Contemporary Clinical Trials Communications 17 (2020) 100529

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: http://www.elsevier.com/locate/conctc

Research paper

Proportional odds assumption for modeling longitudinal ordinal multiple toxicity outcomes in dose finding studies of targeted agents: A pooled analysis of 54 studies

Damien Drubay [a,b,*], Laurence Collette [c], Xavier Paoletti [a,b]

[a] INSERM U1018, CESP, Université Paris-Saclay, UVSQ, Villejuif, F-94805, France
[b] Gustave Roussy, Service de Biostatistique et D'Epidémiologie, Villejuif, F-94805, France
[c] European Organization of Research and Treatment of Cancer (EORTC), Headquarter, Biostatistics Department, 1200, Brussels, Belgium

ARTICLE INFO

Keywords:
Dose finding
Multidimensional data
Proportional odds
Targeted agent

ABSTRACT

Background: Data generated by phase I trials is richer than the classical binary DLT measured at the first cycle used as primary endpoints. Several works developed designs for more informative endpoints, e.g. ordinal toxicity grades and/or longitudinal data which relied however on strong assumptions, in particular the proportional odds (PO) assumption.

Methods: We evaluated this PO assumption for the dose and cycle on a large database of individual patient data from 54 phase I clinical trials of molecularly targeted agents. The PO model is a specific case of the continuation ratio logit model (CRLM) with null parameters. We compared the PO and CRLM models using the widely applicable information criterion (WAIC). We considered a longitudinal multivariate ordinal toxicity outcome (cutaneous, digestive, hematological, general disorders, and other toxicities).

Results: WAIC suggested that the CRLM model (WAIC = 30911.58) outperformed the PO model (WAIC = 31432.10). Deviance from PO assumption for dose was observed for digestive and general disorder toxicities. There was moderate cycle effect with slight deviance from PO assumption for the other type of toxicity.

Conclusions: Designs based on PO for dose should be a useful tool for drug with low expected digestive or general disorder toxicity dose-related incidence.

1. Introduction

In oncology, dose finding phase I clinical trials aim at determining the maximum tolerated dose (MTD) as the dose presenting an acceptable rate of severe toxicity during the first cycle of treatment, also called dose limiting toxicity (DLT). Groups of patients are enrolled at increasing dose levels. A given patient is assigned to a dose that is administered in repeated cycles until treatment failure; intra-patient dose escalation is usually not allowed. At each cycle of treatment, adverse events of various types (digestive, hematological, cutaneous, general disorders, etc.) are measured on a graded scale that ranges from 0 (absence of toxicity) to 4 (severe life threatening toxicity). One of the main limitations of dose finding trials is the limited amount of information extracted from the primary outcome [1,2] that results from (i) the composite nature of the outcome 'worst observed toxicity', (ii) the dichotomization of this graded outcome in presence or absence of severe toxicity (DLT), and (iii) the use of data collected at cycle 1 only although more than 50% of the first severe toxicity occur after cycle 1 [3,4].

Recently the European Medicine Agency underlined the importance of analyzing adverse events at all cycles of treatment in order to refine the risk of toxicity and to consider not only severe toxicity but also intermediate toxicity [5]. Some authors have proposed designs based on the longitudinal ordinal toxicity measurements [6,7]. Markov chain models have been explored [8], and others authors have included multiple toxicity constraints in the dose finding design [9–12]. Alternatively, cumulative logit models have been suggested to model the ordinal nature of the graded toxicity, and to estimate the dose effect [13], possibly adjusted for the treatment cycle [1,14–16]. One of their advantages is to match the assumption of increasing toxicity with dose, to provide easy to interpret coefficients, either in terms of odds ratio or absolute probabilities. However, sample sizes for dose finding trials being typically small; to account for multiple toxicity grades, a natural

* Corresponding author. Gustave Roussy, Service de Biostatistique et D'Epidémiologie, Villejuif, F-94805, France.
E-mail address: damien.drubay@gustaveroussy.fr (D. Drubay).

https://doi.org/10.1016/j.conctc.2020.100529
Received 12 August 2019; Received in revised form 3 January 2020; Accepted 19 January 2020
Available online 25 January 2020

simplification of the cumulative logits is the proportional odds (PO) model that assumes that the effect of covariates (here the dose level and the cycle) is similar on the various cumulative logits. This assumption reduces the number of parameters to estimate and allows using intermediate grades of toxicity to refine the estimate of the risk of severe toxicity [17]. Before we implement such an assumption in prospective clinical trials, we explored the PO assumption on real data. The effects of dose and cycle on various types of toxicity were modeled in a large database of individual data of patients treated in 54 phase I clinical trials of molecularly targeted agents provided to the DLT-TARGETT group, a European Organization for Research and Treatment of Cancer (EORTC)-led initiative. We developed a continuation ratio logit model [18] to account for longitudinal multivariate ordinal toxicity outcomes. In order to draw conclusions applicable to future trials, we based our development on predictive modeling strategy, using the horseshoe shrinkage prior [19] to shrink weak coefficients; we compared the models using the widely applicable information criterion [20] (WAIC).

$$\begin{cases} P(Y_{ijc} > 0 | Y_{ijc} \geq 0) & = \text{expit}(\alpha_{ij} + \beta_{j0} + \gamma_{j0} D_i + \zeta_{j0} c) \\ P(Y_{ijc} > 1 | Y_{ijc} \geq 1) & = \text{expit}(\alpha_{ij} + \beta_{j0} + \beta_{j1} + (\gamma_{j0} + \gamma_{j1}) D_i + (\zeta_{j0} + \zeta_{j1}) c) \\ P(Y_{ijc} > 2 | Y_{ijc} \geq 2) & = \text{expit}(\alpha_{ij} + \beta_{j0} + \beta_{j1} + \beta_{j2} + (\gamma_{j0} + \gamma_{j1} + \gamma_{j2}) D_i + (\zeta_{j0} + \zeta_{j1} + \zeta_{j2}) c) \end{cases} \quad (1)$$

2. Material and methods

2.1. Study design

Full toxicity data of 54 completed phase 1 studies evaluating molecularly targeted agents (MTAs) was provided by four academic institutions (Cancer Research UK (United Kingdom), EORTC, National Cancer Institute-Canada and National Cancer Institute) and three pharmaceutical companies (Pfizer, Roche and Sanofi). The MTAs were administered as single agent to adult patients with solid tumors. All patients who received at least one cycle of treatment were included in the analysis.

The reader may refer to the publication by Postel-Vinay et al. [4] for complete details about the data collection and the study design.

2.2. Toxicity data

All grade 1 or above severity adverse events (AEs) reported as at least 'possibly drug-related', which were not present at baseline and occurred between cycle 1 and cycle 6 were selected; in fact previous data showed that the majority of AEs occurred during the first 6 cycles of treatment in dose finding trials [3]. To ensure comparability of the AEs over trials that used different grading systems, the grade of all reported toxicities was harmonized to the National Cancer Institute Common Terminology Criteria of Adverse Events, version 3.0 and labeled for description according to Medical Dictionary for Regulatory Activities (MedDRA) 15. If the same AE was reported at different severity grades during a given cycle, only the worst toxicity grade was taken into account in the statistical analysis. The grades were grouped in 4 levels: no toxicity, toxicity grade 1, toxicity grade 2 and toxicity grade ≥ 3 (usually considered as dose limiting toxicity (DLT) when they occur during the DLT evaluation period).

In this report, we focused on four different types of AEs which are typical of targeted agents [3]: cutaneous, digestive, general disorder and hematologic toxicities. These categories were defined from the MedDRA classification preferred items in accordance with NCIC experts to better fit the cancer phase 1 trial context and the most frequent toxicity reported in early phase trials of single targeted agents [21]. A fifth (heterogeneous) type was defined that contained all other toxicities and was labeled 'other type'.

2.3. Models

As toxicity data was collected at the end of each cycle of treatment, whose duration may vary across trials, the treatment cycle, as defined per protocol, was used as time unit, irrespective of its duration in days. The dose for the patient i, D_i, was standardized by the MTD, i.e. the ratio between the planned dose and the MTD of this trial, or the maximal allocated dose during this trial when the MTD was not reached ($\approx 25\%$ of trials).

We used mixed effect multivariate continuation ratio logit model [18] with correlated random effects to jointly assess the relationship between the J types of graded toxicity ($J = 5$) denoted $\mathbf{Y}$ and the dose D and cycle C considered as continuous variables. The probability that the patient i presents a toxicity of type j ($j = 1, \cdots, 5$) of grade higher than k ($k = 0, 1, 2$) at the cycle c ($c = 1, \cdots, 6$), was:

with $\beta_{jk} = \sum_{l=0}^{k} \beta_{jl}$ the intercept for the grade k, $\gamma_{jk} = \sum_{l=0}^{k} \gamma_{jl}$ and $\zeta_{jk} = \sum_{l=0}^{k} \zeta_{jl}$ the parameters associated to the dose and the cycle respectively, and $\alpha_{ij} \tilde{} MVN(0, \Sigma)$ a patient specific random effect distributed according to a centered multivariate normal distribution with $J \times J$ covariance matrix Σ. expit denotes the inverse of the logit transform function, i.e. $\text{expit}(x) = \exp(x)/(1 + \exp(x))$. Of note, the random effects have been set at the patient level to account for the possible correlations of the repeated measurements. For model tractability, we did not consider patient-level random effects nested in trial-level random effects. Conditional on α_{ij}, we assumed independence of the longitudinal AEs measures. With this parameterization, labeled *full model* in the rest of the paper, the PO assumption for the dose effect then corresponds to $\gamma_{j1} = 0$ and $\gamma_{j2} = 0$, and for the cycle effect to $\zeta_{j1} = 0$ and $\zeta_{j2} = 0$. The *PO model* can then be written as:

$$P(Y_{ijc} > k | Y_{ijc} \geq k) = \text{expit}(\alpha_{ij} + \beta_{jk} + \gamma_j D_i + \zeta_j c)$$

2.4. Model priors

The model parameters were estimated in a bayesian framework. Despite the large number of trials, patients and observations in this joint analysis, the full model contains 45 fixed parameters plus the random effect covariance matrix. To improve the stability of the estimates, we used Horseshoe shrinkage prior [19] for the P parameters of the fixed effects, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta})$, with $\boldsymbol{\beta} = (\beta_{10}, \beta_{11}, \ldots, \beta_{J2})$, $\boldsymbol{\gamma} = (\gamma_{10}, \gamma_{11}, \ldots, \gamma_{J2})$ and $\boldsymbol{\zeta} = (\zeta_{10}, \zeta_{11}, \ldots, \zeta_{J2})$. A normal prior distribution for each parameter θ_p was elicited, in which the parameters for the variance prior, λ_p and τ, followed the standard half-Cauchy distribution $C^+(0, 1)$:

$$\theta_p | \lambda_p, \tau \sim N\left(0, \lambda_p^2 \tau^2\right)$$

$$\lambda_p \sim C^+(0, 1)$$

$$\tau \sim C^+(0, 1)$$

τ was common to all components of θ, and λ_p was specific of θ_p. This approach has common features with the Bayesian LASSO that uses Laplacian prior distributions [22], but it belongs to the global-local shrinkage prior family [23]: a global prior parameter τ shrinks all the

Table 1
Number of toxicities by type and grade.

Grade	Cutaneous	Digestive	General disorder	Hematologic	Others	Total
1	549	1754	403	1344	1513	5563
2	207	794	433	748	946	3128
≥ 3	31	190	345	200	447	1213
Total	787	2738	1181	2292	2906	9904

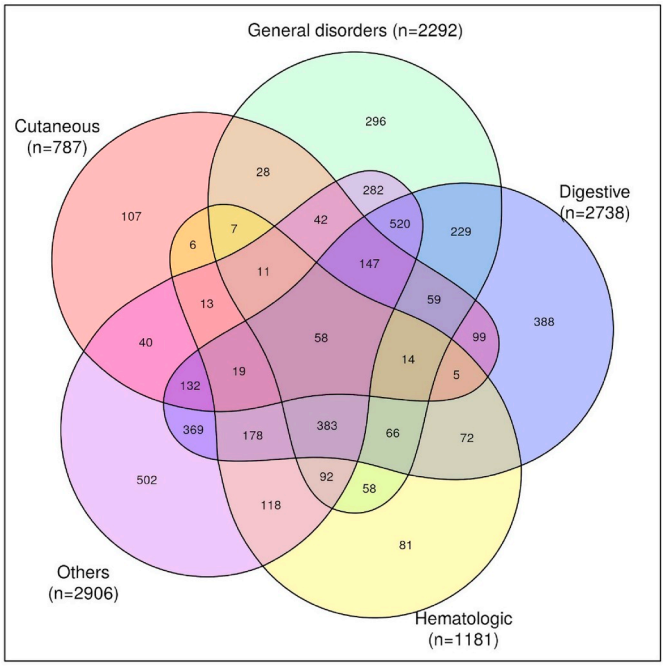

Fig. 1. Co-occurrence of the (n_{total} = 9904) toxicity (all grades) for the 5592 reported cycles.

model parameters towards 0, and local prior parameters λ_p modulates this shrinkage at various degrees. A reparametrisation of τ and λ_p in a κ_p parameter has been proposed to directly quantify the degree of shrinkage [19] by $\overline{\theta_p} = (1 - \kappa_p)\widehat{\theta_p}$, with $\overline{\theta_p}$ the posterior mean of θ_p, $\widehat{\theta_p}$ its maximum likelihood solution, and κ_p the shrinkage coefficient for each parameter p. The distribution of κ_p has a horseshoe ('U') shape, presenting a first point mass at $\kappa_p \approx 0$ (i.e. no shrinkage, $\overline{\theta_p} \approx \widehat{\theta_p}$) and a second at $\kappa_p \approx 1$ (i.e total shrinkage, $\overline{\theta_p} = 0$). The horseshoe prior is therefore expected to apply a more severe shrinkage to zero-coefficients, while avoiding over-shrinkage of larger coefficients, favoring both sparsity and robustness to large signals. Several authors have shown that it outperforms the Laplacian prior for prediction, and provides results comparable to the bayesian model averaging (bayesian gold standard for prediction) without the computation burden [19,23–25].

The covariance matrix of the random effects was modeled as parameters. We used the LKJ prior [26] with the shape parameter $\eta = 1$ (uninformative prior) for the Cholesky factor of the random effect correlation matrix. Parameter posteriors were obtained by an Hamiltonian Monte Carlo sampling scheme [27] (4 chains, 5000 iterations including 1000 burning) using the `Stan` software [28].

2.5. Model selection

In addition of the full and the PO models, a reduced model without the parameters whose 95% credibility intervals that included 0, was fit as a sensitivity analysis. To compare these 3 models, we relied on the widely applicable information criterion [20] (WAIC). It can be viewed as an approximation of cross-validation [20,29] and lower value indicates better compromise between information and model dimension.

Model goodness-of-fit was assessed by graphical representation of the observed proportion of toxicity versus the predicted probabilities of toxicity for the full CLRM and the PO model, extracted from 1000 samples generated from the posterior predictive distribution of the models.

3. Results

3.1. Descriptive results

The 2048 patients in the 54 studies received a total of 5592 cycles. During each cycle, toxicities of various types occurred, resulting in 9904 adverse events detailed by grade in Table 1.

The Venn diagram in Fig. 1 illustrates the co-occurrences of toxicities in the 5592 reported cycles. The most frequent combinations were: digestive/general disorders/other (n = 520), digestive/general disorders/hematologic (n = 383) and digestive/others (n = 369). Distributions of proportions of co-occurrences can be found in the supplementary Table A1.

3.2. Exploring the PO assumption

Table 2 provides the parameter estimates and their 95% credibility intervals (CI) for the full model. The dose significantly increased the risk of all types of toxicities; dose effect ranged from 2.98 (95%CI [2.27: 3.96]) for cutaneous toxicities to 2.16 (95%CI [1.77: 2.65]) for the 'other' type of toxicities. The PO assumption seemed to be plausible for the dose as the 95% credibility intervals of the $Dose_{Grade\geq 2}$ and $Dose_{Grade\geq 3}$ odds ratios included the null value. Conversely, for general disorders the odds ratio increased by 0.37 (95%CI [0.10: 0.67]) for the risk of grade ≥ 2, with additional 0.67 (95%CI [0.05: 1.23]) for the risk

Table 2
Parameter estimates of the full continuation ratio logit model with their 95% credibility interval for each type of toxicity. Bolded figures correspond to parameters with credibility intervals excluding the null value.

Parameter	Cutaneous	Digestive	General disorder	Hematological	Other
$Intercept_{Grade\geq 1}$	**−9.51 [-10.81; -8.38]**	**−1.79 [-2.16; -1.47]**	**−2.66 [-3.14; -2.31]**	**−7.33 [-8.27; -6.62]**	**−1.93 [-2.37; -1.57]**
$Intercept_{Grade\geq 2}$	**−3.11 [-3.68; -2.45]**	**−3.38 [-3.73; -3.00]**	**−3.02 [-3.38; -2.69]**	**−1.95 [-2.39; -1.58]**	**−2.70 [-3.03; -2.39]**
$Intercept_{Grade\geq 3}$	**−4.85 [-6.43; -3.52]**	**−4.61 [-5.24; -3.93]**	**−4.17 [-4.87; -3.42]**	**−2.52 [-3.04; -2.07]**	**−2.70 [-3.14; -2.32]**
$Dose_{Grade\geq 1}$	**2.98 [2.27; 3.96]**	**2.35 [1.96; 2.72]**	**2.16 [1.77; 2.65]**	**2.87 [2.17; 3.75]**	**2.51 [2.12; 2.98]**
$Dose_{Grade\geq 2}$	0.06 [-0.41; 0.43]	0.22 [-0.06; 0.51]	**0.37 [0.10; 0.67]**	0.30 [-0.01; 0.72]	0.08 [-0.19; 0.38]
$Dose_{Grade\geq 3}$	0.33 [-0.56; 1.41]	**1.37 [0.84; 1.87]**	**0.67 [0.05; 1.23]**	0.03 [-0.37; 0.42]	0.26 [-0.03; 0.58]
$Cycle_{Grade\geq 1}$	**0.37 [0.27; 0.48]**	0.04 [-0.01; 0.11]	**0.07 [0.01; 0.14]**	0.08 [-0.00; 0.17]	**0.18 [0.11; 0.24]**
$Cycle_{Grade\geq 2}$	**−0.16 [-0.33; -0.02]**	−0.08 [-0.17; 0.00]	**−0.08 [-0.17; -0.00]**	−0.02 [-0.12; 0.05]	**−0.09 [-0.17; -0.00]**
$Cycle_{Grade\geq 3}$	0.05 [-0.20; 0.37]	−0.01 [-0.13; 0.11]	−0.08 [-0.26; 0.06]	0.00 [-0.10; 0.12]	−0.09 [-0.20; 0.01]

Table 3
Parameter estimates of the proportional odds ratio logit model with their 95% credibility interval for each type of toxicity. Bolded figures correspond to parameters with credibility intervals excluding the null value.

Parameter	Cutaneous	Digestive	General disorder	Hematological	Other
$Intercept_{Grade\geq 1}$	**−9.73 [-10.75; -8.63]**	**−1.90 [-2.25; -1.55]**	**−2.89 [-3.34; -2.49]**	**−7.47 [-8.71; -6.51]**	**−2.01 [-2.37; -1.65]**
$Intercept_{Grade\geq 2}$	**−3.52 [-3.83; -3.22]**	**−3.36 [-3.53; -3.19]**	**−2.94 [-3.11; -2.78]**	**−1.73 [-1.92; -1.56]**	**−2.87 [-3.02; -2.72]**
$Intercept_{Grade\geq 3}$	**−4.40 [-5.15; -3.72]**	**−3.24 [-3.48; -3.01]**	**−3.71 [-4.02; -3.42]**	**−2.49 [-2.75; -2.25]**	**−2.72 [-2.90; -2.55]**
Dose	**3.11 [2.35; 3.87]**	**2.58 [2.19; 2.95]**	**2.55 [2.12; 3.00]**	**3.13 [2.31; 4.01]**	**2.77 [2.36; 3.17]**
Cycle	**0.33 [0.23; 0.42]**	0.01 [-0.04; 0.06]	0.03 [-0.02; 0.09]	0.06 [-0.01; 0.15]	**0.12 [0.07; 0.17]**

Table 4
Random effects correlation matrix of the full continuation ratio logit model (correlation estimates and their 95% credibility interval).

	Cutaneous	Digestive	General disorder	Hematologic	Others
Cutaneous	1				
Digestive	0.22 [0.15; 0.28]	1			
General disorder	0.05 [-0.01; 0.12]	0.45 [0.40; 0.51]	1		
Hematologic	−0.18 [-0.27; −0.07]	0.28 [0.23; 0.35]	0.26 [0.21; 0.33]	1	
Others	0.01 [-0.06; 0.08]	0.42 [0.36; 0.46]	0.41 [0.36; 0.46]	0.37 [0.32; 0.43]	1

of grade ≥ 3 violating the PO assumption. Finally, this assumption did not hold for digestive toxicity as the odds ratio increased by 1.37 (95%CI [0.84: 1.87]) for the risk of grade≥ 3.

Table 3 contains the PO model parameter estimates and their 95%CI. The results were close to that provided by the full model, which is confirmed by the information criteria $WAIC_{PO} = 31432.10$ vs $WAIC_{full} = 30911.58$. Under the PO model, dose effects ranged from 2.55 (95%CI [2.12: 3.00]) for general disorder toxicities to 3.13 (95%CI [2.31: 4.01]) for the 'other' type toxicities. Assuming a proportional odds model would then lead to under-estimate the dose effect on the risk of digestive grade 3 or more toxicity by 35% ($log_{PO}(\text{Odd ratio}) = 2.58$ instead of $log_{Full}(\text{Odd ratio}) = 3.94$), resulting in a large underestimation of the odds ratio. The estimated odds ratio of grade 3 or more hematological toxicity in patients treated at the MTD compared to patients treated at half the MTD is 3.63 (assuming PO) instead of 7.15. This mis-specification had more limited impact for general disorder. The difference of the same estimated odds ratio for grade 2 or more was small

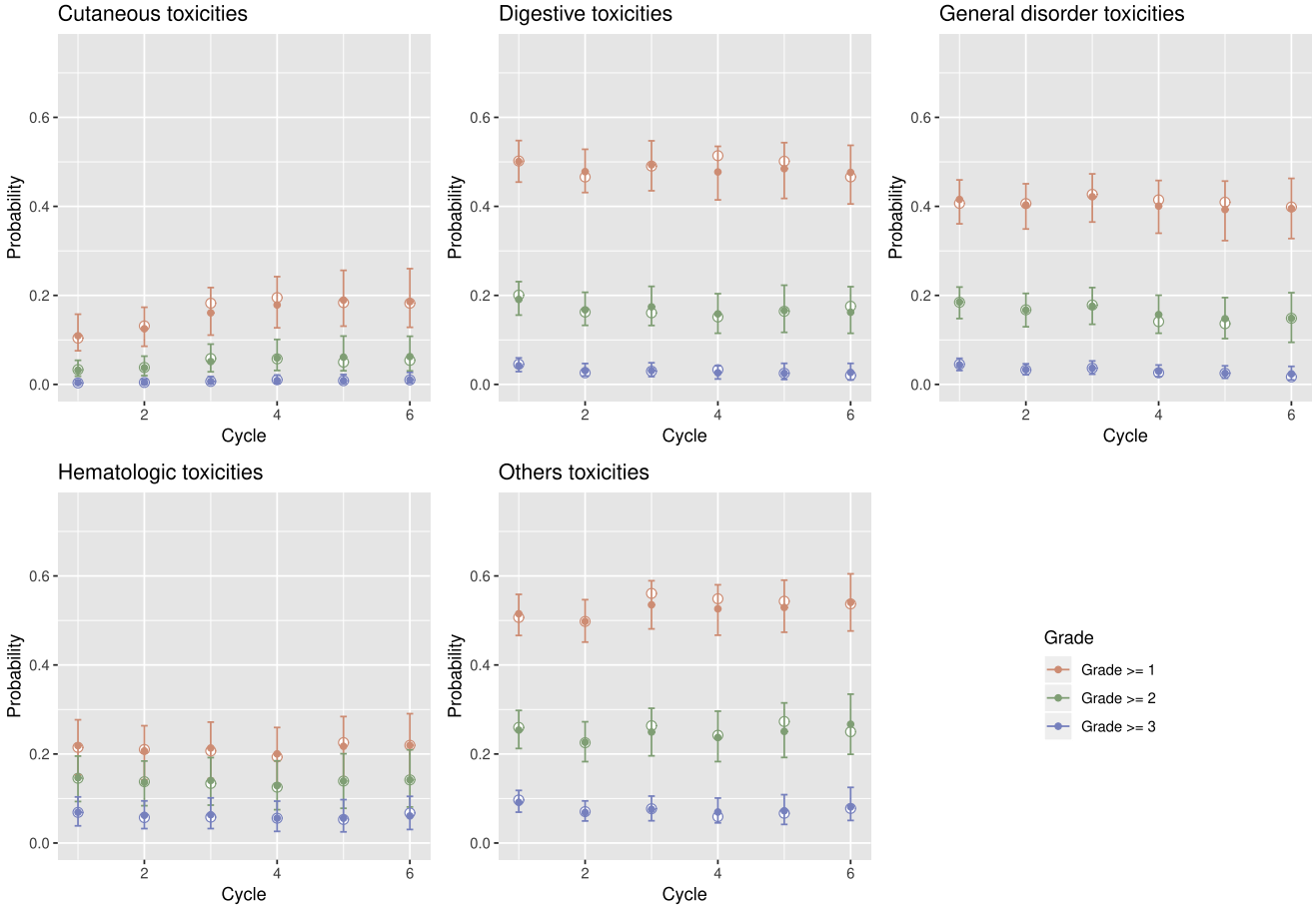

Fig. 2. Observed (empty circle) vs expected conditional probability given the cycle of each type of toxicity at each cycle according to the PO model. The median expected probability (filled circles) and the 95% prediction interval were obtained from 1000 simulations from the posterior predictive distribution of the model.

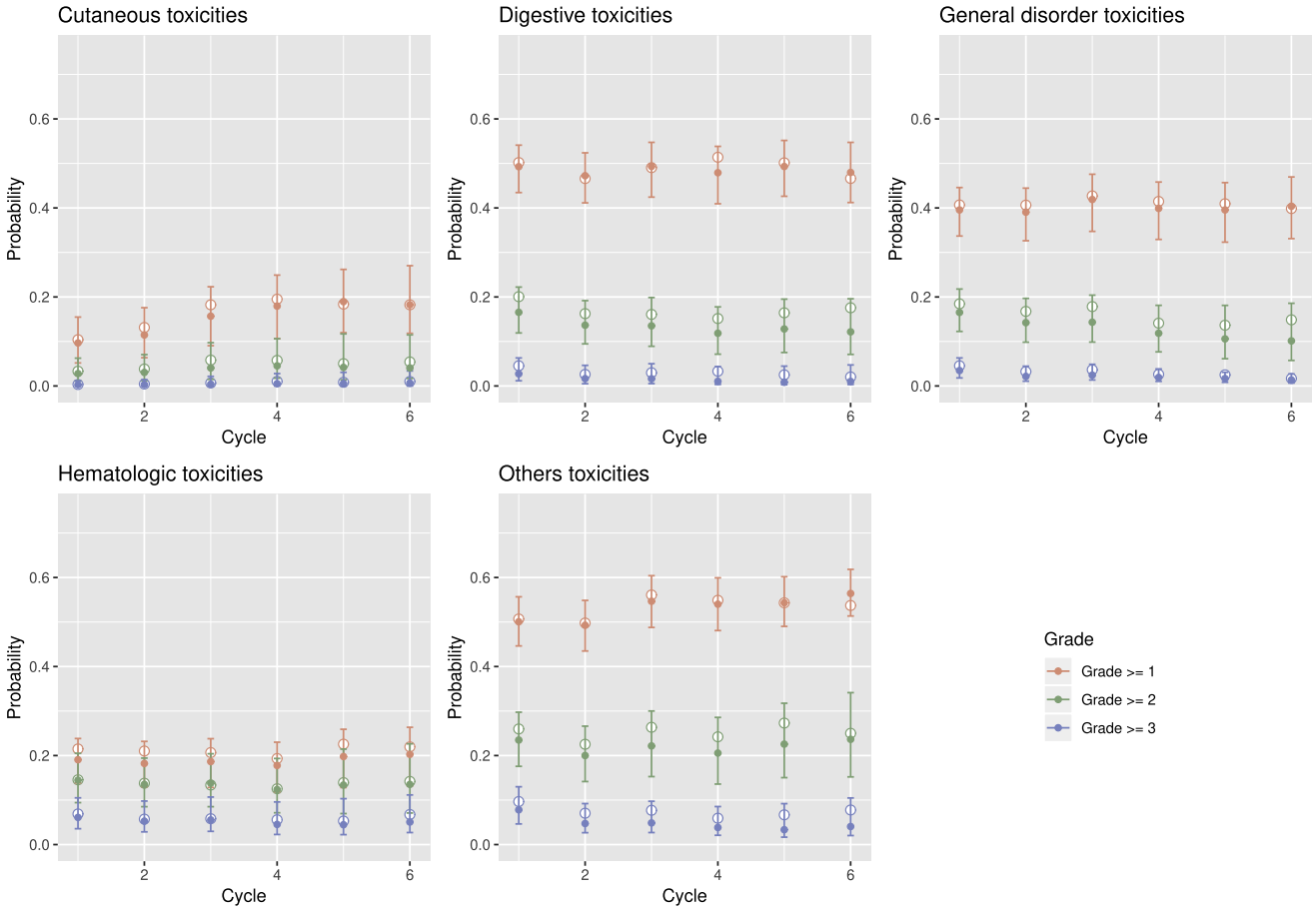

Fig. 3. Observed (empty circle) vs expected conditional probability given the cycle of each type of toxicity at each cycle according to the full model. The median expected probability (filled circles) and the 95% prediction interval were obtained from 1000 simulations from the posterior predictive distribution of the model.

(3.48 instead of 3.44), but it was higher for grade 3 or more (3.58 instead 4.95).

Conversely, additional cycles of treatment did not significantly modify the risk of digestive, general disorder and hematological toxicities, but increased the risk of cutaneous and 'other' toxicities. This is in line with the underlying mechanisms of actions inducing cutaneous rash that is directly related to the drug exposure. Based on the 95% credibility intervals of ζ_{j1}, ζ_{j2}, the PO assumption was plausible for cutaneous and 'other' types of toxicities. Nevertheless, in those later cases, the deviation were of borderline significance as the boundary of intervals for parameters associated with grade 2 or more were 0.02 for the former and −0.00 for the later. Again, the PO assumption would lead to slightly over estimate the cycle effect for grade 2 or higher cutaneous toxicity. One possible explanation for this over-estimation might be the dose reduction which may have been applied to some patients. Indeed, investigators may reduce the dose after several cycles to avoid upcoming toxic side events leading to a systematic bias when we adjusted on the planned dose instead of the actual dose.

The estimated variance-covariance matrix of the random effects of the full CR-model in Table 4 provides some insight on the correlations between the various types of toxicity.

The low to moderate correlations between random effects (from −0.15 to 0.46) suggest that each type of toxicity carries different information, which may also be the consequence of different toxicity profiles according to the investigated agent. This is reassuring that correlation between hematologic and cutaneous toxicity was low as they proceed from different mechanisms. Conversely, general disorders (typically fatigue, mood depressions, pain,…) and digestive toxicity that are commonly associated in clinical practice, were correlated in our data. Of note, the correlations estimated under the PO assumptions were quite similar (cf. Supplementary material Table A2).

3.3. Goodness of fit

Figs. 3 and 2 show that the observed probabilities of toxicity at each cycle were included in the 95% prediction interval drawn from the predictive posterior distributions for both the full and the PO models. The goodness of fit of the model was satisfactory.

4. Discussion

The richness of this large database highlighted some characteristics of the dose-response relationships for different types of toxicities. Our results suggest that the PO assumption may hold in most cases, but this statement cannot be generalized, specifically for digestive and general disorder toxicities. The cycle effect also depends of the type of toxicity as it ranged from no cycle effect for digestive and hematological toxicities, to a moderate effect attenuated for grade 2 or higher of the other types of toxicity.

Those results may have application both at the design and at the analysis level. Assuming proportional odds for the dose effect enables to incorporate the occurrence of intermediate grades of toxicity in the estimate of the risk of toxicity and hence to increase its precision. In particular, the method proposed by Ref. [17] based on PO models appears as a simple and efficient extension of the continual reassessment method (CRM) for ordinal outcomes that may be applied in various situations. Designs based on PO assumption for dose effect may fail for trials applied to a drug which are expected to induce digestive and/or

general disorder toxicities, but they would be more informative than binary CRM designs in the other cases. In case the strict PO assumption appears too strong, an informative prior on the dose parameter in the continuation ratio logit model may be an alternative modeling option. Phase I trials enroll increasing numbers of patients (commonly larger than 100) [30]. More advanced analysis may then be performed to help refining the assessment of the toxicity profile according to the dose and over time. We proposed an analytical tool to explore the dose and the cycle effects on each type of toxicity. In addition, the joint modeling provides estimates of the correlations between the various toxicities, an additional information that is useful for the management of patients during the course of the treatment. One of the statistical issue with the analysis of multiple cycles of treatment relates to the patients who get off-study due to early progressive disease. Follow-up strongly varies across patients. Under the assumption that the risk of early progression is largely independent on the risk of toxicity after adjustment for the dose level, our estimates should not be biased by early drop out. When incorporating repeated cycles of treatment in the analysis, the main limitation is the lack of known model for the relationship between the cumulative drug exposure and the risk of toxicity. As patients are treated at the same dose level over all cycles, the delayed toxicity cannot be disentangled from the cumulative dose effect, if any. Is the toxicity at cycle 3 due to the dose administered at the same cycle or at a previous cycle or is it the consequence of some accumulation? It may depend on the half-life of the compounds, but PK models are often unknown at the time the first in man trials are carried out. This is also the reason why accounting for the actual administered dose is not straightforward. In case of adverse events, some dose reduction may be allowed. How to adjust the model on those dose reductions largely depend on the underlying model for the drug exposure. Would this model be known, our approach could be easily adapted. Our analysis shows that the data generated by phase I trial is richer than the classical binary DLT measured at the first cycle. Analysis of all the collected information is feasible. This may help elaborating new designs with reasonable assumptions for our models to select the optimal dose more reliably.

Declaration of competing interest

All authors declare no conflicts of interests.

Acknowledgement

The authors are indebted to all members of the EORTC-led DLT-TARGETT group, in particular Eliza Rizzo et Jennifer Wong who prepared the data and all the contributors: the NCI, the NCI–C, the EORTC, Cancer research UK, Pfizer, Roch and Sanofi, and all the patients. We thank very much the Cancer Therapy Evaluation Program from the National Cancer Institute (Dr Percy Ivy, NCI) for providing full access to data from several phase I trials sponsored by the NCI.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.conctc.2020.100529.

Funding and support

XP was partly supported by the ITN-Marie Curie IDEAS project; XP and DD were supported by the French NCI (INCA-Optidose program); LC was partly supported by FOCA (Fonds cancer) from Belgium.

Software availability

R and Stan codes for this article are available on the Oncostat team github repository under https://github.com/Oncostat/POP1.

References

[1] Xavier Paoletti, Adélaïde Doussau, Monia Ezzalfani, Elisa Rizzo, Rodolphe Thiébaut, Dose finding with longitudinal data: simpler models, richer outcomes, Stat. Med. 34 (22) (sep 2015) 2983–2998.
[2] Xavier Paoletti, Damien Drubay, Laurence Collette, Dose-finding methods: moving away from the 3 + 3 to include richer outcomes, Clin. Canc. Res. 23 (15) (aug 2017) 3977–3979.
[3] Sophie Postel-Vinay, Carlos Gomez-Roca, L. Rhoda Molife, Bhavesh Anghan, Antonin Levy, Ian Judson, Johann De Bono, Jean-Charles Soria, Stan Kaye, Xavier Paoletti, Phase I trials of molecularly targeted agents: should we pay more attention to late toxicities? J. Clin. Oncol. 29 (13) (may 2011) 1728–1735.
[4] Sophie Postel-Vinay, Laurence Collette, Xavier Paoletti, Elisa Rizzo, Christophe Massard, David Olmos, Camilla Fowst, Bernard Levy, Pierre Mancini, Denis Lacombe, Percy Ivy, Lesley Seymour, Christophe Le Tourneau, Lillian L. Siu, Stan B. Kaye, Jaap Verweij, Jean-Charles Soria, Towards new methods for the determination of dose limiting toxicities and the assessment of the recommended dose for further studies of molecularly targeted agents – dose-limiting toxicity and toxicity assessment recommendation group for early trials of T, Eur. J. Canc. 50 (12) (aug 2014) 2040–2049.
[5] Evaluation of Anticancer Medicinal Products in Man, European Medicines Agency, 2016.
[6] Adélaïde Doussau, Rodolphe Thiébaut, Xavier Paoletti, Dose-finding design using mixed-effect proportional odds model for longitudinal graded toxicity data in phase I oncology clinical trials, Stat. Med. 32 (30) (dec 2013) 5430–5447.
[7] Jun Yin, Xavier Paoletti, Daniel J. Sargent, Sumithra J. Mandrekar, Repeated measures dose-finding design with time-trend detection in the presence of correlated toxicity data, Clin. Trials 14 (6) (dec 2017) 611–620.
[8] Laura L. Fernandes, Jeremy M.G. Taylor, Susan Murray, Adaptive Phase I clinical trial design using Markov models for conditional probability of toxicity, J. Biopharm. Stat. 26 (3) (may 2016) 475–498.
[9] R. Lin, Bayesian optimal interval design with multiple toxicity constraints, Biometrics 74 (4) (dec 2018) 1320–1330. https://doi.org/10.1111/biom.12912.
[10] Graham M. Wheeler, Michael J. Sweeting, Adrian P. Mander, Toxicity-dependent feasibility bounds for the escalation with overdose control approach in phase I cancer trials, Stat. Med. 36 (16) (jul 2017) 2499–2513.
[11] Daniel G. Muenz, Thomas M. Braun, Jeremy Mg Taylor, Modeling adverse event counts in phase I clinical trials of a cytotoxic agent, Clin. Trials (London, England) 15 (4) (aug 2018) 386–397.
[12] Zhengjia Chen, Ye Cui, Taofeek K. Owonikoko, Zhibo Wang, Li Zheng, Ruiyan Luo, Michael Kutner, Fadlo R. Khuri, Jeanne Kowalski, Escalation with overdose control using all toxicities and time to event toxicity data in cancer Phase I clinical trials, Contemp. Clin. Trials 37 (2) (mar 2014) 322–332.
[13] Emily M. Van Meter, Elizabeth Garrett-Mayer, Dipankar Bandyopadhyay, Dose-finding clinical trial design for ordinal toxicity grades using the continuation ratio model: an extension of the continual reassessment method, Clin. Trials: J. Soc. Clin. Trials 9 (3) (jun 2012) 303–313.
[14] Karen Sinclair, Anne Whitehead, A Bayesian approach to dose-finding studies for cancer therapies: incorporating later cycles of therapy, Stat. Med. 33 (15) (jul 2014) 2665–2680.
[15] Pierre Colin, Sandrine Micallef, Maud Delattre, Pierre Mancini, Eric Parent, Towards using a full spectrum of early clinical trial data: a retrospective analysis to compare potential longitudinal categorical models for molecular targeted therapies in oncology, Stat. Med. 34 (22) (sep 2015) 2999–3016.
[16] Juhee Lee, Peter F. Thall, Ji Yuan, Peter Müller, A decision-theoretic phase I-II design for ordinal outcomes in two cycles, Biostatistics (Oxford, England) 17 (2) (apr 2016) 304–319.
[17] M. Emily, Van Meter, Elizabeth Garrett-Mayer, Dipankar Bandyopadhyay, Proportional odds model for dose-finding clinical trial designs with ordinal toxicity grading, Stat. Med. 30 (17) (jul 2011) 2070–2080.
[18] Alan, Agresti and Wiley InterScience (Online Service). *An Introduction to Categorical Data Analysis*, Wiley-Interscience, 2007.
[19] Carlos M. Carvalho, Nicholas G. Polson, James G. Scott, Handling sparsity via the horseshoe, in: David van Dyk, Max Welling (Eds.), Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics, Volume 5 of Proceedings of Machine Learning Research, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA, 2009, pp. 73–80 (PMLR).
[20] Sumio Watanabe, Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory, J. Mach. Learn. Res. 11 (Dec) (2010) 3571–3594.
[21] Christophe Le Tourneau, Albiruni R.A. Razak, Hui K. Gan, Simona Pop, Véronique Diéras, Patricia Tresca, Xavier Paoletti, Heterogeneity in the definition of dose-limiting toxicity in phase i cancer clinical trials of molecularly targeted agents: a review of the literature, Eur. J. Canc. 47 (10) (jul 2011) 1468–1475.
[22] Robert Tibshirani, Regression Shrinkage and Selection via the Lasso, 1996.
[23] Nicholas G. Polson, James G. Scott, Shrink globally, act locally: sparse bayesian regularization and prediction*, in: Bayesian Statistics, vol. 9, Oxford University Press, oct 2011, pp. 501–538.
[24] Jyotishka Datta, Jayanta Ghosh, In search of optimal objective priors for model selection and estimation, in: Current Trends in Bayesian Methodology with Applications, Chapman and Hall/CRC, may 2015, pp. 225–243.
[25] Anindya Bhadra, Jyotishka Datta, Nicholas G. Polson, Brandon T. Willard, Lasso Meets Horseshoe, jun 2017.
[26] Daniel Lewandowski, Dorota Kurowicka, Harry Joe, Generating random correlation matrices based on vines and extended onion method, J. Multivariate Anal. 100 (9) (oct 2009) 1989–2001.

[27] Duane Simon, A.D. Kennedy, Brian J. Pendleton, Roweth Duncan, Hybrid Monte Carlo, Phys. Lett. B 195 (2) (sep 1987) 216–222.
[28] Bob Carpenter, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Brubaker Marcus, Jiqiang Guo, Peter Li, Allen Riddell Stan, A probabilistic programming language, J. Stat. Software 76 (1) (jan 2017) 1–32.
[29] Andrew Gelman, Jessica Hwang, Vehtari Aki, Understanding predictive information criteria for Bayesian models, Stat. Comput. 24 (6) (nov 2014) 997–1016.
[30] Suzanne E. Dahlberg, Geoffrey I. Shapiro, Jeffrey W. Clark, Bruce E. Johnson, Evaluation of statistical designs in phase I expansion cohorts: the dana-farber/harvard cancer center experience, JNCI: J. Natl. Cancer Inst. 106 (7) (2014) jun.