SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Article

Statistical Methods in Medical Research
0(0) 1–14


Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0962280218780619
journals.sagepub.com/home/smm
SAGE

Marginal structural models with dose-delay joint-exposure for assessing variations to chemotherapy intensity

Carlo Lancia,[1] Cristian Spitoni,[2,3] Jakob Anninga,[4] Jeremy Whelan,[5] Matthew R Sydes,[6] Gordana Jovic[6] and Marta Fiocco[1,7]

Abstract

Marginal structural models are causal models designed to adjust for time-dependent confounders in observational studies with dynamically adjusted treatments. They are robust tools to assess causality in complex longitudinal data. In this paper, a marginal structural model is proposed with an innovative dose-delay joint-exposure model for Inverse-Probability-of-Treatment Weighted estimation of the causal effect of alterations to the therapy intensity. The model is motivated by a precise clinical question concerning the possibility of reducing dosages in a regimen. It is applied to data from a randomised trial of chemotherapy in osteosarcoma, an aggressive primary bone-tumour. Chemotherapy data are complex because their longitudinal nature encompasses many clinical details like composition and organisation of multi-drug regimens, or dynamical therapy adjustments. This manuscript focuses on the clinical dynamical process of adjusting the therapy according to the patient's toxicity history, and the causal effect on the outcome of interest of such therapy modifications. Depending on patients' toxicity levels, variations to therapy intensity may be achieved by physicians through the allocation of either a reduction or a delay of the next planned dose. Thus, a negative feedback is present between exposure to cytotoxic agents and toxicity levels, which acts as time-dependent confounders. The construction of the model is illustrated highlighting the high complexity and entanglement of chemotherapy data. Built to address dosage reductions, the model also shows that delays in therapy administration should be avoided. The last aspect makes sense from the cytological point of view, but it is seldom addressed in the literature.

Keywords

Causal inference, marginal structural models, bivariate exposure, osteosarcoma, therapy delay

1 Introduction

This paper presents a detailed discussion of how to analyse longitudinal chemotherapy data in order to assess the causal effect on the outcome of interest of weakening the exposure to cytotoxic agents. The outcome of interest here is Histologic Response (HRe), i.e. the improvement in the appearance of microscopic tissue specimens in a patient after chemotherapy, see Section 2. There are many a rationale behind such a choice. First, it is motivated by a precise clinical question (described below); second, it keeps the complexity of the model to a moderate level and clearly shows both complexity and entanglement of the data; last but not least, HRe is one of the strongest prognostic factors for survival in bone tumours like osteosarcoma[1] or Ewing sarcoma.[2]

[1]Mathematical Institute Leiden University, Leiden, the Netherlands
[2]Mathematical Institute, Utrecht University, Utrecht, the Netherlands
[3]University Medical Center Utrecht, Utrecht, the Netherlands
[4]Raadboud University Medical Center, Nijemegen, the Netherlands
[5]University College Hospital, London, UK
[6]MRC Clinical Trials Unit at UCL, London, UK
[7]Mathematical Institute, Leiden University, Leiden, the Netherlands

Corresponding author:
Carlo Lancia, Mathematical Institute Leiden University, Niels Bohrweg 1, 2333 CA Leiden, the Netherlands.
Email: c.lancia@math.leidenuniv.nl

The analysis of longitudinal chemotherapy data is a complex task because of the presence of a negative feedback between the allocation of the exposure to chemotherapeutic drugs and the toxicities the latter provoke. Toxicities, developed by patients through a chemotherapy cycle, affect subsequent exposure by delaying the next cycle or reducing its dosage. Toxicities are hence *time-dependent confounders*[3]: they are at the same time risk factors[a] for HRe and predictors of subsequent exposure.

Marginal structural models (MSMs) were introduced by J. M. Robins in his seminal paper[6] as a class of models for estimating the causal effect of therapy modifications in presence of time-dependent confounders. MSMs are often based on Inverse-Probability-of-Treatment Weighted (IPTW) estimators,[7,8] which create a pseudo-population by weighting each subject with the inverse of the probability of observing the allocation of dose delays/reductions. In this pseudo-population, toxicity history no longer predicts the next exposure, so a crude analysis estimates the causal effect of exposure modifications. A similar approach is used, for example, when fitting a Cox model in the presence of informative censoring,[9,10] but in this case the method is known as inverse-probability-of-censoring weighted (IPCW) estimation.

Alternative to MSMs is the *mini-trials approach*, described in Klein et al.[11] [§7.4]. However, this method works well only with simple exposure-schemes like starting versus not starting an antiretroviral therapy (and keeping it indefinitely). The key idea of mini-trials is that at each time period, e.g. every month, a fictitious case-control study is started, with the control group formed by all patients that have not yet initiated the therapy. When those patients eventually begin the antiretroviral therapy, they must be artificially censored from each control group they are participating to. Since the censoring is clearly informative, IPCW methods are required to compensate the selection due to the artificial censoring.

Structural nested models (SNMs)[12,13] – and structural nested failure time models (SNFTMs)[14–16] in case of time-to-event data – can also be used in the presence of an exposure-confounder feedback. As the name reveals, these models are nested and require the specification of a quantile–quantile transformation that models the effect of skipping the kth treatment on the outcome of interest. Fitting a SNM is usually much more complicated than a MSM and requires g-estimation.[17,18] Although there are situations where SNMs overcome limitations in the applicability of MSMs,[19] the latter are enough to address the research question stated below.

This paper presents a MSM designed to mimic a randomised trial where the reduction of the exposure intensity is no longer confounded by the toxicity. Since the intensity of the exposure can be weakened by allocating *dose reductions* and/or *cycle delays*, an innovative bivariate exposure model is proposed for the computation of the IPTW estimators. This model is then applied to preoperative data from a large study in resectable osteosarcoma, EURAMOS-1, where the therapy is a multi-agent regimen called MAP (M: Methotrexate (MTX), A: Doxorubicin (DOX), P: Cisplatin (CDDP)). Following clinical input, this paper addresses the question "*what is the effect of reducing MTX by one dose on HRe?*"

Ethical issues make difficult to design clinical trials able to address this research question. Reducing the number of courses of MTX would greatly improve the patients' quality of life because of the unpleasant side-effects produced by this drug at the high dosages administered in osteosarcoma. However, before actually implementing any reduction in the treatment, thorough assessments of the reduction effects are necessary not to compromise patients' prognosis.

MSMs for bivariate exposure-models have been already partially studied in the literature. The earliest example is discussed in Hernán et al.,[20] where the authors estimate the joint causal effect on survival if the start of anti-retroviral therapy and prophylaxis for Pneumocystis Carinii Pneumonia (PCP) is delayed. More recently, the joint effect on mortality of obesity and smoking status was studied in Banack and Kaufman,[21] while Howe et al.[22] address the risk of HIV acquisition due to joint exposure to alcohol and injection drugs. However, this is the first paper where, to the best of the authors' knowledge, MSMs are applied to a class of particularly complex data such as those collected through randomised trials of chemotherapy.

Provided that longitudinal data are available from drug administrations (doses in mg/m^2, treatment dates, and toxicity grades), the model presented here is appropriate to analyse chemotherapy treatments in general. This is particularly true for the treatment of paediatric cancer, where the exposure-confounder feedback is more pronounced.

2 Complexity of chemotherapy data from randomised trials

Multi-agent regimens are designed to maximise the total cell kill within the range of toxicities tolerated by the patient.[23] They also reduce the chance of developing drug-resistant recurrences, which are typical of single-agent regimens.[24] Thus, cytotoxic agents are usually combined to design protocols, which specify rules for administering each drug with respect to both dosage and schedule.

Chemotherapy protocols are very difficult to evaluate because of the large number of alternatives one can test. For example, evidence is in favour of sequential administration of single-agent treatments over multi-drug combinations in advanced breast cancer.[25] In osteosarcoma, the efficacy of combination-chemotherapy is not debated. Nevertheless, adding one more agent to a two-drug protocol significantly improves survival, but the addition of a fourth drug does not.[26] Drug-administration plans, and their organisation in chemotherapy *cycles*, are another aspect that *must* be taken into account. The compression of conventional schedules is thought to achieve greater efficacy by minimising regrowth of tumour cells between treatment cycles.[27]

The situation is further complicated by the dynamical adjustment of the treatment on patients' clinical picture. Exposure to chemotherapy is likely to produce multi-systemic side effects, e.g. organ toxicity or myelosuppression. These side effects are a threat to patient's life and must be controlled by allocating either dose reductions/discontinuations or delays in the administration of the next course. Toxic side effects are often recorded using the Common Terminology Criteria for Adverse Events (CTCAE) grades[28] from 0 to 5. Many different toxicities are recorded in both randomised and non-randomised controlled trials (usually more than 20), and EURAMOS-1 is no exception.

Both dose reductions/discontinuations or delays have the effect of reducing the intensity of subsequent exposure. This is typically measured by the so-called Received Dose Intensity[b] (RDI)

$$\mathrm{RDI} = \frac{\Delta}{\Gamma} = \frac{\dfrac{\text{dose administered}}{\text{dose planned}}}{\dfrac{\text{time for treatment completion}}{\text{planned treatment duration}}} \tag{1}$$

where Δ and Γ are called *standardised dose* and *standardised time*, respectively.[c] On the other hand, physicians are typically eager to restore nominal exposure levels as soon as possible in order to minimise the chance of tumour cell regrowth. Thus, a negative exposure-toxicity feedback is always present, it cannot be removed unless ethical standards are violated, and the efficacy of each protocol can be evaluated through observational studies only.

Another difficulty lies in the therapy variations that may be permitted by a flexible protocol. For example, the EURAMOS-1 protocol allows for the administration of up to 2 extra preoperative doses of MTX. This variation is present in the protocol to enable physicians to adjust preoperative cumulative doses of MTX, e.g. in case of surgery delays. *Caution is required in this situation* to fulfill all hypothesis underlying MSMs. For details, see Section 3.3 below. Also, one has to consider the interconnectedness of drugs: problems with one drug might lead to delays in the administration of the others.

Finally, one has to deal with the way chemotherapy data were collected. In the EURAMOS-1 dataset, for example, only the starting date of each cycle is recorded, so that administrations of DOX, CDDP, and MTX from the same cycle are referred to the same timestamp. In other words, data were thus recorded per-cycle rather than per-administration, that is, it is not possible to distinguish delays that occurred within or between cycles. In other situations chemotherapy data might be even recorded just as overall cumulative exposures and severest toxicities, possibly divided in pre- and post-operative. The way data were collected is a crucial detail and a potential showstopper for observational studies like the one reported here.

3 Materials and methods

This manuscript analyses preoperative data from EURAMOS-1, a large ongoing study in resectable osteosarcoma.[32–35] The preoperative protocol is a three-drug regimen known as MAP (M: Methotrexate (MTX), A: Doxorubicin (DOX), P: Cisplatin (CDDP)), structured in two cycles with a nominal duration of five weeks each. A dose of DOX ($75\,\mathrm{mg/m^2}$) and one dose of CDDP ($120\,\mathrm{mg/m^2}$) are administered at the beginning of week 1. Two doses of MTX ($12000\,\mathrm{mg/m^2}$ each) are administered at the beginning of weeks 4 and 5, respectively. Surgery is expected to be performed by week 11, and the resected specimen evaluated immediately after in order to assess HRe. If less than 10% of the tumour is still viable, then HRe is considered 'good', otherwise 'poor'. Figure 1 shows the organisation of the preoperative treatment and the allocation of therapy modifications for a specific subject.

At clinicians' discretion, one or two extra courses of MTX could have been administered between the end of cycle 2 and surgery to adjust the preoperative exposure to MTX. Depending on the HRe, after surgery subjects were randomised to post-operative treatments having nominal duration of 28 weeks.

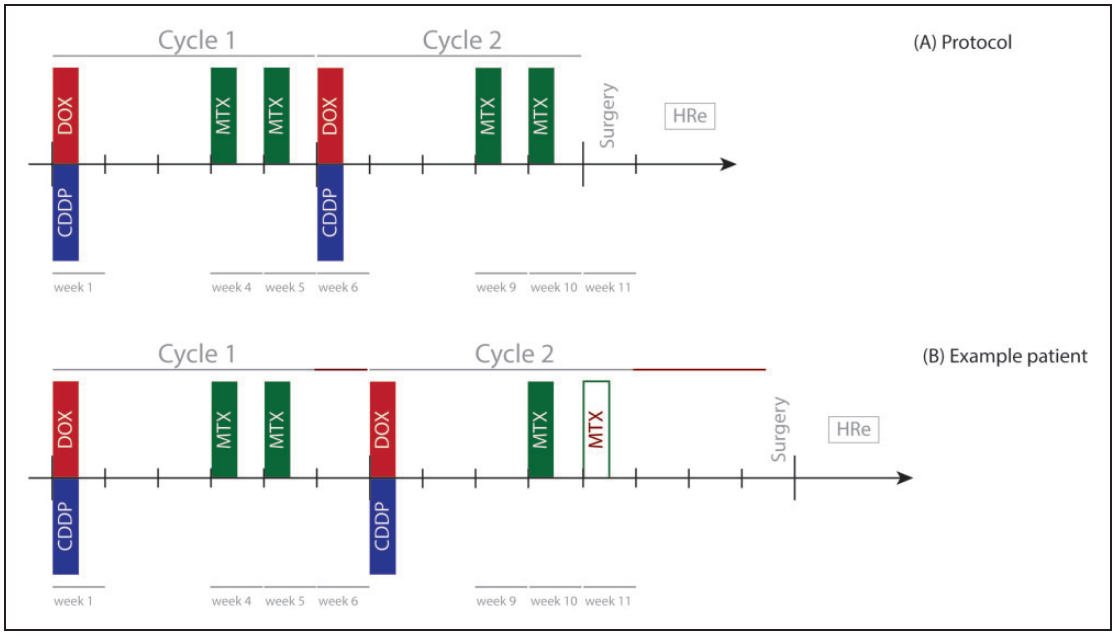

Figure 1. Illustration of the EURAMOS-1 preoperative treatment (a). Example patient (b) experienced a delay of one week in cycle 1 and a discontinuation of the second MTX course and a delay of 18 days in cycle 2.

HRe is a strong prognostic factor for both overall and event-free survival.[1] Although the main results on survival in EURAMOS-1 have not yet been published, the association between HRe and survival is sufficiently well established that it is taken for granted in the osteosarcoma community.

3.1 Dataset description

The initial dataset included 457 patients enrolled between 2005 and 2011 in the branch of EURAMOS-1 coordinated by the European Osteosarcoma Intergroup (EOI), these data relate to the dataset described in Whelan et al.[35] Only the starting date of each cycle was recorded in the chemotherapy registration form. Therefore, the cohort for this analysis was restricted to patients with a registered surgery date because this date was needed to assess delays in completion of cycle 2. To compensate for the time between end of cycle 2 and surgery, the nominal duration of the second chemotherapy cycle was extended to six weeks. For this analysis, patients were excluded also in case of missing HRe and or surgery not performed at the end of cycle 2.[d]

Filtering patients that met the criteria for this analysis gave 364 patients, of which 58% were males and 42% were females. Age in the sample ranged from 4 to 40 following a distribution that is in line with the clinical literature of osteosarcoma. Figure 2 illustrates the patients' selection through a CONSORT-like diagram.

Among the 93 patients excluded from this analysis, 71 have surely undergone surgery because only one between HRe and surgery time is missing. The event/censoring time of the remaining 22 patients is in general much larger than the nominal duration of the preoperative treatment, see Figure 3. In addition, for nearly all of them (20/22), the completion of preoperative treatment could be assessed from chemotherapy data.

In the chemotherapy registration form of EURAMOS-1, it was possible to record (*i*) the CTCAE grade of 21 different adverse events and (*ii*) the indication of whether these produced a delay and/or a reduction. Moreover, the chemotherapy form allowed to grade up to five additional toxicities in the form of free-text descriptions. These free-text toxicities contributed more than 500 unique strings describing preoperative adverse events, each of them associated with the corresponding CTCAE grade and indication of a produced delay/reduction. Both free-text and protocol toxicities were clinically evaluated and mapped[e] by J.A. – an expert oncologist – onto the following four categories: Drug Reaction (DR), Infection (II), Altered Laboratory Values (LV), Myelotoxicity (MT). According to Cole and Hernán[36] [p. 659], this *modus operandi* is preferable over a more detailed specification of the confounders.

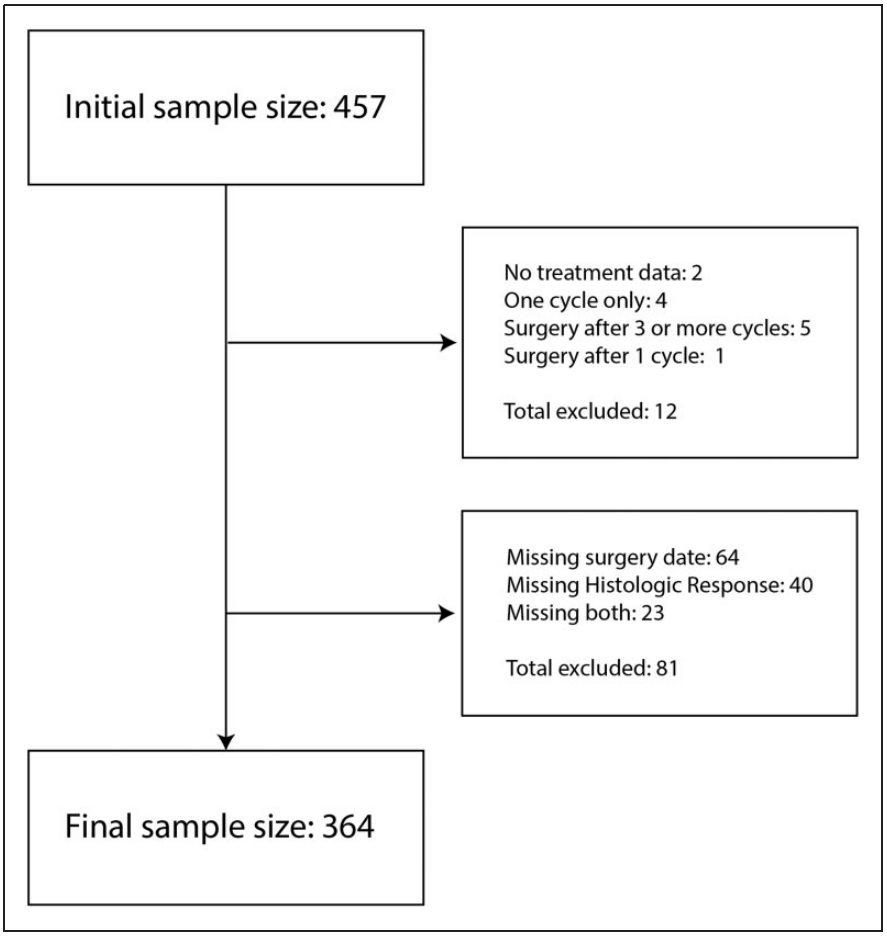

Figure 2. CONSORT-like diagram illustrating selection of patients.

As outlined in Section 1, the clinical question motivating this paper targets the possibility of reducing exposure to MTX. Data of DOX and CDDP was explored first to evaluate the administration of these two drugs. Figure 4 plots, for each cycle, the standardised dose[f] of DOX + CDDP against the standardised dose of MTX. Some properties of the dataset can be derived from this figure:

(1) *Only a negligible fraction of the subjects reported a reduction in the dosage of DOX + CDDP*. This finding permits the exclusion of DOX + CDDP from the definition of the exposure in Section 3.3.
(2) *Only course discontinuations occurred in the adjustment of MTX dosage*. Indeed, the standardised dose per cycle of MTX is tightly located around half-integer values, i.e. no partial doses were administered and the number of MTX courses can be reconstructed from the standardised dose. This finding is in line with what the EURAMOS-1 protocol prescribes for MTX about dose modifications.[33] [§9.1.8.2.8].
(3) *Collected data are not clear about the effect of toxicities on delays*. The chemotherapy registration form included a field named ''did this toxicity cause delay or reduction?''. However, Figure 4 shows in blue (no delay/reduction) some subjects with a marked dose reduction of either DOX + CDDP or MTX. To detect delays, expected cycle durations with the recorded ones were compared.

Figure 5 plots the RDI of MTX against its standardised dose. The solid line satisfies the equation $RDI = \Delta$. Therefore, subjects above this line have a standardised time $\Gamma > 1$ (which means that the end of the cycle was *anticipated*), while subjects below this line correspond to $\Gamma < 1$ (which means that they experienced a delay).

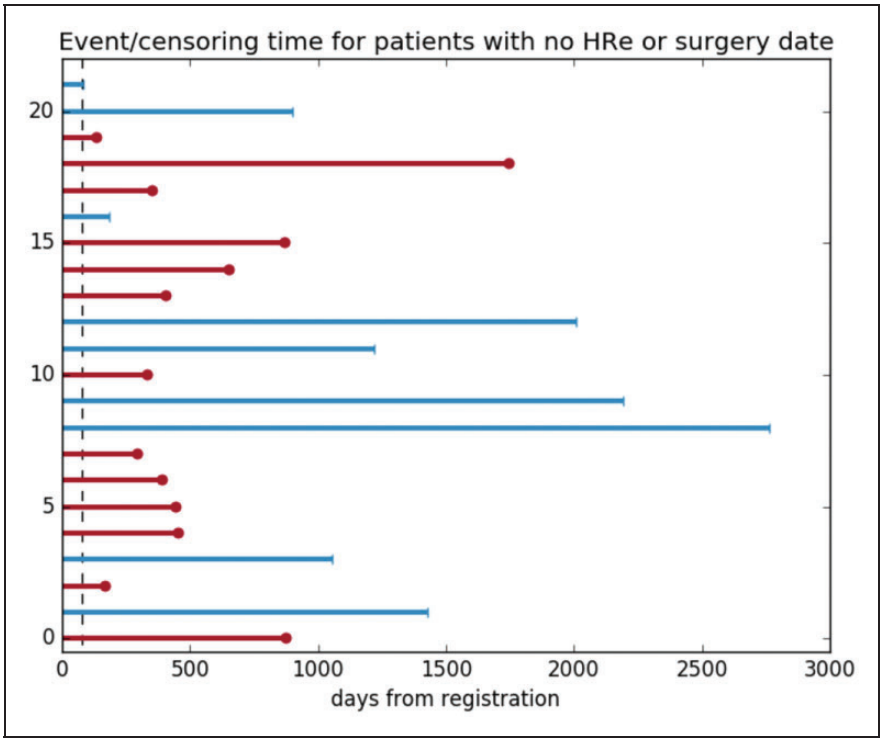

Figure 3. Event/censoring time (red and blue, respectively) for the subjects with both HRe and surgery date missing. The dashed line marks 77 days from registration, i.e. the nominal surgery time; all event/censoring times lie beyond it.

Another two comments can be drawn from this figure:

(4) *The actual duration of cycles presents both anticipations* (up to a week) *and delays* (up to three weeks). This reflects the clinical experience, where variations to the chemotherapy intended-schedule frequently happen often for practical reasons for scheduling.
(5) *Only a negligible fraction of subjects entirely discontinued MTX in cycle 2* ($RDI = 0$). Usually MTX reductions involved only one of the two courses planned in each cycle.

3.2 Modelling the exposure-toxicity feedback

As mentioned above, the EURAMOS-1 preoperative treatment is structured in two five-week cycles. Unfortunately, only the starting date of the cycle was recorded, i.e. the administration date of DOX + CDDP. The two courses of MTX and the severest CTCAE grades for each toxicity category through a cycle are referred to the starting date of the cycle itself. In such a situation, it is quite difficult to model the adjustment of the exposure to a single drug on the past CTCAE history, simply because the inter-cycle causal relationships between exposure and confounders are entangled in the clinical practice and the limitations of the data design.

As pointed out by Property 1 above, dosage of DOX + CDDP is close to nominal for the large majority of the subjects and can be thus excluded from the exposure. Moreover, one should keep in mind that clinical evaluations of toxicity are always performed *before* starting the administration of every dose. A patient must satisfy a number of requirements before the administration of any course of MTX. Among these, *Mucositis* (which falls in the category MT) not worse than CTCAE grade 1; no liver/urinary/ventricular dysfunctions (all included in LV); an adequate White Blood Count (WBC) and platelets count (again MT).

The causal structure proposed here for the EURAMOS-1 chemotherapy data is such that the severest CTCAE grade of each of the four categories DR, II, LV, and MT influences the exposure in terms of number of MTX doses and cycle effective duration. Causal relationships between exposure and confounders are usually depicted in a Direct Acyclic Graph (DAG). Figure 6 shows the one corresponding to the causal structure described above.

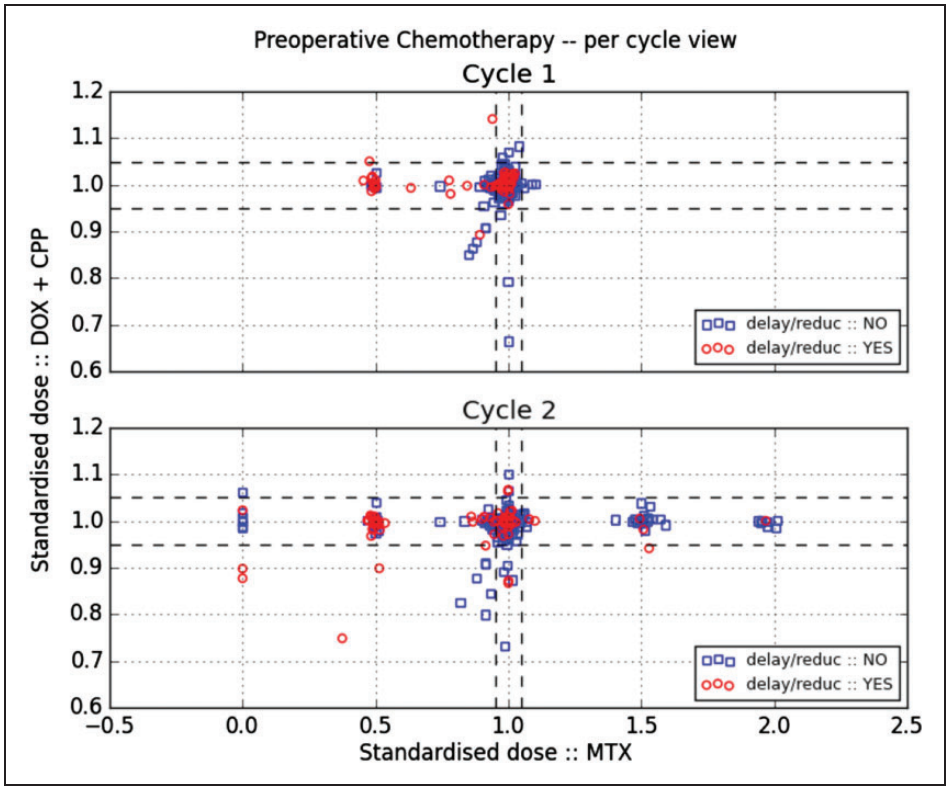

Figure 4. Standardised dose of DOX plus CDDP against standardised dose of MTX. Points lying between the two horizontal (resp. vertical) dotted lines correspond to an administered dose of DOX + CDDP (resp. MTX) within 5% from the intended dosage. Points are coloured in red/blue depending on whether, according to the chemotherapy-registration-form dedicated field, at least a toxicity caused a delay/reduction. On the right part of the lower subfigure, it is possible to notice two groups of patients who received one and two extra courses of MTX.

3.3 Marginal structural model for the histologic response

This section describes the MSM used to measure the causal effect on HRe of reducing by one course the MTX exposure. First, some notation is introduced:

Outcome. Y is a dichotomous variable describing the outcome (HRe), i.e.

$$Y = \mathbb{1}_{\{<10\%\text{ viable tumour}\}} = \begin{cases} 1 & \text{if} < 10\%\text{ viable tumour,} \\ 0 & \text{if} \geq 10\%\text{ viable tumour} \end{cases} \tag{2}$$

where $\mathbb{1}_{\{\cdot\}}$ is the indicator function.

According to the definition of HRe given in Section 3.1, $Y = 1$ is equivalent to a 'good' HRe, while $Y = 0$ denotes a 'poor' HRe.

Exposure. The exposure administered on cycle k is denoted by the vector

$$\begin{aligned} \boldsymbol{A}_k &= \left(A_k^1, A_k^2\right) \\ &\equiv \left(\mathbb{1}_{\{<2\ MTX\ at\ cycle\ k\}},\ \mathbb{1}_{\{>1\ week\ delay\ at\ cycle\ k\}}\right). \end{aligned} \tag{3}$$

According to equation (3), the exposure on cycle k can take on the following four values:

– (0, 0), *at least* two courses of MTX were administered during cycle k,[g] and cycle k was delayed by no more than a week;

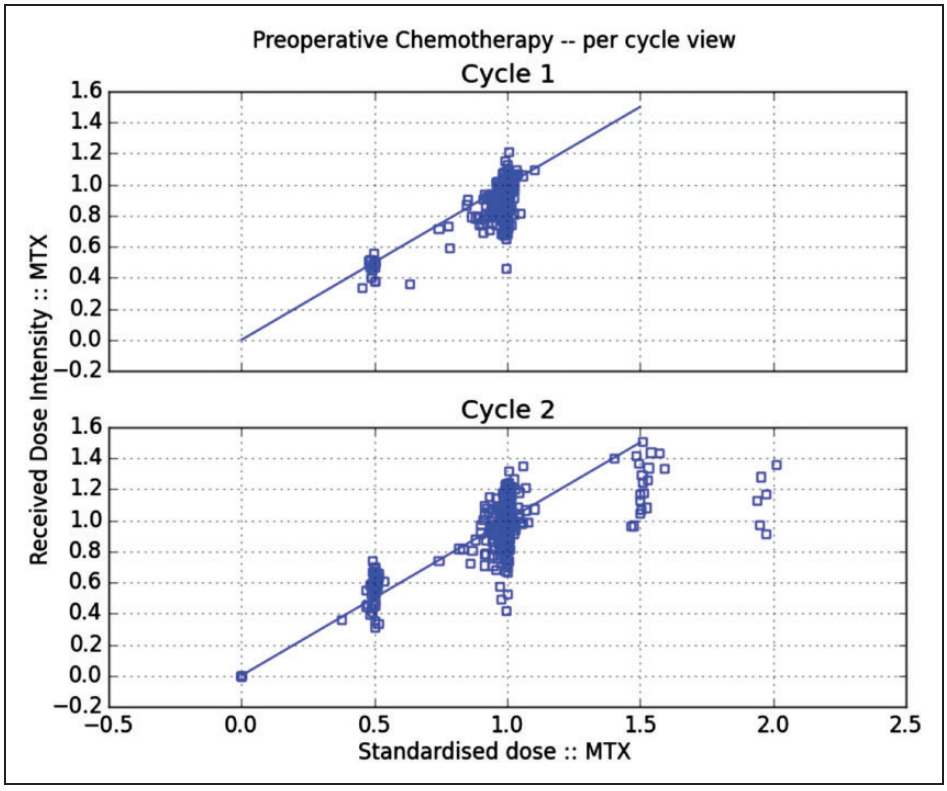

Figure 5. RDI of MTX against standardised dose of MTX. A cycle lasted less than expected for subjects above the solid line and more than expected (delay) for subjects below it.

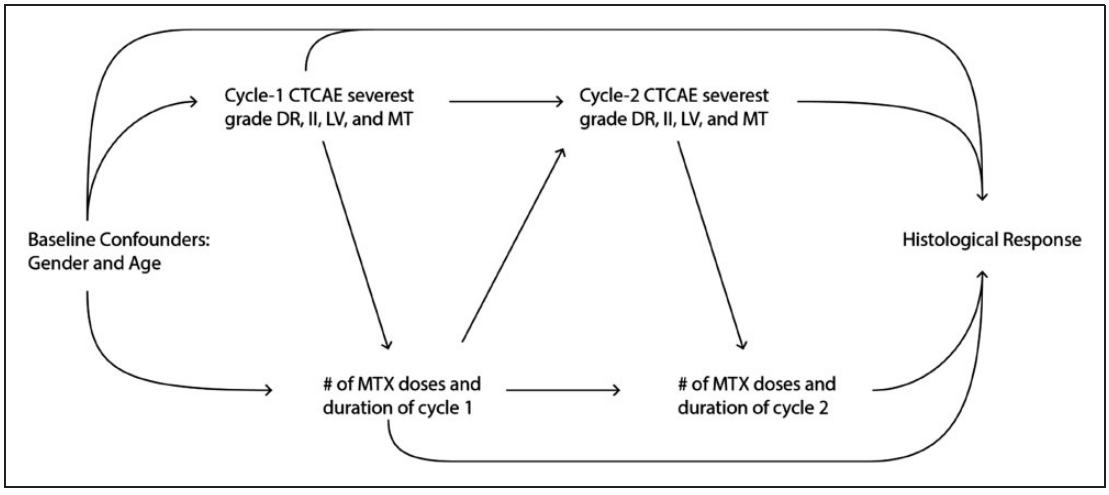

Figure 6. Direct Acyclic Graph (DAG) for the causal relationships between exposure and confounders in EURAMOS-1.

- (0, 1), *at least* two courses of MTX were administered during cycle k, and cycle k was delayed by at least 8 days;
- (1, 0), only one or zero courses of MTX were administered during cycle k, and the cycle was delayed by no more than a week;
- (1, 1), only one or zero courses of MTX were administered during cycle k, and cycle k was delayed by at least eight days.

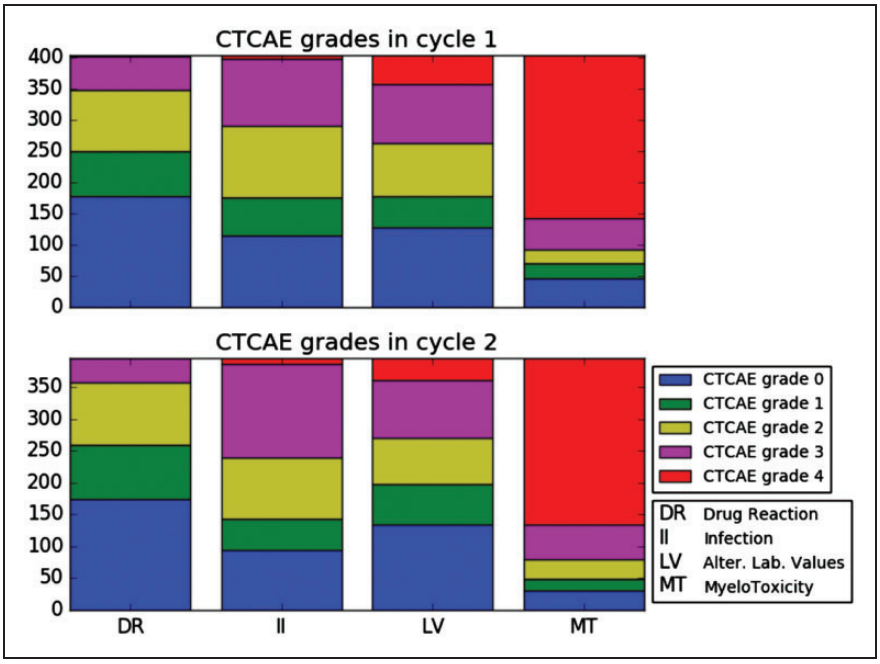

Figure 7. Stacked bar-chart of the preoperative CTCAE grades for each class of toxicity.

Confounders. *According to the literature on MSMs, where the roman capital letter L is used to indicate a confounder, the following variables denote the baseline characteristics and the toxicity categories that influence the allocation of the kth treatment*:

L_k^1: *Drug Reaction (DR) [Toxicity];*
L_k^2: *Infection (II) [Toxicity];*
L_k^3: *Altered Laboratory Values (LV) [Toxicity];*
L_k^4: *Myelotoxicity (MT) [Toxicity];*
L_k^5: *Age at registration [Baseline confounder];*
L_k^6: *Gender [Baseline confounder].*

Figure 7 describes through a stacked bar chart the CTCAE grades measured preoperatively, i.e. in cycle 1 *and* in cycle 2.

The subscript k marks the therapy periods (chemotherapy cycles). As such, L_1^j $(j = 1, \ldots, 6)$ denotes one of the confounders that predict $\boldsymbol{A}_1$, i.e. cycle-1 exposure, while L_2^j denotes one of the confounders that predict $\boldsymbol{A}_2$, i.e. cycle-2 exposure. Finally, the symbol $\boldsymbol{L}_k$ indicates the vector of the six confounders defined above, i.e.

$$\boldsymbol{L}_k = \left(L_k^1, L_k^2, L_k^3, L_k^4, L_k^5, L_k^6\right)$$

The past history of a quantity is denoted by placing an overbar on the corresponding symbol, e.g. $\overline{\boldsymbol{A}}_k \equiv \{\boldsymbol{A}_j\}_{j \le k}$ indicates the sequence of exposure values since beginning of therapy until cycle k. In a similar way, $\overline{\mathbf{L}}_k$ indicates the vector-valued sequence of measured confounders up to (and including) cycle k since the beginning of therapy.

Counterfactual variables are used in causal modelling to describe a *potential outcome*, i.e. an outcome that would be observed had the subject followed, *possibly contrary-to-fact*, a given treatment.

Counterfactual Outcome. *Given the treatment trajectory* $(\bar{\alpha}, \bar{\tau}) = ((\alpha_1, \tau_1), (\alpha_2, \tau_2))$*, let* $Y^{(\bar{\alpha}, \bar{\tau})}$ *be the HRe that would be observed in a subject with exposure history*

$$A_1^1 = \alpha_1, \quad A_1^2 = \tau_1,$$
$$A_1^2 = \alpha_2, \quad A_2^2 = \tau_2, \quad \alpha_j, \tau_j \in \{0, 1\}$$

There are exactly 16 counterfactual outcomes, corresponding to the 16 trajectories $(\bar{\alpha}, \bar{\tau})$ that can be realised by varying $\alpha_1, \alpha_2, \tau_1, \tau_2$ in $\{0, 1\}$.

The model for the histological response is as follows:

Model 1. (MSM logit with bivariate binary-binary exposure)

$$\text{logit } P\left(Y^{(\bar{\alpha},\bar{\tau})} = 1\right) = \beta_0 + \beta_1 \text{cum}(\overline{\boldsymbol{A}})_1 + \beta_2 \text{cum}(\overline{\boldsymbol{A}})_2$$

where $\text{cum}(\overline{\boldsymbol{A}})$ is the cumulative-exposure vector

$$\text{cum}(\overline{\boldsymbol{A}}) = \sum_{k=1}^{2} \boldsymbol{A}_k = \left(\sum_{k=1}^{2} A_k^1, \sum_{k=1}^{2} A_k^2\right).$$

The *dose* component of the cumulative exposure is $\text{cum}(\overline{\boldsymbol{A}})_1 = \sum_{k=1}^{2} A_k^1$ and represents the *number of cycles where MTX was reduced by at least one course*. The *delay* component of the cumulative exposure is $\text{cum}(\overline{\boldsymbol{A}})_2 = \sum_{k=1}^{2} A_k^2$ and represents the *number of cycles that were delayed by more than a week*. The interpretation of the model is then as follows: the quantity e^{β_1} is the *causal* Odds Ratio (OR) associated with reducing the number of courses of MTX in one cycle, while e^{β_2} is the *causal* OR associated with delaying one cycle by more than a week.

Model 1 does not include any baseline covariates. The reason for such a modelling choice is two-fold: on one hand, there is no clinical interest in evaluating the causal effect of exposure modifications within specific population strata, see Robins[19] [§7.1]; on the other, subjects' baseline covariates in the EURAMOS-1 dataset other than age or sex were clinically assessed not to be prognostic factors for HRe.

An interesting model to consider next to Model 1 is the one where extra preoperative MTX is also measured:

Model 2. (MSM logit with bivariate multinomial-binary exposure)

$$\text{logit} P\left(Y^{(\bar{\alpha},\bar{\tau})} = 1\right) = \beta_0 + \beta_1 \text{cum}(\overline{\boldsymbol{A}^*})_1 + \beta_2 \text{cum}(\overline{\boldsymbol{A}^*})_2$$

where $\boldsymbol{A}^*$ is an exposure as defined in equation (3) except that the first component now measures also excesses of MTX

$$A_k^1 = \begin{cases} -1 & \text{if } > 2 \text{ } MTX \text{ doses given at cycle } k, \\ 0 & \text{if } 2 \text{ } MTX \text{ doses given at cycle } k, \\ 1 & \text{if } < 2 \text{ } MTX \text{ doses given at cycle } k \end{cases} \tag{4}$$

The interpretation of Model 2 is similar to that of Model 1.

Under the hypothesis of *No Unmeasured Confounder*,[19,20] both Models 1 and 2 give unbiased estimates of the causal OR of variations in the number of preoperative MTX courses. Being closer to their daily practice, Model 2 is of greater interest for physicians and should better capture the effect of MTX therapy modifications. However, it would not be possible to apply the exposure (4) in a post-operative analysis, while exposure (3) from Model 1 would still work. This is a consequence of the EURAMOS-1 trial-design, which does not allow for extra post-operative MTX except in the very last cycle (for adjusting the *total* cumulative dose). For example, since MTX in excess – observed in cycle 2 – cannot be administered in cycle 3, the hypothesis of *Positivity*[11,37] would be violated by equation (4).

4 Results

The coefficients β_0, β_1, and β_2 in Models 1 and 2 can be correctly estimated via IPTW techniques by a weighted logistic regression with robust variance estimators, e.g. using R package survey.[38] The ith-subject-specfic *stabilised* weights to be used in such a regression are

$$sw_i = \prod_{k=1}^{2} \frac{P\left(A_{i,k}^1, A_{i,k}^2, |\overline{\boldsymbol{A}}_{i,k-1}\right)}{P\left(A_{i,k}^1, A_{i,k}^2, |\overline{\boldsymbol{A}}_{i,k-1}, \overline{\boldsymbol{L}}_{i,k}\right)} \tag{5}$$

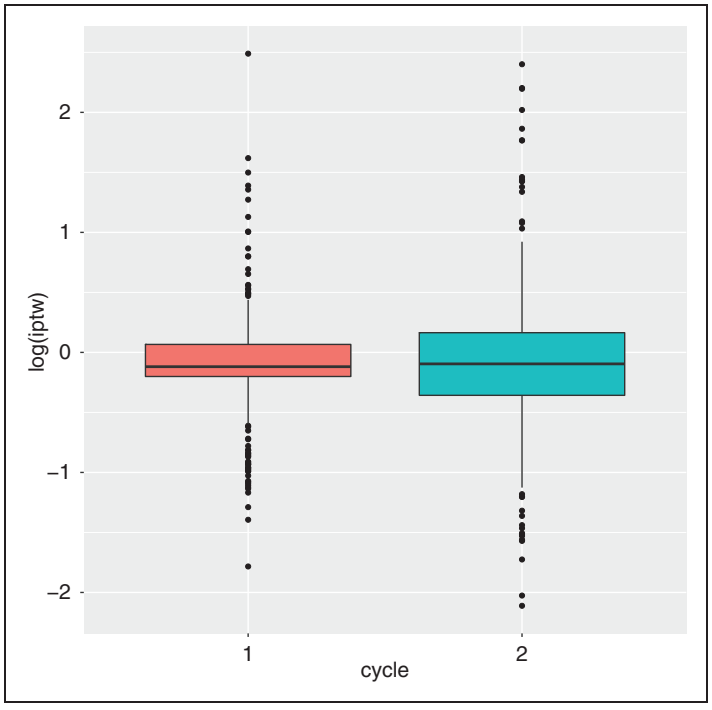

Figure 8. Diagnostic boxplot of subject-specific weights computed via equation (6). The scale on the *y*-axis is logarithmic. The left boxplot corresponds to the terms with $k = 1$ only, while the right one corresponds to the full product.

Table 1. Model 1: Odds ratio (OR) and 95% confidence interval (CI).

	OR	95% CI
(Intercept)	1.300	[0.971, 1.739]
$\text{cum}(\overline{\boldsymbol{A}})_1$	1.295	[0.697, 2.404]
$\text{cum}(\overline{\boldsymbol{A}})_2$	0.636	[0.370, 1.094]

Applying the definition of conditional probability, sw_i is usually recast[20–22] as follows

$$sw_i = \left[\prod_{k=1}^{2} \frac{P\left(A^1_{i,k} | \overline{\boldsymbol{A}}_{i,k-1}\right)}{P\left(A^1_{i,k} | \overline{\boldsymbol{A}}_{i,k-1}, \overline{\boldsymbol{L}}_{i,k}\right)}\right]\left[\prod_{k=1}^{2} \frac{P\left(A^2_{i,k} | \overline{A^1}_{i,k}, \overline{A^2}_{i,k-1}\right)}{P\left(A^2_{i,k} | \overline{A^1}_{i,k}, \overline{A^2}_{i,k-1}, \overline{\boldsymbol{L}}_{i,k}\right)}\right]. \tag{6}$$

The sw_i's can be estimated by use of equation (6) and existing software like the R package ipw.[39] In case of Model 1, the stabilised weights can be alternatively estimated from equation (5) using a Bivariate Logistic Odds-ratio Model (BLOM) – available through the R package VGAM.[40] Fitting a joint exposure-model is not exactly in the original spirit of MSMs, but the computation of the weights with both methods does not show any difference anyway. Figure 8 shows a plot of the weights (equation (6)) used to fit Model 2. The left boxplot corresponds to the product of the factors for $k = 1$ only, while the right boxplot shows the final weights.

Tables 1 and 2 show estimates and Confidence Intervals (CIs) of β_0, β_1, and β_2 for Models 1 and 2, respectively. In both models, decreasing MTX in one cycle has a positive effect on HRe, while delaying one cycle by more than a week has a negative effect. With respect to Model 1, the negative effect of delaying a cycle is in Model 2 more pronounced (OR 0.564 vs. 0.636) and becomes (barely) significant (p-value 0.045 vs. 0.103). Since a negative effect of delays is in accordance with the cytological intuition,[h] this suggests that Model 2 might better capture the routine clinical practice. The formulation of Models 1 and 2 differs only slightly, as the range of $\text{cum}(\overline{\boldsymbol{A}})_1$ is {0, 1, 2} while that of $\text{cum}(\overline{\boldsymbol{A}^*})_1$ is just $\{-1, 0, 1, 2\}$. Nevertheless, the effect of cumulative delay is statistically significant in 2 only. This is attributable to the different way in which equation (6) weights patients with an excess of MTX.

Table 2. Model 2: OR and 95% CI.

	OR	95% CI
(Intercept)	1.289	[0.958, 1.735]
$\text{cum}(\overline{A^*})_1$	1.460	[0.802, 2.656]
$\text{cum}(\overline{A^*})_2$	0.564	[0.323, 0.986]

Although there is a substantial balance between Poor and Good HRe (10 vs. 12) in the original group of patients that received extra MTX, in the pseudo-population this ratio is nearly 3:1 (55.21 vs. 20.13). The observed increase in the OR of the cumulative reduction from Models 1 to 2 is consistent with this remark. The gain of power is a natural consequence of the lower panel of Figure 5.

Based on the results reported, there is no evidence to support the hypothesis that a reduction in the number of courses of MTX might improve the HRe whilst accounting for patients that could have had an increased number of cycles for unrecorded reasons. However, in a clinical scenario where an operative choice need to be taken between

- giving two courses of MTX but delaying the cycle because of high toxicity,
- skipping one course of MTX and not delaying the end of the cycle, the findings of Models 1 and 2 suggest that the latter decision should probably be taken.

5 Conclusions

This paper discussed the application of MSMs, a well-established methodology in causal inference, to a novel class of longitudinal data. These are pre-randomisation data collected in a randomised clinical trial of chemotherapy. This research was motivated by a sharp yet delicate clinical question on the effect of reducing the exposure to one of the chemotherapy agents administered during the EURAMOS-1 trial.

The outcome addressed by the paper, HRe, is acknowledged as one of the strongest prognostic factors for survival in osteosarcoma.[1] Moreover, addressing the HRe allowed the use of data from 95 *non-randomised subjects*. These are patients that were not admitted to or refused to enroll on post-operative treatment.

Concerning the possibility of reducing the number of courses of MTX *tout-court*, ORs listed in Tables 1 and 2 are not statistically significant, so no compelling conclusion can be provided. It would be interesting to repeat the analysis with a separate, larger dataset, where hopefully a stronger fraction of the subjects have a reduced number of courses of MTX and/or delayed cycles. Indeed, Figure 5 shows a strong unbalance among the trajectories $(\bar{\alpha}, \bar{\tau})$, with the majority of subjects having no reductions and no delays (larger than a week). For this manuscript, it was only possible to get access to the EOI-collected data from the EURAMOS-1 trial.

A major achievement of this work is having shown the potentially detrimental effect of chemotherapy delays, an aspect of treatment data seldom addressed in the literature. If confirmed by other studies, especially on survival outcomes, *the negative effect of chemotherapy delays could call for a modification of current clinical practice*.

As a final remark, a validation study using another independent dataset is highly recommended to increase the overall impact of this study. Verification of the negative effect of delaying a therapy cycle in osteosarcoma could be sought in three past European Organisation for Research and Treatment of Cancer (EORTC) studies, namely, trials 80831, 80861, and 80931 (EOI trials BO02, BO03, and BO06, respectively). Subjects randomised to the control arm of these studies are administered identical doses of DOX + CDDP. About the effect of MTX, more insight could be possibly achieved by analysing data from the case arm of EORTC randomised trial 80831 (EOI trial BO02), where the preoperative cumulative dosage of DOX and CDDP are substantially similar to EURAMOS-1 but the dosages of MTX are in ratio 1:3. This is work in progress.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Dutch Foundation KiKa (Stichting Kinderen Kankervrij), grant 163, through the project *Meta-analysis of individual patient data to investigate dose-intensity relation with survival outcome for osteosarcoma patients*.

ORCID iD

Carlo Lancia http://orcid.org/0000-0002-1239-0662
Cristian Spitoni http://orcid.org/0000-0003-0192-606X

Supplemental material

The dataset used in this research can requested via the MRC Clinical Trials Unit at UCL, Institute of Clinical Trials & Methodology, UCL, London, UK.

Notes

a. There is clinical evidence of toxicity being associated with HRe in breast cancer.[4] The very same association is not well-established in osteosarcoma, although there is evidence that toxicity are predictors for survival.[5]
b. Received Dose Intensity (RDI) is historically considered a major prognostic factor[27,29] although recent results contrast this belief for osteosarcoma at least.[1,30,31.]
c. In general $\Delta \leq 1$ and $\Gamma \geq 1$ as a consequence of reductions, delays, and discontinuation. Thus, RDI ≤ 1 with the following interpretation: the smaller the value of RDI, the more evident the delays and/or the reductions occurred throughout the treatment.
d. Reasons for not performing surgery at the end of cycle 2 include progressing disease and tumour localised in bones difficult to operate, e.g. pelvis or mandible.
e. Keeping the highest CTCAE grade.
f. Analogously to the quantity Δ appearing in (1), the standardised dose is computed here as the ratio dose received over dose planned in each cycle (averaged over the number of drugs).
g. Recall that EURAMOS-1 allows for up to 2 extra pre-operative courses of MTX.
h. Roughly speaking, if a delay is introduced in the therapy plan, the chance of tumour regrowth increases.

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 I* 2007; **99**: 112–128.
2. Wunder JS, Paulian G, Huvos AG, et al. The histological response to chemotherapy as a predictor of the oncological outcome of operative treatment of Ewing sarcoma. *J Bone Joint Surg Am* 1998; **80**: 1020–1033.
3. Hernán MA, Brumback BA and Robins JM. Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures. *Stat Med* 2002; **21**: 1689–1709.
4. Singhal V, Singh J, Lyall A, et al. Is drug-induced toxicity a good predictor of response to neo-adjuvant chemotherapy in patients with breast cancer? – a prospective clinical study. *BMC cancer* 2004; **4**: 1.
5. 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**: 703–712.
6. Robins JM. Marginal structural models. In: *Proceedings of the American Statistical Association, Section on Bayesian Statistical Science*, Anaheim, CA, 1997; 1–10. ISBN: 1883276535.
7. Hernán MA, Brumback B and Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of hiv-positive men. *Epidemiology* 2000; **11**: 561–570.
8. Robins JM, Hernán MA and Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; **11**: 550–560.
9. Robins JM and Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* 2000; **56**: 779–788.
10. Willems SJW, Schat A, van Noorden MS, et al. Correcting for dependent censoring in routine outcome monitoring data by applying the inverse probability censoring weighted estimator. *Stat Methods Med Res* 2018; **27**: 323–335.
11. Klein JP, Van Houwelingen HC, Ibrahim JG et al. *Handbook of survival analysis*. Boca Raton, FL: CRC Press, 2013.

12. Almirall D, Coffman CJ, Yancy Jr WS, et al. Structural nested models. In: *Analysis of observational health care data using SAS*. SAS Institute, 2010.
13. Robins JM. Correcting for non-compliance in randomized trials using structural nested mean models. *Commun Stat Theory* 1994; **23**: 2379–2412.
14. Lok JJ. *Statistical modelling of causal effects in time*. PhD Thesis, Leiden University, 2001.
15. Lok J, Gill R, Van Der Vaart A, et al. Estimating the causal effect of a time-varying treatment on time-to-event using structural nested failure time models. *Stat Neerl* 2004; **58**: 271–295.
16. Robins JM. Structural nested failure time models. In: Survival Analysis, Andersen PK, Keiding N (Eds.), The Encyclopedia of Biostatistics, Armitage P, Colton T (Editors). Chichester, UK: Jophn Wiley and Sons, 1998. pp 4372-4389.
17. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math Model* 1986; **7**: 1393–1512.
18. Robins J. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *J Chron Dis* 1987; **40**: 139S–161S.
19. Robins JM. Marginal Structural Models versus Structural nested Models as Tools for Causal inference. In: Halloran ME, Berry D. (eds) *Statistical Models in Epidemiology, the Environment, and Clinical Trials*. The IMA Volumes in Mathematics and its Applications, vol 116. Springer, New York, NY, 2000, pp. 95-133.
20. Hernán MA, Brumback B and Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *J Am Stat Assoc* 2001; **96**: 440–448.
21. Banack HR and Kaufman JS. Estimating the time-varying joint effects of obesity and smoking on all-cause mortality using marginal structural models. *Am J Epidemiol* 2016; **183**: 122–129.
22. Howe CJ, Cole SR, Mehta SH, et al. Estimating the effects of multiple time-varying exposures using joint marginal structural models: alcohol consumption, injection drug use, and HIV acquisition. *Epidemiology* 2012; **23**: 574.
23. Pazdur R, Wagman LD, Camphausen KA, et al. *Cancer management: a multidisciplinary approach: medical, surgical & radiation oncology*. New York, NY: Oncology Group, 2003.
24. Persidis A. Cancer multidrug resistance. *Nat Biotechnol* 1999; **17**: 94–95.
25. Dear RF, McGeechan K, Jenkins MC, et al. Combination versus sequential single agent chemotherapy for metastatic breast cancer. *Cochrane DB Syst Rev* 2013; 12. DOI: 10.1002/14651858.CD008792.pub2
26. Anninga JK, Gelderblom H, Fiocco M, et al. Chemotherapeutic adjuvant treatment for osteosarcoma: where do we stand? *Eur J Cancer* 2011; 47: 2431–2445.
27. Foote M. The importance of planned dose of chemotherapy on time: do we need to change our clinical practice? *Oncologist* 1998; 3: 365–368.
28. US Dept of Health and Human Services. Common terminology criteria Common Terminology Criteria for Adverse Events (CTCAE) 2010, https://www.eortc.be/services/doc/ctc/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf (accessed 1 June 2018).
29. Hryniuk WM. Average relative dose intensity and the impact on design of clinical trials. *Semin Oncol* 1987; **14**: 65–74.
30. Eselgrim M, Grunert H, Kühne T, et al. Dose intensity of chemotherapy for osteosarcoma and outcome in the Cooperative Osteosarcoma Study Group (COSS) trials. *Pediatr Blood Cancer* 2006; **47**: 42–50.
31. Lewis IJ, Weeden S, Machin D, et al. Received dose and dose-intensity of chemotherapy and outcome in nonmetastatic extremity osteosarcoma. *J Clin Oncol* 2000; **18**: 4028–4037.
32. Bielack SS, Smeland S, Whelan J, et al. MAP plus maintenance pegylated interferon {α}-2b (MAPIfn) versus MAP alone in patients with resectable high-grade osteosarcoma and good histologic response to preoperative MAP: first results of the EURAMOS-1 ''good response'' randomization. *J Clin Oncol* 2013; 31(supp 18): LBA10504–LBA10504.
33. EURAMOS. A randomized trial of the European and American Osteosarcoma Study Group to optimize treatment strategies for resectable osteosarcoma based on histological response to a randomized trial of the European and American Osteosarcoma Study Group to optimize treatment strategies for resectable osteosarcoma based on histological response to pre-operative chemotherapy, http://www.euramos.org/media/1258/euramos1_protocol.pdf (accessed 1 June 2018).
34. Marina N, Bielack S, Whelan J, et al. International collaboration is feasible in trials for rare conditions: the euramos experience. In: Jaffe N, Bruland O and Bielack S (eds) *Pediatric and Adolescent Osteosarcoma. Cancer Treatment and Research*, vol 152. Springer, Boston, MA.
35. Whelan J, Bielack S, Marina N, et al. EURAMOS-1, an international randomised study for osteosarcoma: results from pre-randomisation treatment. *Ann Oncol* 2015; **26**: 407–414.
36. Cole SR and Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol* 2008; **168**: 656–664.
37. Robins JM and Hernán MA. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice G, Davidian M, Verbeke G and Molenberghs G (eds) *Longitudinal Data Analysis*, Chapter 23. Chapman & Hall/CRC, 2009, pp. 553–599.
38. Lumley T. Analysis of complex survey samples. *J Stat Softw* 2004; **9**: 1–19.
39. van der Wal WM and Geskus RB. Ipw: an r package for inverse probability weighting. *J Stat Softw* 2011; **43**: 1–23.
40. Yee TW. *Vector generalized linear and additive models with an implementation in R*. New York, NY: Springer-Verlag, 2015.