

Ordinal logistic regression in medical research

ABSTRACT – Medical research workers are making increasing use of logistic regression analysis for binary and ordinal data. The purpose of this paper is to give a non-technical introduction to logistic regression models for ordinal response variables. We address issues such as the global concept and interpretation of logistic models, the model building procedure from a practical point of view, and the assessment of the model adequacy. For illustrative purposes we apply these methods to real data of a study investigating the association between glycosylated haemoglobin and retinopathy. We give some recommendations for the use and assessment of ordinal logistic regression models in medical research.

The application of multiple regression models in medical research has greatly increased in recent years^{1,2}, especially the use of multiple linear regression for continuous response, logistic regression for categorical response, and Cox's proportional hazards model³ for censored response. These models allow one to analyse simultaneously the effect of several explanatory variables on a response variable. This means that adjustments for confounding factors can be made.

The standard logistic regression model is applicable only to binary (yes/no) response variables. The analysis of categorical data with more than two categories requires more complex methods⁴⁻⁸. Statistical standard software for a part of these models has been available for several years. The most popular method for ordinal data is the 'proportional odds model' described by McCullagh⁴.

Like all regression models, the proportional odds model makes assumptions about the nature of the relationship between the response variable and the prognostic factors⁹. If the data do not fulfil the assumptions, the results of a regression applied to them can be misleading or have no meaning at all. Nevertheless, investigating the goodness-of-fit of regression models is rarely done in medical research². It is not uncommon for investigators to throw data into a computer, select a program more or less at random, rummage through the computer output for some p values less than 0.05, and present the results without being aware of the assumptions, problems and pitfalls inherent in the methods used^{9,10}. On the other hand, readers of medical journals frequently do not

understand the basic concept of multiple regression models and are not able to interpret the results.

In this paper we give a non-technical introduction to the proportional odds model for ordinal data. We explain the relations between this model, the standard binary logistic regression model, and the general polytomous logistic model. Special emphasis is given to the model building procedure and the assessment of goodness-of-fit using standard statistical software. We illustrate the application of these methods by analysing real retinopathy data. Finally, we give some recommendations for the use of logistic regression models and reporting of results in medical research papers. We hope that our explanations will help doctors to understand and assess the results of ordinal logistic regression models published in medical journals.

Logistic regression models

Binary logistic regression

To understand ordinal logistic regression, one needs to understand the standard binary regression model. For simplicity, let us first consider the case where the effect of one explanatory variable (co-variate) X on the response variable Y is investigated. If the measurement levels of X and Y are continuous, for example if X = height and Y = forced expiratory volume in 1 second (FEV1), the simplest relationship between Y and X is a straight line given by the simple linear regression model

$$Y = a + \beta X$$

This model assumes that Y is, at least approximately, linearly related to X . If this assumption is invalid, the simple linear regression model is not applicable and other, more complicated, for example nonlinear, models should be considered.

If Y is not continuous but binary (ie only 1/0 type responses such as 'dead/alive' or 'dialysis yes/no' are possible), the simple linear regression model is invalid, because it assumes that Y can take any numeric value between minus infinity and plus infinity. Moreover, if Y is a binary variable, the usual assumption of homogeneous variance is violated¹¹. The key to describing the relationship between Y and X in a valid way is to model the probability of an event, ie $p = P(Y=1)$ instead of Y itself. While Y has only two possible values (1,0), p can take any numeric value between 0 and 1. The odds $p/(1-p)$ can take any positive value and the logarithm of the odds $\ln[p/(1-p)]$, called the *logit*, ranges from

RALF BENDER, PhD, Statistician, Department of Metabolic Diseases and Nutrition, Heinrich-Heine-University of Düsseldorf
ULRICH GROUVEN, PhD, Statistician, Department of Anaesthesiology, Hannover Medical School

minus infinity to plus infinity. Therefore, one can assume a linear relationship between the logit and X :

$$\text{logit} = \ln \left(\frac{p}{1-p} \right) = \alpha + \beta X$$

which is mathematically equivalent to the expression

$$p = \frac{\exp(\alpha + \beta X)}{1 + \exp(\alpha + \beta X)}$$

The term on the right-hand side of this equation is called a logistic function and hence the model is called a logistic regression model. The extension to multiple models is obtained by replacing βX with the linear combination $\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m$ of all covariates.

An important feature of the multiple binary logistic regression model is that odds ratios for the association between Y and X_j , adjusted for all other covariates, can be calculated directly from the logistic coefficients by $OR = \exp(\beta_j)$. However, this simple relationship is true only if the relation between the *logit* and X_j is in fact linear and there are no interactions between the covariates. Other important issues, such as estimation methods, test statistics and numerical algorithms for computations, are beyond the scope of this paper. A simple introduction to the standard binary logistic regression model, and related issues such as odds ratios and interactions, is given by Hall and Round¹², and there are a number of more technical and comprehensive reviews¹³⁻¹⁶.

Polytomous logistic regression

If the response variable Y is discrete with more than two categories, for example $Y = \text{marital status}$ defined in the 3 categories 'married', 'divorced, separated or widowed' and 'single', then the standard binary logistic regression model is not applicable. One possible way to handle such situations is to split the categorical response Y in several ways, for example $Y_1 = \text{'married yes/no'}$, $Y_2 = \text{'single yes/no'}$, and to apply binary logistic regression to each dichotomous variable. However, this will result in several different analyses for only one categorical response. A more structured approach is to formulate one model for the categorical response by means of so-called *generalised logits*. Suppose that Y has $k+1$ categories and the probability for category i is given by $P(Y=i) = p_i$ for $i=1, \dots, k+1$. Then the k *generalised logits* are defined by

$$\text{logit}(Y=i) = \ln \left(\frac{p_i}{1 - (p_1 + \dots + p_k)} \right) = \ln \left(\frac{p_i}{p_{k+1}} \right), \quad i=1, \dots, k$$

This means that the *generalised logits* relate the probabilities p_i for the categories $i=1, \dots, k$ to the reference category $k+1$.

For m covariates the general polytomous logistic regression model becomes

$$\text{logit}(Y=i) = \alpha_i + \beta_{i1} X_1 + \dots + \beta_{im} X_m, \quad i=1, \dots, k$$

Note that the polytomous logistic model is given by k equations if Y has $k+1$ categories and that we have one logistic coefficient β_{ij} for each category/covariate combination. Hence, it is not possible to summarise the effect of a covariate on the response Y by a single measure such as one odds ratio. Although the polytomous model offers the advantage of simultaneously testing the effect of a covariate on all response categories, polytomous logistic regression generates a cumbersome amount of statistical information which is difficult for physicians to understand. Further explanations of this class of models are given by Engel¹⁷ and DeMaris¹⁸.

Ordinal logistic regression

If the response variable Y is ordinal, the categories can be ordered in a natural way such as 'health status good/moderate/bad'. The polytomous logistic regression model can be applied but does not make use of the information about the ordering. One way to take account of the ordering is the use of *cumulative probabilities*, *cumulative odds* and *cumulative logits*. Considering $k+1$ ordered categories, these quantities are defined by

$$P(Y \leq i) = p_1 + \dots + p_i$$

$$\text{odds}(Y \leq i) = \frac{P(Y \leq i)}{1 - P(Y \leq i)} = \frac{p_1 + \dots + p_i}{p_{i+1} + \dots + p_{k+1}}$$

$$\text{logit}(Y \leq i) = \ln \left(\frac{P(Y \leq i)}{1 - P(Y \leq i)} \right), \quad i=1, \dots, k$$

The cumulative logistic model for ordinal response data is given by

$$\text{logit}(Y \leq i) = \alpha_i + \beta_{i1} X_1 + \dots + \beta_{im} X_m, \quad i=1, \dots, k$$

Like the polytomous logistic regression model, we have k model equations and one logistic coefficient β_{ij} for each category/covariate combination. Hence, the general cumulative logistic regression model contains a large number of parameters. However, in some cases a more parsimonious model is possible. If the logistic coefficients do not depend on i , we have only one common parameter β_j for each covariate. It follows that the *cumulative odds* are given by

$$\text{odds}(Y \leq i) = \exp(\alpha_i) \exp(\beta_1 X_1 + \dots + \beta_m X_m), \quad i=1, \dots, k$$

which means that the k odds for each cut-off category i differ only with regard to the intercepts α_i ; in other words, the odds are proportional. Hence, McCullagh⁴ used the term *proportional odds model*. The relatively stringent proportional odds assumption may be especially valid in cases where the ordinal response Y is related to an underlying latent continuous variable⁴,

for example if Y is a grouped continuous variable such as age groups or money income groups. However, categories assessed by an observer are another important type of ordinal variables. Such variables frequently occur in biomedical research. Anderson⁵ pointed out that for assessed ordinal response variables the proportional odds model is not flexible enough to cover the range of problems. He proposed a general class of models for ordinal data called 'stereotype ordered regression models' which include the proportional odds model as a special case⁵. A description of this class of models is beyond the scope of this paper. We believe that, owing to computational difficulties, it will be a long time before the stereotype models are applied in medical research. However, it should be kept in mind that the proportional odds model is the result of the stringent assumption of proportional odds, which is not automatically valid for all ordinal response variables.

The proportional odds model is now the most commonly used logistic regression model for ordinal response, for two reasons. First, it has the convenient feature that the effect of a covariate on Y can be quantified by one regression coefficient, and hence the calculation of one common odds ratio is possible; therefore, the presentation of results is short and simple. Second, standard statistical software with additional features such as stepwise variable selection procedures is now available for calculations^{19,20}.

Other approaches for logistic regression modelling of ordinal response variables can be found in the literature^{6-8,17,21}.

Assessing goodness-of-fit

A short introduction to what is meant by *goodness-of-fit* will underline the importance of assessing the adequacy of statistical models. The purpose of any regression model is to describe the relationship between a response and one or several covariates. Such models can be divided into a systematic component (the regression function) and an error component (the so-called residuals). The error component consists of the deviations of the data from the systematic part. If these residuals are 'large' then the model does not fit well and does not describe the data adequately. In that case, any conclusions drawn from this model are questionable. Hence, assessing goodness-of-fit plays a central role in the model building procedure and should be done before any hypotheses are tested. Important tools for assessing goodness-of-fit of regression models are the residuals and other comparisons of the observed response values with the corresponding predicted values. Assessing goodness-of-fit has two major parts: the global and the individual goodness of fit. Even when the global goodness-of-fit is adequate, there may be still some individual values that do not fit well.

The choice of an appropriate method to assess

goodness-of-fit depends on the regression model used. Harrell *et al*²² give an excellent overview of issues in developing multiple regression models and evaluating model assumptions and goodness-of-fit. We refer here only to the logistic regression models described above. Let us start with the binary logistic model. All goodness-of-fit methods compare the observed conditional event probabilities with the corresponding predicted probabilities. If there are only categorical covariates, and hence a limited number of different covariate patterns, the global goodness-of-fit can be examined by well known methods such as the Pearson chi-square statistic or the likelihood ratio statistic. However, if the number of covariate patterns is large, and hence the number of replicated measurements is small, these methods are invalid because they require a large number of replicated measurements⁷. Note that these methods always fail when the model contains a continuous covariate. Unfortunately, the procedure CATMOD of SAS²³ always prints the results of the chi-square and likelihood ratio goodness-of-fit tests without giving a warning message in cases when they are invalid.

Many current methods developed for logistic regression with continuous covariates are based on pooling the observations according to the predicted probabilities. The most important ones are the goodness-of-fit test of Hosmer and Lemeshow²⁴ implemented in SAS²⁵ and BMDP²⁰ and the test of Brown²⁶ implemented in BMDP²⁰.

Unfortunately, for polytomous and ordinal logistic regression models, no global goodness-of-fit test is yet available in standard statistical software packages. Hence, before the proportional odds model can be applied one should investigate the binary logistic regression models for each dichotomised response. Owing to the stringent model assumption, the proportional odds model is the wrong method to start a valid data analysis⁸. Only if the separate binary models are validated should one proceed and assess the adequacy of the proportional odds model. The proportional odds assumption can be formally tested by means of a score test implemented in SAS^{19,27}. The logistic modelling of the dichotomised responses is also helpful for assessing the validity of the proportional odds assumption²⁸.

A variety of other goodness-of-fit methods has been proposed in the literature, eg the computation of classification rates and graphical methods. To assess the individual goodness of fit, Pregibon²⁹ generalised a number of methods called *regression diagnostics* developed for linear regression to binary logistic regression models. An overview of these methods is given elsewhere^{13,22,30}.

Example

A 6-year follow-up study of type 1 diabetic patients documented the feasibility of translating an intensified

insulin treatment and teaching programme from a specialised university diabetes centre to general internal medicine departments^{31,32}. In this example only a small fraction of the data is considered, for demonstration purposes. For computations the SAS procedures MEANS, UNIVARIATE³³, FREQ²³, and LOGISTIC^{19,25} were used.

In this study, follow-up data for 600 diabetic patients (308 men, 292 women) were available for investigating the association between their retinopathy status at follow up (RSF) and the explanatory variable glycosylated haemoglobin (HbA₁). A complete analysis would require the investigation of other important covariates such as diabetes duration or blood pressure. In order to keep the example simple, only one covariate, namely HbA₁, is considered here. A more detailed analysis of the data is given by Mühlhauser *et al*³² and Bender and Grouven³⁴. Retinopathy status at follow-up was defined by the three ordered categories 'no retinopathy', 'nonproliferative retinopathy' and 'advanced retinopathy or blind'. Table 1 gives a descriptive analysis of the HbA₁ values in the retinopathy states. We dichotomise the response variable RSF by using the variables 'at least nonproliferative retinopathy' (RSF1) and 'at least advanced retinopathy' (RSF2).

The first step in the model building procedure should be a graphical check whether the *logits* of each dichotomised response are linearly related to the covariate HbA₁. This is the basic assumption of any

logistic regression model. To produce adequate plots, HbA₁ must be grouped into intervals so that each interval contains a sufficient number of observations. Unfortunately, such plots are not provided by the standard statistical software. Since many computational steps are necessary to produce these plots, this important step of the model building process is frequently neglected in practice. Figure 1 shows the plots of the *logits* of RSF1 and RSF2 in four HbA₁ groups (quartiles) versus the group midpoints. Both are

Fig 1. Plots of *cumulative logits* of retinopathy status in four HbA₁ groups versus the group midpoints, to check the assumption of linearity.

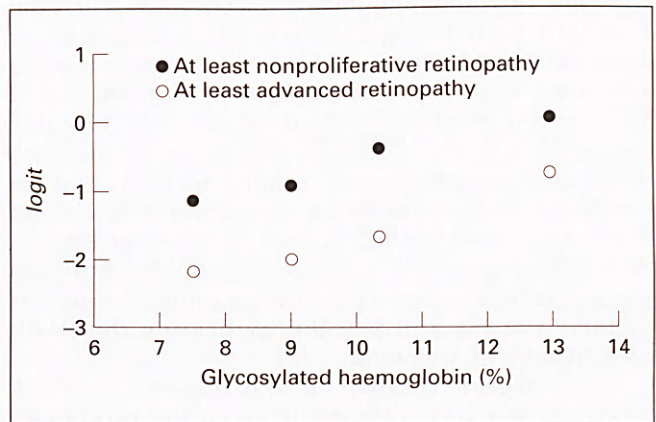


Table 1. Glycosylated haemoglobin and retinopathy status of 600 type 1 diabetic patients.

	Retinopathy status at follow-up			Total n=60
	None n=381	Nonproliferative n=114	Advanced/blind n=105	
HbA ₁ (%)	9.6 (2.1)	10.2 (2.3)	11.0 (2.2)	9.9 (2.2)

Data are given as means (SD).

Table 2. Results of the binary logistic regression models for dichotomised response retinopathy.

Covariate	Logistic coefficient	Standard error	p value	Odds ratio	95% CI
<i>The response 'at least nonproliferative retinopathy'</i>					
Intercept	-2.645	0.418	0.0001		
HbA ₁	0.415	0.086	0.0001	1.23	1.14-1.33
Hosmer-Lemeshow goodness-of-fit test: $p=0.231$					
<i>The response 'at least advanced retinopathy'</i>					
Intercept	-3.932	0.508	0.0001		
HbA ₁	0.232	0.047	0.0001	1.26	1.15-1.38

Hosmer-Lemeshow goodness-of-fit test: $p=0.466$.

Table 3. Results of the proportional odds model using retinopathy as response with three ordered categories.

Covariate	Logistic coefficient	Standard error	<i>p</i> value	Odds ratio	95% CI
Intercept 1	-3.726	0.411	0.0001		
Intercept 2	-2.684	0.395	0.0001		
HbA _{1c}	0.213	0.038	0.0001	1.24	1.15-1.33

Score test of the proportional odds assumption: $p=0.383$.

linear, which means that binary logistic regression is applicable for each dichotomised response. What if the *logits* of Y were not linearly related to X ? Then the standard logistic regression model would not be appropriate and one would have to find either a transformation $Z=g(X)$ so that the *logits* are linearly related to Z , or one should include quadratic (X^2) or cubic (X^3) effects in the model. In our example, the *logits* are approximately linearly related to HbA_{1c} and we could apply the standard logistic model to RSF1 and RSF2. The results of these models are given in Table 2. At first we should look at the goodness-of-fit of the two models. The Hosmer-Lemeshow goodness-of-fit test yielded $p=0.231$ and $p=0.466$, indicating that both models show no lack of fit.

As the logistic coefficients and the odds ratio for HbA_{1c} are not very different between the two binary models, it is possible to apply the proportional odds model to the ordinal response variable (Table 3). The score test for the proportional odds assumption yielded $p=0.383$, indicating that the model is appropriate for the data. What if the proportional odds assumption were not fulfilled by the data? Then more

complicated models should be used, which do not describe the effect of X on Y by means of a common odds ratio^{27,34,35}. In our example, the proportional odds model represents a valid description of the association between HbA_{1c} and the ordinal response retinopathy. The magnitude of the effect of HbA_{1c} on retinopathy can be described by the estimated odds ratio $OR=1.24$ (95% confidence interval: 1.1-1.3). The two lines shown in Fig 2 represent the continuous relationship between the cumulative probabilities of developing retinopathy and glycosylated haemoglobin, given by the estimated common logistic coefficient of the proportional odds model (Table 3). The dots are the observed probabilities in the four HbA_{1c} groups. Figure 2 gives an intuitive impression of how the probability of developing retinopathy increases with increasing HbA_{1c} level.

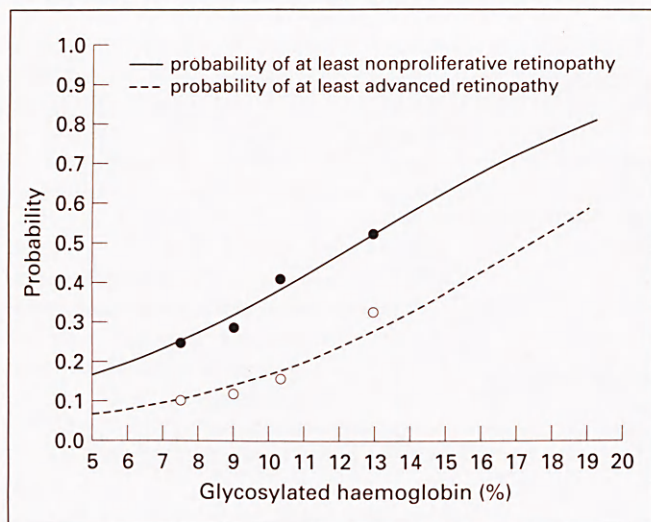
Conclusions and recommendations

Using ordinal logistic regression models in medical research in a valid way is not a simple task. A deep understanding of both the mathematical and the medical background is required. Finding a model that adequately describes the main features of the data is an interactive time-consuming process consisting of initial data analysis, graphical checks, choice and selection of covariates, parameter estimation and assessment of goodness-of-fit. It is impossible to understand and assess the results of such complex statistical methods when only some p values are reported.

We recommend that the following information should be given when results of ordinal logistic regression models are published:

- a thorough description of the response variable and all covariates
- the measurement scales and codes of all variables
- a clear description of which model has been used with a statistical reference
- the software used for computations
- how the covariates and interactions have been selected
- the overall goodness-of-fit of each dichotomised response

Fig 2. Plot of the cumulative probabilities of retinopathy status versus HbA_{1c} estimated from the proportional odds model.



- a statement of how the main model assumptions have been checked
- a table of the final model, containing the logistic coefficients, standard errors, *p* values, and odds ratios with confidence intervals.

Without this information readers are not able to assess the adequacy of the published results and conclusions from ordinal logistic regression models. Journals should provide enough space for the complete presentation of important results.

Acknowledgements

We thank Dr Mühlhauser for providing the data used in the example. The support of the Peter-Klöckner-Stiftung (Duisberg, Germany), through grants to Professor M Berger is gratefully acknowledged.

References

- Altman DG. Statistics in medical journals: development in the 1980s. *Stat Med* 1991;**10**:1897-913.
- Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. *Ann Intern Med* 1993;**118**:201-10.
- Cox DR. Regression models and life tables (with discussion). *J R Stat Soc B* 1972;**34**:187-220.
- McCullagh P. Regression models for ordinal data (with discussion). *J R Stat Soc B* 1980;**42**:109-42.
- Anderson JA. Regression and ordered categorical variables (with discussion). *J R Stat Soc B* 1984;**46**:1-30.
- Armstrong BG, Sloan M. Ordinal regression models for epidemiologic data. *Am J Epidemiol* 1989;**129**:191-204.
- Agresti A. Tutorial on modelling ordered categorical response data. *Psych Bull* 1989;**105**:290-301.
- Greenland S. Alternative models for ordinal logistic regression. *Stat Med* 1994;**13**:1665-77.
- Harrell FE Jr, Lee K, Matchar DB, Reichert TA. Regression models for prognostic prediction: advantages, problems, and suggested solutions. *Cancer Treat Rep* 1985;**69**:1071-7.
- Altman DG. The scandal of poor medical research. *Br Med J* 1994;**308**:283-4.
- Lipsitz SR, Buoncristiani JF. A robust goodness-of-fit test statistic with application to ordinal regression models. *Stat Med* 1994;**13**:143-52.
- Hall GH, Round AP. Logistic regression - explanation and use. *J R Coll Physicians Lond* 1994;**28**:242-6.
- Hosmer DW, Lemeshow S. *Applied logistic regression*. New York: Wiley, 1989.
- van Houwelingen JC, le Cessie S. Logistic regression: a review. *Stat Neerl* 1988;**42**:215-32.
- Imrey PB, Koch GG, Stokes ME. Categorical data analysis: some reflections on the log linear model and logistic regression. Part I. Historical and methodological overview. *Int Stat Rev* 1981;**49**:265-83.
- Lee J. An insight on the use of multiple logistic regression analysis to estimate association between risk factor and disease occurrence. *Int J Epidemiol* 1986;**15**:22-9.
- Engel J. Polytomous logistic regression. *Stat Neerl* 1988;**42**:233-52.
- DeMaris A. A framework for the interpretation of first order interaction in logit modelling. *Psych Bull* 1991;**110**:557-70.
- SAS. SAS[®] technical report P-200 SAS/STAT[®] software: CALIS and LOGISTIC procedures. Release: 6.04, Cary, NC: SAS Institute Inc, 1990.
- BMDP, BMDP statistical software manual. Vol. 2. Berkeley: University Press, 1990.
- Lee J. Cumulative logit modelling for ordinal response variables: applications to biochemical research. *Comput Appl Biosci* 1992;**8**:555-62.
- Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;**15**:361-87.
- SAS. SAS/STAT[®] guide for personal computers. Version 6 Edition. Cary, NC: SAS Institute Inc, 1987.
- Hosmer DW, Lemeshow S. Goodness-of-fit tests for the multiple logistic regression model. *Comm Stat A* 1980;**9**:1043-69.
- SAS. SAS[®] technical report P-229 SAS/STAT[®] software: changes and enhancements. Release 6.07, Cary, NC: SAS Institute Inc, 1992.
- Brown CC. On a goodness-of-fit test for the logistic model based on score statistics. *Comm Stat A* 1982;**11**:1087-105.
- Peterson B, Harrell FE Jr. Partial proportional odds models for ordinal response variables. *Appl Stat* 1990;**39**:205-17.
- Brant R. Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics* 1990;**46**:1171-8.
- Pregibon D. Logistic regression diagnostics. *Ann Stat* 1981;**9**:705-24.
- Hosmer DW, Taber S, Lemeshow S. The importance of assessing the fit of logistic regression models: a case study. *Am J Public Health* 1991;**81**:1630-5.
- Jörgens V, Grüsser M, Bott U, Mühlhauser I, Berger M. Effective and safe translation of intensified insulin therapy to general internal medicine departments. *Diabetologia* 1993;**36**:99-105.
- Mühlhauser I, Bender R, Bott U, Jörgens V, Grüsser M, Wagener W, Overmann H, Berger M. Cigarette smoking and progression of retinopathy and nephropathy in type 1 diabetes. *Diabetic Med* 1996;**13**:536-43.
- SAS. SAS[®] procedures guide for personal computers. Version 6 Edition. Cary, NC: SAS Institute Inc; 1985.
- Bender R, Grouven U. Using binary logistic regression models for ordinal data with non-proportional odds. *J Clin Epidemiol* (submitted for publication).
- Woodward M, Laurent K, Tunstall-Pedoe H. An analysis of risk factors for prevalent coronary heart disease by using the proportional odds model. *Statistician* 1995;**44**:69-80.

Address for correspondence: Ralf Bender, Statistician, Department of Metabolic Diseases and Nutrition, Heinrich-Heine-University of Düsseldorf, P.O. Box 10 10 07, D-40001 Düsseldorf, Germany.