A COMPARISON OF THE LOGISTIC RISK FUNCTION AND THE PROPORTIONAL HAZARDS MODEL IN PROSPECTIVE EPIDEMIOLOGIC STUDIES

by

Manfred S. Green
Department of Epidemiology

and

Michael J. Symons
Department of Biostatistics

University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1432

A COMPARISON OF THE LOGISTIC RISK FUNCTION AND THE PROPORTIONAL HAZARDS MODEL IN PROSPECTIVE EPIDEMIOLOGIC STUDIES

by

Manfred S. Green[1] and Michael J. Symons[2,3]

[1]Department of Epidemiology, School of Public Health, University of North Carolina, Chapel Hill, North Carolina 27514. Present Address: Rekov Hatsabar 6/13; Kiron, Israel.

[2]Department of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill, North Carolina 27514.

[3]To whom reprint requests should be addressed. Accepted for publication in the Journal of Chronic Diseases.

ABSTRACT

A COMPARISON OF THE LOGISTIC RISK FUNCTION AND THE PROPORTIONAL HAZARDS MODEL IN PROSPECTIVE EPIDEMIOLOGIC STUDIES.
M.S. Green and M.J. Symons, University of North Carolina at Chapel Hill, North Carolina 27514.

The logistic regression and proportional hazards models are each currently being used in the analysis of prospective epidemiologic studies examining risk factors in chronic disease applications. The advantages and disadvantages of each are yet to be fully described. However, a theoretical relationship between the two models has been documented.

In this paper the conditions under which results from the two models approximate one another are described. It is shown that where the follow-up period is short and the disease is generally rare at all levels of the risk factors, the regression coefficients of the logistic model approximate those of the proportional hazards model with a constant* underlying hazard rate. Since under the same conditions the likelihood functions approximate one another, the regression coefficients have similar estimated standard errors. Further, estimation of relative risk with these models is contrasted.

These results are illustrated utilizing a previously published data set on metastatic cancer of the breast. With increasing follow-up time, the results from logistic regression became biased and less reliable.

KEY WORDS: Survival Methods, Relative Risk, Rate Ratio, Odds Ratio, Breast Cancer

---

*Postscript: The relationship is slightly more general. By inspection of equation (8), (9), and (10),

$$\beta_0 = \ell n\{\int_0^T \lambda_0(t)dt\} \quad ,$$

where $\lambda_0(t)$ could be a Weibull or Gompertz hazard, for example.

1.0 Introduction and Sketch of Results

The prospective epidemiologic studies on chronic disease that have been carried out in recent years have employed a number of different approaches in the analysis of the data generated. In these studies, a number of variables thought to be possible risk factors for the disease are measured on each individual in a selected cohort at a specified point in time. The cause specific morbidity and/or mortality is then recorded for each individual during the follow-up period.

The data generated from such studies thus consist of values on a dichotomous dependent variable, e.g., presence or absence of disease, alive or dead, and baseline readings of the postulated risk factors. Some years ago, Cornfield [1] suggested the use of the multiple logistic risk function for analysis of such data, and in 1967 Truett, Cornfield and Kannel [2] published a more extensive analysis of the Framingham Heart Study data using this model. Since that time logistic analysis has been widely applied to prospective epidemiologic studies of chronic disease [3,4] and the results of these analyses have contributed considerably to the current knowledge and understanding of the risk factors associated with various chronic diseases.

However, there is an aspect of logistic analysis which challenges its integrity as a model for survival data. This feature is that logistic analysis does not utilize information on the point in time during the follow-up period at which a positive response, i.e., manifestation of disease or death, occurs. Then, for example, a response occurring near the beginning of the follow-up period is given the same weight in the analysis as one occurring near the end of that period. As a result, the significance of the contribution of the postulated risk factor to disease

risk in a logistic analysis may be very dependent upon the length of follow-up period used.

This has several consequences in interpreting the analyses. In particular, comparisons of the significance of risk factors determined from different studies may be complicated by the varying lengths of follow-up periods. Hence, when applying a risk function derived from one study to data from another study, certain extrapolation or correction procedures have been employed [5]. There may be some difficulty in justifying these procedures, given the dependency of the analysis technique on the length of follow-up.

This limiting feature of logistic model was noted in 1973 by Myers, Hankey and Mantel [6]. They commented: "There may then be something questionable about treating our response data as consisting simply of 0's and 1's irrespective of the length of the study period". An additional parameter was introduced into their postulated logistic-exponential model in an attempt to deal with the lack of time-scale invariance of the logistic model. Their approach was illustrated by a portion of the data on a series of 920 disseminated breast cancer patients, analyzed earlier by Cutler, Asire and Taylor [7].

In 1972 Cox [8] proposed a model which incorporates covariates measured on each individual and utilizes the time of response in the analysis. It generalized the well established methods of life table analysis based upon the categorized time of response and is frequently used in the evaluation of mortality and morbidity in different populations. The Cox model uses the exact time of failures and accommodates censoring. This model, and variations of it, has been applied to data from several prospective studies of coronary heart disease and other diseases [9,10].

While both the logistic model and the Cox model are being used in the analysis of data from prospective epidemiologic studies, the relative suitability of each model for such studies is not clear.

Both approaches seem appropriate, with the Cox model using the available information more fully. It would be helpful then to explore the mathematical relation between the two models and to determine under what circumstances, if any, the Cox model and logistic regression yield substantially different results. In particular, it is of considerable interest to determine whether the findings of earlier studies based on logistic analysis would have been any different if the proportional hazards model of Cox had been used.

The mathematical result of Elandt-Johnson [11], derived to demonstrate an approximate relationship between the logistic model and the Cox model likelihoods, is elaborated upon and illustrated in this paper. The two models yield essentially similar results when the disease is rare and/or the length of follow-up is sufficiently short. These analyses are then applied to a data set on metastatic cancer of the breast adapted by Myers et al. [6] from Cutler et al. [7], and the results compared and contrasted. In general at the early stages of follow-up, the two models produce almost identical results. For more lengthy follow-up times the regression coefficients estimated by the two models begin to diverge, with the Cox model explaining more of the variability in the data. The coefficients obtained by the Cox model appear invariant to the time of follow-up and the standard errors decrease with increasing follow-up time. These observations are in distinct contrast to the results obtained from logistic regression.

2.0 Mathematical Relationship Between the Cox and Logistic Models

Let N be the number of individuals in a cohort, each with a set of k postulated risk factors, $\underset{\sim}{z}' = (z_1, z_2, \ldots, z_k)$, measured at the beginning of the follow-up period of length T. Defining a dichotomous response variable Y by

$$Y = \begin{cases} 1, & \text{if the disease manifests itself during follow-up } T; \\ 0, & \text{otherwise;} \end{cases}$$

then the multiple logistic function of risk factors $\underset{\sim}{z}$ with coefficients β_0 and $\hat{\underset{\sim}{\beta}}$ is given as follows:

$$P[Y=1|\underset{\sim}{z}, T] = \{1+\exp[-\beta_0-\underset{\sim}{\beta}'\underset{\sim}{z}]\}^{-1} \quad . \tag{1}$$

The corresponding probability of survival of the follow-up period of length T is

$$P[Y=0|\underset{\sim}{z}, T] = S(T|\underset{\sim}{z}, \underset{\sim}{\beta}) = \exp[-\beta_0-\underset{\sim}{\beta}'\underset{\sim}{z}]\{1+\exp[-\beta_0-\underset{\sim}{\beta}'\underset{\sim}{z}]\}^{-1} \quad . \tag{2}$$

Notice that the follow-up time T is implicit in the forms (1) and (2) and is expected to be the same for all individuals.

The model suggested by Cox [8] is based upon a hazard at positive time t having the form

$$\lambda(t|\underset{\sim}{z}, \underset{\sim}{\alpha}) = \lambda_0(t)\exp(\underset{\sim}{\alpha}'\underset{\sim}{z}) \quad , \tag{3}$$

where the coefficient vector $\underset{\sim}{\alpha}$ for the risk factors $\underset{\sim}{z}$ are not necessarily the same as the coefficients β_0 and $\underset{\sim}{\beta}$ in (1) and (2). The hazard (3) is termed proportional as it is the product of a common underlying hazard, $\lambda_0(t)$, and a function of the risk factors for the disease, usually $\exp(\underset{\sim}{\alpha}'\underset{\sim}{z})$. The corresponding survival function is given by

$$S(T|\underset{\sim}{z}, \underset{\sim}{\alpha}) = \exp\left[-\int_0^T \lambda_0(t)\exp(\underset{\sim}{\alpha}'\underset{\sim}{z})dt\right] \quad . \tag{4}$$

In Cox's model $\lambda_0(t)$ is presumed to be an arbitrary underlying hazard rate corresponding to the risk factors evaluated at zero. Notice that time of follow-up T is explicit in the Cox model survival probability (4) and may be different for each individual.

A relationship between a logistic regression and the special case of the Cox model with $\lambda_0(t) = \lambda$ is now sketched. For the logistic model and form (2) of $S(T|\underset{\sim}{z}, \underset{\sim}{\beta})$,

$$-\ell n[S(T|\underset{\sim}{z}, \underset{\sim}{\beta})] = \ell n\{[1+\exp(-\beta_0-\underset{\sim}{\beta}'\underset{\sim}{z})][\exp(-\beta_0-\underset{\sim}{\beta}'\underset{\sim}{z})]^{-1}\}$$

$$= \ell n(\frac{1+u}{u}) = \ell n(1+\frac{1}{u}) \quad , \tag{5}$$

where $u = \exp[-\beta_0-\underset{\sim}{\beta}'\underset{\sim}{z}]$. By a linearized Taylor series expansion,

$$\ell n(1+\frac{1}{u}) \doteqdot \frac{1}{u} \quad , \tag{6}$$

if u^{-1} is small. The approximation

$$-\ell n[S(T|\underset{\sim}{z}, \underset{\sim}{\beta})] \doteqdot \exp(\beta_0+\underset{\sim}{\beta}'\underset{\sim}{z}) \quad , \tag{7}$$

holds provided $\exp(\beta_0+\underset{\sim}{\beta}'\underset{\sim}{z})$ is small. Therefore, the logistic survival function (7) and Cox model survival function (4) will be approximately equivalent when

$$\exp(\beta_0+\underset{\sim}{\beta}'\underset{\sim}{z}) \doteqdot \int_0^T \lambda_0(t)\exp(\underset{\sim}{\alpha}'\underset{\sim}{z})dt \quad . \tag{8}$$

If the underlying hazard is constant, i.e., $\lambda_0(t) = \lambda$, then

$$\exp(\beta_0+\underset{\sim}{\beta}'\underset{\sim}{z}) \doteqdot \lambda T\exp(\underset{\sim}{\alpha}'\underset{\sim}{z}) \quad . \tag{9}$$

The approximate relationship (9) implies that $\beta_o = \ln(\lambda T)$ and

$$\beta_i \doteq \alpha_i \tag{10}$$

for $i = 1, 2, \ldots, k$. Notice that the length of follow-up T directly affects the intercept coefficient, β_o, in logistic regression. Thus the coefficients of the risk factors in the logistic regression will be approximately equal to the coefficients for the Cox model with a constant underlying hazard and proportional log-linear function of the risk factors. The result holds when $\exp(\beta_o+\beta'z)$ is small, implying that the sum of $\beta_o = \ln(\lambda T)$ plus the linear combination, $\underset{\sim}{\beta}'\underset{\sim}{z}$, of risk factors is small. Hence, neither the cumulative incidence, λT, over the follow-up period nor the combined effects of the covariates, $\underset{\sim}{\beta}'\underset{\sim}{z}$, can be too large. The approximation is therefore enhanced with rare diseases and short follow-up periods, in which the effects of the risk factors are not too great. On the other hand, the approximation will be poorer for more common diseases, longer periods of follow-up, or diseases with risk factors that markedly increase the disease incidence.

Elandt-Johnson [11] has recently shown that the likelihoods for the logistic regression approximates the partial likelihood function of the Cox model when the disease incidence is low, the follow-up period is not too long and the effect of the risk factors for the disease is not too great. With the underlying hazard for the Cox model taken to be constant, this is related to the logistic-exponential model modified by Myers, et al. [6]. When maximum likelihood is used to estimate the corresponding model parameters, similar estimates, standard errors, and significance tests with the approaches can be expected under the conditions just described. However, as the example will show, the parameter estimates for the logistic

regression model can become biased and less precise with increasing lengths of follow-up.

3.0 Three Calculations of Relative Risk With One Risk Factor

Let the event D be the manifestation of disease during a follow-up period of length T with one risk factor present. Then the relative risk of disease for the two levels of risk or exposure, denoted by z_1 and z_2, is defined to be

$$RR = \frac{P(D|z_1)}{P(D|z_2)} \quad . \tag{11}$$

3.1 Relative Risk Calculation with Survivorship Model

For a general survival distribution with hazard $\lambda(t, z)$, the relative risk for the two levels of the risk factor is the ratio

$$RR = \frac{1 - \exp\{-\int_0^T \lambda(t, z_1)dt}{1 - \exp\{-\int_0^T \lambda(t, z_2)dt} \quad , \tag{12}$$

where the cumulative hazard is denoted by

$$\Lambda(T, z) = \int_0^T \lambda(t, z)dt \quad . \tag{13}$$

The Cox's proportional hazards with a Weibull underlying hazard rate is

$$\lambda(t, z) = \lambda p(\lambda t)^{p-1}\exp(\alpha z) \quad . \tag{14}$$

Consequently, the calculation of relative risk (12) with hazard (14) is given by

$$RR = \frac{1 - \exp\{-(\lambda T)^p \exp(\alpha z_1)\}}{1 - \exp\{-(\lambda T)^p \exp(\alpha z_2)\}} \quad . \tag{15}$$

3.2 Ratio of Cumulative Hazards and Rate Ratio Approximations

Under the assumption of low-disease incidence, the exponential functions in the numerator and denominator of (12) can be approximated by a linearized Taylor series, yielding

$$RR \doteq \frac{\int_0^T \lambda(t, z_1)dt}{\int_0^T \lambda(t, z_2)dt} = \frac{\Lambda(T, z_1)}{\Lambda(T, z_2)} \quad , \tag{16}$$

the ratio of cumulative hazards to time T and evaluated at the two levels of risk. Additionally, if the hazard $\lambda(t, z)$ is not a function of time, but of the risk factor only, i.e., exponential survival with hazard $\lambda(z)$, then (16) becomes

$$RR \doteq \frac{T\lambda(z_1)}{T\lambda(z_2)} = \frac{\lambda(z_1)}{\lambda(z_2)} \quad , \tag{17}$$

a rate ratio. Specifically with the proportional hazard of the Cox model (3) and constant underlying hazard λ, then

$$RR \doteq \exp[\alpha(z_1 - z_2] \quad . \tag{18}$$

3.3 Odds Ratio Approximation

The corresponding odds ratio of disease for two levels of risk is

$$OR = \frac{P[D|z_1]}{1 - P(D|z_1)} \Big/ \frac{P(D|z_2)}{1 - P(D|z_2)} \quad . \tag{19}$$

With the usual logistic regression forms in (1) and (2),

$$OR = \frac{\exp(\beta z_1)}{\exp(\beta z_2)} = \exp[\beta(z_1 - z_2)] \quad . \tag{20}$$

Further, if the disease is rare for both levels of the risk factor, then $1 - P(D|z) \doteq 1$ in the denominator and numerator of (19), which becomes

$$OR \doteq \frac{P(D|z_1)}{P(D|z_2)} = RR \quad . \tag{21}$$

Notice that this odds ratio approximation is a statement for prospective studies, while that of Cornfield [12] was for a retrospective, or case-control, research strategy.

3.4 Summary of Three Calculations for Relative Risk

From the definition of relative risk for a survival situation, relative risk is defined by (12). A specific calculational form is given in (15) for Cox's proportional hazard with a Weibull underlying hazard rate and one risk factor. When the disease incidence is generally low for all levels of the risk factor, relative risk can be approximated by the ratio of cumulative hazards (16). If in addition the underlying hazard is exponential, relative risk is approximated by the rate ratio and computed simply by (18). Logistic regression with a generally low disease incidence for all levels of the risk factor leads to a prospective study version of Cornfield's [12] odds ratio approximation to relative

risk (20). Also note that the same conditions for (10) to hold are required of (18) and (20), so then logistic regression and Cox's model with a constant underlying hazard yield similar relative risk approximations.

The calculations of relative risk with two or more risk factors are a straightforward extension of these formulae.

4.0 Illustration and Methods

An example of the effects of increasing length of follow-up on the relationship between logistic regression and Cox models is presented using the data on metastatic cancer adapted by Myers et al. [6] to illustrate their logistic-exponential regression model. These were data on 313 patients with disseminated breast cancer, which is a subset of the data analyzed by Cutler, et al. [7]. The study included a series of patients previously diagnosed as suffering from breast cancer and who subsequently presented with metastases. The patients were then followed up to determine vital status. Two variables were postulated as being significant prognostic factors, specifically, the length of disease-free interval and the number of sites of metastases.

The data were analyzed for periods of follow-up varying from three to 70 months using the logistic regression model, the Cox model, and a Weibull model with proportional hazards. Survivors were included in the logistic regression only if they survived the full length of the identified follow-up period. This restriction was not necessary for the Cox or Weibull model since these two approaches allow for censored survival times due to incomplete follow-up or losses during follow-up.

The parameters for the multiple logistic regression were estimated by the method of maximum likelihood as described by Walker and Duncan [13]

using the SAS program LOGIST [14]. Parameters for the Cox proportional hazards model were estimated by the conditional likelihood procedure [6] with Breslow's modification [15] using the SAS program PHGLM [14]. The data were further analyzed using a proportional hazards model with a Weibull underlying hazard to estimate the regression coefficients. The parameters for the Weibull model were estimated by means of a maximum likelihood procedure using the program MAXLIK written by Kaplan and Elston [16].

5.0 Results

The results of the analyses for the illustration are presented in Tables 1 and 2. Generally, an increased number of sites of disease implied progressively poorer survival and an increased length of disease-free interval was associated with improved survival. The regression coefficients and their standard errors for each of the three models: logistic regression, Cox regression and Weibull proportional hazards regression are presented in Table 1 by increasing lengths of follow-up: 3, 5, 10, 20, 40 and 70 months. Also included are the relative risk estimates corresponding to each risk factor, model and length of follow-up. For illustration, the relative risk estimates are at levels $X_i = 1$ and $X_i = 0$ corresponding to the risk factor specified ($i = 1$ or 2) and a level of zero for the other risk factor. Since the relative risk estimates are straightforward functions of the regression coefficients and model parameters, their pattern of agreement follows that of the regression coefficients which is described next.

As expected from the series approximations in the mathematical development, the regression coefficients for the two risk factors agree,

relative to their estimated standard errors, for the first few months of follow-up. However, for the longer periods of follow-up the coefficient estimates for the logistic model generally diverge and with increasing estimates of standard errors of these regression coefficients. For the Cox model and Weibull proportional hazards model the estimates are fairly stable for all intervals of follow-up and have decreasing standard error estimates. With longer follow-up experience, the additional information would be expected to decrease the standard errors of the estimates as was observed.

The significance of the contribution of the two risk factors to each model is assessed by the standard chi-square approximation, with two degrees of freedom here, to the likelihood ratio test. These results, summarized in Table 2, are consistent with the standard error estimates in Table 1 and suggestive that the Cox model and Weibull proportional hazards model are explaining considerably more variability in the data than logistic regression, especially at the longer intervals of follow-up.

Two important aspects of modelling these data were not explored. The relevance of the proportional hazards model was not examined. The approaches described by Cox [8] or Taulbee [17] could be employed. Further, neither the non-linearity of the risk factors for the categorized scale presented by Myers et al. [6], nor the potential interactive effects of these risk factors was investigated. However, within the framework of a log-linear additive proportional hazards model, clear advantage in modelling by survivorship techniques over logistic regression was demonstrated.

The data of Cutler et al. [7] used by Myers et al. [6] were slightly different from that of this paper. Also, the model devised to circumvent the lack of time invariance by the logistic model displayed insensitivity

to the additional parameter (W) incorporated for this purpose. However, the regression coefficients presented in Table 1 agree quite well with those reported by Myers et al. [6]. Their devised model and the proportional hazards models presented in Table 1 then have provided estimates that are time invariant, as contrasted with those from logistic regression. The standard errors provided in Table 1 at the 70 month follow-up are comparable to those of Myers et al. [6] when their parameter W is set to a fixed value, e.g., W = 1 or 4, but the standard errors are considerably less than those corresponding to the maximum of the likelihood when W is considered unknown. The proportional hazards approaches presented in this paper appear simpler and preferable to the logistic-exponential model based strategy of Myers et al. [6].

6.0 Discussion and Summary

The mathematical approximation between Cox's proportional hazard with a constant underlying hazard and a logistic regression analysis suggests that estimates of risk factor coefficients will be similar, provided the length of the follow-up period is short and the disease incidence is generally rare at all levels of the risk factors. The analytical basis for this has been reviewed and empirically demonstrated in a data set describing survival in a subset of breast cancer patients.

More importantly, the proportional hazards models provided regression coefficients which were relatively stable for increasing lengths of follow-up. The standard errors of these estimates decreased, as expected, with increasing follow-up time. This was in distinct contrast to the relative lack of time invariance of the estimates of the regression coefficients by logistic regression, an observation noted earlier in this data set by Myers et al. [6]. Further, the standard errors of these estimates with

the logistic model generally increased with longer follow-up.

In principle it would appear that the Cox model is superior to the logistic model since it utilizes time of response and hence incorporates more information. From a practical viewpoint, it also yields a somewhat better fit to these data. On the other hand, if the follow-up period is sufficiently short, both models yield essentially the same results so that in such circumstances the choice of model may be decided more on the basis of convenience, availability of computer software, and expertise of the analyst.

Both models yield estimates of relative risk for the risk factors directly from the analyses. One advantage of the logistic model is that it also generates the probability (or risk) of developing the disease for a given level of the risk factor. This is clearly useful for practical applications in the quantification of risk of disease. For the proportional hazards approach, in order to generate probabilities of developing the disease, a survival function must be specified to produce time specific values. This may be problematical, since the form of the underlying hazard function and thus the survival function is assumed to be unknown for the Cox model. However, as has been demonstrated in this study, flexible survival distributions, such as the Weibull used in this paper or Gompertz distributions, can be specified without altering the substantial findings derived from the Cox model. This does, however, make the computations somewhat more complex than those for the logistic model.

The question of when the logistic model can be regarded as a good approximation to the Cox model cannot be answered unequivocally. There is no clearcut point at which the approximation moves abruptly from "good" to "bad". Rather this example suggests that the approximation becomes

poorer with increasing cumulative disease incidence, which is the product of average annual incidence and length of follow-up and also depends upon the effect of the risk factors.

In addition, the odds ratio derived from the logistic model cannot be regarded as a good estimate of relative risk when the risk of disease is not generally small. Relative risk estimates from the Cox model required that the rate of disease incidence be generally low. The length of the follow-up period and the effect of the risk factors must also be considered.

In general, the Cox model should be regarded as more appropriate for the analysis of data when the time of the responses is known. While the Cox model is relatively new as compared to logistic regression, the greater use of the available information suggests its preference over logistic regression. This position was supported by a numerical example, but further analytical work is needed to more fully understand the inter-relationships of logistic regression and Cox's model.

ACKNOWLEDGEMENT

The authors are appreciative of the referees' comments which greatly improved this paper. Partial research support at the University of North Carolina in Chapel Hill for MSG from the Lipid Research Clinic (NHLBI Contract: NO1HV12243-L), Department of Biostatistics, and for MJS from the Epi-stat Unit (NCI Grant #5 P30 CA16086), Cancer Research Center, is gratefully acknowledged.

REFERENCES

1. Cornfield J. Joint dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure: A discriminant function analysis. Fed. Proc., 21: 58-61 (1962).

2. Truett J, Cornfield J, and Kannel W. A multivariate analysis of the risk of coronary heart disease in Framingham. J. Chron. Dis., 20: 511-524 (1967).

3. Gordon T, Castelli WP, Hjortland MC, et al. HDL cholesterol and CHD risk: The Framingham Study. Am J. Med., 62: 707-714 (1977).

4. Goldbourt U and Medalie JH. High density lipoprotein cholesterol and incidence of coronary heart disease: The Israeli Ischemic Heart Disease Study. Am J. Epid., 109: 296-308 (1974).

5. McGee D and Gordon T. The results of the Framingham study applied to four other U.S. based epidemiologic studies of cardiovascular disease. In The Framingham Study. DHEW Publication No. (NIH) 76-1083. Edited by Kannel WB and Gordon T (1976).

6. Myers MH, Hankey BF, and Mantel N. A logistic-exponential model for use with response time data involving regression variables. Biometrics, 29: 257-296 (1973).

7. Cutler SJ, Asire AJ, and Taylor SG. Classification of patients with disseminated cancer of the breast. Cancer, 25: 861-869 (1969).

8. Cox DR. Regression models and life tables (with discussion). J. Royal Stat. Soc., Series B, 34: 187-220 (1972).

9. Dyer AR, Persky V, Stamler J, et al. Heart rate as a prognostic factor for coronary heart disease and mortality: Findings in three Chicago epidemiologic studies. Am. J. Epid., 112: 736-749 (1980).

10. Dyer AR. An analysis of relationship of systolic blood pressures, serum cholesterol and smoking to 14-year mortality in the Chicago People's Gas Company Study, I. J. Chron. Dis., 28: 565-570 (1975).

11. Elandt-Johnson RC. Time dependent logistic models in follow-up studies and clinical trials, I. Binary data. Institute of Statistics Mimeo Series No. 1310, University of North Carolina (1980).

12. Cornfield J. A method of estimating comparative rates from clinical data: applications to cancer of the lung, breast, and cervix. J. National Cancer Inst., 11: 1269-75 (1951).

13. Walker SH and Duncan DB. Estimation of the probability of an event as a function of several independent variables. Biometrics, 54: 167-179 (1967).

14. SAS Supplemental Library User's Guide. SAS Institute, North Carolina (1980).

15. Breslow NE. Covariance analysis of censored survival data. Biometrics, 30: 89-99 (1974).

16. Kaplan EB and Elston RC. A subroutine package for maximum likelihood estimation (MAXLIK). University of North Carolina. Institute of Statistics Mimeo Series No. 823. Revised version (June 1978).

17. Taulbee JD. A general model for the hazard rate with covariables. Biometrics, 35: 439-450 (1979).

TABLE 1

Analysis Results of Metastatic Breast Cancer Mortality in 313 Patients[1]

Months of Follow-up (Fraction of Deaths)[2]	Risk Factor[3]	LOGISTIC REGRESSION $\hat{\beta}_L$	Std. Error	Odds Ratio	COX MODEL $\hat{\beta}_C$	Std. Error	Rate Ratio	RATIO $\beta_L:\beta_C$	WEIBULL MODEL $\hat{\beta}_W$	Std. Error	Risk Ratio
3 (26/309)	X_1	-0.414	0.197	1.51	-0.376	0.187	1.46	1.10	-0.379	0.187	1.46
	X_2	0.283	0.201	1.33	0.249	0.190	1.28	1.14	0.249	0.191	1.28
5 (59/308)	X_1	-0.486	0.143	1.63	-0.402	0.125	1.50	1.21	-0.408	0.125	1.50
	X_2	0.413	0.148	1.51	0.324	0.125	1.38	1.28	0.328	0.124	1.39
10 (116/307)	X_1	-0.537	0.119	1.71	-0.394	0.088	1.48	1.36	-0.413	0.088	1.51
	X_2	0.409	0.128	1.51	0.291	0.089	1.34	1.40	0.307	0.089	1.36
20 (187/305)	X_1	-0.512	0.122	1.67	-0.336	0.069	1.40	1.52	-0.358	0.069	1.43
	X_2	0.442	0.139	1.56	0.249	0.072	1.28	1.77	0.267	0.072	1.31
40 (246/302)	X_1	-0.684	0.169	1.98	-0.343	0.061	1.41	1.99	-0.372	0.061	1.45
	X_2	0.487	0.189	1.63	0.247	0.064	1.28	1.98	0.273	0.064	1.31
70 (284/300)	X_1	-1.277	0.394	3.58	-0.354	0.057	1.43	3.60	-0.379	0.057	1.46
	X_2	0.341	0.323	1.41	0.217	0.061	1.24	1.57	0.242	0.061	1.27

[1] Subset of data from Culter et al. [7] adapted by Myers et al. [6].

[2] The denominator applies only to the logistic model where patients not in the study for the full period were excluded.

[3] The risk factors are: X_1 = Disease free interval = 1, 2, 3, or 4 for intervals of <1, 1-1.9, 2-4.9, and 5+ years.
X_2 = Number of sites = 1, 2, 3, 4 (if 4 or more).

TABLE 2

Model Chi-Squares (2 d.f.) for the Two Risk Factors by Months of Follow-up and Analysis Model for Metastatic Breast Cancer Mortality, Cutler *et al.* [7].

Months of Follow-up	Logistic Regression	Cox Model	Weibull Model
3	6.68	6.17	6.26
5	20.34	18.36	18.86
10	32.30	31.53	34.84
20	29.94	35.49	40.24
40	26.84	44.89	53.14
70	17.46	46.09	57.04