Received: 2 April 2020 | Revised: 31 August 2020 | Accepted: 1 September 2020

DOI: 10.1002/sim.8771

RESEARCH ARTICLE

Statistics in Medicine WILEY

Early phase dose-finding trials in virology

Hakim-Moulay Dehbi[1] | **David M. Lowe**[2] | **John O'Quigley**[3]

[1]Comprehensive Clinical Trials Unit, University College London, London, UK

[2]Institute of Immunity and Transplantation, Royal Free Hospital, London, UK

[3]Department of Statistical Science, University College London, London, UK

Correspondence
Hakim-Moulay Dehbi, Comprehensive Clinical Trials Unit, University College London, 90 High Holborn, 2nd Floor, London WC1V 6LJ, UK.
Email: h.dehbi@ucl.ac.uk

Little has been published in terms of dose-finding methodology in virology. Aside from a few papers focusing on HIV, the considerable progress in dose-finding methodology of the last 25 years has focused almost entirely on oncology. While adverse reactions to cytotoxic drugs may be life threatening, for anti-viral agents we anticipate something different: side effects that provoke the cessation of treatment. This would correspond to treatment failure. On the other hand, success would not be yes/no but would correspond to a range of responses, from small, no more than say 20% reduction in viral load to the complete elimination of the virus. Less than total success matters since this may allow the patient to achieve immune-mediated clearance. The motivation for this article is an upcoming dose-finding trial in chronic norovirus infection. We propose a novel methodology whose goal is twofold: first, to identify the dose that provides the most favorable distribution of treatment outcomes, and, second, to do this in a way that maximizes the treatment benefit for the patients included in the study.

KEYWORDS

continual reassessment method, coronavirus, dose-finding, early phase trials, norovirus, virology

1 | INTRODUCTION

This introduction has three subsections: the first provides a broad description of our goals, the second a summary of the study motivating this current work, and, finally, a subsection making these goals more specific and anticipating the section that follows on statistical methodology.

1.1 | Background and motivation

Antiviral agents can fail to be of benefit to a patient in two different ways. The first is where the toxic side effects are such that the patient is unable to take the full course of treatment, thereby not being in a position to experience treatment benefit. The second type of failure is where the treatment is well tolerated but fails to achieve a meaningful clinical or virological effect, for example, a lessening of symptoms, reduction in viral load, or complete elimination of the virus (or, for some viruses, induction of viral latency). When viral load is the main focus, efficacy may be categorized into three or possibly more outcomes, ranging from a small, insignificant reduction in viral load to a large reduction or total elimination (or suppression) of the virus. Given a new candidate treatment, and some range of possible treatment levels,

usually doses, we may assume that there exists some optimum dose level. As we increase the dose levels we anticipate seeing both higher rates of toxicities as well as higher rates of viral load reduction. While we anticipate that increasing the dose will result in a greater reduction in viral load, we are obliged to take into account that the percentage of patients able to benefit will diminish with dose. This is due to dose limiting toxicity (DLT), in this setting the impossibility for the patient to see through the full course of treatment. In virology, toxicities that would justify stopping the treatment would include significant disturbance of liver function or kidney function, a marked decline in blood cell counts, intolerable neuro-psychiatric side effects, persistent or widespread rash, or intractable gastrointestinal upset. While most of these toxicities are on a continuum, and may be tolerable or treated symptomatically, treatment would need to be stopped once certain thresholds are met (for blood tests) or when the problems become intolerable/untreatable (for clinical symptoms).

In the oncology setting the focus of dose-finding methodology has evolved from that of toxicity alone, to attempts at incorporating efficacy measures via expansion cohorts.[1,2] Other designs have taken as their starting point the aim to model the joint relationships between dose, toxicity, and efficacy.[3-5] In oncology, it is conceivable that a patient suffers a very great degree of toxicity in exchange for treatment benefit and, for this reason, the joint distribution of efficacy and toxicity is worthy of study. However in virology trials, the joint distribution is not a concern. Our focus is on the conditional distribution of viral reduction given the absence of toxicity. Indeed, experiencing a DLT in a virology study means that the patient loses the opportunity to benefit from the treatment. This is because it is of paramount importance to complete the course of treatment to identify a lasting treatment effect in virology. The need to complete the course has been demonstrated in chronic Hepatitis C for example, where a treatment of 6 or 8 weeks instead of 12 weeks increases the risk of relapse markedly.[6,7] This implies that, even though the viral load may (or may not) have reduced sufficiently over 6 weeks, a sustained cure would not be achieved, equating to a lack of efficacy. In terms of toxicity, it is more likely to be a risk early in treatment in virology. Nonetheless toxicity can also occur later as a cumulative dose-related issue (often seen with antibiotics such as linezolid). However, regardless of the timing of DLT, the treatment needs to stop and any potential efficacy is seriously compromised.

The best dose, M, should show an optimal compromise between the goals of minimizing toxicity and maximizing viral load reduction. A successful trial will identify the level M at which we can achieve adequate viral load reductions without side effects in too high a proportion of patients. This assumes that such a level exists. If no such level exists then the aim is a different one, to be in a position to be able to state that the new treatment is ineffective, and once again using as few patients as possible. And, in all cases, each included patient is treated in accordance with ethical principles in as much as, given all the information available, the best candidate dose from the current set of available doses.

1.2 | A dose-finding clinical trial in norovirus

The methodology that we propose was developed for an upcoming study for patients with primary or irreversible secondary immunodeficiency and established chronic norovirus infection resulting in symptoms and/or malabsorption.

Chronic norovirus infection is a serious complication of immune deficiency states leading to considerable morbidity, healthcare utilization, and death.[8] Currently no proven treatment exists.[9] The antiviral medication favipiravir is licensed in Japan for novel or pandemic influenza and has also been used for Ebola virus disease. It has shown promise in pre-clinical and early clinical usage in chronic norovirus infection. Pre-clinical data, especially in a mouse model, is encouraging and suggests that this medication drives lethal mutagenesis and viral extinction.[10,11] Favipiravir was used in one patient chronically infected with norovirus. A positive clinical response was seen (marked reduction in diarrhoea and increase in weight) as well as evidence of viral mutagenesis and shifts in dominant viral populations. This case report was published in the medical literature.[12] However, at the population level, the optimal dose is unknown for chronic norovirus.

The objective of the dose-finding trial will be to establish a tolerable and efficacious dose. A limited number of doses will be considered, up to three. The primary efficacy endpoint will be the reduction of stool viral load. Patients will be treated for 8 weeks in the absence of DLTs. Full elimination of the virus will be the best possible response, and the minimum difference of interest will be an increase in the norovirus cycle threshold value in stool samples (inversely related to viral load) of at least 7.5 units with favipiravir treatment, equivalent to 25% from the mean baseline level in adults.

1.3 | Dose-finding trials in virology

Analogous to the objectives underlying the development of the continual reassessment method (CRM),[13] we would like to construct a design which aims to (1) keep to a minimum the number of patients treated at unacceptably high toxic dose levels; (2) keep to a minimum the number of under-treated patients, that is, patients treated at dose levels producing insufficient viral response; (3) respond quickly to errors in initial estimates, rapidly escalating in the absence of indication of drug activity (viral load reduction) and rapidly de-escalating in the presence of unacceptably high levels of observed toxicity; (4) come to an early closure; either success and a recommended level, or failure of the trial as a whole; and (5) minimize the number of patients needed to complete the study (efficiency).

We do not expect to be able to construct a single design that would achieve all of the above objectives in any optimal way. Instead we develop a class of designs that generally behave well. Designs that "behave well" are designs that broadly meet the above five objectives. Extensive simulations, some of which are recorded here in Section 4, give support to the contention that our proposed designs behave well. The following Sections 2 and 3 focus on the statistical ideas behind the designs and how they can be implemented in practice.

There is an echo in the current work of the method introduced by O'Quigley et al for a specific dose-finding study in HIV.[14] Toxicity amounted essentially to an inability of the patient to maintain the treatment, which is not directly comparable to the presence of serious adverse events as in oncology. In the HIV setting where potentially effective treatments were already available it was possible to amalgamate toxicity with insufficient efficacy into a single category called treatment failure. Indeed only a significant reduction in viral load was seen as a success. For the study here, and for the majority of dose-finding studies in virology, we will not amalgamate these categories. Even if the anti-viral effect is weak or medium, it should be considered to be an improvement on the absence of any effect due to toxicity. Indeed, any non-trivial effect can have value and potentially allow the patient to achieve immune-mediated clearance. This means that, contrary to HIV setting, we need to consider ordered categories of anti-viral effect rather than a yes/no outcome.

2 | STATISTICAL FEATURES OF THE DESIGN

We have available a set of m fixed-order dose levels: $d_1, \ldots d_m$. The probability of encountering toxicity at level d_i is assumed to increase monotonically with dose. This monotonicity requirement is central to the CRM class of designs. We may wish to assume some analogous property of monotonicity across the doses with regard to efficacy. One simple model would be to suppose that, by focusing only on non-toxicities and dividing the efficacy outcomes into two categories of viral load reduction, then the conditional probability of viral load reduction increases monotonically with dose. For more than two categories, we may wish to summarize the distribution of outcomes by its average conditional upon no toxicity, and assume that it is monotonic with dose. Whether or not we make this assumption, overall success, being the absence of toxicity together with viral load reduction, is not necessarily monotonic.

Consider a trial in which $j = 1, \ldots n$ patients may be entered, n being the greatest number of patients that we are prepared to treat in the study. The dose level for the jth patient, X_j can be viewed as random taking discrete values x_j where $x_j \in \{d_1, \ldots d_m\}$. Let Y_j be a binary random variable $(0, 1)$ where 1 denotes a toxic response. We define the true probability of toxic response at $X_j = x_j$ by

$$R(x_j) = \Pr(Y_j = 1 | X_j = x_j). \tag{1}$$

Let V_j be a categorical random variable with k response categories denoting the possible levels of viral load reduction for the jth patient given that no toxicity was observed. Without loss of generality, we assume 3 different levels of viral load reduction given no toxicity; between 0% and 20% (category 1), between 20% and 50% (category 2), and greater than 50% (category 3). Modifying these definitions or adding new intervals is conceptually straightforward. Given dose level $X_j = x_j$, we consider the conditional probability distribution of efficacy given no toxicity, $Q(v|x_j)$, over the three categories. We write:

$$Q(v \mid x_j) = \Pr(V_j \leq v \mid X_j = x_j, \quad Y_j = 0), \quad v = 1, 2, 3. \tag{2}$$

The expected value of the conditional efficacy distribution is given by:

$$E(V_j \mid x_j, Y_j = 0) = \sum_{v=1}^{3} v \times \{Q(v \mid x_j) - Q(v-1 \mid x_j)\}, \quad j = 1, \ldots, n. \tag{3}$$

When $Y_j = 1$ then treatment is incomplete because of toxicity and we assume that no efficacy can be observed. We can then, in a formal way, extend the support of the distribution of V_j to include the value 0, also noting that this conditional distribution is degenerate, that is, $\Pr(V_j = 0 | Y_j = 1) = 1$. As a result, we have that $E(V_j | x_j, Y_j = 1) = 0$. We now have the conditional expectations for efficacy for when $Y_j = 0$ and $Y_j = 1$. Our decisions will be based on the unconditional expectation of efficacy at each dose and this is simply:

$$\begin{aligned} E(V_j \mid x_j) &= R(x_j) \times E(V_j \mid x_j, Y_j = 1) + \{1 - R(x_j)\} \times E(V_j \mid x_j, Y_j = 0) \\ &= \{1 - R(x_j)\} \times E(V_j \mid x_j, Y_j = 0). \end{aligned} \tag{4}$$

The unconditional expectation of efficacy is referred to as the center of mass (CM) of the distribution of V_j. Note that whereas $E(V_j | x_j, Y_j = 0)$ lies between 1 and 3, $E(V_j | x_j)$ will lie between 0 and 3. Low values of $E(V_j | x_j)$, close to zero indicate high toxicity and/or effects that are too weak to be of clinical relevance. Values of $E(V_j | x_j)$ close to 3 indicate almost complete success without toxicity. In practice, we may not be so close to either zero or three and we will select the most promising dose as the one that is closer to 3 than all of the others. It may not be close enough and, of course, that is something that the clinical investigating team will need stipulate ahead of the trial. This can be informal since we are not looking at a comparative trial. Our goal is to locate the best dose, M, and we will take this to be the dose

$$M = \arg\max_i E(V_j \mid d_i), \quad i = 1, \ldots, m.$$

We estimate M sequentially and allocate patients to a dose level based on our current estimate of M. After a fixed number of patients have been treated, we take the last estimated value of M to be our recommended dose for further study.

3 | ADAPTIVE DOSE ALLOCATION

The most direct way to address the estimation of M is to obtain enough empirical observations at each level thereby allowing us to determine, with sufficient precision, estimates of $E(V_j | d_i)$, $i = 1, \ldots, m$. The level ℓ that provides the maximum of these corresponds to our current best estimate for M. In practice, the most direct way to achieve this would be with random allocation between doses. At each observation or group of observations, estimates of $R(d_i)$, $Q(\nu | d_i)$, $E(V | d_i)$ can be updated and, finally, $\hat{M}$ can be derived based on the empirical rates. We would anticipate these as being relatively unstable early on in the study when few patients are treated. The main disadvantage of the random allocation approach is that it requires the inclusion of many patients at levels that may be far from M, which is not in keeping with the objectives of a well-behaving dose-finding algorithm.

3.1 | Dynamic dose-finding algorithm

Instead of using randomization, we propose a dose-finding algorithm that does not depend on modeling to determine, with sufficient precision, estimates of $E(V_j | d_i)$, $i = 1, \ldots m$. The purpose of the algorithm is to locate as efficiently as possible the dose providing the highest CM. It is designed in such a way that it is both cautious and thorough.

The proposed dose-finding algorithm requires the definition of the maximum allowable threshold for toxicity, which we denote as Rmax. This threshold represents the highest proportion of patients with a toxic reaction that can be allowed during the course of the study for the disease-drug combination of interest. In practice we proceed in the following way.

3.1.1 | First cohort of patients at the first dose level d_1

The first cohort of patients is treated at the first (ie, lowest) dose level d_1 and the observed toxicity rate is compared with Rmax. As long as the observed rate is less than the threshold, then the next cohort of patients can be treated at the next dose d_2. If the rate is equal or larger than the threshold, d_1 is repeated. The trial may be stopped early after the first few cohorts of patients if the toxicity rate is estimated to be larger than the threshold with sufficient precision. The amount of certainty required would be defined by the investigators prior to the start of the study.

3.1.2 | Cohorts of patients at subsequent dose levels

At d_2 and higher, Rmax is used as a gatekeeper to determine the set of doses from which to choose for the next cohort:

1. if the observed toxicity rate is less than Rmax, dose escalation becomes an option. The set of potential doses for the next cohort corresponds to all prior doses, the current dose and the next dose level up;
2. if the observed toxicity rate is larger than Rmax, the set of potential doses for the next cohort corresponds to all prior doses (ie, all doses up to, but excluding, the current dose level);
3. if the observed toxicity rate is equal to Rmax, the set of potential doses corresponds to all prior doses and the current dose.

Once the set of potential doses is determined, the dose for the following cohort is the one with the largest probability of being the dose with the largest CM among the set of possible doses (see section below on estimation of the CM for further details).

However, until there is experimentation on all doses, there is no observed data at the doses that have not yet been tested. To allow the dose-finding algorithm to explore higher dose levels, a dose will be tested for the first time if the observed toxicity rate of the dose level below it is less than Rmax. As the process unfolds, we update our running estimates of the toxicity rates at the dose levels. In consequence it can easily occur that doses previously off limits as determined by Rmax become once again available for experimentation. When this happens, we temporarily suspend allocation based on the CM in favor of simple allocation to the newly available dose level. Once Rmax again limits the doses under consideration or, if all doses have been tested, we revert to allocation based on the CM. For example, if at d_4 the observed toxicity rate is higher than Rmax, the next cohort of patients is allocated to the dose with the highest probability of being the one with the largest CM among $\{d_1, d_2, d_3\}$, which could be d_3 for instance. The next cohort would then be treated at d_3 and if the observed toxicity rate (from all patients treated so far at d_3) is less than Rmax, d_4 would be retested. If the observed rate at d_4 (from all patients treated so far at d_4) is now less than Rmax, the next cohort would be treated at d_5. Section 4.1 provides an illustration of this process in a hypothetical dose-finding trial.

This process is repeated until the total sample size is reached. At the end of the dose-finding trial, a comparison between doses is made and a ranking is derived, based on the probability of any given dose being the one with the largest CM.

3.2 | Estimation of the center of mass

Assuming three efficacy categories given no toxicity, and an additional category of no efficacy when there is a toxic reaction, the unconditional distribution of efficacy V_j has four possible categories. To estimate $E(V_j|d_i)\,,\ i=1,\ \dots\ m$, which is the CM of V_j, we exploit the Bayesian Dirichlet-Categorical model. In the Dirichlet-Categorical model, at dose d_i, the distribution of $V_j|P_i$ follows a categorical distribution denoted as $\mathbb{C}(k, P_i)$, where P_i follows a Dirichlet distribution and $k=4$. Effectively P_i is the vector of event probabilities associated with the $k=4$ categories of the distribution at dose d_i, $i=1,\ \dots\ m$.

We define the vector α of hyperparameters for the Dirichlet distribution as a vector of length $k=4$ of the form $\alpha=(1,1,1,1)$. This Dirichlet distribution is the generalization of the uniform distribution, where the sum of the probabilities equals one. In other words, the Dirichlet distribution $\mathbb{D}(4,(1,1,1,1))$ is the prior distribution of P_i. Consequently, the prior mean proportion in each category of V_j at dose d_i is 25%. Given that the CM at dose d_i, defined in Equation (4), is a weighted average of the value of the categories by their associated probabilities, its prior distribution can be calculated by sampling from the $\mathbb{D}(4,(1,1,1,1))$, and it is shown in Figure 1.

The posterior distribution of P_i is a function of the prior distribution and the observed data. We denote by the vector v_i the observed number of occurrences in the k categories at d_i as $v_i=(v_{0i}, v_{1i},\ \dots\ , v_{(k-1)i})$. The posterior distribution of P_i is $P_i|\alpha, v_i \sim \mathbb{D}(k, \alpha+v_i) = \mathbb{D}(k, 1+v_{0i}, 1+v_{1i},\ \dots\ , 1+v_{(k-1)i})$. By sampling from the posterior distribution of P_i, the posterior distribution of the CM at dose d_i can be calculated as a linear combination of the probabilities and the associated values of the categories.

Independent sampling from posterior distributions at all dose levels allows to calculate the probability that each dose is the one with the largest CM, by recording which dose has the largest CM at each sampling iteration.

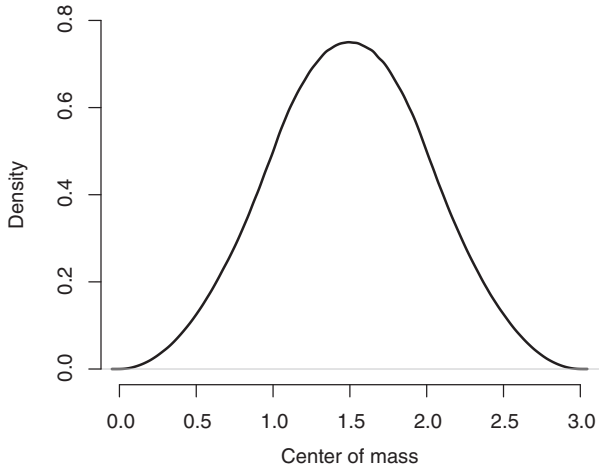

FIGURE 1 Prior distribution of the center of mass (CM) at dose d_i based on a Dirichlet distribution $\mathbb{D}(4, (1, 1, 1, 1))$

3.3 | Impact of the prior distribution on estimation of M

At dose $d_i, i = 1, \ldots, m$, we associate a Dirichlet distribution $\mathbb{D}(k, \alpha_i)$ where $\alpha_i = (\alpha_{1i}, \ldots, \alpha_{ki})$ and $S_i = \sum_{j=1}^{k} \alpha_{ji}$. The expected value of the proportion in each of the k categories at dose d_i is then $E(P_{ki}) = S_i^{-1}\alpha_{ki}$. We characterize the Bayesian Dirichlet-Categorical model via its prior means, each at 25% so that $E(P_{ki}) = 1/4$. The posterior distribution of P_i is a function of the prior distribution and the observed data. Using the same notation as above, the vector v_i denotes the observed number of occurrences in the k categories at d_i as $v_i = (v_{0i}, v_{1i}, \ldots, v_{(k-1),i})$. The posterior distribution of P_i is $P_i|\alpha, v_i \sim \mathbb{D}(k, \alpha + v_i) = \mathbb{D}(k, 1 + v_{0i}, 1 + v_{1i}, \ldots, 1 + v_{(k-1),i})$. The posterior mean of P_i is a combination of the prior means and the observed proportions. The vector of observed proportions is $\pi_i = V_i^{-1} v_{ji}, j = 0, \ldots, k-1$, where $V_i = \sum_{j=0}^{k-1} v_{ji}$. The posterior mean of P_i, $P_i|\alpha, v_i$, is given by

$$\frac{S_i}{n + S_i} \times E(P_i) + \frac{n}{n + S_i} \times \pi_i,$$

where n is the total number of observations. This means that the posterior mean of P_i is a weighted average of the observed proportions and the prior mean proportions, with weights based on the observed sample size and $S_i = \sum_{j=1}^{k} \alpha_{ji}$. Effectively the total effective sample size is the sum of the observed sample size plus S_i. For example, with a sample size of six patients and $\alpha = (1, 1, 1, 1)$, the observed proportions make a contribution of 60% to the posterior mean proportions. In designing any study, we may wish to consider the impact of α_i, at dose levels $d_i, i = 1, \ldots, m$, on estimation early in the study.

4 | SIMULATIONS

4.1 | Illustration of a single hypothetical dose-finding trial

Consider a trial with three potential doses $\{d_1, d_2, d_3\}$, 30 available patients and the cohort size is set at 6 patients. The maximum allowable threshold for toxicity, Rmax, is set at 50%. Four response categories are of interest. The first category, $k = 0$, corresponds to a toxic reaction to the treatment implying failure to observe any efficacy. $k = 1$ corresponds to no toxicity but also no efficacy at all or too little for it to be of clinical relevance, $k = 2$ corresponds to a medium level of efficacy, and $k = 3$ corresponds to high efficacy or complete viral load clearance.

The true conditional efficacy probabilities are provided in Table 1, as well as the CMs. In the Table, the conditional efficacy probabilities are multiplied by (1 − (toxicity rate)). This is because an efficacy response is observed only if a patient does not experience a toxic reaction. For every dose level, the sum of the probabilities equals one. Based on the probabilities in Table 1, a hypothetical trial is simulated.

The first cohort of six patients, treated at d_1, experienced the following: 1 patient had a toxic reaction, four patients did not experience toxicity but did not benefit, and one patient reached a high level of efficacy. Based on these results, the Dirichlet prior of the probabilities associated with d_1 was updated in order to calculate the posterior distribution of its CM, which is plotted in Figure 2A.

TABLE 1 Probabilities of toxicities, efficacy and center of mass (CM) of a hypothetical trial with three dose levels

	k = 0 (toxicity)	k = 1 (no/little efficacy)	k = 2 (medium efficacy)	k = 3 (high efficacy)	CM
d_1	0.10	(1-0.10)*0.70	(1-0.10)*0.20	(1-0.10)*0.10	1.26
d_2	0.20	(1-0.20)*0.40	(1-0.20)*0.30	(1-0.20)*0.30	1.52
d_3	0.30	(1-0.30)*0.10	(1-0.30)*0.20	(1-0.30)*0.70	1.82

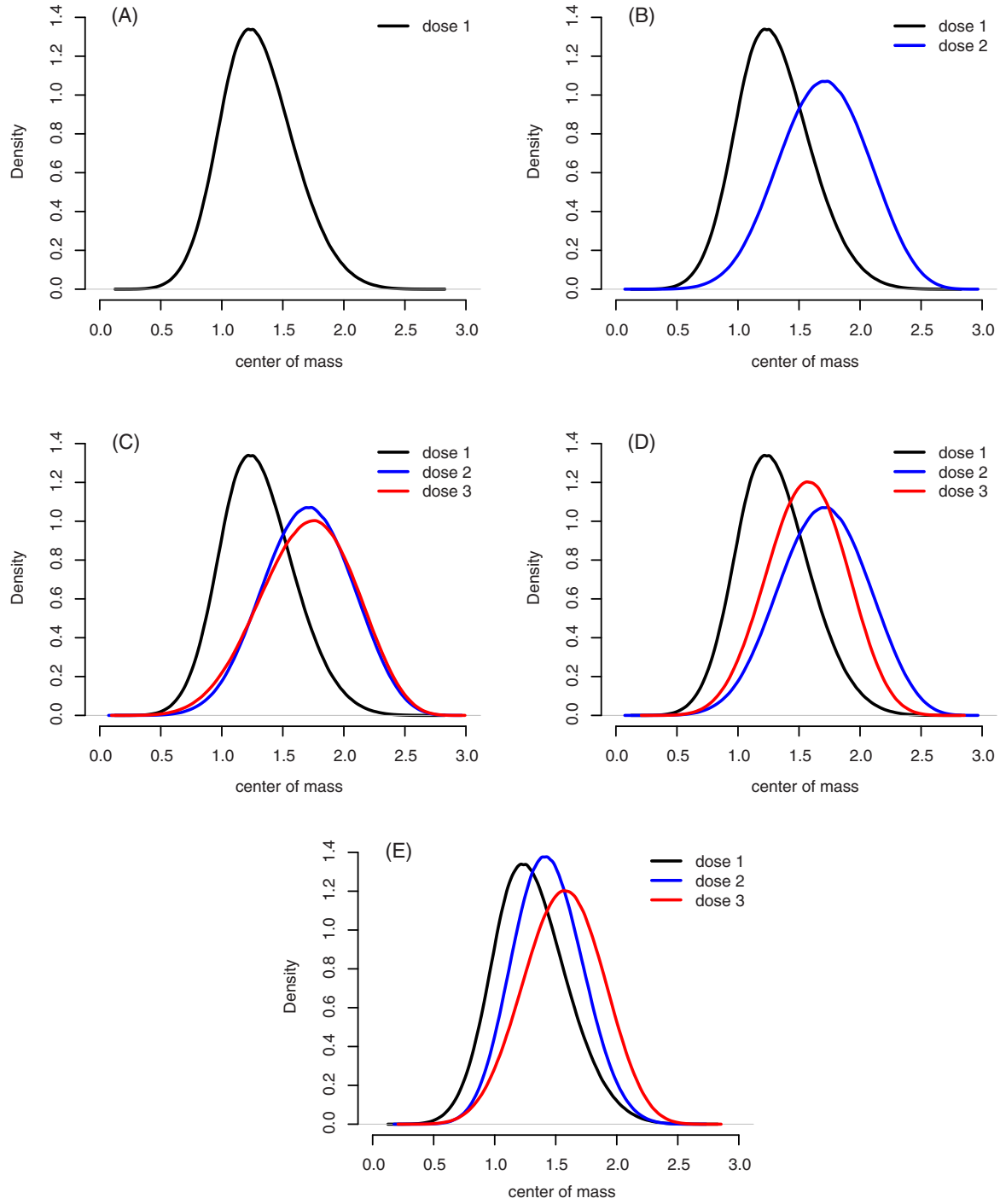

FIGURE 2 Posterior distributions of CMs in hypothetical trial [Color figure can be viewed at wileyonlinelibrary.com]

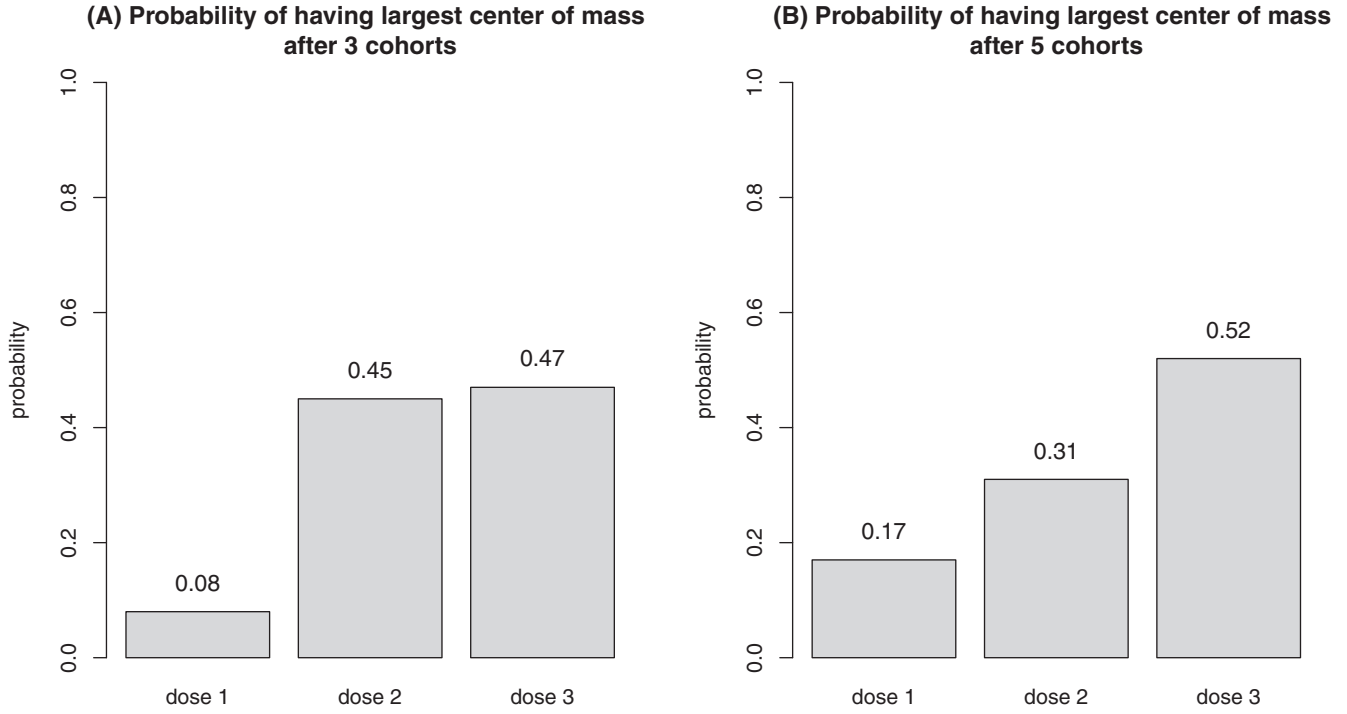

FIGURE 3 Posterior probabilities of having the largest center of mass (CM) in the hypothetical trial

The observed rate of toxicity was 1/6, which is less than 50% (Rmax). Consequently, the second cohort of patient was treated at the second dose level d_2. The following results were observed: one patient experienced toxicity, two patients did not benefit from the treatment, and three patients reached complete viral load clearance. The posterior distributions of the CMs after the first two cohorts is shown in Figure 2B.

The observed toxicity rate at d_2 was below Rmax, and the trial escalated to d_3 with the following observations: two patients experienced toxicity, one reached a medium level of efficacy, and three patients reached a high level of efficacy. The posterior distributions of the CMs at this stage is shown in Figure 2C. The probability that each dose was the dose with the largest CM was calculated, and is shown in Figure 3A.

The next allocation was at the dose level with the highest probability of having the largest CM, given that the observed toxicity rate at d_3 was under the threshold. In this instance, d_3 was repeated for the fourth cohort with the following outcomes: three patients experienced toxicity, one patient reached a medium level of efficacy, and two patients reached a high level of efficacy. Figure 2D shows the posterior distributions of the CMs at this stage of the trial.

Given that the total number of patients with toxicity at d_3 was 5 (two patients from the third cohort and three patients from the fourth cohort experienced toxicity), the observed toxicity rate was less than 50%. Consequently the set of potential doses for the next cohort was still equal to $\{d_1, d_2, d_3\}$. The fifth and last cohorts were allocated to the dose with the highest estimated probability of having the largest CM, which was d_2 with 56% probability. The results were: two patients experienced toxicity, three patients did not reach an efficacy level of clinical significance and one patient reached a high efficacy level. The posterior distributions at the end of the trial are shown in Figure 2E.

At the end of the trial with 30 patients in five cohorts of six patients, d_3 had the highest probability of being the one with the largest CM at 52%, as shown in Figure 3B. Overall six patients were treated at d_1, 12 patients at d_2, and 12 at d_3.

4.2 | Simulations of six different scenarios

The following simulations are based on a trial with three potential doses $\{d_1, d_2, d_3\}$ and a sample size of 30 patients. Cohorts of 6 patients are used for the simulations. For every scenario, Rmax is set at 50%.

We consider six different scenarios, which are quantified in Table 2. Each scenario was simulated 10 000 times. In scenario 1, all doses are acceptable as far as the toxicity rate is concerned, and d_3 is the most efficacious dose. In scenario 2, toxicity rates are higher than in scenario 1, but still under Rmax and d_3 has the largest CM. In scenario 3, although

TABLE 2 Probabilities of toxicities, efficacy and center of mass (CM) for six different scenarios

	k=0 (toxicity)	k=1 (no/little efficacy)	k=2 (medium efficacy)	k=3 (high efficacy)	CM
Scenario 1					
d_1	0.10	(1-0.10)*0.80	(1-0.10)*0.10	(1-0.10)*0.10	1.17
d_2	0.20	(1-0.20)*0.40	(1-0.20)*0.30	(1-0.20)*0.30	1.52
d_3	0.30	(1-0.30)*0.10	(1-0.30)*0.10	(1-0.30)*0.80	1.89
Scenario 2					
d_1	0.15	(1-0.15)*0.75	(1-0.15)*0.20	(1-0.15)*0.05	1.10
d_2	0.30	(1-0.30)*0.60	(1-0.30)*0.30	(1-0.30)*0.10	1.05
d_3	0.45	(1-0.45)*0.05	(1-0.45)*0.20	(1-0.45)*0.75	1.48
Scenario 3					
d_1	0.20	(1-0.20)*0.70	(1-0.20)*0.20	(1-0.20)*0.10	1.12
d_2	0.40	(1-0.40)*0.10	(1-0.40)*0.30	(1-0.40)*0.60	1.50
d_3	0.70	(1-0.70)*0.10	(1-0.70)*0.10	(1-0.70)*0.80	0.81
Scenario 4					
d_1	0.30	(1-0.30)*0.80	(1-0.30)*0.10	(1-0.30)*0.10	0.91
d_2	0.65	(1-0.65)*0.40	(1-0.65)*0.30	(1-0.65)*0.30	0.67
d_3	0.80	(1-0.80)*0.10	(1-0.80)*0.10	(1-0.80)*0.80	0.54
Scenario 5					
d_1	0.55	(1-0.55)*0.80	(1-0.55)*0.10	(1-0.55)*0.10	0.59
d_2	0.75	(1-0.75)*0.40	(1-0.75)*0.30	(1-0.75)*0.30	0.48
d_3	0.90	(1-0.90)*0.10	(1-0.90)*0.10	(1-0.90)*0.80	0.27
Scenario 6					
d_1	0.05	(1-0.05)*0.80	(1-0.05)*0.10	(1-0.05)*0.10	1.24
d_2	0.06	(1-0.06)*0.95	(1-0.06)*0.03	(1-0.06)*0.02	1.01
d_3	0.07	(1-0.07)*0.98	(1-0.07)*0.02	(1-0.07)*0.01	0.96

d_3 is the most efficacious when there is no toxicity, the true toxicity rate at d_3 is greater than Rmax. In this scenario, d_2 is the dose with the largest CM. In scenario 4, the toxicity rate of d_2 and d_3 is greater than Rmax, and d_1 has the largest CM. Scenario 5 is one where all doses have unacceptably high toxicity rates. In scenario 6, the toxicity rates are very low (under 10%) and d_1 has the largest CM.

The results of the simulations are reported in Table 3. In all scenarios, the dose with the largest CM was selected more than 60% of the time. In each scenario, more patients were treated at the dose with the largest CM than at the other two doses. Importantly, in scenario 5 where even the toxicity rate of d_1 is greater than Rmax, the algorithm recommended d_1 81% of the time. However, in practice, the trial may not have reached its full sample size of 30 patients. As noted in Section 3.1, a trial may be stopped early after the first few cohorts of patients if the toxicity rate is estimated to be larger than the threshold with sufficient precision, which is trial-specific.

We investigated the influence of the sample size on the performance of the design for the first three scenarios. In Table 4, we report the percentage of simulations that identified the dose with the largest CM for sample sizes of 24, 30, 36, 42, and 48 patients in cohorts of six patients.

The performance of the dose-finding design increases with sample size. However, the incremental gain in performance declines progressively. Doubling the sample size, from 24 to 48 patients, increased this chance by less than 10 percentage points in all three scenarios. Increasing the sample size from 42 to 48 patients increased the chance to identify the dose with the largest CM by 1 percentage point, in all three scenarios. The proportion of patients receiving the dose with the largest CM also increases with sample size.

TABLE 3 Recommendation and in-trial allocation for the six scenarios

Scenario		d_1	d_2	d_3
1	% recommendation for dose	0.03	0.22	0.75
	% of patients receiving dose	0.21	0.29	0.50
2	% recommendation for dose	0.22	0.15	0.63
	% of patients receiving dose	0.24	0.33	0.43
3	% recommendation for dose	0.20	0.72	0.08
	% of patients receiving dose	0.28	0.49	0.23
4	% recommendation for dose	0.62	0.19	0.19
	% of patients receiving dose	0.45	0.37	0.18
5	% recommendation for dose	0.81	0.15	0.04
	% of patients receiving dose	0.82	0.13	0.05
6	% recommendation for dose	0.71	0.17	0.12
	% of patients receiving dose	0.48	0.28	0.24

TABLE 4 Recommendations for the dose with the largest center of mass (CM) and in-trial allocation for the first three scenarios and different sample sizes

Scenario		n = 24	n = 30	n = 36	n = 42	n = 48
1	% recommendation for d_3	0.72	0.75	0.78	0.79	0.80
	% patients receiving d_3	0.43	0.50	0.54	0.58	0.60
2	% recommendation for d_2	0.60	0.63	0.64	0.65	0.67
	% patients receiving d_2	0.36	0.43	0.46	0.48	0.50
3	% recommendation for d_2	0.69	0.72	0.75	0.77	0.78
	% patients receiving d_2	0.43	0.49	0.53	0.57	0.59

TABLE 5 Recommendations for the dose with the largest CM and in-trial allocation for the first three scenarios with a fixed sample size of 30 and different cohort sizes

Scenario		cohort = 3	cohort = 5	cohort = 6
1	% recommendation for d_3	0.68	0.72	0.75
	% patients receiving d_3	0.54	0.50	0.50
2	% recommendation for d_2	0.60	0.62	0.63
	% patients receiving d_2	0.47	0.43	0.43
3	% recommendation for d_2	0.69	0.70	0.72
	% patients receiving d_2	0.55	0.52	0.49

The effect of the cohort size, for a fixed sample size of 30 patients, was also investigated. In Table 5, we report the operating characteristics of the design for cohort sizes of 3, 5, and 6 patients.

4.3 | Comparison of the dose-finding algorithm with a randomized approach

A competing experimental design might simply randomize patients to the available levels without making any use of the accumulating information on the relationship between these levels and CM in order to guide dose allocation. As far as final recommendation is concerned, such a design ought work reasonably well, but would have the disadvantage of increasing the probability of patients being exposed to potentially unsafe doses. Equally, we may anticipate finding too many patients who have been exposed to levels that are insufficiently effective. It is, nonetheless, instructive, via simulations to compare

TABLE 6 Comparison between dose-finding algorithm and randomization in terms of recommendation for the dose with the largest center of mass (CM) and in-trial allocation for the first three scenarios with three different sample sizes

		n= 24		n= 30		n= 36	
Scenario		Dose-finding	Randomisation	Dose-finding	Randomization	Dose-finding	Randomization
1	% recommendations for d_3	0.72	0.71	0.75	0.73	0.78	0.76
	% of patients receiving d_3	0.43	0.33	0.50	0.33	0.54	0.33
2	% recommendations for d_2	0.60	0.67	0.63	0.70	0.64	0.73
	% of patients receiving d_2	0.36	0.33	0.43	0.33	0.46	0.33
3	% recommendations for d_2	0.69	0.69	0.72	0.72	0.75	0.75
	% of patients receiving d_2	0.43	0.33	0.49	0.33	0.53	0.33

the proposed dose-finding algorithm with a randomized design in terms of performance (ie, proportion of studies that correctly recommend the dose with the largest CM). This exercise was done purely for statistical comparison purposes since, in practice, we would not advocate a randomization strategy that deliberately makes no use of information on the possible relationship between dose level and outcomes during the course of the study.

Three sample sizes were used for this comparison: 24, 30, and 36 patients. In the case of randomization, a third of the patients are randomly allocated to each of the three doses $\{d_1, d_2, d_3\}$. For the algorithmic approach developed in Sections 2 and 3, cohorts of six patients were used. Table 6 reports the results. In scenarios 1 and 3, the performance of the randomized design was very similar to that of the algorithmic design. In scenario 2, the randomized design selected the best dose slightly more frequently than the dose-finding design. For example, where the algorithmic design selected the best dose in 64% of the simulations with six cohorts of six patients, a randomized approach with 12 patients per dose recommended the best dose in 73% of the simulations. However, the algorithmic design allocated 46% of the patients on average to the best dose, while the randomized approach allocated only a third of the patients to that dose.

5 | DISCUSSION

The simulations show that the performance of the proposed dose-finding methodology is very encouraging in a range of realistic scenarios. The optimal dose was selected in more than 60% of the simulations. Considering the dose with the largest center of mass and the second best dose, the algorithm recommended one of these two doses more than 75% of the time.

Increased precision and more effective allocation may result from a greater degree of parameterization in our proposed class of design. Dose-finding in virology, given the nature of the anticipated potential adverse reactions, will differ from oncology in that there is not such a great need to appeal to models to help the investigators keep a strict control on these adverse events. It would seem sufficient to work with empirical estimates and to simply ensure that, as in the current study, we can keep the adverse events (corresponding to the inability to complete and obtain any benefit from the treatment) to below some level. For the particular study of interest that motivated this publication, that of combating the norovirus, the investigators felt that in the search for as much efficacy as possible, any amount of adverse events lower than 50% would be acceptable. This helped guide our initial dose escalation strategy. In virology, efficacy is our driving concern and, here, introducing some parametric structure into the algorithm may potentially increase precision and more effective allocation. Efficiency gains could follow but, as always, these would be contingent upon a satisfactory fit and this would necessarily form a part of the approach.

We assessed the effect on the operating characteristics of the sample size as well as of the cohort size. Doubling the size from 24 patients to 48 patients increased the percentage of simulations that recommended the best dose, but by less than 10 percentage points across the range of scenarios. It also increased the percentage of patients allocated to the best dose. Changing the cohort size from six patients to five or three patients decreased the recommendation percentage of the best dose, but only by a small amount of 7 percentage points. Simultaneously this increased the percentage of patients allocated to the best dose. Consequently when designing a study, there is a need to consider the appropriate

balance between cohort size and operating characteristics on the one hand, and other logistical considerations on the other hand. Indeed a smaller cohort size may render the trial more adaptive to unfolding results as well as potential external changes.

The simulations showed that the performance was close, as far as the accuracy of dose-finding recommendations is concerned, to that of a randomized approach in two of the three scenarios that we considered, and was marginally less good in one scenario. This is due to the adaptive nature of the dose-finding approach. The allocation of patients during the study depends on previous observations. While the estimates of toxicity and efficacy may be unstable at the beginning of the study, these may converge quickly if the doses have markedly different toxicity and efficacy profiles. In comparison the fixed nature of the randomized approach prevents the progressive convergence towards the region where the best dose lies. From the patient's perspective, this is a significant improvement as the patient has a much higher chance than in the randomized design to be allocated to an optimal dose, or at least a dose near the optimal region.

We anticipate that monotonicity, both in terms of efficacy and toxicity with respect to the doses, will apply in the majority of circumstances. If, among a set of doses, there exists an optimal dose M that provides the best balance between toxicity and efficacy that can be identified, then we assume that doses higher than M will show too high an increased rate of toxicity to compensate for any gains in viral load reduction whereas, for doses lower than M, despite anticipated lower rates of toxicity, the rate of viral load reduction may not be sufficient. Nonetheless, we have simulated one scenario where monotonicity was not respected. In this scenario with three doses, the second dose was less toxic and more efficacious than the third dose. It was found that the second dose was selected in approximately 70% of the simulations. Our sixth scenario is also one where monotonicity was not respected efficacy-wise.

The sixth scenario of our simulations reflects a situation where the recommended dose may be safe but of little value as far as efficacy is concerned. In such a situation, in practice the Data Safety and Monitoring Board (DSMB) may decide, based on the totality of the evidence, that the dose-finding trial should be stopped early for lack of benefit. The role of the DSMB is also critical in cases where even the first dose level (or other dose levels) may be too toxic. We recommend that trialists specify in advance the statistical trigger(s) that may terminate the study early due to toxicity concerns.

The proposed dose-finding methodology exploits the Bayesian Dirichlet-Categorical model to estimate the center of mass of the distribution of responses. It is important to note that this means that the total effective sample size is the sum of the observed sample size plus the sum of the prior parameters of the Dirichlet distribution. It is possible to follow the dynamic dose-finding algorithm described in Section 3.1 purely on the observed proportions of patients in the different response categories. However, the introduction of the Bayesian prior allows the quantification, and visual representation, of the uncertainty about the distribution of responses and its center of mass.

There are few publications in virology on the important topic of dose-finding methodology. In the study of O'Quigley et al,[14] efficacy was dichotomized into a yes/no variable representing sufficient viral load reduction. In the method described by Mason et al,[15] a CRM approach was employed but instead of looking at DLTs, inefficacy, defined as a binary variable, was the main endpoint. To our knowledge, our approach is the first to allow efficacy to be represented differently in the context of early phase dose-finding trials in virology. Given the specificities of virology, in which grades of efficacy are expected, this development provides a more accurate representation of the clinical reality. As previously stated, even medium levels of efficacy may allow immune-mediated clearance, or form the basis for the introduction of additional complementary treatment. At the end of any given dose-finding trial, there remains uncertainty in the selection of the best dose. In our simulated hypothetical trial described in Section 4, the dose that was recommended at the end of the trial had a 52% chance of being the dose with the largest center of mass. It may be advisable to complement a dose-finding trial in virology with a randomized comparison of the doses that are ranked first, second, and perhaps even third, at the end of the dose-finding study. In the hypothetical trial, by selecting the second and third doses for a randomized comparison, the probability that one of these doses is the one with the largest CM is 84%. In other words, we can see the dose-finding trial as an elimination exercise prior to randomization. Doses that show themselves to be poor candidates for clinical application would not enter the second phase of experimentation. The calculation of the required sample size for this second phase is currently a matter of ongoing research in our group. Dehbi and Hackshaw[16] have published on sample size calculation for binary endpoints in selection trials. This method may be extended to categorical endpoints. Alternatively the randomized phase may be based on a binary endpoint, possibly the proportion of patients that reach a satisfactory reduction in viral load that ensures a lasting effect.

5.1 | More aggressive viruses including SARS-CoV-2

The methods described here apply generally to dose-finding in virology. The more aggressive viruses associated with significant mortality would come under that heading. At the same time, further improvements to the accuracy and overall performance of the dose-finding design can be found by incorporating any new known specificities into the design structure. We mention here some of the features of more aggressive viruses that might better guide our construction of a dose-finding algorithm.

1. *Urgency*. A more deadly, highly contagious, virus raises the issue of urgency. We may be facing a major public health crisis for which the usual clinical trials paradigm—Phase I, Phase II, and then Phase III—simply fails. It is too slow and, therefore, not fit for purpose. Recent developments from dose finding in oncology, in particular the inclusion of expansion cohorts blurs to some degree the distinction between the early phase and the later phase trials.[1,2] The inclusion of randomization to the early phase trial allows us to greatly accelerate the development process, to learn as we proceed, treating as best we can those patients in the trial while obtaining ever greater precision on our estimates of the most effective dose to be used in practice.
2. *Endpoints*. The usual endpoints are toxicity and efficacy. Already, toxicity is defined differently in virology than in oncology and we may again need consider the definition for more aggressive viruses. The next bullet point is also important in that the definition of toxicity may not be the same across a whole population. As far as efficacy measures are concerned, for norovirus, we have worked with some simple groupings. These groupings describe the degree of reduction in viral load. For large studies that may be called for in the context of a serious public health issue, these groupings could be refined. The impact of new definitions on the ability of the design to accurately identify the most effective dose would need to be studied.
3. *Patient heterogeneity*. The more aggressive and deadly viruses are likely to behave differently with respect to different patient groups. Those at greatest risk may well be very much in the minority. Potentially such a group would have the most to gain by effective treatment. As a result the high risk group is likely to have a different toxicity threshold than the low risk group: there being more at stake they are likely to be willing to accept a higher degree of side effects. We might refer to this as heterogeneity in the toxicity threshold. But we may also need consider heterogeneity in the criteria by which we judge efficacy since, for example, a younger patient population may respond well by a relatively modest reduction in viral load but that such a reduction, in a high risk group, may be inadequate. Carrying out separate studies in the different risk groups is always an option. However, when possible, bridging between related studies can provide considerable gains in efficiency. Several examples from oncology could be a useful starting point.

ACKNOWLEDGMENTS

The authors would like to thank the reviewers and editors of Statistics in Medicine, whose detailed comments have improved the manuscript markedly.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study. We only used simulated data.

ORCID

Hakim-Moulay Dehbi https://orcid.org/0000-0002-0816-0178

REFERENCES

1. Iasonos A, O'Quigley J. Clinical trials: early phase clinical trials-are dose expansion cohorts needed? *Nat Rev Clin Oncol*. 2015;12(11):626.
2. Iasonos A, O'Quigley J. Dose expansion cohorts in phase I trials. *Stat Biopharm Res*. 2016;8(2):161-170.
3. Asakawa T, Hamada C. A pragmatic dose-finding approach using short-term surrogate efficacy outcomes to evaluate binary efficacy and toxicity outcomes in phase I cancer clinical trials. *Pharm Stat*. 2013;12(5):315-327.

4. Brock K, Billingham L, Copland M, Siddique S, Sirovica M, Yap C. Implementing the EffTox dose-finding design in the Matchpoint trial. *BMC Med Res Methodol*. 2017;17(1):112.
5. Chiuzan C, Garrett-Mayer E, Nishimura MI. An adaptive dose-finding design based on both safety and immunologic responses in cancer clinical trials. *Stat Biopharm Res*. 2018;10(3):185-195.
6. Jacobson IM, Lawitz E, Gane EJ, et al. Efficacy of 8 weeks of sofosbuvir, velpatasvir, and voxilaprevir in patients with chronic HCV infection: 2 phase 3 randomized trials. *Gastroenterology*. 2017;153(1):113-122.
7. Emmanuel B, Wilson EM, O'Brien TR, Kottilil S, Lau G. Shortening the duration of therapy for chronic hepatitis C infection. *Lancet Gastroenterol Hepatol*. 2017;2(11):832-836.
8. Brown LAK, Clark I, Brown JR, Breuer J, Lowe DM. Norovirus infection in primary immune deficiency. *Rev Med Virol*. 2017;27(3):e1926.
9. Brown LAK, Ruis C, Clark I, et al. A comprehensive characterization of chronic norovirus infection in immunodeficient hosts. *J Allergy Clin Immunol*. 2019;144(5):1450-1453.
10. Arias A, Thorne L, Goodfellow I. Favipiravir elicits antiviral mutagenesis during virus replication in vivo. *Elife*. 2014;3:e03679.
11. Jin Z, Tucker K, Lin X, et al. Biochemical evaluation of the inhibition properties of favipiravir and 2'-C-methyl-cytidine triphosphates against human and mouse norovirus RNA polymerases. *Antimicrob Agents Chemother*. 2015;59(12):7504-7516.
12. Ruis C, Brown LAK, Roy S, et al. Mutagenesis in norovirus in response to favipiravir treatment. *N Engl J Med*. 2018;379(22):2173-2176.
13. 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.
14. O'Quigley J, Hughes MD, Fenton T. Dose-finding designs for HIV studies. *Biometrics*. 2001;57(4):1018-1029.
15. Mason AJ, Gonzalez-Maffe J, Quinn K, et al. Developing a Bayesian adaptive design for a phase I clinical trial: a case study for a novel HIV treatment. *Stat Med*. 2017;36(5):754-771.
16. Dehbi HM, Hackshaw A. Sample size calculation in randomised phase II selection trials using a margin of practical equivalence. *Trials*. 2020;21:1-7.

How to cite this article: Dehbi H-M, Lowe DM, O'Quigley J. Early phase dose-finding trials in virology. *Statistics in Medicine*. 2020;1–14. https://doi.org/10.1002/sim.8771