MOX-Report No. 76/2022

Longitudinal Latent Overall Toxicity (LOTox) profiles in osteosarcoma: a new taxonomy based on latent Markov models

Spreafico, M.; Ieva, F.; Fiocco, M.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it https://mox.polimi.it

Longitudinal Latent Overall Toxicity (LOTox) profiles in osteosarcoma: a new taxonomy based on latent Markov models

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]National Center for Healthcare Research and Pharmacoepidemiology, Milan 20126, Italy

[4]Health Data Science Center, Human Technopole, Milan 20157, Italy

[5]Department 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

[*]marta.spreafico@polimi.it

Abstract

Due to the presence of multiple types of adverse events with different levels of severity, the analysis of longitudinal toxicity data is a difficult task in cancer studies. In this work, a novel approach based on latent Markov models and compositional data techniques is proposed. The latent status of interest is the Latent Overall Toxicity (LOTox) condition of each patient. The main objectives are to identify different latent states of overall toxicity burden and to investigate the evolution of individual toxicity risk during cancer treatment. This methodology is applied to osteosarcoma treatment data to provide novel techniques that may support medical decisions in childhood cancer therapy.

Key-words: Latent Markov models; Longitudinal data; Compositional Data; Toxicity; Osteosarcoma

1 Introduction

Osteosarcoma is a malignant bone tumour mainly affecting children and young adults. Although osteosarcoma is the most common primary malignant bone cancer, it is a rare disease and has an annual incidence of 3-4 patients per million (Smeland et al., 2019). Multidisciplinary management including neoadjuvant and adjuvant chemotherapy with aggressive surgical resection (Ritter and Bielack, 2010) or intensified chemotherapy has improved clinical outcomes but over the past 40 years there have been no further improvements in survival (Anninga et al., 2011).

In cancer trials, the relationship between chemotherapy dose and clinical efficacy outcomes are problematic to analyse due to the presence of negative feedback between exposure to cytotoxic drugs and other aspects, such as latent accumulation of chemotherapy-induced toxicity. Toxicities, developed by patients through chemotherapy, are time-dependent confounders for the effect of chemotherapy on patient's status (Lancia et al., 2019). Toxicities affect subsequent exposure by delaying the next cycle or reducing chemotherapy doses (Souhami et al., 1997), being at the same time risk factors for mortality and predictors for future exposure levels. According to the Common Terminology Criteria for Adverse Events (CTCAE) (U.S. Department of Health and Human Services, 2006), a multimodality grading system for the standardized classification of adverse events (AEs) in cancer therapy, nominal grades of AEs severity range from minor to life-threatening injuries or death (Trotti et al., 2003). Since patients may have multiple AEs with different levels of severity, identifying the actual extent of toxic burden and investigating the evolution of patient's overall toxicity status during treatment represent challenging problems in cancer research.

Due to the complexity of longitudinal chemotherapy data, no standard method is available for analysing AEs data. Toxicity data are mainly used in cancer studies as summary indexes, such as maximum toxicity over time, maximum grade among events, or weighted sums of individual toxic effects (Bekele and Thall, 2004; Rogatko et al., 2004; Trotti et al., 2007; Lee et al., 2012; McTiernan et al., 2012; Sivendran et al., 2014; Thanarajasingam et al., 2015, 2016; Zhang et al., 2016; Carbini et al., 2018). Although these methods can summarise data over time, substantial amount of information (e.g., isolated vs repeated events, single vs multiple episodes, longer-lasting lower-grade toxicities, toxic events timing) are discarded. As neglecting the time component may give an inaccurate depiction of toxicity, alternative methods for a longitudinal analysis of AEs have been proposed (Trotti et al., 2007; Thanarajasingam et al., 2016, 2020; Hirakawa et al., 2019; Spreafico et al., 2021). These approaches are not suitable for the nominal CTCAE grades still they provide more insights into treatment-related toxicity, suggesting that longitudinal methods should become routine in future analyses of cancer trials. Models to deal with both longitudinal and categorical aspects of toxicity levels progression are then necessary, still not well developed.

Longitudinal data are often of interest in a wide range of research fields, such as social, economic and behavioural sciences, education or public health. In many applications involving longitudinal data, the interest lies in analysing the evolution of a latent characteristic of a group of individuals over time, rather than in studying their observed attributes (Bartolucci et al., 2014). The phenomenon which affects the distribution of the response variables that are relevant for the problem under consideration may not be directly observable. In a clinical context, this latent characteristic may reflect patients' quality-of-life and could contain valuable information related to patient's health status and disease progression.

In the statistical literature many models have been proposed for the analysis of longitudinal data; for a concise review see Fitzmaurice et al. (2009). For longitudinal categorical data, where the interest is in describing individual changes with respect to a latent status, Latent Markov (LM) models can be used (Wiggins, 1973; Bartolucci et al., 2013). These models study the evolution of an individual characteristic of interest, when it is not directly observable. The idea behind a LM model is that the latent process fully explains the observable behaviour of a subject, assuming that the response variables are conditionally independent given the latent process. The latent process follows a Markov chain with a finite number of states, which represent different conditions of the latent characteristic of interest. LM models can also account for the effect of observable covariates, serial dependence between observations, measurement errors, or unobservable heterogeneity. For

a detailed overview on LM models see Bartolucci et al. (2013, 2014).

Motivated by the need to improve methods for summarising and quantifying the overall toxicity level and its evolution during treatment, in this work a novel procedure based on LM models for longitudinal toxicity data is proposed. The latent status of interest is the Latent Overall Toxicity (LOTox) condition of a patient, which affects the distribution of the observed categorical toxic grades measured over treatment. The proposed approach aims at identifying different latent states of overall toxicity burden (*LOTox states*) and investigating how patients move between states during chemotherapy treatment.

A LM model for longitudinal toxicity data assumes that at each time occasion for each patient a vector of probabilities of being in the various LOTox states is given. Since the probability elements of each vector are non-negative coordinates whose sum is one, these vectors are naturally confined to a suitably dimensioned simplex, thus being *Compositional Data* (CoDa) or *compositions*. In statistics, CoDa are quantitative descriptions of the parts of some whole, carrying relative information. In this context, Aitchison (1986) developed a methodology based on log-ratio transformations of CoDa, which nowadays represent the mainstream approach in the analysis of compositions formed by probabilities or percentages. Among the developed transformations, the *additive log-ratios* consider a specific reference part in contrast with all the others. In this work, this approach is exploited to compare over time a reference "good" overall toxicity condition (i.e., the LOTox state characterized by the lowest toxicity burden) in contrast with all the other LOTox states, characterized by worsening overall toxicity. In this way, the dynamic risk of experiencing "worse" overall toxicity statuses relative to a "good" toxic condition over time is investigated.

Three are the main novelties presented in this work: (i) the introduction of a new method based on LM models to summarize and quantify multiple AEs and their evolution during treatment, where both longitudinal and categorical aspects of the observed toxic levels are included in the model; (ii) the identification of groups of patients with a common distribution for the observed toxic categories, and thus a similar overall toxicity burden; (iii) the reconstruction of personalized *longitudinal LOTox profiles*, which represent the probability over time of being in a specific LOTox state or the relative risk with respect to a reference "good" toxic condition; this allow to study the individual overall toxic risk evolution during treatment for each subject. The proposed approach is applied to osteosarcoma treatment to provide novel techniques which could support clinicians in planning new protocols and guidelines for childhood cancer therapy. Provided that longitudinal CTCAE-graded toxicity data are available, the developed procedure is a flexible approach that can be adapted and applied to other cancer studies.

The article is organized as follows. Data from the MRC BO06/EORTC 80931 Randomized Controlled Trial for patients with osteosarcoma (Lewis et al., 2007) are described in Section 2. Statistical methods are introduced in Section 3. Results for MRC BO06 data are presented in Section 4. Section 5 ends with a discussion of strengths and limitations of the proposed approach, identifying some possible developments for future research.

2 MRC BO06 randomized clinical trial data

In childhood cancer research, the development of new evidence-based guidelines to support clinical decisions in tailored interventions for an effective management of adverse symptoms and treatments is still a key issue. Analysing the evolution of toxicities in patients who have completed the treatment could lead to new insights into the progression and tolerance of toxic AEs during therapy.

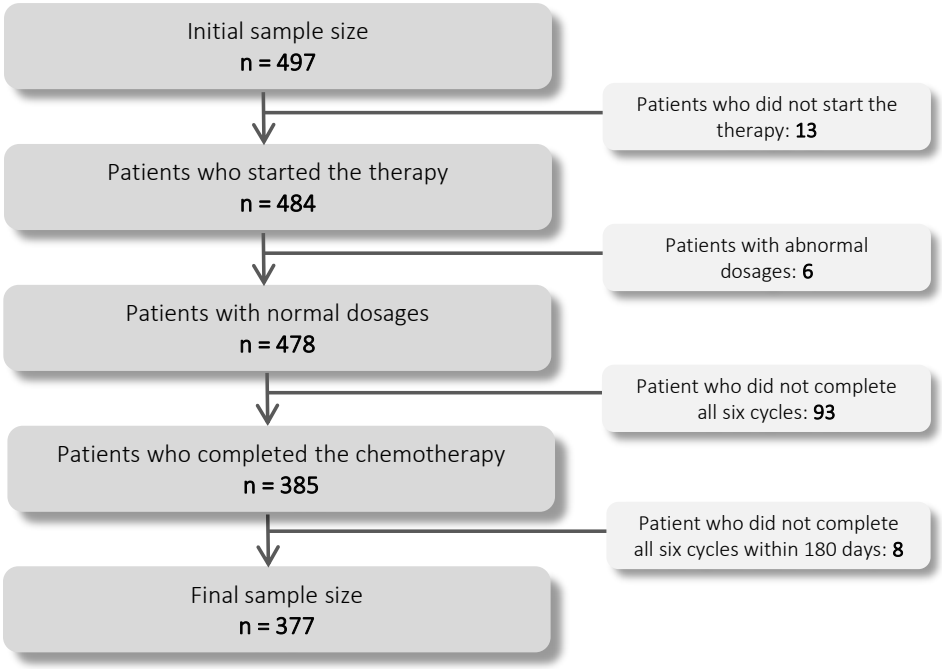

Figure 1: Flowchart of cohort selection.

In Section 2.1 the selected cohort of patients is illustrated. Longitudinal chemotherapy data and patient characteristics are presented in Section 2.2.

2.1 Data illustration

Data from the MRC BO06/EORTC 80931 randomized clinical trial for patients with non-metastatic high-grade osteosarcoma recruited between 1993 and 2002 were analysed (Lewis et al., 2007). Patients were randomized between conventional (*Reg-C*) and dose-intense (*Reg-DI*) regimens. Both arms had six cycles of the same course of doxorubicin and cisplatin with different time schedule (3-weekly vs 2-weekly, supported by granulocyte colony stimulating factor). Details concerning the trial protocol are provided in Appendix A.

The dataset included 497 eligible patients; 19 patients who did not start chemotherapy (13) or reported an abnormal dosage (i.e., +25% higher than planned) for a single or both agents (6) were excluded. Patients who did not complete all six cycles of chemotherapy (93) and did not terminate the last cycle within 180 days after randomization (8) were excluded. The final cohort of 377 patients included in the analyses (75.9% of the initial sample) is shown in the consort diagram in Figure 1.

2.2 Longitudinal chemotherapy data

During the trial treatment, case report forms were used to document across cycles all the information required by the MRC BO06/EORTC 80931 trial protocol for each patient.

Patients baseline characteristics (age, gender, allocated chemotherapy regimen, site and location of the tumour) were registered at randomization. Among 377 patients, 229 (60.7%) were males and *Reg-DI* was allocated in 52.3% of the patients (197). Median age was 15 years (IQR [11;

18]). Treatment-related factors (administered dose of chemotherapy, cycles delays, chemotherapy-induced toxicity, haematological parameters) were collected at each cycle of chemotherapy. For each patient i and cycle t, chemotherapy dose was analysed as percentage of achieved chemotherapy dose up to cycle t, i.e., the percentage of the cumulative drugs administrated up to cycle t divided by the cumulative drugs planned up to t. Non-haematological chemotherapy-induced toxicity for nausea/vomiting (*naus*), infection (*inf*), oral mucositis (*oral*), cardiac toxicity (*car*), ototoxicity (*oto*) and neurological toxicity (*neur*) were graded according to the Common Terminology Criteria for Adverse Events Version 3 (CTCAE v3.0) (U.S. Department of Health and Human Services, 2006), with grades ranging from 0 (none) to 4 (life-threatening) (see Appendix A for further details). Nausea/vomiting, infection and oral mucositis were classified as *generic* toxicities since they represent common adverse events for chemotherapeutic treatments in general. Cardiac toxicity, ototoxicity and neurological toxicity, which could also cause irreversible conditions (see Appendix Table A.1), were classified as *drug-specific* toxicities since they are related to the use of cisplatin or doxorubicin (Al-malky et al., 2020; dos Santos et al., 2020).

Considering the CTCAE-grades registered over cycles for each non-haematological adverse event, *generic* toxicities were more frequent than *drug-specific* ones, as expected. 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% in cycle 1 to 52.5% in cycle 6. The percentages of patients who reported oral mucositis or infections were more stable over cycles: 30.5%–43.3% for mucositis, with 78% (294/377) reporting mucositis at least once, and 23.8%–31.3% for infection, with 69% (260/377) reporting an infection at least once. Ototoxicity was reported at least once in 21.5% (81/377), cardiac toxicity in 14.1% (53/377) and neurological toxicity in 11.7% (44/377). At each cycle, CTCAE-grade 4 for *generic* toxicities and CTCAE-grades ≥ 2 for *drug-specific* toxicities were reported in less than 5% of patients. Low-frequency classes were merged and toxic categories were represented according to the degree of severity or as present or not, depending on the type of toxicity as follows:

- the severity of the toxic event for *generic* toxicities: *none* (CTCAE-grade 0), *mild* (CTCAE-grade 1), *moderate* (CTCAE-grade 2), and *severe* (CTCAE-grades 3 or 4);
- the absence or the presence of toxic event for *drug-specific* toxicities: *no* (CTACE-grade 0) and *yes* (CTACE-grades ≥ 1).

These categories identified for each toxicity constitute the item-response elements selected to model the latent process representing the "true" overall toxic status. Table 1 shows the observed frequencies (and percentages) of the selected categories for each toxicity over cycles for the final cohort. The observed responses for each patient are then given by the longitudinal toxic categories measured along the cycles, which are then used to evaluate the LOTox condition during treatment.

3 Statistical Methods

In the following Sections, the novel Latent Markov (LM) approach for modelling the Latent Overall Toxicity (LOTox) condition of each patient starting from the observed longitudinal toxic categories measured during chemotherapy treatment is introduced. In Section 3.1 motivations for the proposed approach for treating the longitudinal toxicity data are discussed. Mathematical details are provided in Section 3.2. Model selection procedure and longitudinal profiles are presented in Sections 3.3 and 3.4, respectively.

Table 1: Frequencies of toxic categories over the six cycles. For nausea, infection and mucositis ($j = 1, 2, 3$), the set of toxic categories indicating the severity of the toxic event is defined as $\mathcal{C}_j = \{none; mild; moderate; severe\}$. For cardiotoxicity, otoxocity and neurological toxicity ($j = 4, 5, 6$), the set of toxic categories indicating the presence or the absence of the toxic event is defined as $\mathcal{C}_j = \{no; yes\}$.

Toxicity	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Cycle 6
Nausea						
none	57 (15.1%)	88 (23.3%)	115 (30.5%)	126 (33.4%)	146 (38.7%)	179 (47.5%)
mild	74 (19.6%)	87 (23.1%)	76 (20.2%)	72 (19.1%)	86 (22.8%)	74 (19.6%)
moderate	117 (31.1%)	117 (31.1%)	114 (30.2%)	113 (30.0%)	96 (25.5%)	87 (23.1%)
severe	129 (34.2%)	85 (22.5%)	72 (19.1%)	66 (17.5%)	49 (13.0%)	37 (9.8%)
Infection						
none	259 (68.7%)	287 (76.1%)	268 (71.1%)	265 (70.3%)	268 (71.1%)	286 (75.9%)
mild	30 (7.9%)	24 (6.4%)	26 (6.9%)	31 (8.2%)	23 (6.1%)	16 (4.3%)
moderate	64 (17.0%)	45 (11.9%)	61 (16.2%)	54 (14.3%)	52 (13.8%)	45 (11.9%)
severe	24 (6.4%)	21 (5.6%)	22 (5.8%)	27 (7.2%)	34 (9.0%)	30 (8.0%)
Mucositis						
none	265 (70.3%)	228 (60.5%)	234 (62.1%)	237 (62.9%)	214 (56.8%)	262 (69.5%)
mild	54 (14.3%)	46 (12.2%)	59 (15.6%)	52 (13.8%)	62 (16.4%)	44 (11.7%)
moderate	44 (11.7%)	54 (14.3%)	43 (11.4%)	55 (14.6%)	63 (16.7%)	50 (13.2%)
severe	14 (3.7%)	49 (13.0%)	41 (10.9%)	33 (8.7%)	38 (10.1%)	21 (5.6%)
Cardiotoxicity						
no	374 (99.2%)	361 (95.8%)	362 (96.0%)	359 (95.2%)	357 (94.7%)	355 (94.2%)
yes	3 (0.8%)	16 (4.2%)	15 (4.0%)	18 (4.8%)	20 (5.3%)	22 (5.8%)
Ototoxicity						
no	357 (94.7%)	361 (95.8%)	350 (92.8%)	342 (90.7%)	346 (91.8%)	326 (86.5%)
yes	20 (5.3%)	16 (4.2%)	27 (7.2%)	35 (9.3%)	31 (8.2%)	51 (13.5%)
Neurological toxicity						
no	371 (98.4%)	367 (97.3%)	362 (96.0%)	367 (97.3%)	356 (94.4%)	363 (96.3%)
yes	6 (1.6%)	10 (2.7%)	15 (4.0%)	10 (2.7%)	21 (5.6%)	14 (3.7%)

3.1 Motivations for latent Markov models for longitudinal toxicity data

LM models are statistical methods employed for the analysis of longitudinal (categorical) data specifically designed to study the evolution of an individual characteristic of interest, when it is not directly observable (Wiggins, 1973; Bartolucci et al., 2013). A LM approach for longitudinal toxicity data assumes the existence of a latent process representing the "true" LOTox status, which affects the distribution of the response variables, in our case the observed toxicities. Two main motivations justify the use of LM models to quantify the toxic risk in cancer studies: (i) account for *measurement errors* in the observed toxicity variables, and (ii) identify different *LOTox sub-populations* (i.e., the latent states) in the global population (i.e., the patients' cohort) and their changes over time.

Since therapy protocol is adapted at each cycle depending on patient's reaction to treatment, it is reasonable to assume that the latent variables follow a first-order Markov chain, so that the "true" level of overall toxicity at a given cycle is influenced only by the previous level. Non-hematological toxicities do not depend directly on each other as they relate to different systems and functions of the human body (i.e., nausea/vomiting is part of the stomach-gastrointestinal system, infections of the immune system, oral mucositis of the mouth-gastrointestinal system, cardiotoxicity of the cardiovascular system, ototoxicity of the auditory-sensory system and neurotoxicity of the nervous system). Therefore, the response toxicity variables can be assumed conditionally independent, as each observed response is expected to depend only on the corresponding "true" LOTox level.

In this context, a LM model may be seen as an extension of the latent class model (Collins

and Lanza, 2010), where patients are allowed to move between latent states during the observation period. LM models for longitudinal toxicity data are characterized by several parameters: the initial probability of each LOTox state, the transition probabilities among different states over chemotherapy cycles, and the conditional response probabilities given the latent variable. Individual covariates (if available) can be included in the latent model and may affect the initial and transition probabilities of the Markov chain (Bartolucci et al., 2009), as explained in Section 3.2.

A LM approach is appropriate to both identify the actual overall toxicity burden and investigate its evolution during treatment for each patient. On one hand, patients that at a specific time result in the same sub-population are characterized by a common distribution for the observed toxic categories, and by a similar overall toxicity burden. On the other hand, individual dynamic changes among latent states allow to evaluate the LOTox evolution during treatment for each subject.

3.2 Latent Markov model with covariates

Let $\mathcal{J}$ be the set of $J = |\mathcal{J}|$ categorical response variables measured at each time $t = 1, \dots, T$. Denote by $Y_{ij}^{(t)}$ the response variable $j \in \{1, \dots, J\}$ for subject $i \in \{1, ..., n\}$ at time t, with set of categories $\mathcal{C}_j$ coded from 0 to $c_j - 1$. Let $\boldsymbol{Y}_i^{(t)} = \left(Y_{i1}^{(t)}, ..., Y_{iJ}^{(t)}\right)$ denote the observed multivariate response vector at time t for patient i and $\widetilde{\boldsymbol{Y}}_i = \left(\boldsymbol{Y}_i^{(1)}, \dots, \boldsymbol{Y}_i^{(T)}\right)$ be the corresponding complete response vector. Denote by $\widetilde{\boldsymbol{X}}_i = \left(\boldsymbol{X}_i^{(1)}, \dots, \boldsymbol{X}_i^{(T)}\right)$ the complete vector of individual covariates, where elements $\boldsymbol{X}_i^{(t)} = \left(\boldsymbol{S}_i, \boldsymbol{Z}_i^{(t)}\right)$ are the vectors of time-fixed $\boldsymbol{S}_i$ and time-varying $\boldsymbol{Z}_i^{(t)}$ covariates for subject i at occasion t. The general LM model assumes the existence of a latent process $\boldsymbol{U}_i = \left(U_i^{(1)}, \dots, U_i^{(T)}\right)$ which affects the distribution of the response variables $\widetilde{\boldsymbol{Y}}_i$. The latent process follows a first-order Markov chain with state space $\{1, \dots, k\}$, where k is the total number of *latent states*. LM models usually assume that the response vectors $\boldsymbol{Y}_i^{(1)}, \dots, \boldsymbol{Y}_i^{(T)}$ are conditionally independent given the latent process $\boldsymbol{U}_i$ (*local independence of the response vectors*) and that the elements $Y_{ij}^{(t)}$ are conditionally independent given $U_i^{(t)}$ (*conditional independence of elements*). The motivation of these assumptions is that the latent process fully explains the observable behaviour of a subject, as explained in Section 3.1.

LM models are made by two components: the *measurement model* concerns the conditional distribution of the response variables given the latent process, and the *latent model* is related to the distribution of the latent process (i.e., initial and transition probabilities). The latent process represents an individual characteristic of interest that is not directly observable that may evolve over time, also depending on observable covariates. The main research interest hence lies in modelling the latent process and the effect of covariates on its dynamic. LM models where both the initial and the transition probabilities of the latent process may depend on covariates is considered. Three different sets of probabilities (i.e., parameters) can be defined.

- *Conditional response probability* (or item-response probability) $\phi_{jy|u}^{(t)}$ is the probability of observing a response y for variable j at time t, given the latent status $u \in \{1, ..., k\}$:

$$\mathrm{P}\left(Y_{ij}^{(t)} = y \middle| U_i^{(t)} = u\right) = \phi_{jy|u}^{(t)} \qquad j = 1, \dots, J \quad y = 0, ..., c_j - 1.$$

  To ensure that the interpretation of the latent states remains constant over time, conditional response probabilities are assumed time-homogeneous, i.e., $\phi_{jy|u}^{(t)} = \phi_{jy|u}$ $\forall t = 1, \dots, T$.

Given the estimated $\hat{\phi}_{jy|u}$, the latent states can be characterized in terms of observed response categories.

- *Initial latent states prevalence* $\delta_{u|\boldsymbol{x}_i^{(1)}}$ is the probability of membership in latent state $u \in \{1, \ldots, k\}$ at time $t = 1$, given the vector of covariates $\boldsymbol{x}_i^{(1)}$ for individual i:

$$\mathrm{P}\left(U_i^{(1)} = u | \boldsymbol{X}_i^{(1)} = \boldsymbol{x}_i^{(1)}\right) = \delta_{u|\boldsymbol{x}_i^{(1)}}.$$

The estimated $\hat{\delta}_{u|\boldsymbol{x}_i^{(1)}}$ may be interpreted as quantities proportional to the size of each latent state at the first time-occasion, given the covariates. A natural way to allow the initial probabilities of the LM chain to depend on individual covariates is a multinomial logit parametrization:

$$\log \frac{\mathrm{P}\left(U_i^{(1)} = u \mid \boldsymbol{X}_i^{(1)} = \boldsymbol{x}_i^{(1)}\right)}{\mathrm{P}\left(U^{(1)} = 1 \mid \boldsymbol{X}_i^{(1)} = \boldsymbol{x}_i^{(1)}\right)} = \log \frac{\delta_{u|\boldsymbol{x}_i^{(1)}}}{\delta_{1|\boldsymbol{x}_i^{(1)}}} = \beta_{0u} + \boldsymbol{x}_i^{(1)\top} \boldsymbol{\beta}_{1u} \tag{1}$$

where $u = 2, ..., k$ and $\boldsymbol{\beta}_u = \left(\beta_{0u}, \boldsymbol{\beta}_{1u}^{\top}\right)^{\top}$ are the parameters vectors to be estimated.

- *Transition probability* $\tau^{(t)}_{u|\bar{u}\boldsymbol{x}_i^{(t)}}$ is the probability of a transition to latent state u at time t, conditional on membership in latent state $\bar{u}$ at time $t-1$, given the individual vector of covariates $\boldsymbol{x}_i^{(t)}$ (if available):

$$\mathrm{P}\left(U_i^{(t)} = u \mid U_i^{(t-1)} = \bar{u}, \boldsymbol{X}_i^{(t)} = \boldsymbol{x}_i^{(t)}\right) = \tau^{(t)}_{u|\bar{u}\boldsymbol{x}_i^{(t)}}$$

where $t = 2, \ldots, T$ and $u, \bar{u} = 1, \ldots, k$. The estimated $\hat{\tau}^{(t)}_{u|\bar{u}\boldsymbol{x}_i^{(t)}}$ reflect changes or persistence in the various states over time, given the individual covariates whose effects can be modelled through a multinomial logit parametrization:

$$\log \frac{\mathrm{P}\left(U_i^{(t)} = u \mid U_i^{(t-1)} = \bar{u}, \boldsymbol{X}_i^{(t)} = \boldsymbol{x}_i^{(t)}\right)}{\mathrm{P}\left(U_i^{(t)} = \bar{u} \mid U_i^{(t-1)} = \bar{u}, \boldsymbol{X}_i^{(t)} = \boldsymbol{x}_i^{(t)}\right)} = \log \frac{\tau^{(t)}_{u|\bar{u}\boldsymbol{x}_i^{(t)}}}{\tau^{(t)}_{\bar{u}|\bar{u}\boldsymbol{x}_i^{(t)}}} = \gamma_{0\bar{u}u} + \boldsymbol{x}_i^{(t)\top} \boldsymbol{\gamma}_{1\bar{u}u} \tag{2}$$

for $t = 2, ..., T$ and $\bar{u}, u = 1, ..., k$ with $\bar{u} \neq u$. $\boldsymbol{\gamma}_{\bar{u}u} = \left(\gamma_{0\bar{u}u}, \boldsymbol{\gamma}_{1\bar{u}u}^{\top}\right)^{\top}$ are the parameters vectors to be estimated.

Under the assumptions of *local* and *conditional independence*, the *manifest distribution* of the response variables (i.e., the conditional distribution of $\widetilde{\boldsymbol{Y}}_i$ given $\widetilde{\boldsymbol{X}}_i$) is given by:

$$\begin{aligned}
\mathrm{P}(\widetilde{\boldsymbol{y}}_i \mid \widetilde{\boldsymbol{x}}_i) &= \mathrm{P}\left(\widetilde{\boldsymbol{Y}}_i = \widetilde{\boldsymbol{y}}_i \mid \widetilde{\boldsymbol{X}}_i = \widetilde{\boldsymbol{x}}_i\right) = \\
&= \sum_{\boldsymbol{u}} \mathrm{P}\left(\widetilde{\boldsymbol{Y}}_i = \widetilde{\boldsymbol{y}}_i \mid \widetilde{\boldsymbol{X}}_i = \widetilde{\boldsymbol{x}}_i, \boldsymbol{U}_i = \boldsymbol{u}\right) \times \mathrm{P}\left(\boldsymbol{U}_i = \boldsymbol{u} \mid \widetilde{\boldsymbol{X}}_i = \widetilde{\boldsymbol{x}}_i\right) = \\
&= \sum_{\boldsymbol{u}} \mathrm{P}\left(\boldsymbol{U}_i = \boldsymbol{u} \mid \widetilde{\boldsymbol{X}}_i = \widetilde{\boldsymbol{x}}_i\right) \times \mathrm{P}\left(\widetilde{\boldsymbol{Y}}_i = \widetilde{\boldsymbol{y}}_i \mid \boldsymbol{U}_i = \boldsymbol{u}\right) = \\
&= \sum_{\boldsymbol{u}} \delta_{u^{(1)}|\boldsymbol{x}_i^{(1)}} \prod_{t=2}^{T} \tau^{(t)}_{u^{(t)}|u^{(t-1)}\boldsymbol{x}_i^{(t)}} \times \prod_{t=1}^{T} \prod_{j=1}^{J} \phi_{jy_{ij}^{(t)}|u^{(t)}}
\end{aligned} \tag{3}$$

where $\boldsymbol{u} = (u^{(1)}, \ldots, u^{(T)})$. The vector $\widetilde{\boldsymbol{y}}_i = \left(\boldsymbol{y}_i^{(1)}, \ldots, \boldsymbol{y}_i^{(T)}\right)$ is a realization of $\widetilde{\boldsymbol{Y}}_i$, where $\boldsymbol{y}_i^{(t)}$ is a realization of $\boldsymbol{Y}_i^{(t)}$ with elements $y_{ij}^{(t)}$. The vector $\widetilde{\boldsymbol{x}}_i = \left(\boldsymbol{x}_i^{(1)}, \ldots, \boldsymbol{x}_i^{(T)}\right)$ is a realization of $\widetilde{\boldsymbol{X}}_i$, where $\boldsymbol{x}_i^{(t)} = \left(\boldsymbol{s}_i, \boldsymbol{z}_i^{(t)}\right)$ is a realization of $\boldsymbol{X}_i^{(t)} = \left(\boldsymbol{S}_i, \boldsymbol{Z}_i^{(t)}\right)$.

Parameters estimation is performed maximizing the log-likelihood for a sample of n independent units, i.e., $\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log P(\widetilde{\boldsymbol{y}}_i \mid \widetilde{\boldsymbol{x}}_i)$, using an Expectation-Maximization algorithm (Bartolucci et al., 2013, 2014, 2015). Deterministic and random initializations are implemented to reach the global maximum of $\ell(\boldsymbol{\theta})$ and prevent identifiability issue related to the multimodality of the likelihood function.

3.3 Model selection

The choice of the final LM model for the application consists of two steps: (i) identification of the number of latent states k, and (ii) selection of the covariates to be included in the final model. When the number of latent states k can not be a priori defined based on clinical indications, it can be selected according different measures. Akaike information criterion (AIC) by Akaike (1973) or the Bayesian information criterion (BIC) by Schwarz (1978), defined as

$$\text{AIC} = -2\hat{\ell} + 2g \qquad \text{and} \qquad \text{BIC} = -2\hat{\ell} + \log(n)g,$$

where $\hat{\ell}$ is the maximum of the log-likelihood of the model of interest and g denotes the number of free parameters, are used. In particular, the smaller the values of the above criteria, the better the model represents the optimum compromise between goodness-of-fit and complexity. If the two criteria lead to selecting a different number of states, BIC is usually preferred (Bacci et al., 2014; Bartolucci et al., 2017).

Basic LM models (i.e., LM models with time-heterogeneous transitions and no covariates - named M1) were fitted increasing the value of k from 1 to 10, and the number of latent states k was selected according to the minimum BIC. Once k was determined, a forward strategy was adopted to identify the covariates to be included in the final model. In particular, the smallest basic LM model with k latent states and time-homogeneous transitions (i.e., the LM model restricted to the case in which initial and transition probabilities are parametrized by multinomial logit without covariates - named M2) was initially fitted and then the effect of each covariate on initial and/or transition probabilities (models M3-M12) was added. Only the covariates whose effect reduces the value of the BIC index were included in the final LM model.

3.4 Longitudinal profiles: latent probability and relative risk

In LM models literature, once the model has been estimated, a decoding procedure is usually implemented to obtain a path prediction for each subject, i.e., finding the most likely sequence of latent states on the basis patient-specific observed data (Bartolucci et al., 2013, 2014). However, this sequence represents a summary of how the entire latent process evolves over time, as it only provides information about the most-likely condition without giving details about other states (see Appendix B). To obtain more insights into the entire latent process and its evolution, longitudinal information related to each latent state can be reconstructed for each subject. For each patient-specific observed data $(\widetilde{\boldsymbol{x}}_i, \widetilde{\boldsymbol{y}}_i)$, the Expectation-Maximization algorithm provides the *posterior*

probabilities of variables $U_i^{(t)}$

$$p_{iu}^{(t)} = \mathrm{P}\left(U_i^{(t)} = u \middle| \widetilde{\boldsymbol{Y}}_i = \widetilde{\boldsymbol{y}}_i, \widetilde{\boldsymbol{X}}_i = \widetilde{\boldsymbol{x}}_i\right) \qquad t = 1, \ldots, T \quad u \in \{1, ..., k\}, \tag{4}$$

which can be estimated using recursions and involving the *manifest* distribution in Equation (3). For each latent state $u \in \{1, \ldots, k\}$, probabilities in (4) can be used to reconstruct the *longitudinal latent probability profile* of the i-th subject, as follows:

$$\boldsymbol{p}_{iu} = \left\{ p_{iu}^{(t)} = \mathrm{P}\left(U_i^{(t)} = u \middle| \widetilde{\boldsymbol{Y}}_i = \widetilde{\boldsymbol{y}}_i, \widetilde{\boldsymbol{X}}_i = \widetilde{\boldsymbol{x}}_i\right), \quad t = 1, \ldots, T \right\}. \tag{5}$$

Each profile $\boldsymbol{p}_{iu}$ represents the probability over time t of being in latent state u for individual i, given the observed complete response $\widetilde{\boldsymbol{y}}_i$ and covariates $\widetilde{\boldsymbol{x}}_i$ (if available). Applying this procedure, k *longitudinal latent probability profiles* (one for each latent state) are obtained for each subject i, which can be expressed as a $k \times T$ matrix

$$\boldsymbol{P}_i = \begin{bmatrix} \boldsymbol{p}_{i1} \\ \cdots \\ \cdots \\ \boldsymbol{p}_{ik} \end{bmatrix} = \begin{bmatrix} p_{i1}^{(1)} & p_{i1}^{(2)} & \ldots & p_{i1}^{(T)} \\ \cdots & & & \cdots \\ \cdots & & & \cdots \\ p_{ik}^{(1)} & p_{ik}^{(2)} & \ldots & p_{ik}^{(T)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{p}_i^{(1)} & \boldsymbol{p}_i^{(2)} & \ldots & \boldsymbol{p}_i^{(T)} \end{bmatrix}$$

with longitudinal latent probability profiles $\boldsymbol{p}_{iu}$ as row-components. Columns of $\boldsymbol{P}_i$ represent the vectors $\boldsymbol{p}_i^{(t)}$ of posterior probabilities over time $t = 1, \ldots, T$ and can be seen as Compositional Data (CoDa) vectors belonging to the k-part Aitchison-Simplex $\mathcal{S}^k$ (Aitchison, 1986), i.e.,

$$\boldsymbol{p}_i^{(t)} \in \mathcal{S}^k = \left\{ \boldsymbol{p} = [p_1, ..., p_k] \in \mathbb{R}^k \middle| p_u > 0, u = 1, \ldots, k; \sum_{u=1}^{k} p_u = 1 \right\}. \tag{6}$$

Due to the sum constraint in Equation (6), elements $p_{iu}^{(t)}$ of the composition $\boldsymbol{p}_i^{(t)}$ are mutually dependent features which only carry relative information. In this context, Aitchison (1986) introduced a methodology based on log-ratio transformations of CoDa, which are required to remove constraints and eventually to map the composition to a real space, allowing standard statistical techniques to be applied to the transformed data. In most practical settings, the choice of transformation will depend on the preferred interpretation.

In the current framework, rather than considering the absolute individual elements $p_{iu}^{(t)}$, it could be of interest to examine over time the relative risk of being in a reference latent state $u = R$ in contrast with all the other latent states. Among the transformations introduced by Aitchison (1986), this can be done considering the *additive log-ratios* of each CoDa vector $\boldsymbol{p}_i^{(t)}$, as follows:

$$\begin{aligned} \mathrm{alr}\left(\boldsymbol{p}_i^{(t)}\right) &= \left[\log \frac{p_{i1}^{(t)}}{p_{iR}^{(t)}} \ldots \log \frac{p_{iR-1}^{(t)}}{p_{iR}^{(t)}} \ \log \frac{p_{iR+1}^{(t)}}{p_{iR}^{(t)}} \ldots \log \frac{p_{ik}^{(t)}}{p_{iR}^{(t)}} \right]^T \\ &= \left[r_{i1}^{(t)} \ \ldots \ r_{iR-1}^{(t)} \ r_{iR+1}^{(t)} \ \ldots \ r_{ik}^{(t)} \right]^T \\ &= \boldsymbol{r}_i^{(t)} \in \mathbb{R}^{k-1} \end{aligned} \tag{7}$$

where R is the reference latent state which can be chosen arbitrary among $\{1, \ldots, k\}$. Note that this transformation maps each bounded sample into a real space $\left(\text{alr: } \mathcal{S}^k \to \mathbb{R}^{k-1}\right)$ and if one of the $p_{iu}^{(t)}$ elements is exactly zero, a zero-handling procedure would be needed before applying

the transformation. In that case, an easily applicable possibility would be to replace each zero with a small appropriate value, modifying the non-zero values of the relative composition in a multiplicative way in order to satisfy the sum constraint requirement. For further details see Martín-Fernández et al. (2011). Applying this procedure to each compositions, $k-1$ *longitudinal relative risk profiles* (one for each non-reference state) are obtained for each subject i, which can be expressed as a $(k-1)\times T$ matrix

$$\boldsymbol{R}_i = \begin{bmatrix} \boldsymbol{r}_i^{(1)} & \boldsymbol{r}_i^{(2)} & \dots & \boldsymbol{r}_i^{(T)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{r}_{i1} \\ \dots \\ \boldsymbol{r}_{iR-1} \\ \boldsymbol{r}_{iR+1} \\ \dots \\ \boldsymbol{r}_{ik} \end{bmatrix}$$

where column-element $\boldsymbol{r}_i^{(t)}$ are given by Equation (7) and row-element $\boldsymbol{r}_{iu}$ with $u \neq R$ are the *longitudinal relative risk profile* of state u for subject i, as follows:

$$\boldsymbol{r}_{iu} = \left\{ r_{iu}^{(t)} = \log \frac{p_{iu}^{(t)}}{p_{iR}^{(t)}}, \quad u \neq R,\, t = 1, \dots, T \right\}. \tag{8}$$

Each profile $\boldsymbol{r}_{iu}$ represents the relative risk (in logarithmic scale) over time t of being in latent state $u \neq R$ with respect to the reference state R for individual i. Since this procedure is a transformation-based analysis, transformed elements $r_{iu}^{(t)}$ must then be interpreted with respect to the chosen reference. In particular, a positive (negative) value $r_{iu}^{(t)}$ at time t means that the risk for subject i of being in latent state $u \neq R$ is $\exp\left\{r_{iu}^{(t)}\right\}$ times higher (lower) than being in reference state R.

For the application discussed in this work, the *LOTox states* summarize different levels of overall toxicity burden, representing a proxy for patient's quality of life. Therefore, for each patient i, longitudinal latent probability profile in Equation (5) represents the probability over time of being in the LOTox state u (i.e., the probability over time of developing an overall toxic burden quantified by state u) given patient's history: observed toxicity categories $\widetilde{\boldsymbol{y}}_i$ and personal characteristics $\widetilde{\boldsymbol{x}}_i$ over treatment. Once the LOTox states have been identified, it is reasonable to analyse and interpret the different results in relation to the state characterized by the lowest overall toxicity burden (i.e., "good" toxic condition), which is chosen as the reference R. In this way, the longitudinal relative risk profile in Equation (8) represents the risk of being in LOTox condition $u \neq R$ compared to the lowest toxic status. By reconstructing the *longitudinal LOTox profiles*, it is possible to (i) describe patient's response to therapy over cycles, (ii) quantify the overall toxicity burden evolution over treatment cycles given patient's history and (iii) investigate the individual dynamic changes among latent states, detecting differences in health status and quality of life among patients.

4 Data application

In this section, the results obtained from the application of the proposed LM model to the MRC BO06/EORTC 80931 randomized clinical trial are reported. Statistical analyses were performed in the `R`-software environment (R Core Team, 2020), using `LMest` package by Bartolucci et al. (2017). `R` code for the current study is available at `https://github.com/mspreafico/BO06-LOTox`.

Table 2: Results for Latent Markov (LM) model selection for longitudinal toxicity data with different values of latent states k and different restrictions. The maximum log-likelihood of each model is denoted by $\hat{\ell}$ and g is the number of free parameters. WBC, PLT and NEUT in models M10-12 refers to white blood cell, platelets and neutrophils counts, respectively.

Latent Markov (LM) model	k	g	$\hat{\ell}$	AIC	BIC
M1: Unrestricted LM model without covariates	1	18	-8794.91	17625.81	17696.59
	2	35	-8420.19	16910.38	17048.01
	3	68	-8216.99	16569.98	16837.37
	4	111	-8035.21	16292.42	16728.90
	5	164	-7902.59	16133.18	16778.07
	6	227	-7793.14	16040.29	16932.91
	7	300	-7688.12	15976.24	17155.91
	8	383	-7603.30	15972.61	17478.66
	9	476	-7530.49	16012.98	17884.73
	10	579	-7462.34	16082.68	18359.45
M2: Multinomial logit LM model without covariates	4	63	-8069.21	16264.43	16512.16
M3: M2 + regimen effect on initial prob.	4	66	-8065.49	16262.97	16522.50
M4: M2 + gender effect on initial prob.	4	66	-8061.73	16255.45	16514.98
M5: M2 + age effect on initial prob.	4	66	-8055.35	16242.69	16502.22
M6: M2 + regimen effect on transition prob.	4	75	-8063.37	16276.74	16571.66
M7: M2 + gender effect on transition prob.	4	75	-8060.33	16270.66	16565.58
M8: M2 + age effect on transition prob.	4	75	-8061.07	16272.14	16567.06
M9: M2 + time-var chemotherapy dose on both prob.	4	78	-8045.55	16247.10	16553.82
M10: M2 + time-var WBC count on both prob.	4	78	-8062.53	16281.07	16587.78
M11: M2 + time-var PLT count on both prob.	4	78	-8047.15	16250.30	16557.02
M12: M2 + time-var NEUT count on both prob.	4	78	-8062.67	16281.34	16588.05

4.1 Latent Markov model for longitudinal toxicity data

For each cycle $t = 1, \ldots, 6$, let $\mathcal{J} = \{naus, inf, oral, car, oto, neur\}$ be the set of non-haematological toxicities, representing response variables $Y_{ij}^{(t)}$. The relative sets of response categories identified in Section 2.2 were coded from 0 to $c_j - 1$, as follows:

$$\mathcal{C}_j = \{0 : none, 1 : mild, 2 : moderate, 3 : severe\} \quad \text{for } generic \text{ toxicities } (j = 1, 2, 3),$$
$$\mathcal{C}_j = \{0 : no, 1 : yes\} \quad \text{for } drug\text{-}specific \text{ toxicities } (j = 4, 5, 6).$$

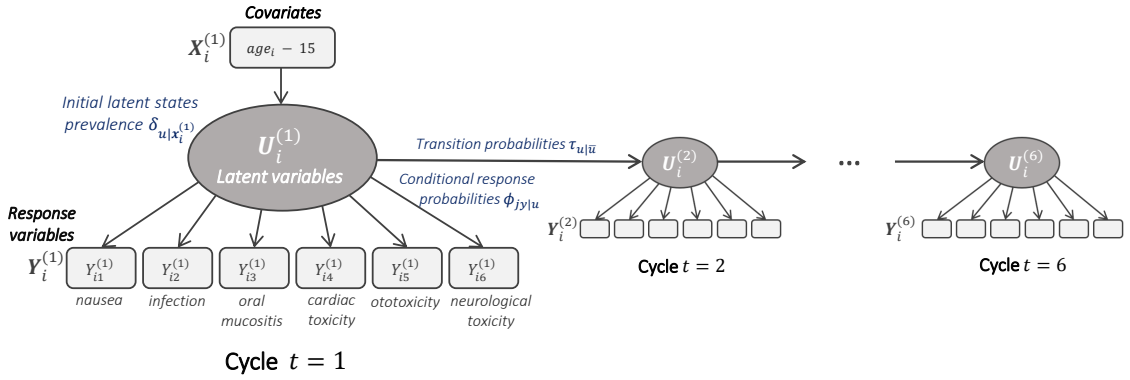

Figure 2: Path diagram for a given subject i under the latent Markov model M5 with non-haematological toxicities as response variables, time-homogeneous transitions and *age* at randomization as covariate affecting the initial probabilities of the latent variables.

The procedure described in Section 3.3 was applied to first identify the number of latent states

k and then select the covariates to be included in the final model. Age, gender and allocated regimen at randomization were considered as time-fixed covariates, while percentage of achieved chemotherapy dose up to cycle t, white blood cell, neutrophils and platelets counts measured at each cycle were considered as time-varying ones. Results are shown in Table 2. The unrestricted LM model without covariates (M1) with the minimum BIC (16728.90) was obtained for $k = 4$, identifying a latent process with four *LOTox states*. Moreover, the basic model M2 with initial and transition probabilities parametrized by multinomial logit was preferable (BIC = 16512.16) to the unrestricted model M1 with the same number of latent states. Several models (M3-M12) with four latent states, obtained from M2 adding covariates effect to initial and/or transition probabilities, were fitted. By comparing models M3-M12 with M2, age (centred with respect to the mean) at randomization was the only covariate leading to a significant improvement in terms of both BIC and AIC (M5). Model M5, whose path diagram for a given subject i is shown in Figure 2, was then selected as final model:

- initial probabilities were associated with patient's *age* at randomization and Equations (1) for a patient i became

$$\log \frac{\delta_{u|age_i}}{\delta_{1|age_i}} = \beta_{0u} + \beta_{1u} \cdot (age_i - 15) \qquad u = 2, 3, 4; \tag{9}$$

- transition probabilities were assumed time-homogeneous and Equations (2) became

$$\log \frac{\tau_{u|\bar{u}}}{\tau_{\bar{u}|\bar{u}}} = \gamma_{0\bar{u}u} \qquad \bar{u}, u = 1, 2, 3, 4 \text{ with } \bar{u} \neq u. \tag{10}$$

Figure 3 shows the estimated conditional response probabilities $\hat{\phi}_{jy|u}$ for each type of non-haematological toxicity under the selected model M5, which can be used for interpreting the latent states. In each toxicity-panel, each column refers to a different latent state $u \in \{1, 2, 3, 4\}$. People in good conditions are allocated in state 1, since for all non-haematological toxicities the most probable category was the absence of the adverse event. State 2 seems to correspond to patients with non-severe nausea and it was the only state where *drug-specific* toxicities occurred with a relevant probability, especially for ototoxicity where $\hat{\phi}_{51|2} = 0.429$. State 3 seems to be characterized by patients undergoing only nausea or vomiting, mostly moderate or severe. In State 4 people with multiple *generic* toxicities - mostly severe or moderate - with the certainty of having nausea ($\hat{\phi}_{10|4} = 0$) are present. Based on these results, the following LOTox states labelling were derived:

- State 1: quite good conditions (non-toxic) $\rightarrow$ *no LOTox*
- State 2: non-severe nausea with possible *drug-specific* AEs $\rightarrow$ *moderate LOTox*
- State 3: moderate/severe nausea/vomiting only $\rightarrow$ *low LOTox* (limited to nausea)
- State 4: multiple severe/moderate *generic* toxicities $\rightarrow$ *high LOTox*.

Note that the states numbering (from 1 to 4) does not correspond with the progressive severity of overall toxicity burden (from *no* to *high*).

Table 3 displays the estimated regression parameters $\hat{\boldsymbol{\beta}}_u = \left(\hat{\beta}_{0u}, \hat{\beta}_{1u}\right)$ for the initial probabilities in Equation (9) and the estimated transition probabilities $\hat{\tau}_{u|\bar{u}}$ in Equation (10). The estimated intercepts indicates that for 15-year patients the most prevalent state at cycle 1 was *low LOTox*

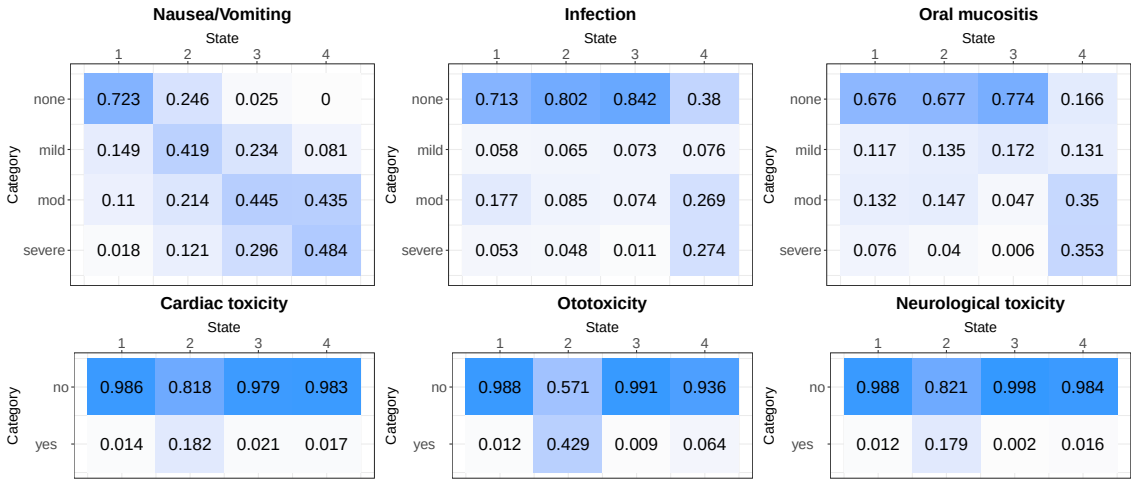

Figure 3: Estimated conditional response probabilities $\hat{\phi}_{jy|u}$ for the final LM model in Figure 2. Each panel refers to a different toxicity $j \in \mathcal{J} = \{1 : naus, 2 : inf, 3 : oral, 4 : car, 5 : oto, 6 : neur\}$. Each row refers to a response categories $y \in \{none; mild; moderate; severe\}$ for $j = 1, 2, 3$ (*generic* toxicities) and $y \in \{no; yes\}$ for $j = 4, 5, 6$ (*drug-specific* toxicities). Each column refers to a latent states $u \in \{1, 2, 3, 4\}$.

Table 3: Estimated regression parameters affecting the distribution of the initial probabilities in Equation (9) and estimated transition probabilities in Equation (10).

Regression parameters for initial probabilities				
	u	2	3	4
Intercept	$\hat{\beta}_{0u}$	-1.2679	1.0138	-0.3031
Age	$\hat{\beta}_{1u}$	0.1858	0.0014	0.0512
Transition probabilities from $\bar{u}$ to u ($\hat{\tau}_{u\|\bar{u}}$)				
$\bar{u} \setminus u$	1	2	3	4
1	0.9674	0.0167	0.0032	0.0127
2	0.0525	0.9214	0.0245	0.0016
3	0.1070	0.0526	0.7581	0.0824
4	0.1555	0.0356	0.0868	0.7221

state 3 (limited to nausea), followed by *no LOTox* state 1, *high LOTox* state 4 and *moderate LOTox* state 2. The estimates for *age* were all positive, indicating that older individuals reported a higher overall severity at the first cycle compared to younger patients. The estimated transition probabilities $\hat{\tau}_{u|\bar{u}}$ shows a quite high persistence in the same state, especially for non-toxic state 1 and moderate state 2, where *drug-specific AEs* may also lead to permanent conditions (see Appendix Table A.1). The highest transition probability was 15.6% and was observed from the *high LOTox* state 4, where the effects of *generic AEs* are reversible and temporary, to the first non-toxic state. Other transitions were observed from *high LOTox* state 4 to nausea/vomiting only in state 3 (8.7%) and from *low LOTox* state 3 (limited to nausea) to *no LOTox* state 1 (10.7%) or *high LOTox* state 4 (8.2%). The remaining transition probabilities were always lower than 8%.

Starting from these parameter estimates, Figure 4 (left panel) displays the estimated vectors of initial probabilities $\hat{\boldsymbol{\delta}}_i = \left(\hat{\delta}_{1|age_i}, \hat{\delta}_{2|age_i}, \hat{\delta}_{3|age_i}, \hat{\delta}_{4|age_i}\right)$ for patients aged 10, 15 and 20 years old and the vector $\bar{\boldsymbol{\delta}} = (\bar{\delta}_1, \bar{\delta}_2, \bar{\delta}_3, \bar{\delta}_4) = (0.202, 0.093, 0.557, 0.148)$ obtained as average of vectors $\hat{\boldsymbol{\delta}}_i$ over all the 377 subjects in the sample. On average, at cycle 1 *low LOTox* state 3 of subjects with nausea/vomiting only had the largest dimension (55.7%), followed by 20.2% of individuals for *no*

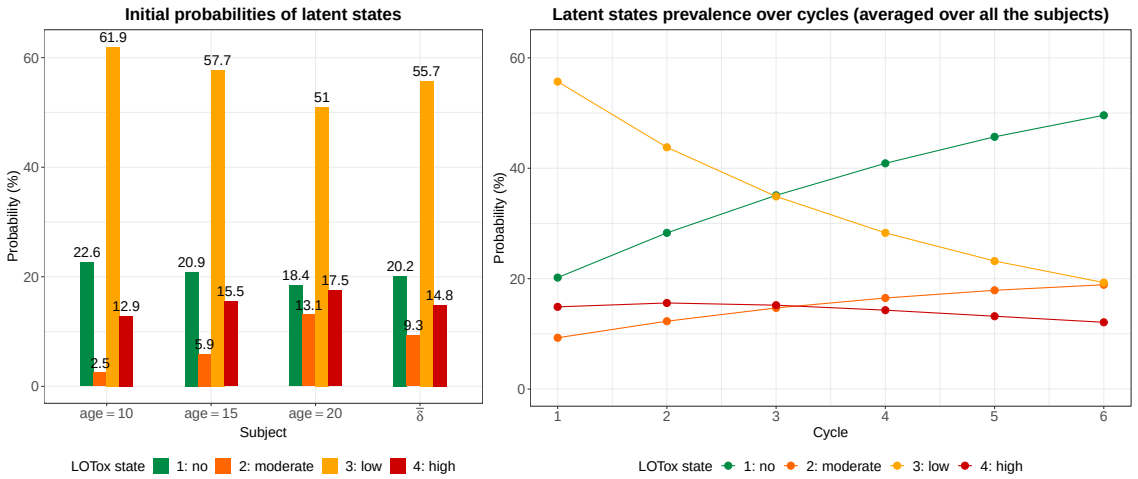

Figure 4: Left panel: estimated initial probabilities of latent states for patients aged 10, 15 and 20 years old and average $\bar{\boldsymbol{\delta}}$ of the initial probabilities over all the 377 subjects in the sample.
Right panel: latent states prevalences over cycles $t = 1, \ldots, 6$ averaged over all the subjects.
Different colours refer to different Latent Overall Toxicity (LOTox) state (green: *no LOTox state 1*; yellow: *low LOTox state 3*; orange: *moderate LOTox state 2*; red: *high LOTox state 4*).

LOTox state 1. *No* and *low LOTox* states together, representing the states with the lowest overall toxic severity, accounted for more than 75% of the patients, whereas less than 25% belonged to the latent states corresponding to the worst toxic conditions (*moderate* and *high LOTox* states 2 and 4).

Right panel in Figure 4 shows the estimated average probability of each latent state at each time-occasion, i.e., the latent states prevalences averaged over all the subjects at each cycle. On average, the presence of *low* overall severity limited to nausea (state 3) decreased over cycles from 55.7% to 19.3% ($t = 6$), whereas *no* and *moderate* overall toxicity (state 1 and 2, respectively) increased from 20.2% to 49.7% and from 9.2% to 18.9%. The presence in *high* overall severity (state 4) was rather stable over cycles ranging in 10.1%-15.6%, with peaks at cycles 2 and 3.

4.2 Longitudinal profiles of Latent Overall Toxicity

Once the parameters were estimated for the final LM model, the longitudinal latent probability profiles $\boldsymbol{p}_{iu}$ were reconstructed for each patient i and latent state u, as explained in Section 3.4. In case of longitudinal toxicity data, profiles $\boldsymbol{p}_{iu}$ in Equation (5) can be also named *longitudinal Probability profiles of LOTox* (*P-LOTox*) since they represent the probability over cycles $t = 1, 2..., 6$ of being in the LOTox state $u \in \{1, 2, 3, 4\}$ for each patient i, given the observed toxic categories over treatment and individual characteristics (i.e., the age at randomization).

Figure 5 shows the longitudinal P-LOTox profiles $\boldsymbol{p}_{iu}$ for four patients $i = \{A, B, C, D\}$ aged 15 years old and with different observed toxic categories over cycles, as reported in Appendix C. Each panel refers to a different patient and displays the individual realisations of the latent process over cycles. Different patterns of overall toxicity evolution during treatment can be observed between subjects, based on patient-specific observed toxicity data. As an example, right panel shows that at cycle 1 patient D had probabilities 79.6% of being in *low LOTox* state, 15.5% of having a non-toxic condition, 4.5% and 0.4% of *high* and *moderate LOTox*, respectively. Then, the probabilities evolved over the cycles, as shown by the four profiles, ending with a 99.7% probability of being in quite good conditions at the end of treatment.

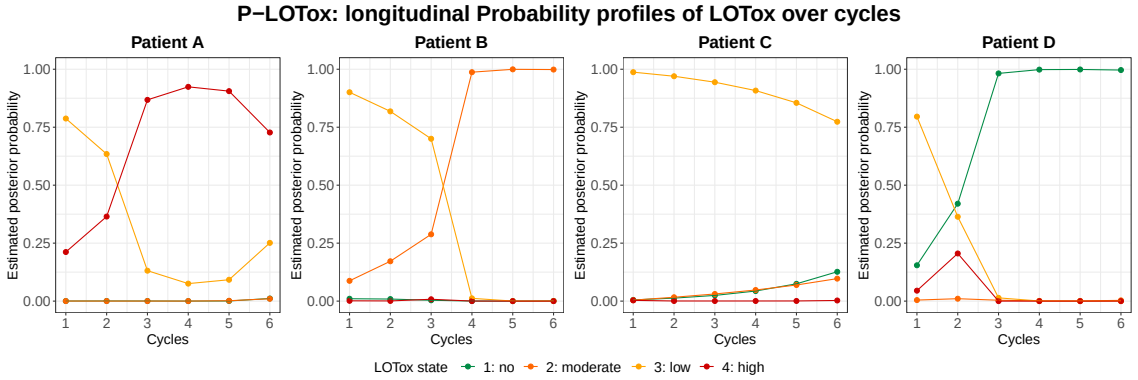

Figure 5: Longitudinal Probability profiles of Latent Overall Toxicity (P-LOTox) $\boldsymbol{p}_{iu}$ for four 15 year-old patients at randomization. Each panel refers to a different patient $i = \{A, B, C, D\}$ in Appendix C. Different colours refer to different latent states $u \in \{1, 2, 3, 4\}$ (green: *no LOTox state 1*; yellow: *low LOTox state 3*; orange: *moderate LOTox state 2*; red: *high LOTox state 4*).

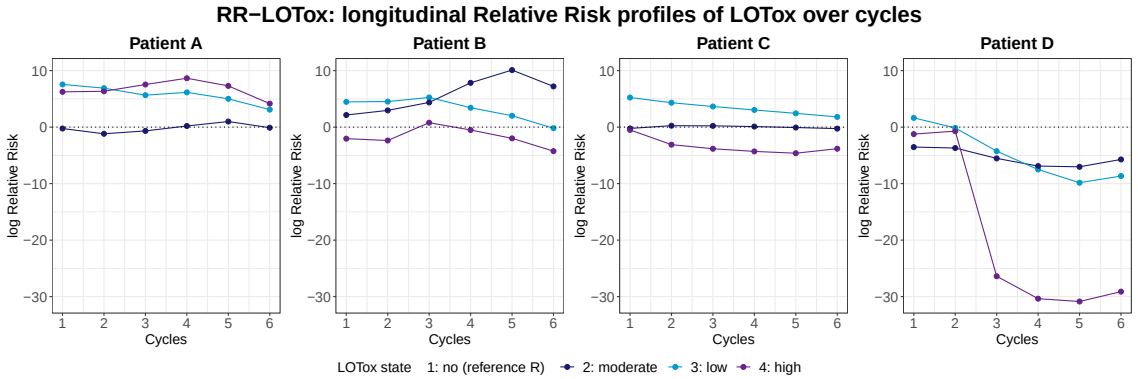

Figure 6: Longitudinal Relative Risk profiles of Latent Overall Toxicity (RR-LOTox) $\boldsymbol{r}_{iu}$ for four 15 year-old patients at randomization. Each panel refers to a different patient $i = \{A, B, C, D\}$ in Appendix C. Reference LOTox state is *no LOTox* state $R = 1$. Different colours refer to different non-reference latent states $u \in \{2, 3, 4\}$ (light-blue: *low LOTox state 3* vs *no LOTox*; blue: *moderate LOTox state 2* vs *no LOTox*; purple: *high LOTox state 4* vs *no LOTox*).

As mentioned in Section 4.1, the lowest toxic burden is represented by the non-toxic state 1 of patients in quite good conditions, which was chosen as reference state ($R = 1$: *no LOTox*) to reconstruct the longitudinal latent relative risk profiles $\boldsymbol{r}_{iu}$ for each patient i and latent state $u \in \{2, 3, 4\}$, as explained in Section 3.4. In case of longitudinal toxicity data, profiles $\boldsymbol{r}_{iu}$ in Equation (8) can be also called *longitudinal Relative Risk profiles of LOTox* (*RR-LOTox*) since they represent for each patient i the relative risk (in logarithmic scale) over cycles $t = 1, 2..., 6$ of being in the LOTox state $u \in \{2, 3, 4\}$ rather than in the non-toxic state $R = 1$, given the observed toxic categories over treatment and individual characteristics (i.e., the age at randomization).

Figure 6 shows the longitudinal RR-LOTox profiles $\boldsymbol{r}_{iu}$ for patients $i = \{A, B, C, D\}$ in Appendix C. Each panel refers to a different subject. Different toxic risk progressions during treatment may be observed among patients, depending on their observed toxicity data. For example, right panel shows that at first cycle patient D's risk of being in *low LOTox* state was 5.14 times higher the risk of having a non-toxic condition, whereas risks of *high* and *moderate LOTox* were 0.29 and 0.03 times lower, respectively. Then, RR-LOTox profiles evolved over the cycles, as shown by the four trajectories, ending up with negligible relative risks (< 0.01) for *low/moderate/high LOTox*

conditions compared with a non-toxic condition at the end of treatment.

Both longitudinal P-LOTox and RR-LOTox profiles summarize and quantify the overall toxic risk over time for each patient based on observed individual characteristics, capturing differences in the overall history of toxicity across patients. P-LOTox profiles reflect the absolute size of the probabilities over time for each latent state, whereas RR-LOTox profiles focus on the relative risk with respect to the clinically desirable condition, i.e., the non-toxic one. Both information in cooperation with medical staff could lead to improvements in the definition of new guidelines and useful tools for healthcare assessment and treatment planning.

5 Discussion

Due to the presence of multiple types of Adverse Events (AEs) with different levels of severity, identifying the actual extent of toxic burden and investigating the evolution of patient's overall toxicity represent challenging problems in cancer research. AEs are one of the main factors determining clinical decisions in medical interventions and treatment planning, playing a fundamental role in health assessment and patient monitoring. The development of statistical methods able to summarize multiple AEs and to deal with the complexity of chemotherapy data, considering both the longitudinal and categorical aspects of toxicity levels progression, is then necessary and clinically relevant.

This paper proposed a new taxonomy based on LM model with covariates and CoDa methods to provide novel techniques for investigating the evolution of the latent overall toxicity condition for each patient over chemotherapy treatment. This is important in light of the need to develop new tools to support clinical decisions in tailored interventions for effective management of adverse symptoms and treatments. This approach was applied to longitudinal chemotherapy data for osteosarcoma patients from MRC BO06/EORTC 80931 Randomized Controlled Trial, where toxic AEs were registered over cycles according to CTCAE grades (U.S. Department of Health and Human Services, 2006) through apposite case report forms, as indicated by the protocol (Lewis et al., 2007).

By assuming the existence of a LM chain for the LOTox condition of a patient, the proposed taxonomy identified sub-populations of patients characterized by a common distribution of toxic categories, and by a similar overall toxicity burden. Four LOTox states were found, representing different levels of multiple AEs severity: (i) people in quite good conditions (*no LOTox state 1*), (ii) patients undergoing only nausea or vomiting - mostly moderate or severe - (*low LOTox state 3*), (iii) subjects with non-severe nausea and the possibility to develop *drug-specific* AEs (*moderate LOTox state 2*), or (iv) people with multiple severe/moderate *generic* toxicities (*high LOTox state 4*). The LM approach estimated the initial prevalence of each state and the probability of individual changes over time. This allowed to reconstruct the patient-specific longitudinal LOTox profiles to assess the dynamic evolution of overall toxicity burden during treatment for each subject.

Both longitudinal P-LOTox and RR-LOTox profiles captured the individual realisations of the latent process over cycles, showing different patterns of overall toxicity evolution during treatment among patients. On one hand, P-LOTox profiles illustrated the latent process using absolute terms, giving insights into the actual probabilities of being in the various LOTox states over cycles. On the other hand, RR-LOTox profiles – obtained by additive log-ratios transformation – reported relative risk measures to emphasize the difference between low/moderate/high LOTox states and the clinically desirable non-toxic condition. These aspects can not be investigated using a simple

path prediction (see Appendix B). Together, absolute probabilities and relative risks provide a complete picture of the individual LOTox dynamics during treatment, which may be considered as a proxy for patient's quality of life and used to describe patient's response to therapy over cycles in terms of toxic AEs.

This retrospective exploratory analysis has some limitations. The procedure used to select the final model may miss the best available one, since not all possible models have been fitted. However, it is computationally efficient and follows a standard stepwise forward selection approach. The analysis was performed on a single trial in osteosarcoma, considering only non-haematological toxicities. Other factors of potential interest were not routinely recorded during the trial, including among others nephrotoxicity, lymphocytes count or tumour size. To get more information about the robustness of the model developed in this study, it should be applied to other osteosarcoma data provided that the toxicity are longitudinally recorded.

On the other hand, this LM procedure can be adapted and applied to other cancer studies. Provided that toxicities are recorded according to the CTCAE scale or an analogous grading system, the LM approach represents a general and flexible method to quantify the personal evolution of overall toxic risk during chemotherapy. Moreover, this work opens doors to further researches, both in the field of statistical methodology development as well as in cancer research. The additive log-ratios transformation allowed to remove non-negative and sum-to-one constraints of the CoDa vectors, mapping the compositions to a real space and opening up the possibility to apply standard statistical techniques to the transformed data. Patients can be stratified in different risk groups to be used during treatment, based on their different LOTox dynamics. Association between longitudinal LOTox profiles and survival outcomes may provide new insights in the treatment effect during the evolution of the disease.

In conclusion, the proposed approach provided novel techniques to summarise and quantify patient's overall toxic risk and its evolution during treatment. In cooperation with medical staff, these techniques could lead to improvements in the definition of new guidelines and useful tools for healthcare assessment and treatment planning, trying to reduce the impact of therapies in terms of toxic AEs.

Acknowledgements

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

Data Availability Statement

Data are not publicly available due to privacy restrictions. Access to the full dataset of MRC BO06 trial can be requested to MRC Clinical Trials Unit at UCL, Institute of Clinical Trials and Methodology, UCL, London.

References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Chapman & Hall Ltd.

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In Petrov, B. N. and Csaki, F., editors, *Proceedings of the 2nd International Symposium on Information Theory*, pages 267–281. Budapest: Akademiai Kiado.

Al-malky, H. S., Harthi, S. E. A., and Osman, A.-M. M. (2020). Major obstacles to doxorubicin therapy: Cardiotoxicity and drug resistance. *Journal of Oncology Pharmacy Practice*, 26(2):434–444.

Anninga, J. K., Gelderblom, H., Fiocco, M., Kroep, J. R., Taminiau, A. H., Hogendoorn, P. C., and Egeler, R. M. (2011). Chemotherapeutic adjuvant treatment for osteosarcoma: Where do we stand? *European Journal of Cancer*, 47(16):2431–2445.

Bacci, S., Pandolfi, S., and Pennoni, F. (2014). A Comparison of Some Criteria for States Selection in the Latent Markov Model for Longitudinal Data. *Advances in Data Analysis and Classification*, 8(2):125–145.

Bartolucci, F., Farcomeni, A., and Pennoni, F. (2013). *Latent Markov Models for Longitudinal Data*. Chapman & Hall/CRC, Boca Raton.

Bartolucci, F., Farcomeni, A., and Pennoni, F. (2014). Latent Markov models: a review of a general framework for the analysis of longitudinal data with covariates. *TEST*, 23:433–465.

Bartolucci, F., Lupparelli, M., and Montanari, G. E. (2009). Latent Markov model for longitudinal binary data: An application to the performance evaluation of nursing homes. *The Annals of Applied Statistics*, 3(2):611 – 636.

Bartolucci, F., Montanari, G. E., and Pandolfi, S. (2015). Three-step estimation of latent Markov models with covariates. *Computational Statistics & Data Analysis*, 83:287–301.

Bartolucci, F., Pandolfi, S., and Pennoni, F. (2017). LMest: An R Package for Latent Markov Models for Longitudinal Categorical Data. *Journal of Statistical Software*, 81(4):1–38.

Bekele, B. N. and Thall, P. F. (2004). Dose-Finding Based on Multiple Toxicities in a Soft Tissue Sarcoma Trial. *Journal of the American Statistical Association*, 99(465):26–35.

Carbini, M., Suárez-Fariñas, M., and Maki, R. G. (2018). A Method to Summarize Toxicity in Cancer Randomized Clinical Trials. *Clinical Cancer Research*.

Collins, L. M. and Lanza, S. (2010). *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*. John Wiley and Sons Inc.

dos Santos, N. A. G., Ferreira, R. S., and dos Santos, A. C. (2020). Overview of cisplatin-induced neurotoxicity and ototoxicity, and the protective agents. *Food and Chemical Toxicology*, 136:111079.

Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G. (2009). *Longitudinal data analysis*. Chapman & Hall, CRC, London.

Hirakawa, A., Sudo, K., Yonemori, K., Sadachi, R., Kinoshita, F., Kobayashi, Y., Okuma, H. S., Kawachi, A., Tamura, K., Fujiwara, Y., Rubinstein, L., and Takebe, N. (2019). A Comparative Study of Longitudinal Toxicities of Cytotoxic Drugs, Molecularly Targeted Agents, Immunomodulatory Drugs, and Cancer Vaccines. *Clinical Pharmacology & Therapeutics*, 106(4):803–809.

Juang, B. and Rabiner, L. (1991). Hidden Markov models for speech recognition. *Technometrics*, 33:251–272.

Lancia, C., Spitoni, C., Anninga, J., Whelan, J., Sydes, M. R., Jovic, G., and Fiocco, M. (2019). Marginal structural models with dose-delay joint-exposure for assessing variations to chemotherapy intensity. *Statistical Methods in Medical Research*, 28(9):2787–2801. PMID: 29916309.

Lee, S., Hershman, D., Martin, P., Leonard, J., and Cheung, Y. (2012). Toxicity burden score: a novel approach to summarize multiple toxic effects. *Annals of Oncology*, 23(2):537–541.

Lewis, I., Nooij, M. A., Whelan, J., Sydes, M. R., Grimer, R., Hogendoorn, P. C. W., Memon, M. A., Weeden, S., Uscinska, B. M., van Glabbeke, M., Kirkpatrick, A., Hauben, E. I., Craft, A. W., Taminiau, A. H. M., and On behalf of MRC BO06 and EORTC 80931 collaborators and European Osteosarcoma Intergroup (2007). Improvement in Histologic Response But Not Survival in Osteosarcoma Patients Treated With Intensified Chemotherapy: A Randomized Phase III Trial of the European Osteosarcoma Intergroup. *JNCI: Journal of the National Cancer Institute*, 99(2):112–128.

Martín-Fernández, J. A., Palarea-Albaladejo, J., and Olea, R. A. (2011). Dealing with Zeros. In Pawlowsky-Glahn, V. and Buccianti, A., editors, *Compositional Data Analysis*, chapter 4, pages 43–58. John Wiley & Sons, Ltd.

McTiernan, A., Jinks, R. C., Sydes, M. R., Uscinska, B., Hook, J. M., van Glabbeke, M., Bramwell, V., Lewis, I. J., Taminiau, A. H., Nooij, M. A., Hogendoorn, P. C., Gelderblom, H., and Whelan, J. S. (2012). Presence of chemotherapy-induced toxicity predicts improved survival in patients with localised extremity osteosarcoma treated with doxorubicin and cisplatin: A report from the European Osteosarcoma Intergroup. *European Journal of Cancer*, 48(5):703–712.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ritter, J. and Bielack, S. (2010). Osteosarcoma. *Annals of Oncology*, 21(suppl 7):vii320–vii325.

Rogatko, A., Babb, J. S., Wang, H., Slifker, M. J., and Hudes, G. R. (2004). Patient Characteristics Compete with Dose as Predictors of Acute Treatment Toxicity in Early Phase Clinical Trials. *Clinical Cancer Research*, 10(14):4645–4651.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464.

Sivendran, S., Latif, A., McBride, R. B., Stensland, K. D., Wisnivesky, J., Haines, L., Oh, W. K., and Galsky, M. D. (2014). Adverse Event Reporting in Cancer Clinical Trial Publications. *Journal of Clinical Oncology*, 32(2):83–89.

Smeland, S., Bielack, S. S., Whelan, J., Bernstein, M., Hogendoorn, P., Krailo, M. D., Gorlick, R., Janeway, K. A., Ingleby, F. C., Anninga, J., Antal, I., Arndt, C., Brown, K., Butterfass-Bahloul, T., Calaminus, G., Capra, M., Dhooge, C., Eriksson, M., Flanagan, A. M., Friedel, G., ..., and Marina, N. (2019). Survival and prognosis with osteosarcoma: outcomes in more than 2000 patients in the EURAMOS-1 (European and American Osteosarcoma Study) cohort. *European Journal of Cancer*, 109:36–50.

Souhami, R. L., Craft, A. W., Van der Eijken, J. W., Nooij, M., Spooner, D., Bramwell, V. H., Wierzbicki, R., Malcolm, A. J., Kirkpatrick, A., Uscinska, B. M., Van Glabbeke, M., and Machin, D. (1997). Randomised trial of two regimens of chemotherapy in operable osteosarcoma: a study of the European Osteosarcoma Intergroup. *The Lancet*, 350(9082):911–917.

Spreafico, M., Ieva, F., Arlati, F., Capello, F., Fatone, F., Fedeli, F., Genalti, G., Anninga, J., Gelderblom, H., and Fiocco, M. (2021). Novel longitudinal Multiple Overall Toxicity (MOTox) score to quantify adverse events experienced by patients during chemotherapy treatment: a retrospective analysis of the MRC BO06 trial in osteosarcoma. *BMJ Open*, 11(12):e053456.

Thanarajasingam, G., Atherton, P. J., Novotny, P. J., Loprinzi, C. L., Sloan, J. A., and Grothey, A. (2016). Longitudinal adverse event assessment in oncology clinical trials: the Toxicity over Time (ToxT) analysis of Alliance trials NCCTG N9741 and 979254. *The Lancet Oncology*, 17(5):663–670.

Thanarajasingam, G., Hubbard, J. M., Sloan, J. A., and Grothey, A. (2015). The Imperative for a New Approach to Toxicity Analysis in Oncology Clinical Trials. *JNCI: Journal of the National Cancer Institute*, 107(10). djv216.

Thanarajasingam, G., Leonard, J. P., Witzig, T. E., Habermann, T. M., Blum, K. A., Bartlett, N. L., Flowers, C. R., Pitcher, B. N., Jung, S.-H., Atherton, P. J., Tan, A., Novotny, P. J., and Dueck, A. C. (2020). Longitudinal Toxicity over Time (ToxT) analysis to evaluate tolerability: a case study of lenalidomide in the CALGB 50401 (Alliance) trial. *The Lancet Haematology*, 7(6):e490–e497.

Trotti, A., Colevas, A., Setser, A., Rusch, V., Jaques, D., Budach, V., Langer, C., Murphy, B., Cumberlin, R., Coleman, C., and Rubin, P. (2003). CTCAE v3.0: development of a comprehensive grading system for the adverse effects of cancer treatment. *Seminars in Radiation Oncology*, 13(3):176–181.

Trotti, A., Pajak, T. F., Gwede, C. K., Paulus, R., Cooper, J., Forastiere, A., Ridge, J. A., Watkins-Bruner, D., Garden, A. S., Ang, K. K., and Curran, W. (2007). TAME: development of a new method for summarising adverse events of cancer treatment by the Radiation Therapy Oncology Group. *The Lancet Oncology*, 8(7):613–624.

U.S. Department of Health and Human Services (2006). *Common Terminology Criteria for Adverse Events v3.0 (CTCAE)*. *https://www.eortc.be/services/doc/ctc/ctcaev3.pdf*.

Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13(2):260–269.

Wiggins, L. (1973). *Panel analysis: latent probability models for attitude and behaviour processes*. Elsevier, Amsterdam.

Zhang, S., Chen, Q., and Wang, Q. (2016). The use of and adherence to CTCAE v3.0 in cancer clinical trial publications. *Oncotarget*, 7(40):65577–65588.

Appendix

A MRC BO06/EORTC 80931 RCT protocol

Data from the MRC BO06/EORTC 80931 Randomized Controlled Trial (RCT) for patients with non-metastatic high-grade osteosarcoma recruited between 1993 and 2002 were analysed (Lewis et al., 2007). The trial randomised patients between conventional treatment with doxorubicin (DOX) and cisplatin (CDDP) given every 3 weeks (*Reg-C*) versus a dose-intense regimen of the same two drugs given every 2 weeks (*Reg-DI*), supported by granulocyte colony-stimulating factor. Chemotherapy was administered for six cycles (a cycle is a period of either 2 or 3 weeks depending on the allocated regimen), before and after surgical removal of the primary osteosarcoma. In both arms, DOX (75 mg/m^2) plus CDDP (100 mg/m^2) were given over six cycles. Surgery to remove the primary tumour was scheduled at week 6 after starting treatment in both arms, that is, after 2 cycles (2 × [DOX+CDDP]) in regimen-C and after 3 cycles (3 × [DOX+CDDP]) in regimen-DI. Postoperative chemotherapy was intended to resume 2 weeks after surgery in both arms. Figure A.1 shows the trial design. Laboratory tests were usually performed before each cycle of chemotherapy (in some cases also during and after the cycle) in order to monitor patient's health status and the development of toxicities or adverse events. Non-haematological chemotherapy-induced toxicity for nausea/vomiting, infection, oral mucositis, cardiac toxicity, ototoxicity and neurological toxicity were graded according to the Common Terminology Criteria for Adverse Events Version 3 (CTCAE v3.0) by U.S. Department of Health and Human Services (2006), with grades ranging from 0 (none) to 4 (life-threatening), as shown in Table A.1. Delays or chemotherapy dose reductions during

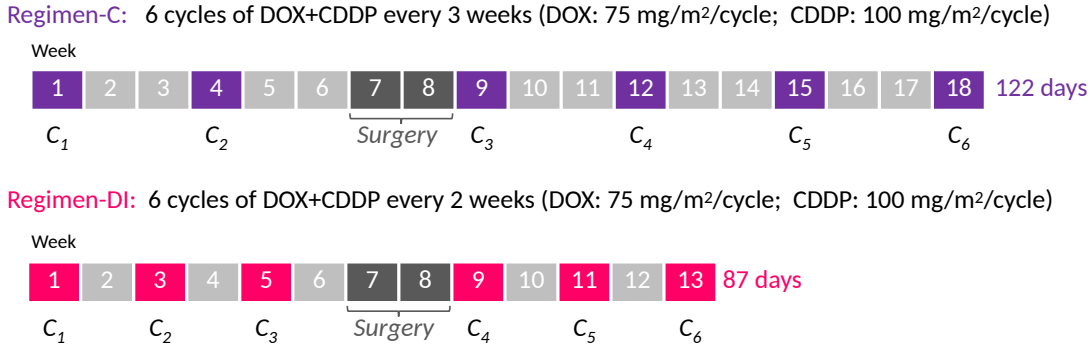

Figure A.1: Patients are randomized at baseline to one of the two regimens, with the same anticipated cumulative dose but different duration.

Table A.1: Toxicity coding based on Common Terminology Criteria for Adverse Events (CTCAE) v3.0 by U.S. Department of Health and Human Services (2006) for non-haematological chemotherapy-induced toxicity related to nausea/vomiting, infection, oral mucositis, cardiac toxicity, ototoxicity and neurological toxicity.

Toxicity	**Grade 0**	**Grade 1**	**Grade 2**	**Grade 3**	**Grade 4**
Nausea/Vomiting	None	Nausea	Transient vomiting	Continuative vomiting	Intractable vomiting
Infection	None	Minor infection	Moderate infection	Major infection	Major infection with hypotension
Oral Mucositis	No change	Soreness or erythema	Ulcers: can eat solid	Ulcers: liquid diet only	Alimentation not possible
Cardiac toxicity	No change	Sinus tachycardia	Unifocal PVC arrhythmia	Multifocal PVC	Ventricular tachycardia
Ototoxicity	No change	Slight hearing loss	Moderate hearing loss	Major hearing loss	Complete hearing loss
Neurological toxicity	None	Paraesthesia	Severe paraesthesia	Intolerable paraesthesia	Paralysis

treatment were possible in case of toxicity. Additional details can be found in the primary analysis of the trial by Lewis et al. (2007).

B Path prediction for latent Markov models

In latent Markov models literature, once the model has been estimated, a decoding procedure is usually implemented to obtain a path prediction $\overset{*}{\boldsymbol{u}}_i = \left(\overset{*}{u}_i^{(1)}, \ldots, \overset{*}{u}_i^{(T)}\right)$ for each subject i, on the basis patient-specific observed data over time.

Among the developed procedures, *local decoding* finds the most likely state occupied by a subject at any time point t: elements of $\overset{*}{\boldsymbol{u}}_i$ can be obtained by maximizing the *posterior probabilities* at

each time t in Equation (4), as follows

$$\overset{*}{u}_i^{(t)} = \max_{u \in \{1,\dots,k\}} p_{iu}^{(t)} \quad \text{for all } t = 1, ..., T.$$

As an alternative, *global decoding* finds the most likely sequence of latent states for a given subject on the basis of the responses he/she provided. It is based on an adaptation of the Viterbi algorithm (Viterbi, 1967; Juang and Rabiner, 1991) which maximises the *joint conditional probability* for each subject i, i.e.,

$$\overset{*}{\boldsymbol{u}}_i = \arg\max_{\boldsymbol{u}} \mathrm{P}\left(\boldsymbol{U}_i = \boldsymbol{u} | \widetilde{\boldsymbol{Y}}_i = \widetilde{\boldsymbol{y}}_i, \widetilde{\boldsymbol{X}}_i = \widetilde{\boldsymbol{x}}_i\right),$$

through a forward-backward recursion. For further details see Bartolucci et al. (2013, 2014).

B.1 *Data Application: LOTox sequences*

In case of longitudinal toxicity data, path prediction $\overset{*}{\boldsymbol{u}}_i$ represents the sequence of LOTox states over time for subject i. Let us consider the four patients aged 15 years old with different observed toxic categories over cycles reported in Appendix C. The *LOTox sequences* for patients $i = \{A, B, C, D\}$ can be then obtained as

(i) the sequences of the most probable LOTox states at each cycle t (i.e., *local decoding*)

$$\overset{*}{\boldsymbol{u}}_A = (3,3,4,4,4,4), \quad \overset{*}{\boldsymbol{u}}_B = (3,3,3,2,2,2), \quad \overset{*}{\boldsymbol{u}}_C = (3,3,3,3,3,3), \quad \overset{*}{\boldsymbol{u}}_D = (3,1,1,1,1,1),$$

(ii) the sequences of the most likely LOTox states across cycles (i.e., *global decoding*)

$$\overset{*}{\boldsymbol{u}}_A = (3,3,4,4,4,4), \quad \overset{*}{\boldsymbol{u}}_B = (3,3,3,2,2,2), \quad \overset{*}{\boldsymbol{u}}_C = (3,3,3,3,3,3), \quad \overset{*}{\boldsymbol{u}}_D = (3,3,1,1,1,1).$$

Differences between (i) and (ii) are due to the different types of probabilities that are maximized, respectively posterior and joint conditional probabilities. The individual *LOTox sequence* allows to predict the LOTox state to which every patient belongs at a given occasion. However, it represent a summary of how the entire latent process evolves over time for a patient, as it only provides information about the most-likely condition without giving details about other states.

C Observed toxic categories over cycles for patients A–D

Table C.1 reports the observed toxic categories over cycles related to four 15-year patients named A, B, C and D, whose relative *longitudinal probability/relative risk profiles of LOTox* are shown in Figures 5 and 6, respectively .

Table C.1: Observed toxicity categories over cycles $t = 1, ..., 6$ for four random patients $i \in \{A, B, C, D\}$ aged 15 years old. Categories for *generic* toxicities (nausea, infection and oral mucositis) are $\{0 : none, 1 : mild, 2 : moderate, 3 : severe\}$ $(j = 1, 2, 3)$. Categories for *drug-specific* toxicities (cardiac toxicity, ototoxicity and neurological toxicity) are $\{0 : no, 1 : yes\}$ $(j = 4, 5, 6)$. For each patient i the complete response vector is $\tilde{\boldsymbol{y}}_i = \left(\boldsymbol{y}_i^{(1)}, \ldots, \boldsymbol{y}_i^{(1)}\right)$ where $\boldsymbol{y}_i^{(t)} = \left(y_{i1}^{(t)}, \ldots, y_{i6}^{(t)}\right)$.

Patient i	Cycle t	age_i	Naus $y_{i1}^{(t)}$	Inf $y_{i2}^{(t)}$	Oral $y_{i3}^{(t)}$	Car $y_{i4}^{(t)}$	Oto $y_{i5}^{(t)}$	Neur $y_{i6}^{(t)}$
A	1	15	3	0	1	0	0	0
	2		3	1	0	0	0	0
	3		3	3	0	0	0	0
	4		3	2	1	0	0	0
	5		3	0	2	0	0	0
	6		3	0	1	0	0	0
B	1	15	1	0	0	0	0	0
	2		1	0	0	0	0	0
	3		3	0	0	0	0	0
	4		1	0	0	0	1	0
	5		1	0	0	0	1	0
	6		1	0	0	0	1	0
C	1	15	2	0	0	0	0	0
	2		1	0	0	0	0	0
	3		1	0	0	0	0	0
	4		1	0	0	0	0	0
	5		1	0	0	0	0	0
	6		1	0	0	0	0	0
D	1	15	2	0	0	0	0	0
	2		2	0	2	0	0	0
	3		0	0	0	0	0	0
	4		0	0	0	0	0	0
	5		0	0	0	0	0	0
	6		0	0	0	0	0	0

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

75/2022 Rea, F.; Savaré, L; Franchi, M.; Corrao, G; Mancia, G
Adherence to Treatment by Initial Antihypertensive Mono and Combination Therapies

71/2022 Calabrò, D.; Lupo Pasini, M.; Ferro, N.; Perotto, S.
A deep learning approach for detection and localization of leaf anomalies

74/2022 Salvador, M.; Regazzoni, F.; Dede', L.; Quarteroni, A.
Fast and robust parameter estimation with uncertainty quantification for the cardiac function

70/2022 Andrini, D.; Balbi, V.; Bevilacqua, G.; Lucci, G.; Pozzi, G.; Riccobelli, D.
Mathematical modelling of axonal cortex contractility

69/2022 Franco, N.R; Manzoni, A.; Zunino, P.
Learning Operators with Mesh-Informed Neural Networks

68/2022 Orlando, G.; Benacchio, T.; Bonaventura, L.
An IMEX-DG solver for atmospheric dynamics simulations with adaptive mesh refinement

72/2022 Spreafico, M.; Ieva, F.; Arlati, F.; Capello, F.; Fatone, F.; Fedeli, F.; Genalti, G.; Anninga, J.; Gelderblom, H.; Fiocco, M.
Novel longitudinal multiple overall toxicity score to quantify adverse events experienced by patients during chemotherapy treatment: a retrospective analysis of the MRC BO06 trial in osteosarcoma

73/2022 Spreafico, M.; Gasperoni, F.; Barbati, G.; Ieva, F.; Scagnetto, A.; Zanier, L.; Iorio, A.; Sinagra, G.; Di Lenarda, A.
Adherence to disease-modifying therapy in patients hospitalized for HF: findings from a community-based study

67/2022 Alghamdi, M. M.; Boffi, D.; Bonizzoni, F.
A greedy MOR method for the tracking of eigensolutions to parametric elliptic PDEs

63/2022 Corti, M.; Antonietti, P.F.; Dede', L.; Quarteroni, A.
Numerical Modelling of the Brain Poromechanics by High-Order Discontinuous Galerkin Methods