Biometrical Journal **41** (1999) 3, 305–319

Quantitative Risk Assessment in Epidemiological Studies Investigating Threshold Effects

RALF BENDER

Department of Metabolic Diseases and Nutrition
Heinrich-Heine-University of Düsseldorf
Germany

Summary

In this paper a method for quantitative risk assessment in epidemiological studies investigating threshold effects is proposed. The simple logistic regression model is used to describe the association between a binary response variable and a continuous risk factor. By defining acceptable levels for the absolute risk and the risk gradient the corresponding benchmark values of the risk factor can be calculated by means of nonlinear functions of the logistic regression coefficients. Standard errors and confidence intervals of the benchmark values are derived by means of the multivariate delta method. The proposed approach is compared with the threshold model of ULM (1991) for assessing threshold values in epidemiological studies.

Key words: Benchmark values; Binary data; Epidemiological studies; Logistic regression; Quantitative risk assessment; Threshold model.

Zusammenfassung

In dieser Arbeit wird eine Methode der quantitativen Risikoabschätzung in epidemiologischen Studien zur Untersuchung von Schwellenwerteffekten vorgeschlagen. Zur Beschreibung der Assoziation zwischen einer binären Zielvariablen und einem stetigen Risikofaktor wird das einfache logistische Regressionsmodell verwendet. Indem akzeptable Schwellen für das absolute Risiko und den Risikogradienten definiert werden, können die entsprechenden Referenzschwellen für den Risikofaktor mit Hilfe von nicht-linearen Funktionen der logistischen Regressionskoeffizienten bestimmt werden. Standardfehler und Konfidenzintervalle für die Referenzschwellen werden mit Hilfe der multivariaten Delta-Methode abgeleitet. Der vorgeschlagene Ansatz wird mit dem Modell von ULM (1991) zur Abschätzung von Schwellenwerten in epidemiologischen Studien verglichen.

1. Introduction

In several medical areas such as toxicology and occupational epidemiology it is often of interest to assess whether an explanatory factor has a threshold effect on a specific response variable. Recently, in diabetes research a number of authors investigated whether glycosylated hemoglobin (HbA_{1c}) has a threshold effect on the

risk of microvascular diabetic complications (DIABETES CONTROL AND COMPLICATIONS TRIAL RESEARCH GROUP, 1993, 1995, 1996; DANNE et al., 1994; KLEIN et al., 1994; KROLEWSKI et al., 1995; REICHARD, 1995; ORCHARD et al., 1997). A similar problem is the identification of glycemic threshold levels suitable to prevent complications in pregnant diabetic patients (LANGER, 1996).

Unfortunately, for calculation of threshold values no standard method is available. Hence, a number of different approaches have been applied to identify glycemic threshold values leading to controversial results. Besides the invalid application of methods for continuous response variables to determine break points in linear regressions (JONES and MOLITORIS, 1984), these controversial results are the consequence of an unclear research hypothesis, insufficient amount of data, unreflected use of statistical methods and unjustified conclusions. In the problem of identifying specific glycemic levels, a threshold value is quite generally described and interpreted as a value, "below which the risk gradient of diabetes complications is minimal or flat" (ORCHARD et al., 1997). It should be recognized that the identification of such "threshold values" requires methods for quantitative risk assessment. First of all, it is required to specify a suitable statistical model describing the dose-response relationship between a continuous risk factor (HbA_{1c}) and a binary response variable (diabetic complication yes/no). Secondly, acceptable levels of the absolute risk or the risk gradient must be defined within the framework of the chosen statistical model. Thirdly, these values have to be estimated and the uncertainty of the estimations have to be described by means of confidence intervals. However, in all papers investigating the existence of glycemic thresholds, conclusions were presented without definition of what is meant by a "threshold value".

While a number of approaches to determine break points in regression models suitable for continuous response variables are presented in the literature (JONES and MOLITORIS, 1984; EDLER and BERGER, 1985), methods for binary response data are rare. Recently, ULM (1991) described a modification of the logistic model for estimating and testing a threshold value in epidemiological studies. This model assumes that the risk of an event is constant below the threshold and increases according to the logistic equation above the threshold. If this assumption is reasonable for the data considered the application of this method is justified. However, even in dose-response relationships without a threshold value according to ULM (1991), it is possible that there is a value below which the risk gradient is "flat". To identify such a value it has to be defined what is meant by "flat".

A number of approaches for quantitative risk assessment have been developed for quantal response data of toxicological and teratological experiments (MEISTER, 1994). Especially, the general concept of benchmark doses (CRUMP, 1984; CHEN and KODELL, 1989) seems to be appropriate for the problem of identifying glycemic threshold values. However, this concept has to be adapted to the non-experimental data situation of epidemiological studies. CRUMP (1984) defined the benchmark dose to be the lower 95% confidence limit of the dose corresponding to a

specified percent increase in an adverse effect over the background level. Thus, to calculate the benchmark dose according to CRUMP (1984) the background level has to be known. In the case of experimental quantal data the background level is given by the risk of the control group receiving the dose zero of a chemical agent. However, in non-experimental data situations such as epidemiological studies, there is no clear background risk as there is no clear "zero dose". Another issue is whether a benchmark dose should be defined in terms of a confidence limit. This concept has been criticized by MURRELL, PORTIER, and MORRIS (1998). A confidence limit is a reflection of the variance and sample size of the data rather than a reflection of the measured dose-response curve. At first the best estimate of a measure should be provided accompanied by an adequate information about its uncertainty. Thus, a benchmark value should be defined as a point estimate characterizing the dose-response relationship accompanied by confidence limits to indicate the uncertainty of the estimate.

In this paper a quantitative risk assessment approach suitable for binary response data of epidemiological studies investigating threshold effects is presented. Similar to the concept of benchmark doses for experimental quantal data, two types of benchmark values corresponding to an acceptable risk and an acceptable risk gradient are calculated on the basis of the simple logistic regression model. As these benchmark values represent nonlinear functions of the logistic regression coefficients, standard errors and confidence intervals are derived by means of the multivariate delta method (BISHOP, FIENBERG, and HOLLAND, 1975). The benchmark approach is compared to the threshold model of ULM (1991).

The proposed method is illustrated by application to a study investigating the association between glycosylated hemoglobin and the development of diabetic nephropathy and to a study investigating the risk of makrosomia in pregnancies of type 1 diabetic women.

2. Quantitative Risk Assessment Based Upon Logistic Regression

The standard method in epidemiological studies to analyze the effect of a continuous risk factor X on a binary response variable Y (e.g. disease yes/no) is logistic regression (HOSMER and LEMESHOW, 1989). Let $p(x) = P(Y = 1 \mid Y = x)$ be the probability of an event given $X = x$ then the simple logistic regression model assumes that the log odds (also called 'logits') of p are linearly related to X, i.e. $\log(p/(1-p)) = \alpha + \beta x$, which is mathematically equivalent to

$$p(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} . \tag{1}$$

Hence, the probability p of having an event is a continuous function of X which is differentiable at each point. In the following it is assumed that there is a positive relation ($\beta > 0$) between p and X. The following definitions and formula can be

transferred to negative associations and also to other generalized linear models using other link functions.

Generally speaking, a benchmark value is a characteristic point of the dose-response curve at which the risk of an event rises so steeply that this point has practical relevance. The difficulty is to define what is meant by "so steeply". There are at least two possibilities to define benchmark values based on the logistic curve.

At first, a benchmark can be defined as the *value of an acceptable risk level* (VARL), where the acceptable risk level is given by a probability p_0 (e.g. $p_0 = 0.05$). This means that for values of X below VARL the risk of an event is lower than p_0 (Figure 1). It follows from (1) that

$$\text{VARL} = p^{-1}(p_0) = \frac{1}{\beta}\left(\log\left(\frac{p_0}{1-p_0}\right) - \alpha\right). \tag{2}$$

The definition of the VARL is quite simple and can be easily communicated. However, the shape of the relation between p and X is not taken into account. It may be that for values of X above VARL the risk is above p_0 but only by such a small amount which has no practical relevance.

A suitable definition which takes the shape of the response curve into account can be formulated by means of the derivative of the logistic function (1), which is given by

$$d(x) = p'(x) = \frac{\beta \exp(\alpha + \beta x)}{(1 + \exp(\alpha + \beta x))^2}. \tag{3}$$

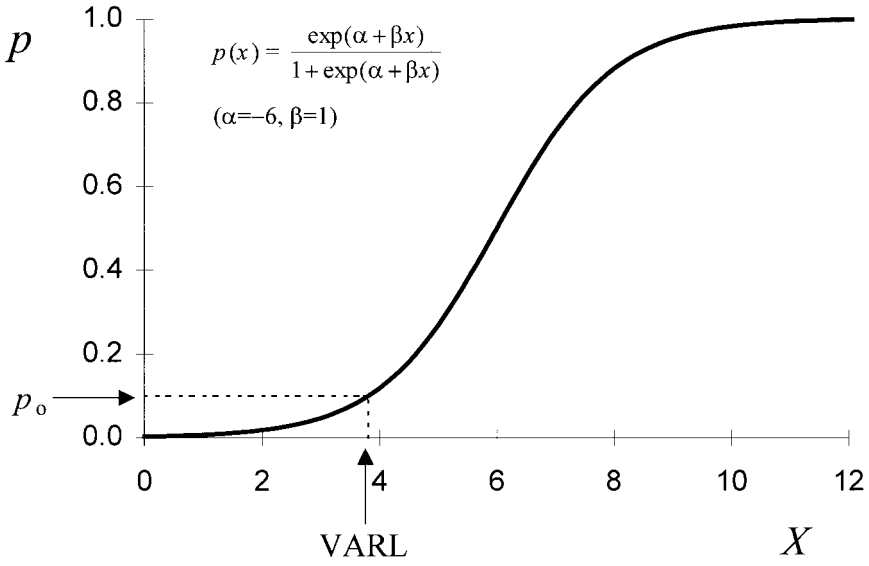

Fig. 1. Benchmark VARL defined as the value of the risk factor X at which the risk of an event based on the logistic curve is equal to the risk level p_0

The derivative (3) of a logistic curve represents an unimodal and symmetrical curve with a maximum value of $d_{\max} = \beta/4$ at $x_{\max} = -\alpha/\beta$ (Figure 2). Hence, for any given value d_0 (except of $d_{\max}$) there exist two corresponding x values, one below and one above $x_{\max}$.

A gradient type benchmark can be defined as the lower *value of an acceptable risk gradient* (VARG), where the acceptable risk gradient is given by a value $d_0 \leq \beta/4$ (e.g. $d_0 = 0.2$). If the chosen gradient d_0 is larger than $\beta/4$ no corresponding VARG exists. According to this definition for values of the risk factor X below VARG the risk increase per unit increase of X is lower than d_0. The meaning of the gradient type benchmark VARG is illustrated in Figures 2 and 3. It follows from (3) that

$$\mathrm{VARG} = d^{-1}(d_0) = \frac{1}{\beta}\left(\log\left(\frac{\beta - 2d_0 - \sqrt{\beta^2 - 4d_0\beta}}{2d_0}\right) - \alpha\right)$$

$$(0 < d_0 \leq \beta/4)\,. \tag{4}$$

Of course, the limits p_0 and d_0 are subjective and it may be difficult to give generally acceptable reasons for the chosen limits. However, if a benchmark value should have a clinical rather than a pure mathematical meaning the specification of limits is unavoidable. Other more complicated benchmark definitions are possible, perhaps based upon a combination of limits for the risk and the risk gradient.

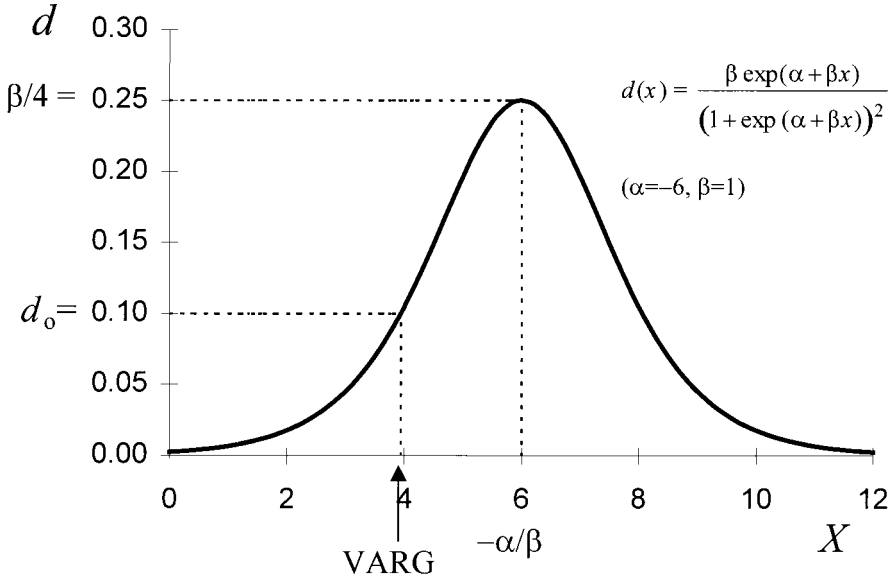

Fig. 2. Derivative of the logistic curve with maximum and benchmark VARG defined as the lower value of the risk factor X at which the risk gradient is equal to the gradient level d_0

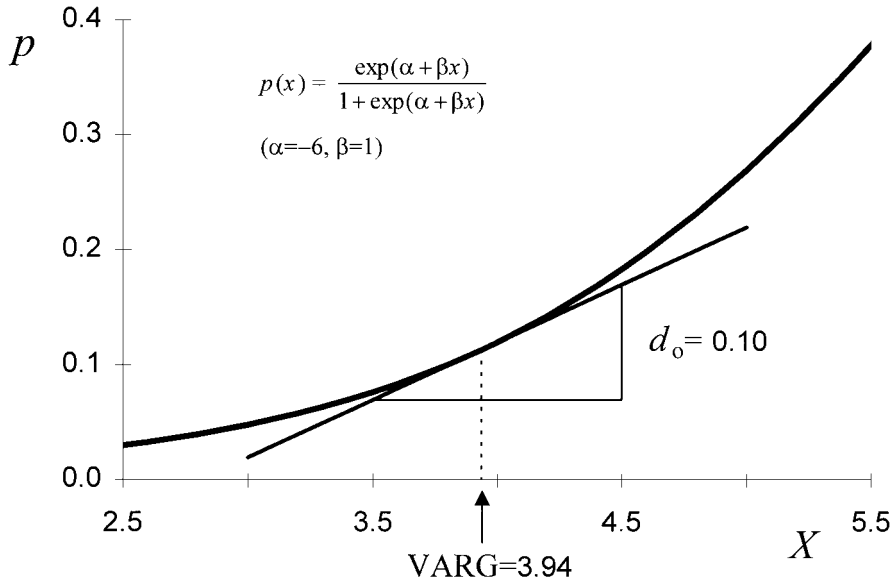

Fig. 3. Benchmark VARG defined as the lower value of the risk factor X at which the risk gradient of a logistic curve is equal to the gradient level d_0

3. Standard Errors and Confidence Intervals

In applications, VARL and VARG can be estimated by means of formulas (2) and (4), in which the logistic coefficients α and β are replaced by their maximum likelihood estimators (MLEs). As the benchmarks VARL and VARG are nonlinear functions of the logistic coefficients, the standard errors (SEs) of the corresponding estimators can be calculated approximately by means of the multivariate delta method (BISHOP et al., 1975). Let $\hat{\boldsymbol{\theta}}$ be the MLE of the parameter vector $\boldsymbol{\theta} = (\alpha, \beta)'$ and let $\boldsymbol{\Sigma} = (\sigma_{ij})_{i,j=1,2}$ be the covariance matrix of $\hat{\boldsymbol{\theta}}$ then the standard errors of the estimators of VARL and VARG can be calculated approximately by

$$\mathrm{SE}\,(\widehat{\mathrm{VARL}}) = \frac{1}{\hat{\beta}}\,\sqrt{\hat{\sigma}_{11} + 2\,\widehat{\mathrm{VARL}}\,\hat{\sigma}_{12} + (\widehat{\mathrm{VARL}})^2\,\hat{\sigma}_{22}}\,, \tag{5}$$

$$\mathrm{SE}\,(\widehat{\mathrm{VARG}})$$

$$= \frac{1}{\hat{\beta}}\,\sqrt{\hat{\sigma}_{11} + 2\left(\widehat{\mathrm{VARG}} + \sqrt{\hat{\beta}^2 - 4d_0\hat{\beta}}^{\,-1}\right)\hat{\sigma}_{12} + \left(\widehat{\mathrm{VARG}} + \sqrt{\hat{\beta}^2 - 4d_0\hat{\beta}}^{\,-1}\right)^2\hat{\sigma}_{22}}\,. \tag{6}$$

As a MLE is asymptotically normally distributed and a function of MLEs is the MLE of the corresponding parameter, the estimators of VARL and VARG are MLEs and asymptotically normally distributed. Let $\hat{\nu}$ be the estimator of VARL or VARG, then an approximate $(1-\alpha)\times 100\%$ confidence interval of the bench-

mark value ν can be calculated by

$$\hat{\nu} \pm z_{1-\alpha/2}\, \mathrm{SE}(\hat{\nu})\,, \tag{7}$$

where z_q denotes the q-quantile of the standard normal distribution.

For estimation of the benchmarks VARL and VARG and the calculation of standard errors and confidence intervals, I have prepared a SAS/IML (SAS, 1985) program, which is available via the internet ("http://www-public.rz.uni-duesseldorf.de/~bender/softw.htm") or from the author on request.

4. Threshold Model

In order to incorporate the concept of threshold effects directly into the statistical analysis ULM (1991) developed a threshold model within the framework of logistic regression. He modified the logistic model to

$$\log\left(\frac{p}{1-p}\right) = \begin{cases} \alpha & \text{for} \quad X \le \tau \\ \alpha + \beta(x-\tau) & \text{for} \quad X > \tau \end{cases}, \tag{8}$$

where τ is the threshold value of the risk factor X. This model assumes that the risk of an event is constant below the threshold τ and increases according to the logistic equation above τ. Instead of the logits other link functions can also be used. ULM (1991) proposed a maximum likelihood procedure based upon a grid search to estimate the unknown parameters τ, α, β and a likelihood ratio test for the hypothesis that the threshold value τ is larger than the observed minimum of X, i.e.

$$\mathrm{H}_0 : \tau \le \min(x) \quad \text{vs.} \quad \mathrm{H}_1 : \tau > \min(x)\,. \tag{9}$$

An exact algorithm for estimating the threshold value was developed by KÜCHENHOFF (1997). Based upon the likelihood ratio statistic an approximate confidence interval for τ can be constructed (ULM, 1991). For performing the iterative optimization procedure specialized software is required (e.g. GLIM, SAS/IML). For the calculations of the example given below I have used SAS/IML (SAS, 1985). Note that in Ulm's model the threshold is defined as non-differentiable break point of a continuous response curve regardless of the amount of risk increase at the threshold point.

5. Examples

5.1 *Glycemic threshold for diabetic nephropathy*

The methods described above were applied to data of a follow-up study of a prospective multicenter trial, which documented the feasibility to translate an in-

tensified insulin treatment and teaching program from a specialized University diabetes center to general hospitals (MÜHLHAUSER et al., 1996). In this study 784 consecutively referred patients with type 1 diabetes had taken part in the same 5-day treatment and teaching program for intensified insulin therapy in one of 10 participating hospitals and were re-examined after 1, 2, 3, and 6 years. Here, only the data of the subgroup of patients ($n = 480$) with no nephropathy at baseline and complete follow-up data are presented. The response variable considered here is development of nephropathy after 6 years of follow-up (NEPH). A patient was classified as having nephropathy if proteinuria >50mg/l or serum creatinine >133 μmol/l or a renal replacement therapy was applied. The main explanatory variable is glycosylated hemoglobin (HbA_{1c}) calculated as mean value of the measurements at baseline, 1, 2, and 3-year examination. Hence, the effect of glycemic control during the first 3 years of observation on the 6 year risk of diabetic nephropathy was investigated. HbA_{1c} levels were measured by the Diamat® HPLC-method (Biorad, München, Germany) which has a reference range for non-diabetic subjects of 4.3–6.1%. As important predictor for nephropathy the mean diastolic blood pressure (DBP) of the measurements at baseline and the first 3 years also is considered (MÜHLHAUSER et al., 1996). Table 1 summarizes this set of data.

Table 1
Characteristics of the sample of diabetic patients with no nephropathy at baseline

	Nephropathy at 6-year follow-up		
	No $n = 391$	Yes $n = 89$	Total $n = 480$
Gender (male/female)	$n = 210/n = 181$	$n = 49/n = 40$	$n = 259/n = 221$
Age at baseline (years)	26.51 (6.81)	25.88 (7.13)	26.39 (6.87)
Diabetes duration at baseline (years)	7.50 (6.97)	7.94 (6.82)	7.58 (6.93)
Mean systolic blood pressure (mmHg)	133.9 (11.4)	136.4 (12.4)	134.4 (11.6)
Mean diastolic blood pressure (mmHg)	77.9 (6.5)	79.8 (5.8)	78.3 (6.4)
Mean glycosylated hemoglobin (%)	7.52 (1.16)	8.36 (1.66)	7.67 (1.31)

data are given as means (SD) or numbers

At first, a simple and a multiple logistic regression analysis were performed. The results are given in Table 2. As the results of the two models concerning the effect of HbA_{1c} on NEPH do not differ, in the following only HbA_{1c} is considered as single explanatory variable. Nearly the same results were obtained by using the complementary log-log (-2 log-likelihood $= 431.51$) and the probit link function (-2 log-likelihood $= 432.90$).

Table 2
Logistic regression analysis for development of diabetic nephropathy

Risk factor	Regression coefficient	Standard error	p-value	Difference for odds ratio	Odds ratio	95% Confidence interval
			Simple logistic regression model			
Intercept	−5.089	0.731	0.0001			
HbA_{1c}	+0.457	0.089	0.0001	1%	1.58	1.3–1.9
	−2 log-likelihood = 432.52, Hosmer-Lemeshow goodness-of-fit test: $p = 0.965$					
			Multiple logistic regression model			
Intercept	−8.978	1.735	0.0001			
HbA_{1c}	+0.463	0.091	0.0001	1%	1.59	1.3–1.9
DBP	+0.049	0.019	0.0112	5 mmHg	1.28	1.1–1.5
	−2 log-likelihood = 426.04, Hosmer-Lemeshow goodness-of-fit test: $p = 0.924$					

sample size: $n = 480$, number of events: $n = 89$

Secondly, the benchmarks VARL and VARG were estimated as described above. Reasonable choices for p_0 are 0.01, 0.05, and 0.10. However, the corresponding benchmark values for $p_0 = 0.01$ and $p_0 = 0.05$ lay outside the observation range of HbA_{1c}. Hence, for the acceptable risk limit only the value $p_0 = 0.10$ was considered, yielding VARL = 6.3 (SE = 0.434, 95%-CI = 5.5–7.2). At first

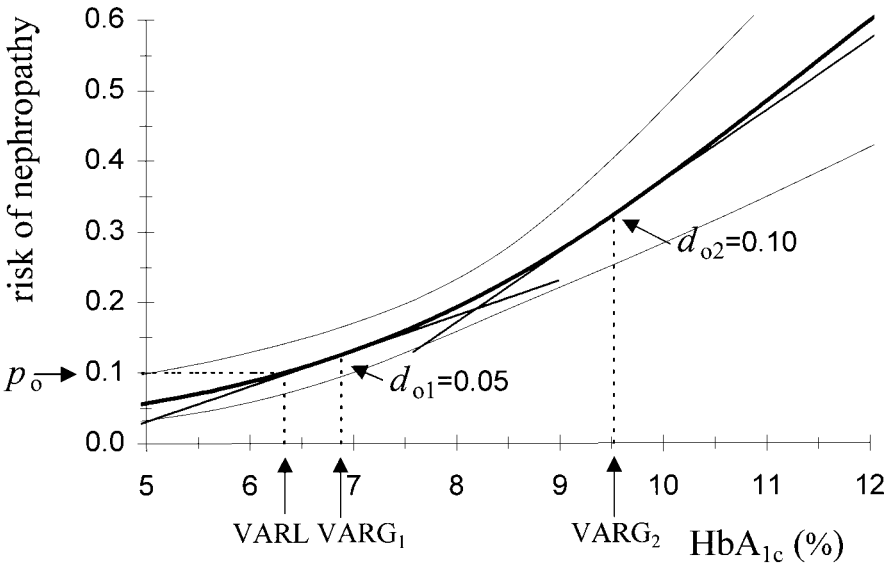

Fig. 4. Estimated logistic curve with 95% confidence band and corresponding benchmarks VARL (for $p_0 = 0.10$) and VARG (for $d_0 = 0.05$, 0.10) for the nephropathy data (sample size: $n = 480$, events: $n = 89$)

sight the choice of the acceptable risk limit $p_0 = 0.10$ appears to be high, but one has to consider that the 6 year risk of developing diabetic nephropathy is high in the considered population. The result of an HbA_{1c}-VARL of 6.3% means that a patient should be normoglycemic to have a risk below $p_0 = 0.10$ of developing nephropathy. However, the corresponding wide confidence interval makes a clear-cut clinical recommendation concerning gylcemic control difficult. The maximal value of the derivative of the estimated logistic response curve was $d_{\max} = 0.114$. Thus, as acceptable risk gradients the values $d_{01} = 0.05$ and $d_{02} = 0.10$ were considered. The corresponding VARG estimates were $VARG_1 = 6.9$ (SE = 0.427, 95%-CI = 6.0–7.7) and $VARG_2 = 9.5$ (SE = 1.514, 95%-CI = 6.6–12.5), respectively. This means that below an HbA_{1c} of 6.9% and 9.5% the risk gradient for developing nephropathy is below 0.05 and 0.10, respectively. As above, the wide confidence intervals indicate a large uncertainty of these estimates. These results are illustrated in Figure 4.

Thirdly, the threshold model of ULM (1991) was applied to the same data. A threshold value could be estimated with this approach yielding $\hat{\tau} = 5.9$ (95%-CI = 4.9–7.6). The fit of the threshold model (-2 log-likelihood = 431.64) was slightly better than that of the simple logistic model (-2 log-likelihood = 432.52) but was not significant ($p = 0.175$). Nearly the same result was obtained by using the complementary log-log link ($\hat{\tau} = 5.9$, 95%-CI = 4.9–8.1, -2 log-likelihood = 430.81, $p = 0.201$). The response curves of the logistic model and the threshold model are compared in Figure 5 demonstrating that there is no clinically relevant difference

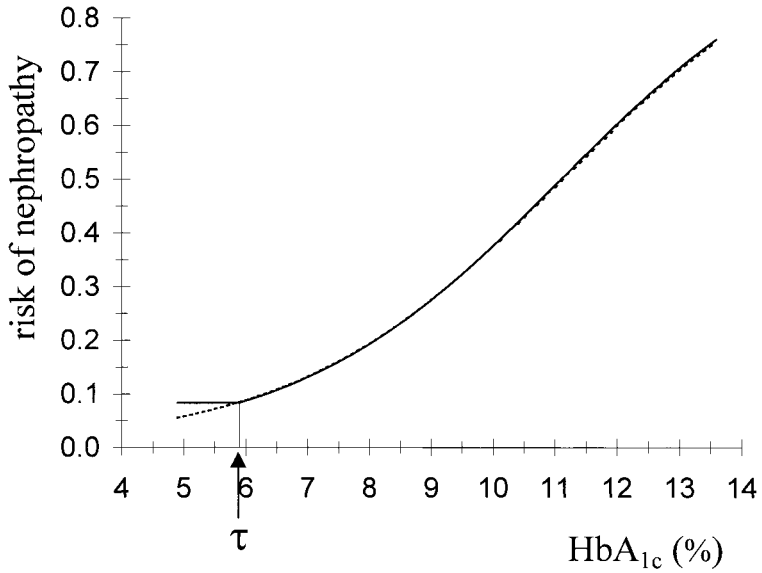

Fig. 5. Comparison of the estimated response curves from the logistic model (- - - - -) and the threshold model of ULM (1991) (———) for the nephropathy data (sample size: $n = 480$, events: $n = 89$)

between the curves. The data can be adequately described by both models. Whether the correct response curve contains a break point or not would lead to the same clinical recommendations concerning glycemic control.

5.2 *Glycemic threshold for fetal macrosomia*

A similar problem as in example 1 is the identification of glycemic threshold values which are suitable for the prevention of complications in pregnant diabetic patients (LANGER, 1996). One of these complications is fetal macrosomia, which is frequently defined as birth weight >90th percentile of the reference population adjusted for gender and gestational age. As the choice of the 90th percentile for definition of macrosomia is arbitrary, it is implausible to assume that a glycemic threshold defined as non-differentiable break point in the association between HbA_{1c} and the risk of macrosomia exists. However, as – by definition – the risk of macrosomia in the general population is 10%, an important clinical issue is whether there is a target value for glycemic control below which the risk of macrosomia is normalized, i.e. is 10%. This question leads directly to the calculation of the benchmark VARL using $p_0 = 0.10$.

For illustration, the data of a cohort study of type 1 diabetic pregnant women were used (KIMMERLE, BENDER, and BERGER, 1998). In this study information about macrosomia and the maternal third trimester HbA_{1c} (HPLC-method, normal range for pregnant women: 3.7–5.4%) was available for $n = 111$ women. The number of infants with macrosomia was $n = 27$.

A significant association between third trimester HbA_{1c} and the risk of macrosomia could be demonstrated. The results of a simple logistic regression analysis are shown in Table 3. The estimated benchmark VARL for the normal risk value in the general population of $p_0 = 0.10$ was VARL = 4.6 (SE = 0.495, 95%-CI = 3.6–5.6). As the confidence interval of the VARL estimate lies in the normoglycemic range of pregnant women, this result indicates that diabetic pregnant women should be normoglycemic to have a normalized risk of fetal macrosomia.

Table 3
Logistic regression analysis for fetal macrosomia

Risk factor	Regression coefficient	Standard error	*p*-value	Difference for odds ratio	Odds ratio	95% Confidence interval
			Simple logistic regression model			
Intercept	−6.408	1.852	0.0005			
HbA_{1c}	+0.917	0.315	0.0036	1%	2.5	1.3–4.6
	−2 log-likelihood = 113.42, Hosmer-Lemeshow goodness-of-fit test: $p = 0.785$					

sample size: $n = 111$, number of events: $n = 27$

6. Discussion

In this paper an approach of quantitative risk assessment for epidemiological studies investigating threshold effects of continuous risk factors is presented within the framework of logistic regression. This approach is compared with the threshold model of ULM (1991).

The demonstration of a threshold effect between a continuous risk factor and a binary response variable is not a simple task. It goes far beyond the output of a simple logistic regression analysis where only the strength of an association is assessed without the goal of describing the relationship in detail. Before a threshold effect can be established, the meaning of the threshold value has to be defined. ULM (1991) defined a threshold as non-differentiable break point of the risk response curve with constant risk below and increasing risk above the break point. In a similar but more general way PASTOR and GUALLAR (1998) defined a threshold as change point in a two-segmented logistic regression with different risk increases below and above the change point. These approaches have the advantage that they represent objective criteria and do not require a quantitative risk assessment based upon subjective limits. However, break and change points have first of all only a mathematical meaning. It is possible that a value can be found which fulfills the mathematical criteria but has no clinical relevance because the risk increase beyond the threshold point is negligible. On the other hand, there can be situations in which a benchmark value with clinical relevance exists but which is not conform with the mathematical definition of a break or change point. A quantitative risk assessment based upon the standard logistic curve requires the specification of acceptable limits for the absolute risk or the risk gradient but this procedure guarantees that the estimated benchmark values correspond with a clearly defined amount of risk increase. Hence, clinical decisions are possible and the meaning of the estimated benchmark values can be communicated.

The main drawback of all statistical models for estimating threshold values is that they make assumptions about the thresholds themselves. Hence, the results can be model dependent. ULM (1991) pointed out that for assessing of threshold effects always more than one model should be applied. One should be aware that the use of other models or even the use of data transformations can result in other estimates of the threshold value. Conclusions about threshold effects should only be made if the model assumptions, e.g. a constant risk below the threshold, is biologically plausible. In cases, where a threshold value defined as break or change point cannot be identified or is implausible, the quantitative risk assessment approach proposed in this paper represents a useful alternative.

In the example of the association between glycosylated hemoglobin and diabetic nephropathy no significant threshold value according to ULM (1991) could be estimated. That means that the data do not suggest the existence of a break point. Moreover, even if the estimated threshold would have been significant, the

clinical relevance of this threshold would be questionable as there is nearly no difference between the response curves of the threshold model and the logistic model (Figure 5). On the other hand, different benchmark values could be estimated in dependence on the limits of an acceptable risk and an acceptable risk gradient. However, no clear-cut conclusion could be drawn because the confidence intervals of all estimated benchmark values were too wide. It remained unclear whether diabetic patients should maintain their HbA_{1c} in the normoglycemic range or if a HbA_{1c} value beyond the upper limit of the normoglycemic range is sufficient to avoid a risk of diabetic complications above the chosen acceptable limits. Nevertheless, the benchmarks VARL and VARG give useful information which HbA_{1c} values are associated with which risk levels and risk gradients. The application of the quantitative risk assessment approach forces investigators to define the benchmark values which should be identified. It is not surprising that there is a continuous controversy whether a glycemic threshold for diabetic complications exists (ORCHARD et al., 1997) due to the lack of a clear definition of what is meant by a "threshold effect".

In the example of the association between the maternal third trimester HbA_{1c} and the risk of macrosomia the assumption of a break point was principally implausible. The quantitative risk assessment approach represents the natural method to identify the target value for glycemic control below which the risk of macrosomia is normalized. Although the confidence limits of the estimated benchmark VARL = 4.6 (95%-CI = 3.6–5.6) were wide, a clinically important conclusion is possible. As the confidence interval lies in the normal range of pregnant women (3.7–5.4%), the data suggest that diabetic women with normoglycemia during pregnancy have a normalized risk of fetal macrosomia.

In summary, the quantitative risk assessment approach proposed in this paper is a useful tool to estimate benchmark values of continuous risk factors in epidemiological studies investigating threshold effects. The approach is based upon the simple logistic regression model which is one of the most often used methods in epidemiology. It is required to specify acceptable levels for the absolute risk or the risk gradient, which guarantees the clinical relevance of the estimated benchmark values for the considered problem.

Acknowledgements

I thank Prof. Dr. I. Mühlhauser, PD Dr. R. Kimmerle, and Prof. Dr. M. Berger (Düsseldorf, Germany) for providing the data used in the examples. I thank Prof. H. KÜCHENHOFF (München, Germany) for the valuable comments. The support of the P. Klöckner Stiftung (Duisburg, Germany) is gratefully acknowledged.

References

BISHOP, Y. M. M., FIENBERG, S. E., and HOLLAND, P. W., 1975: *Discrete Multivariate Analysis: Theory and Methods*. MIT Press, Cambridge, MA.

CHEN, J. C. and KODELL, R. L., 1989: Quantitative risk assessment for teratological effects. *Journal of the American Statistical Association* **84**, 966–971.

CRUMP, K. S., 1984: A new method for determining allowable daily intake. *Fundamental and Applied Toxicology* **4**, 854–871.

DANNE, T., WEBER, B., HARTMANN, R., ENDERS, I., BURGER, W., and HOVENER, G., 1994: Long-term glycemic control has a nonlinear association to the frequency of background retinopathy in adolescents with diabetes. *Diabetes Care* **17**, 1390–1396.

DIABETES CONTROL AND COMPLICATIONS TRIAL RESEARCH GROUP, 1993: The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *New England Journal of Medicine* **329**, 977–986.

DIABETES CONTROL AND COMPLICATIONS TRIAL RESEARCH GROUP, 1995: The relationship of glycemic exposure (HbA_{1c}) to the risk of development and progression of retinopathy in the Diabetes Control and Complications Trial. *Diabetes* **44**, 968–983.

DIABETES CONTROL AND COMPLICATIONS TRIAL RESEARCH GROUP, 1996: The absence of a glycemic threshold for the development of long-term complications: The perspective of the Diabetes Control and Complications Trial. *Diabetes* **45**, 1289–1298.

EDLER, L. and BERGER, J., 1985: Computational methods to determine a break point in linear regression. *EDV in Medizin und Biologie* **16**, 128–134.

HOSMER, D. W. and LEMESHOW, S., 1989: *Applied Logistic Regression*. Wiley, New York.

JONES, R. H. and MOLITORIS, B. A., 1984: A statistical method for determining the breakpoint of two lines. *Analytical Biochemistry* **141**, 287–290.

KIMMERLE, R., BENDER, R., and BERGER, M., 1998: Prädiktion von *large-gestational-age*-Neugeborenen (LGA) bei Schwangerschaften von Typ-1-Diabetikerinnen: Nützlichkeit des HbA_{1c} (HPLC) im 3. Trimenon. *Diabetes und Stoffwechsel* **7** (Suppl. 1), 120–121.

KLEIN, R., KLEIN, B. E. K., MOSS, S. E., and CRUICKSHANKS, K. J., 1994: Relationship of hyperglycemia to the long-term incidence and progression of diabetic retinopathy. *Archives of Internal Medicine* **154**, 2169–2178.

KROLEWSKI, A. S., LAFFEL, L. M. B., KROLEWSKI, M., QUINN, M., and WARRAM, J. H., 1995: Glycosylated hemoglobin and the risk of microalbuminuria in patients with insulin-dependent diabetes mellitus. *New England Journal of Medicine* **332**, 1251–1255.

KÜCHENHOFF, H., 1997: An exact algorithm for estimating breakpoints in segmented generalized linear models. *Computational Statistics* **12**, 235–247.

LANGER, O., 1996: Is normoglycemia the correct threshold to prevent complications in the pregnant diabetic patient? *Diabetes Reviews* **4**, 2–10.

MEISTER, R., 1994: Biometrical basis and principles of quantitative risk assessment based on animal experiments – Models, benchmarks or NOELs for reproductive toxicity evaluation? *Informatik, Biometrie und Epidemiologie in Medizin und Biologie* **25**, 275–282.

MURRELL, J. A., PORTIER, C. J., and MORRIS, R. W., 1998: Characterizing dose-response I: Critical assessment of the benchmark dose concept. *Risk Analysis* **18**, 13–26.

MÜHLHAUSER, I., BENDER, R., BOTT, U., JÖRGENS, V., GRÜSSER, M., WAGENER, W., OVERMANN, H., and BERGER, M., 1996: Cigarette smoking and progression of retinopathy and nephropathy in type 1 diabetes. *Diabetic Medicine* **13**, 536–543.

ORCHARD, T. J., FORREST, K. Y. Z., ELLIS, D., and BECKER, D. J., 1997: Cumulative glycemic exposure and microvascular complications in insulin-dependent diabetes mellitus. The glycemic threshold revisited. *Archives of Internal Medicine* **157**, 1851–1856.

PASTOR, R. and GUALLAR, E., 1998: Use of two-segmented logistic regression to estimate change-points in epidemiologic studies. *American Journal of Epidemiology* **148**, 631–642.

REICHARD, P., 1995: Are there any glycemic thresholds for the serious microvascular diabetic complications? *Journal of Diabetes and Its Complications* **9**, 25–30.

SAS, 1985: *SAS/IML User's Guide, Version 5 Edition*. SAS Institute Inc., Cary, NC.

ULM, K., 1991: A statistical method for assessing a threshold in epidemiological studies. *Statistics in Medicine* **10**, 341–349.

Dr. RALF BENDER
Klinik für Stoffwechselkrankheiten und Ernährung
Heinrich-Heine-Universität Düsseldorf
Postfach 10 10 07
D-40001 Düsseldorf
Germany
E-mail: bender@uni-duesseldorf.de

Received, April 1998
Revised, January 1999
Accepted, February 1999