

A tutorial on selecting and interpreting predictive models for ordinal health-related outcomes

Maria Guzman-Castillo¹ · Sally Brailsford² · Michelle Luke³ · Honora Smith²

Received: 14 August 2014 / Revised: 5 June 2015 / Accepted: 11 June 2015 /
Published online: 24 June 2015
© Springer Science+Business Media New York 2015

Abstract Ordinal variables are very often objects of study in health sciences. However, due to the lack of dissemination of models suited for ordinal variables, users often adopt other practices that result in the loss of statistical power. In this tutorial, different models from the family of logistic regression models are introduced as alternatives to handle and interpret ordinal outcomes. The models that were considered include: ordinal regression model (ORM), continuation ratio model (CRM), adjacent category model (ACM), generalised ordered logit model, sequential model, multinomial logit model, partial proportional odds model, partial continuation ratio model and stereotype ordered regression model. By using the relationship of hospital length of stay in a public hospital in Mexico with patient characteristics as an example, the models were used to describe the nature of such relationship and to predict the length of stay category to which a patient is most likely to belong. After an initial analysis, the ORM, CRM and ACM proved to be unsuitable for our data due to the transgression of the parallel regression assumption. The rest of the models were estimated in STATA. The results suggested analogous directionality of the parameter estimates between models, although the interpretation of the odds ratios varied from one model to another. Performance measurements indicated that the models had similar prediction performance. Therefore, when there is an interest in exploiting the ordinal nature of an outcome, there is no reason to maintain practices that ignore such nature since the models discussed here proved to be computationally inexpensive and easy to estimate, analyse and interpret.

Keywords Logistic regression · Ordinal variables · Generalised ordered logit · Partial proportional odds · Multinomial logit · Stereotype ordered regression

✉ Maria Guzman-Castillo
M.Guzman-Castillo@liverpool.ac.uk

¹ Department of Public Health and Policy, University of Liverpool, Whelan Bld., Liverpool L69 3GB, UK

² University of Southampton, Southampton, UK

³ University of Sussex, Flamer Brighton, UK

1 Introduction

Ordinal variables allow assigning numbers to classify characteristics of subjects into categories that are ordered in some meaningful way. There are two broad categories of ordinal variables: the first is a discretised version of a continuous variable, which is split into different intervals according to some specific criteria and where each interval corresponds to a discrete category. For example, patients can be grouped according to their hospital length of stay (LoS) into categories such as “short LoS”, “medium LoS” and “long LoS”. The other type of ordinal variable is originated by an assessing process which evaluates an indeterminate amount of information before providing a grade or score of the ordinal variable (Anderson 1984). Examples of this type of ordinal variable are the Lansky score (Lansky et al. 2006) and Barthel scale (Mahoney and Barthel 1965) to assess general well-being and performance in daily living activities for children with cancer and geriatric patients, the score systems for patient dependency at an intensive care unit (Flaatten et al. 2002) or the triage classification used in emergency departments.

Ordinal variables are very often objects of study. A classical approach is to create models to describe them in relation to other variables by looking for rules of classification based on data. In this context, there are a variety of models that can accommodate ordinal variables, including the classical and well-established logistic regression models. However, in practice, there is a very limited dissemination of the range of models that can be used. Consequently, it is very common to use logistic models that are designed for nominal variables (unordered categories) with ordinal variables (Keski-Rahkonen et al. 2003; Takazawa et al. 2003; Newman et al. 2005; Mcelroy et al. 2002; Walston et al. 2002; Mäntyselkä et al. 2003). Another common approach is to amalgamate adjacent categories of the ordinal outcome into two broad categories and use standard binary logistic regression (Bender and Grouven 1998). However, this often results in the loss of information, description and statistical power (Ananth and Kleinbaum 1997; Armstrong and Sloan 1989). Amalgamating adjacent categories could be acceptable if when estimating a binary logistic regression (using the categories that one desire to combine), it is found that all the slopes in the model are simultaneously equal to zero.

The purpose of this tutorial is to provide an introduction to models from the family of logistic regression that are suitable for the analysis of ordinal data, putting particular emphasis on understanding how each model can be used to answer different research questions. The logistic regression models considered here are: the ordinal regression model, continuation ratio model, adjacent category model, generalised regression model, sequential model, multinomial logit model, partial proportional odds model, partial continuation ratio model and stereotype ordered regression model.

2 Data

The illustrative example of an ordinal outcome variable in this tutorial is a discretised version of LoS, whereby patients were classified according to three categories: “short LoS” (patients with LoS up to 3 days), “medium LoS” (patients with LoS from 4 to 11 days) and “long LoS” (patients with LoS from 12 days onwards). This classification was made based on empirical observation and personal judgment. Figure 1 shows the distribution of the continuous variable LoS, and its basic descriptive statistics are given in Table 1. Figure 2 depicts the discretised version of LoS split into three categories, where in

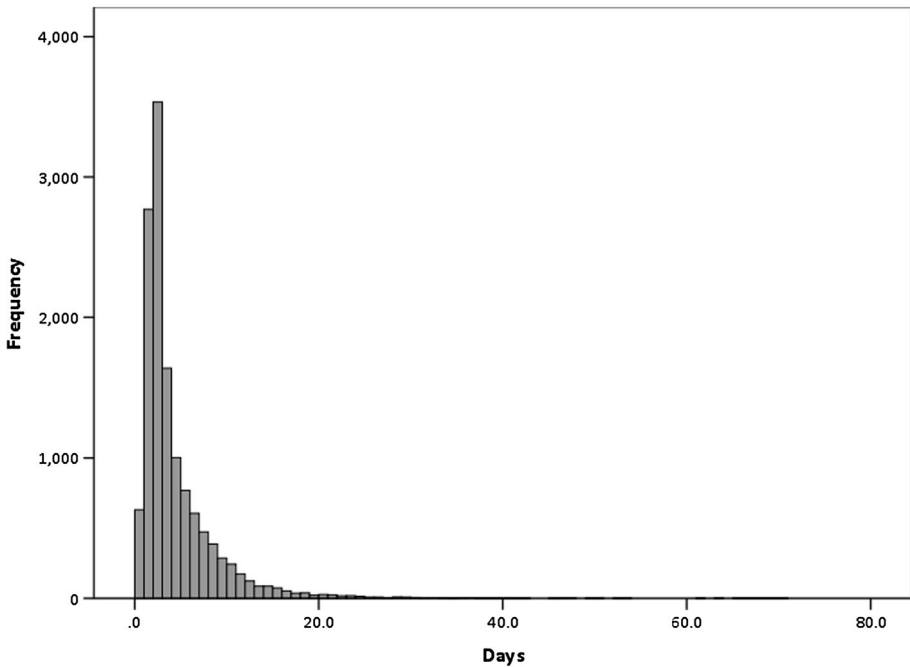


Fig. 1 Distribution of the LoS as continuous variable

Table 1 Descriptive statistics for LoS

Mean	Std. Deviation	Variance	Minimum	Maximum	Skewness	Kurtosis
3.97	4.55	20.77	0.50	196	4.31	34.73

accordance with Fig. 1, the vast majority of patients have a short LoS and just few of them experience a long LoS.

Potential predictor variables of a patient’s LoS category are described in Table 2. The data were extracted from routine patient records in a general public hospital in Mexico. The hospital belongs to the Secretariat of Health, is located in the heart of an urban area and is open to the general population, making it the preferable option for people who cannot afford private medical services or who are not affiliated to another healthcare provider. It is a 148-bed second-level hospital, which means it offers outpatient walk-in clinics and hospitalisations for basic medical specialties, such as adult medicine, paediatrics, obstetrics and gynaecology, and general surgery. Hospitals that correspond to this level of care have operating rooms and equipment suitable for performing surgery of low and medium level of complexity.

The data correspond to almost 13,300 patient records from the years 2005 to 2009. The variables “diagnosis” and “surgical procedure” originally contained, respectively, around 800 and 200 different ICD codes (International Classification of Diseases codes version 10 for diagnoses and version 9 for surgical procedures), which would complicate their

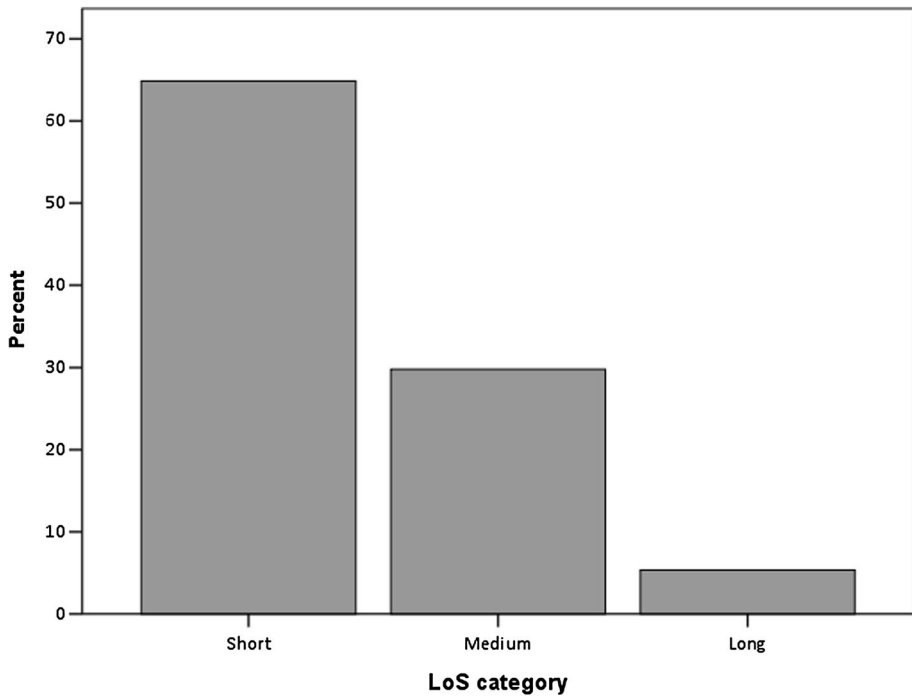


Fig. 2 Distribution of the LoS as discretised variable

Table 2 Description of variables

Variables	Description
Gender	Female is the reference category
Previous admissions	Number of previous hospitalisations
Age	Age of patient
Ward	Ward of treatment: adult medicine and general surgery
Origin	Area of arrival in hospital: accident emergency and outpatient clinic
Number of surgical procedures	Total number of surgical procedures undergone
Diagnosis	Main health problem or disease, cause of the hospitalisation
Surgical procedure	Main surgical procedure
Length of stay	Number of nights spent at hospital

inclusion for further statistical analysis. To reduce the number of ICD codes of such variables, we used hierarchical cluster methods based on the χ^2 dissimilarity measure (Rezanková 2009). The diagnosis codes were grouped into five clusters using complete linkage algorithm (Defays 1977), and the surgical procedure codes were grouped into four main categories using Ward's algorithm (Ward 1963). For more details, the reader is referred to Guzman Castillo (2012). Tables 3 and 4 contain some examples of the most common ICD codes for each diagnosis and surgical procedure category.

Table 3 Some common diagnoses within each of the three diagnosis categories

Diagnosis category 1	Diagnosis category 2	Diagnosis category 3
Chronic cholecystitis	Non-insulin-dependent diabetes mellitus	Gastritis and duodenitis
Unspecified appendicitis	Stroke	Abscess of intestine
Acute appendicitis	Gastrointestinal haemorrhage	Ventral Hernia
Diarrhoea and gastroenteritis	Hepatic failure	Decubitus ulcer
	Alcohol cirrhosis of liver	Pleural effusion

Table 4 Some common surgical procedures within each of the four surgical procedure categories

Surgical procedure category 1	Surgical procedure category 2	Surgical procedure category 3	Surgical procedure category 4
Exploratory laparotomy	Amputation of toe	Appendectomy	Peritoneal dialysis
Abdominal hysterectomy	Irrigation of wound	Cholecystectomy	Repair of inguinal hernia
Fasciotomy	Partial excision of large intestine	Endoscopy of small intestine	Aspiration of skin and subcutaneous tissue
Open reduction of fracture	Other incision of pleura	Prostatectomy	Umbilical herniorrhaphy
Exploratory laparotomy	Other lysis of peritoneal adhesions	Excision of haemorrhoids	Circumcision

3 Models

Fullerton (2009) presented a classification of models for ordinal data based on how they deal with the proportional odds assumption or parallel regression assumption. Most of the models presented here can be thought as $J - 1$ simultaneous binary logistic regressions, where J is the number of categories of the ordinal-dependent variable. The parallel regression assumption means that the beta coefficients are equal across the simultaneous regressions. Assuming the equality of slopes among categories allows interpreting the models in the same way for all categories, making more parsimonious models.

The models that are based on that assumption are the ordinal regression model, the continuation ratio model and the adjacent category model. However, there is a general consensus that this assumption is quite stringent and the chance of all the dependent variables in the model having identical slope coefficients is likely to be quite rare (Lall et al. 2002). Consequently, other models have been presented in statistical literature as alternatives.

The generalised ordered logit, the sequential and multinomial models have all slope coefficients not delineated by the parallel regression assumption. One of the drawbacks of these models is that they include many more parameters, as a result of setting free all variables from parallel line constraints. Although it is very common to find that the parallel regression assumption has been violated, usually not all the slope coefficients of the model transgress the assumption.

The partial proportional odds model, partial continuation ratio and stereotype ordered model are models that impose constraints for parallel lines only where they are needed, i.e.

some slope coefficients can be the same for all the J categories, while others can differ, hence avoiding including unnecessary extra parameters in the model.

This selection of models is by no means comprehensive. In particular, one can find other partially constrained models for the ordinal regression, the continuation ratio and the adjacent category models (Cole and Ananth 2001; Hauser and Andrew 2006; Fullerton 2009).

Following this classification, the definition of the models in the next section has been grouped into models that hold the parallel regression assumption for every independent variable, for no independent variables and for some independent variables.

3.1 Parallel assumption for every independent variables

3.1.1 Ordinal regression model

The ordinal regression model (ORM), commonly known as the cumulative odds model (Walker and Duncan 1967) or proportional odds model (McCullagh 1980), was the first model developed exclusively for ordinal outcomes. The ORM can be defined as a probability model:

$$\ln\left(\frac{\Pr(y \leq j|\mathbf{x})}{\Pr(y > j|\mathbf{x})}\right) = \tau_j - \mathbf{x}\beta, \quad j = 1, \dots, J - 1,$$

where \mathbf{x} is the vector of independent variables, β_s are the slope coefficients, τ_j are the thresholds, and J is the number of categories of the ordinal-dependent variable. The predicted probabilities of belonging to a certain category are defined as:

$$\begin{aligned} \Pr(y = 1|\mathbf{x}) &= \frac{\exp(\tau_1 - \mathbf{x}\beta)}{1 + \exp(\tau_1 - \mathbf{x}\beta)}, \\ \Pr(y = j|\mathbf{x}) &= \frac{\exp(\tau_j - \mathbf{x}\beta)}{1 + \exp(\tau_j - \mathbf{x}\beta)} - \frac{\exp(\tau_{j-1} - \mathbf{x}\beta)}{1 + \exp(\tau_{j-1} - \mathbf{x}\beta)}, \quad j = 2, \dots, J - 1. \\ \Pr(y = J|\mathbf{x}) &= 1 - \frac{\exp(\tau_{J-1} - \mathbf{x}\beta)}{1 + \exp(\tau_{J-1} - \mathbf{x}\beta)}, \end{aligned}$$

Furthermore, the ORM is often formulated as a latent variable model, defined as:

$$\begin{aligned} y'_i &= \mathbf{x}\beta + \epsilon_i. \\ y_i = j &\text{ if } \tau_{j-1} \leq y'_i < \tau_j, \quad j = 1, \dots, J, \end{aligned}$$

where y'_i is the latent variable ranging from ∞ to $-\infty$, and ϵ_i is the random error. The thresholds τ_1 through τ_{J-1} are parameters to estimate, assuming that $\tau_0 = -\infty$ and $\tau_J = \infty$. In the context of LoS, the continuous latent variable y'_i can be thought of as the propensity of a patient to belong to a certain category. For example, the LoS category now relies on the latent variable:

$$\begin{aligned} y_i = \textit{short} &\text{ if } \tau_0 \leq y'_i < \tau_1. \\ y_i = \textit{medium} &\text{ if } \tau_1 \leq y'_i < \tau_2. \\ y_i = \textit{long} &\text{ if } \tau_2 \leq y'_i < \tau_3. \end{aligned}$$

Thus, when the latent variable crosses a threshold τ_j , the patient category changes.

3.1.2 Continuation ratio model

A special type of ordinal model is the continuation ratio model (CRM) proposed by Fienberg (1977) in which the categories represent levels, where the lowest level must occur before the second, the second before the third and so forth until the highest level (Hilbe 2009). It can be thought as stages in some process through which an individual can advance. A key characteristic of the process is that an individual must pass through each stage (Long and Freese 2006). This special characteristic suits the nature of the patient journey through the hospital where the patient can evolve from a short to a medium LoS and so on. The CRM is defined as:

$$\ln\left(\frac{\Pr(y = m|\mathbf{x})}{\Pr(y > m|\mathbf{x})}\right) = \tau_m - \mathbf{x}\beta, \quad m = 1, \dots, J - 1,$$

where m is the stage and J is the number of categories of the outcome variable. The predicted probabilities are calculated by:

$$\Pr(y = m|\mathbf{x}) = \frac{\exp(\tau_m - \mathbf{x}\beta)}{\prod_{j=1}^m [1 + \exp(\tau_j - \mathbf{x}\beta)]} \quad m = 1, \dots, J - 1.$$

$$\Pr(y = J|\mathbf{x}) = 1 - \sum_{j=1}^{J-1} \Pr(y = j|\mathbf{x}).$$

3.1.3 Adjacent category model

As the name indicates, the probability of interest in the adjacent category model (ACM) is the adjacent probability: the probability of having a short LoS against the probability of having a medium LoS or the probability of having a medium LoS against the probability of having a long LoS. The ACM formulated by Goodman (1983) is defined as:

$$\ln\left(\frac{\Pr(y = m|\mathbf{x})}{\Pr(y = m + 1|\mathbf{x})}\right) = \tau_m - \mathbf{x}\beta, \quad m = 1, \dots, J - 1,$$

The predicted probabilities are calculated by:

$$\Pr(y = m|\mathbf{x}) = \frac{\exp\left(\sum_{j=m}^{J-1} (\tau_j - \mathbf{x}\beta)\right)}{1 + \sum_{q=1}^{J-1} \left[\exp\left(\sum_{m=q}^{J-1} (\tau_m - \mathbf{x}\beta)\right)\right]} \quad m = 1, \dots, J - 1.$$

$$\Pr(y = J|\mathbf{x}) = 1 - \sum_{j=1}^{J-1} \Pr(y = j|\mathbf{x})$$

3.2 No parallel assumption for independent variables

3.2.1 Generalised ordered logit model

Described by Clogg and Shihadeh (1994), the generalised ordered logit model (GOLM) allows the slope coefficients to differ for each of $J-1$ binary regressions as represented in the following equation: $\ln\left(\frac{\Pr(y \leq j|\mathbf{x})}{\Pr(y > j|\mathbf{x})}\right) = \tau_j - \mathbf{x}\beta_j, \quad \text{for } j = 1 \text{ to } J - 1.$

The predicted probabilities are calculated as:

$$\begin{aligned}\Pr(y = 1|\mathbf{x}) &= \frac{\exp(\tau_1 - \mathbf{x}\beta_1)}{1 + \exp(\tau_1 - \mathbf{x}\beta_1)} \\ \Pr(y = j|\mathbf{x}) &= \frac{\exp(\tau_j - \mathbf{x}\beta_j)}{1 + \exp(\tau_j - \mathbf{x}\beta_j)} - \frac{\exp(\tau_{j-1} - \mathbf{x}\beta_{j-1})}{1 + \exp(\tau_{j-1} - \mathbf{x}\beta_{j-1})}, \quad j = 2, \dots, J - 1. \\ \Pr(y = J|\mathbf{x}) &= 1 - \frac{\exp(\tau_{J-1} - \mathbf{x}\beta_{J-1})}{1 + \exp(\tau_{J-1} - \mathbf{x}\beta_{J-1})}\end{aligned}$$

Notice that the equations for the GOLM are similar to the ORM. GOLM retains the nature of the ORM by considering simultaneously the effects of a set of independent variables across successive dichotomisations of the outcome (O'connell 2010), yet setting free the slope coefficients β_j to vary across the categories.

3.2.2 Sequential model

Also known as the unconstrained continuation ratio model, the sequential model (SeqM) presented in (Kahn and Morimune 1979) and (Mare 1979) describes an ordinal outcome as a sequence of decisions or steps. It can be expressed as:

$$\ln\left(\frac{\Pr(y = m|\mathbf{x})}{\Pr(y > m|\mathbf{x})}\right) = \tau_m - \mathbf{x}\beta_j, \quad m = 1, \dots, J - 1,$$

The predicted probabilities are calculated by:

$$\begin{aligned}\Pr(y = m|\mathbf{x}) &= \frac{\exp(\tau_m - \mathbf{x}\beta_j)}{\prod_{j=1}^m [1 + \exp(\tau_j - \mathbf{x}\beta_j)]} \quad m = 1, \dots, J - 1. \\ \Pr(y = J|\mathbf{x}) &= 1 - \sum_{j=1}^{J-1} \Pr(y = j|\mathbf{x}).\end{aligned}$$

3.2.3 Multinomial logit model

Luce (1959) proposed the multinomial logit model (MNL) as an extension of the binary logistic regression model to handle polytomous outcomes, where the categories are no longer considered as ordered and the effects of the independent variables are allowed to differ for each outcome (Hilbe 2009). Although it was defined for nominal outcomes, it is often used for ordinal data. The MNL can be expressed as:

$$\ln\left(\frac{\Pr(y = m|\mathbf{x})}{\Pr(y = b|\mathbf{x})}\right) = \mathbf{x}\beta_{m|b}, \quad m = 1, \dots, J,$$

where b is the reference category or the comparison group. The predicted probabilities are calculated by:

$$\Pr(y = m|\mathbf{x}) = \frac{\exp(\mathbf{x}\beta_{m|b})}{\sum_{j=1}^J \exp(\mathbf{x}\beta_{j|b})}.$$

The MNL relies on the assumption of independence of irrelevant alternatives (IIA) (Luce 1959; Arrow 1963), where the odds do not depend on other alternative outcomes that

are available. In other words, adding or deleting outcome categories does not affect the odds among other outcomes. It is plausible to assume that the categories of LoS are independent, because the odds of belonging to certain LoS category do not change if the other two categories are omitted. Alternatively, the Hausman–McFadden test (Hausman and Mcfadden 1984) and the Small–Hsiao test (Small and Hsiao 1985) can be used to evaluate IIA. However, there is evidence suggesting that both tests often give inconsistent results and provide little guidance to violations of the IIA assumption (Long and Freese 2006)

3.3 Parallel assumption for some independent variables

3.3.1 Partial proportional odds model

The partial proportional odds model (PPOM) formulated by Peterson and Harrell (Peterson and Harrell 1990) imposes constraints for parallel lines only where they are needed. The GOLM equation is now extended to accommodate the unconstrained parameters which violated the assumption:

$$\ln\left(\frac{\Pr(y \leq j|\mathbf{x})}{\Pr(y > j|\mathbf{x})}\right) = \tau_j - (\mathbf{x}\beta + T\gamma_j), \quad j = 1, \dots, J - 1.$$

Here \mathbf{x} is the vector containing the full set of independent variables. \mathbf{T} is a vector containing a subset of independent variables which violate the parallel assumption, and γ_j are the regression coefficients associated with the variables in \mathbf{T} . The predicted probabilities of belonging to a certain category are defined as:

$$\begin{aligned} \Pr(y = 1|\mathbf{x}) &= \frac{\exp(\tau_1 - (\mathbf{x}\beta + T\gamma_1))}{1 + \exp(\tau_1 - (\mathbf{x}\beta + T\gamma_1))}. \\ \Pr(y = j|\mathbf{x}) &= \frac{\exp(\tau_j - (\mathbf{x}\beta + T\gamma_j))}{1 + \exp(\tau_j - (\mathbf{x}\beta + T\gamma_j))} - \frac{\exp(\tau_{j-1} - (\mathbf{x}\beta + T\gamma_{j-1}))}{1 + \exp(\tau_{j-1} - (\mathbf{x}\beta + T\gamma_{j-1}))} \quad j = 2, \dots, J - 1. \\ \Pr(y = J|\mathbf{x}) &= 1 - \frac{\exp(\tau_{J-1} - (\mathbf{x}\beta + T\gamma_{J-1}))}{1 + \exp(\tau_{J-1} - (\mathbf{x}\beta + T\gamma_{J-1}))} \end{aligned}$$

3.3.2 Partial continuation ratio model

The partial continuation model (PCRM) extends the equation for the CRM by adding coefficients for those variables that violate the parallel assumption:

$$\ln\left(\frac{\Pr(y = m|\mathbf{x})}{\Pr(y > m|\mathbf{x})}\right) = \tau_m - (\mathbf{x}\beta + T\gamma_m), \quad m = 1, \dots, J - 1,$$

The predicted probabilities are calculated by:

$$\begin{aligned} \Pr(y = m|\mathbf{x}) &= \frac{\exp(\tau_m - (\mathbf{x}\beta + T\gamma_j))}{\prod_{j=1}^m [1 + \exp(\tau_j - (\mathbf{x}\beta + T\gamma_j))]} \quad m = 1, \dots, J - 1. \\ \Pr(y = J|\mathbf{x}) &= 1 - \sum_{j=1}^{J-1} \Pr(y = j|\mathbf{x}). \end{aligned}$$

3.3.3 Stereotype ordered regression model

The stereotype ordered regression model (SORM) can be thought of as imposing ordering constraints on a multinomial model (Lunt 2005). It was proposed by Anderson (1984) in response to the restrictive parallel regression assumption of the ORM. The model was originally defined for ordinal variables originated by an assessing process, but it is not restricted to this type of variables only. The SORM is defined as:

$$\ln \frac{\Pr(y = m|\mathbf{x})}{\Pr(y = b|\mathbf{x})} = (\theta_m - \theta_b) - (\phi_m - \phi_b)\mathbf{x}\beta,$$

where θ s are the intercepts and ϕ s are scale factors associated with the outcome categories and b is the reference category or the comparison group. The model allows the coefficients associated with each independent variable to differ by a scale factor that depends on the pair of outcomes on the left-hand side of the equation. Similarly, the θ s allow different intercepts for each pair of outcomes. If the relationship between the independent variables and dependent variable is ordinal, then $\phi_1 > \phi_2 > \dots > \phi_{J-1} > \phi_J$.

Constraints need to be added to the model to make it identifiable: $\phi_1 = 1$, $\phi_J = 0$, $\theta_1 = 1$ and $\theta_J = 0$. The predicted probabilities of belonging to a certain category are defined as:

$$\Pr(y = m|\mathbf{x}) = \frac{\exp(\theta_m - \phi_m\mathbf{x}\beta)}{\sum_{j=1}^J (\exp(\theta_j - \phi_j\mathbf{x}\beta))}.$$

The model presented here is known as a one-dimensional stereotype ordered regression. Anderson (1984) also presents an extension of the equations to model ordinal variables that are constructed by multiple domains. For example, the coronary heart disease risk score (Wilson et al. 1998) classifies individuals according to their risk of having a coronary heart disease event in the next 10 years. The score is constructed by adding the scores of different risk factors (e.g. diabetes, cholesterol and blood pressure) indicating level or risk (very low, low, moderate, high and very high)

4 Odds ratios and interpretation

The interpretation of the logistic regression models can be more manageable if it comes in terms of odds ratios (ORs). The odds of an event occurring are defined as the probability of an event occurring divided by the probability of that event not occurring. In terms of logistic regression models, the odds ratio then compares the change in the odds that results from a unit change in the predictor.

The models previously described differ in how an event (or non-event) is defined. Table 5 summarises how the odds in each type of approach and models are interpreted.

The adjacent category, multinomial logit and stereotype ordered models are similar in the sense they perform one by one comparison of the categories. For example, the adjacent category model compares one category against the next higher category. Because the MNLM ignores the ordering of the outcomes, it compares a given category against the reference category. The SORM interpretation of the odds could be similar to the MNLM, but when estimated using the default options of STATA, it compares the highest category versus the lowest category.

Table 5 Odds ratio interpretation for each logistic regression model

As the predictor increases by one unit, the odds of _____ increase, holding constant the effects of the other predictors

ORM	Being in a higher category versus a category m or lower
GOLM	
PPOM	
CRM	Progressing to a higher category versus category m (given you have progressed from the lowest)
SEQ	
PCRM	
ACM	Being in the next higher category versus being in a category m
MNLM	Being in a category m versus the reference category
SORM	Being in the highest category occurring versus the lowest category

The odds ratios can be computed from the models' parameter estimates by exponentiating the β coefficient. When the odds ratio is greater than 1 (i.e. computed from positive β s), it indicates that, as the predictor increases, the odds of the event occurring increase by a factor of $\exp(\beta)$, holding all the other variables constant. Conversely, a value lower than 1 (i.e. computed from negative β s) indicates that, as the predictor increases, the odds of the event occurring decrease by a factor of $\exp(\beta)$, holding all the other variables constant.

To facilitate the interpretation, when a predictor has a negative effect on the event occurring, instead of calculating the odds of the event occurring, the odds of the event not occurring can be computed by simply taking the inverse of the effect on the odds of the event occurring (i.e. $\exp(\beta)^{-1}$). Alternatively, the models can be interpreted as per cent change: as the predictor increases, the odds of the event occurring increase by $(100 \times [\exp(\beta) - 1])\%$, holding all other the variables constant.

5 Model estimation

As in common practice, two-thirds of the LoS data set, named the training set, was allocated for estimation purposes and the remaining third, named the validation set, was used for testing (Dobbin and Simon 2011). All the models described above were fitted to the training set using STATA. The first step is to evaluate the parallel assumption. We fitted the ORM using **ologit** followed by the command **brant** which performs a Brant test. This test compares the beta coefficients from $J-1$ binary logits and gives a list of which variables are violating the parallel assumption. This was useful later to estimate the partially constrained models (i.e. PPOM, PCRM and SORM) where constraints needed to be imposed on those variables where the assumption is not violated. The commands **ocratio** and **adjcatlogit** were used to estimate the CRM and ACM; the **gologit** command with the **npl** option was used for GOLM; and the commands **ucrlogit**, **mlogit** and **slogit** were used to estimate the SeqM, MNLM and SORM, respectively.

The PPOM was estimated using the **gologit2** command with the **autofit** option to impose constraints on the variables where the parallel assumption is not violated. Finally, the PCRM was estimated using the **seqlogit**, but the constraints needed to be added manually. See the appendix for the full STATA code and alternative commands for estimation.

All the models were estimated through the procedure of maximum likelihood estimation. Maximum likelihood estimates are the values of the parameters that have the “maximum likelihood” of generating the observed sample.

To compare the models discussed here, along with the log-likelihood, the Akaike information criterion (AIC) and Bayesian information criterion (BIC) were calculated. A model with a higher log-likelihood should be considered as a better-fitting model. Models with the smaller absolute AIC and/or BIC values should be preferred.

In addition, model performance was measured through accuracy rates (per category and overall performance) to express the percentage of times the patient membership (i.e. observed category to which they belong) matches with the membership predicted by the models discussed here. The predicted category was assigned using the highest probability method (Anderson and Philips 1981), which allocates a patient to the category for which he or she got the highest probability estimate.

Furthermore, an analysis of the residuals might be useful in identifying the data points for which the models fit poorly. However, recent research conducted (Hosmer and Lemeshow 2010) has highlighted the difficulties of defining what a large residual is in the context of logistic models, as it highly depends on the type of data involved. Therefore, the analysis of the residuals was left out of the scope of this tutorial.

6 Results

Table 6 gives the results of the Brant test. Notice that the χ^2 test, at the top of the table, indicates that the null hypothesis, stating that the model parameters are equal across categories (i.e. parallel regression assumption), can be rejected at the 0.0001 level. In general, when statistical assumptions, such as the parallel regression assumption, are broken as they are in this data set, models based on the parallel assumption for all the independent variables cannot be accurately applied to the whole population (the parameters of the model are said to be biased). In other words, it is not possible to draw conclusions about the population, although valid estimates of the models can be generated. Consequently, these models where the parallel assumption is imposed to all their independent variables (ORM, CRM and ACM) were excluded from further analysis. Notice that not all

Table 6 Brant test for parallel assumption

	χ^2	$p > \chi^2$	df
All	153.17	0	11
Female	19.23	0	1
Previous admissions	0.23	0.628	1
Outpatient clinic	2.57	0.109	1
General surgery ward	4.89	0.027	1
Diagnosis category 2	11.65	0.001	1
Diagnosis category 3	31.97	0	1
S. procedure category 2	2.47	0.116	1
S. procedure category 3	0.78	0.377	1
S. procedure category 4	5.82	0.016	1
The results in bold correspond to the coefficients for all variables being tested simultaneously	Number of s. procedures	13.97	0
	Age centred	1.84	0.175

the variables violate the assumption. This should be taken into account when imposing constraints to the partial models.

Table 7 displays the odds ratios estimates for the six remaining models. Notice that for all models, gender, surgical procedure category 3 and number of surgical procedures to undergo were not significant predictors of short to medium LoS. However, gender and number of surgical procedures to undergo then became significant predictors of long LoS. For example, it seems that gender might not be an initial factor when predicting whether a patient will be more than 4 days hospitalised. But being female might increase the odds of staying more than 12 days.

The first two columns of Table 7 display the parameter estimates for the GOLM. The first column contrasts short LoS with categories medium and long LoS, and the second column contrasts categories short and medium LoS with category long LoS. In terms of interpretation, an $OR \geq 1$ indicates that higher values in the predictors make it more likely that the patient belongs to an upper LoS category than the current one, while an $OR < 1$ indicates that higher values on the independent variable increase the likelihood of belonging to the current category or to a lower one (Williams 2007). For example, the OR of diagnosis category 2 (e.g. insulin-dependent diabetes mellitus, hepatic failure and stroke) is higher in the first section, indicating that a patient with a diagnosis category 2 is more likely (2 times more) to have a medium or long LoS rather than a short one. The OR of outpatient clinic is less than 1, so to ease the interpretation, we take their inverse. The higher value in the second column indicates that a patient who is referred to hospitalisation from the outpatient clinic is more likely (4.2 times more) to have a short or medium LoS rather than a long one.

The next two columns of Table 7 display the odds ratios for the SEQM. The first column contrasts short LoS with categories medium or long LoS, and the second column contrasts category medium LoS with category long LoS. In terms of interpretation, an $OR \geq 1$ indicates that higher values in the predictor make it more likely that the patient progresses to an upper LoS category than the current one, while an $OR < 1$ indicates that higher values on the independent variable increase the likelihood of not progressing to the next category. For example, the OR of diagnosis category 2 (e.g. insulin-dependent diabetes mellitus, hepatic failure and stroke) is higher in the first section, indicating that a patient with a diagnosis category 2 is more likely (2 times more) to progress to a medium or long LoS rather than a short LoS. The OR value of the coefficient of outpatient clinic is higher in the first section (after taking the inverse), indicating that a patient who is referred to hospitalisation from the outpatient clinic is more likely (3 times more) to have a short LoS rather than a medium or long one.

The next model is the MNLM where short LoS is the reference category (i.e. STATA usually picks the category with the highest frequency to be the reference category, but this can be easily modified). A patient with a diagnosis category 3 is three times more likely to have a short LoS rather than a medium LoS. A patient who is referred to hospitalisation from the outpatient clinic is 5 times more likely to have a short LoS than a long one. A patient undergoing a surgical procedure category 4 is 2.5 times more likely to have a long LoS rather than a short one.

The next two models (PPOM and PCRM) have beta coefficients varying across the categories only for those variables that violate the parallel assumption, i.e. gender, general surgery ward, diagnosis category 2 and 3, surgical procedure category 4 and number of surgical procedures to undergo. Thus, the parameter estimates for the rest of the variables (i.e. constrained variables) are the same in both columns. In the PPOM, a patient undergoing a surgical procedure category 4 is almost 60 % more likely to have a longer LoS. The odds of having a shorter LoS are three times higher for patients who are referred to

Table 7 Odds ratio estimates for all models

	GOLM		SEQM		MNLM		PPOM		PCRM		SORM	
	Medium or long LoS versus short LoS	Long LoS versus short LoS	Long LoS versus medium LoS	Long LoS versus short LoS	Medium LoS versus short LoS	Long LoS versus short LoS	Medium or long LoS versus short LoS	Long LoS versus short LoS	Long or medium LoS versus short LoS	Long LoS versus medium LoS	Long LoS versus short LoS	
Female	1.07	1.52*	1.07	1.50*	1.01	1.51*	1.07	1.51*	1.07	1.54*	1.08	
Previous admissions clinic	0.97*	0.97*	0.97*	0.99*	0.97*	0.96*	0.97*	0.97*	0.97*	0.97*	0.97*	
Outpatient clinic	0.34*	0.24*	0.34*	0.50*	0.38*	0.18*	0.34*	0.34*	0.36*	0.36*	0.33*	
General surgery ward	0.61*	0.81	0.60*	1.06*	0.60*	0.62*	0.6*	0.83	0.60*	1.11	0.59*	
Diagnosis category 2	2.04*	1.46*	2.04*	1.05*	2.04*	2.10*	2.03*	1.48*	2.06*	1.02*	2.11*	
Diagnosis category 3	0.36*	0.74*	0.36*	1.48*	0.33*	0.52*	0.36*	0.77*	0.34*	1.85*	0.34*	
S. procedure category 2	0.71*	0.91	0.70*	1.24*	0.68*	0.85	0.72*	0.72*	0.77*	0.77*	0.69*	
S. procedure category 3	0.89	1.01	0.89	1.12*	0.87	1.01	0.90	0.90	0.92	0.92	0.89	
S. procedure category 4	1.50*	2.07*	1.50*	1.66*	1.38*	2.45*	1.56*	1.56*	1.52*	1.52*	1.54*	
Number of s. procedures	0.96	0.67*	0.96	0.67*	1.01	0.68*	0.95	0.72*	0.94	0.74*	0.95	
Age centred cons	1.01*	1.01*	1.01*	1.01*	1.01*	1.02*	1.01*	1.01*	1.01*	1.01*	1.01*	
	1.03	0.07	1.04	0.15	0.90	0.13	1.03	0.07	1.04	0.14		

* indicates significant coefficients at the 0.05 level

hospitalisation from the outpatient clinic. Conversely, more specific comparisons can be made for the variables which were set free of constraints: the higher coefficient of diagnosis category 2 (e.g. insulin-dependent diabetes mellitus, hepatic failure and stroke) in the first column indicates that patient with a condition classified under that diagnosis category is two times more likely to have a medium or long LoS rather than a short one.

The next model is the PCRM, from which estimates were almost identical to the PPOM. It is just the interpretation which slightly changes. For example, a patient with a diagnosis category 2 is two times more likely to progress to medium or long LoS.

The last two columns show the output for the SORM. The parameters can be interpreted in terms of the odds of the reference category versus the first category. STATA usually selects the last category (long LoS) as the reference category, but this can be easily changed in the command line. For example, the odds of having a long LoS versus a short LoS are twice as high for patients with a disease classified in the diagnosis category 2 (e.g. insulin-dependent diabetes mellitus, hepatic failure and stroke), The odds of having a short versus a long LoS are three times higher for patients who are referred to hospitalisation from the outpatient clinic.

Table 8 displays the goodness of fit and performance measurement. The model that represents the best fit according to the log-likelihood and AIC (by a negligible margin) is the MNLM. However, in terms of BIC, the best model is the PPOM. The absolute lower value of BIC could indicate a better fit or the presence of fewer parameters; it penalises free parameters more strongly than AIC. The second part of the table shows the accuracy rates on the validation set, giving an idea of how well the models do in predicting new patients. The six models perform well in predicting patients with short LoS. However, the models failed to predict any patient with a long LoS. Correct discrimination between categories is only possible if categories are essentially different in terms of the predictors (Ashby et al. 1986), which may suggest that collecting more data is appropriate for our illustrative example. Only when more data are not available is it acceptable to combine adjacent categories (e.g. medium and long LoS) and use conventional binary logistic regression instead to improve classification performance.

7 Discussion

When choosing the most appropriate model, there are some main points to consider: MNLM is a very popular model, and there is a wide selection of software on the market available for its implementation, which is naturally an advantage. However, the biggest

Table 8 Comparative chart logistic regression models

Model	Log-likelihood	AIC	BIC	Accuracy rate			
				Short (%)	Medium (%)	Long (%)	Overall (%)
GOLM	−8959.0	17,966.0	18,145.7	86.6	48.5	0	70.6
SeqM	−8955.2	17,958.3	18,138.1	80.6	60.6		73.6
MNLM	−8954.2	17,956.4	18,136.2	86.6	48.5	0	70.6
PPOM	−8963.8	17,963.5	18,098.3	86.6	48.5	0	70.6
PCRM	−8966.7	17,969.5	18,104.3	80.6	60.6		73.6
SORM	−8991.9	18,011.8	18,116.6	86.7	48.2	0	

drawback of the MNLM is that it ignores the clearly ordinal nature of the data which hinders the ability to assess effect directionality and progression (Cliff 1996).

The models based on the parallel assumptions are the only ones that strictly maintain the ordinality of the outcomes (McCullagh 1980). The rest of the models only retain the ordinal nature of the outcome to some extent (except for the MNLM that completely ignores the ordinal nature). Another advantage of the models based on the parallel assumption is their parsimony compared to most of the models presented here. However, the parallel assumption is rarely fulfilled and these models cannot be applied. SORM is another parsimonious model and easy to analyse, but it has been frequently associated with highly biased estimates and dubious identifiability when more than four outcome categories are involved (Holtbrugge and Schumacher 1991). The partially constrained models have less parameter than the fully unconstrained models. Both AIC and BIC take into account parsimony when evaluating the model goodness of fit. For our particular case, the more parsimonious models did not outperform those models with more parameters.

Furthermore, the choice of model should depend on the research question: if the main goal is classification and prediction, it is not imperative to select a model that preserves ordinality but to find which minimises cost of misclassification. For example, when categories are ordered, misclassification to an adjacent category should be always preferred to misclassification to a more extreme one (Ashby et al. 1986). There is strong evidence that suggests the MNLM outperforms ordinal models under a variety of circumstances (Campbell et al. 1991). For our illustrative example, more data need to be collected if any of these models were to be used for prediction and classification purposes. If collecting more data is not feasible, combining adjacent categories (e.g. medium and long LoS) and using conventional of binary logistic regression instead might be acceptable.

However, if the main goal is to understand the nature and direction of the predictor effects by exploiting the ordinal nature of the outcome, one should pick the model that best interprets the event of interest under study, especially in our case where there was little difference in terms of goodness of fit and performance between the six models.

The ACM, MNLM and SORM provide one by one comparison of categories. These models can be of particular use when hospitals plan their resources around a specific type of patients like short stays. It might be the case that patients with short and medium LoS are no different in terms of the level of care they need. Therefore, planners might be more interested in identifying factors for those patients who might have a long LoS. These three models allow direct comparisons with that category.

The ORM, GOLM and PPOM reveal the likelihood of moving in a certain direction (the patient moves towards either longer LoS or shorter LoS). They are very useful in identifying trends in the odds, and for the illustrative example shown here, understanding patient progression regarding to their LoS category, either upward or downward, brought more valuable information than comparing against a fixed category (e.g. MNLM or SORM).

The CRM, SeqM and PCRM reveal the likelihood of moving in certain direction assuming that patients must have pass through lower categories. This might be particularly useful if there is an interest in understanding patient pathways at hospital.

8 Conclusion

This tutorial presents a synthesised review of nine different models from the family of logistic regression for the analysis of ordered outcomes. The three first models ORM, CRM and ACM are developed under a strong assumption (i.e. the parallel regression assumption)

that is rarely fulfilled. Accordingly, other models are presented as viable alternatives, including GOLM, SEQM, MNLM, PPOM, PCRM and SORM. During model validation, performance measurements indicated that the six models had analogous performance when predicting new data. Moreover, the direction of the effects estimates were very similar between models, although the interpretation of the odds ratios varied from one model to another.

Consequently, other factors should be taken into account when choosing the most appropriate model, such as simplicity, number of model parameters or software availability. Most importantly, the choice should be based on the user's research question and event under study. The models presented here have proved to be computationally inexpensive and easy to estimate, analyse and interpret. Therefore, there is no reason to maintain frequently used practices, such as using nominal models or binary regression models, on combined adjacent categories.

Acknowledgments This research was supported by Consejo Nacional de Ciencias y Tecnología (CON-ACyT), Mexico.

References

- Ananth, C., Kleinbaum, D.: Regression models for ordinal responses: a review of methods and applications. *Int. J. Epidemiol.* **26**(6), 1323–1333 (1997)
- Anderson, J.: 'Regression and ordered categorical variables'. *J R Stat Soc Ser B (Methodol)* **46**(1), 1–30 (1984)
- Anderson, J., Philips, P.: Regression, discrimination and measurement models for ordered categorical variables. *Appl. Stat.* **30**(1), 22–31 (1981)
- Armstrong, B., Sloan, M.: Ordinal regression models for epidemiologic data. *Am. J. Epidemiol.* **129**(1), 191–204 (1989)
- Arrow, K.: *Social choice and individual values*, 2nd edn. Yale University Press, New Haven (1963)
- Ashby, D., Pocock, S., Shaper, A.: Ordered polytomous regression: an example relating serum biochemistry and haematology to alcohol consumption. *Appl. Stat.* **35**(6), 289–301 (1986)
- Bender, R., Grouven, U.: Using binary logistic regression models for ordinal data with non-proportional odds. *J. Clin. Epidemiol.* **51**(10), 809–816 (1998)
- Campbell, M., Donner, A., Webster, K.: Are ordinal models useful for classification? *Stat. Med.* **10**(3), 383–394 (1991)
- Cliff, N.: *Ordinal methods for behavioral data analysis*. Lawrence Erlbaum Associates, Mahwah (1996)
- Clogg, C., Shihadeh, E.: *Statistical models for ordinal variables*. Sage, Thousand Oaks (1994)
- Cole, S.R., Ananth, C.V.: Regression models for unconstrained, partially or fully constrained continuation odds ratios. *Int. J. Epidemiol.* **30**(6), 1379–1382 (2001)
- Defays, D.: An efficient algorithm for a complete link method. *Comput. J.* **20**(4), 364–366 (1977)
- Dobbin, K., Simon, R.: Optimally splitting cases for training and testing high dimensional classifiers. *BMC Med. Genomics* **4**(1), 31 (2011)
- Fienberg, S.: *The analysis of cross-classified categorical data*. MIT Press, Cambridge (1977)
- Flaatten, H., Bonde, J., Ruukonen, E., Winsø, O.: Classification for coding procedures in the intensive care unit. *Acta Anaesthesiol. Scand.* **46**(8), 994–998 (2002)
- Fullerton, A.S.: A conceptual framework for ordered logistic regression models. *Sociol. Methods Res.* **38**(2), 306–347 (2009)
- Goodman, L.A.: 'The analysis of dependence in cross-classifications having ordered categories, using log-linear models for frequencies and log-linear models for odds'. *Biometrics* 149–160 (1983)
- Guzman Castillo, M.: *Modelling Patient Length of Stay in Public Hospitals in Mexico*. Ph.D. thesis, University of Southampton (2012)
- Hauser, R.M., Andrew, M.: 1. Another look at the stratification of educational transitions: the logistic response model with partial proportionality constraints. *Sociol. Methodol.* **36**(1), 1–26 (2006)
- Hausman, J., Mcfadden, D.: Specification tests for the multinomial logit model. *Econom. J. Econom. Soc.* **52**(5), 1219–1240 (1984)
- Hilbe, J.: *Logistic regression models*. CRC Press, Boca Raton (2009)

- Holtbrugge, W., Schumacher, M.: A comparison of regression models for the analysis of ordered categorical data. *Appl. Stat.* **40**(2), 249–259 (1991)
- Hosmer, D., Lemeshow, S.: *Applied logistic regression*. John Wiley, New York (2010)
- Kahn, L.M., Morimune, K.: 'Unions and employment stability: a sequential logit approach'. *Int. Econ. Rev.* **21**(7), 217–235 (1979)
- Keski-Rahkonen, A., Kaprio, J., Rissanen, A., Virkkunen, M., Rose, R.: Breakfast skipping and health-compromising behaviors in adolescents and adults. *Eur. J. Clin. Nutr.* **57**(7), 842–853 (2003)
- Lall, R., Campbell, M., Walters, S., Morgan, K.: A review of ordinal regression models applied on health-related quality of life assessments. *Stat. Methods Med. Res.* **11**(1), 49–67 (2002)
- Lansky, S., List, M., Lansky, L., Ritter-Sterr, C., Miller, D.: The measurement of performance in childhood cancer patients. *Cancer* **60**(7), 1651–1656 (2006)
- Long, S., Freese, J.: *Regression models for categorical dependent variables using Stata*. StataCorp LP, College Station (2006)
- Luce, R.: *Individual choice behavior : a theoretical analysis*. John Wiley & Sons Inc, New York (1959)
- Lunt, M.: Prediction of ordinal outcomes when the association between predictors and outcome differs between outcome levels. *Stat. Med.* **24**(9), 1357–1369 (2005)
- Mahoney, F., Barthel, D.: Functional evaluation: the Barthel index. *Md. State Med. J.* **14**, 61–65 (1965)
- Mäntyselkä, P., Turunen, J., Ahonen, R., Kumpusalo, E.: Chronic pain and poor self-rated health. *JAMA* **290**(18), 2435–2442 (2003)
- Mare, R.D.: Social background composition and educational growth. *Demography* **16**(1), 55–71 (1979)
- McCullagh, P.: 'Regression models for ordinal data'. *J. R. Stat. Soc. Ser. B (Methodol.)* **42**(2), 109–142 (1980)
- McElroy, S.L., Frye, M.A., Suppes, T., Dhavale, D., Keck, P.E., Leverich, G.S., Althuler, L., Denicoff, K.D., Nolen, W.A., Kupka, R., Grunze, H., Walden, J., Post, R.M.: Correlates of overweight and obesity in 644 patients with bipolar disorder. *J. Clin. Psychiatry* **63**(3), 207–213 (2002)
- Newman, A., Foster, G., Givelber, R., Nieto, F., Redline, S., Young, T.: Progression and regression of sleep-disordered breathing with changes in weight: the sleep heart health study. *Arch. Intern. Med.* **165**(20), 2408–2413 (2005)
- O'Connell, A.A.: *Logistic regression models for ordinal response variables*. Sage, Thousand Oaks (2010)
- Peterson, B., Harrell, F.: Partial proportional odds models for ordinal response variables. *Appl. Stat.* **39**(2), 205–217 (1990)
- Rezanková, H.: Cluster analysis and categorical data. *Statistika* **89**(9), 216–232 (2009)
- Small, K., Hsiao, C.: Multinomial logit specification tests. *Int. Econ. Rev.* **26**(3), 619–627 (1985)
- Takazawa, K., Arisawa, K., Honda, S., Shibata, Y., Saito, H.: Lower-extremity muscle forces measured by a hand-held dynamometer and the risk of falls among day-care users in Japan: using multinomial logistic regression analysis. *Disabil. Rehabil.* **25**(8), 399–404 (2003)
- Walker, S., Duncan, D.: Estimation of the probability of an event as a function of several independent variables. *Biometrika* **54**(1–2), 167–179 (1967)
- Walston, J., McBurnie, M., Newman, A.: Frailty and activation of the inflammation and coagulation systems with and without clinical comorbidities: results from the cardiovascular health study. *Arch. Intern. Med.* **162**(20), 2333–2341 (2002)
- Ward, J.: Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58**(301), 236–244 (1963)
- Williams, R.: Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *Stata J.* **6**(1), 58–82 (2007)
- Wilson, P.W.F., D'Agostino, R.B., Levy, D., Belanger, A.M., Silbershatz, H., Kannel, W.B.: Prediction of coronary heart disease using risk factor categories. *Circulation* **97**(18), 1837–1847 (1998)