

A COMPUTER PROGRAM FOR THE GENERALIZED CHI-SQUARE ANALYSIS OF COMPETING RISKS GROUPED SURVIVAL DATA (CRISCAT)

William M. STANISH, G.Y.H. CHI, William D. JOHNSON *, Gary G. KOCH, J. Richard LANDIS ** and Shelley LIU-CHI

*Department of Biostatistics, University of North Carolina, Chapel Hill, NC27514; * Department of Biometry, Louisiana State University Medical Center, New Orleans, LA 70112 and ** Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA*

CRISCAT is a computer program for the analysis of grouped survival data with competing risks via weighted least squares methods. Competing risks adjustments are obtained from general matrix operations using many of the strategies employed in a previously developed program (GENCAT) for multivariate categorical data. CRISCAT computes survival rates at several time points for multiple causes of failure, where each rate is adjusted for other causes in the sense that failure due to these other causes has been eliminated as a risk. The program can generate functions of the adjusted survival rates, to which asymptotic regression models may be fit. CRISCAT yields test statistics for hypotheses involving either these functions or estimated model parameters. Thus, this computational algorithm links competing risks theory to linear models methods for contingency table analysis and provides a unified approach to estimation and hypothesis testing of functions involving competing risks adjusted rates.

Competing risks	Categorical data	Weighted least squares	Linear models	Computer program
Multivariate analysis	Survival analysis			

1. Introduction

Methods for the analysis of survival data have challenged those involved in the statistical interpretation of research data for many years. One class of problems which has received considerable attention pertains to the analysis of survival in the presence of competing risks. The competing risks situation arises when units under study are at risk of failure due to any of several causes but only one is observable since failure from a specific cause ends the unit's life. The researcher is frequently interested in characterizing the survival profile corresponding to a particular cause when it is the only risk of failure, but from a practical point of view he can only observe data in a competing risks environment. In this connection, the unadjusted or crude probabilities of survival are adjusted for competing risks to yield net probabilities.

Competing risks arise in the study of length of life of human beings and other biological systems, machines and their component parts, consumer products, and elsewhere. When survival times are classified into categories, the methods described in [1,2] are

appropriate under certain assumptions for analyzing competing risks survival data. Until recently, however, there has not been a convenient formulation for carrying out the analysis of competing risks adjusted data in terms of a general approach to model fitting, estimation, and hypothesis testing. This paper describes a computer program called CRISCAT which performs the analysis of competing risks categorical survival data via weighted least squares methods for contingency table analysis and provides a unified analysis in the spirit of traditional analysis of variance.

Section 2 contains a brief summary of the theoretical results which form the basis of the methodology employed by CRISCAT. A discussion of the theory is given in [4]. Section 3 describes the general structure of survival data in terms of contingency tables.

Section 4 develops the estimation procedure and computational algorithm. Basically, the competing risks adjustments are obtained from general matrix operations such as those used in the program GENCAT [5] for the analysis of multivariate categorical data. In addition to using the linear, logarithmic and exponential transformations of GENCAT,

CRISCAT also uses:

1. Operators which transform part of the vector to logarithmic or exponential functions and leave the rest of the vector unchanged;
2. A permutation operator which allows the user to rearrange elements in a vector so as to restrict the size of input matrices required in intermediate calculations.

Section 5 gives a characterization of asymptotic regression methodology, which is discussed extensively in [10] and applied to a varied array of problems in [5–7]. Also given are various functions of the adjusted survival rates which facilitate the fitting of certain probabilistic models.

Section 6 contains the operating instructions for CRISCAT which, together with [5], enable the user to employ a wide range of options resulting in a general methodology for the analysis of competing risks survival data. Section 7 contains example problems and the corresponding input to CRISCAT.

2. Adjusted survival probabilities

In the following definitions suppose that failures may occur due to any one of $k = 1, 2, \dots, c$ causes and let M denote a unit that is operative at time $t = 0$.

(i) Crude probabilities:

$$\xi_k(t) = \xi_k = \Pr \left(\begin{array}{l} M \text{ will fail in the interval } (0, t) \\ \text{due to the } k\text{-th cause when there} \\ \text{is risk of failure due to any one} \\ \text{of } c \text{ causes} \end{array} \right)$$

(ii) Net probabilities:

$$q_k(t) = q_k = \Pr \left(\begin{array}{l} M \text{ will fail in the interval } (0, t) \\ \text{when there is risk of failure due} \\ \text{to only the } k\text{-th cause} \end{array} \right)$$

$$p_k(t) = p_k = (1 - q_k) = \Pr \left(\begin{array}{l} M \text{ will survive the} \\ \text{interval } (0, t) \text{ when} \\ \text{there is risk of fail-} \\ \text{ure due to only the} \\ k\text{-th cause} \end{array} \right)$$

(iii) Total probabilities:

$$q(t) = q = \sum_{k=1}^c \xi_k(t) = \Pr \left(\begin{array}{l} M \text{ will fail in the interval} \\ (0, t) \text{ when there is risk} \\ \text{of failure due to any one} \\ \text{of } c \text{ causes} \end{array} \right)$$

$$\xi_0(t) = \xi_0 = 1 - q = \Pr \left(\begin{array}{l} M \text{ will survive the inter-} \\ \text{val } (0, t) \text{ when there is} \\ \text{risk of failure due to any} \\ \text{of } c \text{ causes:} \end{array} \right)$$

If the risks act independently, then:

$$\xi_0 = S(t) = \exp \left\{ - \int_0^t \lambda(x) dx \right\} \tag{2.1}$$

where

$$\lambda(x) = \sum_{k=1}^c \lambda_k(x)$$

is the total hazard function and $\lambda_k(x)$ is the hazard due to the k -th cause. Following Chiang [1,2], we assume that:

$$\frac{\lambda_k(x)}{\lambda(x)} = \psi_k \text{ for all } x \text{ in } (0, t) \tag{2.2}$$

Under these conditions, the following relationship can be established:

$$p_k(t) = \{\xi_0(t)\}^{\{\xi_k(t)/q(t)\}} = \{\xi_0(t)\}^{\psi_k} \tag{2.3}$$

Thus, this fundamental equation expresses the net survival probabilities in terms of the crude failure probabilities, the total failure probability and the total survival probability.

3. Data structure

Let $i = 1, 2, \dots, s$ index a set of categories which correspond to distinct sub-populations as defined in terms of pertinent independent variables. Similarly, let $j = 1, 2, \dots, t$ index a set of time intervals during which survival status is observed and $k = 1, 2, \dots, c$ index a set of causes from which each study unit is subject to failure. Then let n_{ijk} denote the number of units in the i -th sub-population that fail during the

j -th time interval due to the k -th cause, and let:

$$n_{ij0} = (n_{i.} - \sum_{l=1}^j \sum_{k=1}^c n_{ilk})$$

denote the respective number of units in the i -th subpopulation who have survived through the end of the j -th exposure period. With this notation, a typical collection of grouped survival data with competing risks of failure can be summarized in the life table format shown in table 1. Koch, Johnson, and Tolley [8] have shown that such data from follow-up studies can be investigated equivalently from the point of view of a strict contingency table format. Similarly, Johnson and Koch [4] demonstrated that competing risks data can be analyzed in terms of matrix operations on both the strict contingency table and the life table format. However, implementation of the contingency table format may not be feasible computationally when t and c are both moderately large (e.g., $t \geq 10, c \geq 3$) because of the size and cumbersome manipulations required for various matrices. Although this source of difficulty is inherent in the analysis of competing risks, its impact can be reduced considerably by performing calculations on the basis of the life table format.

Table 1
Life table format for competing risks survival data

Sub-population	Time interval	Cause of failure				Alive at end of interval
		C_1	C_2	...	C_c	
1	1	n_{111}	n_{112}	...	n_{11c}	n_{110}
1	2	n_{121}	n_{122}	...	n_{12c}	n_{120}
...
1	t	n_{1t1}	n_{1t2}	...	n_{1tc}	n_{1t0}
2	1	n_{211}	n_{212}	...	n_{21c}	n_{210}
2	2	n_{221}	n_{222}	...	n_{22c}	n_{220}
...
2	t	n_{2t1}	n_{2t2}	...	n_{2tc}	n_{2t0}
...
s	1	n_{s11}	n_{s12}	...	n_{s1c}	n_{s10}
s	2	n_{s21}	n_{s22}	...	n_{s2c}	n_{s20}
...
s	t	n_{st1}	n_{st2}	...	n_{stc}	n_{st0}

4. Estimation of net probabilities

Based on the data structure in table 1, an estimate of the net probability that a unit in the i -th subpopulation survives the k -th cause of failure through the j -th time interval, given that failure does not occur before that time due to some other cause, is given by:

$$\tilde{F}_{ijk} = \prod_{l=1}^j \left(\frac{n_{ilt0} + \sum_{l'=l+1}^t \sum_{k'=1}^c n_{il'l'k'}}{n_{ilt0} + \sum_{l'=l}^t \sum_{k'=1}^c n_{il'l'k'}} \right) \{n_{ilk} / \sum_{k'=1}^c n_{ilk'}\} \tag{4.1}$$

The functions $\{\tilde{F}_{ijk}\}$ can be generated efficiently from a computer programming point of view by identifying an appropriate matrix formulation. In this context, the computations are applied to the concatenated vector ξ of life table conditional rates:

$$\hat{\xi}_{ijk} = (n_{ijk} / \sum_{k'=0}^c n_{ijk'}) \text{ for } k = 1, 2, \dots, c, 0 \tag{4.2}$$

in the row-wise order displayed in table 1; i.e.:

$$\hat{\xi}' = (\hat{\xi}'_{11}, \hat{\xi}'_{12}, \dots, \hat{\xi}'_{1t}, \dots, \hat{\xi}'_{s1}, \hat{\xi}'_{s2}, \dots, \hat{\xi}'_{st}) \tag{4.3}$$

where:

$$\hat{\xi}'_{ij} = (\hat{\xi}_{ij1}, \hat{\xi}_{ij2}, \dots, \hat{\xi}_{ijc}, \hat{\xi}_{ij0}) \tag{4.4}$$

The $\{\hat{\xi}_{ijk}\}$ represent estimates of probabilities for subjects from the i -th subpopulation during the j -th time interval, conditional on surviving the first $(j - 1)$ intervals. For $k = 1, 2, \dots, c$, the statistics estimate the conditional probability of failure due to the k -th cause; the $\{\hat{\xi}_{ij0}\}$ estimate the conditional probability of surviving all causes. Within the scope of the strict contingency table framework, the vectors $\{\hat{\xi}_{ij}\}$ are asymptotically uncorrelated. Thus, a consistent estimator for the covariance matrix of $\hat{\xi}$ is the block diagonal matrix $\tilde{V}_{\hat{\xi}}$ for which the blocks on the main diagonal are:

$$\tilde{V}_{\hat{\xi}_{ij}} = [D_{\hat{\xi}_{ij}} - \hat{\xi}_{ij}\hat{\xi}'_{ij}] / n_{i,j-1,0} \tag{4.5}$$

The overall estimator $\tilde{F}(\hat{\xi})$ for the net survival prob-

abilities is formulated as the compound vector:

$$\tilde{F} = \tilde{F}(\tilde{\xi}) = \exp(-\tilde{J}) \exp[\tilde{L}_1, \tilde{L}_2] \log_e \begin{bmatrix} \tilde{A}_1 \cdot \tilde{\xi} \\ \tilde{K}[\log_e(\tilde{A}_2 \tilde{\xi})] \end{bmatrix} \quad (4.6)$$

for which:

$$\begin{aligned} \tilde{A}_1 &= \begin{bmatrix} I_c & \mathbf{0}_c \\ \mathbf{1}'_c & 0 \end{bmatrix} \otimes I_{st}, \tilde{A}_2 = [\mathbf{0}'_c, 1] \otimes I_{st} \\ \tilde{L}_1 &= [I_c, -\mathbf{1}_c] \otimes I_{st}, \tilde{L}_2 = \mathbf{1}_c \otimes I_{st} \\ \tilde{J} &= (I_c \otimes T_{1,t}) \otimes I_s, \tilde{K} = -I_{st} \end{aligned} \quad (4.7)$$

where I_c is a c -dimensional identity matrix, $\mathbf{0}_c$ is a $(c \times 1)$ vector of 0's, $\mathbf{1}_c$ is a $(c \times 1)$ vector of 1's, $T_{1,t}$ is a t -dimensional lower triangular matrix of 1's, and \otimes denotes Kronecker product. $\tilde{A}_1 \cdot \tilde{\xi}$ yields the estimates $\{\hat{\xi}_{ijk}\}$ in (4.4) for $k \neq 0$, $\tilde{A}_2 \tilde{\xi}$ yields the estimates of $\{\hat{\xi}_{ij0}\}$ in (4.4), \tilde{L}_1 and \tilde{L}_2 generate the $\{p_{ik}(j)\}$ in (2.3) on the $\log(-\log)$ scale, which are then exponentiated, and $-\tilde{J}$ generates the log survival probability estimates which, in turn, are exponentiated for \tilde{F} . The function vector \tilde{F} is arranged according to cause within time within subpopulation. For purposes of model fitting it is usually convenient to rearrange the vector elements according to time within cause within subpopulation. This rearrangement can be obtained by permuting the elements of \tilde{F} to obtain the vector F shown in (4.8).

The calculations implied by (4.6) are straightforward and can proceed in the same sequential manner which has been described previously for GENCAT. However, from a program user point of view, their implementation can be somewhat cumbersome. CRISCAT provides further advantages with respect to savings in user time and effort by efficient use of Kronecker products for specifying matrices with common component blocks which arise in competing risks analysis, and by use of a permutation (perm) operator, a split-log (slog) operator, and a split-exponential (sexp) operator. The condensed formulation of the vector of estimated survival probabilities adjusted for competing risks is given by:

$$F = \exp \bar{A}_4 \text{ perm} \exp \bar{A}_3 \log_e \bar{A}_2 \text{ slog} \bar{A}_1 \tilde{\xi} \quad (4.8)$$

for which:

$$\bar{A}_1 = \begin{bmatrix} I_c & \mathbf{0}_c \\ \mathbf{1}'_c & 0 \\ \mathbf{0}'_c & 1 \end{bmatrix} \otimes I_{st} \quad (4.9)$$

slog = **slog** ($c + 2, 1$) takes the natural log of the last element of each segment of the function vector, which is divided into segments of equal length ($c + 2$). That is, slog forms the log of the total survival probability in each (subpopulation, time period) combination:

$$\bar{A}_2 = \begin{bmatrix} I_{c+1} & \mathbf{0}_{c+1} \\ \mathbf{0}'_{c+1} & -1 \end{bmatrix} \otimes I_{st} \quad (4.10)$$

$$\bar{A}_3 = [I_c, -\mathbf{1}_c, \mathbf{1}_c] \otimes I_{st} \quad (4.11)$$

perm = **perm** (c_1, c_2, \dots, c_u) permutes elements of the function vector, which is divided into segments of equal length u . The constants c_1, c_2, \dots, c_u , are specified by the program user so as to rearrange elements of the function vector according to time within cause within subpopulation:

$$\bar{A}_4 = -T_{1,t} \otimes I_{sc} \quad (4.12)$$

where $T_{1,t}$ is a t -dimensional lower triangular matrix of 1's.

The matrix \bar{A}_1 yields estimates of the crude failure probability for each cause and the total failure and total survival probabilities within each (subpopulation, time period) combination. The split-log transformation and \bar{A}_2 transform the estimate of each total survival probability to its negative logarithm and leave the other functions unchanged. Application of the log transformation, matrix \bar{A}_3 , and the exponential transformation yields estimates of the negative logarithms of the conditional net survival probabilities. Following permutation of the vector elements, matrix \bar{A}_4 yields, for each (subpopulation \times cause \times time period) combination, an estimate of the logarithm of the net survival probability adjusted for competing risks. Finally, the vector is exponentiated to obtain estimates on the probability scale.

The principal reason for expressing F in such a matrix framework is that it facilitates the construction of the consistent estimator V_F for the covariance matrix of F which is based on the first order linear

Taylor series (i.e., the δ -method). In other words, by applying matrix differentiation methods similar to those used in [3,6], it follows that the matrix H of first partial derivatives of the functions F evaluated at $\hat{\xi}$ can be written as:

$$H = \left. \left(\frac{dF(x)}{dx} \right) \right|_{x=\hat{\xi}} = D_F \bar{A}_4 P D_{y_3} \bar{A}_3 D_{y_2}^{-1} \bar{A}_2 D_{y_1}^{-1} \bar{A}_1 \quad (4.13)$$

where the vector y_1 contains elements of the vector $\bar{A}_1 \hat{\xi}$ in the positions corresponding to functions which are log-transformed, and 1's elsewhere:

$$y_2 = \bar{A}_2 [\mathbf{slog}(\bar{A}_1 \hat{\xi})], \quad y_3 = \exp \bar{A}_3 \log y_2 \quad (4.14)$$

D_y is a diagonal matrix with elements of y on the main diagonal, and P is the matrix for which $\text{perm}(y_3) = P y_3$ (P is an identity matrix with similarly permuted rows and columns). It follows that:

$$V_F = H V_{\hat{\xi}} H' \quad (4.15)$$

is a consistent estimator of the covariance matrix of the vector of adjusted survival rates.

5. Modeling and testing

The function vector F is a consistent estimator of $F(\xi)$. Hence, the variation among the elements of $F(\xi)$ can be investigated by fitting linear regression models by the method of weighted least squares. This phase of the analysis can be characterized by writing:

$$E_A \{F\} = E_A \{F(\xi)\} = F(\xi) = X\beta \quad (5.1)$$

where X is a pre-specified ($u \times v$) design (or independent variable) matrix of known coefficients with full rank $v \leq u$, β is an unknown ($v \times 1$) vector of parameters, and ' E_A ' means 'asymptotic expectation'. Alternatively, the transformed functions:

$$G = \log_e F \quad (5.2)$$

or

$$R = \log_e \{-\log_e F\} \quad (5.3)$$

can be investigated. In particular, the function vector G is of interest for fitting exponential probability models via the design matrix:

$$X = -t \otimes I_{sc} \quad (5.4)$$

where $t' = (t_1, t_2, \dots, t_r)$ represents the vector of right-hand endpoints of the t time intervals. As a result, the parameter vector represents the hazard rates for the various causes within each of the sub-populations. Similarly, the function vector R is of interest for fitting Weibull probability models with common shape parameter via:

$$X = \{[1_t \otimes I_c], [\log_e(t) \otimes 1_c]\} \otimes I_s \quad (5.5)$$

Somewhat more general formulations of the design matrix can be used to fit piecewise exponential, piecewise Weibull, and certain other special types of probability models. For each design matrix specified, CRISCAT prints a goodness of fit statistic for the model, estimated model parameters, predicted values, and residuals. In addition, the user may test any number of hypotheses of the form $C\beta = 0$ by specifying the C -matrices of interest. Input format for the X - and C -matrices is flexible, as with GENCAT.

6. Use of CRISCAT

Because CRISCAT is an extension of GENCAT, and the use of the latter program has been thoroughly described elsewhere [5], this section is directed at details regarding the extended features of CRISCAT. The extensions which facilitate the generalized analysis of competing risks data are contained in the FUNCTION FORMULATION CARDS. CRISCAT allows three additional transformations which GENCAT does not use:

- (i) Permutation of elements in the function vector
- (ii) Split-logarithmic transformations
- (iii) Split-exponential transformations

The split-log and split-exponential transformations are designed for the following framework. Suppose that the function vector may be divided into segments of equal length u , and within each segment one wishes to transform u_1 functions to their natural logs (or antilogs), and leave the remaining $u - u_1$ functions unchanged. Furthermore, suppose that the u_1 functions to be changed in each segment are the last u_1 functions of the segment. The split-log [$\text{slog} = \text{slog}(u, u_1)$] and split-exponential [$\text{sexp} = \text{sexp}(u, u_1)$] transformations perform this function. The variable values of u and u_1 are specified by the

user. Note that the length of a segment may equal the length of the function vector, so that it is permissible to apply the transformation to the full (unsegmented) function vector.

The purpose of the permutation transformation is to provide for rearrangement of the elements of the function vector. This is useful because it allows:

- (1) The efficient use of Kronecker products in matrix manipulations;
- (2) Application of the split-log and split-exponential transformations to any elements of the function vector.

Again, suppose that the function vector may be divided into segments of equal length u , and one wishes to permute the elements of each segment in the same way. The permutation transformation ($\mathbf{perm} = \mathbf{perm}(c_1, c_2, \dots, c_u)$) provides this capability. The constants c_1, c_2, \dots, c_u specify the desired permuted order of the elements and are supplied by the user. For example, if a segment of length $u = 5$ were represented by $(x_1, x_2, x_3, x_4, x_5)$, application of the transformation $\mathbf{perm}(1, 3, 2, 4, 5)$ would yield the segment $(x_1, x_3, x_2, x_4, x_5)$. As before, the assumption of a segmented function vector involves no loss in generality, since the length of a segment may equal the length of the entire vector.

The permutation transformation is equivalent to a linear transformation in which the linear operator is an identity matrix with rows permuted in the same manner as columns. For example, the permutation given in the preceding paragraph is equivalent to a linear transformation in which the basic block of the block diagonal linear operator is:

$$A^* = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Most of the additional input required to use these transformations is specified on the transformation card ([5] p. 206). However, for the permutation transformation, the user supplies an additional card following the transformation card which specifies the permutation order (c_1, c_2, \dots, c_u) . The descriptions of these cards are as follows:

TRANSFORMATION CARD

Columns	Information Contained
5	Type of transformation 0 = Permutation 1 = Linear 2 = Logarithmic, or split-logarithmic 3 = Exponential, or split-exponential 4 = Addition of a vector of constants
10	For linear operator matrix: 1 = Entire matrix will be read in by rows 2 = Basic block of a block diagonal matrix (with identical blocks) will be read in by rows 3 = Main diagonal of a diagonal matrix will be read in as a vector For permutation, split-logarithmic, and split-exponential transformations: Number of functions u in each segment
11–15	For linear operator matrix: Number of rows of the linear operator matrix (including all blocks if column 10 = 2) For split-log or split-exponential: Number of functions u_1 in each segment which will be transformed to their logarithms (or antilogs)
16–20	For block diagonal linear operator matrix: Number of rows in the basic block of the block diagonal matrix
25 (optional)	Print options: 0 (or blank) = Print resulting covariance matrix 1 = Suppress printing of resulting covariance matrix
30 (optional)	Save options: 0 (or blank) = Do not save resulting vector and its covariance matrix

Columns	Information Contained
	1 = Save resulting vector and its covariance matrix for subsequent analysis in the same run
	2 = Write resulting vector and its covariance matrix to unit 2 (typically punched cards)
33–80 (optional)	For: (1) Each row of the linear operator matrix, or (2) The vector of diagonal elements of the linear operator, or (3) The vector of constants to be added, or (4) The vector of constants defining permutation order Format by which vector(s) will be read [default = (16F5.1)].

PERMUTATION CARD

This card is input only if the transformation card specifies a permutation (column 5 = 0). This card contains the vectors of constants c_1, c_2, \dots, c_u , that indicate the order in which the elements of a segment are to be permuted (see related text for example). The vector is entered according to the format specified either by default or in columns 33–80 of the transformation card.

7. Examples and sample input cards

In this section two examples of the use of CRISCAT are presented with primary attention directed at the preparation of required control cards. Certain parts of the description of these examples and the results of analysis have been reproduced from [4], where further details concerning the choice of appropriate functions and relevant hypotheses to be tested can be found.

7.1 An example using ulcer data

This example is based on a 5 year follow-up study of patients receiving 1 of 4 operations for duodenal ulcer [9]. The operations selected on a random basis for treatment of duodenal ulcer were:

- (1) Vagotomy and drainage;
- (2) Vagotomy and antrectomy;
- (3) Vagotomy and hemigastrectomy;
- (4) Gastric resection.

Antrectomy involved removal of 20–30% of the distal stomach, hemigastrectomy involved removal of ~50% of the stomach, and gastrectomy without vagotomy involved resection of 66–75% of the stomach. Patients were observed at 6, 24, and 60 months subsequent to operation. Results in terms of failure due to (recurrence or death) versus (reoperation or lost to follow-up) are given in table 2.

In this example, death was confirmed in most cases by death certificate; however, recurrence was diagnosed only during hospital visit. Hence, exact time of failure is unknown for patients lost to follow-up or with recurrences. There is interest only in the survival curve with respect to recurrence or death after making a competing risk adjustment for reoperation or lost to follow-up. Analyses are undertaken in terms of the corresponding subset of elements of the estimated log net survival probability vector $G = \log_e F$.

Since G was not known a priori to be characterized by any specific type of model, preliminary analysis focused on tests for hypotheses pertaining to comparisons of the respective operations. For this purpose, the identity model $X = I_{12}$ was used so that the parameters:

$$\beta' = \{\beta_{ij}\} = [\beta_{11}, \beta_{12}, \beta_{13}, \beta_{21}, \beta_{22}, \beta_{23}, \beta_{31}, \beta_{32}, \beta_{33}, \beta_{41}, \beta_{42}, \beta_{43}]$$

correspond to the vector of natural logarithms of the cumulative net survival probabilities for the i -th operation at the end of the j -th time interval with risk of failure due only to death or recurrence. The corresponding observed survival rates F and their estimated standard errors are obtained via $F = \exp(G)$ and are given in table 3 together with statistical tests for hypotheses concerning operation comparisons. These tests suggest that differences among the results of the four operations become progressively larger over time, with statistical significance attained only after the 5 year follow-up. If a stochastic model can be found for describing these trends, then the variation of the elements comprising the vector G can be described in terms of a reduced model with fewer parameters.

Table 2
Observed life table for ulcer data

Subpopulation (operation)	Interval of failure (months)	Cause of failure		Patient status satisfactory at end of interval
		Death or recurrence	Reoperation or lost	
Vagotomy and drainage	6	10	10	317
	24	13	16	288
	60	26	36	226
Vagotomy and antrectomy (25% Resection)	6	9	9	313
	24	16	7	290
	60	18	36	236
Vagotomy and hemigastrectomy (50% Resection)	6	9	5	329
	24	5	17	307
	60	10	24	273
(75% Resection)	6	9	8	329
	24	15	11	303
	60	24	37	242

For this purpose, one potentially appropriate model is the suppressed exponential as described in [8]. The results of fitting this intermediate model are given in table 3. The residual lack of fit test indicates that the reduced model is appropriate for the ulcer data. However, the results given for tests of significance involving the intercept parameters $\{\lambda_{i0}\}$ suggest that further model simplification through the specification of a common intercept is warranted. While the significance tests also indicate differences among the hazard parameters $\{\lambda_{i1}\}$ for the 4 operations, inspection of the parameter estimates suggests a common hazard parameter is appropriate for the first, second, and fourth operation. The results of fitting this final model are given at the bottom of table 3. The residual lack of fit is non-significant ($\alpha = 0.25$). Moreover, the test of the hypothesis $H_0: \lambda_{11} = \lambda_{31}$ is significant ($\alpha = 0.01$) and, hence, the final model and its estimated parameters represent a concise characterization of the time until failure process for the 4 operations.

The quantity $\exp\{-\lambda_{*0}\} = 0.984$ represents the estimator for the probability of being alive without recurrence shortly after treatment regardless of type of operation; $\exp\{-\lambda_{11}\} = 0.998$ represents the estimator for the rate at which the probability of survival or non-recurrence decreases for each additional

month of risk subsequent to receiving the first, second, and fourth operations; and $\exp\{-\lambda_{31}\} = 0.999$ represents the estimator for the rate at which the probability of survival or non-recurrence decreases for each additional month of risk subsequent to receiving the third operation (vagotomy and hemigastrectomy).

The input cards necessary for applying CRISCAT to produce these estimates and test results are shown in fig. 1–5 and described briefly in the following paragraphs. For detailed descriptions of card preparation, see [5]. Data input (fig. 1):

This phase of the program is the same as GENCAT. There are $r = 3$ response profiles within each of the $s = 12$ subpopulations, so that $r*s = 36 \leq 80$ which allows the data to be entered as case 1 frequency data. The 12 data cards containing the observed frequencies are entered immediately after the Parameter Card for Frequency Data. Transformations (fig. 2):

The sequence of transformations needed to generate the observed log survival rates and observed survival rates is shown in fig. 2. The sequence for computing F is that given in eq. (4.8), except that the permutation transformation is not included. The reason for this is that adjusted survival probabilities were computed for one cause only, and thus, there is no difference between the nesting sequence (cause within time within operation) and the sequence (time within cause

Table 3
Summary of results from competing risks analysis of ulcer data

Subpopulation (operation)	Interval of failure (months)	Observed estimates				Final model predicted estimates			
		Survival rate	Estimated standard error	(log _e) Survival rate	Estimated standard error	(log _e) Survival rate	Estimated standard error	Survival rate	Estimated standard error
Vagotomy and drainage	6	0.9699	0.0094	-0.0306	0.0097	-0.0302	0.0046	0.9702	0.0045
	24	0.9290	0.0143	-0.0736	0.0154	-0.0737	0.0065	0.9290	0.0060
	60	0.8392	0.0211	-0.1753	0.0252	-0.1606	0.0136	0.8516	0.0116
Vagotomy and antrectomy (25% resection)	6	0.9724	0.0091	-0.0280	0.0093	-0.0302	0.0046	0.9702	0.0045
	24	0.9222	0.0150	-0.0811	0.0162	-0.0737	0.0065	0.9290	0.0060
	60	0.8609	0.0197	-0.1497	0.0229	-0.1606	0.0136	0.8516	0.0116
Vagotomy and hemi-gastrectomy (50% resection)	6	0.9736	0.0087	-0.0268	0.0089	-0.0214	0.0048	0.9788	0.0047
	24	0.9584	0.0109	-0.0425	0.0114	-0.0385	0.0071	0.9622	0.0068
	60	0.9259	0.0146	-0.0770	0.0158	-0.0727	0.0146	0.9298	0.0136
(75% resection)	6	0.9737	0.0087	-0.0267	0.0089	-0.0302	0.0046	0.9702	0.0045
	24	0.9285	0.0141	-0.0742	0.0151	-0.0737	0.0065	0.9290	0.0060
	60	0.8499	0.0200	-0.1626	0.0236	-0.1606	0.0136	0.8516	0.0116

Preliminary model	Hypothesis	d.f.	Q
$X = I_{12}$	$\beta_{11} = \beta_{21} = \beta_{31} = \beta_{41}, \beta_{12} = \beta_{22} = \beta_{32} = \beta_{42}, \beta_{13} = \beta_{23} = \beta_{33} = \beta_{43}$	9	22.97
	$\beta_{11} = \beta_{21} = \beta_{31} = \beta_{41}$ (no difference due to operations, time interval 1)	3	0.11
	$\beta_{12} = \beta_{22} = \beta_{32} = \beta_{42}$ (no difference due to operations, time interval 2)	3	5.50
	$\beta_{13} = \beta_{23} = \beta_{33} = \beta_{43}$ (no difference due to operations, time interval 3)	3	16.95

Intermediate model	Parameters	Estimated parameters	Estimated standard error	Hypothesis	d.f.	Q	
$X =$	-1 -6 0 0 0 0 0 0 0	λ_{10}	0.0147	0.0100	$\lambda_{10} = \lambda_{20} = \lambda_{30} = \lambda_{40}$ (intercepts equal)	3	0.59
	-1 -24 0 0 0 0 0 0 0	λ_{11}	0.0026	0.0004			
	-1 -60 0 0 0 0 0 0 0	λ_{20}	0.0148	0.0096			
	0 0 -1 -6 0 0 0 0 0	λ_{21}	0.0022	0.0004	$\lambda_{11} = \lambda_{21} = \lambda_{31} = \lambda_{41}$ (hazard rates equal)	3	21.10
	0 0 -1 -24 0 0 0 0 0						
	0 0 -1 -60 0 0 0 0 0	λ_{30}	0.0212	0.0090	$\lambda_{10} = \lambda_{20} = \lambda_{30} = \lambda_{40}$ $\lambda_{11} = \lambda_{21} = \lambda_{31} = \lambda_{41}$	6	21.21
	0 0 0 0 -1 -6 0 0 0						
	0 0 0 0 -1 -24 0 0 0						
	0 0 0 0 -1 -60 0 0 0						
	0 0 0 0 0 0 -1 -6	λ_{31}	0.0009	0.0002			
0 0 0 0 0 0 -1 -24	λ_{40}	0.0115	0.0092				
0 0 0 0 0 0 -1 -60	λ_{41}	0.0025	0.0004				
				Residual lack of fit	4	1.78	

Table 3 (continued)

Final model	Parameters	Estimated parameter	Estimated standard error	Hypothesis	d.f.	Q	
X =	-1 -6 0	λ_{*0}	0.01574	0.00472	$\lambda_{11} = \lambda_{31}$	1	19.92
	-1 -24 0						
	-1 -60 0	λ_{11}	0.00241	0.00023			
	-1 0 -6	λ_{31}	0.00095	0.00024	Residual lack of fit	9	3.07
	-1 0 -24						
	-1 0 -60						
	-1 -6 0						
	-1 -24 0						
	-1 -60 0						
	-1 -6 0						
-1 -24 0							
-1 -60 0							

COLUMN	1			2			3			4			5			6			7			8		
1234567890123456789012345678901234567890123456789012345678901234567890	5	1	1																					
	12	3																						
	10	10	317																					
	13	16	288																					
	26	36	226																					
	9	9	313																					
	16	7	290																					
	18	36	236																					
	9	5	329																					
	5	17	307																					
	10	24	273																					
	9	8	329																					
	15	11	303																					
	24	37	242																					

Fig. 1. Data input.

COLUMN	1			2			3			4			5			6			7			8		
1234567890123456789012345678901234567890123456789012345678901234567890	1	2	36	3																				
	1																							
	1 1																							
		1																						
		2	3	1																				
		1	2	36	3																			
	1																							
		1																						
		-1																						
		2																						
		1	2	12	1																			
	1-1	1																						
		3																						
		1	2	12	3																			
	-1																							
	-1-1																							
	-1-1-1																							
		3																						
		2																						

Fig. 2. Transformations.

COLUMN	1			2			3			4			5			6			7			8		
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890			
7	4	12																						
8	1	3																						
1	-1																							
1		-1																						
1			-1																					
8	1	3																						
1	-1																							
1		-1																						
1			-1																					
8	1	3																						
1	-1																							
1		-1																						
1			-1																					
8	1	9																						
1	-1																							
1		-1																						
1			-1																					
1	-1																							
1		-1																						
1			-1																					
1	1	-1																						
1			-1																					
1				-1																				

Fig. 3. Preliminary model.

COLUMN	1			2			3			4			5			6			7			8		
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890			
7	2	8	2																					
1	1	1																						
6	24	60																						
8	1	3																						
1	-1																							
1		-1																						
1			-1																					
8	1	3																						
1	-1																							
1		-1																						
1			-1																					
8	1	6																						
1	-1																							
1		-1																						
1			-1																					

Fig. 4. Intermediate model.

COLUMN	1			2			3			4			5			6			7			8		
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890			
7	1	3																						
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
6	24	60	6	24	60																			
8	1	1																						
1	-1																							

Fig. 5. Final model.

within operation). The linear operators used in this example to compute adjusted survival probabilities for a single cause are simply the appropriate submatrices of those defined in eq. (4.9) through eq. (4.12). The split-log transformation card specifies $\text{slog}(3,1)$, which means that the function vector is divided into segments of 3 elements, and that the last function of each segment (the total survival probability) is to be transformed to its natural log. The last transformation in the entire sequence is a log transformation which yields $G = \log_e(F)$. Preliminary model (fig. 3):

The design matrix parameter card indicates that the program is to generate an identity matrix of order 12 as the design matrix. The contrast matrix cards are for the 4 hypothesis tests shown in table 3.

The intermediate model is shown in fig. 4; the final one, in fig. 5. The contrast matrix cards are for testing the corresponding hypotheses shown in table 3.

7.2 Weibull model synthetic data

The data in table 4 are from a hypothetical experiment in which survival times have been synthetically generated as the minimum values for successive sets of independent random variables from Weibull distributions with the same shape parameter. These distributions were constructed so that the median survival times $\phi_k^{(0.50)}$ for the 3 causes and the first quartile $\phi_1^{(0.25)}$ for cause 1 were approximately as follows:

$$\begin{aligned} \phi_1^{(0.25)} &= 12 \text{ months} \\ \phi_1^{(0.50)} &= 34.5 \text{ months} \\ \phi_2^{(0.50)} &= 36 \text{ months} \\ \phi_3^{(0.50)} &= 96 \text{ months} \end{aligned} \tag{7.1}$$

The specifications in (7.1) are equivalent to the following scale (λ_k) and shape (δ) parameters for Weibull models:

$$\begin{aligned} \log_e \lambda_1 &= -3.35, \log_e \lambda_2 = -3.39, \log_e \lambda_3 = -4.21, \\ \delta &= 0.843 \end{aligned} \tag{7.2}$$

Since the data are arranged according to the life table format, the compound functions (4.8) and (5.3) can be used to obtain R , the vector of competing risks adjusted probabilities on the log(-log) scale. The resulting estimators are shown in table 5 together with their estimated standard errors.

Table 4
Observed life table for synthetic data generated from Weibull model with three independent causes of failure

Interval of failure (months)	Cause of failure			Alive at end of interval
	C ₁	C ₂	C ₃	
1	14	13	7	366
2	12	9	5	340
3	8	6	1	325
4	10	8	4	303
5	5	8	4	286
6	5	10	2	269
9	16	16	9	228
12	16	10	6	196
18	17	20	8	151
24	9	11	10	121
36	18	18	7	78
48	11	8	7	52
60	7	7	2	36
96	6	16	1	13

The equal shape parameter model has the following form:

$$E_A(\hat{R}) = X\beta = [X_1 \otimes I_3, (\log X_2) \otimes 1_3] \begin{bmatrix} \log_e \lambda_1 \\ \log_e \lambda_2 \\ \log_e \lambda_3 \\ \delta \end{bmatrix} \tag{7.3}$$

where $X_1 = 1_{14}$, and $X_2 = [1,2,3,4,5,6,9,12,18,24,36,48,60,96]'$

The weighted least squares estimators b for β in (7.3) and their estimated standard errors are given in table 6, together with the estimators for the proportional hazard rates ψ_k which are obtained by transformation as:

$$\psi_k = \lambda_k / (\lambda_1 + \lambda_2 + \lambda_3). \tag{7.4}$$

The input cards necessary to procedure the previous results are shown in fig. 6-9. Figure 6 contains the data for the example. Figure 7 contains the cards which specify the transformations of eq. (4.8) and eq. (5.3). In this case, matrix A_4 is the negative of that given in (4.12), and the last transformation is \log instead of \exp in order to obtain probability estimates on the log(-log) scale. Figure 8 contains the cards which specify the design matrix in (7.3). The cards of fig. 9 transform the estimates of β in

Table 5
Observed and predicted estimates of log-log survival rates and their standard errors from competing risks of synthetic Weibull data

Interval of failure (months)	Observed estimates			Model predicted estimates		
	Cause 1	Cause 2	Cause 3	Cause 1	Cause 2	Cause 3
1	-3.31 (0.27)	-3.38 (0.28)	-4.00 (0.38)	-3.11 (0.14)	-3.06 (0.14)	-3.85 (0.16)
2	-2.65 (0.20)	-2.82 (0.21)	-3.43 (0.29)	-2.59 (0.12)	-2.55 (0.12)	-3.33 (0.15)
3	-2.36 (0.17)	-2.56 (0.19)	-3.34 (0.28)	-2.29 (0.11)	-2.25 (0.11)	-3.03 (0.14)
4	-2.07 (0.15)	-2.27 (0.17)	-3.03 (0.24)	-2.08 (0.10)	-2.03 (0.10)	-2.82 (0.13)
5	-1.94 (0.14)	-2.04 (0.15)	-2.78 (0.22)	-1.91 (0.10)	-1.87 (0.10)	-2.65 (0.13)
6	-1.82 (0.14)	-1.79 (0.14)	-2.67 (0.21)	-1.77 (0.10)	-1.73 (0.10)	-2.51 (0.13)
9	-1.49 (0.12)	-1.47 (0.12)	-2.25 (0.18)	-1.47 (0.09)	-1.43 (0.09)	-2.21 (0.12)
12	-1.20 (0.11)	-1.28 (0.11)	-2.01 (0.17)	-1.26 (0.09)	-1.22 (0.09)	-2.00 (0.12)
18	-0.92 (0.10)	-0.93 (0.10)	-1.71 (0.15)	-0.96 (0.08)	-0.91 (0.08)	-1.70 (0.12)
24	-0.76 (0.10)	-0.74 (0.10)	-1.37 (0.14)	-0.74 (0.08)	-0.70 (0.08)	-1.48 (0.12)
36	-0.43 (0.10)	-0.42 (0.10)	-1.12 (0.14)	-0.44 (0.08)	-0.40 (0.08)	-1.18 (0.12)
48	-0.20 (0.10)	-0.24 (0.10)	-0.83 (0.14)	-0.23 (0.08)	-0.19 (0.08)	-0.97 (0.12)
60	-0.02 (0.10)	-0.06 (0.10)	-0.73 (0.14)	-0.06 (0.08)	-0.02 (0.08)	-0.80 (0.12)
96	0.22 (0.12)	0.50 (0.13)	-0.64 (0.16)	0.29 (0.09)	0.33 (0.09)	-0.45 (0.12)

COLUMN	1				2				3				4				5				6				7				8			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
5	1	1																														
14	4																															
14	13	7	366																													
12	9	5	340																													
8	6	1	325																													
10	8	4	303																													
5	8	4	286																													
5	10	2	269																													
16	16	9	228																													
16	10	6	196																													
17	20	8	151																													
9	11	10	121																													
18	18	7	78																													
11	8	7	52																													
7	7	2	36																													
6	16	1	13																													

Fig. 6.

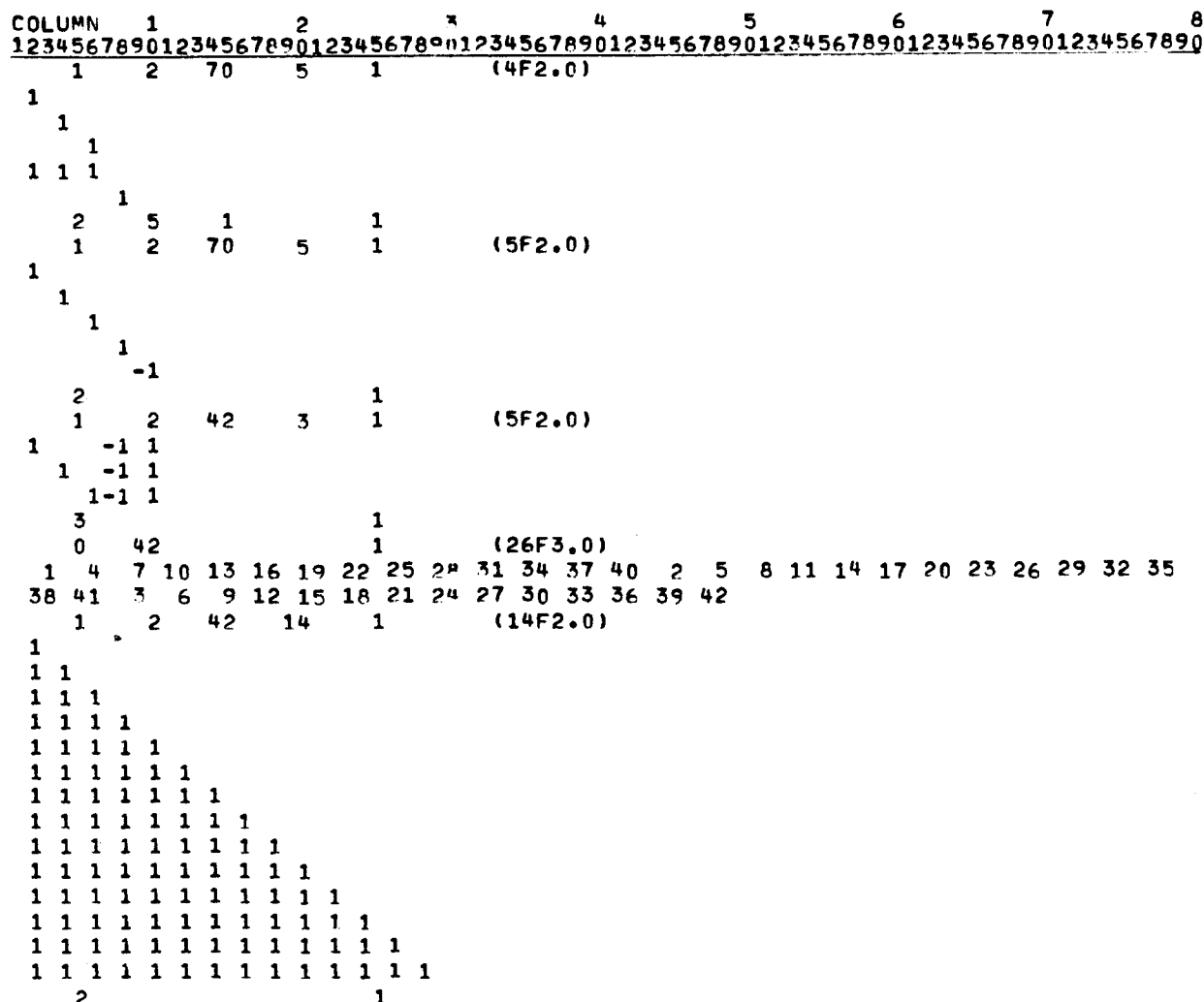


Fig. 7.

(7.3) to the proportional hazard rates of (7.4), while preserving the estimate of δ .

8. Hardware specifications

CRISCAT is written in double precision in IBM System 360/370 FORTRAN IV, which incorporates a few extensions to American National Standard (ANS) FORTRAN. As a result, minor modifications of the source code may be required to use the program on other machines.

Table 6
Estimated parameters and corresponding estimated standard errors from equal shape parameter model for Weibull synthetic data

Parameter	True value	Fitted Weibull	
		Estimate	Estimated standard error
$\log_e \lambda_1$	-3.35	-3.11	0.14
$\log_e \lambda_2$	-3.39	-3.06	0.14
$\log_e \lambda_3$	-4.21	-3.85	0.16
δ	0.843	0.744	0.033
ψ_1	0.419	0.397	0.025
ψ_2	0.404	0.414	0.025
ψ_3	0.177	0.189	0.020

Residual lack of fit: $Q = 35.61, d.f. = 38$.

10. Disclaimer

Although CRISCAT has been tested extensively, no warranty, expressed or implied, is made to the accuracy and functioning of the program. No responsibility is assumed by the authors. However, if specific problems or questions do arise, contact Dr. J. Richard Landis at the Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA.

Acknowledgements

The authors wish to thank Jean McKinney, Jean Harrison, and Cristina Ramirez for their conscientious typing of the original manuscript and subsequent revisions.

References

- [1] C.L. Chiang, *Biometrics* 17 (1961) 57–78.
- [2] C.L. Chiang, *Introduction to Stochastic Processes in Biostatistics* (John Wiley and Sons, 1968).
- [3] J.E. Grizzle, C.F. Starmer and G.G. Koch, *Biometrics* 23 (1969) 489–504.
- [4] W.D. Johnson and G.G. Koch, *Int. Stat. Rev.* 46 (1978) 21–51.
- [5] J.R. Landis, W.M. Stanish, J.L. Freeman and G.G. Koch, *Comput. Prog. Biomed.* 6 (1976) 196–231.
- [6] R.N. Forthofer and G.G. Koch, *Biometrics* 29 (1973) 143–157.
- [7] R.N. Forthofer and G.G. Koch, *Comput. Prog. Biomed.* 3 (1974) 237–248.
- [8] G.G. Koch, W.D. Johnson and H.D. Tolley, *J. Am. Stat. Assoc.* 67 (1972) 783–796.
- [9] W.D. Johnson, J.E. Grizzle and R.W. Postlethwait, *Arch. Surg.* 101 (1970) 391–395.
- [10] G.G. Koch, J.R. Landis, J.L. Freeman, D.H. Freeman and R.G. Lehnen, *Biometrics* 33 (1977) 133–158.