



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



DYNAMIC CLASSIFICATION THROUGH THREE-STEP ESTIMATION: EVIDENCE FROM A MULTINATIONAL LONGITUDINAL STUDY OF MYASTHENIA GRAVIS PATIENTS

LUCA BRUSA¹, FULVIA PENNONI¹,
FRANCESCO BARTOLUCCI², LORENZO MAGGI³
(luca.brusa@unimib.it)

¹University of Milano-Bicocca, Department of Statistics and
Quantitative Methods (Milan, Italy)

²University of Perugia, Department of Economics (Perugia, Italy)

³Foundation IRCCS Neurological Institute Carlo Besta, Neuroimmunology and Neuromuscular
Diseases Unit (Milan, Italy)

Outline

- 1 Three-step estimation approach
- 2 The data: Myasthenia Gravis
- 3 Results
- 4 References

General setting

- **Response variables:** $\mathbf{Y}_i^{(t)} = (Y_{i1}^{(t)}, \dots, Y_{ir}^{(t)})'$
 - $Y_{ij}^{(t)}$ results from the administration of item j to subject i at time t
 - $Y_{ij}^{(t)}$ is categorical with c_j categories, labeled from 0 to $c_j - 1$
 - subsets of items measure different macro-areas
- **Individual covariates:** $\mathbf{X}_i^{(t)}$, observed for subject i at time t
- Hidden Markov (HM) model estimated using a three-step approach:
 - 1 multidimensional **Latent Class Graded Response (LCGR)** model
 - 2 **assignment** of each individual to the predicted latent class
 - 3 weighted estimation of the **effects of covariates** on the classification
- Main advantages:
 - overcoming of the challenges of full maximum likelihood estimation
 - inclusion of the multidimensional structure of the data

1st step - LCGR model: Notation

- **Responses** provided by the **same subject** at **different time occasions** are considered **independent**, as referred to distinct individuals:

$$\begin{array}{ccc}
 \mathbf{Y}_i^{(t)} & \longleftrightarrow & \mathbf{Y}_h \\
 \begin{array}{l} i=1,\dots,n \\ t=1,\dots,T \end{array} & & h=1,\dots,M
 \end{array}$$

- M is the total number of observations across all subjects and times
- $h = 1, \dots, M$ is defined on the basis of a one-to-one mapping of (i, t)
- **Vector of latent variables:** $\mathbf{U}_h = (U_{h1}, \dots, U_{hD})'$
 - corresponding to the D latent traits measured by the J items
 - assumed to follow a discrete distribution with k support points denoted as ξ_1, \dots, ξ_k

1st step - LCGR model: Parameters

- Model parameters:
 - **class probabilities**: $\pi_u = p(\mathbf{U}_h = \xi_u)$
 - **conditional response probabilities**: $\phi_{jy|u} = p(Y_{hj} = y | \mathbf{U}_h = \xi_u)$
- **Parameterization** ($u = 1, \dots, k, j = 1, \dots, J, y = 1, \dots, c_j - 1$):

$$\log \frac{p(Y_{hj} \geq y | \mathbf{U}_h = \xi_u)}{p(Y_{hj} < y | \mathbf{U}_h = \xi_u)} = \log \frac{\phi_{j,y|u} + \dots + \phi_{j,c_j-1|u}}{\phi_{j,0|u} + \dots + \phi_{j,y-1|u}} = \sum_{d=1}^D \xi_{ud} \delta_{jd} - \beta_{jy},$$

where:

- ξ_{ud} is the **support point** (or **ability level**) for latent class u and trait d
- δ_{jd} is an indicator variable equal to 1 if item j measures latent trait d
- β_{jy} is the **difficulty parameter** for category y of item j

2nd step - Assignment of subjects to latent classes

- Based on the model estimates, the **posterior distribution** of the latent variables given the responses is computed, for each observation $h = 1, \dots, M$:

$$p(\mathbf{U}_h = \xi_u | \mathbf{Y}_h = \mathbf{y}) = \frac{p(\mathbf{Y}_h = \mathbf{y} | \mathbf{U}_h = \xi_u) \pi_u}{p(\mathbf{Y}_h = \mathbf{y})}$$

- Longitudinal structure** of the data:

$$\hat{q}_{hu} = p(\mathbf{U}_h = \xi_u | \mathbf{Y}_h = \mathbf{y}) \quad \longleftrightarrow \quad \hat{q}_{iu}^{(t)} = p(\mathbf{U}_i^{(t)} = \xi_u | \mathbf{Y}_i^{(t)} = \mathbf{y})$$

- Units are assigned to a certain latent class at each time t , following the maximum a-posteriori approach:

$$\hat{u}_i^{(t)} = \arg \max_{u=1, \dots, k} \hat{q}_{iu}^{(t)}$$

3rd step - Effect of covariates

- Let $V_i^{(t)}$ denote the categorical variable with k categories referred to the latent class of subject i at time occasion t
- The distribution of $\mathbf{V}_i = (V_i^{(1)}, \dots, V_i^{(T)})$ is based on the **initial and transition probabilities**, which are affected by the covariates:

$$\log \frac{p(V_i^{(1)} = u | \mathbf{x}_i^{(1)})}{p(V_i^{(1)} = 1 | \mathbf{x}_i^{(1)})} = \beta_{0u} + \beta'_{1u} \mathbf{x}_i^{(1)}$$

$$\log \frac{p(V_i^{(t)} = v | V_i^{(t-1)} = u, \mathbf{x}_i^{(t)})}{p(V_i^{(t)} = u | V_i^{(t-1)} = u, \mathbf{x}_i^{(t)})} = \gamma_{0uv} + \gamma'_{1uv} \mathbf{x}_i^{(t)}$$

- The parameters $\mathbf{B} = (\beta_{0u}, \beta_{1u})$ and $\mathbf{\Gamma} = (\gamma_{0uv}, \gamma_{1uv})$ are then estimated **fitting two weighted multinomial logit models**

Outline

- 1 Three-step estimation approach
- 2 The data: Myasthenia Gravis
- 3 Results
- 4 References

Myasthenia Gravis

- **Myasthenia Gravis (MG)** is a neuromuscular autoimmune disorder characterized by extensive weakness and disability in the voluntary muscles (eye, bulbar, respiratory, axial, and limb muscles)
- **Symptoms:**
 - muscle weakness in the neck or limbs
 - difficulty swallowing or speaking
 - ocular symptoms (ptosis and diplopia)
 - excessive fatigue due to frequent activities
- The data for this analysis are based on an **observational study** (MyRealWorld-MG) conducted among adult patients with MG
- The study is multinational (nine countries), digital, and **longitudinal**; the sample is made of **1,702 patients** who repeatedly reported a variety of symptoms during the **years 2020-2023**

The 84 polytomous items

- Responses are provided for **84 polytomous** items grouped into 8 macro-areas

Section	Description
MG-ADL (8)	Symptoms severity of MG in activities of daily livings
HADS (14)	Levels of hospital anxiety and depression
EQ-5D-5L (11)	Generic quality of life
MG-QOL (15)	MG-related quality of life
SLEEP (6)	Overall sleep quality and disturbance
DYSPNEA (5)	Frequency, severity and duration of shortness of breath
FACIT (13)	Fatigue levels and impact days
HUI (12)	Health status and health-related quality of life

- Each item j has **a varying number c_j of categories** scored on an ordinal scale from 0 to $c_j - 1$ (lowest and highest burden of MG)

Outline

- 1 Three-step estimation approach
- 2 The data: Myasthenia Gravis
- 3 Results
- 4 References

Results: First step

- The LCGR model is estimated considering the 84 items; the **number of latent classes** is fixed a-priori and equal to $k = 3$
- To characterize the three patient **risk profiles** we use the following measure (**item mean score**):

$$\hat{\mu}_{j|u} = \frac{1}{c_j - 1} \sum_{y=0}^{c_j-1} y \hat{\phi}_{jy|u}$$

- Values of $\hat{\mu}_{j|u}$ **close to 0** correspond to **low probabilities that subjects in latent class u experience adverse conditions related to item j**
- Values of $\hat{\mu}_{j|u}$ **close to 1** correspond to **high probabilities that subjects in latent class u experience adverse conditions related to item j**

Results: Item mean score - Some examples

MG-QOL: worklimitations	0.23	0.62	0.81
MG-QOL: walking	0.06	0.34	0.66
MG-QOL: speaking	0.03	0.21	0.51
MG-QOL: socialactivity	0.14	0.50	0.83
MG-QOL: publicplaces	0.05	0.31	0.67
MG-QOL: plans	0.30	0.65	0.88
MG-QOL: overwhelmed	0.05	0.32	0.64
MG-QOL: independence	0.07	0.39	0.76
MG-QOL: hobbies	0.23	0.59	0.88
MG-QOL: grooming	0.02	0.17	0.49
MG-QOL: frustration	0.18	0.50	0.78
MG-QOL: family	0.07	0.39	0.75
MG-QOL: eyes	0.09	0.42	0.75
MG-QOL: eating	0.03	0.23	0.54
MG-QOL: depressed	0.06	0.34	0.63
	First latent class	Second latent class	Third latent class

- Patients with low risk (**1st class**) are associated with the **lowest** $\hat{\mu}_{j|u}$ **values** \Rightarrow represent individuals with the **best health conditions**
- Patients with high risk **3rd class** are associated with the **highest** $\hat{\mu}_{j|u}$ **values** \Rightarrow represent individuals with the **worst health conditions**

Results: Item mean score - 1st vs 3rd class

- Main items characterizing the 1st and 3rd latent classes:

1st risk profile

FACIT: energy	0.34	EQ-5D-5L: tiredness	0.20
FACIT: frustrated	0.28	FACIT: finishing	0.10
HADS: fright	0.26	HADS: worry:	0.16
SLEEP: problem	0.22	HUI: pain:	0.15

3rd risk profile

FACIT: social	0.76	SLEEP: qual	0.59
HADS: fright	0.64	MG-QOL: publicplaces	0.58
MG-QOL: overwhelmed	0.64	FACIT: frustrated	0.57
MG-QOL: walking	0.63	FACIT: fatigue	0.56

Results: Third step - Initial probabilities

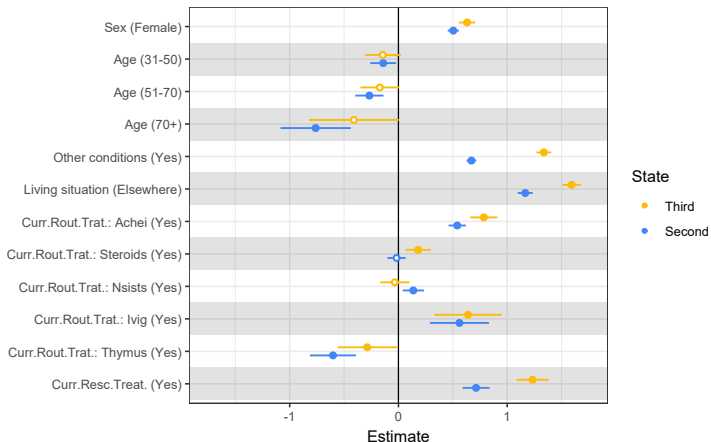
- **Estimated initial probabilities** $\hat{\pi}_u$ (average over all covariates configurations)

1st state	2nd state	3rd state
0.266	0.460	0.274

- Effect of the **covariates** on the initial probabilities, including:
 - **sex** (Male, Female)
 - **age** (4 categories: 18-30, 31-50, 51-70, 70+)
 - **other medical conditions** (No, Yes)
 - **living situation** (Own home, Elsewhere)
 - current/past routine/rescue **medical treatments**

Results: Third step - Initial probabilities

- Confidence intervals (confidence level: 95%) for some of the used covariates



Outline

- 1 Three-step estimation approach
- 2 The data: Myasthenia Gravis
- 3 Results
- 4 References

References I

- BARTOLUCCI, F., BACCI, S., AND GNALDI, M. (2016). *Statistical Analysis of Questionnaire*. Chapman & Hall/CRC, Boca Raton, FL.
- BARTOLUCCI, F., MOINTANARI, G. E., AND PANDOLFI, S. (2015). Three-step estimation of latent markov models with covariates. *Computational Statistics and Data Analysis*, **83**, 287–301.
- DEWILDE, S., PHILLIPS, G., PACI, S., DE RUYCK, F., TOLLENAAR, N., AND JANSSEN, M. (2024). The burden patients with myasthenia gravis experience in terms of breathing, fatigue, sleep, mental health, discomfort and usual activities in comparison to the general population. *Advances in Therapy*, **41**, 271–291.
- DRESSER, L., WLODARSKI, R., REZANIA, K., AND SOLIVEN, B. (2021). Myasthenia gravis, epidemiology, pathophysiology and clinical manifestations. *Journal of clinical medicine*, **10**, 2235.

References II

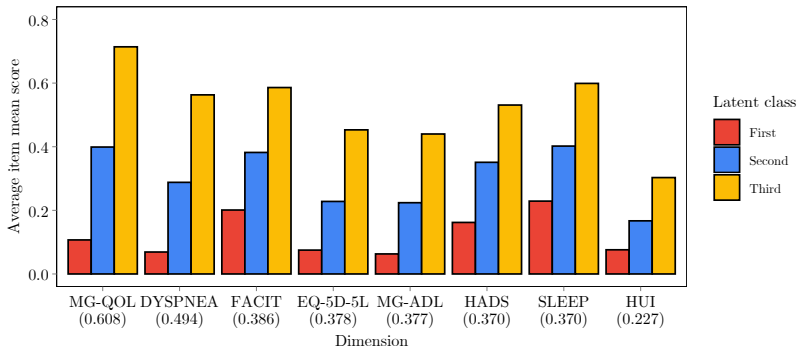
- NAKAI, M. AND PENNONI, F. (2020). Identifying groups with different traits using fourteen domains of social consciousness: A multidimensional latent class graded item response theory model. In *Advanced Researches in Behaviormetrics and Data Science (Springer)*, Essays in Honor of Akinori Okada, pp. 233-252.
- SALARI, N., FATAHI, B., BARTINA, Y., KAZEMINIA, M., FATAHIAN, R., MOHAMMADI, P., SHOHAIMI, S., AND MOHAMMADI, M. (2021). Global prevalence of myasthenia gravis and the effectiveness of common drugs in its treatment: a systematic review and meta-analysis. *IEEE Inform. Theory Soc. Newsl.*, **19**, 10–13.
- SAMEJIMA, F. (1969). *Estimation of ability using a response pattern of graded scores*. Psychometric Society, Richmond, VA.

Acknowledgments

Acknowledgment: L. Brusa, F. Bartolucci, and F. Pennoni acknowledge the financial support from the grant “Hidden Markov Models for Early Warning Systems” of Ministero dell’Università e della Ricerca (PRIN 2022TZEXKF) funded by European Union - Next Generation EU, Mission 4, Component 2, CUP J53D23004990006.

Results: Item mean score - Macro-areas averages

- Average** value of the **item mean scores** $\hat{\mu}_{j|u}$ **for each macro-area**; values in parentheses are the differences between the averages in the third and in the first latent classe



Results: Third step - Initial probabilities

- **Women** have **odds of belonging to latent state 3 rather than latent state 1** that are almost double (1.879) those of men
- **Patients who do not live in their own home** have **odds** of belonging to latent state 3 rather than latent state 1 that are nearly 5 times higher (4.911) than those of patients living in their own home
- **Most current treatments have a positive effect**; for example, for the treatment known as ACHEI (Acetylcholinesterase Inhibitors):
 - The odds of belonging to latent state 2 rather than latent state 1 increase among those undergoing this treatment (with a multiplicative increase of 1.717)
 - The odds of belonging to latent state 3 rather than latent state 1 increase among those undergoing this treatment (with a multiplicative increase of 2.194)

Results: Third step - Transition probabilities

- **Estimated transition probabilities** $\hat{\pi}_{v|u}$ (average over all covariates configurations)

	$v = 1$	$v = 2$	$v = 3$
$u = 1$	0.794	0.186	0.020
$u = 2$	0.203	0.609	0.188
$u = 3$	0.025	0.280	0.695

- Only current routine treatments are included as covariates in the corresponding multinomial logit model
- As an example, the estimated **parameters for steroid-based treatment are always negative**: patients receiving this treatment have higher odds of remaining with the same health conditions (same state) over time