Electronic Journal of Applied Statistical Analysis
EJASA (2012), Electron. J. App. Stat. Anal., Vol. 5, Issue 3, 387 – 392
e-ISSN 2070-5948, DOI 10.1285/i20705948v5n3p387

LONGITUDINAL RASCH MODEL FOR SELF ASSESSMENT OF SIDE EFFECTS IN CHEMOTHERAPY CYCLES

Laura Pagani[(1)], Maria Chiara Zanarotti[*(2)]

[(1)]*Department of Statistical Sciences, University of Udine, Italy*
[(2)]*Department of Statistical Sciences, Catholic University of Milan, Italy*

Received 01 August 2012; Accepted 09 November 2012
Available online 16 November 2012

***Abstract**: In this paper longitudinal responses data are considered and it is showed how to benefit of the use of multilevel setting to model data structure. In fact, both longitudinal and item responses data can be ascribed to hierarchical generalized linear model, so that the whole data complexity is effectively modeled in a 3-level multilevel structure. An application to chemotherapy side effects is showed.*

***Keywords**: Longitudinal data, hierarchical generalized linear model, multilevel rasch model, chemotherapy side effects.*

1. Introduction

Longitudinal data belong to the more general family of correlated data. Correlated data embraces a large set of data structures, like multivariate observations, clustered data, repeated measures, spatially correlated data and, in particular, longitudinal data [9]. Despite of the common membership, statistical techniques for longitudinal data analysis are specifically devoted for this particular research perspective. A dataset is longitudinal if it records the same type of information on the same statistical units at multiple points in time and time itself is object of scientific investigation. Even if there are several statistical modeling techniques to analyze repeated measures data, Multilevel Model (MLM, or Hierarchical Linear Model) appears as a very helpful tool (see, for example, [1] and [8]) because it takes into account of the hierarchical setting of longitudinal data: the development over time is modelled by a linear regression equation and each individual gets his own growth curve, specified by individual regression coefficients that may depend on individual attributes [3]. Another advantage of MLM in

[*] E-mail: maria.zanarotti@unicatt.it

longitudinal data analysis is the fact that MLM allows to managing incomplete time data where number of time points can vary across individuals and time points need not to be equally spaced. Data considered in this contribution are obtained by administering, on multiple occasions, a set of items included in a questionnaire devoted to measure, through ordinal responses, an individual latent trait, *i.e.* a personal characteristics not directly observable. Goals of the analysis are to obtain person and item measures that locate each individual and each item into a unique continuous interval scale and to study the behavior of such measures across time points. To obtain items and persons measures on the same interval scale, Rasch Model (RM) appears to be a good tool because of its capacity to estimate both persons and items parameters on an interval scale. Also, as pointed out by several authors, RM can be interpreted as a member of Hierarchical Generalized Linear Model (HGLM) considering items as first-level units and persons as second-level units, in a fully crossed design (see, for example, [4], [7] and [2]). The belonging of RM to HGLM has many advantages, and, in particular, the possibility to easily add others levels to the model to take into account the dependency among observations when lower-level units are nested within over-setting. Since data considered are repeated measures taken in different time points, the HGLM for Repeated Measures [5] with three levels appears a suitable model to accommodate data structure, where: first level is represented by items, second level is represented by time (taking values 0, 1, 2, …) and third level is represented by persons. The HGLM for Repeated Measures is described in next section and an application of the model to investigate malaise severity and site effects during chemotherapy cycle will close the paper.

2. Longitudinal Ordinal Rasch model in MLM framework

Kamata [4] has established the equivalence between dichotomous Rasch model and two levels HGLM. After his contribution, other authors (see, for example, [5]) have extended Kamata's result to Polytomous Rasch Model (PRM), and, in particular, to the case in which response categories are ordinal. Denoting with Y_{ij} the response of person i (i=1,…, n) to item j (j=1,…, J) and with k (k=1,…, K) the k-th ordinal response category, RM implies that conditional probability of Y_{ij} is a function of three sets of parameters (person, item and threshold parameters). In particular, considering the so called *cumulative logit model* and denoting with p_{ijk} the probability that Y_{ij} takes values higher than or equal to k, *i.e.* $p_{ijk}=P(Y_{ij}\geq k)$, PRM model assumes the form:

$$\ln\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\alpha_i-\theta_j+\tau_k \qquad (1)$$

This model is the so called *rating scale model* (RSM), where α_i-s are person parameters (measuring the so called latent trait), θ_j-s are item parameters (constant across persons) and τ_k-s are threshold parameters (constant across persons and across items). The HGLM formulation of RSM is obtained by considering, for person i, a dummy variable X_{qij} taking value 1 when q=j and 0 if $q\neq j$ for item j, q=1,…, J-1 (the dummy variable for the last item is dropped to obtain a full rank matrix). Then the PRM assume the form:

$$\ln\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\beta_{0i}+\delta_{ki}+\sum_{q=1}^{J-1}\beta_{qi}X_{qij} \qquad \text{and} \qquad \begin{cases}\beta_{0i}=\gamma_{00}+u_{0i}\\ \beta_{qj}=\gamma_{q0}\\ \delta_{ki}=\delta_{k}\end{cases} \tag{2}$$

Right side of model (2) alone is a structural model, called *level-1* or *item-level model*: at this level β-s are item parameters, were β_{0i} (the intercept term) is associated to dropped item (the last) and can be interpret as the expected item effect of this item for person i. The other β_{qi} parameters are specific effects of items j (j=1,..., J-1), expressed in term of difference form β_{0i}. Threshold parameters are denoted by δ_{ki}. Note that, at item-level model, item parameters are not constant across persons. A level-2 model is obtained adding the left-side of the (2): this two-level logistic model is equivalent to RSM where u_{0i} are person parameters, fixed across items, γ-s are item parameters, fixed across persons, and δ-s are threshold parameters, fixed across both persons and items. While in RM both item and person are fixed effect, in HGLM person effects u_{0i} are random with distribution N[0,var(u_{0i})]. Putting all together the two sides of the (2), the following model for item q is obtained:

$$\ln\left(\frac{p_{iqk}}{1-p_{iqk}}\right)=(\gamma_{00}+\gamma_{q0})+\delta_{k}+u_{0i} \tag{3}$$

where parameters $(\gamma_0+\gamma_{q0})$, δ_k and u_{0i} are equivalent, respectively, to α_i , τ_k and θ_j of model (1). Finally, to consider longitudinal data, a third level is added. The three-level model assumes the form:

$$\ln\left(\frac{p_{ijkm}}{1-p_{ijkm}}\right)=\beta_{0mi}+\delta_{kmi}+\sum_{q=1}^{J-1}\beta_{qmi}X_{qjmi}\,,\ \begin{cases}\beta_{0mi}=\gamma_{00i}+\gamma_{01i}d_{mi}+u_{0mi}\\ \beta_{qmi}=\gamma_{q0i}+\gamma_{q1i}d_{mi}\\ \delta_{kmi}=\delta_{ki}\end{cases} \quad \text{and} \quad \begin{cases}\gamma_{00i}=\pi_{000}+r_{00i}\\ \gamma_{01i}=\pi_{010}+r_{01i}\\ \gamma_{q0i}=\pi_{q00}\\ \gamma_{q1i}=\pi_{q10}\\ \delta_{ki}=\delta_{k}\end{cases} \tag{4}$$

where: d_{mi} (m=1,…, M) is time variable coded so that it takes value 0 for m=1; subscript m indicates second-level units; u_{0mi}~N[0,var(u_{0mi})], with variance of u_{0mi} constant for every third-level units; note that, at level-2, item location β_{qmi} are supposed nonrandom function of time. Model (4) allows the third-level mean effect to vary across units, so that it models variation in growth trajectories among persons. In fact, while item location's intercept and linear coefficient for time remain constant across both the second and the third-level units, the latent trait parameters can vary randomly across persons. The π-s parameters represent: π_{000} the average overall latent trait at time 0; π_{010} the overall linear time effect; π_{q00} overall item location at time 0; π_{q10} is the overall change in item location over time. The random effects r_{00i} and r_{01i} are assumed $\mathbf{N}[\mathbf{0},\boldsymbol{\Sigma}]$, where $\boldsymbol{\Sigma}$ is the variance-covariance matrix. Combining the three sides of the (4), model for item q can be obtained in a similar way as in (3).

3. Application

The model proposed in the previous section is applied to observational cancer data. A group of 88 women with breast cancer were asked to fill a questionnaire, administrated at the end of six consecutive chemotherapy cycles, to self assess their level of malaise. The tool consists of 15 items that refer to both physical (pain, nausea and vomiting, fatigue, dry mouth, swallowing difficulty, diarrhoea, constipation, alopecia) and psychological (insomnia, sleepiness, anger, anguish, sadness, anxiety, depression) side effects. Responses regard, for each side effect, the level of malaise in a 4-point Likert scale: 1=not at all, 2=slight, 3=a lot and 4=very much. There are no missing data. The number of observations in the data set is 7920 (responses of 88 women to 15 items in 6 cycles). The aim of this analysis is to investigate how the patient malaise (the latent trait or person parameters) changes over time assuming that side effects severity (item parameters) are fixed across chemotherapy cycles and patients. The program GLLAMM [6], was used to perform the analysis. Although model (4) is very flexible, with the possibility to easily add higher levels and individual and/or contextual covariates, it is very computer intensive. So it is useful to follow a step-by-step procedure to estimate the model: at each step estimated parameters are saved and then used as initial values for the next, more complex, model.

Table 1. Parameters estimate of the final model.

Fixed effects	**Coefficient (SE)**	**Standard Error**
Swallowing difficulty	-2.994	0.123
Diarrhea	-2.600	0.112
Insomnia	-2.325	0.109
Constipation	-2.087	0.102
Dry mouth	-0.632	0.083
Sadness	0.610	0.080
Anger	0.641	0.081
Fatigue	0.720	0.087
Sleepiness	1.170	0.083
Depression	1.305	0.082
Anguish	1.558	0.082
Anxiety	2.226	0.084
Alopecia	2.412	0.085
Nausea	2.935	0.086
Chemotherapy cycle 2	0.761	0.085
Chemotherapy cycle 3	0.958	0.086
Chemotherapy cycle 4	1.563	0.087
Chemotherapy cycle 5	2.073	0.089
Chemotherapy cycle 6	2.855	0.094
Thresholds paramaters	**Threshold**	**Standard Error**
δ_1	0.414	0.071
δ_2	2.702	0.079
δ_3	5.477	0.099
Random Effects	**Parameter**	**Standard Error**
$Var(r_{00i})$	0.083	0.031
$Var(r_{01i})$	0.002	0.001
$Cov(r_{00i},r_{01i})$	0.012	0.006
Deviance	14059.16	

Note: Reference categories are item: Pain, Chemotherapy cycle: cycle 1.

After a preliminary analysis of models deviances (not reported for brevity), the final model has: random intercept, random slope, overall item location parameters at time 0, the side effects and five different overall linear time effect parameters. Last parameters express the chemotherapy cycle effects (note that the time variable d_m is a categorical one with six levels). The parameters of the change in item location over time are not significant. Table 1 shows estimation for fixed parameters (side and chemotherapy cycle effects), thresholds and the variances of the random intercept and slope parameters. Results indicate that side effects, during the whole therapy, are all significant, even if psychological aspects are more severe than those physical, except for alopecia (see also Figure 1.). The coefficients of chemotherapy cycle effects indicate that the level of malaise increase during the therapy.

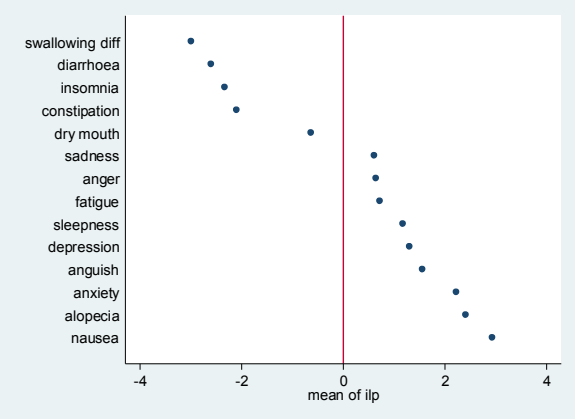

Figure 1. Side effects from the lowest to the highest intensity.

With regard to the random effects, the Wald tests, for the variance of the intercepts and the slops, are significant. This means that the variability between patients regarding the intercept (the average level of malaise at the beginning of the therapy) and the linear time effect (the change in malaise intensity as therapy occasions increase) is relevant. Note that the variance of the random slope is very small: this suggests that the trend of malaise severity is almost the same for all patients (see Figure 2).

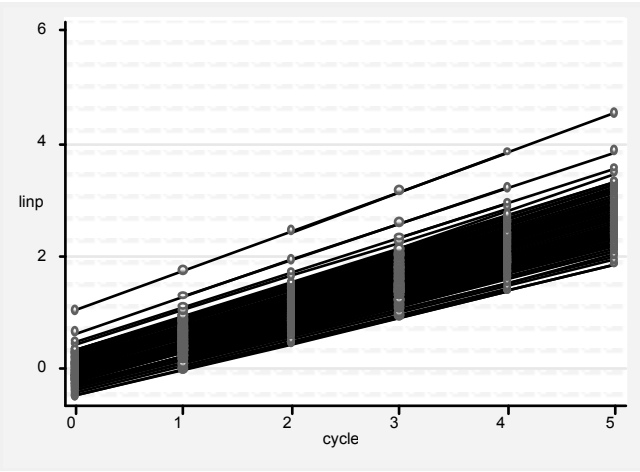

Figure 2. Patients latent growth curves.

The relationship between intercept and slope of patient growth curves ($Cor(r_{00i},r_{01i})=0.856$) is positive suggesting that patients with high initial level of malaise have a faster rate than those with initial moderate or low levels of malaise.

4. Concluding remarks

The advantage of considering the longitudinal Rasch model in the framework of the multilevel models is to allow flexibility in the data structure, with the possibility to easily add higher levels and individual and/or contextual covariates, and also to investigate the relationship between intercept and slope of the individual latent growth curves.

References

[1]. Bryk, A.G., Raudenbush, S.W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, 101,147-158.

[2]. Doran, H., Bates, D., Bliese, P. Dowling, M. (2007). Estimating the Multilevel Rash Model: With the lme4 Package, *Journal of Statistical Software*, 20, 1-18.

[3]. Hox, J.J. (2009). In: T.D. Little, K.U. Schnabel, & J. Baumert (Eds.) *Modeling longitudinal and multiple-group data: Practical issues, applied approaches, and specific examples*. Hillsdale, NJ: Erlbaum.

[4]. Kamata, A. (2001). Item Analysis by the Hierarchical Generalized Linear Model, *Journal of Educational Measurement*, 38, No.1, 79-93.

[5]. Pastor, D.A., Beretvas, S.A. (2006). Longitudinal Rasch Modeling in the Context of Psychotherapy Outcomes Assessment, *Applied Psychological Measurement*, 30, 100-120.

[6]. Rabe-Hesketh, S., Skrondal, A., Pickles, A. (2004) *GLLAMM Manual*. U.C. Berkeley Division of Biostatistics Working Paper Series, Working Paper 160

[7]. Raudenbush, S.W., Johnson, C., Sampson, R. J. (2003). A Multivariate, Multilevel Rasch model with application to self-reported criminal behavior, *Sociological Methodology*, 3 (1), 169-211.

[8]. Snijders, T. (1996). Analysis of longitudinal data using the hierarchical linear model, *Quality & Quantity*, **30**, 405-426.

[9]. Verbeke, G, Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York, Springer-Verlag.