



Ordinal outcome model with medico- administrative data

Valéry Bocquet, PhD

Competence Center in Methodology and Statistics
Centre de Recherche Public de la Santé (CRP-Santé)
Luxembourg

Table of contents

1) Medico-administrative database ?

2) Ordinal data : usual data ?

3) Presentation of the different regression models applied to ordinal data

4) Application to a medico-administrative database

1) Medico-administrative database

Definition

Specificities

- Big database
- Longitudinal data
- Observations at non constant frequencies.

2) Ordinal data

For instance : *health status* with good / moderate / bad

Advantages of ordinal models vs binary logistic regression :

- increased power,
- a measure of effect that applies to all dichotomies of the outcome.

3) Regression models

Proportionality of predictors ?

Yes

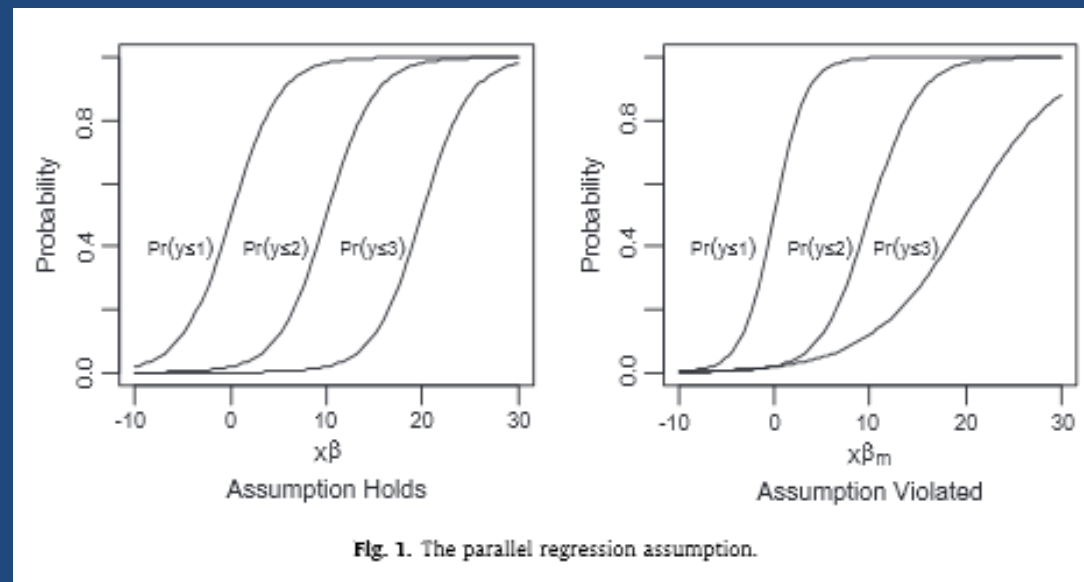


No



Proportional odds assumption

Do the effects of one or more independent variables significantly vary across cutpoints equations in the model?



(Fullerton, 2012)

3) Regression models

Proportionality of predictors ?

Yes

No

Cumulative logit model

(Walker & Duncan, 1967)

Or also called

Proportional odds model

(Mc Cullagh, 1980)

Continuation ratio model

(Feinberg, 1980)

Non proportional odds model:

Generalized Ordered Logit

(Mc Cullagh and Nelder, 1989)

Partial proportional odds model

(Peterson and Harrell, 1990)

IF YES

Cumulative logit model (Walker and Duncan, 1967)

Proportional odds model (Mc Cullagh, 1980)

$$\text{logit} [\pi(Y \leq j | x_1, x_2, \dots, x_p)] = \ln \left(\frac{\pi(Y \leq j | x_1, x_2, \dots, x_p)}{\pi(Y > j | x_1, x_2, \dots, x_p)} \right) = \alpha_j + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p;$$

Advantage :

Easily interpretable. Not affected by the direction chosen for modeling the response variable.

Drawback :

- Large samples can lead to a significant proportionality test.

IF YES

Continuation ratio model (Feinberg, 1980)

$$\ln \left(\frac{\pi(Y \geq j | x_1, x_2, \dots, x_p)}{\pi(Y = j | x_1, x_2, \dots, x_p)} \right) = \alpha_j + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

(Agresti, 2007; Allison, 1999; O'Connell, 2006)

- Analogous to the proportional hazards model of Cox but in discrete time.
- Well-suited for failure time data and outcomes measuring threshold points.
- Homogeneity of effect across cut-points => only one OR
- Affected by the direction chosen for modeling the response variable.

IF NO

Generalized ordered logit *(Mc Cullagh and Nelder, 1989)*

$$\text{logit} [\pi(Y \leq j | x_1, x_2, \dots, x_p)] = \ln \left(\frac{\pi(Y \leq j | x_1, x_2, \dots, x_p)}{\pi(Y > j | x_1, x_2, \dots, x_p)} \right) = \alpha_j + \beta_{1j}X_1 + \beta_{2j}X_2 + \dots + \beta_{pj}X_p;$$

Drawback :

- Numerous parameters to estimate. Problematic with small samples.

IF NO

Partial Proportional Odds Model

(Peterson and Harrell, 1990)

$$\text{logit} [\pi(Y \leq j | x_1, x_2, \dots, x_p)] = \ln \left(\frac{\pi(Y \leq j | x_1, x_2, \dots, x_p)}{\pi(Y > j | x_1, x_2, \dots, x_p)} \right) = \alpha_j + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p;$$

Advantage :

- Keep the ordinal aspect of the outcome
- Allowed to assume proportional odds for some predictors while not for others.

Application

(Renard, Bocquet et al., Plos One, 2013)

Aim : determine the attributes of the adherence to diabetes guidelines

Type of study : cohort

Sample size : more than 1 million reimbursements corresponding to between 12,000 and 17,000 patients each year

Length of follow-up : from 2000 to 2006

Software used : SAS System for Windows, version 9.2; SAS Institute Inc., Cary, NC

Outcome : adherence to 6 criteria over the 2000-2006 period

Explanatory variables :

Type of variables	Category	Subcategory
Continuous	time since the first A10 reimbursed in the dataset	
Binary	sex	Female, male
Categorical	type of A10 treatment	Insulin only, oral only, mixed
Categorical	type of treating physician	Diabetologist, Internist-D, GP-ID, Other-GID
Categorical	year	2000 to 2006

Methodology used

STEP 1 : descriptive analyses

STEP 2 : multivariable analysis

1) Linearity of continuous variables : As the linearity test was significant for each variable, these variables were not categorized.

2) Multiple imputation : As few data were missing, a multiple imputation was performed (PROC MI in SAS),

3) Ordinal logistic model

Outcome : adherence to 6 criteria over the 2000-2006 period

Type of outcome : ordinal and longitudinal

Methodology : Ordinal logistic model for repeated measures (PROC GENMOD repeated)



Analysis

Interaction : between time and several other explanatory variables



a year stratification was needed



the used methodology was an ordered logit model for each year of the period. (PROC LOGISTIC)

Proportionality : the assumption of proportional odds ratio was not met.



For each year, use of a partial proportional odds model (Peterson and Harrell, 1990). (PROC NLMIXED)

It means that the odds ratios can not be interpreted as constant across all possible cut points of the outcome.

Combine results from multiple datasets (PROC MIANALYZE).

Table: Odds ratios (95%CI) from cumulative odds models of adherence to guidelines (from 0: any adherence to 6: full adherence) between 2000 and 2006

Category	2000				2006			
	N	OR (95% CI)			N	OR (95% CI)		
		6 vs others	3,4,5,6 vs others	1,2,3,4,5,6 vs 0		6 vs others	3,4,5,6 vs others	1,2,3,4,5,6 vs 0
A10 duration, y	12626	1.04 (0.71-1.52)	1.03 (1.01-1.06)	1.00 (0.97-1.02)	16602	0.99 (0.99-1.00)	1.09 (1.03-1.15)	1.04 (1.03-1.05)
Sexe								
Female	6231	1	1	1	7718	1	1	1
Male	6395	0.30 (0.08-1.08)	0.79 (0.71-0.87)	0.81 (0.74-0.88)	8884	0.87 (0.83-0.92)	0.87 (0.83-0.92)	0.87 (0.83-0.92)
TP								
D	948	1	1	1	1422	1	1	1
I-D	3201	0.62 (0.15-2.60)	0.95 (0.80-1.14)	0.91 (0.73-1.12)	4100	0.75 (0.42-1.32)	0.61 (0.53-0.69)	0.72 (0.59-0.89)
G-ID	7872	0.15 (0.03-0.83)	0.68 (0.57-0.80)	0.56 (0.46-0.68)	10423	0.36 (0.20-0.65)	0.47 (0.42-0.53)	0.50 (0.41-0.60)
O-GID	605	0.53 (0.02-14.46)	0.56 (0.41-0.75)	0.83 (0.63-1.10)	657	0.18 (0.03-1.10)	0.40 (0.32-0.49)	0.42 (0.32-0.55)
A10 treatment								
Insulin only	1302	1	1	1	1658	1	1	1
Oral only	9974	0.83 (0.15-4.52)	1.11 (0.94-1.30)	1.41 (1.22-1.64)	13142	1.48 (1.35-1.63)	1.48 (1.35-1.63)	1.48 (1.35-1.63)
Mixed	1350	1.29 (0.22-7.44)	1.12 (0.92-1.37)	0.96 (0.80-1.17)	1802	1.35 (1.20-1.52)	1.35 (1.20-1.52)	1.35 (1.20-1.52)

A10 duration : A10 treatment duration; TP : Treating Physician; D : Diabetologist; I-D: Internist and not diabetologist; G-ID : General practitioner and not I nor D; O-GID : other practitioner and not G, I, D.

Comparison of the estimated odds ratio according to different models

		2000 – 3,4,5,6 vs others			2006 – 3,4,5,6 vs others		
		Partial proportional odds model	Non proportional odds model	Proportional odds model	Partial proportional odds model	Non proportional odds model	Proportional odds model
A10 duration		1,03 [1,01-1,16]	0,97 [0,64-1,48]	1,05 [0,70-1,57]	1,09 [1,03-1,15]	1,08 [1,02-1,15]	1,08 [1,01-1,15]
Sexe							
	Female	1	1	1	1	1	1
	Male	0,79 [0,71-0,87]	0,79 [0,71-0,87]	0,29 [0,08-1,05]	0,87 [0,83-0,92]	0,86 [0,80-0,91]	0,91 [0,59-1,43]
TP							
	D	1	1	1	1	1	1
	I-D	0,95 [0,80-1,14]	0,93 [0,78-1,12]	0,12 [0,03-0,49]	0,61 [0,53-0,69]	0,60 [0,53-0,68]	0,39 [0,20-0,75]
	G-ID	0,68 [0,57-0,80]	0,66 [0,56-0,79]	0,02 [0,00-0,10]	0,47 [0,42-0,53]	0,47 [0,42-0,53]	0,17 [0,09-0,32]
	O-GID	0,56 [0,41-0,75]	0,54 [0,40-0,73]	0,01 [0,00-4,93]	0,40 [0,32-0,49]	0,39 [0,32-0,48]	0,05 [0,01-0,32]
A10 treatment							
	Insulin only	1	1	1	1	1	1
	Oral only	1,11 [0,94-1,30]	1,12 [0,95-1,32]	0,81 [0,16-4,13]	1,48 [1,35-1,63]	1,43 [1,27-1,60]	2,58 [1,20-5,57]
	Mixed	1,12 [0,92-1,37]	1,14 [0,94-1,40]	0,77 [0,13-4,70]	1,35 [1,20-1,52]	1,30 [1,13-1,50]	4,00 [1,80-8,89]

Conclusion

The ordinal outcome is very frequent in medical studies.

It is very important to use an adapted method (wrong estimation).

A methodology available in SAS and in other statistical softwares.

References

- Agresti A. *A survey of models for repeated ordered categorical response data*. Stat Med; 8:1209-1224, 1989.
- Feinberg S. *The analysis of cross-classified categorical data*. Cambridge, MA: MIT Press; 110-116, 1980.
- McCullagh P. *Regression models for ordinal data (with discussion)*. J R Stat Soc Series B; 42:109-42, 1980.
- McCullagh P and Nelder JA. *Generalized Linear Models*. Second Edition, Chapman & Hall, 37, 1989.
- Peterson B and Harrell FE. *Partial proportional odds models for ordinal response variables*. Appl Stat, 39:205-17, 1990.
- Renard L, Bocquet V, Vidal-Trecan G, Lair ML, Blum-Boisgard C. *Adherence to International Follow-up Guidelines in Type 2 Diabetes: A Longitudinal Cohort Study in Luxembourg*. Plos One, 2013.
- Walker S and Duncan D. *Estimation of the probability of an event as a function of several independant variables*. Biometrika; 54: 167-179, 1967.