Modelling longitudinal latent toxicity profiles evolution in osteosarcoma patient

Spreafico, M.; Ieva, F.; Fiocco, M.; Perna, C.; Salvati, N.; Schirripa Spagnolo, F.

Modelling longitudinal latent toxicity profiles evolution in osteosarcoma patients

Modellazione dell'evoluzione di profili longitudinali latenti di tossicità in pazienti con osteosarcoma

Marta Spreafico, Francesca Ieva and Marta Fiocco

Abstract In cancer trials, the analysis of longitudinal chemotherapy data is a difficult task due to the complex registration and evolution of toxicity levels during treatment. Models to deal with both the longitudinal and the categorical aspects of toxicity level progression are necessary, still not well developed. In this work, a Latent Transition Analysis (LTA) procedure to identify and reconstruct the longitudinal latent profiles of toxicity evolution of each patient over time is proposed. The latent variables determining the progression of the observed toxicity levels can be thought of as the outcomes of an underlying latent process. This methodology has never been applied to osteosarcoma treatment and provides new insights for supporting decisions in childhood cancer therapy.

Abstract *Negli studi sul cancro, analizzare i dati longitudinali di chemioterapia è problematico a causa della complessa evoluzione dei livelli di tossicità durante il trattamento. Modelli in grado di tenere conto sia degli aspetti longitudinali che di quelli categoriali dell'evoluzione dei livelli delle tossicità sono necessari, ma non ancora ben sviluppati. In questo lavoro viene proposta una procedura di analisi a transizioni latenti per identificare e ricostruire i profili latenti longitudinali relativi all'evoluzione delle tossicità di ogni paziente nel tempo. Le variabili latenti che determina l'evoluzione dei livelli di tossicità osservati possono essere pensate come i risultati di un processo latente sottosante. Questo approccio non è mai stato applicato al trattamento dell'osteosarcoma e fornisce nuove intuizioni per lo studio del cancro infantile.*

Key words: latent markov models, latent transition analysis, longitudinal data, toxicity, osteosarcoma

Marta Spreafico[1,2,3] Francesca Ieva[1,3,4] Marta Fiocco[2,5,6]
[1]MOX – Department of Mathematics, Politecnico di Milano, Milan 20133, Italy
[2] Mathematical Institute, Leiden University, Leiden, The Netherlands
[3]CHRP, National Center for Healthcare Research and Pharmacoepidemiology, Milan 20126, Italy
[4]CADS, Center for Analysis Decisions and Society, Human Technopole, Milan 20157, Italy
[5] Dept. of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands
[6] Trial and Data Center, Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands
e-mail: marta.spreafico@polimi.it francesca.ieva@polimi.it m.fiocco@math.leidenuniv.nl

1 Introduction

In many clinical applications involving longitudinal data, the interest lies in the analysis of the evolution of similar latent profiles related to subgroups of individuals rather than in the study of their observed attributes [1, 2]. These latent characteristics may reflect patients' quality-of-life, including valuable information related to patient's health status and disease progression.

In cancer trials, the analysis of longitudinal chemotherapy data is a complex task due to the presence of negative feedbacks between exposure to cytotoxic drugs and the toxicities the latter provoke. Toxic adverse events are at the same time risk factors for mortality and predictors of future exposure levels, representing time-dependent confounders for the effect of chemotherapy on patient's status [3]. Toxicity data are usually considered in very simplistic ways in cancer studies, where they act as fixed covariate over treatment [4, 5], discarding substantial amount of information (e.g., isolated vs repeated events, single vs multiple episodes, toxic events timing). Methods for longitudinal adverse events have also been proposed [4, 5] but they improperly treated toxicity levels as numerical values, as a simplifying hypothesis due to the complexity of the problem. Indeed, since multiple types of adverse events with different extents of toxicity burden occur simultaneously, studying the toxicity evolution during treatment is a challenging problem in cancer research.

In this work, a novel procedure based on *Latent Transition Analysis* (LTA) [1], a special case of first-order *Latent Markov* (LM) *models for longitudinal data* [2], is proposed to identify and reconstruct the longitudinal latent profiles of toxicity evolution. LM models for longitudinal data have been successfully applied in several fields, such as social, economic and behavioural sciences, education and public health, criminology or marketing [1, 2]. Clinical examples include, among others, the evolution of psycho-physic conditions in elderly individuals, the course of emotions among anorectic patients, or the analysis of pneumococcal carriage in children to study interactions between co-colonizing serotypes [2]. The idea behind this approach is that the latent variables can be thought of as the outcomes of a latent process which determines the evolution of the observed toxicity levels. This approach properly allows to take into account both the longitudinal and categorical aspects of toxicity level progression and the corresponding toxic risk evolution in oncology.

This approach has never been applied to osteosarcoma treatment and provides new insights for childhood cancer therapy. The presented procedure is really flexible and appropriate to analyse cancer chemotherapy treatment in general. Data from the MRC BO06/EORTC 80931 randomized controlled trial for osteosarcoma [6], a malignant bone tumour mainly affecting children and young adults, are analysed.

2 Methods

Let us consider a set $\mathscr{J}$ of categorical response variables measured at $t = 1, \ldots, T$ occasions, with $J = |\mathscr{J}|$. For each $j = 1, \ldots, J$, let $Y_j^{(t)}$ denote the j-th response variable at time t, with set of possible categories $\mathscr{C}_j$. Let $\widetilde{\boldsymbol{Y}} = (\boldsymbol{Y}^{(1)}, \ldots, \boldsymbol{Y}^{(T)})$ be the

complete response vector, where $\boldsymbol{Y}^{(t)}$ is the observed multivariate response vector at time t. The general Latent Transition Analysis (LTA) formulation is a Latent Markov (LM) model that assumes the existence of a latent process which affects the distribution of the response variables. This latent process, denoted by $\boldsymbol{U} = (U^{(1)}, \ldots, U^{(T)})$, follows a first-order Markov chain with state space $\{1, \ldots, k\}$, where the number of latent states k can be a priori defined or selected according to the Bayesian information criterion (BIC). Three different sets of model parameters $\boldsymbol{\theta}$ can be defined:

- the **item-response probability** $\phi^{(t)}_{jy|u}$, i.e., the probability of a particular observed response y on variable j at time t, conditional on latent class u membership:

$$\phi^{(t)}_{jy|u} = P(Y^{(t)}_j = y \big| U^{(t)} = u) \qquad y \in \mathscr{C}_j \quad j = 1, \ldots, J \quad u \in \{1, \ldots, k\};$$

- the **initial latent status prevalence** δ_u, i.e., the probability of membership in latent state u at time $t = 1$:

$$\delta_u = P(U^{(1)} = u) \qquad u \in \{1, \ldots, k\};$$

- the **transition probability** $\tau^{(t)}_{u|\bar{u}}$, i.e., the probability of a transition to latent state u at time t, conditional on membership in latent state $\bar{u}$ at time $t-1$:

$$\tau^{(t)}_{u|\bar{u}} = P(U^{(t)} = u \big| U^{(t-1)} = \bar{u}) \qquad t = 2, \ldots, T \quad u, \bar{u} \in \{1, \ldots, k\}.$$

Assuming *local independence*, i.e., the observed variables are independent conditional on the latent class, the *manifest distribution* of the response variables is:

$$P(\widetilde{\boldsymbol{Y}} = \widetilde{\boldsymbol{y}}) = \sum_{\boldsymbol{u}} P(\boldsymbol{U} = \boldsymbol{u}) \times P(\widetilde{\boldsymbol{Y}} = \widetilde{\boldsymbol{y}} \big| \boldsymbol{U} = \boldsymbol{u}) = \sum_{\boldsymbol{u}} \delta_{u^{(1)}} \prod_{t=2}^{T} \tau^{(t)}_{u^{(t)}|u^{(t-1)}} \times \prod_{t=1}^{T} \prod_{j=1}^{J} \phi^{(t)}_{jy^{(t)}_j|u^{(t)}}$$

where $\widetilde{\boldsymbol{y}}$ is a realization of $\widetilde{\boldsymbol{Y}}$ made by the subvectors $(\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(T)})$, $\boldsymbol{y}^{(t)}$ is a realization of $\boldsymbol{Y}^{(t)}$ with elements $y^{(t)}_j$ and $\boldsymbol{u} = (u^{(1)}, \ldots, u^{(T)})$.
Parameters estimation $\hat{\boldsymbol{\theta}}$ is performed maximizing the log-likelihood for a sample of n independent units, i.e., $\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log P(\widetilde{\boldsymbol{Y}}_i = \widetilde{\boldsymbol{y}}_i)$, using an Expectation-Maximization (EM) algorithm. For further details see [2, 7].

The EM algorithm also provides the estimated *posterior* probabilities of $U^{(t)}$ [2], which can be used to reconstruct the ***Longitudinal Latent Probability Profile*** (LLPP) for each latent state $u \in \{1, \ldots, k\}$ and subject $i \in \{1, ..., n\}$, as follows:

$$\boldsymbol{p}_{u,i} = \left\{ p^{(t)}_{u,i} = P\left(U^{(t)} = u \big| \widetilde{\boldsymbol{Y}}_i = \widetilde{\boldsymbol{y}}_i\right), \quad t = 1, \ldots, T \right\} \tag{1}$$

LLPP in Eq. (1) represents the probability over time t of being in latent state u for individual i, given the observed response $\widetilde{\boldsymbol{y}}_i$. Applying this procedure, a k-variate longitudinal latent profile $\left(\boldsymbol{p}_{1,i}, \ldots, \boldsymbol{p}_{k,i}\right)$ such that $\sum_u p^{(t)}_{u,i} = 1$ for each $t = 1, \ldots, T$ can be reconstructed for each subject i.

3 Data and Patients

Data from MRC BO06/EORTC 80931 randomised controlled trial for patients with non-metastatic high-grade osteosarcoma [6] were analysed. Patients were randomized at baseline between Conventional (*Reg-C*) or Dose-Intense (*Reg-DI*) regimens of six cycles of chemotherapy, with identical anticipated cumulative dose but different duration. Non-haematological chemotherapy-induced toxicity for nausea/vomiting (*naus*), infection (inf), mucositis (*oral*), cardiac toxicity (*car*), ototoxicity (oto) and neurological toxicity (*neur*) were registered at each cycle and graded according to the Common Terminology Criteria for Adverse Events (CTCAE) v3.0 [8], with grades ranging from 0 (none) to 4 (life-threatening). Additional details can be found in the primary analysis of the trial [6].

In the study cohort $n = 377$ patients that completed the chemotherapy within 180 days from randomization were included. Nausea/vomiting was reported at least once over cycles in 97.3% of patients (367/377), with a percentage that decreased over cycles from 84.9% (327/377) in cycle 1 to 52.5% (198/377) in cycle 6. The percentages of patients that reported oral mucositis or infections were more stable over cycles: 30.5%-43.3% for mucositis and 23.8%-31.3% for infection. Other toxicities were less frequent ($<10\%$), especially for grades above 1.

4 Application and Results

Latent Markov (LM) model formulation presented in Sect. 2 is now applied to chemotherapy-induced longitudinal categorical toxicity data presented in Sect. 3, as shown in Fig. 1. For each cycle $t = 1,\ldots,6$, let $\mathscr{J} = \{naus, inf, oral, car, oto, neur\}$ be the set of categorical response variables $Y_j^{(t)}$ with possible sets of response categories $\mathscr{C}_j = \{0 : none, 1 : mild, 2 : moderate, 3/4 : severe\}$ for $j = 1,2,3$ and $\mathscr{C}_j = \{0 : none, 1 : mild, 2/3/4 : mod/sev\}$ for $j = 4,5,6$. The item-response probabilities were assumed to be time homogeneous, i.e. $\phi_{jy|u}^{(t)} = \phi_{jy|u} \forall t$, that is a common parameters restriction in latent transition analysis [1]. Due to the multimodality of the log-likelihood function $\ell(\boldsymbol{\theta})$, different random initializations of EM algorithm were used and the final estimate $\hat{\boldsymbol{\theta}}$ was the one corresponding to the highest $\ell(\boldsymbol{\theta})$. Statistical analyses were performed using the `R` package `LMest` [7].

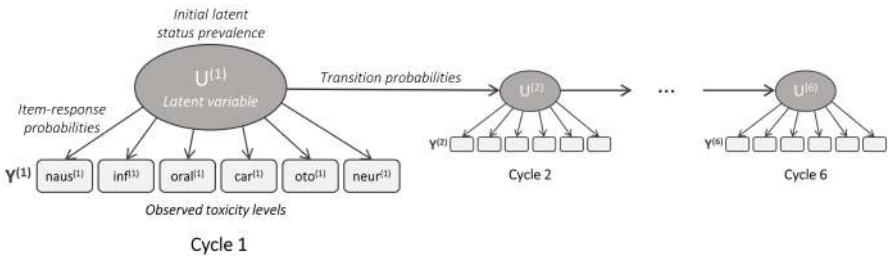

Fig. 1 Structure of the Latent Markov (LM) model for longitudinal data of toxicity levels.

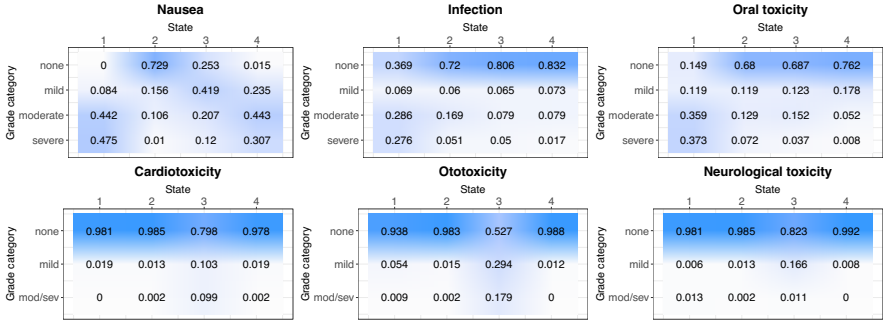

Fig. 2 Estimated item-response probabilities. Each panel refers to a different toxicity variable in $\mathscr{J} = \{naus, inf, oral, car, oto, neur\}$ with grade response categories $\{0 : none, 1 : mild, 2 : moderate, 3/4 : severe\}$ for $\{naus, inf, oral\}$ and $\{0 : none, 1 : mild, 2/3/4 : mod/sev\}$ for $\{car, oto, neur\}$.

According to minimum BIC, the selected number of latent states was $k = 4$. Fig. 2 shows the estimated item-response probabilities $\hat{\phi}_{jy|u}$ for each toxicity, which provided the basis for the interpretation of the latent states. From these results the following latent state labelling was derived:

- State 1 *Severe/Moderate a-specific toxic state*
- State 2 *Non-toxic state*
- State 3 *Mild nausea (with possible specific toxicity)*
- State 4 *Severe/Moderate nausea only*.

The estimated initial latent status prevalence showed that the prevailing state at time $t = 1$ was *Severe/Moderate nausea only* (61.8%), followed by *Non-toxic state* (19.5%), *Severe/Moderate a-specific toxic state* (12.1%) and *Mild nausea* (6.5%). In particular, the prevalence of *Severe/Moderate nausea only* decreased over cycles from 61.8% to 20.6% ($t = 6$), whereas the ones of *Non-toxic state* and *Mild nausea* increased from 19.5% to 49.3% and from 6.5% to 19.2%, respectively. Latent state prevalence of *Severe/Moderate a-specific toxic state* was more stable over cycles ranging in 10.8%-16.5%, with peaks in cycles 2 and 3.

Longitudinal latent probability profiles (LLPPs) were finally reconstructed according to Eq. (1). Fig. 3 shows the LLPPs $\boldsymbol{p}_{u,i}$ related to each latent state $u = \{1,2,3,4\}$ for seven random patients. LLPPs are able to capture the individual realisations of the latent process over cycles through a customized reconstruction, showing different patterns of toxicity evolution over treatment among patients.

5 Conclusion

The proposed approach allowed (i) to move from complex chemotherapy data to a set of different subgroups of individuals that exhibit similar patterns of toxicity grades progression over cycles by identifying the latent states and (ii) to reconstruct

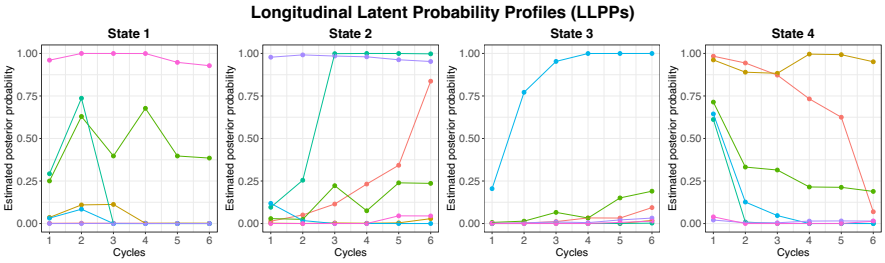

Fig. 3 Reconstructed longitudinal latent probability profiles $\boldsymbol{p}_{u,i}$ for seven random patients. Each panel refers to a different state $u = \{1,2,3,4\}$. Different colours refer to different patients.

and provide Longitudinal Latent Probability Profiles (LLPPs) related to toxicity evolution in a tailored way. This procedure represents a novelty for osteosarcoma treatment and, more generally, for cancer studies, providing new insights for childhood therapy.

This work opens doors for many further developments. The LM model could be enriched by considering other relevant clinical information as adjusting covariates. Moreover, it could be of clinical interest to study the association between LLPPs and time-to-event outcomes. This is a non-trivial modelling task, representing a challenging problem both for clinical and statistical research.

Acknowledgements The authors thank Medical Research Council for sharing the dataset used in this work.

References

1. Collins, L.M., Lanza, S.T.: Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences. John Wiley and Sons Inc (2010).
2. Bartolucci, F., Farcomeni, A., Pennoni, F.: Latent Markov Models for Longitudinal Data. Chapman & Hall/CRC, Boca Raton (2013).
3. Lancia, C. et al.: Marginal structural models with dose-delay joint-exposure for assessing variations to chemotherapy intensity. Stat. Methods Med. Res. **28**(9), 2787–2801 (2019).
4. Trotti, A. et. al.: TAME: development of a new method for summarising adverse events of cancer treatment by the Radiation Therapy Oncology Group. Lancet Oncol. **8**, 613–24 (2007).
5. Thanarajasingam, G. et al.: Longitudinal adverse event assessment in oncology clinical trials: the Toxicity over Time (ToxT) analysis of Alliance trials NCCTG N9741 and 979254. Lancet Oncol. **17**(5), 663–670 (2016).
6. Lewis, I.J. et al.: Improvement in Histologic Response But Not Survival in Osteosarcoma Patients Treated With Intensified Chemotherapy: A Randomized Phase III Trial of the European Osteosarcoma Intergroup. J. Natl. Cancer Inst. **99**(2), 112–128 (2007).
7. Bartolucci, F., Pandolfi, S., Pennoni, F.: LMest: An R Package for Latent Markov Models for Longitudinal Categorical Data. J. Stat. Soft. **81**(4), 1–38 (2017).
8. U.S. Department of Health and Human Services: Common Terminology Criteria for Adverse Events v3.0 (CTCAE). (2006) URL: https://www.eortc.be/services/doc/ctc/ctcaev3.pdf