 37830

57. Dr. David Hall, Statistics, Systems Department, Pacific Northwest Laboratory, P. O. Box 999, Richland, Washington 99352

58. Dr. Elmer Hall, Biometry, Emory University, Atlanta, Georgia 30322

59. Dr. James Hanley, Epidemiology and Health, McGill University, 3775 University Street, Montreal, Quebec, CANADA H3A 2B4

60. Professor D. Herbert, Department of Radiology, College of Medicine, University of South Alabama, 240 Cancer Center/Clinical Building, Mobile, Alabama 36688

61. Professor Gary Koch, Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina 27514

62. Dr. D. Krewski, Environmental Health Center, Room 117, Tunney's Pasture, Ottawa, Ontario, CANADA K1A 0L2

63. Dr. Steven Lagakos, Department of Biostatistics, Harvard University, 677 Huntington Avenue, Boston, Massachusetts 02115

64. Dr. Charles Land, Radiation Studies Section, National Cancer Institute, LAN-3016, Bethesda, Maryland 20205

65. Dr. L. G. Littlefield, Medical Division, Oak Ridge Associated Universities, P. O. Box 117, Oak Ridge, Tennessee 37830

66. Dr. C. C. Lushbaugh, Medical Division, Oak Ridge, Associated Universities, P. O. Box 117, Oak Ridge, Tennessee 37830

67. Dr. Stella C. Machado, Radiation Studies Section, National Cancer Institute, LAN-3016, Bethesda, Maryland 20205

68. Dr. George P. McCabe, Jr., Department of Statistics, Purdue University, West Lafayette, Indiana 47907

69. Professor Peter McCullagh, Department of Mathematics, Imperial College, London, ENGLAND SW7 2AZ

70. Professor Carl Morris, Department of Mathematics, University of Texas, Austin, Texas 78712

71. Professor J. A. Nelder, Department of Statistics, Rothamsted Experimental Station, Harpenden, Herts, ENGLAND

72. Dr. Ronald Peierls, Applied Mathematics Department, Brookhaven National Laboratory, Upton, New York 11973

73. Dr. O. J. Pendleton, Accident Analysis Division, Texas Transportation Institute, College Station, Texas 77840

74. Dr. Daryl Pregibon, Room 2C-265, Bell Lab, Murray Hill, New Jersey 07974

75. Dr. Ross Prentice, Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, Washington 98104

76. Dr. Douglas S. Robson, Biometrics Unit, Cornell University, Ithaca, New York 14850

77. Professor Harald H. Rossi, Department of Radiology, Radiological Research Laboratory, Columbia University, New York, New York 10032

78. Dr. Daniel L. Solomon, Department of Statistics, North Carolina State University, P. O. Box 5457, Raleigh, North Carolina 27650

79. Dr. George Stapleton, E-201, EV-32, Germantown, U. S. Department of Energy, Washington, D. C. 20545

80. Professor Bruce W. Turnbull, 356 Upson Hall, Cornell University, Ithaca, New York 14853

81. Dr. Ray A. Waller, S-1, Statistics, Los Alamos National Laboratory, P. O. Box 1663, Los Alamos, New Mexico 87545

82. Dr. Jerome Wilson, Medical Division, Oak Ridge Associated Universities, P. O. Box 117, Oak Ridge, Tennessee 37830

83. Office of Assistant Manager for Energy Research and Development, Department of Energy, Oak Ridge Operations Office, Oak Ridge, Tennessee 37830

84-265. Given Distribution as shown in TIC-4500 under Mathematics and Computers Category