Statistics in Medicine

Early Phase Dose-Finding Trials in Virology

Journal:	*Statistics in Medicine*
Manuscript ID	SIM-20-0236
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	02-Apr-2020
Complete List of Authors:	Dehbi, Hakim-Moulay; University College London, Comprehensive Clinical Trials Unit Lowe, David; University College London, Comprehensive Clinical Trials Unit O'Quigley, John; University College London, Comprehensive Clinical Trials Unit
Keywords:	virology, dose-finding, continual reassessment method, Norovirus, Coronavirus

ORIGINAL ARTICLE

Journal Section

Early Phase Dose-Finding Trials in Virology

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

[1]Comprehensive Clinical Trials Unit at UCL, University College London

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

Correspondence

Hakim-Moulay Dehbi PhD, Comprehensive Clinical Trials Unit at UCL, 90 High Holborn 2nd Floor London WC1V 6LJ, United Kingdom

Email: h.dehbi@ucl.ac.uk

Funding information

Among the many lessons we are learning from the crisis surrounding COVID-19 is the near absence of dose-finding methods specifically designed for trials in virology. Aside from one or two papers focused on HIV, the considerable progress in dose-finding methodology of the last 25 years has focused almost entirely on oncology. While adverse reactions to cytotoxics 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. 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 paper is an upcoming dose-finding trial in chronic norovirus infection. We propose a new 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 maximises the treatment benefit for the patients included in the study. Finally, we add a section that discusses the relevance of this work to more aggressive viruses, including COVID-19.

KEYWORDS

virology, dose-finding, early phase clinical trials, continual reassessment method, Norovirus, Coronavirus

1 | INTRODUCTION

This introduction has 3 subsections: the first provides a broad description of our goals, the second a detailed 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.

The focus of dose-finding methodology in oncology 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, 4, 5]. In oncology it is conceivable that a patient suffer 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. In the case of interest here, 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, toxicity in a virology study means that the patient loses the opportunity to benefit from the treatment. Clearly the rate of toxicity and its dependence on dose is a major concern but, given this rate to be acceptable, our efforts will only look at rates of viral load reduction and the dependence of this on dose.

The best dose, M, should show an optimal compromise between the goals of minimizing toxicity and maximising 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, they are allocated a dose believed to be a good candidate dose for being the best from the current set of available doses.

1.2 | Motivating example: 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 utilisation and death [6]. Currently no proven treatment exists [7]. 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 [8, 9]. 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 [10]. 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 four. The primary efficacy endpoint will be the reduction of stool viral load. Patients will be treated for 8 weeks in the absence of dose limiting toxicities. Full elimination of the virus will be the best possible response, and the minimum difference of interest will be an increase in the Norovirus CT 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) [11], 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, i.e. 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; (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, Hughes and Fenton (2001) [12] for a specific dose-finding study in HIV. 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 DESIGNS

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 also assume a monotonicity property to hold across the doses with regard to the conditional probability of achieving efficacy in terms of viral load reduction given no toxicity. 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 enter. The dose level for the j^{th} 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 j^{th} 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, $Q(v \mid x_j)$, over the 3 categories. We write:

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

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

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

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, i.e., $\Pr(V_j = 0 | Y_j = 1) = 1$. As a result we have that $E(V_j \mid x_j, Y_j = 1) = 0$. 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:

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

The unconditional expectation of efficacy is referred to as the centre of mass (CM) of the distribution of V_j. Note that whereas $E(V_j \mid x_j, Y_j = 0)$ lies between 1 and 3, $E(V_j \mid x_j)$ will lie between 0 and 3. Low values of $E(V_j \mid x_j)$, close to zero indicate high toxicity and/or effects that are too weak to be of clinical relevance. Values of $E(V_j \mid x_j)$ close to 3 indicate effects that are being met with 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 three 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

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

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 $E(V_j \mid d_i), \; i = 1, ..., m$. The level ℓ that provides the maximum of these corresponds to our current best estimate for M. In practice such an empirical approach could be achieved by random allocation between doses. At each observation or group of observations estimates of $R(d_i)$, $Q(v \mid d_i)$, $E(V \mid d_i)$ can be updated and, finally, $\hat{M}$ based on the empirical rates. We would anticipate these as being relatively unstable early on in the study when few patients are treated. Bayesian statistics may be helpful in this instance,

via the addition of prior distributions which may even lean on expert input. The main disadvantage of this 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 randomisation, we propose a dose-finding algorithm that does not depend on modeling to determine, with sufficient precision, estimates of $E(V_j \mid d_i)$, $i = 1, \ldots m$.

The proposed dose-finding algorithm requires the definition of the maximum allowable threshold for toxicity, which we denote as R_{max}. This threshold represents the highest proportion of patients with a toxic reaction that can be allowed for the disease-drug combination of interest.

First cohort of patients at the first dose level d_1

The first cohort of patients is treated at the first (i.e. lowest) dose level d_1 and the observed toxicity rate is compared with R_{max}. As long as the 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. This process would be defined by the investigators prior to the start of the study.

First cohort of patients at the second dose level d_2 and subsequent cohorts

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

1. if the rate is less than R_{max}, 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 rate is larger than R_{max}, the set of potential doses for the next cohort corresponds to all prior doses (i.e. all doses up to, but excluding, the current dose level);
3. if the rate is equal to R_{max}, 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 estimated CM among the possible dose set. However, until there is experimentation on all doses, the estimated CM is unknown at

the doses that have not yet been tested. To allow the dose-finding algorithm to proceed, a dose is allowed to be tested for the first time if the observed toxicity rate of the dose level below it is less than R_{max}, even though its CM is unknown at that point. For example, if the first and second cohort are treated at d_1 and d_2 respectively, the third cohort would be treated at d_3 as long as the toxicity rate at d_2 was less than R_{max}.

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 the each dose being the one with the largest CM.

Estimation of the centre 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 $E(V_j \mid d_i),\ i = 1, \ldots m$, 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 $V_j \mid P_i$ follows a categorical distribution denoted as Cat (k, P_i), where P_i follows a Dirichlet distribution and $k = 4$. 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 generalisation of the uniform distribution, with the additional characteristic that the sum of the probabilities equals one. In other words, for the prior distribution of P_i we use a flat Dirichlet distribution, $P_i \mid \alpha \sim Dir(k, \alpha)$.

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 $v_i = (v_{0i}, v_{1i}, , v_{(k-1)i})$. The posterior distribution of P_i is $P_i \mid \alpha, v_i \sim Dir(k, \alpha + v_i) = Dir(k, 1 + v_{0i}, 1 + v_{1i}, , 1 + v_{(k-1)i})$.

By sampling from the posterior distribution of P_i, the posterior distribution of the CM at d_i can be derived by recording the value of the linear combination presented in equation (4). Independent sampling from posterior distributions at all dose levels allows the derivation of the probability that each dose is the one with the largest CM by recording which dose has the largest CM at each sampling iteration.

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, R_{max}, is set at 50 percent. 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 6 patients, treated at d_1, experienced the following: 1 patient had a toxic reaction, 4 patients did not experience toxicity but did not benefit, and 1 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 d_1s CM, which is plotted in Figure 1a).

The observed rate of toxicity was 1/6, which is less than 50% (R_{max}). 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 1b).

The observed toxicity rate at d_2 was below R_{max}, 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 1c). The probability that each dose was the dose with the largest CM was calculated, and is shown in Figure 2a).

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 1d) 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 (2 patients from the third cohort and 3 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 cohort was 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 CMs posterior distributions at the end of the trial is shown in Figure 1e).

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 2b). Overall six patients were treated at d_1, twelve patients at d_2 and twelve at d_3.

4.2 | Simulations

The following simulations are based on a trial with 3 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, R_{max} is set at 50%.

We consider three 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, although d_3 is the most efficacious when there is no toxicity, the true toxicity rate at d_3 is larger than R_{max}. In this scenario, d_2 is the dose with the largest CM. In scenario 3, d_2 has the largest CM. The results of the simulations are reported in Table 3.

In all three scenarios the dose with the largest CM was selected more than 60% of the time. In each scenario more patients are treated at the dose with the largest CM than at the other two doses.

We investigated the influence of the sample size on the performance of the design. 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 6 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.

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 the three scenarios and cohort sizes of 3, 5 and 6 patients.

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

Randomisation at the start of a dose-finding study in virology has the crucial disadvantage of exposing patients to potentially unsafe doses. Equally, too many patients may be exposed to inefficacious doses. Nonetheless, via simulations we compared the proposed dose-finding algorithm with a randomised design in terms of performance (i.e. proportion of studies that correctly recommend the dose with the largest CM). This comparison was made purely for statistical comparison purposes. In practice we do not advocate randomisation at the start of a dose-finding trial in virology.

Three sample sizes were used for this comparison: 24, 30 and 36 patients. In the case of randomisation, a third of the patients are randomly allocated to each of the three doses $\{d_1, d_2, d_3\}$. For the dose-finding approach cohorts of 6 patients were used. Table 6 reports the results.

In scenarios 1 and 3, the performance of the randomised design was very similar to that of the dose-finding design. In scenario 2, the randomised design selected the best dose slightly more frequently than the dose-finding design. For example, where the dose-finding design selected the best dose in 64% of the simulations with 6 cohorts of 6 patients, a randomised approach with 12 patients per dose recommended the best dose in 73% of the simulations. However, the dose-finding design allocated 46% of the patients on average to the best dose, while the randomised approach allocated only a third of the patients.

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 centre 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. For example, using multinomial regression, the ordered outcomes may be modelled. 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 6 patients to 5 or 3 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 also showed that the dose-finding design's performance was close to that of a randomised 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 randomised approach prevents the progressive convergence towards the region where the best dose lies. From the patient's perspective, this may be a significant improvement as the patient has a much higher chance than in the randomised 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 to assess the performance of the proposed dose-finding methodology in such circumstances. 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.

There are few publications in virology on the important topic of dose-finding methodology. In the study of O'Quigley, Hughes and Fenton (2001) [12], efficacy was dichotomized into a yes/no variable representing sufficient viral load reduction. In the method described by Mason et al. (2016) [13], 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. 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 mentioned, 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 centre of mass. It may be advisable to complement a dose-finding trial in virology with a randomised 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 randomised 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 randomisation. 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.

5.1 | More aggressive viruses including COVID-19

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, in particular COVID-19, 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 1 followed by Phase 2 followed by Phase 3 - simply fails. It is much too slow and, therefore, not fit for purpose. Recent developments from dose finding in oncology, in particular the inclusion of expansion cohorts (refs Iasonos etc) blurs to some degree the distinction between the early phase and the later phase trials. 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 any tolerance 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 (Iasonos and O'Quigley 2016).

CONFLICT OF INTEREST

No conflicts of interest to report.

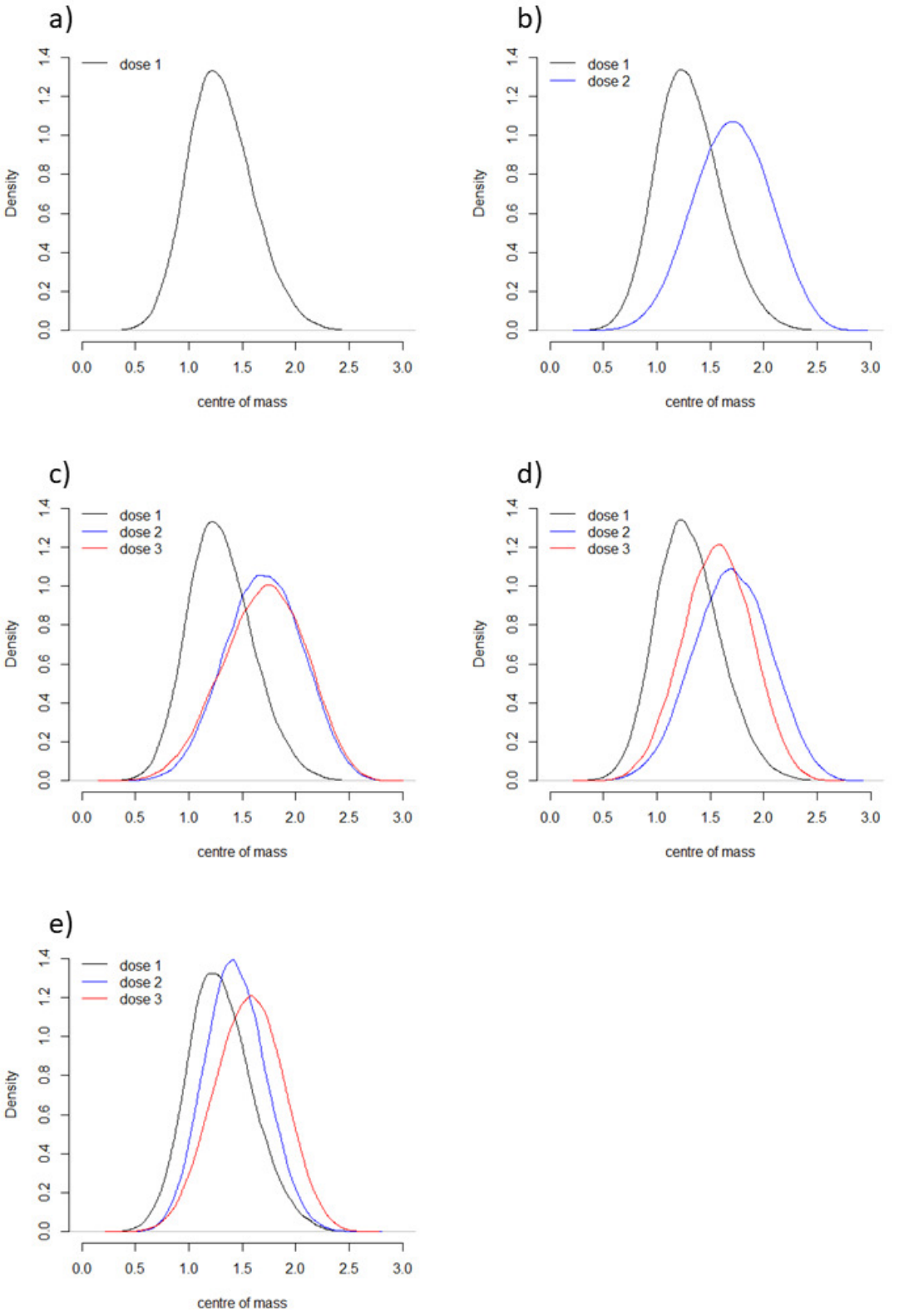

FIGURE 1 Posterior distributions of CMs in hypothetical trial

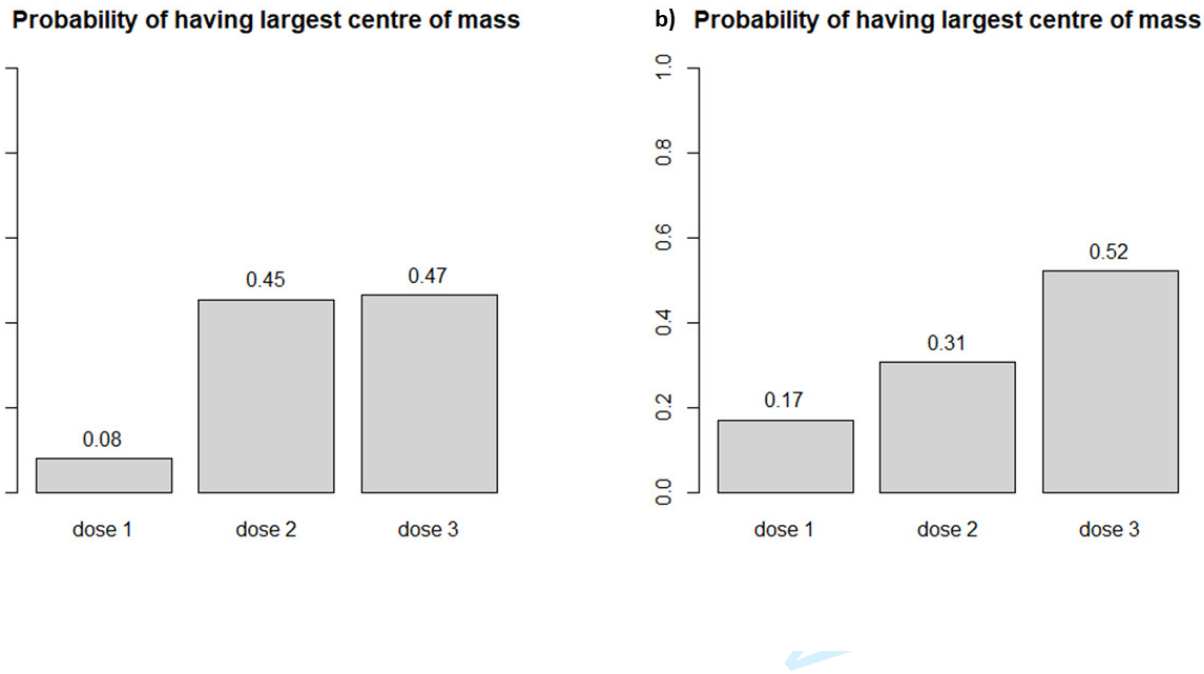

FIGURE 2 Posterior probabilities of having the largest CM

REFERENCES

[1] Iasonos A, O'Quigley J. Clinical trials: early phase clinical trials—are dose expansion cohorts needed? Nature reviews Clinical oncology 2015;12(11):626.

[2] Iasonos A, O'Quigley J. Dose expansion cohorts in phase I trials. Statistics in biopharmaceutical research 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. Pharmaceutical statistics 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 medical research methodology 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. Statistics in biopharmaceutical research 2018;10(3):185–195.

[6] Brown LAK, Clark I, Brown JR, Breuer J, Lowe DM. Norovirus infection in primary immune deficiency. Reviews in medical virology 2017;27(3):e1926.

[7] Brown LAK, Ruis C, Clark I, Roy S, Brown JR, Albuquerque AS, et al. A comprehensive characterization of chronic norovirus infection in immunodeficient hosts. Journal of Allergy and Clinical Immunology 2019;144(5):1450–1453.

[8] Arias A, Thorne L, Goodfellow I. Favipiravir elicits antiviral mutagenesis during virus replication in vivo. Elife 2014;3:e03679.

[9] Jin Z, Tucker K, Lin X, Kao CC, Shaw K, Tan H, et al. Biochemical evaluation of the inhibition properties of favipiravir and 2'-C-methyl-cytidine triphosphates against human and mouse norovirus RNA polymerases. Antimicrobial agents and chemotherapy 2015;59(12):7504–7516.

[10] Ruis C, Brown LAK, Roy S, Atkinson C, Williams R, Burns SO, et al. Mutagenesis in norovirus in response to favipiravir treatment. New England Journal of Medicine 2018;379(22):2173–2176.

[11] O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. Biometrics 1990;p. 33–48.

[12] O'Quigley J, Hughes MD, Fenton T. Dose-finding designs for HIV studies. Biometrics 2001;57(4):1018–1029.

[13] Mason AJ, Gonzalez-Maffe J, Quinn K, Doyle N, Legg K, Norsworthy P, et al. Developing a Bayesian adaptive design for a phase I clinical trial: a case study for a novel HIV treatment. Statistics in medicine 2017;36(5):754–771.

TABLE 1 Probabilities of toxicities, efficacy and centre of mass (CM) of a hypothetical trial with three doses

	k=0 (toxicity)	k=1 (no/little efficacy)	k=2 (medium efficacy)	k=3 (high efficacy)	CM
d_1	10%	(1-10%)*70%	(1-10%)*20%	(1-10%)*10%	1.26
d_2	20%	(1-20%)*40%	(1-20%)*30%	(1-20%)*30%	1.52
d_3	30%	(1-30%)*10%	(1-30%)*20%	(1-30%)*70%	1.82

TABLE 2 Probabilities of toxicities, efficacy and centre of mass (CM) for three different scenarios

	k=0 (toxicity)	k=1 (no/little efficacy)	k=2 (medium efficacy)	k=3 (high efficacy)	CM
Scenario 1					
d_1	10%	(1-10%)*80%	(1-10%)*10%	(1-10%)*10%	1.17
d_2	20%	(1-20%)*40%	(1-20%)*30%	(1-20%)*30%	1.52
d_3	30%	(1-30%)*10%	(1-30%)*10%	(1-30%)*80%	1.89
Scenario 2					
d_1	15%	(1-15%)*75%	(1-15%)*20%	(1-15%)*5%	1.10
d_2	30%	(1-30%)*60%	(1-30%)*30%	(1-30%)*10%	1.05
d_3	45%	(1-45%)*5%	(1-45%)*20%	(1-45%)*75%	1.48
Scenario 3					
d_1	20%	(1-20%)*70%	(1-20%)*20%	(1-20%)*10%	1.12
d_2	40%	(1-40%)*10%	(1-40%)*30%	(1-40%)*60%	1.50
d_3	70%	(1-70%)*10%	(1-70%)*10%	(1-70%)*80%	0.81

TABLE 3 Recommendation and in-trial allocation for the three 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

TABLE 4 Recommendations for the dose with the largest centre of mass (CM) and in-trial allocation for the 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 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

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

Scenario		n = 24 Dose-finding	Randomisation	n = 30 Dose-finding	Randomisation	n = 36 Dose-finding	Randomisation
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