Special Issue Paper

Statistics in Medicine

Received 21 January 2014, Accepted 17 May 2015 Published online 24 June 2015 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6552

Dose finding with longitudinal data: simpler models, richer outcomes

Xavier Paoletti[a*†‡], Adélaïde Doussau,[a,b] Monia Ezzalfani,[a] Elisa Rizzo[c] and Rodolphe Thiébaut[b]

Phase I oncology clinical trials are designed to identify the optimal dose that will be recommended for phase II trials. This dose is typically defined as the dose associated with a certain probability of severe toxicity at cycle 1, although toxicity is repeatedly measured over cycles on an ordinal scale. Recently, a proportional odds mixed-effect model for ordinal outcomes has been proposed to (i) identify the optimal dose accounting for repeated events and (ii) to provide some framework to explore time trend. We compare this approach to a method based on repeated binary variables and to a method based on an under-parameterized model of the dose–time toxicity relationship. We show that repeated binary and ordinal outcomes both improve the accuracy of dose-finding trials in the same proportion; ordinal outcomes are, however, superior to detect time trend even in the presence of nonproportional odds models. Moreover, less parameterized models led to the best operating characteristics. These approaches are illustrated on two dose-finding phase I trials. Integration of repeated measurements is appealing in phase I dose-finding trials. Copyright © 2015 John Wiley & Sons, Ltd.

Keywords: ordinal; binary; mis-specified; continual reassessment method

1. Background

Phase I oncology clinical trials are the first step of clinical development for a new drug. They are designed to evaluate the toxicity profile of the drug and to recommend a dose whose activity will be investigated in further trials. For decades, the fundamental underlying assumption in oncology for cytotoxic agents was ‘more is better‘. According to this assumption, the dose recommended for phase II is based on the dose level corresponding to the maximum tolerated dose (MTD). Because of the high proportion of life-threatening events in this setting, cancer patients are enrolled sequentially in phase I trials at increasing dose levels, starting with a dose level having a low probability of severe toxicity. Treatment is usually administered in cycles that are repeated up to the progression of the disease or occurrence of unacceptable toxicity or patients refusal. Cycle duration commonly lasts 3 to 4 weeks, even in trials with continuous drug administration. The main endpoint is the toxicity induced by the treatment. Severity of toxicity in cancer clinical trials is graded according to the common terminology criteria for adverse events from the National Cancer Institute [1], which ranges from 1 (mild adverse event) to 4 (life-threatening) and 5 (toxic death). It is measured after each cycle of treatment, providing repeated data. Classically, the MTD is a dose associated with a predefined probability (between 20% to 30% [2]) of severe grade 3 or 4 nonhematological toxicity or grade 4 hematological toxicity, called dose-limiting toxicity (DLT) evaluated on the first cycle of treatment. The main outcome is then a single binary variable providing limited information. Two families of methods have been developed to find this dose, sometimes called algorithmic and model-based dose-escalation designs. One of the attractive method for dose finding is the continual reassessment method (CRM or CRML when likelihood inference is used) [3], an adaptive method that relies on models of the risk of toxicity at cycle 1. As for the standard algorithmic method

[a]*INSERM U900, Institut Curie, Paris, France*
[b]*Centre de recherche INSERM U897, & INRIA SISTM & Université de Bordeaux, ISPED, France*
[c]*European Organization for Research and Treatment of Cancer Headquarters, Brussels, Belgium*
**Correspondence to: Xavier Paoletti, Gustave Roussy, Biostatistics and Epidemiology Department, 114 rue Ed Vaillant 94805 Villejuif cedex, France.*
†*E-mail: xavier.paoletti@gustaveroussy.fr*
‡*Current address: Biostatistics and Epidemiology department, INSERM U1018, Gustave Roussy, Villejuif France.*

(so-called 3 + 3), any information beyond cycle 1 is discarded. Alternatively, the accelerated titration design, a commonly used extension of the 3 + 3 [4], is one of the only methods that proposes the use of toxicity data from subsequent cycles in the design. Nevertheless, after the first DLT, the design is similar to the 3 + 3, and information beyond cycle 1 is not formally analyzed to identify the MTD.

New classes of molecules, the targeted agents, which first appeared in the late 1990s, raise specific issues compared with cytotoxic agents: (i) the maximum tolerated dose may not be the best choice for the recommended dose, because the hypothesis of an increasing relationship between dose and activity may no longer be appropriate [5]; (ii) the definition of dose-limiting toxicities may also include mild to moderate toxicity, because the toxicity profile is different from what is commonly observed with cytotoxic agents [6]; (iii) the risk of cumulative or late toxic side events may considerably limit the chance of success of these agents as they are administered over long periods (even until disease progression for treatment of advance stages), while cytotoxic were commonly administered for a maximum of six or eight cycles. Due to these specificities, the definition of the MTD for the targeted agents has then been challenged by the DLT-TARGETT task force led by the European Organization for Research and Treatment of Cancer (EORTC) [7, 8]. One of the main conclusions highlighted the importance of analyzing data collected at all cycles in identifying the MTD. The authors also stressed the need to fine-tune the assessment of the MTD in an expansion cohort.

When the DLTs can occur at different treatment cycles, several definitions of the MTD can be drawn. One possible measure is the time to DLT first occurrence. The MTD is then defined as a dose associated with a predefined risk of *cumulative* toxicity over a given period and the method of choice is the time to event CRM (tite-CRM) [9]. An alternative measure is the risk of DLT at each cycle. The MTD is then the dose associated with a predefined risk of DLT *per cycle*. The present study focuses on this last setting. A practical dose-finding approach derived from the CRML and based on the risk of toxicity at each cycle (the proportional odds mixed-effect regression model (POMM)-CRML) has been proposed by Doussau *et al.* [10]. A proportional odds (PO) relationship is assumed to model longitudinal ordinal outcomes (mild, moderate, or severe toxicity). The two main advantages of using longitudinal data are: (i) to substantially improve the ability to identify the MTD in the absence of strong intra-patient correlation; (ii) to set a framework for estimating the probability of toxicity over time, which would suggest that there is a risk of late or cumulative effects. However, proportional odds assumption has been showed to be of limited interest to conduct trials based on cycle 1 only [11, 12]. Furthermore, our previous work showed the difficulty to fit proportional odds mixed-effect models to scarce longitudinal data.

In this communication, we explore the interest of modeling longitudinal ordinal outcomes compared with longitudinal binary outcomes and the impact of using under-specified working models to estimate the probability of toxicity. Simpler under-parameterized odds proportional models may help tackle the estimation difficulties and improve the operating characteristics of the dose-finding approach. In Section 2, we introduce two motivating examples. In Section 3, we present the general approach for the use of longitudinal ordinal or binary toxicity data. In Section 4, we compare the various approaches using either binary or ordinal data as well as several models with special attention to the detection of a time trend. Finally, in Section 5, we apply the methods to the two examples.

2. Motivating examples

In the two phase I clinical trials of targeted agents used as motivating examples, data collected after cycle 1 were not formally included in the process to recommend a dose for phase II. The impact of this complementary information is investigated in Section 5.

2.1. The ITCC/erlotinib-RT trial

The European consortium for innovative therapies for children with cancer (ITCC) carried out a phase I trial of erlotinib, a tyrosine kinase inhibitor, in combination with radiotherapy in children with glioblastoma [13]. An adaptation of the CRML [14] was used to identify the dose associated with a 20% probability of DLT. The DLT assessment period was taken over the first two cycles (6 weeks) of treatment. Twenty children were evaluated at three increasing doses of erlotinib ranging from 75 to 125mg/m^2. Two DLTs were observed: fatal grade 5 seizures, and grade 3 skin rash and pruritus. The probability of DLT at $d_3 = 125$mg/m^2 after all patients had been included was 16% (95% CI: 4–45%), and this dose was recommended for phase II studies. A total of 96 cycles were delivered to 20 patients; 12 children completed six cycles of treatment. Six children (26 cycles) received 75mg/m^2, six children (34 cycles)

received 100mg/m^2, and eight children (36 cycles) received 125mg/m^2. Nineteen cases of grade 2 toxicity and seven cases of grade 3 to 5 toxicity (in six patients) were recorded during the first six cycles of treatment, including six severe toxicity after the first cycle.

2.2. The EORTC/R-Viscum Trial

The EORTC carried out a phase I trial of intravenous aviscumine, in adult patients with solid tumors [15]. The CRML was used to identify the dose associated with a 20% probability of DLT during the first 3 weeks of treatment [16]. A total of 41 patients were evaluated at 14 increasing doses ranging from 10 to 6400ng/kg. Four DLTs were observed: one case of fatigue and three cases of hepatitis. The probability of DLT at $d_{13} = 5600$ng/kg after all patients had been included was 16% (95% CI: 7–37%), which was the dose recommended for phase II studies. A total of 97 cycles were administered (94 cycles 1 to 6); three patients completed six cycles of treatment. The worst grades were grade 2 in 34 cycles and grade 3 in seven cycles, including two severe toxicity that occurred after the first cycle.

3. Methods for longitudinal data

3.1. Notations

Let us assume that n patients are to be sequentially enrolled in a dose-finding trial with K dose levels, $d_1, \dots, d_K$, d_k representing the dose at level k. Considering that a patient i is treated at the same dose throughout the trial at all cycles $j = 1, \dots, J$, we denote $X_i = X_{i1} = \dots = X_{iJ}$ the dose level administered and t_{ij} the times associated with day 1 of each cycle j. Let Y_{ij} denote an ordinal variable with three levels, representing the severity of the worst adverse events occurring at cycle j of treatment. The outcome Y_{ij} takes value 1 if no toxicity or grade 1 toxicity is observed, 2 for moderate grade 2 toxicity, and 3 for severe toxicity greater than grade 3. The severe toxicity at cycle 1, $Y_{i1} = 3$, then corresponds to the usual definition of DLT. Let $p_{jg}(d_k, t_{ij}) = \Pr(Y_{ij} = g | X_i = d_k, t_{ij})$ be the probabilities of outcome g, $g = 1$ to 3 at cycle j and dose d_k.

Definition of the MTD depends on the type of endpoint considered. In the usual framework, this is the dose with probability of toxicity at cycle 1 closest to the target τ. If time to event data are considered, this is the dose associated with a cumulative risk of DLT over the predefined period of DLT assessment (for instance, six cycles). With longitudinal data, the MTD is the dose with a probability of DLT per cycle closest to τ and hence verifying $\text{argmin}_{d_k} |\Pr(Y_{ij} = 3 | X_i = d_k, t_{ij}) - \tau|$ in the absence of a time effect. If the risk of DLT is modified with cumulative treatment, then we assume that the dose recommended for phase 2 cannot be defined using toxicity as unique endpoint. Of note, the MTD definition does not integrate moderate toxicity.

3.2. General model and estimation

Doussau *et al*. (2013) estimated the probabilities of graded toxicity using the POMM [17, 18]. The risk of toxicity is assumed to be subject specific, and a random intercept u_i accounts for the expected inter-patient variability for the risk of DLT at a dose. Let us denote $q_{jg}(d_k, t_{ij}) = P(Y_{ij} \leqslant g | X_i = d_k, t_{ij})) = \sum_{\ell=1}^{g} p_{j\ell}(d_k, t_{ij})$. As $q_3 = 1$, the cumulative logits can be described with two functions:

$$\text{logit}\, q_{jg}\left(d_k, t_{ij}\right) = \text{logit}\left(P\left(Y_{ij} \leqslant g | X_{ij} = d_k, t_{ij}\right)\right) = \alpha_g - \beta_1 d_k - h\left(t_{ij}\right) - u_i, g = 1, 2 \tag{1}$$

where u_i follows a normal density f of variance σ_0^2 , $u_i \sim \mathcal{N}(0, \sigma_0^2)$, and h is some function of time that takes value 0 at t_{i1}. We denote θ the vector of all parameters. As patients are assumed to be repeatedly treated at the same dose level, delayed or late dose effects are confounded with cumulative effect. Model (1) can be rewritten in terms of X_{ij} the allocated dose at each cycle:

$$q_{jg} = \alpha_g - \beta_1 x_{i1} - h\left(x_{ij}\right) - u_i, j = 2 \dots J \tag{2}$$

where h captures now the cumulative dose effect. For a given agent, functional form of h should depend upon the pharmacological properties of the agent; for instance, the risk of toxicity may be proportional to the drug exposure at a given cycle and h would be an expression of the clearance. Such model would be quite difficult to specify in first-in-man trials, but it may be extremely useful for dose-finding trials carried out in sub-populations such as in children, where h may be known prior to the trial initiation. In

the following, we stick to the formulation (1). In theory, h may take any form, bearing in mind that less than 10% of the patients receive more than six cycles and only very simple relationships can be explored. The probabilities of severe and of moderate or severe toxicity respectively are monotonically increasing functions of the dose and are assumed to be related by a proportional odds model.

Given the observations of the dose and event outcomes at a given time point of the trial, (x_i, y_{ij}, t_{ij}), the likelihood for the parameter vector θ, is

$$L\left(y_{ij}, x_i, t_{ij}|\theta\right) = \prod_{i,j} \left(p_{1j}\left(x_i, t_{ij}|u_i\right)\right)^{I_{y_{ij}=1}} \times \left(p_{2j}\left(x_i, t_{ij}|u_i\right)\right)^{I_{y_{ij}=2}} \times \left(p_{3j}\left(x_i, t_{ij}|u_i\right)\right)^{I_{y_{ij}=3}} \times f\left(u_i, \sigma_0^2\right) \quad (3)$$

where $I_{[y_{ij}=g]}$ takes value 1 if $Y_{ij} = g$ and 0 otherwise. As the random effect is unknown, evaluation of L must integrate the random effect distribution f. No closed form is available, and maximization is obtained using Laplace approximation and adaptive Gauss–Hermite quadrature.

Fitting such a model to scarce data is challenging. Complex models are usually tackled with the help of Bayesian inference and slightly informative priors [19]. Alternatively, in the spirit of the CRML, simpler 'working' models can be explored. Shen and O'Quigley [20], followed by Cheung *et al.* [21] showed the good asymptotic properties of the CRML under model mis-specification. In particular, the recommended dose converges to the true MTD, and the estimate of the risk of DLT converges to its true value when the sample size goes to infinity under some constraints.

The key point is that adaptive sampling at the best current estimate of the MTD entails accumulation of information on a very limited number of doses that in turn represent the main contribution to the likelihood of the model, making local model goodness-of-fit sufficient. Reasons for not working with a richer CRML model are outlined in several contributions [20–22]. Using the same reasoning, we propose simplified models to conduct the trial using longitudinal data.

3.3. *Simplified models for ordinal endpoint*

Doussau *et al.* [10] used model (4) for conducting the patients' allocation and model (5) that assumed a simple linear function of time $h(t_{ij}) = \beta_2 \times t_{ij}$ for estimating and testing a time effect after completion of the trial:

$$\text{logit}\, q_{jg}\left(d_k, t_{ij}\right) = \alpha_g - \beta_1 \times d_k - u_i \quad (4)$$

$$\text{logit}\, q_{jg}\left(d_k, t_{ij}\right) = \alpha_g - \beta_1 \times d_k - \beta_2 \times t_{ij} - u_i \quad (5)$$

where $u_i \sim \mathcal{N}(0, \sigma_0^2)$, $t_{i1} = 0$ and $g = 1, 2$. In the absence of time effect, the cumulative probability per cycle, $q_{jg}(d_k, t_{ij})$ is denoted $q_g(d_k)$.

This model is more flexible than the class of single-parameter models developed with the CRML. We will then compare this model with a reparameterized model where the parameter β_1 is supposed a known constant, β_c.

3.4. *Simple models for the binary endpoint*

In the previous section, the target dose was defined in terms of the risk of severe toxicity for an ordinal variable. We now consider the case of an binary outcome that denotes severe toxicity. We then estimate the risk of severe toxicity at each cycle. Model (6) is used for conducting the patients' allocation, and model (7) provides estimate and test of the time effect after completion of the trial.

$$\text{logit}\, q_3\left(d_k, t_{ij}\right) = \alpha_2' - \beta_1' \times d_k - u_i \quad (6)$$

$$\text{logit}\, q_{j3}\left(d_k, t_{ij}\right) = \alpha_2' - \beta_1' \times d_k - \beta_2' \times t_{ij} - u_i \quad (7)$$

where $u_i \sim \mathcal{N}(0, \sigma_0'^2)$ and $t_{i1} = 0$. As previously, β_1' can be either estimated from the data or taken as a known value $\beta_1' = \beta_c$.

3.5. Dose escalation procedure

The general dose-escalation procedure is similar whether outcomes are ordinal or binary. Following the principle of adaptive design for oncology dose-finding trials, estimates $\hat{p}_3(d_k)$ obtained either from models (4) or (6) are used to conduct dose allocation. The maximum likelihood arises at the boundary of the parameter space provided at least one outcome of each category g has been observed and measurements have been collected over more than one cycle. This is generally called the heterogeneity requirement. The design then consists of two steps with a run-in that is driven by using a 'classic' CRML based on data from cycle 1.

Patients are sequentially enrolled in the trial starting at the lowest dose. After heterogeneity in the outcomes has been observed, the decision criteria is to select the dose d_k minimizing $|\hat{p}_3(d_k) - \tau|$. A new patient is included when the previous patients have completed at least one cycle of treatment. The overall duration of the trial is therefore not altered compared with the use of cycle 1 only; inclusions can be grouped. The algorithm can be described as follows:

(1) Run-in: enroll patients according to the CRML based on the first cycle only until sufficient data have been collected to fit a dose-time-toxicity model.
(2) Before each new inclusion,
    (a) fit a mixed-effects model to all collected data, that is, to the outcomes at all cycles for all patients previously included available at the time of the new inclusion, which in turn provides estimates of the probability of toxicity for mean patient $u_i = 0$.
    (b) evaluate the decision criteria and identify the dose for which the estimate of the risk of severe toxicity per cycle is closest to the target τ;
    (c) the new patient is treated at this current recommended dose;
(3) The trial is terminated when the maximum number of patients have been treated or after certain stopping rules have been reached [23, 24].

Upon completion of the trial, the time trend can be investigated from models (5) or (7), which is tested with likelihood ratio test. In the absence of a significant time effect, the dose closest to the target is recommended. At this dose, the accuracy of the estimate of the probability of severe toxicity is estimated using the delta method. In the following, we denote the two CRM designs based on longitudinal ordinal data and longitudinal binary data as POMM-CRML and based on the logistic mixed effect model (LMM-CRML).

3.6. Illustrating example

In a step-by-step simulation, we illustrate the process of a prospective trial carried out using either a proportional odds model (POMM-CRML) for repeated ordinal data or a logistic model where the dose parameter is offset (LMM-CRML$_{\bar{d}}$) for repeated binary data; a maximum of six cycles was assumed. The target per cycle probability of toxicity was 0.25. A latent variable, z_{ij}, has been generated from the uniform distribution over (0,1) for each patient at each cycle that represents the sensitivity of the patient to the risk of toxicity at any dose level [25]. This variable has then been categorized either as an ordinal variable with three modalities or as a binary variable, according to which of the two methods we applied. The same data were then used for both approaches even if the doses' allocation was different between the two approaches. For instance, suppose that the latent variable $z_{71} = 0.38$ describes subject 7 at cycle 1; suppose that following the POMM-CRML, this subject is allocated at $x_7 = d_3$ with $p_3(d_3) = 0.20$ and $p_2 = 1 - q_1(d_3) = 0.42$; then subject 7 would have grade 2 toxicity. Conversely, suppose that with the LMM-CRML$_{\bar{d}}$, patient 7 would be allocated to $x_7 = d_4$ with $p_3(d_4) = 0.45$; then this same subject 7 would have had grade 3 toxicity.

The history of this simulation is given in Table I. A total of 16 subjects are presented. We took the cycle as a unit of time and we assumed that a new patient was enrolled after each new cycle. Therefore, time equals the number of simulated patients. At each time point, up to six new observations were collected. For instance, after time 3, three patients had been enrolled who where followed up for three, two, and one cycles, respectively. Patient 2 went off study after grade 3 toxicity and had no data after cycle 1. Black and gray colors alternate to facilitate the visualization of observations at the same time points.

Escalation stage consisted in a sequence of one patient at each of the three first dose levels. Then a first severe toxicity occurred.

Table I. Simulation of a trial history with the POMM-CRML and LMM-CRML$_{\tilde{d}}$.

	POMM-CRML										LMM-CRML$_{\tilde{d}}$								
Time	pt i	x_i	y_{i1}	y_{i2}	y_{i3}	y_{i4}	y_{i5}	y_{i6}	(α_1, α_2)	β_1	σ_0^2	x_i	y_{i1}	y_{i2}	y_{i3}	y_{i4}	y_{i5}	y_{i6}	α_1	σ_0^2
1	1	d_1	1	1	2	1	1	1	-	-	-	d_1	0	0	0	0	0	0	-	-
2	2	d_2	1	1	1	1	1	1	-	-	-	d_2	0	0	0	0	0	0	-	-
3	3	d_3	3						-	-	-	d_3	1						-0.88	-
4	4	d_3	1	1	1	1	1	1	-	-	-	d_2	0	0	0	0	1		-1.37	-
5	5	d_3	2	1	3				-	-	-	d_3	0	0	1				-1.97	-
6	6	d_3	1	1	1	1	1	1	(-2.0, -2.83)	-*	-	d_3	0	0	0	0	0	0	-4.22	-
7	7	d_4	1	2	3				(-4.49, -5.44)	5.37	14.1	d_4	0	0	1				-2.32	-
8	8	d_4	1	3					(-3.42, -4.48)	4.31	5.6	d_3	0	0	0	0	0	0	-2.70	-
9	9	d_4	3						(-3.35, -4.04)	6.31	1.5	d_4	1						-2.27	-
10	10	d_3	1	1	1	3			(-3.68, -4.36)	6.73	1.59	d_3	0	0	0	1			-2.42	-
11	11	d_3	1	3					(-4.15, -4.83)	7.39	1.83	d_4	0	1					-2.64	-
12	12	d_3	1	3					(-3.57, -4.11)	6.41	1.31	d_4	0	1					-2.87	-
13	13	d_3	1	1	1	1	1	1	(-2.68, -3.05)	4.89	0.34	d_4	0	0	0	0	0	0	-2.62	-
14	14	d_3	3						(-2.54, -2.89)	4.83	0.60	d_4	1						-2.60	0.08
15	15	d_3	2	2	3				(-2.47, -2.98)	4.96	0.78	d_4	1						-2.58	0.97
16	16	d_3	3						(-2.35, -3.01)	5.46	2.27	d_4	1						-2.50	4.05

*POMM could not be estimated and a model with an offset on the slope was fit instead. POMM, proportional odds mixed effect; CRML, continual reassessment method with likelihood reference. regression model

POMM-CRML. After three patients, the model could not be fitted and despite the first DLT, dose level 3 was maintained up to time 6. A total of six cycles of follow-up had then been observed for patients 1 and, respectively, 5, 4, 3, and 2 for subsequent patients. Fitting the POMM on the 18 first observations (six patients) was possible only if the dose effect was offset. Estimates of the risk of severe toxicity for a patient $u_7 = 0$ were as follows: 0.01, 0.04, 0.11, 0.25, 0.41, and 0.64. Based on these estimates, d_4 was selected as the best current MTD. At time 7, four additional observations were available, and the POMM could be fitted. The time-dose-toxicity relationship was updated after each new observation was collected. After the 16 subjects had been included and followed up to six cycles or first severe toxicity, the final estimates of the risk of severe toxicity for a patient $u_i = 0$ were 0, 0.02, 0.20, 0.69, 0.92, and 0.99; the risk of moderate toxicity was low: 0, 0.01, 0.13, 0.12, 0.04, and 0; dose level 3 was recommended for phase II.

LMM-CRML$_{\bar{d}}$. The same simulation was conducted with the LMM-CRML$_{\bar{d}}$ with β_1 fixed to $\beta_c = 2.42$. Despite the same outcomes as in the previous simulation were used, dose allocation rapidly differed and $x_4 = d_2$. After subject 10, dose allocation settled down at d_4 that was eventually identified as the MTD. Estimate of the variance of the random intercept was almost 0 until 14 subjects had been included, which illustrates the lack of reliability of such estimates with very low sample sizes. The final estimates of the risk of severe toxicity were as follows: 0.01, 0.05, 0.15, 0.31, 0.48, and 0.71, which was markedly different from the estimate obtained using POMM. Interestingly, had we fit a more flexible logistic model on the final data set, we would have obtained the following estimates: $\beta_1 = 8.1$, $\alpha_1 = -5.2$, and $\sigma_0^2 = 7.3$, and the risk of severe toxicity for a patient $u_i = 0$ would then have been as follows: 0, 0, 0.06, 0.62, 0.85, and 0.99.

3.7. Time-to-event CRM

Longitudinal analysis is contrasted with the tite-CRM that has been described in length elsewhere [9,26]. The outcome of interest is the first severe toxicity during a predefined follow-up period. Consider that the binary outcome Y_i denoting DLT is now measured over a period T corresponding to several cycles of treatment. At a given time point $t < T$ of the trial, Y_i may be viewed as censored if no DLT has been observed. Cheung and Chappell proposed to extend the CRM by considering a model of the dose-toxicity relation weighted by each patient's individual follow-up. The authors investigated a simple linear weight function $w(t, T) = \frac{t}{T}$ that assumes that the hazard of DLT is uniform over T. An adaptive weighting scheme independent of the dose effect is an alternative option, with weights depending on the timing of events in previous patients. This weighting scheme is preferable if the censoring rate is high [27]. Design and analysis are then very similar to those of CRM, except that incomplete data can be used and new patients can be enrolled even when some patients are still on study. In fact, the tite-CRM would lead to exactly the same inference as a classic CRM approach that would use a DLT assessment period of T without early dropout for reasons other than toxicity.

4. Operating characteristics

A simulation study was conducted to assess the operating characteristics of using simplified models with binary data to identify the MTD and to detect time trends compared with the use of ordinal data. As a benchmark, we also ran the CRML based on the first cycle only. We focused on the distribution of the final recommended dose level, the distribution of the allocated doses (that is the risk of under and over dosing of the patients), and the ability to detect a time trend when it exists.

4.1. Simulation setting

The following approaches based on DLT at cycle 1 or based on repeated ordinal or binary outcomes were compared:

CRML: using likelihood inference on the binary outcome observed on cycle 1. A logistic working model with the dose parameter set was chosen. Following [22], this slope parameter was set to $\beta_c = 2.42$ so that if the risk of DLT at a given dose was 0.25, then the predicted value at the next higher dose was about 0.40 and at the next lower dose was 0.11. Before any heterogeneity among the outcomes had been observed, dose levels were escalated after each patient had tolerated the previous dose level.

LMM-CRML: CRML was used for the run-in. Then, once heterogeneity in the observations had been reached, model (6) was fitted after each new observation. Two parameterizations were compared: the dose parameter β_1 was either freely estimated or it was fixed to $\beta_1 = \beta_c = 2.42$; note that β_c is different from the true parameters β_1 in all three scenarios.

POMM-CRML: CRML was used for the run-in. Model (4) was then fitted after each new observation. Systematic diagnostic tools were applied to detect estimation issues; apart from nonconvergence indicated by the optimization criteria, standard deviation of the estimates for the parameters had to be no more than 10-fold higher than the true parameter. Two parameterizations were compared where dose parameter β_1 was either estimated or set to $\beta_c = 2.42$.

Finally, we also report the possibility of running a trial with the CRML based on cycle 1 only, to collect the data at all cycles and to retrospectively analyze the longitudinal data after trial completion. This would correspond to a phase I trial targeting the MTD defined on the first cycle only with a recommended dose for phase 2 defined on all cycles. We will compare the performances of this retrospective assessment of longitudinal data with those of a prospective use.

Data generation. We assumed that a maximum of $K = 6$ increasing dose levels taking the values 4.1, 4.8, 5.3, 5.7, 6.0, and 6.4 could be escalated. These doses correspond to the log of a dose-toxicity relationship used in [28]. Ordinal toxicity at each cycle of treatment $j = 1, \ldots, J$ was generated from a proportional odds mixed-effect model that related the risk of each category of toxicity to the dose and the cycle; cycle was included as a categorical variable associated with $J-1$ parameters. Three dose-toxicity relationships denoted A, B, and C were explored, assuming either steep or smooth dose slope as this is a key driver of the performance of any methods [22]. The MTDs (dose at which the risk of DLT per cycle is closest to the target, $\tau = 0.25$) were d_4, d_6, and d_2 for the scenarios A, B, and C, respectively. As an illustration, the true risks of toxicity at cycle 1 for an average patient $u_i = 0$ in scenario A is plotted in Figure 1 (plain lines). The figure also displays the 'working' model that we built by fixing the slope parameter β_1 to $\beta_c = 2.42$ (dashed line). As it is unlikely that the true shape of the dose-toxicity curve is known in advance in first-in human trials, none of the true relationships corresponds to the 'working' model. In scenarios B and C, all data were generated without time effect. Conversely, in scenario A, in addition to the case where risk of toxicity was stable in time, four relations between time and the risk of toxicity were investigated that were characterized by: (*i*) a log-linear time trend (odd ratio (OR)=1.3 per additional cycle); (*ii*) a piecewise time trend with no effect before cycle 3 and increased effect thereafter (OR=1.3 at cycles 4 to 6 compared with cycle 1); and (*iii* and *iv*) nonproportional odds time-toxicity relations. All parameters used for generating data are listed in Table II.

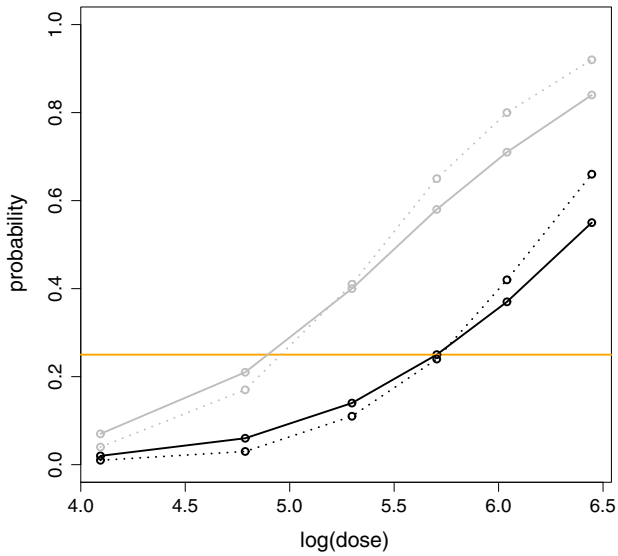

Figure 1. True relations between dose and risk of severe (dark line) and moderate or severe (gray line) toxicity in the absence of time effect for scenario A (plain lines) and the related mis-specified 'working' models (dashed lines). The horizontal line represents the target percentile.

Table II. True parameters' values. In scenario $A(ii)$, cycle is treated as categorical variable with five binary indicators. Scenarios $A(iii)$ and $A(iv)$ break the proportional odds hypothesis and β_2 is provided for the two outcome categories.

Scenario	α_1	α_2	β_1	β_2	σ^2
A	9.9	11.28	1.78	0	0.5
B	9.9	11.28	1.58	0	0.5
C	9.9	11.28	2.11	0	0.5
$A(i)$	9.9	11.28	1.78	0.26	0.5
$A(ii)$	9.9	11.28	1.78	(0,0,0.26,0.19,0.21)	0.5
$A(iii)$	9.9	11.28	1.78	$\beta_{2,g=2} = 0.26$ $\beta_{2,g=1} = 0.13$	0.5
$A(iv)$	9.9	11.28	1.78	$\beta_{2,g=2} = 0.13$ $\beta_{2,g=1} = 0.26$	0.5

Patients could receive a maximum of six cycles of treatment. We further assumed that conditional on a random intercept; the risks of toxicity at the various cycles were independent for a given patient. We did not consider random dose or random cycle effects. The value of the random intercept, u_i, was drawn from a normal distribution with variance $\sigma_0^2 = 0.5$.

A simulated trial was completed when 30 patients had been sequentially included and evaluated. We assumed that a new patient entered the trial after each cycle. Furthermore, the patient's follow-up ended after onset of the first DLT or after completion of the six cycles.

Simulations were carried out with an R-program and packages `ordinal` for fitting POMM and `lme4` for fitting logistic model on repeated binary data. Each simulated trial was repeated 1000 times, providing distributions of the parameters used for evaluation.

4.2. Results

Absence of a cycle effect. Distributions of the final recommendations for the various models and designs are listed in Table III. The mean number of observations was between 100 and 120, corresponding to an average of 3.5 to 4 cycles per patient. Fitting POMM at some point in the simulated trial could be performed in 85% to 100% of the simulations, except in scenario B. In the absence of a time effect, for all scenarios, analysis of longitudinal data considerably improved identification of the correct MTD. A greater improvement was obtained for models in which the slope was not estimated from the data. The absolute gain in accuracy (percentage of correct selection) was as high as 16% compared with the CRML in scenario B. Conversely using ordinal longitudinal data did not significantly increase the accuracy rate compared with using simple longitudinal binary data. This difference is in line with previous reports that investigated the added value of ordinal outcomes on the first cycle only [11, 12] and concluded to very limited gains.

When using LMM-CRML in scenario B (MTD= d_6), the logistic model with four parameters (the two intercepts, the dose effect, and the variance of the random intercept) could not be fitted in a large proportion of simulations. This can be partly explained by the fact that if no severe toxicity was observed before d_6, events were concentrated at that level as no higher dose levels could be explored, and the model was not identifiable.

In scenario C, in which all doses were associated with high risks of severe toxicity, the probability of picking up the right dose was as high as 80% or 90% after 30 patients, which indicates that an early stopping rule would be needed.

Lastly, retrospective analysis of longitudinal data collected after the CRML based on DLT observed during first cycle demonstrated fairly similar performances, except that convergence issues were more frequent in case the model was constraint with a constant dose parameter. This reinforces the idea that significant improvement can be obtained even without modifying the process of a trial, by just analyzing all the collected data as long as adaptive design is used.

Distribution of dose allocation was in line with the final recommendations; the risk of under-dosing or over-dosing was reduced when we accounted for repeated data compared with using cycle 1 only (Table IV). Analysis of ordinal data also contributed to enroll a higher frequency of patients at doses adjacent to the MTD than the analysis of binary data. The risk of over-dosing was somehow controlled;

Table III. Distribution of final recommendations after 1000 simulations. POMM-CRML$_{\bar{d}}$ and LMM-CRML$_{\bar{d}}$ denote the POMM-CRML and the LMM-CRML respectively in which the dose parameter is set up and not estimated. P/R indicates whether longitudinal data were used prospectively (P) or retrospectively (R) after completion of data collection using a classic CRM. ♯ cy. stands for number of cycles and CV indicates the percentage of simulations that converged; $p_3(d_k)$ and $p_{2+}(d_k)$ are the probabilities of severe and moderate or severe toxicity per cycle in the absence of a cycle effect for an 'average' patient $u_i = 0$. Bold entries indicates the true target dose.

Scenario	Model	P/R	d_1	d_2	d_3	d_4	d_5	d_6	♯ cy.	CV
	$p_3(d_k)$		0.02	0.06	0.14	**.24**	0.35	0.53		
	$p_{2+}(d_k)$		0.09	0.21	0.40	**0.59**	0.71	0.82		
A	CRML	P	0	4	28	**48**	19	1	30	100
	POMM-CRML$_{\bar{d}}$	P	0	0	16	**73**	11	0	107	89
		R	0	0	20	**71**	9	0	107	84
	POMM-CRML	P	0	1	23	**61**	14	0	117	96
		R	0	2	22	**57**	20	0	110	100
	LMM-CRML$_{\bar{d}}$	P	0	0	17	**71**	12	1	110	100
		R	0	0	20	**68**	12	0	110	100
	LMM-CRML	P	0	1	23	**60**	13	3	110	98
		R	0	0	27	**63**	10	0	110	100
	$p_3(d_k)$		0.01	0.02	0.05	0.09	0.14	**0.24**		
	$p_{2+}(d_k)$		0.03	0.09	0.18	0.29	0.40	**0.55**		
B	CRML	P	0	0	2	6	32	**60**	30	100
	POMM-CRML$_{\bar{d}}$	P	0	0	0	1	17	**82**	118	69
		R	0	0	0	1	25	**74**	124	80
	POMM-CRML	P	0	0	1	1	21	**77**	119	96
		R	0	0	0	2	20	**78**	124	100
	LMM-CRML$_{\bar{d}}$	P	0	0	0	0	19	**81**	118	100
		R	0	0	0	1	24	**75**	124	100
	LMM-CRML	P	0	0	0	3	30	**64**	119	87
		R	0	0	0	3	23	**74**	124	95
	$p_3(d_k)$		0.07	**0.24**	0.48	0.68	0.80	0.88		
	$p_{2+}(d_k)$		0.22	**0.56**	0.78	0.89	0.94	0.97		
C	CRML	P	16	**77**	7	0	0	0	30	100
	POMM-CRML$_{\bar{d}}$	P	1	**98**	1	0	0	0	101	84
		R	1	**98**	1	0	0	0	107	65
	POMM-CRML	P	17	**82**	2	0	0	0	107	85
		R	7	**87**	8	1	0	0	107	92
	LMM-CRML$_{\bar{d}}$	P	3	**96**	1	0	0	0	103	100
		R	2	**96**	2	0	0	0	107	100
	LMM-CRML	P	26	**72**	2	0	0	0	110	85
		R	13	**82**	4	0	0	0	118	93

in scenarios *A* and *C*, the proportion of patients who received a dose higher than the MTD was lower with methods making full use of all cycles than with the CRML. No difference was observed among methods based on longitudinal data.

Power for detecting a cycle effect. Table V reports the percentage of simulations in which the test of the time effect was not null at the 5% level. Tests were carried out after simulation of 30 patients. The test for the time effect was significant in 46% of simulations of ordinal data (POMM-CRML) and 38% of simulations of binary data (LMM-CRML) in scenario $A(i)$ (OR=1.3 at cycle j compared with cycle $j-1$ and $j > 1$). As expected, in the case of a non log-linear relationship between time and toxicity (scenario $A(ii)$), the power was lower. Loss was even more marked when modeling binary data; results with simple logistic models were fairly disappointing. When using the POMM-CRML, violation of the PO assumption (scenario $A(iii)$ and $A(iv)$) also entailed a loss in power, but detection of time trend was still more effective with ordinal data than with binary data. This loss was more marked when the time

Table IV. Distribution of patients' allocations after 1000 simulations. POMM-CRML$_{\bar{d}}$ and LMM-CRML$_{\bar{d}}$ denote the POMM-CRML and the LMM-CRML respectively in which the dose parameter is set up and not estimated. $p_3(d_k)$ and $p_{2+}(d_k)$ are the probabilities of severe and moderate or severe toxicity per cycle in the absence of a cycle effect for an 'average' patient $u_i = 0$. Bold entries indicate the true target dose.

Scenario	Model	d_1	d_2	d_3	d_4	d_5	d_6
	$p_3(d_k)$	0.02	0.06	0.14	**0.24**	0.35	0.53
	$p_{2+}(d_k)$	0.09	0.21	0.40	**0.59**	0.71	0.82
A	CRML	6	12	27	**31**	19	5
	POMM-CRML$_{\bar{d}}$	3	6	25	**46**	16	4
	POMM-CRML	4	9	28	**41**	16	5
	LMM-CRML$_{\bar{d}}$	4	7	24	**46**	15	4
	LMM-CRML	5	9	26	**41**	15	5
	$p_3(d_k)$	0.01	0.02	0.05	0.09	0.14	**0.24**
	$p_{2+}(d_k)$	0.03	0.09	0.18	0.29	0.40	**0.55**
B	CRML	4	5	7	11	28	**44**
	POMM-CRML$_{\bar{d}}$	3	4	5	7	21	**59**
	POMM-CRML	3	5	6	9	23	**53**
	LMM-CRML$_{\bar{d}}$	4	4	5	8	23	**57**
	LMM-CRML	4	4	5	11	26	**51**
	$p_3(d_k)$	0.07	**0.24**	0.48	0.68	0.80	0.88
	$p_{2+}(d_k)$	0.22	**0.56**	0.78	0.89	0.94	0.97
C	CRML	29	**55**	13	2	1	0
	POMM-CRML$_{\bar{d}}$	9	**79**	10	2	1	0
	POMM-CRML	21	**67**	9	2	1	0
	LMM-CRML$_{\bar{d}}$	16	**72**	10	2	0	0
	LMM-CRML	28	**60**	9	2	0	0

Table V. Percentages of simulations where a significant time effect was detected for different models applied on scenario *A* . 'PO' stands for proportional odds; 'Y' and 'N' indicates wether the time-toxicity relationship follows proportional odds assumption. The lines (*i*) and (*ii*) correspond to scenarios with log linear and stepwise cycle effect respectively, and (*iii*) and (*iv*) to non odds proportional scenarios.

Scenario	PO relation	POMM-CRML$_{\bar{d}}$	POMM-CRML	LMM-CRML$_{\bar{d}}$	LMM -CRML
no time effect	Y	0.07	0.05	0.05	0.05
A(*i*)		0.64	0.54	0.42	0.36
A(*ii*)		0.43	0.35	0.27	0.24
A(*iii*)	N	0.53	0.43	-	-
A(*iv*)		0.33	0.26	-	-

effect was stronger on severe (less frequent) toxicity than on moderate or severe toxicity compared with the opposite scenario *A*(*iii*). This again shows the added value of richer variables.

The mis-specified models in which the dose effect was set up and not estimated outperformed more flexible models; for example, in scenario *Ai*, the time test was significant in 64% of simulations, compared with 46% with a model comprising an additional dose parameter to be estimated. The same trend was obtained with binary data as with ordinal data and for all scenarios. The false-positive rate in the absence of a time effect was 5−7% of simulations, indicating an acceptable type I error rate control with likelihood ratio tests despite the limited sample sizes.

5. Applications

The data introduced in Section 2 were reanalysed to identify the RP2D and to detect a time trend. The targeted probability of severe toxicity per cycle was set at 20%, to match the target used in the two trials. As

Table VI. Erlotinib+RT reanalysis: observed and predicted per cycle probability of graded toxicity; G3: Severe toxicity, G2+: Moderate or severe toxicity;$p_g(d_k)$ is the probability of toxicity per cycle in the absence of a time effect for an 'average' patient $u_i = 0$. ♯ stands for number; the bold column corresponds to the final recommended dose.

d_k	d_1	d_2	$\mathbf{d_3}$
♯ of patients	6	6	**8**
♯ of cycles	26	34	**36**
♯ of G2	4	8	**7**
♯ of G3	2	1	**4**
% of G3 (per cycle)	8	3	**11**
% of G2+ (per cycle)	23	26	**31**
Analysis using POMM			
$\hat{p}_{2+}(d_k)$, in % per cycle	23	27	**31**
95%CI ($\hat{p}_{2+}(d_3)$)			**19-46**
$\hat{p}_3(d_k)$, in % per cycle	6	7	**9**
95%CI ($\hat{p}_3(d_3)$)			**4-19**
Analysis using logistic model with random intercept (LMM)			
$\hat{p}_3(d_k)$, in % per cycle	5	7	**9**
95%CI ($\hat{p}_3(d_3)$)			**3-24**

retrospective analysis of adaptive designs is not directly feasible and requires further assumptions, only final estimates of the probability of toxicity are provided, bearing in mind that if methods for longitudinal data had been used to conduct the trial, the dose allocation would not have been the same; analyzing data derived from a different design results in a certain degree of loss of efficiency. Time-to-event CRM was also retrospectively applied to estimate the cumulative risk of DLT over the first six cycles.

5.1. The ITCC/ erlotinib +RT trial

Models (4) and (5) were estimated by adjusting for dose. Estimates of fixed intercepts, time, and dose were then $\hat{\theta} = (\alpha_1 = 1.78, \alpha_2 = 3.34, \beta_1 = 0.8, \beta_2 = -0.03)$. The variance of the random effect, σ_0^2, could not be estimated. The probability of toxicity did not significantly vary over time ($p = 0.83$), suggesting the absence of delayed effect; estimates of the model with dose only gave $\hat{\theta} = (\alpha_1 = 1.85, \alpha_2 = 3.41, \beta_1 = 0.80)$. The estimated probabilities of toxicity per cycle at each dose are shown in Table VI. The estimated probability of severe or moderate toxicity per cycle was 30.8% at the highest level tested. The estimated probability of severe toxicity was 8.6% (95%CI=3.7 − 18.7%). These estimates are closed to the crude percentage of grade 3 toxicity per cycle calculated as the mean ratio of the number of patients with grade 3/4 over the patients at risk at a given cycle. This risk appears lower compared with the estimates reported in Section 2 (16% at the highest level with 95%CI: 4 − 45%), but one should keep in mind that the DLT assessment period was two cycles, mechanically increasing the risk of toxicity compared with a risk calculated over a period of one cycle only. The precision measured by half the length of the confidence interval was much narrower with a POMM compared with a logistic model based on cycle 1 only (7.5 vs 20). The added value of ordinal longitudinal data over binary longitudinal data was also reflected by the accuracy of the estimate with precision of the 95% confidence interval of 7.5 versus 10.6, respectively.

The risk for a child to experience severe toxicity over the first six cycles was 37% (95%CI=12–72%) at the recommended dose d_3, as obtained from tite-CRM.

5.2. The EORTC/R-Viscum Trial

In this trial, doses were transformed on the (0, 1) scale as in the original trial and took the following values: 0.0035, 0.005, 0.009, 0.015, 0.024, 0.035, 0.05, 0.07, 0.11, 0.2, 0.33, 0.48, 0.62, and 0.74. Estimates of fixed intercepts, time, dose, and variance of the random intercept from model (5)were $\hat{\theta} = (\alpha_1 = 0.68, \alpha_2 = 3.36, \beta_1 = 2.09, \beta_2 = -0.024, \sigma_0^2 = 0.80)$. The probability of toxicity did not vary significantly with time ($p = 0.27$). Estimates of the model with dose as the only covariate gave $\theta = (\alpha_1 = 1.0, \alpha_2 = 3.78, \beta_1 = 2.48, \sigma_0^2 = 1.23)$. The estimated probabilities of toxicity per cycle at each dose are shown in

Table VII. r-Viscum trial reanalysis: observed and predicted probability of graded toxicity per cycle. G3: Severe toxicity, G2+: Moderate or severe toxicity; symbol ♯ stands for numbers and cy. for cycle; $p_g(d_k)$ is the probability of toxicity per cycle in the absence of a time effect for a patient $u_i = 0$. The four lowest dose levels were collapsed into a single column. The bold column corresponds to the final recommended dose.

	Dose (ng/kg)										
	$d_1 - d_4$	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}	d_{13}	$\mathbf{d_{14}}$
	10-100	200	400	800	1600	2400	3200	4000	4800	**5600**	6400
♯ of patients	4	1	1	1	1	1	4	6	10	**7**	5
♯ of cycles	15	2	1	2	2	2	13	11	25	**13**	8
♯ of G2	3	0	0	0	1	2	4	4	10	**5**	4
♯ of G3	2	0	1	0	1	0	0	1	2	**1**	2
% of G3 (per cycle)	0	20	0	100	0	50	0	9	8	**8**	25
% of G2+ (per cycle)	35	0	100	0	50	50	31	46	48	**46**	75
Analysis using POMM											
$\hat{p}_{2+}(d_k)$ in % per cycle		32	32	33	34	36	40	46	54	**61**	66
95%CI ($\hat{p}_{2+}(d_{13})$)										**49-84**	
$\hat{p}_3(d_k)$ in % per cycle		4	4	4	5	5	6	8	10	**14**	17
95%CI ($\hat{p}_3(d_{13})$)										**6-28**	
Analysis using logistic model with random intercept											
$\hat{p}_3(d_k)$ in % per cycle		8	8	8	8	9	10	12	14	**17**	19
95%CI ($\hat{p}_3(d_{13})$)										**5-40**	

Table VII. The first four columns were collapsed to form a single column labeled '0 – 100'. According to the POMM, the recommended dose is associated with a risk of severe toxicity of 14% (95%CI=6–28%) per cycle and a risk of moderate or severe toxicity of 61% (95%CI=49–84%) per cycle. Point estimate was close to the one obtained with binary data only (17% with a 95% confidence from 5% to 40%). Longitudinal binary data did not allow for a gain in precision over the estimate based on cycle 1 only. Of note, the model of the CRML used in the trial was less-parameterized with a single parameter, which may explain the smaller variance. As expected due to the limited number of cycles that could be administered, accuracy of the estimate was not strongly increased when using repeated data. The relative gain in precision computed as the ratio of the precisions of the two confidence intervals was 27% over the classic CRML and 45% over the LMM-CRML.

Analysis of data collected during all cycles improved the accuracy of the estimate of the risk of severe toxicity. As all toxic side effects were reversible, a longitudinal model is appealing even though patients could receive only very few cycles. Additional information is not difficult to interpret and is consistent with the usual estimates from the first treatment cycle. No strong time trend could be detected in the examples we analyzed.

6. Discussion

Methodological research on dose-finding studies has been very active in the last decade; the main fields of statistical developments include optimization of the CRM [29] and investigation of various allocation rules [30], joint modeling of toxicity and activity endpoints [31–33], and integration of time into assessment of the endpoint [4, 9]. However, new proposals have usually been accompanied by increased complexity of models raising estimation and robustness issues in the context of scarce data. Although estimations are usually tackled with a Bayesian approach and mildly informative priors, the additional variability introduced by the numerous parameters together with the poor information contained in binary variables limit the performances of the dose-finding process. For instance, Simon *et al.* developed a *K*-max model to retrospectively analyze longitudinal data of phase I trials [4], but Legezda and Ibrahim showed the model to be intractable in a prospective dose-finding study [34]. They proposed simplified pharmacokinetic–pharmacodynamic models to estimate the best dose to be allocated to each patient on the basis of a binary outcome, if they received numerous cycles of treatment without going off study.

In this study, we demonstrate that using all cycles of treatment increase the accuracy of MTD identification and allows detection of a time trend, providing essential information to recommend a dose for phase II. The benefit of using information from repeated cycles is much greater than the optimal performance that could be theoretically obtained when using cycle 1 only [25, 35]. Furthermore, modeling

ordinal grades of toxicity increases the power to detect time trends compared with binary data, even if the time function is mis-specified and if the PO assumption is not verified. Toxicity scores as developed by Yuan [36], Rogatko [37], or Ezzalfani [38] could be even more powerful. Models for longitudinal ordinal data do not significantly improve the ability to pick the correct dose compared with models for longitudinal binary data; the benefit of richer variables is somehow counterbalanced by the additional parameters to be estimated. In addition, as the target dose is defined as a risk of severe toxicity, moderate adverse events only provide indirect information.

The second finding of this study is that simpler models constraining some of the parameters turn out to be more efficient than more parameterized and more flexible models, in terms of selection of the correct dose. The sampling rule that consists of systematically recommending the current estimate of the MTD limits the need for a global goodness-of-fit for all doses [9, 20]. Even though parameters estimates are likely to be biased, estimates of the probability of toxicity at the final recommended dose should converge to the true MTD, providing sampling at the dose closest to the target. On the contrary, mis-specified time-toxicity relationship reduces the power of the test on time; greater attention should be devoted to this aspect of modeling. In particular, as a simplifying assumption, we assumed that patients could receive the same dose until they got off-study. In practice, although intra-patient dose escalation is commonly ruled out in protocols, moderate or severe toxicity may lead to reduce the dose or delay the administration of a new cycle of treatment. A simple way to account for dose reduction would be to use the administered dose instead of the planned dose at each cycle as in model (2). However, this may not be consistent with the pharmacokinetic–pharmacodynamic model that would relate the drug concentration at a time point to the risk of toxicity. In practice, this model is largely unknown before the first-in-man trial is completed. The very simple modeling used here may therefore not be optimal. Similar issues are also observed with the tite-CRM as the time to toxicity is associated to a planned dose and not to the cumulative dose of treatment received at the time of toxicity. This model may be mis-specified if there is a cumulative risk of toxicity as with radiotherapy treatments. In this setting, finding the correct balance between model complexity and biases is challenging and should be decided on a case by case basis.

Random effect models are powerful tools to estimate dose effects on longitudinal data. In phase I trials, patients go off study because of severe toxicity or of progression. When a patient goes off study because of severe toxicity, outcomes at subsequent cycles are then missing. We assume that the risk of progression at a given dose is independent of the DLT. We then may consider that the missingness process is independent of the unmeasured toxicity value. Generalized linear mixed model with maximum likelihood estimator naturally addresses missing at random data. However, estimation of the variance of the random effect is challenging when few repeated measurements are available. Lack of accuracy in the estimate of this parameter may impact the dose recommendation. For instance, in scenario *A*, we fit a proportional odds model (no random effect) on all data collected from a trial that would have used the CRM to guide dose escalation. The distribution of final dose recommendations at d_3, d_4, and d_5 was as follows: 24%, 55%, and 13%, that was slightly less good than the corresponding simulation of POMM-CRML (R) in Table III (22, 57, and 13, respectively).

Phase I trials provide much more information than a simple binomial outcome. Since the early 2000's, new classes of molecules have been developed that raise specific issues that cannot be efficiently addressed using basic methods. There is a converging need from both statisticians and physicians to improve dose-finding studies. This improvement will be made possible by the use of more informative variables.

Acknowledgements

We are indebted to Andrew Kramar (O. Lambret Cancer Center, Lille, France) for his stimulating discussions and clever suggestions. We thank Ben Bolker (McMaster University, Hamilton, Canada) for his valuable help with `lme4` package. We would like to express our gratitude to Pr Patrick Schöeffski (University Hospital Leuven, Belgium), Pr Pierre Fumoleau (Dijon Cancer Center and University, France), and Laurence Collette (EORTC-HQ Brussels, Belgium) who provided the individual patient data for the EORTC R-Viscum study. Likewise, we are grateful to Dr Birgit Geoerger and Dr Marie-Cécile le Deley (Institut Gustave Roussy, Villejuif, France) for providing data for the ITCC erlotinib trial. We also thank Dr Anthony Saul for reviewing the English language. This work was supported by the National Cancer Institute of France (INCa); SIRIC & Optidose(no. 2010-1-PL SHS-06-IC-1).

References

1. National Cancer Institute UNIoHCTEP. *Common Terminology Criteria for Adverse Events (CTCAE) v4.03.*, 2010. http://ctep.cancer.gov/protocolDevelopment/electronic applications/ctc.htm.
2. Storer BE. Small-sample confidence sets for the MTD in a phase I clinical trial. *Biometrics* 1993; **49**(4):1117–1125. PMID: 8117905.
3. O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics* 1990; **46**(1):33–48. PMID: 2350571.
4. Simon R, Freidlin B, Rubinstein L, Arbuck SG, Collins J, Christian MC. Accelerated titration designs for phase I clinical trials in oncology. *Journal of the National Cancer Institute* 1997; **89**(15):1138–1147. PMID: 9262252.
5. Gupta S, Hunsberger S, Boerner SA, Rubinstein L, Royds R, Ivy P, LoRusso P. Meta-analysis of the relationship between dose and benefit in phase I targeted agent trials. *Journal of the National Cancer Institute* 2012; **104**(24):1860–1866. PMID: 23169991.
6. Le Tourneau C, Diéras V, Tresca P, Cacheux W, Paoletti W. Current challenges for the early clinical development of anticancer drugs in the era of molecularly targeted agents. *Targeted Oncology* 2010; **5**(1):65–72. PMID: 20361265.
7. Paoletti X, Le Tourneau C, Verweij J, Siu L, Seymour L, Postel-Vinay S, Collette L, Rizzo E, Ivy P, Olmos D, Massard C, Lacombe D, Kaye S, Soria JC. Defining dose-limiting toxicity for phase 1 trials of moleularly targeted agents: results of a dlt-targett international survey. *European Journal of Cancer* 2014; **50**(12):2050–2056.
8. Postel-Vinay S, Collette L, Paoletti X, Rizzo E, Massard C, Olmos D, Fowst C, Levey B, Mancini P, Lacombe D, Ivy P, Seymour L, Le Tourneau C, Siu L, Kaye S, Verweij J, Soria JC. Towards new methods for the determination of dose limiting toxicities and the assessment of the recommended dose for further studies of non-cytotoxic agents. *European Journal of Cancer* 2014; **50**(12):2040–2049.
9. Cheung YK, Chappell R. Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics* 2000; **56**(4):1177–1182. PMID: 11129476.
10. Doussau A, Thiébaut R, Paoletti X. Dose-finding design using mixed effect proportional odds model for longitudinal graded toxicity data in phase i oncology clinical trials. *Statistics in Medicine* 2013; **32**(30):5430–5447.
11. Van Meter EM, Garrett-Mayer E, Bandyopadhyay D. Proportional odds model for dose-finding clinical trial designs with ordinal toxicity grading. *Statistics in Medicine* 2011; **30**(17):270–280.
12. Iasonos A, Zohar S, O'Quigley J. Incorporating lower grade toxicity information into dose finding designs. *Clinical Trials* 2011; **8**(4):370–379.
13. Geoerger B, Hargrave D, Thomas F, Ndiaye A, Frappaz D, Andreiuolo F, Varlet P, Aerts I, Riccardi R, Jaspan T, Chatelut E, Le Deley MC, Paoletti X, Saint-Rose C, Leblond P, Morland B, Gentet JC, Méresse V, Vassal G. Innovative therapies for children with cancer pediatric phase I study of erlotinib in brainstem glioma and relapsing/refractory brain tumors. *Neuro-oncology* 2011; **13**(1):109–118. PMID: 20974795.
14. O'Quigley J, Shen LZ. Continual reassessment method: a likelihood approach. *Biometrics* 1996; **52**(2):673–684 (eng). PMID: 8672707.
15. Schöffski P, Riggert S, Fumoleau P, Campone M, Bolte O, Marreaud S, Lacombe D, Baron B, Herold M, Zwierzina H, Wilhelm-Ogunbiyi K, Lentzen H, Twelves C, European Organization for Research and Treatment of Cancer New Drug Development Group. Phase I trial of intravenous aviscumine (rViscumin) in patients with solid tumors: a study of the european organization for research and treatment of cancer new drug development group. *Annals of Oncology* 2004; **15**(12):1816–1824. PMID: 15550588.
16. Paoletti X, Baron B, Schöffski P, Fumoleau P, Lacombe D, Marreaud S, Sylvester R. Using the continual reassessment method: lessons learned from an EORTC phase I dose finding study. *European Journal of Cancer* 2006; **42**(10): 1362–1368. PMID: 16740385.
17. Ezzet F, Whitehead J. A random effects model for ordinal responses from a crossover trial. *Statistics in Medicine* 1991; **10**(6):901–906; discussion 906–907. PMID: 1876780.
18. Molenberghs G, Verbeke G. *Models for Discrete Longitudinal Data*. Springer-Verlag: New York, 2005.
19. Houédé N, Thall PF, Nguyen H, Paoletti X, Kramar A. Utility-based optimization of combination therapy using ordinal toxicity and efficacy in phase I/II trials. *Biometrics* 2010; **66**(2):532–540. PMID: 19673865.
20. Shen L, O'Quigley J. Consistency of continual reassessment method under model misspecification. *Biometrika* 1996; **83**(2):395–405.
21. Cheung YK, Chappell R. A simple technique to evaluate model sensitivity in the continual reassessment method. *Biometrics* 2002; **58**(3):671–674. PMID: 12230003.
22. Paoletti X, Kramar A. A comparison of model choices for the continual reassessment method in phase I cancer trials. *Statistics in Medicine* 2009; **28**(24):3012–3028. PMID: 19672839.
23. Zohar S, Chevret S. The continual reassessment method: comparison of bayesian stopping rules for dose-ranging studies. *Statistics in Medicine* 2001; **20**(19):2827–2843.
24. O'Quigley J, Reiner E. Miscellanea. A stopping rule for the continual reassessment method. *Biometrika* 1998; **85**(3): 741–748.
25. Paoletti X, O'Quigley J, Maccario J. Design efficiency in dose finding studies. *Computational Statistics & Data Analysis* 2004; **45**(2):197–214.
26. Braun TM. Generalizing the tite-crm to adapt for early- and late-onset toxicities. *Statistics in Medicine* 2006; **25**(12): 2071–2083.
27. Chevret S. *Statistical Methods for Dose-finding Experiments*. Wiley, 2006.
28. Zhou Y, Whitehead J. Practical implementation of Bayesian dose-escalation procedures. *Drug Information Journal* 2003; **37**(1):45–59.
29. O'Quigley J, Conaway M. Extended model-based designs for more complex dose-finding studies. *Statistics in Medicine* 2011; **30**(17):2062–2069.

30. Neuenschwander B, Branson M, Gsponer T. Critical aspects of the Bayesian approach to phase I cancer trials. *Statistics in Medicine* 2008; **27**(13):2420–2439. PMID: 18344187.
31. Thall PF, Cook JD. Dose-finding based on efficacy-toxicity trade-offs. *Biometrics* 2004; **60**(3):684–693.
32. Braun T. The bivariate continual reassessment method. extending the CRM to phase I trials of two competing outcomes. *Control Clinical Trials* 2002; **23**(3):240–256.
33. Zhong W, Koopmeiners JS, Carlin BP. A trivariate continual reassessment method for phase I/II trials of toxicity, efficacy, and surrogate efficacy. *Statistics in Medicine* 2012; **31**(29):3885–3895.
34. Legezda ATR, Ibrahim JG. Longitudinal design for phase I clinical trials using the continual reassessment method. *Controlled Clinical Trials* 2000; **21**(6):574–588.
35. Wages NA, Conaway MR, O'Quigley J. Performance of two-stage continual reassessment method relative to an optimal benchmark. *Clinical Trials* 2013; **10**(6):862–875.
36. Yuan Z, Bailey H, Chappell R. The continual reassessment method for multiple toxicity grades: a Bayesian quasi-likelihood approach. *Biometrics* 2007; **63**(1):173–179.
37. Rogatko A, Babb J, Wang H, Slifker M, Hudes GR. Patient characteristics compete with dose as predictors of acute treatment toxicity in early phase clinical trials. *Cancer Clinical Reseach* 2004; **10**(14):4645–4651.
38. Ezzalfani M, Zohar S, Qin R, Mandrekar SJ, Le Deley MC. Dose finding designs using a novel quasi-continuous endpoint for multiple toxicities. *Statistics in Medicine* 2013; **32**(16):2728–2746.