

BMJ Open 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

Marta Spreafico ^{1,2,3} Francesca Ieva ^{1,3,4} Francesca Arlati,¹ Federico Capello,¹ Federico Fatone,¹ Filippo Fedeli ¹ Gianmarco Genalti,¹ Jakob Anninga,⁵ Hans Gelderblom,⁶ Marta Fiocco^{2,7,8}

To cite: Spreafico M, Ieva F, Arlati F, *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. *BMJ Open* 2021;**11**:e053456. doi:10.1136/bmjopen-2021-053456

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-053456>).

Received 13 May 2021

Accepted 01 November 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Ms Marta Spreafico;
marta.spreafico@polimi.it

ABSTRACT

Objectives This study aims at exploring and quantifying multiple types of adverse events (AEs) experienced by patients during cancer treatment. A novel longitudinal score to evaluate the Multiple Overall Toxicity (MOTox) burden is proposed. The MOTox approach investigates the personalised evolution of high overall toxicity (high-MOTox) during the treatment.

Design Retrospective analysis of the MRC-BO06/EORTC-80931 randomised controlled trial for osteosarcoma.

Setting International multicentre population-based study.

Participants A total of 377 patients with resectable high-grade osteosarcoma, who completed treatment within 180 days after randomisation without abnormal dosages (+25% higher than planned).

Interventions Patients were randomised to six cycles of conventional versus dose-intense regimens of doxorubicin and cisplatin. Non-haematological toxicity data were collected prospectively and graded according to the Common Terminology Criteria for Adverse Events (CTCAE).

Main outcome measures The MOTox score described the overall toxicity burden in terms of multiple toxic AEs, maximum-severity episode and cycle time-dimension. Evolution of high-MOTox was assessed through multivariable models, that investigated the impact of personalised characteristics (eg, achieved chemotherapy dose, previous AEs or biochemical factors) cycle-by-cycle.

Results A cycle-by-cycle analysis identifies different evolutions of MOTox levels during treatment, detecting differences in patients' health. Mean MOTox values and percentages of patients with high-MOTox decreased cycle-by-cycle from 2.626 to 1.953 and from 57.8% to 36.6%, respectively. High-MOTox conditions during previous cycles were prognostic risk factors for a new occurrence (ORs range from 1.522 to 4.439), showing that patient's history of toxicities played an important role in the evolution of overall toxicity burden during therapy. Conventional regimen may be preferred to dose-intense in terms of AEs at cycles 2–3 ($p<0.05$).

Strengths and limitations of this study

- The Multiple Overall Toxicity (MOTox) score describes the overall toxicity burden in terms of multiple toxic adverse events, maximum-severity episode and cycle time-dimension.
- Cycle-by-cycle multivariable models are used to assess the evolution of high MOTox over therapy, based on patients' characteristics.
- The MOTox approach can be applied to any cancer studies with Common Terminology Criteria for Adverse Events-graded toxicity data.
- The study retrospectively analysed MRC BO06 osteosarcoma trial data, limited to a young population with a rare tumour.
- External validations are needed to evaluate the application of the novel score to guide treatment decisions, both for osteosarcoma and other cancer types.

Conclusions The novel longitudinal method developed can be applied to any cancer studies with CTCAE-graded toxicity data. After validation in other studies, the MOTox approach may lead to improvements in healthcare assessment and treatment planning.

Trial registration number ISRCTN86294690; Post-results.

INTRODUCTION

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.¹ The introduction of the Common Terminology Criteria for Adverse Events (CTCAE)² multimodality grading system greatly facilitated the standardised reporting of AEs and the comparison of outcomes between trials and institutions.^{3,4} According to CTCAE, AEs range in severity from minor, asymptomatic changes to life-threatening injuries or death.³ Characterisation of toxicity is of interest to patients and clinicians engaged in shared decision making about a treatment strategy.⁵ 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.⁶ Incorporating time into analysis of toxicity is important for the comparison between different chemotherapy regimens or even multiple toxicities from the same regimen.⁷ Therefore, it is crucial to provide an effective tool to assess the evolution of overall toxicity over chemotherapy treatment to guide the therapy strategy.

Since patients might have different types and number of AEs, to summarise 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).^{3,4,7-10} 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, that is, weighted sums of individual toxic effects,¹⁰⁻¹⁴ 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 insights on AEs and their evolution over time.⁵ This information could give new insights in cancer treatment.

In this framework, alternative methods of longitudinal toxic event evaluation have been proposed^{5,10,15-17} but none of them is focused on analysing the evolution of high overall toxicity (high-MOTox) over treatment using a cycle-by-cycle approach. To quantify risk for each patient, 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 studied using logistic regression models to predict high-MOTox at the end of the cycle using achieved chemotherapy dose, previous toxicities, biochemical and haematological factors over time. A new method to summarise multiple toxic AEs including time-dimension, that is, the longitudinal MOTox score, and a cycle-by-cycle analysis of high-MOTox over treatment using personalised characteristics are proposed.

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 randomised controlled trial¹⁸ (RCT) for the treatment of osteosarcoma, a malignant bone tumour mainly affecting children and young adults with an annual incidence of 3-4 patients per million.^{19,20} Patients were treated with cisplatin (CDDP) and doxorubicin (DOX), two cytotoxic drugs commonly used in the treatment of various types of human cancers and are characterised by various toxic AEs: apart from nausea, specific renal and neurotoxicity^{21,22} for CDDP or cardiotoxicity^{23,24} for DOX. Longitudinal MOTox scores over therapy were computed considering non-haematological toxicity. Demographics, treatment-related and biochemical characteristics were used to examine high-MOTox over cycles.

MRC BO06 trial was chosen due to the accessibility of the data by the authors. Provided that longitudinal CTCAE-graded toxicity data are available, the novel MOTox scores can be applied to analyse data from any cancer treatment.

METHODS

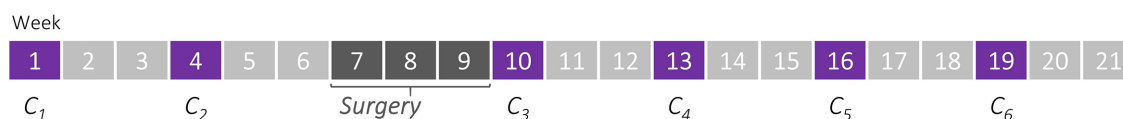
Patients and data

Data from the MRC BO06/EORTC 80931 RCT for patients with non-metastatic high-grade osteosarcoma recruited between 1993 and 2002 were analysed.¹⁸ Patients were randomised 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, see [figure 1](#)). Details concerning the trial protocol are provided in online supplemental appendix A.

The dataset included 497 eligible patients; 19 patients who did not start chemotherapy (13) or reported an abnormal dosage (ie, +25% higher than planned) of one 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 randomisation (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 2](#).

Patients baseline characteristics (age, sex, allocated chemotherapy regimen, site and location of the tumour) were registered at randomisation. Treatment-related

Regimen-C: DOX+CDDP every 3 weeks (DOX: 75 mg/m²/week; CDDP: 100 mg/m²/week)



Regimen-DI: DOX+CDDP every 2 weeks (DOX: 75 mg/m²/week; CDDP: 100 mg/m²/week)

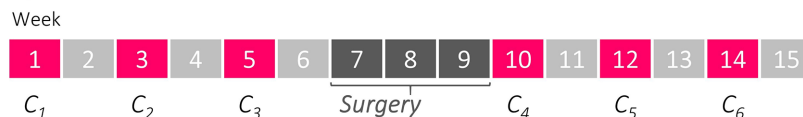


Figure 1 Patients are randomised at baseline to one of the two regimens, with the same anticipated cumulative dose (doxorubicin DOX 25 mg/m²/day for 3 days+cisplatin CDDP 100 mg/m² as a continuous 24 hours infusion on day 1) but different duration (3-weekly vs 2-weekly cycles). CDDP, cisplatin; DOX, doxorubicin.

factors (administered dose of chemotherapy, cycles delays, haematological and biochemical parameters, chemotherapy-induced toxicity and histological response to preoperative chemotherapy) were collected prospectively at each cycle of chemotherapy.¹⁸ Levels of renal clearance, alkaline phosphatase (ALP), lactate dehydrogenase, calcium (Ca) and magnesium were measured at the beginning of each cycle (ie, before the drugs administration) according to local practice. Blood counts (white

cell count (WCC), neutrophils (N), platelets (PLT)) 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*). Non-haematological chemotherapy-induced toxicity for nausea/vomiting, mucositis, neurological toxicity, cardiac toxicity, ototoxicity and infection were graded according to the CTCAE V.3.0,² with grades ranging from 0 (none) to 4 (life-threatening) (see [table 1](#)). For each patient and cycle, chemotherapy dose was analysed as percentage of

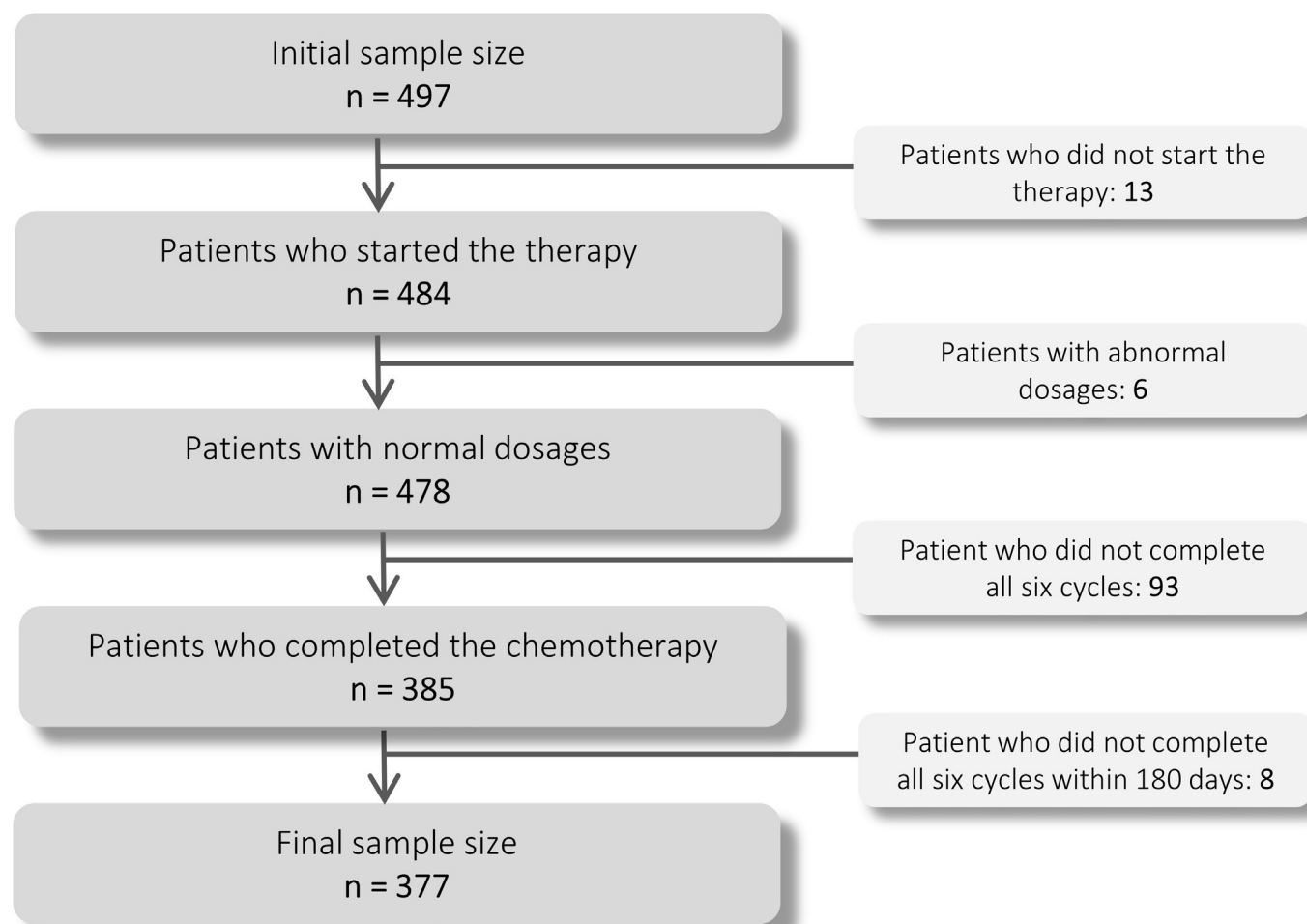


Figure 2 Flow chart of cohort selection.

Table 1 Toxicity coding based on Common Terminology Criteria for Adverse Events grade V.3.0² for non-haematological chemotherapy-induced toxicity related to nausea/vomiting, mucositis, neurological toxicity, cardiac toxicity, ototoxicity and infection

Toxicity	Grade 0	Grade 1	Grade 2	Grade 3	Grade 4
Nausea/Vomiting	None	Nausea	Transient vomiting	Continuative vomiting	Intractable vomiting
Oral Mucositis	No change	Soreness/erythema	Ulcers: can eat solid	Ulcers: liquid diet only	Alimentation not possible
Infection	None	Minor infection	Moderate infection	Major infection	Major infection with hypotension
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.

achieved chemotherapy dose up to cycle k , that is, the percentage of the cumulative drugs administrated up to cycle k divided by the cumulative drugs planned up to k as explained in online supplemental appendix B.

Patient and public involvement

There was no patient or public involvement in the original RCT, which was run before this was recognised as good practice, nor in this study; patient and public involvement is not common in methodological studies.

Longitudinal MOTox scores and outcomes

The longitudinal chemotherapy-induced MOTox score is a cycle-dependent longitudinal mean-max index that quantifies multiple types of AEs experienced by a patient during treatment. For each subject, multiple AEs are recorded along the cycles according to CTCAE scale, with values from 0 to 4. At each cycle k , the *longitudinal MOTox score* for a patient i is defined as

$$MOTox_i^k = \text{average toxic level}_i^k + \text{worst grade}_i^k,$$

where the average toxic level is 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 (see online supplemental appendix C).

As a mean-max index of CTCAE-graded toxicity levels ranging from 0 to 4 each, the *longitudinal MOTox score* $MOTox_i^k$ ranges from 0 to 8. A MOTox value equal to 0 reflects a patient with all toxicities equal to CTCAE-grade 0, that is, a patient that did not experience any kind of toxicity for the cycle under analysis. A MOTox value of 8 represents a subject with all toxicities equal to CTCAE-grade 4, that is, a subject that experienced the highest level of toxicity burden for each type of toxic AE. 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 choice of a longitudinal mean-max index was made to include the cycle-time component in the analysis and to take into account several aspects: (1) multiple lower-grade chronic toxicities, which may have impact on patient's quality of life; (2) huge level in a specific toxicity, which can cause severe effects and permanent consequences for the patient (delayed interval or reduced chemotherapy dose, resulting in overall lower dose intensity). Considering all these aspects, the MOTox score can detect differences in health status among patients, providing more informative compare than traditional methods.

The median value of MOTox scores over all the patients in all the cycles was defined as global median MOTox value (τ) and it was used as a threshold to define the cycle-dependent longitudinal binary score for high-MOTox, named longitudinal high-MOTox score (see online supplemental appendix C). In particular, binary variable $high - MOTox_i^k$ indicates if patient i experienced *high* MOTox with respect to the global median MOTox value τ at cycle k , that is, it distinguishes patients with *low* (high MOTox equal to 0) or *high* (high MOTox equal to 1) overall toxicity burden over therapy. Binary high-MOTox scores over cycles represent the clinical endpoints used as outcome measures for high-MOTox over treatment.

Statistical analysis

A retrospective analysis to examine prognostic factors for binary *high-MOTox* scores over cycles was conducted. Base-line and treatment-related characteristics were examined. A two-sided significance level of 5% was adopted. R software was used for the analyses.²⁵

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 (ie, histological response, biochemical and haematological markers), missing values were imputed using multiple imputations by chained equations algorithm.²⁶

At each cycle, the impact of factors on binary high-MOTox was examined using multivariable logistic regression models and expressed by ORs.²⁷ An OR>1.0 indicates a greater risk of achieving a high-MOTox in case of a 1-unit increase for numerical characteristics or compared with 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).²⁸

RESULTS

Patient and treatment-related characteristics

Patient characteristics at randomisation are shown in table 2. 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) with 30.2% of adults²⁹ (114). A summary of the biochemical and haematological values measured over the entire dataset is shown in table 3. Grades of chemotherapy-induced non-haematological toxicity over cycles are reported in figure 3. 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).

Non-haematological longitudinal Overall Toxicity scores

For each patient, non-haematological chemotherapy-induced toxicity related to nausea/vomiting, mucositis, neurological toxicity, cardiac toxicity, ototoxicity and infection were considered to compute the longitudinal MOTox score over cycles. MOTox scores (figure 4—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 , that is, 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. The global mean MOTox value τ is reported as solid black line. Different evolution patterns of longitudinal MOTox score

Table 2 Patients' characteristics at randomisation and histological responses

Baseline characteristic	
Patients	377
Age* (years)	
Median (IQR)	15 (11–18)
Minimum/maximum	3/40
Child*	109 (28.9%)
Adolescent*	154 (40.9%)
Adult*	114 (30.2%)
Sex	
Female	148 (39.3%)
Male	229 (60.7%)
Allocated treatment	
Regimen-C	180 (47.7%)
Regimen-DI	197 (52.3%)
Site of tumour	
Femur	227 (60.2%)
Fibula	22 (5.8%)
Humerus	37 (9.8%)
Radius	3 (0.8%)
Tibia	87 (23.1%)
Ulna	1 (0.3%)
Location of tumour	
Distal	217 (57.6%)
Mid-shaft	11 (2.9%)
Proximal	148 (39.2%)
Missing (NA)	1 (0.3%)
Histological response†	
Poor	186 (49.3%)
Good	144 (38.2%)
Missing (NA)	47 (12.5%)

*Age groups were defined according to Collins *et al*²⁹: child (male: 0–12 years; female: 0–11 years), adolescent (male: 13–17 years; female: 12–16 years) and adult (male: 18 or older; female: age 17 years or older).

†The resected specimen was examined histologically to assess response to preoperative chemotherapy.¹⁸ Good histological response was defined as $\geq 90\%$ necrosis in the tumour resected; 10% or more viable tumour after preoperative chemotherapy was defined poor.¹⁸ NA, not available.

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 characterised by high toxicity over cycles, table 4 reports the means of MOTox scores at each cycle k for patients allocated in Reg-DI and Reg-C, $MOTox_{DI}^k$ and $MOTox_C^k$, respectively. In cycles 2–3,

Table 3 Descriptive of biochemical and haematological values over the entire dataset

Biomarkers	Mean (SD)	Median (IQR)	Min/Max
White cell 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/1.73 m 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

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 6** shows the mean values of each non-haematological toxicity along with 95% Bonferroni's CIs 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.

The global median MOTox value τ was then used to compute the longitudinal dichotomous high-MOTox scores over cycles. Right panel in **figure 4** 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-MOTox at cycles 2–3 ($p < 0.05$) was found, supporting results shown in **table 4**. At each cycle, high-MOTox was

strongly associated with low/high MOTox at previous cycles.

Multivariable logistic regression models for high-MOTox 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 WCC, N, PLT, ALP and Ca were considered before the beginning of each cycle (ie, the administration of the course). Only WCC values were considered at the planned nadir of

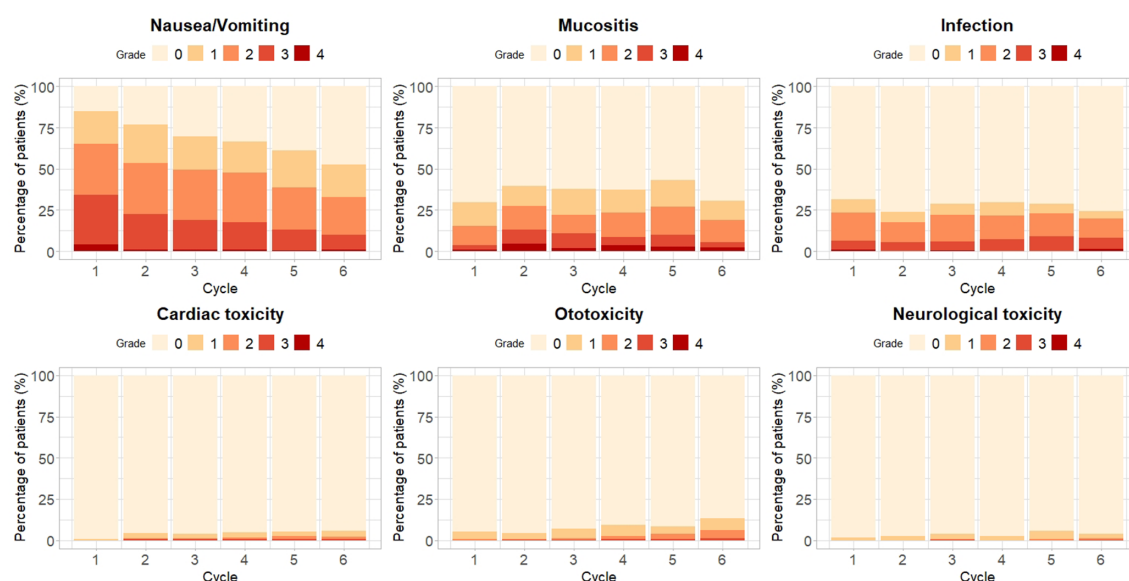


Figure 3 Bar plots of chemotherapy-induced toxicity CTCAE grades over cycles (wheat: 0; light-orange: 1; orange: 2; red: 3; dark-red: 4). Each panel refers to a different type of toxicity: nausea/vomiting (top-left), mucositis (top-centre), infection (top-right), cardiac toxicity (bottom-left), ototoxicity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

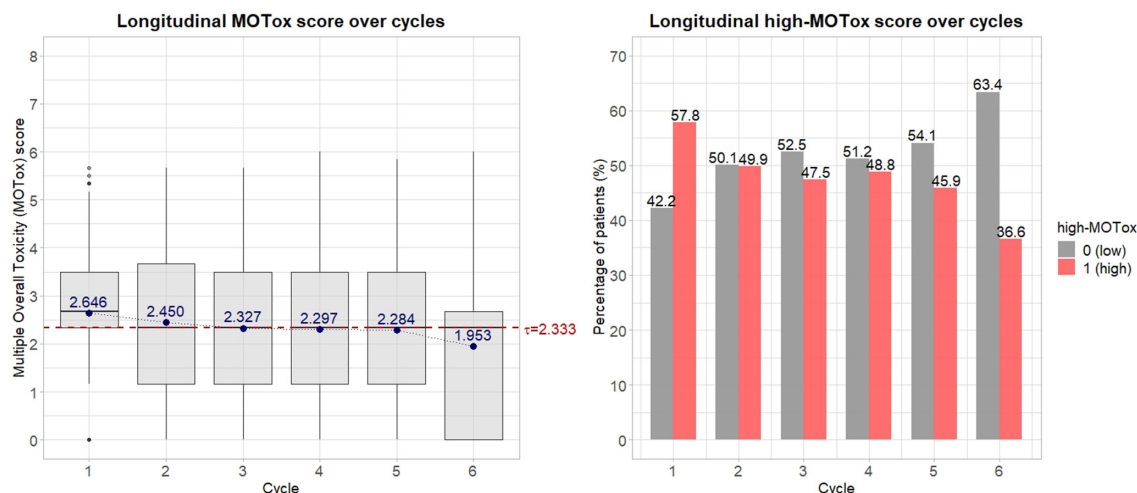


Figure 4 Left panel: Boxplots of longitudinal MOTox scores over cycles. Blue points refers to the mean MOTox values per cycle. Dashed red line refers to the global median MOTox value $\tau = 2.333$. Right panel: Bar plots of longitudinal high-MOTox scores over cycles (grey: 0 or low; magenta: 1 or high). MOTox, Multiple Overall Toxicity.

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 randomisation. Neutrophils-Platelets Score, 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: $n > 7.5 \times 10^9 / L$ or $PLT > 400 \times 10^9 / L$; poor: $n > 7.5 \times 10^9 / L$ and $PLT > 400 \times 10^9 / L$),³⁰ and N-WCC Ratio (NWR, ie, the N

count divided by the WCC) 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 shows estimated ORs along with 95% CIs and overall performances (ie, 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

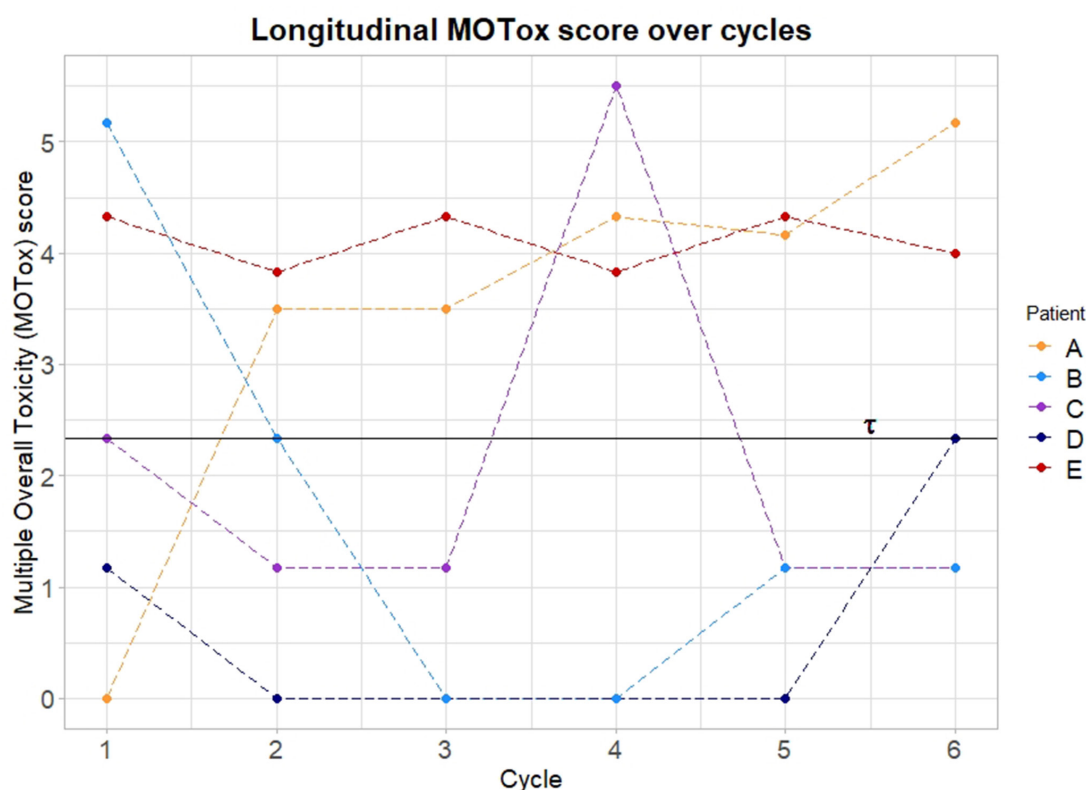


Figure 5 Example of evolution of longitudinal MOTox scores over cycles for five patients from the study cohort. solid black line refers to the global median MOTox value $\tau = 2.333$. MOTox, Multiple Overall Toxicity.

Table 4 Overall toxicity differences between dose-intense (DI) and conventional (C) regimens

	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Cycle 6
\overline{MOTox}_{DI}^k	2.552	2.653	2.488	2.240	2.261	1.920
\overline{MOTox}_C^k	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

\overline{MOTox}_{DI}^k and \overline{MOTox}_C^k are the means of MOTox scores at cycle k for patients allocated in Reg-DI and Reg-C, respectively. MOTox, Multiple Overall Toxicity.

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 WCC at nadir had a protective role on the risk of having high-MOTox (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-MOTox 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-MOTox burden.

The performed analyses were finally used to develop a demo webapp (<http://osteowebapp.prod.s3-website.eu-central-1.amazonaws.com/>). The demo shows how the multivariable models developed to predict high-MOTox index at each cycle could be used as a support tool for clinical decision making. The webapp is presented in online supplemental appendix D.

eu-central-1.amazonaws.com/). The demo shows how the multivariable models developed to predict high-MOTox index at each cycle could be used as a support tool for clinical decision making. The webapp is presented in online supplemental appendix D.

DISCUSSION

Due to the presence of multiple types of 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.

Principal findings

This paper 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

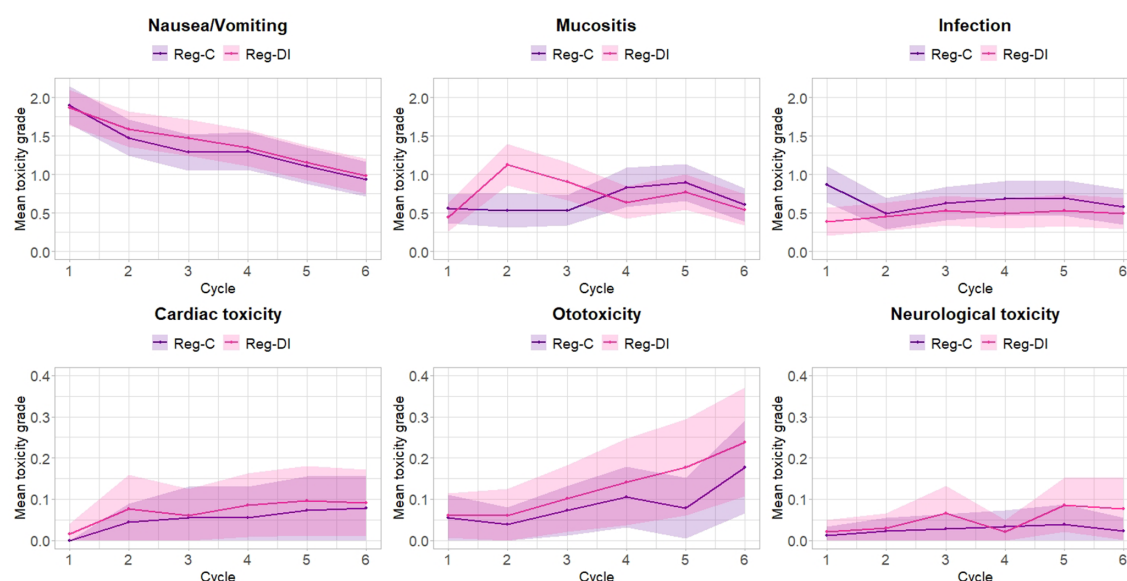


Figure 6 Mean value of chemotherapy-induced toxicity during cycles along with 95% Bonferroni CIs, stratified by the regimens (purple: Reg-C; pink: Reg-DI). Each panel refers to a different type of toxicity: nausea/vomiting (top-left), mucositis (top-centre), infection (top-right), cardiac toxicity (bottom-left), ototoxicity (bottom-centre) and neurological toxicity (bottom-right).

Table 5 Multivariable logistic regression model for each cycle $k \in \{2, 3, 4, 5, 6\}$ with estimated ORs along with their 95% CIs

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

When not specified, haematological factors were computed before the administration of the course.

WCC and PLT are included in the models as difference between the current logarithmic measure and the logarithmic value measured at randomisation.

AUC, Area Under the receiver operating characteristic Curve; DI, dose-intensive; MOTox, Multiple Overall Toxicity; NWR, Neutrophils-White cell count Ratio; PLT, platelets; WCC, white cell count.

longitudinal mean-MOTox score over therapy. Starting from recorded CTCAE grades, the MOTox score summarised the occurrence of repeated AEs allowing to (1) describe the overall toxicity burden, (2) consider the most severe collateral effect and (3) 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 with traditional methods (ie, max-grade or max-time). The cycle-by-cycle longitudinal evolution of high-MOTox was analysed using multivariable logistic regression models to predict binary high-MOTox at the end of the cycle in terms of previous toxicity levels 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-MOTox 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 (ie, 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.¹² 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 (eg, 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.

Strengths and limitations

The results of this study present several strengths and limitations. The presented MOTox and binary high-MOTox scores represent new indices to summarise and quantify patient's overall toxic risk in oncology. These indices can be used to (1) describe patient's response to therapy over cycles, (2) predict the upcoming overall toxicity level given patient's history and (3) support clinical decisions, trying to reduce the impact of therapies in terms of toxic AEs. This study also shows the importance

of collecting well-documented data on chemotherapy-induced AEs during treatment. Provided that toxicities are recorded according to the CTCAE scale, the developed procedure can be applied to any cancer treatment and can be tailored according to a specific study. The longitudinal MOTox score represents a flexible method to quantify the personalised evolutions of the overall toxicity burden in cancer patients.

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 (QOL) were not collected. Although over the last 20 years the main chemotherapy protocol has been used, some aspects of osteosarcoma treatment and supportive care have changed from current measures,⁹ such as the prophylaxis of nausea and vomiting. Such changes are not always easily identifiable and are difficult to account for in retrospective analyses.⁹ 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 it 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.

The novel MOTox approach can become a useful tool for healthcare assessment and cancer treatment planning. As shown in the demo webapp presented in online supplemental appendix D, once validated, the multivariable models could be used to set up a support tool to

predict high-MOTox 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. Added to a cooperation with medical staff, this could lead to new insights for cancer treatment.

Future research

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 (eg, 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 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 health-related QOL or patient-reported outcome 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. The complexity of the problem requires development of new methodologies able to deal with the dynamic nature of the data and all aspects of chemotherapy treatment, such as therapy discontinuation. This is not a trivial task, but it could lead to new improvements for subject-specific predictions and personalised treatment.

CONCLUSIONS

The manuscript introduces 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 with traditional approaches. External validations are needed to evaluate the application of the score in order to guide clinical decisions on treatments prospectively. 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 generalisation to many different settings, added to a cooperation with medical staff, could lead to improvements in the definition of useful tools for healthcare assessment and treatment planning.

Author affiliations

¹MOX, Department of Mathematics, Politecnico di Milano, Milan, Italy

²Mathematical Institute, Leiden University, Leiden, Netherlands

³CHRP – National Centre for Healthcare Research and Pharmacoepidemiology, University of Milano-Bicocca, Milan, Italy

⁴CHDS – Centre for Health Data Science, Human Technopole, Milan, Italy

⁵Department of Solid Tumors, Princess Máxima Center for Pediatric Oncology, Utrecht, Netherlands

⁶Department of Medical Oncology, Leiden University Medical Center, Leiden, Netherlands

⁷Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, Netherlands

⁸Trial and Data Center, Princess Máxima Center for Pediatric Oncology, Utrecht, Netherlands

Twitter Marta Fiocco @marta_fiocco

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

Contributors All authors contributed to the literature search. All authors have accessed and verified the data. MS, FA, FC, FFa, FFg, GG and FI performed the statistical analyses. MS drafted the manuscript and prepared tables and figures. MS, FI, JA, HG and MF interpreted the results and prepared the manuscript. All authors approved the final version of the article. MF acts as guarantor.

Funding None declared. The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Disclaimer The lead author (MS) affirms that the manuscript is an honest, accurate and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

Competing interests None declared.

Patient consent for publication Not required.

Ethical approval Permission to recruit patients to the MRC BO06/EORTC 80931 protocol was provided by the appropriate national and local regulatory and local committees. Link-anonymised data were used for the purposes of this study, and the use of the data was consistent with the consent taken. Because it involved retrospective review, the project was exempt from Human Subjects protection requirements by the Institutional Review Board of the European Organisation for the Research and Treatment of Cancer.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. 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.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Marta Spreafico <http://orcid.org/0000-0002-7773-9976>

Francesca Ieva <http://orcid.org/0000-0003-0165-1983>

Filippo Fedeli <http://orcid.org/0000-0001-5991-2029>

REFERENCES

- 1 Souhami RL, Craft AW, Van der Eijken JW, *et al.* Randomised trial of two regimens of chemotherapy in operable osteosarcoma: a study of the European osteosarcoma intergroup. *Lancet* 1997;350:911–7.
- 2 U.S. Department of Health and Human Services. Common terminology criteria for adverse events v3.0 (CTCAE), 2006. Available: https://www.eortc.be/services/doc/ctc/ctcae_v3.pdf
- 3 Trotti A, Colevas A, Setser A, *et al.* CTCAE v3.0: development of a comprehensive grading system for the adverse effects of cancer treatment. *Semin Radiat Oncol* 2003;13:176–81.
- 4 Zhang S, Chen Q, Wang Q. The use of and adherence to CTCAE v3.0 in cancer clinical trial publications. *Oncotarget* 2016;7:65577–88.
- 5 Thanarajasingam G, Atherton PJ, Novotny PJ, *et al.* Longitudinal adverse event assessment in oncology clinical trials: the toxicity over time (toxT) analysis of alliance trials NCCTG N9741 and 979254. *Lancet Oncol* 2016;17:663–70.
- 6 Lancia C, Spitoni C, Anninga J, *et al.* Marginal structural models with dose-delay joint-exposure for assessing variations to chemotherapy intensity. *Stat Methods Med Res* 2019;28:2787–801.
- 7 Thanarajasingam G, Hubbard JM, Sloan JA, *et al.* The imperative for a new approach to toxicity analysis in oncology clinical trials. *J Natl Cancer Inst* 2015;107. doi:10.1093/jnci/djv216. [Epub ahead of print: 01 Aug 2015].
- 8 Sivendran S, Latif A, McBride RB, *et al.* Adverse event reporting in cancer clinical trial publications. *J Clin Oncol* 2014;32:83–9.
- 9 McTiernan A, Jinks RC, Sydes MR, *et al.* 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. *Eur J Cancer* 2012;48:703–12.
- 10 Trotti A, Pajak TF, Gwede CK, *et al.* Tame: development of a new method for summarising adverse events of cancer treatment by the radiation therapy Oncology Group. *Lancet Oncol* 2007;8:613–24.
- 11 Bekele BN, Thall PF. Dose-Finding based on multiple toxicities in a soft tissue sarcoma trial. *J Am Stat Assoc* 2004;99:26–35.
- 12 Rogatko A, Babb JS, Wang H, *et al.* Patient characteristics compete with dose as predictors of acute treatment toxicity in early phase clinical trials. *Clin Cancer Res* 2004;10:4645–51.
- 13 Lee SM, Hershman DL, Martin P, *et al.* Toxicity burden score: a novel approach to summarize multiple toxic effects. *Ann Oncol* 2012;23:537–41.
- 14 Carhini M, Suárez-Fariñas M, Maki RG. A method to summarize toxicity in cancer randomized clinical trials. *Clin Cancer Res* 2018;24:4968–75.
- 15 Le-Rademacher JG, Hillman S, Storrick E, *et al.* Adverse event burden Score-A versatile summary measure for cancer clinical trials. *Cancers* 2020;12:3251.
- 16 Hirakawa A, Sudo K, Yonemori K, *et al.* A comparative study of longitudinal toxicities of cytotoxic drugs, molecularly targeted agents, immunomodulatory drugs, and cancer vaccines. *Clin Pharmacol Ther* 2019;106:803–9.
- 17 Thanarajasingam G, Leonard JP, Witzig TE, *et al.* Longitudinal toxicity over time (toxT) analysis to evaluate tolerability: a case study of lenalidomide in the CALGB 50401 (Alliance) trial. *Lancet Haematol* 2020;7:e490–7.
- 18 Lewis IJ, Nooij MA, Whelan J, *et al.* Improvement in histologic response but not survival in osteosarcoma patients treated with intensified chemotherapy: a randomized phase III trial of the European osteosarcoma intergroup. *J Natl Cancer Inst* 2007;99:112–28.
- 19 Smeland S, Bielack SS, Whelan J, *et al.* Survival and prognosis with osteosarcoma: outcomes in more than 2000 patients in the EURAMOS-1 (European and American osteosarcoma study) cohort. *Eur J Cancer* 2019;109:36–50.
- 20 Anninga JK, Gelderblom H, Fiocco M, *et al.* Chemotherapeutic adjuvant treatment for osteosarcoma: where do we stand? *Eur J Cancer* 2011;47:2431–45.
- 21 Gregg RW, Molepo JM, Monpetit VJ, *et al.* Cisplatin neurotoxicity: the relationship between dosage, time, and platinum concentration in neurologic tissues, and morphologic evidence of toxicity. *J Clin Oncol* 1992;10:795–803.
- 22 Aldossary SA. Review on pharmacology of cisplatin: clinical use, toxicity and mechanism of resistance of cisplatin. *Biomed. Pharmacol. J.* 2019;12:07–15.
- 23 Zhou S, Starkov A, Froberg MK, *et al.* Cumulative and irreversible cardiac mitochondrial dysfunction induced by doxorubicin. *Cancer Res* 2001;61:771–7.
- 24 Zhao L, Zhang B. Doxorubicin induces cardiotoxicity through upregulation of death receptors mediated apoptosis in cardiomyocytes. *Sci Rep* 2017;7:44735.
- 25 R Core Team. R: A Language and Environment for Statistical Computing. In: *R foundation for statistical computing*. Vienna, Austria, 2018. URL: <https://www.R-project.org/>
- 26 Buuren Svan, Groothuis-Oudshoorn K. mice : Multivariate Imputation by Chained Equations in R. *J Stat Softw* 2011;45:1–67.
- 27 McCullagh P, Nelder J, Models GL. Monographs on statistics and applied probability. *Chapman & Hall* 1989.
- 28 Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett* 2006;27:861–74.
- 29 Collins M, Wilhelm M, Conyers R, *et al.* Benefits and adverse events in younger versus older patients receiving neoadjuvant chemotherapy for osteosarcoma: findings from a meta-analysis. *J Clin Oncol* 2013;31:2303–12.
- 30 Liu B, Huang Y, Sun Y, *et al.* Prognostic value of inflammation-based scores in patients with osteosarcoma. *Sci Rep* 2016;6:39862.

A 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

Spreafico M, Ieva F, Arlati F, Capello F, Fatone F, Fedeli F, Genalti G, Anninga JK, Gelderblom H, Fiocco M.

Supplementary Material

A. MRC BO06/EORTC 80931 Randomized Clinical Trial (RCT) data

Data from the MRC BO06/EORTC 80931 Randomized Clinical Trial (RCT) for 497 patients with non-metastatic high-grade osteosarcoma recruited between 1993 and 2002 were analysed.¹ 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, before and after surgical removal of the primary osteosarcoma. In both arms, DOX (25 mg/m²/d for 3 days) plus CDDP (100 mg/m²) 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 two cycles in *Reg-C* and after three cycles in *Reg-DI*. Postoperative chemotherapy was intended to resume three weeks after surgery in both arms (see Figure 1 in the Manuscript).

Patients baseline characteristics were registered at randomization. The resected specimen was examined histologically to assess response to pre-operative chemotherapy. *Good* histological response was defined as ≥90% necrosis in the tumour resected; 10% or more viable tumour after pre-operative chemotherapy was defined *poor*. 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*) in order to monitor patient's health status and the development of toxicities or adverse events. Delays or chemotherapy dose reductions during treatment were possible in case of toxicity. Non-haematological chemotherapy-induced toxicity for nausea/vomiting, mucositis, neurological toxicity, cardiac toxicity, ototoxicity and infection were graded according to the Common Terminology Criteria for Adverse Events Version 3 (CTCAE v3.0),² with grades ranging from 0 (none) to 4 (life-threatening) (see Table 1 in the Manuscript). Additional details related to the trial protocol can be found in the primary analysis of the trial by Lewis *et al.* (2007).¹

B. Percentage of achieved chemotherapy dose up to cycle k

Chemotherapy dose given at cycle k was analysed as *percentage of achieved chemotherapy dose up to cycle k* . For each patient i , it was defined as the percentage of the cumulative drugs administrated up to cycle k divided by the cumulative drugs planned up to k

$$p\delta_i^k = \frac{\text{cumulative drugs administrated up to cycle } k}{\text{cumulative drugs planned up to cycle } k} \cdot 100\%$$

$$= \frac{\sum_{c=1}^k (DOX_i^c + CDDP_i^c) / sa_i^c \text{ [mg/m}^2\text{]}}{175 \text{ [mg/m}^2\text{]} \cdot k} \cdot 100\%$$

where $k \in \{1, \dots, 6\}$ 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.

C. Longitudinal Multiple Overall Toxicity (MOTox) scores

The longitudinal chemotherapy-induced Multiple Overall Toxicity (MOTox) score is introduced. Let T be the set of different toxicity categories. Let k be the cycle index (which takes value $k = \{1, \dots, 6\}$) 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:

$$\begin{aligned} MOTox_i^k &= \text{average toxic level}_i^k + \text{worst grade}_i^k \\ &= \frac{1}{|T|} \sum_{j \in T} tox_{ij}^k + \max_{j \in T} (tox_{ij}^k) = \text{mean}(tox_{ij}^k) + \max_{j \in T} (tox_{ij}^k) \in [0, 8]. \end{aligned} \quad (C.1)$$

The MOTox score is a *cycle-dependent longitudinal mean-max* index that quantifies the 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

$$\tau = \text{median}_{i,k}(MOTox_i^k),$$

was defined as *global median MOTox value* and it was used as a threshold to define a longitudinal binary score for *high* (or *low*) overall toxicity:

$$\text{high} - MOTox_i^k = \begin{cases} 1 & \text{if } MOTox_i^k > \tau \\ 0 & \text{otherwise.} \end{cases} \quad (C.2)$$

MOTox and high-MOTox scores represent new indices to measure patients' overall toxicity related to multiple types of AEs.

C.1 Toxicity data in MRC BO06/EORTC 80931 RCT

In the MRC BO06/EORTC 80931 RCT, non-haematological chemotherapy-induced toxicity related to

- *naus* Nausea/vomiting
- *oral* Mucositis (inflammation of the mucosae of the gastrointestinal tract, especially the oral ones)
- *inf* Infection
- *car* Cardiac toxicity (heart dysfunctions).
- *oto* Ototoxicity (hearing loss)
- *neur* Neurological toxicity

were registered at each cycle and graded according to CTCAE scale (see Table 1 in the Manuscript). Therefore, in our retrospective analysis, set $T = \{naus, oral, neur, car, oto, inf\}$ was considered to compute MOTox and high-MOTox scores over cycles for each patient, according to Equation (C.1) and Equation (C.2).

C.2 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 (C.1) – as well as the global median OTox 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 i 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. For MRC BO06 trial data, median MOTox values by arms (*Reg-DI* or *Reg-C*) τ_{DI} and τ_C were both equal to 2.333. Therefore, we considered a unique global median value $\tau = 2.333$ defined for all patients.

Binary variable *high* – $MOTox_i^k$ in Equation (C.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 MOTox equal to 0) or *high* (high MOTox equal to 1) overall toxicity burden over treatment.

As a graphical tool, Figure 5 in the Manuscript represents how the MOTox score is able to detect differences in the evolution of the overall toxicity burden among patients during cancer treatment, as shown for patients A-E. It also shows how the global mean MOTox value τ is used to create cycle-by-cycle the binary high-OTox indexes analysed with multivariable logistic regression models.

D. Demo OsteoWebApp

The demo *OsteoWebApp*, available at <http://osteowebapp.prod.s3-website.eu-central-1.amazonaws.com/>, 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 the Manuscript could be used as a support tool for clinical decision making.

The application is implemented through *Amazon Web Services*³ tools and executes the R⁴ code related to the models in Table 6 of the Manuscript. 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 S1. 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 S1 shows that a patient in *Reg-DI* with *high*-MOTox at cycle 1, a cumulative administrated dose of 350 mg/m² (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.

About	2nd Cycle	3rd Cycle	4th Cycle	5th Cycle	6th Cycle
Previous Toxicity (Level) *					
High Toxicity					
Treatment (Regimen) *					
Reg-DI					
Achieved Cumulative Dose (mg/m ²) *					
350					
White Blood Count at randomization (x 10 ⁹ /L) *					
7.65					
White Blood Count During Current (x 10 ⁹ /L) *					
3.90					
Probability of High Toxicity: 0.7356 Sensitivity: 0.68085 Specificity: 0.66667					
<input type="button" value="Predict Toxicity Index"/>					

Supplemental Figure S1. Example of user interface for *OsteoWebApp*.

References

1. Lewis IJ, Nooij MA, Whelan J et al. Improvement in histologic response but not survival in osteosarcoma patients treated with intensified chemotherapy: a randomized phase III trial of the European Osteosarcoma Intergroup. *J Natl Cancer Inst* 2007; **99**(2):112–128. doi: 10.1093/jnci/djk015
2. 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/ctcae3.pdf>
3. Amazon Web Services (AWS) website. URL: <https://aws.amazon.com/it/>
4. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. 2018. URL: <https://www.R-project.org/>