Statistical modelling of time-varying covariates for survival data

Spreafico, M.

Chapter 5

A novel longitudinal method for quantifying multiple overall toxicity

This chapter has been published in *BMJ Open*, 11:e053456, 2021 as M. Spreafico, F. Ieva, F. Arlati, *et al.* "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" [190].

In cancer trials the relationship between chemotherapy dose and clinical efficacy outcomes is 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. Toxic Adverse Events (AEs), developed by patients through a chemotherapy cycle, affect subsequent exposure by delaying the next cycle or reducing its dosage, representing one of the principal reasons for treatment discontinuation [186]. The introduction of the Common Terminology Criteria for Adverse Events (CTCAE) [208] multimodality grading system greatly facilitated the standardized reporting of AEs and the comparison of outcomes between trials and institutions [204, 226]. According to CTCAE, AEs range in severity from minor, asymptomatic changes to life-threatening injuries or death [204]. Characterisation of toxicity is of interest to patients and clinicians engaged in shared decision making about a treatment strategy [198]. Toxicities 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 [112]. Incorporating time into analysis of toxicity is important for the comparison between different chemotherapy regimens or even multiple toxicities from the same regimen [199]. Therefore, it is crucial to provide an effective tool to assess the evolution of overall toxicity over chemotherapy treatment and to guide the therapy strategy.

Since patients might have different types and number of AEs, to summarize toxicity during treatment and investigate the true extent of toxic burden represent challenging problems in cancer research. Due to the complexity of longitudinal chemotherapy data, no standard method is available for summarising AEs data into a concise score of overall risk. Toxicity data are usually analysed in cancer prediction models by looking at the maximum toxicity over time (max-time) or maximum grade among events (max-grade) [204, 226, 199, 184, 140, 205]. Although both methods can summarise data over time, a lot of information are not used. The max-time method summarises longitudinal data

into a single AE profile by using the worst (maximum-severity) grade over treatment for each toxic event, without distinguish between isolated and repeated episodes. The max-grade method summarises all the toxic AEs through the maximum grade among all types of events, without discerning between single or multiple episodes. Other methods, i.e., weighted sums of individual toxic effects [205, 28, 172, 117, 35], have also been proposed to consider longer-lasting lower-grade chronic toxicities, which may have impact on patient's quality of life. However, these approaches do not provide information about AEs timing or severity at a given cycle during treatment. The inclusion of time-related information could provide intuitions on AEs and their evolution over time [198], giving new insights in cancer treatment.

In this framework, alternative methods of longitudinal toxic event evaluation have been proposed [198, 205, 114, 84, 200] but none of them is focused on analysing the evolution of high overall toxicity over treatment using a cycle-by-cycle approach. To quantify risk for each patient including a time-dimension, in this chapter a new longitudinal Multiple Overall Toxicity (MOTox) score is proposed. At each cycle, this score summarises multiple CTCAE-graded AEs, and describe the overall toxic status along with the most severe risk event. The evolution of high MOTox scores over cycles is then studied using logistic regression models to predict high overall toxicity at the end of the cycle using personalized characteristics, achieved chemotherapy dose, previous toxicities, biochemical and haematological factors over time. To illustrate the use of the longitudinal MOTox procedure to quantify how chemotherapy-induced toxicities may evolve in cancer patients, a retrospective analysis was conducted on MRC BO06/EORTC 80931 Randomized Controlled Trial (RCT) for the treatment of osteosarcoma [119]. As mentioned in Chapter 4, patients were treated with cisplatin (CDDP) and doxorubicin (DOX), two cytotoxic drugs commonly used in the treatment of various types of human cancers. Both DOX and CDDP are characterized by various toxic AEs: apart from nausea, specific renal and neurotoxicity [68, 9] for CDDP or cardiotoxicity [228, 227] for DOX. Longitudinal MOTox scores over therapy were computed considering non-haematological toxicity. Demographics, treatment-related and biochemical characteristics were used to examine high overall toxicity over cycles. Provided that longitudinal CTCAE-graded toxicity data are available, the novel MOTox scores can be applied to analyse data from any cancer treatment.

The rest of this chapter is organized as follows. In Section 5.1 non-haematological toxicity data in BO06 trial are described. Longitudinal MOTox scores and statistical methodologies are introduced in Section 5.2. Results are presented in Section 5.3. Section 5.4 ends with a discussion of strengths and limitations of the current approach, identifying some developments for future research.

5.1. BO06 data

Data from the MRC BO06/EORTC 80931 RCT for the treatment of osteosarcoma [119] were analysed. Specifically, we focused on the final cohort of 377 patients who completed all six cycles of chemotherapy within 180 days after randomization without abnormal

dosages (i.e., +25% higher than planned). This cohort is the same as the one analysed in Chapter 4, under TVCM analysis (see Figure 4.3 in Chapter 4). Patient characteristics at randomization have been reported in Table 4.1 in Chapter 4.

5.1.1. Treatment-related factors

As reported in Section 4.2.1 in Chapter 4, BO06 RCT protocol established that treatment-related factors (administered dose of chemotherapy, cycles delays, haematological and biochemical parameters, chemotherapy-induced toxicity and histological response to pre-operative chemotherapy) were collected prospectively at each cycle of chemotherapy [119].

Laboratory tests were performed over cycles in order to monitor patient's health status and the development of toxicities or adverse events. Specifically, levels of alkaline phosphatase, renal clearance, lactate dehydrogenase, calcium and magnesium were measured at the beginning of each cycle (i.e., before the drugs administration) according to local practice. Blood counts (white blood cells, neutrophils, platelets) were obtained before each cycle and at the expected nadir of the course (day 10 of the cycle in *Reg-C*, day 8 in *Reg-DI*). A summary of the biochemical and haematological values measured for the selected cohort over the entire dataset is shown in Table 5.1.

Delays or chemotherapy dose reductions during treatment were possible in case of toxicity. Specifically, non-haematological chemotherapy-induced toxicity related to nausea/vomiting (*naus*), infection (*inf*), oral mucositis (*oral*, i.e., inflammation of the mucosae of the gastrointestinal tract, especially the oral ones), cardiac toxicity (*car*, i.e., heart dysfunctions), ototoxicity (*oto*, i.e., hearing loss) and neurological toxicity (*neur*) were registered at each cycle and graded according to the Common Terminology Criteria for Adverse Events Version 3 (CTCAE v3.0) [208], with grades ranging from 0 (none) to 4 (life-threatening) (see Table 5.2). Grades of chemotherapy-induced non-haematological toxicity over cycles recorded for the selected cohort are reported in Figure 5.1. 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:

Table 5.1. Descriptive of biochemical and haematological values over the entire dataset.

Biomarkers*	**Mean (s.d.)**	**Median (IQR)**	**Min/Max**
White Blood Count $[\times 10^9/L]$	7.36 (8.25)	5.00 (3.10; 8.20)	0.10/117
Neutrophils $[\times 10^9/L]$	4.74 (6.93)	2.60 (1.12; 5.30)	0/83.38
Platelets $[\times 10^9/L]$	219.8 (157.5)	190 (99; 311)	2/999
Renal Clearance $[ml/min/1.73m^2]$	112.3 (34.9)	110 (90; 132)	8/396
Alkaline Phosphatase $[IU/L]$	238.5 (279.1)	162.5 (98.0; 267.2)	14/3680
Lactate Dehydrogenase $[IU/L]$	447.0 (264.2)	394.0 (298.8; 531.0)	4/4310
Calcium $[mmol/l]$	2.34 (0.36)	2.35 (2.25; 2.45)	0.21/9.70
Magnesium $[mmol/l]$	0.71 (0.24)	0.69 (0.57; 0.80)	0.07/3.06

* Haematological and biochemical laboratory tests were usually performed before each cycle of chemotherapy; for blood count also at the expected nadir of the course, that is day 10 of the cycle in *Reg-C* or day 8 in *Reg-DI*.

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).

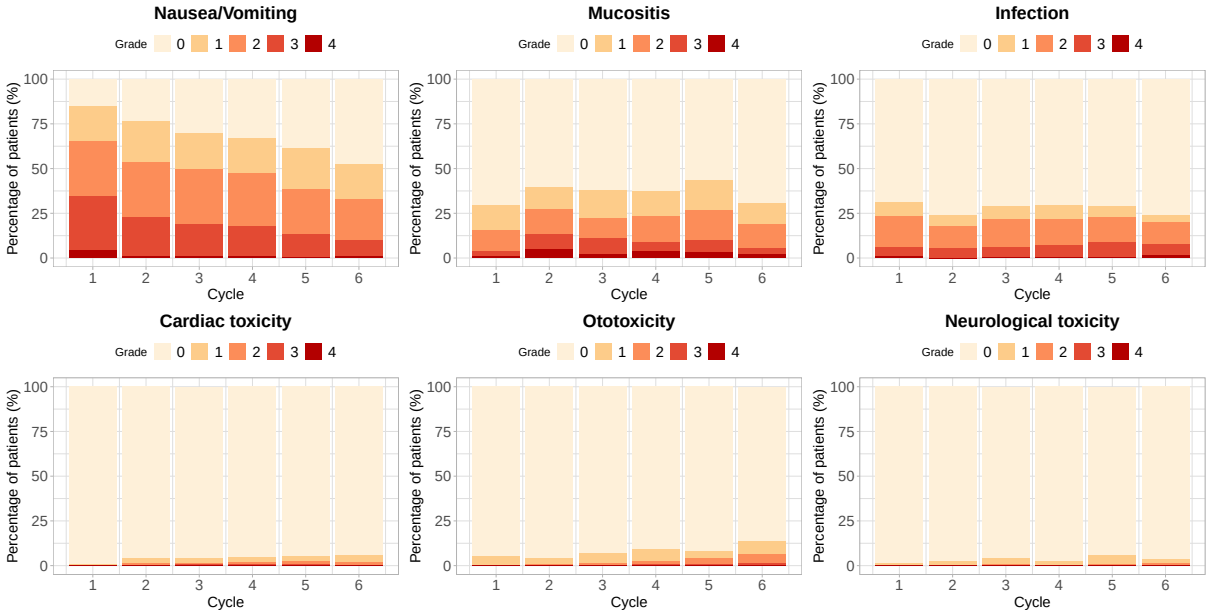

Figure 5.1. Bar-plots of chemotherapy-induced toxicity CTCAE grades over cycles (wheat: 0; light-orange: 1; orange: 2; red: 3; dark-red: 4). Each panel refers to a different type of toxicity: nausea/vomiting [top-left], mucositis [top-centre], infection [top-right], cardiac toxicity [bottom-left], ototoxicity [bottom-centre] and neurological toxicity [bottom-right].

Table 5.2. Toxicity coding based on Common Terminology Criteria for Adverse Events (CTCAE) v3.0 by [208] 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 or 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

PVC = premature ventricular contraction

5.2. Statistical Methodologies

5.2.1. Longitudinal Multiple Overall Toxicity (MOTox) scores and outcomes

The longitudinal chemotherapy-induced Multiple Overall Toxicity (MOTox) score is introduced. Let $\mathcal{T}$ be the set of different toxicity categories. Let k be the cycle index (which takes value $k \in \{1, ..., 6\}$) and $tox_{ij,k}$ be the j-th toxicity level for the i-th patient at the k-th cycle with value from 0 to 4. The chemotherapy-induced MOTox score for the i-th patient at cycle k is defined as:

$$\begin{aligned} MOTox_{i,k} &= \text{average toxic level}_{i,k} + \text{worst grade}_{i,k} \\ &= \frac{1}{|\mathcal{T}|} \sum_{j \in \mathcal{T}} tox_{ij,k} + \max_{j \in T} (tox_{ij,k}) \in [0, 8] \end{aligned} \tag{5.1}$$

where the *average toxic* level is hence the arithmetic mean of the grades related to all the toxic AEs registered for the patient at cycle k, and the *worst grade* is the maximum CTCAE-grade among all the toxic AEs experienced by the patient at the cycle under analysis.

The MOTox score is a *cycle-dependent longitudinal mean-max index* that quantifies the multiple types of Adverse Events (AEs) experienced by patient i during cycle k. This choice was made to include the cycle-time component in the analysis and to take into account that (i) multiple lower-grade chronic toxicities may have impact on patient's quality of life and (ii) huge level in one specific toxicity can cause severe effects and permanent consequences for the patient. MOTox score can detect differences in health status among patients, providing more information with respect to traditional methods.

This novel score only requires that the different types of toxicity necessary for the computation, are recorded according to the CTCAE grading system. In this way, this definition can be applied to different groups of CTCAE-graded toxicities and applied to any cancer treatment.

The median value of MOTox scores over all the patients in all the cycles, computed as

$$\tau = \underset{i,k}{median} \, (MOTox_{i,k}),$$

is defined as *global median MOTox value*. It is used as a threshold to define a longitudinal binary score for *high* (or *low*) overall toxicity, named *longitudinal high-MOTox score* :

$$high\text{-}MOTox_{i,k} = \begin{cases} 1 & \text{if } MOTox_{i,k} > \tau \\ 0 & \text{otherwise.} \end{cases} \tag{5.2}$$

that indicates if patient experienced high MOTox with respect to the global median MOTox value τ at cycle k ($high\text{-}MOTox_{i,k} = 1$) or not ($high\text{-}MOTox_{i,k} = 0$).

MOTox and high-MOTox scores represent new indices to measure patients' overall toxicity related to multiple types of AEs. Binary high-MOTox scores over cycles represent the clinical endpoints used as outcome measures for high overall toxicity over treatment.

Interpretation of longitudinal MOTox scores

As a *mean-max* index of CTCAE-graded toxicity levels ranging from 0 to 4 each, the MOTox score $MOTox_{i,k}$ in Equation (5.1) – as well as the global median MOTox value τ – ranges from 0 to 8. A MOTox score equal to 0 reflects a patient i who did not experience any kind of toxicity for the cycle k under analysis, i.e., a patient with all the toxicities equal to CTCAE-grade 0 at cycle k. Conversely, a MOTox score equal to 8 represents a subject i who experienced the highest level of toxicity burden for each type of toxic AE for the cycle k under analysis, i.e., a subject i with all toxicities equal to CTCAE-grade 4 at cycle k.

The global median MOTox τ represents the median value of MOTox scores computed over all the patients in all the cycles, i.e., the median overall toxicity related to multiple AEs experienced by all the patients over the entire chemotherapy treatment. If required by the needs of the study, different median MOTox values breakdown by arms/regimens represent an easily-applicable alternative to a global τ in order to study other treatment regimen/cancer types where different arms/regimens are characterized by significantly different overall toxicity burden.

Binary variable $high-MOTox_{i,k}$ in Equation (5.2) indicates if patient i experienced high MOTox with respect to the global median MOTox value τ at cycle k, i.e., it distinguishes patients with *low* ($high\text{-}MOTox_{i,k} = 0$) or *high* ($high\text{-}MOTox_{i,k} = 1$) overall toxicity burden over treatment.

5.2.2. Statistical analysis

A retrospective analysis to examine prognostic factors for binary high-MOTox scores over cycles was conducted. Baseline and treatment-related characteristics were examined. In particular, chemotherapy dose given at cycle k was analysed as percentage of *achieved chemotherapy dose up to cycle k* for each patient i, defined as the percentage of the cumulative drugs administrated up to cycle k divided by the cumulative drugs planned up to k:

$$\begin{aligned} \mathrm{p}\delta_{i,k} &= \frac{\text{cumulative drugs administrative up to cycle } k}{\text{cumulative drugs planned up to cycle } k} \cdot 100\% \\ &= \frac{\sum_{c=1}^{k} \left(DOX_{i,c} + CDDp_{i,c}\right) / sa_{i,c}\,[mg/m^2]}{175\,[mg/m^2] \cdot k} \cdot 100\% \end{aligned} \tag{5.3}$$

where $k \in \{1, ..., 6\}$ is the cycle index, sa is patient's surface area in m^2, DOX and $CDDP$ are the administrated mg of doxorubicin and cisplatin, respectively. A two-sided significance level of 5% was adopted. R software [161] was used for the analyses.

Data on non-haematological toxicity were not available for 1.25% of measurements, which were considered as CTCAE 0-grade according to clinical indication. For treatment-related missing values (i.e., histologic response, biochemical and haematological markers), missing values were imputed using multiple imputations by chained equations algorithm [209].

At each cycle, the impact of factors on high overall toxicity (binary *high-MOTox*) was examined using multivariable logistic regression models and expressed by odds ratios (OR) [137]. An $OR > 1.0$ indicates a greater risk of achieving a high overall toxicity in case of a 1-unit increase for numerical characteristics or compared to the baseline category for categorical ones. Covariates with more than 15% of missing values in the original data were not included in the multivariable models. A stepwise backward selection procedure was applied to select the best set of covariates at each cycle based on Akaike Information Criterion (AIC). Variance Inflation Factor (VIF) was also used to remove non-significant and highly collinear covariates. Predictive capacities of models were assessed by sensitivity and specificity metrics and Area Under the receiver operating characteristic Curve (AUC) [55].

5.3. Results

5.3.1. Non-haematological longitudinal Overall Toxicity scores

For each patient, non-haematological chemotherapy-induced toxicity related to nausea, mucositis, infection, neurological toxicity, cardiac toxicity, and ototoxicity, i.e., set $\mathcal{T} = \{naus, oral, inf, car, oto, neur\}$, were considered to compute the longitudinal MOTox scores over cyclesfor each patient, according to Equation (5.1) and Equation (5.2). MOTox scores (Figure 5.2 – left panel) ranged between 0 and 6 and the mean values (blue points) decreased over cycles from 2.626 (cycle 1) to 1.953 (cycle 6). The global median MOTox value τ, i.e., the median value of overall toxicity over all the patients in all the cycles, was 2.333 (dashed red line). An example of longitudinal MOTox scores over cycles for five random patients from the study cohort is shown in Figure 5.3. The global mean MOTox value τ is reported as solid black line. Different evolution patterns of longitudinal MOTox score over cycles are presented: increasing pattern (*orange:* patient A), decreasing pattern (*light blue:* patient B), isolated severe status (*violet:* patient C), low-values (*blue:* patient D) and high-values (*red:* patient E) over cycles.

To evaluate which regimens is characterized by high toxicity over cycles, Table 5.3 reports the means of MOTox scores at each cycle for patients allocated in *Reg-DI* and *Reg-C*, and respectively. In cycles 2-3, mean overall toxicity for patients in *Reg-DI* was higher than for those in *Reg-C* ($p<0.05$), whereas from cycle 4 the difference was not statistically significant. Figure 5.4 shows the mean values of each non-haematological toxicity along with 95% Bonferroni's confidence intervals over cycles, stratified by regimens. Each panel refers to a different type of toxicity: nausea/vomiting, mucositis, infection, cardiac toxicity, ototoxicity and neurological toxicity. The biggest contribution to the difference in the mean MOTox scores by regimes was given by mucositis, significantly higher in *Reg-DI* than in *Reg-C* at cycles 2 and 3.

Median MOTox values by arms (*Reg-DI* or *Reg-C*) τ_{DI} and τ_C were both equal to 2.333. Therefore, the global MOTox median value $\tau = 2.333$ was then used to compute the longitudinal dichotomous high-MOTox scores over cycles. Right panel in Figure 5.2 shows

the percentages of patients with high-MOTox, which decrease from 57.8% (218/377) at cycle 1 to 36.6% (138/377) at cycle 6. Association between chemotherapy regimens and high overall toxicity at cycles 2–3 ($p<0.05$) was found, supporting results shown in Table 5.3. At each cycle, high overall toxicity was strongly associated with *low/high* MOTox at previous cycles.

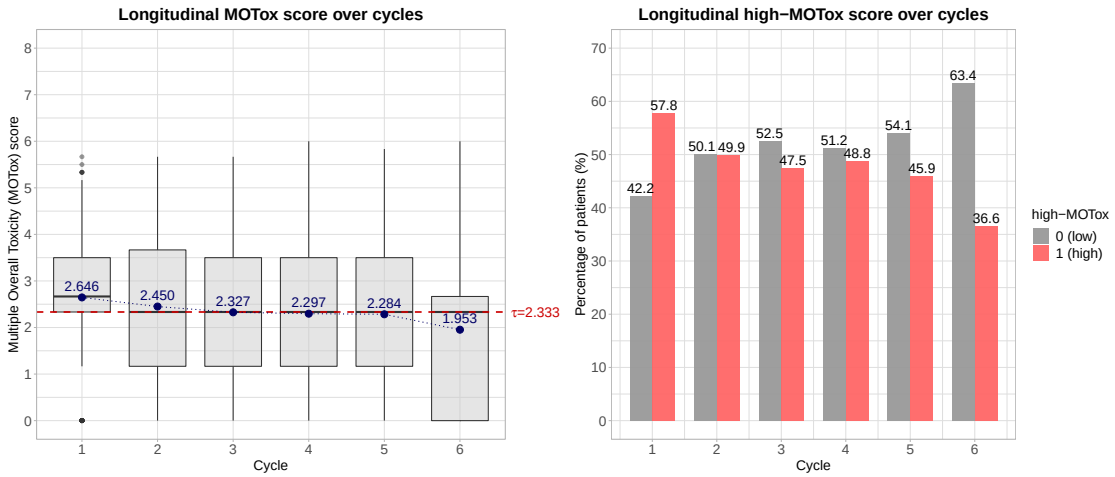

Figure 5.2. Left panel: Boxplots of longitudinal MOTox scores over cycles. Blue points refers to the mean MOTox values per cycle. Dashed red line refers to the global median MOTox value =2.333. Right panel: Bar-plots of longitudinal high-MOTox scores over cycles (grey: 0 or low; magenta: 1 or high).

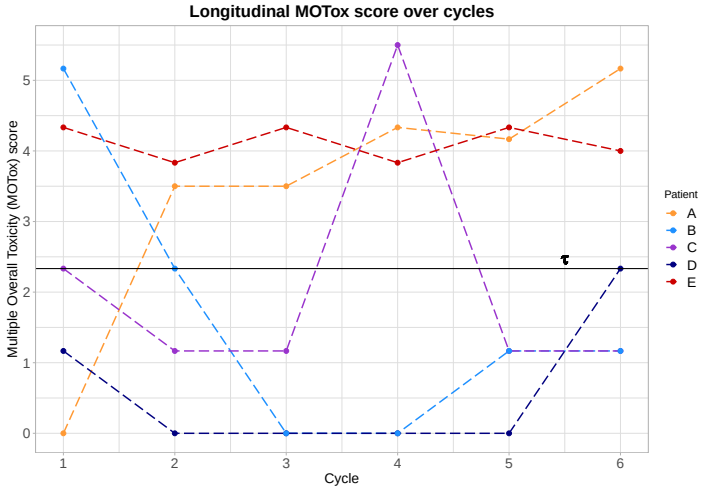

Figure 5.3. Example of evolution of longitudinal Multiple Overall Toxicity (MOTox) scores over cycles for five patients from the study cohort. Solid black line refers to the global median MOTox value $\tau = 2.333$.

Table 5.3. Overall toxicity differences between Dose-Intense (DI) and Conventional (C) regimens. $\overline{MOTox}^k_{DI}$ and $\overline{MOTox}^k_{C}$ are the means of MOTox scores at cycle k for patients allocated in *Reg-DI* and *Reg-C*, respectively.

	Cycle 1	**Cycle 2**	**Cycle 3**	**Cycle 4**	**Cycle 5**	**Cycle 6**
$\overline{MOTox}^k_{DI}$	2.552	2.653	2.488	2.240	2.261	1.920
$\overline{MOTox}^k_{C}$	2.782	2.229	2.150	2.359	2.309	1.989
p-value of test	0.045	0.003	0.018	0.437	0.737	0.657

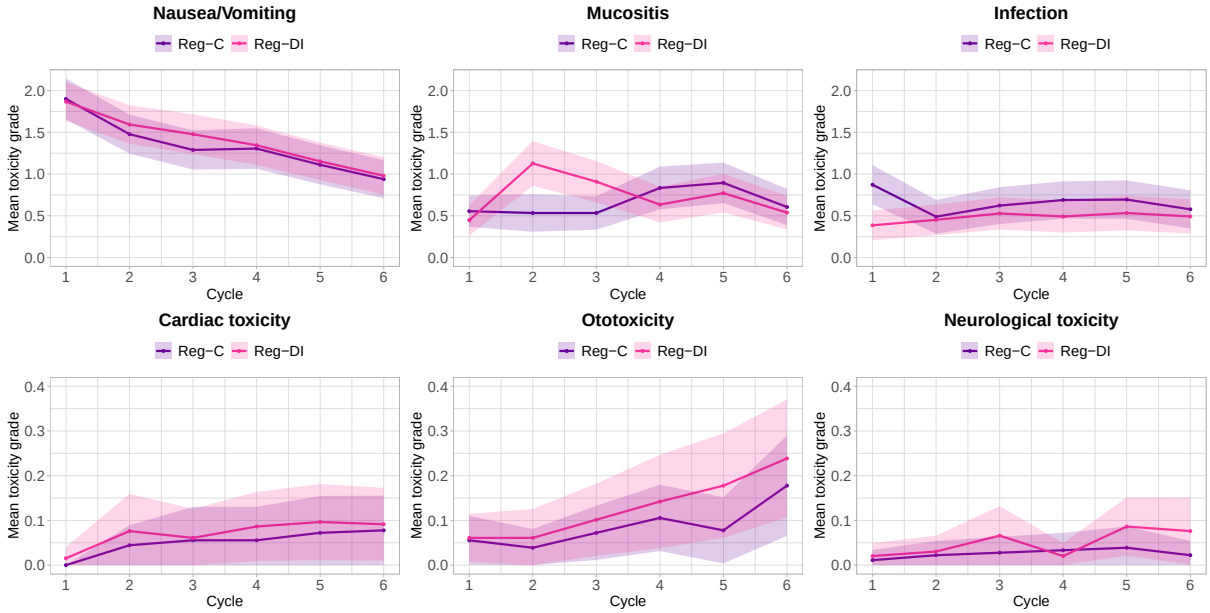

Figure 5.4. Mean value of chemotherapy-induced toxicity during cycles along with 95% Bonferroni confidence intervals, stratified by the regimens (purple: *Reg-C*; pink: *Reg-C*). Each panel refers to a different type of toxicity: nausea/vomiting [top-left], mucositis [top-centre], infection [top-right], cardiac toxicity [bottom-left], ototoxicity [bottom-centre] and neurological toxicity [bottom-right].

5.3.2. Multivariable logistic regression models for high overall toxicity over cycles

The evolution of longitudinal binary *high-MOTox* score over cycles defined was analysed through multivariable logistic regression models, using a cycle-by-cycle approach. Starting from the second cycle, each logistic regression modelled the binary dependent variable *high-MOTox* at the of the cycle in terms of patient's characteristics and previous toxicity levels. Baseline and treatment-related information with less than 15% of missing values in the original dataset were considered as possible prognostic factors for toxicity. In particular, among haematological and biochemical factors, measurements of white blood count (WBC), neutrophils (N), platelets (PLT), alkaline phosphatase (ALP) and calcium (Ca) were considered before the beginning of each cycle (i.e., the administration of the course). Only WBC values were considered at the planned nadir of each cycle, due to the high percentage of missing values (>15%) for other blood counts. Due to the skewed nature of biomarkers distributions, haematological and biochemical factors were included in the models as difference between the logarithmic measure and the logarithmic value measured at randomization. Neutrophils–Platelets Score (NPS), a three-level systemic inflammation-based score (*good*: N$\leq 7.5 \times 10^9/L$ and PLT$\leq 400 \times 10^9/L$; *intermediate*: N$> 7.5\times10^9/L$ or PLT$> 400\times10^9/L$; *poor*: N$> 7.5\times10^9/L$ and PLT$> 400\times10^9/L$) [127], and Neutrophils-White blood count Ratio (NWR, i.e., the neutrophils count dived by the white blood cell count) were also considered. For each model, multicollinear variables with VIF greater than 5 were removed. Then, stepwise backward procedures were used to select covariates according to AIC. The selected models were fitted on the whole dataset.

Table 5.4 shows estimated Odds Ratios (ORs) along with 95% Confidence Intervals (CIs)

and overall performances (i.e., specificity, sensitivity and AUC) of each logistic regression model. All the models have similar overall performances: sensitivity and specificity values ranged between 0.66 and 0.77; AUCs were between 0.72 and 0.79. No sex effect was found. In cycle 2 and 3, higher percentage of achieved chemotherapy dose is associated to the risk of high toxicity, especially for patient in *Reg-DI* (cycle 2). Haematological factors were selected in each model. Both PLT before the administration of the course and WBC at nadir had a protective role on the risk of having high overall toxicity ($OR < 1$). In particular, an increase in the dynamic difference between the logarithmic levels decreased the risk of high toxicity. Patients with previous *high-MOTox* had higher risk to experience again high overall toxicity with respect to patients with previous *high-MOTox* ($OR > 1$), showing that *high-MOTox* conditions during previous cycles were risk factors for the occurrence of *high-MOTox* at the current cycle. In particular, toxicity information related to different previous cycles were selected and statistically significant in the final models, meaning that patient's global history – and not only the last condition – had impact on his/her current *low/high* overall toxicity burden.

The performed analyses were finally used to develop a demo webapp availbale at `http://osteowebapp.prod.s3-website.eu-central-1.amazonaws.com/`. The demo shows how the multivariable models developed to predict high overall toxicity index at each cycle could be used as a support tool for clinical decision making. The webapp is presented in Appendix B.1.

5.4. Final remarks

Due to the presence of multiple types of Adverse Events (AEs) with different levels of toxicity burden, to study the overall toxicity progression during chemotherapy is a difficult problem in cancer research. The development of statistical methods able to deal with the complexity of longitudinal chemotherapy data and to provide a methodology to use the information of AEs data into a score of overall risk is necessary and of clinical relevance.

This chapter explored the evolution of chemotherapy-induced toxicity over treatment in patients with osteosarcoma. First, a novel approach to analyse longitudinal chemotherapy data was discussed, the cycle-dependent longitudinal mean-max Multiple Overall Toxicity (MOTox) score over therapy. Starting from recorded CTCAE grades, the MOTox score summarised the occurrence of repeated AEs allowing to (i) describe the overall toxicity burden, (ii) consider the most severe collateral effect, and (iii) incorporate the time-component of treatment cycles. Results showed that the inclusion of worst-graded events, multiple lower-grade chronic toxicities, and time-dimension related to chemotherapy cycles allowed to consider different evolutions of overall toxic levels over treatment. This approach investigates in more details the effect of AEs on patients' life compared to traditional methods (i.e., max-grade or max-time). The cycle-by-cycle longitudinal evolution of high overall toxicity was analysed using multivariable logistic regression models to predict binary high-MOTox at the end of the cycle in terms of previous toxicity levels

Table 5.4. Multivariable logistic regression model for each cycle $k \in \{2, 3, 4, 5, 6\}$ with estimated Odds Ratios (ORs) along with their 95% Confidence Intervals (CIs).

	Cycle 2		**Cycle 3**		**Cycle 4**		**Cycle 5**		**Cycle 6**	
Covariates	**ORs**	**95% CIs**	**ORs**	**95% CIs**	**ORs**	**95% CIs**	**ORs**	**95% CIs**	**ORs**	**95% CIs**
Baseline										
Sex (male)					1.458	[0.912; 2.33]	1.548	[0.967; 2.478]		
Regimen (Reg-DI)	2.379	[1.455; 3.889]								
Treatment-related factors										
Achieved dose (%)	1.112	[1.042; 1.187]	1.056	[1.008; 1.106]						
WBC							1.664	[0.978; 2.831]		
WBC at nadir	0.778	[0.615; 0.983]			0.701	[0.569; 0.864]	0.637	[0.499; 0.813]	0.748	[0.589; 0.950]
PLT			0.535	[0.375; 0.765]	0.625	[0.392; 0.996]			0.629	[0.424; 0.932]
NWR					0.367	[0.094; 1.436]				
Previous toxicities										
High MOTox ($k = 1$)	4.439	[2.788; 7.070]	1.561	[0.968; 2.516]	1.522	[0.953; 2.430]				
High MOTox ($k = 2$)			4.429	[2.666; 6.772]	1.569	[0.972; 2.532]			1.743	[1.044; 2.910]
High MOTox ($k = 3$)					2.701	[1.696; 4.304]	2.639	[1.664; 4.186]	1.580	[0.938; 2.661]
High MOTox ($k = 4$)							3.718	[2.346; 5.893]	2.542	[1.523; 4.244]
High MOTox ($k = 5$)									3.341	[2.001; 5.580]
Sensitivity	0.681		0.704		0.674		0.699		0.717	
Specificity	0.667		0.661		0.715		0.701		0.766	
AUC [95% CI]	0.733	[0.683; 0.784]	0.743	[0.694; 0.793]	0.728	[0.677; 0.780]	0.756	[0.707; 0.805]	0.787	[0.737; 0.837]

WBC = white blood count; PLT = platelets; NWR = neutrophils-white blood count ratio; MOTox = multiple overall toxicity.
When not specified, haematological factors were computed before the administration of the course.
WBC and PLT are included in the models as difference between the current logarithmic measure and the logarithmic value measured at randomization.

and patient's characteristics. At each cycle, previous toxicity levels were selected: high-MOTox during previous cycles were risk factors for the occurrence of high-MOTox at the next cycle. The highest impact on the risk was observed for the last available toxic condition. Patient's history of toxicity played a fundamental role in the risk of high overall toxicity burden during cycles and, consequently, on patient's health status during the therapy. This analysis also suggested that the Conventional Regimen might be preferred to the Dose-Intense in terms of life conditions during the first half of the therapy (i.e., up to the third cycle): mean MOTox values in *Reg-DI* were statistically higher than in *Reg-C* during cycles 2-3 and *Reg-DI* was a risk factor for the occurrence of high-MOTox at the end of the second cycle. However, in terms of survival, a beneficial effect of low level (grade 1-2) platelet and nausea/vomiting toxicity and more severe (grade 3-4) mucositis on survival in osteosarcomas was previously shown [172]. Appraisal of the experienced toxicity against survival encourages the genetic exploration of the individual sensitivity to both adverse effects as well as the sensitivity of the tumour to chemotherapy.

Different statistical and machine learning methods for high/low binary classification were considered, among others support vector machines or ensemble methods (e.g., random forests or XGBoost). More complex methods showed no significant improvements in terms of predictive performances with respect to logistic regression models. Therefore, the choice was driven by the clinical interpretability offered by the cycle-by-cycle logistic regression approach.

The presented MOTox and binary high-MOTox scores can be used to (i) describe patient's response to therapy over cycles, (ii) predict the upcoming overall toxicity level given patient's history and (iii) support clinical decisions, trying to reduce the impact of therapies in terms of toxic AEs. Provided that longitudinal CTCAE-graded toxicity data are available from drug administrations, the new approach is a flexible procedure that can be adapted and applied to other cancer studies. The possible generalization to many different settings, added to a cooperation with medical staff, could lead to improvements in the definition of useful tools for health care assessment and treatment planning. As shown in the demo webapp presented in Appendix B.1, once validated, the multivariable models could be used to set up a support tool to predict high overall toxicity at the end of each cycle. This would allow to monitor patient's toxic burden during treatment and to inform dose reductions or dose delays to make treatment more tolerable.

This retrospective exploratory analysis comes with some challenges and limitations. Although the toxicity data were recorded using the standardised CTCAE grading system, heterogeneity in assessing non-haematological toxicity is present in the data, especially considering that MRC BO06 RCT is limited to a young population with a rare tumour. The analysis was performed on a single RCT in osteosarcoma, where only non-haematological toxicities were recorded according to CTCAE. Other factors such as nephrotoxicity, lymphocytes count, tumour size, CTCAE-graded haematological toxicities or quality-of-life were not collected. Although over the last twenty years the main chemotherapy protocol has been used, some aspects of osteosarcoma treatment and supportive care have changed from current measures [140], such as the prophylaxis of nausea and vomiting. Such changes are not always easily identifiable and are difficult to account

for in retrospective analyses [140]. Finally, this work focused on the quantification and evolution of overall toxicity in patients who completed all 6 cycles of chemotherapy treatment. This choice was due to a specific research question. However, this may lead to bias selection due to the exclusion of patients who may have had high toxicity levels as the reason for treatment discontinuation. Since the definition of the MOTox score is general, it can be computed also for those excluded patients, but alternative statistical methods to multivariable logistic models must be developed to also take into account therapy discontinuation. In fact, subsequent analyses should include patients who have discontinued treatment to better understand if MOTox is a potential measure of treatment tolerability and if may be associated with treatment discontinuation.

External validation is needed to evaluate the application of the novel score in order to guide prospective treatment decisions in clinical practice, both for osteosarcoma and for other types of cancer. On one hand, integration with data from other osteosarcoma studies could help in further investigating the performance of the models and in examining whether the analysis should be integrated with more information on toxicity or other potential predictors. On the other hand, to apply the developed procedure to the clinical decision-making process in different treatment regimen/cancer types, the multivariate methods need to be tailored according to each specific study.

This work opens doors to many further developments, both in the field of statistical methodology and in cancer research. From a clinical point of view, the interest may lie in identifying patients with extremely high or extremely low overall toxicity with respect to intermediate toxic conditions. As consequences multiple MOTox categories related to different levels of overall toxicity (e.g., extremely-high/high/intermediate/low/extremely-low MOTox) are defined. Thresholds to establish the MOTox ranges for the different categories needs to be created. This is not a trivial task which requires a proper external validation. Furthermore, the comparison between the MOTox score and Quality-Of-Life (QOL) represents a challenging area of investigation in clinical research. MRC BO06 trial did not collect QOL data, but it would be of interest to evaluate MOTox in the context of rigorously collected Heath-related QOL (HrQOL) or Patient-Reported Outcome (PRO) data to investigate the role of the developed tool in better understanding treatment tolerability. Therefore, future analyses must focus on data where QOL is properly measured and reported. From a statistical point of view, (i) the CTCAE-grades of toxicity could be analysed in greater depth through an appropriate longitudinal approach to categorical data, and (ii) an adequate modelling of the intricate mechanism between toxicity, chemotherapy dose and survival, which is still lacking in the medical literature, represents a major challenge of clinical relevance. Due to the complexity of the problem, both aspects are not straightforward and ask for the developments of new methodologies, as we will see in Chapters 6 and Chapter 7, respectively.

In summary, this chapter introduced a novel longitudinal method to explore and quantify AEs experienced by patients during cancer treatment. Preliminary results from the retrospective analysis of MRC BO06 RCT showed that longitudinal methods should be considered in future analyses of cancer trials, since they could lead to new insights into chemotherapy-induced toxicity compared to traditional approaches. For this reason, in

the next chapter we develop a new taxonomy based on latent Markov models [22] and compositional data [6] to model the evolution of latent overall toxicity burden on the basis of nominal CTCAE-grades observed over cycles.

B. Appendix to Chapter 5

B.1. Demo OsteoWebApp

The demo *OsteoWebApp* displays how the novel MOTox approach can become a useful tool for health care assessment and cancer treatment planning. In particular, it shows how the multivariable models to predict high overall toxicity at the end of each cycle developed in Chapter 5 could be used as a support tool for clinical decision making. It is available at: `http://osteowebapp.prod.s3-website.eu-central-1.amazonaws.com/`.

The application is implemented through Amazon Web Services (`https://aws.amazon.com/it/`) tools and executes the R [161] code related to the models in Table 5.4. Thanks to the intuitive interface, the webapp is easy to use and complete in the information it provides.

An example of the user interface, showing the inputs and results for model related to cycle 2, is reported in Figure 5.5. The top bar shows the cycle of chemotherapy of interest. The main form asks a series of information, depending on the variables selected for each cycle. The *"Predict Toxicity Index"* button in blue allows to get the results of the prediction, which are provided in terms of probability of develop a high overall toxicity level. Sensitivity and specificity of each model are also reported. Example in Figure 5.5 shows that a patient in *Reg-DI* with *high*-MOTox at cycle 1, a cumulative administrated dose of 350 mg/m^2 (which corresponds to a 100% of achieved dose), WBC values of 7.65 $[\times 10^9/L]$ at randomization and of 3.9 $[\times 10^9/L]$ at nadir has 73.5% probability to be in *high-MOTox* status at the end of cycle 2.

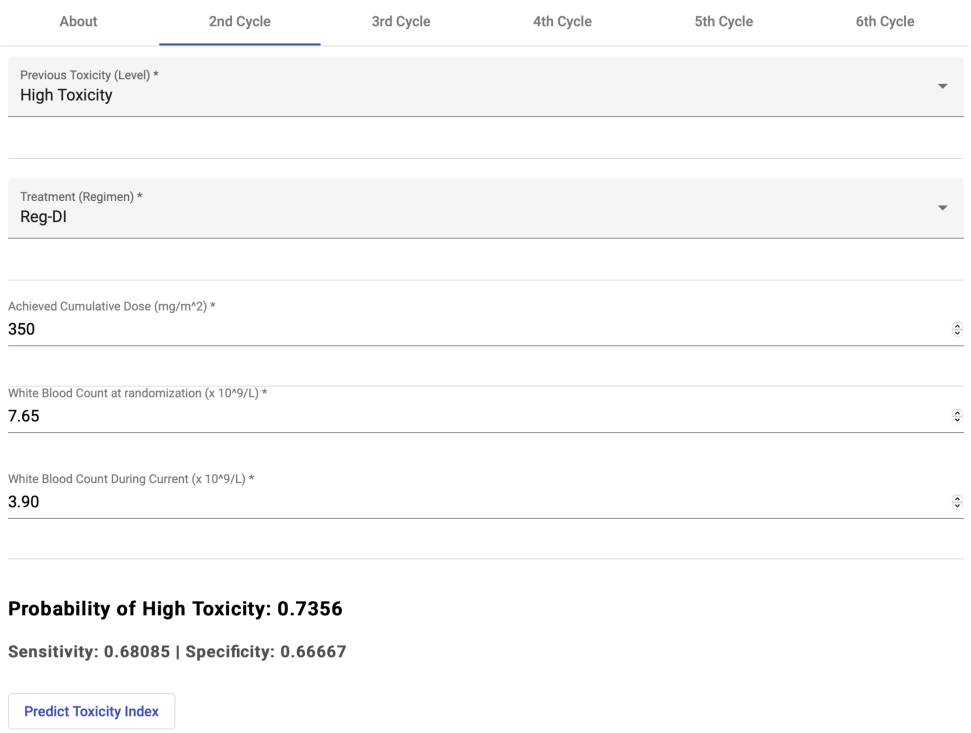

Figure 5.5. Example of user interface for *OsteoWebApp*.

CHAPTER 5