Listy Biometryczne - Biometrical Letters
Vol. 35(1998), No. 2, 95-106

The role of covariates in experimental carcinogenesis*

Christos P. Kitsos

Department of Statistics, Athens University of Economics and Business,
76 Patission Str., Athens 10434, Greece

e-mail: kitsos@aueb.gr

SUMMARY

The target of this paper is to discuss the effect of covariates in the family of multistage models, which are used to describe the experimental carcinogenesis. The prediction near zero is rather difficult as any model in the neighborhood of zero can be approximated with a linear function. The class of Multistage Models was adopted to approximate the data. The effect of covariates was examined and their influence on prediction. For a real data set the logistic model was applied and the effect of the covariate "age" was examined.

KEY WORDS: percentile, prediction, covariates, response, efficiency, hazard function, link function.

1. Introduction

In quantitative toxicology the terms "dose" (i.e. the amount of a chemical or energy in a radiological situation administered to or received by exposed subject), "effect" (i.e. an action as a result of a stimulus received through a receptor), "response" (i.e. any detectable change) are approached and linked through a statistical model.

Eventually there are different statistical models to describe a process by which a normal cell becomes malignant through, at least one, transformation. When the malignancy is referred to a tumour we are referred to cancer. Sometimes the interest is focused on the "growth rate" of affected tissues as malignant tumours are capable of floating away and forming new malignant growths in other sites. Humans are certainly exposed to carcinogens such as nicotine but in principle risk assessments are based on

*The paper was submitted on the occasion of 70-th birthday of Professor Tadeusz Caliński.

animals rather than humans. This is also accepted for two main reasons: there are no reliable estimates for "safe doses" and the epidemiological methods are insensitive to a small increase in cancer. It is assumed that a "safe dose" is the dose which will not increase the current cancer incidence rate by more than an "acceptable" low risk level, see the early work the Hartley and Sielken (1977). The earlier suggestion to estimate the "virtually safe dose" (VSD) was based on confidence intervals and it was rather mechanistic (Crump et al., 1977; Armitage, 1982). The upper confidence limit for the proportion of tumours was calculated and the dose-response curve was extrapolated towards zero, for the investigated animal species. Although a different terminology is adopted for the percentile point L_p, $p \in (0,1)$, like MTD – maximum tolerance dose, TD – tumourigenic dose, ED – estimated dose, LD – lethal dose, the point remains the same: adopt that model which will provide the best downwards extrapolation to L_p, as it has no meaning to perform experiments below an unknown level of dose.

The crucial issue when fitting a multistage model is prediction, and therefore the effect of covariates is essential. There is a special interest in low-dose estimation in experimental carcinogenesis as the effect of a low dose is difficult to be investigated, i.e. a "small" dose might provide no response. Moreover, different models appear to be "linear" in the neighborhood of zero, see section 4, and therefore it is difficult to choose the appropriate model for prediction. In this paper the multistage models (MM) are briefly reviewed in Section 2 and the effect of covariates in experimental carcinogenesis is discussed in Section 3.

2. Multistage models (MM)

Different nonlinear models have been developed and applied under the name "multistage models". For the class of models which is to be studied in this paragraph, the target is to describe with a mathematical model the process leading to cancer, i.e. the process by which a normal cell is transformed to a malignant one. The class of MM has been applied for the analysis of a large number of epidemiological data (Armitage, 1985).The crucial issue when fitting an MM model is prediction.. To be more specific, the task is to predict through the, assumed correct, model the low-dose effects in a risk assessment. That is, extrapolation downwards in the neighbourhood of zero, while prediction in statistics is rather related to a forward extrapolation.

Because tumour incidence data is titled to a prescribed dose-response relationship, the approach to these data by toxicologists and environmentalists is rather empirical, without a reference to an explicit dose-response model.

The dose-response curve, $F(\cdot)$ say, is a result of a binary response problem. From a statistical point of view $F(x)$ is the cumulative distribution function, describing, through a probability model, the phenomenon, with x being the dose level. Moreover,

$F(x)$ is rather an assumed approximation, than a known deterministic mechanism for the phenomenon which it describes.

This is true as the outcome $Y_i = 1$ or 0, success or failure, is linked with the covariates and parameters through a model, F say, in the form

$$P(Y_i = 1) = F(x_i, \theta) = 1 - P(Y_i = 0),$$

where the covariate x_i is going with the observation Y_i, $i = 1, 2, \ldots n$, with n denoting the number of observations and θ the involved parameters.

The function $F(x)$ can be considered as a cumulative distribution function (cdf) of a random variable Z, say, defined through the random variable Y as follows

$$Y = Y(x) = \begin{cases} 1 & \text{if} \quad Z \leq x, \\ 0 & \text{if} \quad Z > x. \end{cases}$$

It is trivial then that $P(Y = 1) = P(Z \leq x) = F(x)$ and $E(Y) = F(x)$, $Var(Y) = F(x)(1 - F(x))$.

When it is assumed that cancer is the result of a single event (or "hit") in a single cell, the one-parameter model

$$F(x) = 1 - \exp(-\theta x), \theta > 0 \tag{2.1}$$

is considered, known as the one-hit model.

When a fixed number, k say, of (identical) "hits" occur in a tissue the multi-hit model is assumed to describe the phenomenon and the corresponding $F(x)$ is approximated by the assumed correct model

$$F(x) = \frac{1}{(k-1)!} \int_0^{\theta x} x^{k-1} \exp(-x) dx \; . \tag{2.2}$$

When a suplinear relationship is assumed,the one-hit model is transformed to the Weibull model with the shape parameter $s > 1$ and

$$F(x) = 1 - \exp(-\theta x^s). \tag{2.3}$$

With $s < 1$ a supralinear Weibull model is considered in (2.3).

When it is assumed that the susceptible cell can be transformed through k distinct stages in order to be a malignant one, the phenomenon can be described by the multistage model of Armitage-Doll (1954). The main assumption is that the transformation rate from each stage to the next one is linear. Eventually, the cdf of developing cancer from exposure to a dose x, within a fixed time period, is given by

$$F(x) = 1 - \exp[-(\theta_0 + \theta_1 x + \ldots + \theta_k x^k)], \tag{2.4}$$

where θ_i, $i = 0, 1, \ldots, k$, are defined through the coefficients of the linear transformations assumed between stages, i.e. $\theta_i = \theta_i(x) = \lambda_i + \mu_i x$ with x being a constant, continuously applied dose during the fixed sequence of k stages. It is assumed that $\lambda_i > 0$ and $\mu_i \geq 0$. Actually,

$$\theta_0 = \prod_{i=1}^{k} \lambda_i, \qquad \theta_1 = [\sum_{i=1}^{k} (\lambda_i/\mu_i)] \prod_{i=1}^{k} \lambda_i, \quad \ldots, \quad \theta_k = \prod_{i=1}^{k} \mu_i$$

The most usual forms of (2.4) are the multistage linear model and the multistage model. Notice that model (2.4) is developed on a different biological insight and not as a general mathematical form of the previous models.

The Logit and Probit models (McCullagh and Nelder, 1989), known as tolerance distribution models in cancer risk assessment, are also useful to toxicology and are included in the MM class.

The MM class appeared earlier (Armitage and Doll, 1954), and is based on the assumption that a single normal cell may become fully malignant when a sequence of k, say, irreversible heritable mutation-like changes occurred.. Now, under the assumption that the intermediate cells are subject to a stochastic birth-death process for cell proliferation and cell differentiation, when $k = 2$ the Biologically Based Models (BBM) were created by Moolgavkar and his associates, see Moolgavkar and Venzon (1979), and developed by a series of papers by Luebeck and Moolgavkar (1989, 1991, 1992). For an optimal design approach for estimating the percentiles L_p in the class of MM, see Kitsos (1998).

In principle, the parameter estimation is based on the method of maximum likelihood, see for details Dobson (1990, chapter 4) among others. The nice property is that the method of scoring provides an iterative weighted least squares scheme, which converges to the vector of estimates of θ, $\hat{\theta}$ say.

Example 1. Consider the one-hit model as in (2.1). For estimating the unknown θ the log-likelihood $l(\theta)$ is formed and is proportional to

$$l(\theta) \propto -\theta \sum (m_i - y_i) x_i + \sum y_i \log(1 - \exp(-\theta x_i)),$$

where y_i is the number of responses among m_i animals after a predetermined time when treated with dose x_i, $i = 1, 2, \ldots n$. The Newton-Raphson iteration $\theta_{v+1} = \theta_v - l'(\theta_v)/l''(\theta_v)$ provides the MLE $\hat{\theta}$. For the numerical evaluation of MLE through the Newton-Raphson scheme for the one-hit model, see Kitsos (1998).

The two families of models, MM and BBM, are, in principle, based on different hazard functions. Indeed: if the mutation rates are very small and independent of

time the hazard function of cancer for the Armitage-Doll model is

$$\lambda(t) = c(t - t_0)^{k-1}, \quad c > 0, \tag{2.5}$$

where k is the number of stages and t_0 is a fixed and positive number for the growth of tumour. At time t_0 there are no cancer tumours, while at time $t \geq t_0$ a cancer tumour is developed.

When interest is focused on identifying etiological agents of cancer and developing the appropriate statistics for risk assessment of environmental agents, then the most appropriate hazard function is the one defined by Cox as

$$\lambda(t) = \lambda_0(t)S(W, \beta) \tag{2.6}$$

and this is essential in the BBM class of models. With $\lambda_0(t) > 0$ known as the baseline hazard function, $S(W, \beta)$ is the risk function which relates the environmental factor W, i.e. the covariates, and the vector of unknown parameters β.

The BBM are based on a Poisson stage-to-stage process. For example, Moolgavkar and Venzon (1979) assumed a Poisson process with birth rate at the i-th cell $b_i(t) = ib$ and death rate at the i-th cell $d_i(t) = id$, i.e. a homogenous birth-rate process.

In this paper we restrict our interest to the MM class of models and we are focused on the effect of covariates. When the target is prediction this effect is essential and it is discussed in Section 3.

3. The effect of covariates for prediction

In most bioassays and at the experimental carcinogenesis as well, the target is to compare two different therapies/factors or to evaluate the prognostic factors. Because the population under study is rather heterogeneous with respect to prognosis, it is asked to adjust the covariate effect describing the above-mentioned heterogeneity (Cox and Snell, 1989). Let x_1 be the factor of interest and x_2 the covariate and θ_1, θ_2 be the corresponding regression parameters (for the classical regression model see Seber, 1977). The full model with link function g (McCullagh and Nelder, 1989), is [recall that $F(x) = E(Y)$ in Section 2]

$$F(x_1, x_2) = F(x) = E(Y/x_1, x_2) = g^{-1}(\theta_0 + \theta_1 x_1 + \theta_2 x_2), \tag{3.1}$$

while the restricted model with estimate θ_1^* is

$$G(x_1) = E(Y/x_1) = g^{-1}(\theta_0^* + \theta_1^* x_1). \tag{3.2}$$

Note that the models (3.1) and (3.2), although nonlinear, are intrinsic linear.

The variances corresponding to models (3.1) and (3.2) are

$$\text{Var}(Y/x_1, x_2) = \sigma^2_{Y.12}, \ \text{Var}(Y/x_1) = \sigma^2_{Y.1}. \tag{3.3}$$

Therefore, the relative efficiency of $\hat{\theta}_1$ to $\hat{\theta}_1^*$ can be defined as

$$RE(\hat{\theta}_1, \hat{\theta}_1^*) = \frac{\sigma^2_{Y.12}}{\sigma^2_{Y.1}} = \frac{1-\rho^2_{12}}{1-\rho^2_{Y2.1}}, \tag{3.4}$$

with $\rho_{12} = Cor(x_1, x_2)$ and $\rho_{Y2.1}$ being the partial correlation between Y and x_2 with x_1 assumed to be fixed, i.e. the effect of x_2 on Y. From (3.4) is easy to see that

$$RE(\hat{\theta}_1, \hat{\theta}_1^*) = \begin{cases} = 1 & \text{if} \ \ | \rho_{Y2.1} | \ = \ | \rho_{12} |, \\ < 1 & \text{if} \ \ \rho_{12} > \rho_{Y2.1}, \\ > 1 & \text{if} \ \ \rho_{12} < \rho_{Y2.1}. \end{cases} \tag{3.5}$$

In principle, our interest is concentrated on a randomized treatment effect, i.e. $\rho_{12} = 0$. That is, the emphasis is given to adjustment, as eventually $RE(\hat{\theta}_1, \hat{\theta}_1^*) \geq 1$. The question if $\theta_2 = 0$, which is actually a statistical null hypothesis, or $\theta_2 \neq 0$, the alternative, is crucial on misspecification by omitting or including x_2. Indeed: if x_2 is adjusted for, the assumed correct model (3.1) is fitted if $\theta_2 \neq 0$. But if $\theta_2 = 0$ this leads to overspecification of the model. If x_2 is not included, the (3.2) model is correct if $\theta_2 = 0$, while if $\theta_2 \neq 0$ the model (3.2) is underspecified.

If the link function is the logistic function the models (3.1) and (3.2) are reduced to

$$\log \frac{P_{1.12}}{1-P_{1.12}} = \theta_0 + \theta_1 x_1 + \theta_2 x_2, \tag{3.6}$$

$$\log \frac{P_{1.1}}{1-P_{1.1}} = \theta_0^* + \theta_1^* x_1, \tag{3.7}$$

with $P_{1.12} = P[Y = 1 \mid x_1, x_2]$ and $P_{1.1} = P(Y = 1 \mid x_1)$.

The above discussion is valid for $x_2 = x^2$ and also when the logistic function is replaced with function (2.4).

Note that for $g(\cdot)$ being the logistic model, the curvature of $1/g'(\cdot)$ is a convex function, leading to a downward bias of $\hat{\theta}_1^*$, i.e. $\lim \mid \hat{\theta}_1^* \mid < \mid \theta_1 \mid$. Therefore, the bias tends to zero only when $\theta_1 = 0$.

There is a similarity between the logistic and the Cox model (Prentice and Kalbfleisch, 1979; Schumacher et al., 1987; Lagakos and Schoenfeld, 1984), while the behaviour of the variances of the adjusted and unadjusted estimators needs, we think, more investigation in this particular problem. Note that $\text{Var}(\theta_0^*)$ and $\text{Var}(\theta_1^*)$ in (3.7) can be reduced if a D-optimal design approach is adopted, but still (3.5) holds with θ^* being the D-optimum vector of parameters.

Example 2. We are referring to the Ille-et-Vilaine study of oesophagel cancer, see Table 6.1 of Breslow and Day (1980). There are 12 risk categories defined by a two-level exposure (variable *alc*) and six age strata (variable *age*). The exposure is the average alcohol consumption: 0-79 g/day is considered as "unexposed" while over this amount means "exposed". The strata of ages come from the grouped ages 25-34, 35-44, 45-54, 55-64, 65-74, 75+. The data are summarized in Table 1 (Table 6.1 in the reference).

Table 1. The Ille-et-Vilaine study, Breslow and Day (1980)

alc	1	0	1	0	1	0	1	0	1	0	1	0
age	1	1	2	2	3	3	4	4	5	5	6	6
Cases	1	0	4	5	25	21	42	34	19	36	5	8
Total[1]	10	106	30	169	54	159	69	173	37	124	5	39

[1]The sum of cases and controls

The GLIM package was used to form models (3.6) and (3.1). The best fitting model (3.6) is

$$\log \frac{p_{1\cdot 12}}{1-p_{1\cdot 12}} = -4.100 + 1.780\ alc + 0.616\ age.$$

The log-likelihood ratio statistic, also known as scaled deviance D, say D_1 for this case, was evaluated as $D_1 = 31.92$. The variances of the estimates $\hat{\theta}_i$, $i = 0, 1, 2$, are 0.09862, 0.005315 and 0.03500, respectively.

When fitting model (3.7), we get

$$\log \frac{p_{1\cdot 1}}{1-p_{1\cdot 1}} = -1.857 + 1.730\, alc.$$

The scaled deviance was evaluated as $D_2 = 115.17$. The variances of the estimates $\hat{\theta}_1^*$, $i = 0, 1$, are 0.01111 and 0.3070, respectively. D follows a X^2_{n-p} distribution with n being the number of observations and p the number of parameters involved. When a good description of the model is provided then we can expect that D is almost equal to $n - p$. Using the GLIM program an estimate of the variance can be obtained as $D/(n-p)$. Therefore, the values corresponding to (3.3) are

$$\begin{aligned} \text{estimated } \sigma^2_{Y\cdot 12} &= 31.92/9 = 3.54, \\ \text{estimated } \sigma^2_{Y\cdot 2} &= 115.17/10 = 11.52. \end{aligned}$$

From (3.4), the relative efficiency of $\hat{\theta}_1$ to $\hat{\theta}_1^*$ is estimated as $11.52/3.54 = 3.25 > 1$. The effect of the covariate *age* can also be considered by evaluating the fitted $\hat{p}_i$ values. When the model (3.6) is assumed, $\hat{p}_i$, $i = 1, 2, \ldots 12$, cover the interval $[0.029774, 0.7982]$, while when the model (3.7) is fitted $\hat{p}_i$ are either around 0.1305 or

0.4683, as *alc* is a binary variable. That is, for prediction purposes the effect of x_2 is essential on evaluating $\hat{p}_i$.

Example 3. Consider the whole data set of Ille-et-Vilaine study, as in Appendix I of Breslow and Day (1980). There are six age groups, as in Example 2 (variable *age*), 4 alcohol groups : $0-39, 40-79, 80-119, 120+$ in g/day (variable *alc)* and 4 tobacco groups: $0-9, 10-19, 20-29, 30+$ in g/day (variable *tob).* Therefore, in total we have 96 observations.

Different logistic models were fitted. We report only the following results:

	Variables in model	Scaled deviance D	Expected D or df	Estimate of variance
1.	*age, alc, tob*	100.40	84	1.195
2.	*alc, tob*	221.42	89	2.485

Therefore, the relative efficiency of the covariate *age* is $2.487/1.195 = 2.08 > 1$. Note that if all variables are included in the model (fit *age-alc-tob* in GLIM terms) there are zero degrees of freedom.

With x_1 and x_2 as above, the hazard function in the Cox model is

$$\lambda(t) = \lambda_0(t) \exp(\theta_1 x_1 + \theta_2 x_2), \tag{3.8}$$

with $\lambda_0(t)$ the unknown baseline hazard function (see Section 2). When the covariate x_2 is omitted the Cox model is

$$\lambda(t) = \lambda_0(t) \exp(\theta_1^* x_1). \tag{3.9}$$

Asymptotically, this leads to underestimation of the randomized effect in the Cox model, as in the logistic model. The effect θ_1 is estimated with no bias only when $\theta_1 = 0$. For an extended discussion for hazard rate models with covariates, see Prentice and Kalbfleisch (1979).

The above discussion we think provides evidence that in the MM class the effect of covariates can be essential in the same way as in the BBM class, especially when the target is prediction.

4. Discussion

A general prediction problem has been discussed under a different approach by Kitsos (1993), with some applications to carcinogenesis. In this paper it is pointed out that the effect of covariates is crucial when the target is to solve the equation $G(L_p) = p$

from model (3.2) rather than from model (3.1) and the VSD L_p has to be estimated. The effect of covariate omission in experimental carcinogenesis is introduced and the relative efficiency is evaluated. The semiparametric nature of the Cox model does not allow to apply completely the techniques for generalized linear models which are an extension of the same techniques for the general linear model.

Now, the Risk Above Background (RAB) for a dose level x is defined for the MM as

$$RAB(x) = \frac{F(x) - F(0)}{1 - F(0)} \quad .$$

It is easy to see that for the one hit model or the Weibul model

$$RAB(x) = F(x)$$

and therefore for these models $RAB(L_p) = p$.

For the generalized multistage model it is easy to see that at dose level L_p we have

$$RAB(L_p) = 1 - (1 - p)e^{\theta_0}.$$

Moreover, for the generalized multistage model

$$\left.\frac{d\ RAB(x)}{dx}\right|_{x=0} = \theta_1,$$

that is, for low dose the incremental risk for a "small" dose level in the neighborhood of zero is approximately linear. This approximation to the unknown curve is equivalent to the tangent to the dose-response curve at the point $x = 0$. Actually, this is true for any k in the model (2.4), and thus the prediction near zero is a hazardous procedure. The effect of covariates is essential, as the target is the "best in terms of prediction" model to be adopted.

Certainly, the discussed case was referred to uncorrelated covariates and therefore an open problem might be the multicolinear predictive covariates, with application to experimental carcinogenesis.

Acknowledgments

The author would like to express his thanks to CCMS / NATO for a generous support through a pilot study on experimental carcinogenesis. The referee's comments are very much appreciated.

References

Armitage, P. (1982). The assessment of low dose carcinogenicity. *Biometrics* **28**, 119-129.

Armitage, P. (1985). Multistage models of carcinogenesis. *Environmental Health Perspectives* **63**, 195-201.

Armitage, P., Doll, R. (1954). The age distribution of cancer and a multi-stage theory of carcinogenesis. *Brit. J. Cancer* **8**, 1-12.

Begg, M.D., Lagakos, S. (1993). Loss in efficiency caused by omitting covariates and misspecifying exposure in logistic regression models. *Journal of the American Statistical Association* **88**, 166-170.

Bieler, G.S., Williams, R.L. (1993). Ratio estimates, the delta method and quantal response tests for increased carcinogenicity.. *Biometrics* **49**, 793-801.

Bowman, D., Chen, J.J., George, E.O. (1995). Estimating variance function in developmental toxicity studies. *Biometrics* **51**, 1523-1528.

Breslow, N.E., Day, N.E (1980). *Statistical Methods in Cancer Research.* Volume 1 - The Analysis of Case - Control Studies. IARC.

Bucchi, A.R., Gabrielle, M., Lupi, C. and Zapponi, G.A. (1988). Dose-response relationships in rodents of promoter carcinogens: A tentative interpretation of some downward trends. *Biomedical and Environmental Sciences* **1**, 184-193.

Chen, C., Gibb, H., Moini, A. (1991). Models for analyzing data in initiation-promotion studies. *Environmental Health Respectives* **90**, 287-292.

Cox, D.R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* **74**, 187-220.

Cox, D.R., Snell, E.J. (1989). *Analysis of binary data.* Second Edition, Chapman and Hall.

Crump, K.S., Guess, H.A. and Deal, K.L. (1977). Confidence intervals and tests of hypotheses concerning dose response relations inferred from animal carcinogenicity data. *Biometrics* **33**, 437-451.

Dobson, A.J. (1990). *An introduction to generalized linear models.* Chapman and Hall.

Doll, R. (1971). The age distribution of cancer: implications for models of carcinogenesis. *Journal of the Royal Statistical Society, Series A* **134**, 133-166.

Doll, R. (1978). An epidemiological perspective on the biology of cancer. *Cancer Res.* **38**, 3573-3583.

Hartley, H.O., Sielken, R.L. (1977). Estimation of 'Safe dose' in carcinogenic experiments. *Biometrics* **33**, 1-30.

Kafadar, K., Tukey, J.W. (1993). U.S. Cancer death rates: a simple adjustment for urbanization. *International Statistical Review* **61**, 257-281.

Kitsos, C.P. (1992). Adopting sequential procedures for biological experiments. In: *Model Oriented Data Analysis*, Muller, Wynn, Zhigljavsky (Eds), 3-9, Physica-Verlag.

Kitsos, C.P. (1993). An algorithm for constructing the best predictive model. In: *Advances in Statistical Software*, F. Faulbau (Ed.), 535-539, ZUMA Publication.

Kitsos, C.P. (1995). Sequential assays for experimental carcinogenesis. ISI 50th Session, Book 1, 625-626, Beijing, 21-29 Aug. 1995.

Kitsos, C.P (1998). Optimal Designs for Estimating the Percentiles of the Risk in Multistage Models of Carcinogenesis. To appear.

Lagakos, S.W., Schoenfeld, D.A. (1984). Properties of proportional – hazard score tests under misspecified regression models. *Biometrics* **40**, 1037-1048.

Luebeck, G.E., Moolgarkar, S.H. (1989). Two-event model for carcinogenesis: biological, mathematical and statistical considerations. *Risk Analysis* **10**, 323-341.

Luebeck, G.E., Moolgarkar, S.H. (1991). Stochastic analysis of intermediate lesions in carcinogenesis experiments. *Risk Analysis* **11**, 149-157.

Luebeck, G.E., Moolgavkar, S.H. (1992). Multistage carcinogenesis: population-based model for colon cancer. *Journal of National Cancer Institute*, 610-618.

McCullagh, P., Nelder, J.A. (1983). *Generalized linear models.* Chapman and Hall, London.

Moolgavkar, S. Venzon, D. (1979). Two-event models for carcinogenesis: incidence curves for childhood and adult tumours. *Mathematical Biosciences* **47**, 55-77.

Petersen, T. (1986). Fitting parametric survival models with time-dependent covariates. *Appl. Statist.* **35**, 281-288.

Prentice, R.L., Gloeckler, L.A. (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics* **34**, 56-67.

Prentice, R.L., Kalbfleisch, J.D. (1979). Hazard rate models with covariates. *Biometrics,* 25-39.

Schumacher, M., Olschewski, M., Schmoor, C. (1987). The impact of heterogeneity on the comparison of survival times. *Statistics in Medicine* **6**, 773-784.

Seber, G.A.F. (1977). *Linear regression analysis.* John Wiley and Sons.

Travis, C.C., White, R.K. (1988). Interspecific scaling for toxicity data. *Risk Analysis* **8**, 119-125.

Travis, C.C., White, R.K., Ward, R.C. (1990). Interspecies extrapolation of pharmacokinetics. *J. Theor. Biol.* **42**, 285-304.

Wosniok, W., Kitsos, C., Watanabe, K. (1998). Statistical issues in the application of Multistage and Biologically Based Models. To appear.

Wu, C.F. (1985). Efficient sequential designs with binary data. *Journal of the American Statistical Association* **80**, 974-984.

Zapponi, G.A., Loizzo, A., Valente, P. (1989). Carcinogenic risk assessment; some comparisons between risk estimates derived from human and animal data. *Experimental Pathology* **37**, 210-218.

Received 10 June 1998; revised 6 October 1998

O roli zmiennych towarzyszących w doświadczalnych badaniach nad rakiem

STRESZCZENIE

Celem pracy jest przedyskutowanie wpływu zmiennych towarzyszących w pewnej rodzinie modeli wieloetapowych stosowanych w experymentalnych badaniach nad rakiem. Predykcja w pobliżu zera jest trudna gdyż dowolny model może być w tym obszarze przybliżony funkcją liniową. Do aproksymacji danych wykorzystano klasę modeli wieloetapowych. Zbadano wpływ zmiennych towarzyszących na predykcję. Dla pewnych danych rzeczywistych zastosowano model logistyczny oraz przeanalizowano wpływ zmiennej towarzyszącej "wiek".

SŁOWA KLUCZOWE: percentyl, predykcja, reakcja, efektywność, funkcja ryzyka, funkcja łącząca.