Biometrical Journal **42** (2000) 6, 677–699

Calculating Ordinal Regression Models in SAS and S-Plus

RALF BENDER

Institute of Epidemiology and Medical Statistics
School of Public Health
University of Bielefeld
Germany

AXEL BENNER

Biostatistics
German Cancer Research Center
Heidelberg
Germany

Summary

Although a number of regression models for ordinal responses have been proposed, these models are not widely known and applied in epidemiology and biomedical research. Overviews of these models are either highly technical or consider only a small part of this class of models so that it is difficult to understand the features of the models and to recognize important relations between them. In this paper we give an overview of logistic regression models for ordinal data based upon cumulative and conditional probabilities. We show how the most popular ordinal regression models, namely the proportional odds model and the continuation ratio model, are embedded in the framework of generalized linear models. We describe the characteristics and interpretations of these models and show how the calculations can be performed by means of SAS and S-Plus. We illustrate and compare the methods by applying them to data of a study investigating the effect of several risk factors on diabetic retinopathy. A special aspect is the violation of the usual assumption of equal slopes which makes the correct application of standard models impossible. We show how to use extensions of the standard models to work adequately with this situation.

Key words: Logistic Regression; Ordinal Data; Proportional Odds Model; Continuation Ratio Model; Equal Slopes Assumption; SAS; S-Plus; Diabetic Retinopathy.

Zusammenfassung

Obwohl verschiedene Regressionsmodelle für ordinale Zielvariablen existieren, sind diese Modelle in der epidemiologischen und biomedizinischen Forschung noch nicht sehr verbreitet. Übersichtsarbeiten hierzu sind entweder sehr technisch ausgerichtet oder beinhalten nur einen kleinen Teil dieser Modelle, so daß es schwierig ist deren Eigenschaften und Beziehungen untereinander zu verstehen. In dieser Arbeit geben wir einen Überblick über logistische Regressionsmodelle für ordinale Daten, die auf ku-

mulativen und bedingten Wahrscheinlichkeiten basieren. Wir zeigen, wie die bekanntesten Modelle, nämlich das Proportional Odds Modell und das Continuation Ratio Modell, sich in den Rahmen der verallgemeinerten linearen Modelle einbetten lassen. Wir beschreiben die Eigenschaften und Interpretationen dieser Modelle und zeigen, wie sie sich mit Hilfe von SAS und S-Plus berechnen lassen. Die Modelle werden veranschaulicht und verglichen, indem sie auf Daten einer Studie angewendet werden, in welcher der Einfluß von Risikofaktoren auf die diabetische Retinopathie untersucht wird. Ein spezieller Aspekt ist die Verletzung der gewöhnlichen Annahme identischer Regressionskoeffizienten, welche die sinnvolle Anwendung von Standardmodellen unmöglich macht. Es werden Möglichkeiten aufgezeigt, wie Erweiterungen der gewöhnlichen Modelle für eine adäquate Datenanalyse verwendet werden können.

1. Introduction

Many variants of regression models for analyzing ordinal response variables have been developed and described during the past years (McCULLAGH, 1980; ANDERSON, 1984; GREENLAND, 1985; ASHBY, POCOCK, and SHAPER 1986; GREENWOOD and FAREWELL, 1988; McCULLAGH and NELDER, 1989; AGRESTI, 1989; ARMSTRONG and SLOAN, 1989; ASHBY, WEST, and AMES 1989; HASTIE, BOTHA, and SCHNITZLER 1989; PETERSON and HARRELL, 1990; HOLTBRÜGGE and SCHUMACHER, 1991; LEE, 1992; GREENLAND, 1994; COX, 1995; ANANTH and KLEINBAUM, 1997; COX, 1997; SCOTT, GOLDBERG, and MAYO 1997; BENDER and GROUVEN, 1998; HARRELL et al., 1998b). Compared to frequently used methods for binary and nominal data ordinal regression models have the advantage that they make full use of ranked data (LEE, 1992; SCOTT et al., 1997; ANANTH and KLEINBAUM, 1997). Nevertheless, these models have been underutilized in biomedical and epidemiological research. As standard statistical software for some of the models is available for several years the main reasons for the underutilization of ordinal regression models may be problems in understanding and interpreting of these models and their results. Additionally, it is frequently not clear how available software programs can be used to perform the calculations for these models.

In this paper the two most popular families of ordinal regression models based upon cumulative and conditional probabilities are described and explained in detail. It is shown how the proportional odds model (McCULLAGH, 1980) and the continuation ratio model (McCULLAGH and NELDER, 1989) are embedded in the general framework of generalized linear models. The characteristics, connections, and interpretations of the models for practical applications are outlined. It is shown in detail how the required calculations can be performed by means of SAS and S-Plus.

2. Methods

2.1 *The Framework of Generalized Linear Models*

To understand the principle of the class generalized linear models (McCULLAGH and NELDER, 1989) let us at first consider a binary response variable *Y*. For simpli-

city, we consider only one explanatory factor X (binary or continuous). The question of an investigator may then be whether X has an effect on Y. As Y itself has only 2 values (1 = yes and 0 = no) we analyze whether the probability of an event $\pi(x) = P(Y = 1 \mid X = x)$ is associated with X by means of an appropriate model. The class of generalized linear models in this case is given by

$$f\{\pi(x)\} = \alpha + \beta x \tag{1}$$

where f is an appropriate function (called link function), α is the intercept, and β is the regression coefficient for X. In the case of $m \geq 2$ explanatory factors $X_1,\ldots, X_m$, βx is replaced by the linear combination $\beta_1 x_1 + \ldots + \beta_m x_m$. For the analysis of binary and ordinal response data the following two link functions have been widely used.

1. the logit link: $$f(\pi) = \log\{\pi/(1-\pi)\} \tag{2}$$

2. the complementary log-log link: $$f(\pi) = \log\{-\log(1-\pi)\}. \tag{3}$$

More link functions are discussed (McCullagh and Nelder, 1989), but in this paper we consider only these two. The logit link leads to the ordinary logistic regression model since $\log\{\pi(x)/(1\pi(x))\} = \alpha + \beta x$ is equivalent to

$$\pi(x) = \exp(\alpha + \beta x)/\{1 + \exp(\alpha + \beta x)\}, \tag{4}$$

which is the logistic function. The complementary log-log (cloglog) link leads to the model in which π and X are related via the extreme value function since $\log\{-\log(1-\pi(x))\} = \alpha + \beta x$ is equivalent to

$$\pi(x) = 1 - \exp\{-\exp(\alpha + \beta x)\}. \tag{5}$$

2.2 *Grouped Continuous Models*

Let Y now be a categorical response variable with k+1 ordered categories. We consider a response variable such as 'disease status' given by ordered categories with higher values belonging to more severe disease states, for example

$$Y = \begin{cases} 0 & = & \text{healthy} \\ 1 & = & \text{slightly ill} \\ 2 & = & \text{moderately ill} \\ \vdots & = & \vdots \\ k & = & \text{seriously ill} \end{cases} \tag{6}$$

Let $\pi_j(x) = P(Y = j \mid X = x)$ be the probability for the realization of $Y = j$ given $X = x$, $j = 0,1,\ldots, k$. The class of grouped continuous models is based upon the cumulative probabilities

$$\gamma_j(x) = \mathrm{P}(\mathrm{Y} \geq j \mid X = x) = \pi_j(x) + \ldots + \pi_k(x), \qquad j = 1,\ldots, k. \tag{7}$$

The name "grouped continuous model" can be explained by the view that Y is a discretized variable of an underlying latent continuous trait defined by cut-off points j. It is then natural to formulate a model by means of the cumulative probabilities γ_j. It is, however, not essentially necessary to suppose the existence of an underlying continuous variable in order to use the cumulative probabilities for the description of ordered categories. The class of grouped continuous models is obtained by the generalized linear model (1) in which the cumulative probabilities are used instead of π

$$f\{\gamma_j(x)\} = \alpha_j + \beta_j x, \qquad j = 1,\ldots, k. \tag{8}$$

Note that we have k model equations and k regression coefficients to describe the relationship between Y and X. The standard assumption in most applications is that the regression coefficient does not depend on j, that means the model

$$f\{\gamma_j(x)\} = \alpha_j + \beta x, \qquad j = 1,\ldots, k \tag{9}$$

is considered. Thus, it is assumed that for the considered link function f the corresponding regression coefficients are equal for each cut-off point j. The adequacy of this 'equal slopes assumption' has to be evaluated carefully before this model can be applied. We return to this important issue in the application section and concentrate now on the different models we get for the different link functions.

Different models result from the use of different link functions. If the logit link (2) is used model (9) becomes

$$\begin{aligned} \log\{\gamma_j(x)/(1\gamma_j(x))\} &= \alpha_j + \beta x \\ \Leftrightarrow \gamma_j(x) &= \exp(\alpha_j + \beta x)/\{1 + \exp(\alpha_j + \beta x)\}, \end{aligned} \tag{10}$$

which is the well-known proportional odds (PO) model (McCULLAGH, 1980), also called ordinal logistic model (SCOTT et al., 1997), cumulative logit model (LEE, 1992; ANANTH and KLEINBAUM, 1997), cumulative odds model (ARMSTRONG and SLOAN, 1989; GREENLAND, 1994), or McCullagh's grouped continuous model (GREENWOOD and FAREWELL, 1988). The number of different names for the same models causes unnecessarily much confusion about ordinal regression models. If in the case of several explanatory factors $X_1,\ldots, X_m$ the more general model (8) or a mixture of (8) and (9) is used in which the equal slopes assumption holds for some of the explanatory factors, the resulting model is the partial proportional odds (PPO) model (PETERSON and HARRELL, 1990).

If the complementary log-log link (3) is used model (9) becomes

$$\log\{\log(1\gamma_j(x))\} = \alpha_j + \beta x \Leftrightarrow \gamma_j(x) = 1 - \exp\{-\exp(\alpha_j + \beta x)\}, \tag{11}$$

which is the discrete proportional hazards (PH) model (McCULLAGH, 1980; GREENLAND, 1994). Corresponding to the PPO model, model (11) releasing the equal slopes assumption for some of several explanatory factors may be called partial proportional hazards (PPH) model.

The difference between model (10) and model (11) is simply the use of different scales. Both models have similar properties and are in practice frequently indistinguishable. The PO model, however, has some appealing features. At first, it is invariant under a reversal of the categories, as only the signs of the regression coefficients change. Secondly, it is invariant under collapsibility of the ordered categories, as the regression coefficients do not change when response categories are collapsed or the category definitions are changed. Thirdly, it produces the most easily interpretable regression coefficients, as $\exp(\beta)$ is the homogenous odds ratio over all cut-off points summarizing the effect of the explanatory factor X on the response Y in one single frequently used measure. Due to these reasons, the PO model is by far the most used regression model for ordinal data.

2.3 *Continuation Ratio Models*

The class of continuation ratio (CR) models is based upon the conditional probabilities of being in category j among all subjects who are in category j or lower

$$\begin{aligned}\delta_j(x) &= \mathrm{P}(Y = j \mid Y \leq j, X = x) = \mathrm{P}(Y = j \mid X = x)/\mathrm{P}(Y \leq j \mid X = x) \\ &= \pi_j(x)/[\pi_0(x) + \ldots + \pi_j(x)], \qquad j = 1,\ldots, k. \end{aligned} \tag{12}$$

The δ_j are called continuation ratios. In this paper we use the backward formulation of the continuation ratios although the forward continuation ratios $\varphi_j(x) = \mathrm{P}(Y = j \mid Y \geq j, X = x)$ are used more frequently. The use of forward continuation ratios makes sense especially if the response Y represents discrete survival times. As high risk patients have short survival times, forward continuation ratios represent a comparison of high risk patients with low risk patients in this case. However, if Y represents a disease status given by ordered categories with higher values belonging to more severe disease states, the opposite is true. We want to estimate the odds of severe disease states compared with lower disease states and not vice versa. Hence, in the case considered here we need the backward continuation ratios. The two models for the backward and forward continuation ratios are not equivalent and yield different results. Thus, one has to be careful which continuation ratios should be used.

The class of CR models is obtained by the generalized linear model (1) in which the continuation ratios δ_j are used instead of π

$$f\{\delta_j(x)\} = \alpha_j + \beta_j x, \qquad j = 1,\ldots, k. \tag{13}$$

Similar to the grouped continuous models the basic assumption in most applications is the equal slopes assumption, that means the model

$$f\{\delta_j(x)\} = \alpha_j + \beta x, \qquad j = 1,\ldots, k \tag{14}$$

with a homogenous regression coefficient for X is considered. If the logit link (2) is used model (14) becomes

$$\log \{\delta_j(x)/(1 - \delta_j(x))\} = \alpha_j + \beta x$$
$$\Leftrightarrow \delta_j(x) = \exp (\alpha_j + \beta x)/\{1 + \exp (\alpha_j + \beta x)\}, \qquad (15)$$

which is called continuation ratio model (GREENWOOD and FAREWELL, 1988; ARMSTRONG and SLOAN, 1989; ANANTH and KLEINBAUM, 1997) or logistic continuation ratio model (GREENLAND, 1994) or proportional logit hazard model (COX, 1997). If in the case of several explanatory factors $X_1,\ldots, X_m$ the more general model (14) or a mixture of (14) and (15) is used in which the equal slopes assumption is released for some of the X_l the resulting model is called extended continuation ratio (ECR) model (HARRELL et al., 1998b).

If the complementary log-log link (3) is used model (14) becomes

$$\log \{-\log (1 - \delta_j(x))\} = \alpha_j + \beta x \Leftrightarrow \delta_j(x) = 1 - \exp \{-\exp (\alpha_j + \beta x)\}, \qquad (16)$$

which is also called continuation ratio model. An interesting connection between the class of grouped continuous models and the class of continuation ratio models is given by the fact that model (11) and model (16) are equivalent in the sense that the same estimates for the regression coefficient will be obtained if both models are fitted to the same data (LÄÄRÄ and MATTHEWS, 1985).

3. Computational Issues

3.1 *General Comments*

While a number of statistical software packages provide binary logistic regression at least for the logit link, regression models for ordinal data are not so often implemented. However, it is possible to use any software for binary logistic regression to estimate the parameters of CR models. The CR models (14) are based upon conditional probabilities for specific patient strata, which are conditionally independent. If the original data set is appropriately restructured by repeatedly including the corresponding data subset and two new variables are added, the cut-point and the dichotomous status of the response at that cut-point, binary logistic regression can be applied to the restructured data (ARMSTRONG and SLOAN, 1989; BERRIDGE and WHITEHEAD, 1991; SCOTT et al., 1997). In the following, the data restructuring for the backward continuation ratios is described.

Let Y be a categorical response variable with $k + 1$ ordered categories $0,1,\ldots, k$ and corresponding frequencies $n_0, n_1,\ldots, n_k$. The first subset of data for the continuation ratio of the highest category k ($Y \leq k$) contains all observations of the original data set with $\delta_k = P(Y = k \mid Y \leq k) = \pi_k$. The value of the new cut-point variable CP defining this subset is k for all observations. The n_k observations with

response category $Y = k$ are assigned the new binary response $BR = 1$, and the remaining observations are assigned a value of $BR = 0$. The second subset of data contains only the observations with $Y \leq k - 1$ because $\delta_{k-1} = P(Y = k - 1 \mid Y \leq k - 1)$. The value of CP is now $k - 1$, the n_{k-1} observations with response category $Y = k - 1$ are assigned the binary response $BR = 1$, and the remaining observations are assigned a value of $BR = 0$. This procedure continues until category $j = 1$. The last data subset contains the observations with $Y = 1$ and $Y = 0$. The values for the new variables are $CP = 1$ and $BR = 1$ if $Y = 1$ and $BR = 0$ if $Y = 0$. The restructured data set contains a total of $k(n_0 + n_1) + (k - 1)\, n_2 + (k - 2)\, n_3 + \ldots + n_k$ observations. Applying binary logistic regression to the restructured data set with BR as response and $k - 1$ dummy variables representing the cut-point variable yields the regression coefficients of the continuation ratio model for ordinal data.

The use of binary logistic regression to the restructured data has the advantage that it is possible to work with the more general models (13) without the usual but restrictive equal slopes assumption. By including interactions between the dummy variables describing the strata and the considered explanatory factor X, a model with different regression coefficients for each cut-point is obtained. This allows to fit models where the effects of a subset of explanatory factors is different for different levels of Y but where another subset of explanatory factors could be modeled assuming equal slopes. Such a procedure is not possible for the grouped continuous models. For calculations of the models classes (8) and (9) appropriate software capable of ordinal regression models is required.

3.2 *Computation by Means of SAS*

Several procedures of SAS can be used for binary and ordinal logistic regression models. The procedure PROBIT (SAS, 1987) was developed to analyze binomial and multinomial biological assay data. It can also be used for other discrete event data but has no specific features for ordinal response data and is not considered here. The procedure CATMOD (SAS, 1987) provides a wide variety of categorical data analyses including logistic regression for binary, multinomial and ordinal response. CATMOD is, however, mainly an analysis of variance procedure for multinomial response data. The standard SAS procedure for binary and ordinal logistic regression models is LOGISTIC (SAS, 1990), which is considered here.

To calculate logistic regression models by means of LOGISTIC principally only a few SAS statements are required. A typical SAS procedure call is given by

```
PROC LOGISTIC DATA=SASDATA DESCENDING;
  MODEL Y = X1 X2 X3 / LINK = LOGIT;
RUN;
```

Hence, in SAS one has only to invoke the LOGISTIC procedure, to specify the SAS data set containing the data, and to specify the binary or ordinal response

variable Y and the explanatory factors X_l. The explanatory factors must be continuous or binary. As LOGISTIC contains no CLASS statement, multinomial explanatory factors can only be included by means of dummy variables created in a preceding data step. In most applications one has to specify the DESCENDING option in the PROC statement because the LOGISTIC procedure models by default the probability of $Y = 0$ in the case of a binary response. The LINK option in the model statement is only required if other than the logit link should be used. For the complementary log-log link one has to specify LINK=CLOGLOG. Note that in the case of an ordinal response LOGISTIC automatically calculates the PO model (10) if LINK=LOGIT or the PH model (11) if LINK=CLOGLOG is specified, regardless of whether the equal slopes assumption is fulfilled by the data or not.

The CR models can be calculated by means of LOGISTIC if the original data set is restructured by means of SAS data steps. The data restructuring for the backward continuation ratios in the case of $k + 1 = 4$ ordered categories 0, 1, 2, 3 can be achieved by means of the following SAS code.

```
DATA CR1; SET SASDATA; IF Y<=3; CP=3; IF Y=3 THEN BR=1; ELSE BR=0;
DATA CR2; SET SASDATA; IF Y<=2; CP=2; IF Y=2 THEN BR=1; ELSE BR=0;
DATA CR3; SET SASDATA; IF Y<=1; CP=1; IF Y=1 THEN BR=1; ELSE BR=0;
DATA CRM; SET CR1 CR2 CR3;
  IF CP=3 THEN CP3=1; ELSE CP3=0;
  IF CP=2 THEN CP2=1; ELSE CP2=0;
RUN;
```

The CR model (15) using the logit link can then be calculated by means of

```
  PROC LOGISTIC DATA=CRM DESCENDING;
    MODEL BR = CP2 CP3 X1 X2 X3 / LINK=LOGIT;
  RUN;
```

The CR model (16) using the complementary log-log link can be calculated by the same SAS code using the MODEL option LINK=CLOGLOG. This procedure will result in the same parameter estimates as the calculations for model (11) by applying PROC LOGISTIC to the original SASDATA. However, one has to be very careful of using the correct continuation ratios, the correct definition of the response, and the correct use of the DESCENDING option in LOGISTIC to get the correct result. With two possible continuation ratios (backward and forward), two possible definitions of the response (1 and 0), and two possible uses of the DESCENDING option (yes and no) there are 8 possibilities to calculate CR models in SAS yielding 4 different model variants from which only one is correct.

The class of ECR models can be calculated by including the corresponding product terms between the CP_j and X_l variables. These product terms have to be produced in a preceding SAS data step.

3.3 *Computation by Means of S-Plus*

Ordinal logistic regression is not part of standard S-Plus but can be calculated via the *Design* library by use of the function *lrm* (HARRELL, 1998a) . For binary and proportional odds logistic models general penalized maximum likelihood estimation is used. A sample function call is given by

```
lrm(y ~ x1 + x2 + x3, data=splusdata).
```

Interaction terms or nonlinear functions of the predictors in the model formula are supported, e.g. specifying a smoothing spline fit of $x1$ and including and interaction term for $x2$ and $x3$ can be done by reformulating

```
lrm(y ~ s(x1) + x2 + x3 + x2:x3, data=splusdata).
```

Continuation ratio logistic models can be applied by using the data restructuring procedure described above. The function *cr.setup* as provided with the *Design* library does this for forward CR automatically. To get the backward continuation ratios one has to reverse the order of the response variable *Y.* A function *crback.setup* to be used for backward CR models is provided by the authors. The backward CR model can then be fit to the data using the standard *glm* function (family binomial) or the *lrm* function from the *Design* library, e.g.

```
attach(splusdata)
up <- crback.setup(y)
y <- up$y
cohort <- up¢ohort
lrm(y ~ cohort + x1 + x2 + x3, data=splusdata[up$subs,])
```

or

```
glm(y ~ cohort + x1 + x2 + x3, family=binomial, data=splusdata[up$subs,]).
```

3.4 *Model Diagnostics Offered by SAS*

Diagnostic tools to assess the goodness-of-fit of the calculated model are offered by SAS only for binary response models. The option LACKFIT in the MODEL statement performs the Hosmer and Lemeshow goodness-of-fit test (HOSMER and LEMESHOW, 1980). The regression diagnostics developed by PREGIBON (1981) can be requested by specifying the INFLUENCE option. The IPLOTS option requests the printing of an index plot (a scatter plot with case number as the horizontal axis) for each regression diagnostic statistic. Although these tools are available only for binary response models they represent valuable tools for model building in the case of ordinal responses as they can be applied in a preceding step to each dichotomized response.

In the case of ordinal responses no diagnostic tools to assess the global goodness-of-fit are offered by SAS. However, the score test for testing the equal slopes

assumption developed by PETERSON and HARRELL (1990) is calculated and printed by default. In the case of the PO model the test is called ‘‘score test for the proportional odds assumption,. In the case of the PH model the term ‘‘score test for the equal slopes assumption, is used. As the score test is anti-conservative and often yields p-values which are far too small (HARRELL et al., 1998b) other techniques are required to assess the validity of the equal slopes assumption. Especially, graphical techniques are useful, which can be produced more comfortably by means of S-Plus.

3.5 *Model Diagnostics Offered by S-Plus*

Ordinary PO and CR models assume equal slopes for all levels of the response for each explanatory factor. A simple graphical check is done by plotting the means of the explanatory factors for each level of the response together with the expected values of the explanatory factors given the level of the response under the ordinal logistic regression model which includes only this one explanatory factor. The means should be ordered if the equal slopes assumption is true.

Let y_i and x_i for $i = 1,\ldots,n$ are the observed values of n subjects of the response Y and the explanatory factor X, respectively. The expected value of X given $Y = j$ is then estimated by

$$\hat{E}(X \mid Y = j) = \sum_{i=1}^{n} x_i \frac{\hat{\pi}_j(x_i)}{\sum_{i=1}^{n} I(y_i = j)} \tag{17}$$

since

$$E(X \mid Y = j) = \sum_{x} xP(X = x \mid Y = j) \tag{18}$$

and

$$P(X = x \mid Y = j) = \frac{P(Y = j|X = x)\, P(X = x)}{P(Y = j)} \tag{19}$$

A function *plot.xmean.ordinaly* which produces plots of the means of all predictors X given the levels of Y together with their expected values for PO and CR models is provided with the *Design* package for the S-language (HARRELL, 1998a). A revised version of this function which includes backward CR models is provided by the authors. For plotting the expected values of the predictors for PO and CR models, and for simulations we need to compute the estimates of the probabilities $\pi_j(x) = \mathrm{P}(Y = j \mid X = x)$. In the situation of PO models this is done straightforward by

$$\begin{aligned} P(Y = j \mid X = x) &= P(Y \geq j \mid X = x)\; P(Y \geq j + 1 \mid X = x) \\ P(Y = k \mid X = x) &= P(Y \geq k \mid X = x). \end{aligned} \tag{20}$$

For CR models we get

$$P(Y = k \mid X = x) = P(Y = k \mid Y \leq k, X = x)$$

$$P(Y = k-1 \mid X = x) = P(Y = k-1 \mid Y \leq k-1,\ X = x)\ P(Y \leq k-1 \mid X = x)$$

$$= P(Y = k-1 \mid Y \leq k-1,\ X = x)\ [1 - P(Y = k \mid X = x)]$$

and then recursively

$$P(Y = j \mid X = x) = P(Y = j \mid Y \leq j, X = x) \left[1 - \sum_{i=j+1}^{k} P(Y = i \mid X = x)\right],$$

$$j = 0,\ldots, k-1 \tag{21}$$

If linearity and additivity of the explanatory factors are correct then the odds ratio of two levels of an explanatory factor is always independent from levels of *Y*. In this situation partial residuals are useful for the graphical analysis of ordinal logistic regression. Partial residuals for the *l*th explanatory factor, the *i*th subject, and the *j*th response level are given by

$$r_{lij} = \hat{\beta}_l x_{li} + \frac{I(y_i \geq j) - \hat{\pi}_{ij}}{\hat{\pi}_{ij}(1 - \hat{\pi}_{ij})}\,. \tag{22}$$

The function *residuals.lrm* from the *Design* library computes ordinary, score, Pearson, deviance and partial residuals for a PO model by use of option *'type'*. Plotting methods are available for score and partial residuals by use of the option *'plot'*. In case of partial residuals smoothed scatterplots are drawn, where the default smoothing technique is *lowess*

```
fit.po <- lrm(y ~ x1 + x2 + x3, data=splusdata)
residuals(fit.po, type="partial", plot=T).
```

Using scatterplot smoothing can help to find out which explanatory factor may not be linearly and/or ordinally related to the response. Plots of binary logistic score residuals produced e.g. by

```
residuals(fit.po, type="binary.score", plot=T)
```

are also useful to check the PO assumption.

The function *plot.lrm.partial* plots partial residuals for a sequence of binary logistic model separately for each predictor. It can be used to check CR models in the same way as *residuals.lrm* is used for PO models, e.g.

```
f2 <- lrm(y=2 ~ x1 + x2 + x3, data=splusdata)
f1 <- lrm(y=1 ~ x1 + x2 + x3, data=splusdata, subset=y<=1)
plot.lrm.partial(f2, f1, center=T) .
```

4. Application of Ordinal Regression Models

4.1 *Data and Methods*

The data come from a 6-year follow-up study of type 1 diabetic patients participating in an intensified insulin treatment and teaching program (JÖRGENS et al., 1993; MÜHLHAUSER et al., 1996). In this example only a small fraction of the variables is considered for demonstration purposes. In this study follow-up data of 613 diabetic patients are available and an interesting question is whether there are associations between the retinopathy status at follow-up (*RS*) and the risk factor smoking (*SM*), adjusted for the known risk factors diabetes duration (*DD*), glycosylated hemoglogin (*H1C*), and diastolic blood pressure (*DBP*). An adjustment for retinopathy status at baseline would also be possible but this would perhaps dilute the effect of smoking if smoking and baseline retinopathy are associated. Hence, it was decided not to adjust for baseline retinopathy in this analysis. *RS* is defined by $k + 1 = 3$ ordered categories $0 =$ 'no retinopathy', $1 =$ 'nonproliferative retinopathy', and $2 =$ 'advanced retinopathy or blindness'. *SM* is a binary variable indicating whether a patient smoked during the study period or not. *DD* is a continuous variable measured at follow-up. For *H1C* and *DBP* their mean values for the study period are used (5 measurements: baseline, 1, 2, 3, and 6 year follow-up). In Table 1 a descriptive overview of the data is given.

BENDER and GROUVEN (1998) used these data to compare several logistic regression models (PO model, separate binary logistic regression models, polytomous logistic regression, PPO model) in the case of non-proportional odds. The equal slopes assumption is not fulfilled for the risk factor smoking (BENDER and GROUVEN, 1998). Hence, the application of the PO model (10) yielded misleading results. As the separate binary regression approach is very close to that of the PPO model the careful application of separate binary logistic regression models represents a simple and adequate tool to analyze ordinal data with non-proportional

Table 1
Descriptive statistics of the diabetes data

Variable name	Variable label	Retinopathy status at follow-up (*RS*)			Total
		None $n = 388$	Nonproliferative $n = 118$	Advanced/blind $n = 107$	$n = 613$
	Any retinopathy at baseline	19 (5%)	20 (17%)	74 (69%)	113 (18%)
SM	Ever smoking	197 (51%)	76 (64%)	52 (49%)	325 (53%)
DD	Diabetes duration (yr)	12.7 (6.7)	16.7 (6.1)	21.6 (6.5)	15.0 (7.4)
H1C	Glycosylated hemoglobin (%)	7.5 (1.2)	8.0 (1.2)	8.3 (1.4)	7.8 (1.3)
DBP	Diast. blood pressure (mmHg)	78.6 (6.6)	80.9 (6.7)	84.5 (7.4)	80.1 (7.1)

data are given as means (SD) or numbers (%)

odds. In this paper, additionally the class of CR models with and without the equal slopes assumption is investigated and compared with the other approaches presented by BENDER and GROUVEN (1998).

4.2 *Results*

The results of all standard models (PO, PH, and CR) are quite similar even for the estimated regression coefficients although they have a different meaning in the different models (Table 2).

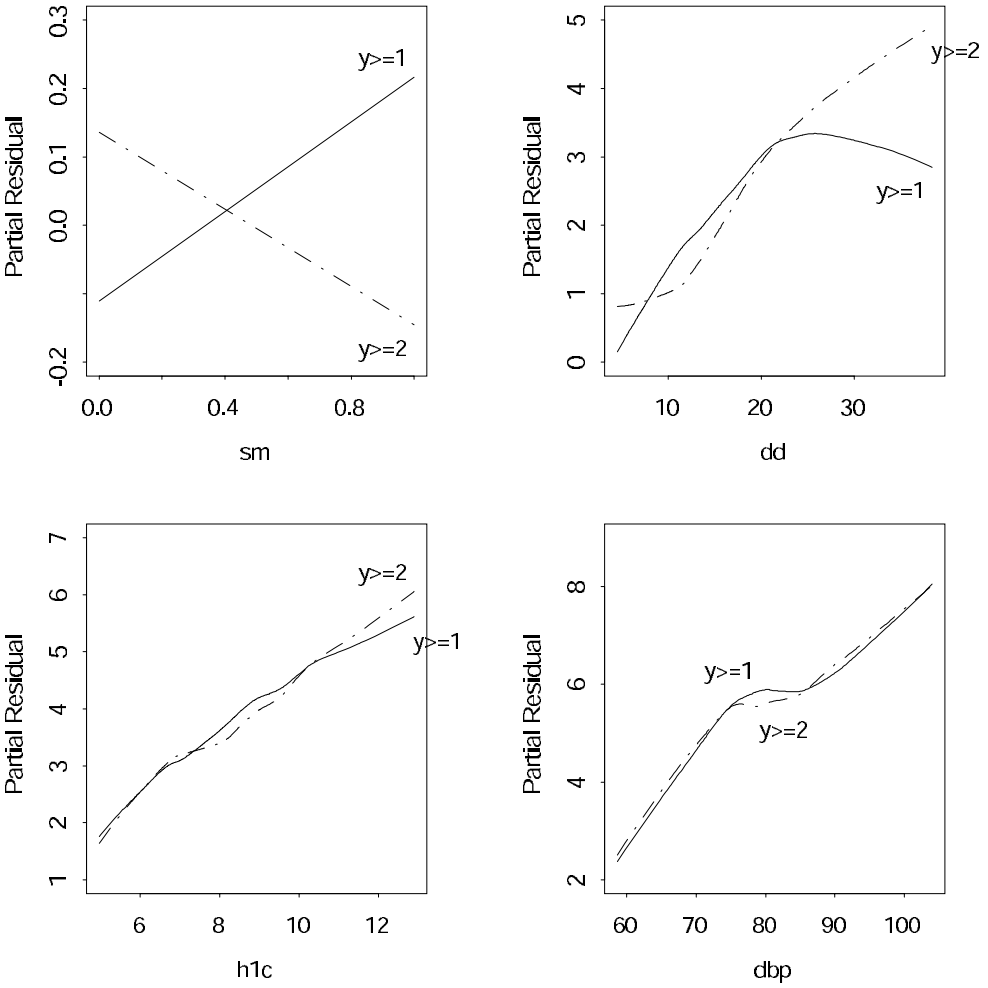

Fig. 1. Smoothed partial residual plots for the predictors of the standard PO model applied to the diabetes data. For smoking (sm) the equal slopes assumption does not seem to be adequate. For diabetes duration (dd) a quadratic functional relationship may be recommended.

Table 2
Results of the inappropriate standard models applied to the diabetes data

Class	Grouped continuous						Continuation ratio					
Link	logit			cloglog			logit			cloglog		
Model	PO model			PH model			CR model			CR = PH model		
	β_l	SE	p	β_l	SE	p	β_l	SE	p	β_l	SE	p
SM	0.255	0.192	0.1842	0.226	0.143	0.1135	0.263	0.175	0.1327	0.226	0.144	0.1165
DD	0.140	0.014	0.0001	0.092	0.009	0.0001	0.119	0.012	0.0001	0.092	0.009	0.0001
H1C	0.460	0.074	0.0001	0.344	0.053	0.0001	0.418	0.068	0.0001	0.344	0.053	0.0001
DBP	0.072	0.014	0.0001	0.060	0.010	0.0001	0.069	0.012	0.0001	0.060	0.010	0.0001

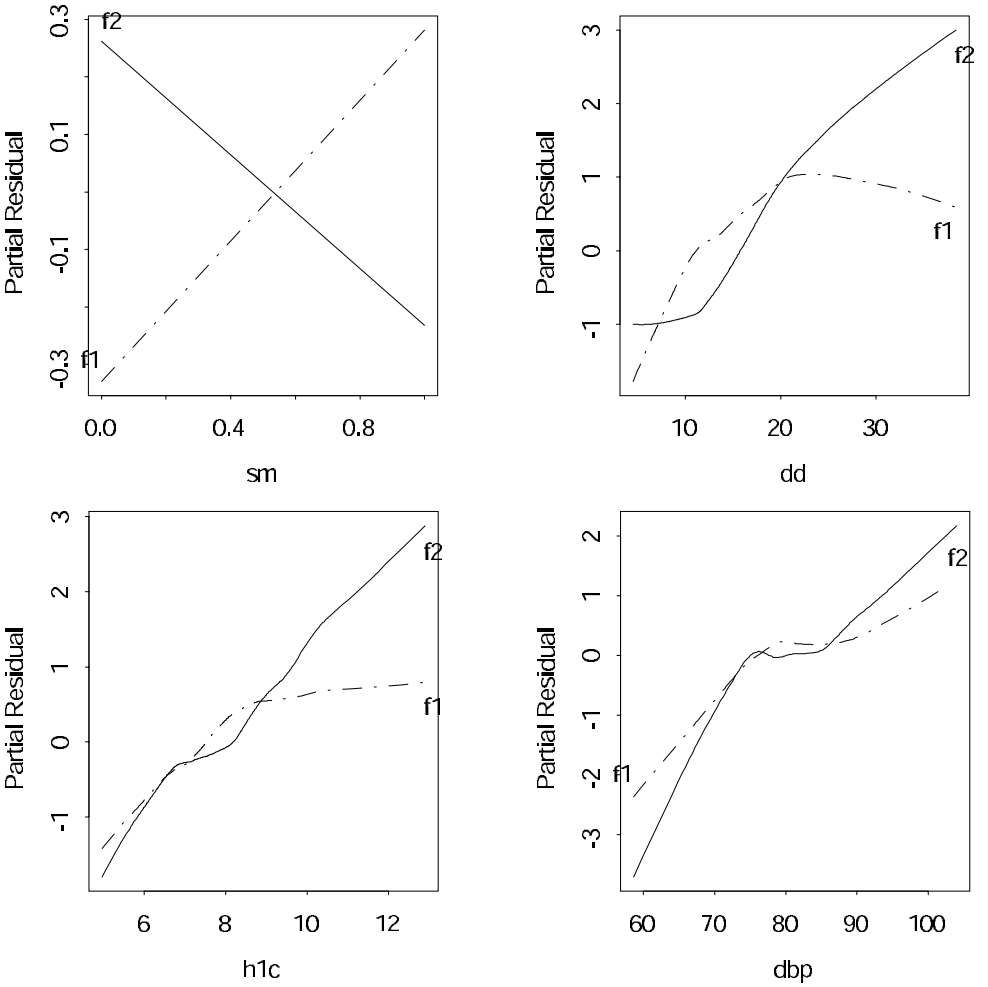

Fig. 2. Smoothed partial residual plots for the predictors of the standard CR model applied to the diabetes data. The results are similar to that of the PO model (Figure 1) but show greater fluctuations resulting from the use of subset-modeling (see section 3.5).

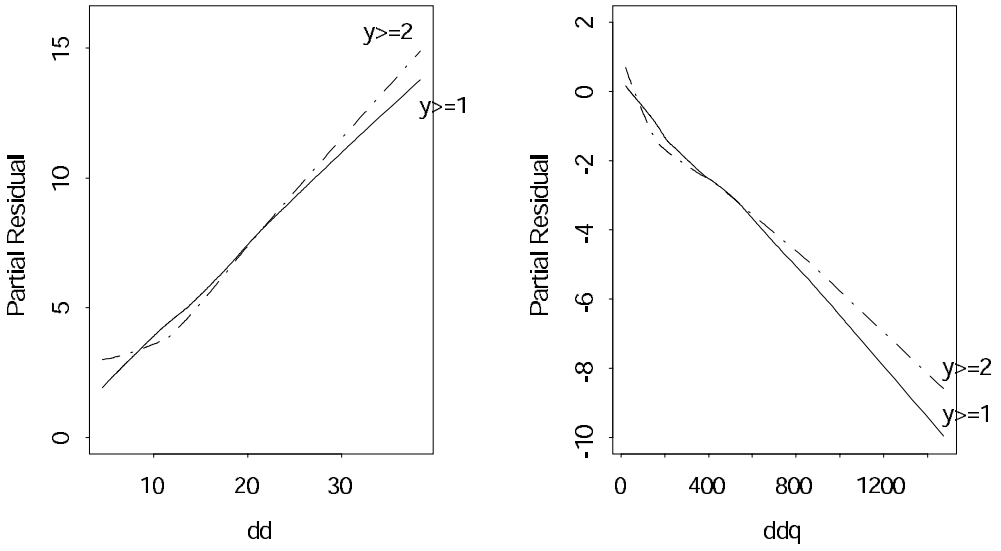

Fig. 3. Smoothed partial residual plots of the PO model applied to the diabetes data for the linear (dd) and quadratic effects (ddq) of diabetes duration.

The application of all standard models is inappropriate as the equal slopes assumption is not fulfilled for the risk factor smoking and because diabetes duration is not linearly related to the response neither on the logit nor on the cloglog scale. This can be seen in Figures 1 and 2 where the smoothed partial residuals are shown for the PO and the CR model.

Table 3
Results of the adequate separate and extended models applied to the diabetes data

Class	Grouped continuous						Continuation ratio					
Link	logit			logit			logit			cloglog		
Model	separate binary models			PPO model			ECR model			ECR model		
	β_{lj}	SE	p	β_{lj}	SE	p	β_{lj}	SE	p	β_{lj}	SE	p
SM	0.42	0.210	0.0476	0.444	0.200	0.0266	0.730	0.249	0.0034	0.586	0.203	0.0039
	−0.12	0.264	0.6436	−0.148	0.261	0.5707	−0.162	0.243	0.5036	−0.134	0.197	0.5003
DD	0.41	0.066	0.0001	0.377	0.061	0.0001	0.360	0.058	0.0001	0.315	0.049	0.0001
	0.39	0.096	0.0001									
DDQ	−0.01	0.002	0.0001	−0.006	0.002	0.0001	−0.006	0.001	0.0001	−0.006	0.001	0.0001
	−0.01	0.002	0.0105									
H1C	0.43	0.084	0.0001	0.461	0.076	0.0001	0.419	0.070	0.0001	0.339	0.055	0.0001
	0.56	0.105	0.0001									
DBP	0.06	0.015	0.0001	0.066	0.014	0.0001	0.061	0.013	0.0001	0.049	0.010	0.0001
	0.08	0.018	0.0001									

Regardless of whether a grouped continuous or a continuation ratio model should be used it is required to include unequal slopes for smoking and the quadratic effect of diabetes duration (*DDQ*). That the relation between the response and diabetes duration is adequately described by using the linear and the quadratic effect of *DD* is supported by Figure 3.

The equal sloped assumption seems to be adequate for all explanatory variables except of smoking for the PO as well as for the CR model (Figures 4 and 5). From the expected means there is no clear preference for one of the models. Both models seem to be compatible with the data provided that unequal slopes are included for smoking.

The inclusion of unequal slopes for smoking leads to the model variants summarized in Table 3.

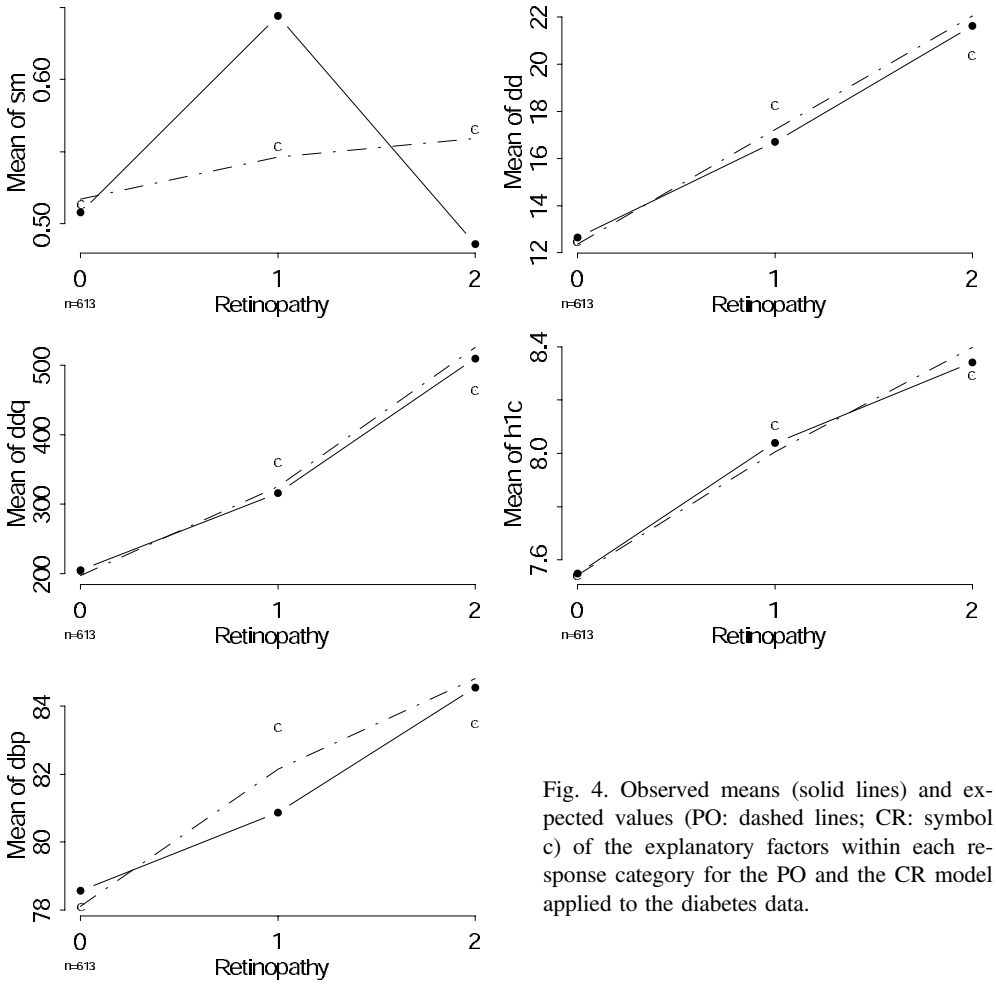

Fig. 4. Observed means (solid lines) and expected values (PO: dashed lines; CR: symbol c) of the explanatory factors within each response category for the PO and the CR model applied to the diabetes data.

Again there is little difference between the results of the different models with the exception that the parameter estimates of global extended models (PPO and ECR) have lower standard errors and lower p-values than the separate binary logistic models due to the lower number of estimated parameters. However, the conclusions concerning significant effects of the explanatory variables are the same from all models. The estimated effects of the explanatory factors calculated by means of the ECR model are shown in Figure 6.

5. Simulations

To investigate the bias of the estimated regression coefficients calculated from a standard model assuming equal slopes if actually an extended model is true, two

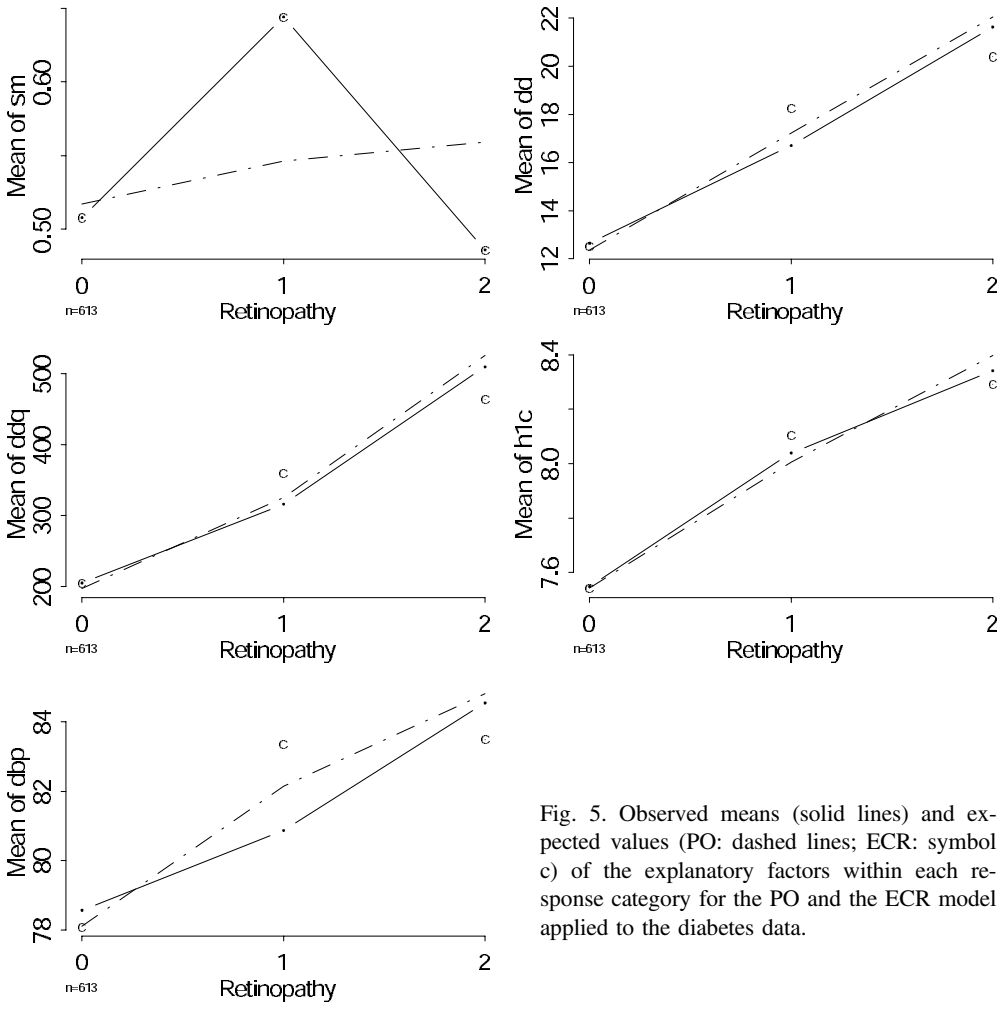

Fig. 5. Observed means (solid lines) and expected values (PO: dashed lines; ECR: symbol c) of the explanatory factors within each response category for the PO and the ECR model applied to the diabetes data.

simulation studies were performed. At first, simulations were performed by fitting the PO, the CR, and the ECR model to data built from the same models, that means the true models were fitted. Secondly, the CR and the ECR model were fitted to data built from the ECR model.

The diabetes data set is used for simulations as follows. First the PO model is fit with all relevant prognostic factors, then the individual probabilities $P(Y_i \geq j \mid X = x_i)$ are estimated using the parameter estimates of the model

```
fit.po <- lrm(ret6 ~ sm + dd + ddq + h1c + dbp, x=T, y=T, data=crm).
```

From this the matrix of estimated probabilities for $Y = 0,1,2$ is computed and by random sampling from (0, 1, 2) using these probabilities a new set of response values is generated. This was done 1000 times.

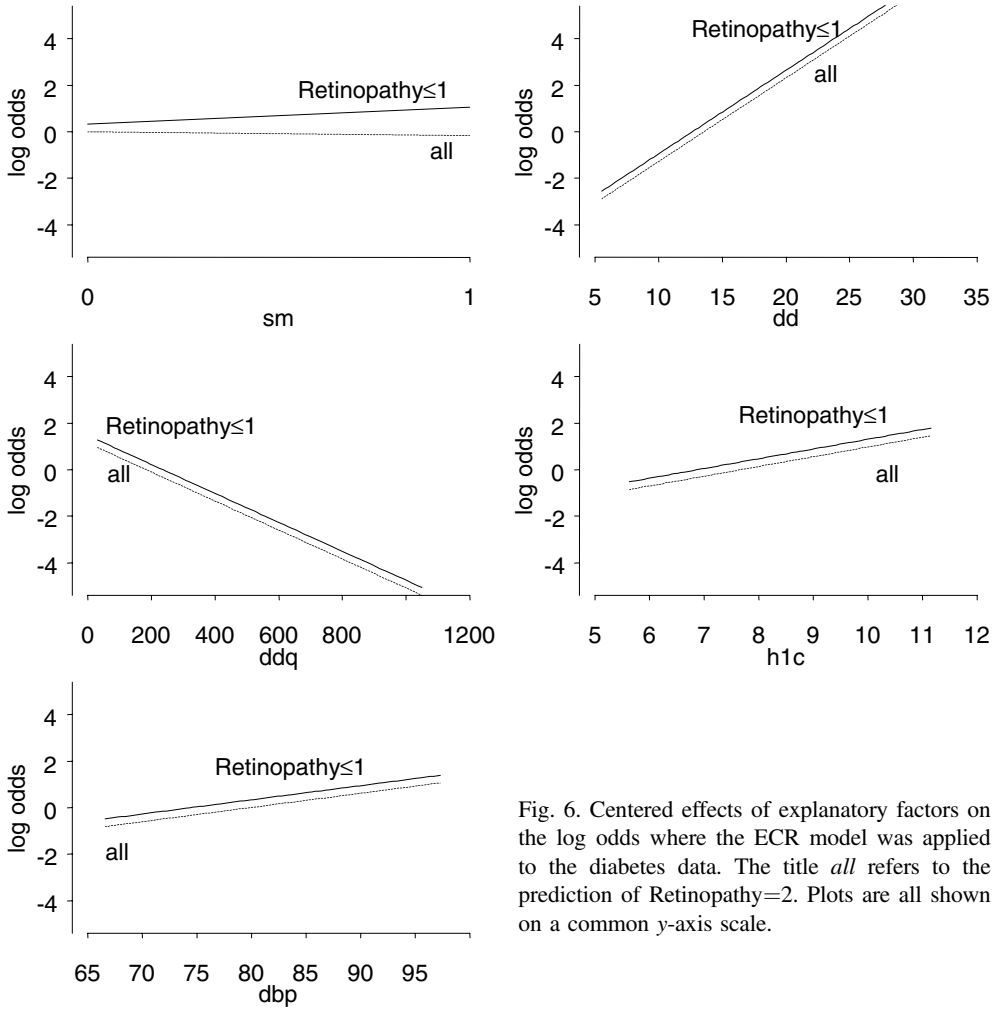

Fig. 6. Centered effects of explanatory factors on the log odds where the ECR model was applied to the diabetes data. The title *all* refers to the prediction of Retinopathy=2. Plots are all shown on a common *y*-axis scale.

Table 4
Bias of the regression coefficients of smoking calculated from 1000 simulations of the diabetes data

	PO model	CR model	ECR model
SM	0.29%	1.81%	0.30%
Interaction	–	–	1.96%

Comparing the parameter estimates of the original model with the means of estimates from 1000 simulation runs gives estimates of bias and standard errors. The same procedure was done with the CR model, with and without interaction between smoking status and response level,

```
attach(crm)
up <- crback.setup(ret6)
y <- up$y
cohort <- up$cohort
fit.cr <- lrm(y ~ cohort + sm + dd + ddq + h1c + dbp, data=crm[up$subs,])
fit.ecr <- lrm(y ~ cohort*sm + dd + ddq + h1c + dbp, data=crm[up$subs,]).
```

The results for the factor smoking are given in Table 4. As expected, the bias of the estimated regression coefficients is low in all models.

Another simulation study was performed using an artificial data set built from the ECR model with

$$P(Y = 2 \mid Y \leq 2, X_1 = x_1, X_2 = x_2)$$
$$= \exp(\alpha + \beta_1 x_1 + \beta_2 x_2)/\{1 + \exp(\alpha + \beta_1 x_1 + \beta_2 x_2)\}$$
$$P(Y = 1 \mid Y \leq 1, X_1 = x_1, X_2 = x_2)$$
$$= \exp(\alpha + \theta + (\beta_1 + \gamma) x_1 + \beta_2 x_2)/\{1 + \exp(\alpha + \theta + (\beta_1 + \gamma) x_1 + \beta_2 x_2)\}$$

and parameters $(\alpha, \theta, \beta_1, \beta_2, \gamma) = (-0.5, 0.5, 1, 1, -0.5)$ where X_1 and X_2 are binary random variables and Y was computed as above. The sample size is $n = 500$ and the number of simulation runs is 1000. The results are given in Table 5.

Table 5
Bias of the regression coefficients calculated from 1000 simulations of artificial data built from the ECR model

	CR model	ECR model
β_1	35.3%	2.13%
γ	–	1.05%
β_2	0.45%	0.93%

As before, the bias of the regression coefficients is low if the true model is fitted to the data. However, the bias of the parameter β_1 estimated from the standard CR model is quite large as the equal slopes assumption does not hold for this coefficient.

6. Discussion

We agree with SCOTT et al. (1997) and ANANTH and KLEINBAUM (1997) that ordinal regression models should be more widely used in epidemiology and biomedical research. However, for adequate use one has to be vary careful about goodness-of-fit and validity of the model assumptions. If the usual assumption of equal slopes for all ordinal response levels is fulfilled by the data the standard models (proportional odds model and continuation ratio model) represent powerful tools producing easily interpretable parameters which summarize the effects of explanatory factors over all response levels. However, if the equal slopes assumption is violated, extensions of the standard models are required. As in practice in the beginning of data analysis it is unknown which model if any describes the data adequately the use of diagnostic tools for investigating goodness-of-fit and checking of model assumptions is required in any case. We show how the statistical software packages SAS and S-Plus can be used for model building and computation of ordinal regression models even in the case of unequal slopes.

SAS offers an easy access to the standard ordinal regression models such as the proportional odds and the proportional hazards model. For these models SAS contains a lot of model building options concerning the selection of explanatory factors. However, the user of SAS should be aware of the fact that the standard models assuming equal slopes are computed even in the case where this assumption is violated. Diagnostic tools for investigating goodness-of-fit are implemented only for binary response variables. In the case of ordinal responses much more effort by the users themselves is required to find models describing the data adequately. In SAS one has the option to dichotomize the ordinal response and apply the diagnostic tools offered by SAS for each dichotomized response (BENDER and GROUVEN, 1998). A more easy access to graphical tools assessing the goodness-of-fit of ordinal regression models is given by the *Design* package for S-Plus (HARRELL, 1998a). Especially, smoothed partial residual plots are useful to assess simultaneously the adequacy of the equal slopes assumption and the linearity assumption of the expanatory factors for the considered link function.

We found only negligible differences between the two model classes of grouped continuous and continuation ratio models as well as of the different link functions. This can be explained by the fact that the grouped continuous model and the continuation ratio model using the cloglog link are equivalent in any case (LÄÄRÄ and MATTHEWS, 1985). On the other hand the link functions logit and cloglog are quite similar at least for small probabilities (MCCULLAGH and NELDER, 1989). As

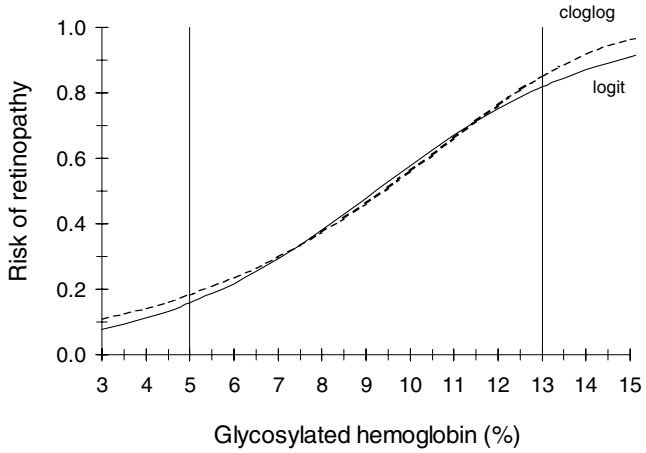

Fig. 7. Estimated response curves for the simple binary regression models using the logit link (solid lines) and the cloglog link (dashed lines) applied to the diabetes data. The vertical lines represent the interval for which data are available.

in practical applications the response probabilities are not large different link functions would usually not lead to quite different estimated associations between the explanatory factors and the response. This is shown in Figure 7 where the estimated response curves from binary logistic regression models using the logit and the cloglog link applied to the diabetes data are compared. There are differences between the curves in the upper and lower area. However, in the interval for which data are available ($HbA1_c$ between 5 and 13%) no suitable differences are observed.

As the grouped continuous and continuation ratio models are equivalent for the cloglog link and the logit and the cloglog link are comparable in practical applications, all models considered here would not result in quite different estimated response curves. Hence, it can be expected that the conclusions from all models would be similar in practice. However, the main model assumptions have to be checked carefully. If the equal slopes assumption is valid the standard models can be used; if this assumption is violated the respective extended models must be applied.

The choice for one of the standard models depends on whether the investigator wants to model the cumulative probabilities or the continuation ratios. In the authors' opinion the proportional odds model is easier to use and easier to interpret than the proportional hazards and the continuation ratio model. It may be difficult to explain a researcher that the results are not the same after a reversal of the response categories. Only the PO model has the appealing feature that only the signs of the regression coefficients change after a reversal of the response categories.

In the case of unequal slopes the application of standard ordinal regression models such as the PO and the CR model leads to questionable results. These

models represent no adequate description of the true association between the explanatory factors and the response. In the case of the diabetes data the application of all standard models would lead to the invalid conclusion that the effect of smoking has no suitable effect on the development of diabetic retinopathy. The simulations showed that the bias of the regression coefficient is large if the standard CR model is applied but the true model has unequal slopes. To perform a valid data analysis either separate logistic regressions or the PPO or ECR models can be used. For model choice the same issues than above are relevant. An additional point in this case is the availability of software. In this point the ECR model has the advantage that any software suitable for binary logistic regression can be used.

In summary, there are no differences of practical relevance between the results of grouped continuous and continuation ratio models. All models have similar assumptions which must be checked carefully before a model can be applied adequately. If the equal slopes assumption is not fulfilled more complicated extensions of the models should be applied. The choice of a model depends on the investigator's preference for cumulative probabilities or continuation ratios and the availability of software for calculations.

References

AGRESTI, A., 1989: Tutorial on modeling ordered categorical response data. *Psychological Bulletin* **105**, 290–301.

ANANTH, C. V. and KLEINBAUM, D. G., 1997: Regression models for ordinal responses: A review of methods and applications. *International Journal of Epidemiology* **26**, 1323–1333.

ANDERSON, J. A., 1984: Regression and ordered categorical variables (with discussion). *Journal of the Royal Statistical Society* B **46**, 1–30.

ARMSTRONG, B. and SLOAN, M., 1989: Ordinal regression models for epidemiologic data. *American Journal of Epidemiology* **129**, 191–204.

ASHBY, D., POCOCK, S. J., and SHAPER, A. G., 1986: Ordered polytomous regression: An example relating serum biochemistry and haematology to alcohol consumption. *Applied Statistics* **35**, 289–301.

ASHBY, D., WEST, C. R., and AMES, D., 1989: The ordered logistic regression model in psychatry: Rising prevalence of dementia in old people's homes. *Statistics in Medicine* **8**, 1317–1326.

BENDER, R. and GROUVEN, U., 1998: Using binary logistic regression models for ordinal data with non-proportional odds. *Journal of Clinical Epidemiology* **51**, 809–816.

BERRIDGE, D. M. and WHITEHEAD, J., 1991: Analysis of failure time data with ordinal categories of response. *Statistics in Medicine* **10**, 1703–1710.

COX, C., 1995: Location-scale cumulative odds models for ordinal data: A generalized non-linear model approach. *Statistics in Medicine* **14**, 1191–1203.

COX, C., 1997: Multinomial regression models based on continuation ratios. *Statistics in Medicine* **7**, 435–441.

GREENLAND, S., 1985: An application of logistic models to the analysis of ordinal responses. *Biometrical Journal* **27**, 189–197.

GREENLAND, S., 1994: Alternative models for ordinal logistic regression. *Statistics in Medicine* **13**, 1665–1677.

GREENWOOD, C. and FAREWELL, V., 1988: A comparison of regression models for ordinal data in an analysis of transplant-kidney function. *Canadian Journal of Statistics* **16**, 325–335.

HARRELL, F. E. Jr., 1998a: Design: S functions for biostatistical/epidemiologic modeling, testing, estimation, validation, graphics, and prediction. Functions available on the Web in the StatLib repository of statistical software at "http://lib.stat.cmu.edu/S/Harrell/".

HARRELL, F. E. Jr., MARGOLIS, P. A., GOVE, S., MASON, K. E., MULHOLLAND, E. K., LEHMANN, D., MUHE, L., GATCHALIAN, S., and EICHENWALD, H. F., 1998b: Tutorial in biostatistics: Development of a clinical prediction model for an ordinal outcome: The World Health Organization Multicentre Study of Clinical Signs and Etiological Agents of Pneumonia, Sepsis and Meningitis in Young Infants. *Statistics in Medicine* **17**, 909–944.

HASTIE, T. J., BOTHA, J. L., and SCHNITZLER, C. M., 1989: Regression with an ordered categorical response. *Statistics in Medicine* **8**, 785–794.

HOLTBRÜGGE, W. and SCHUMACHER, M., 1991: A comparison of regression models for the analysis of ordered categorical data. *Applied Statistics* **40**, 249–259.

HOSMER, D. W. and LEMESHOW, S., 1980: Goodness-of-fit tests for the multiple logistic regression model. *Communications in Statistics – A: Theory and Methods* **9**, 1043–1069.

JÖRGENS, V., GRÜSSER, M., BOTT, U., MÜHLHAUSER, I., and BERGER, M., 1993: Effective and safe translation of intensified insulin therapy to general internal medicine departments. *Diabetologia* **36**, 99–105.

LÄÄRÄ, E. and MATTHEWS, J. N. S., 1985: The equivalence of two models for ordinal data. *Biometrika* **72**, 206–207.

LEE, J., 1992: Cumulative logit modelling for ordinal response variables: Applications to biomedical research. *Computer Applications in Biosciences* **8**, 555–562.

MCCULLAGH, P., 1980: Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society – B* **42**, 109–142.

MCCULLAGH, P. and NELDER, J. A., 1989: *Generalized Linear Models*. Chapman and Hall, New York.

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.

PETERSON, B. and HARRELL, F. E. Jr., 1990: Partial proportional odds model for ordinal response variables. *Applied Statistics* **39**, 205–217.

PREGIBON, D., 1981: Logistic regression diagnostics. *Annals of Statistics* **9**, 705–724.

SAS, 1987: *SAS/STAT Guide for Personal Computers, Version 6 Edition*. SAS Institute Inc., Cary, NC.

SAS, 1990: *SAS Technical Report P-200, SAS/STAT Software: CALIS and LOGISTIC Procedures, Release 6.04*. SAS Institute Inc., Cary, NC.

SCOTT, S. C., GOLDBERG, M. S., and MAYO, N. E., 1997: Statistical assessment of ordinal outcomes in comparative studies. *Journal of Clinical Epidemiology* **50**, 45–55.

Dr. RALF BENDER (Dipl. Stat.)
AG 3 – Epidemiologie und Medizinische Statistik
Fakultät für Gesundheitswissenschaften
Universität Bielefeld
Postfach 10 01 31
D-33501 Bielefeld
Germany
Fax: 0521/106–6465
Email: Ralf.Bender@uni-bielefeld.de

Received, August 1999
Revised, May 2000
Accepted, May 2000