

Designing and evaluating dose-escalation studies made easy: the MoDEsT web app

Journal Title
XX(X):1-7
©The Author(s) 2017
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/


Philip Pallmann¹, Fang Wan², Adrian P Mander³, Graham M Wheeler^{3,4}, Christina Yap⁵, Sally Clive⁶, Lisa V Hampson⁷ and Thomas Jaki²

Abstract

Background/aims: Dose-escalation studies are essential in the early stages of developing novel treatments, when the aim is to find a safe dose for administration in humans. Despite their great importance, many dose-escalation studies use study designs based on heuristic algorithms with well-documented drawbacks. Bayesian decision procedures provide a design alternative that is conceptually simple and methodologically sound, but very rarely used in practice, at least in part due to their perceived statistical complexity. There are currently very few easily accessible software implementations that would facilitate their application.

Methods: We have created MoDEsT, a free and easy-to-use web application for designing and conducting single-agent dose-escalation studies with a binary toxicity endpoint, where the objective is to estimate the maximum tolerated dose. MoDEsT uses a well-established Bayesian decision procedure based on logistic regression. The software has a user-friendly point-and-click interface, makes changes visible in real time, and automatically generates a range of graphs, tables, and reports. It is ~~specifically~~ aimed at clinicians ~~as well as statisticians with limited expertise in model-based dose-escalation designs who may have limited access to statistical methodology support~~, and does not require ~~an in-depth understanding of statistics~~ any statistical programming skills to evaluate the operating characteristics of, or implement, the Bayesian dose-escalation design.

Results: MoDEsT comes in two parts: a 'Design' module to explore design options and simulate their operating characteristics, and a 'Conduct' module to guide the dose-finding process throughout the study. We illustrate the practical use of both modules with data from a real phase I study in terminal cancer.

Conclusion: Enabling ~~both methodologists and~~ clinicians to understand and apply model-based study designs with ease is a key factor towards their routine use in early-phase studies. We hope that MoDEsT will enable incorporation of Bayesian decision procedures for dose escalation at the earliest stage of clinical trial design, thus increasing their use in early-phase trials.

Keywords

Phase I clinical trial, dose-finding study, logistic model, Bayesian statistics, graphical user interface, shiny app, R software

Introduction

The primary aim of many phase I dose-escalation studies is to estimate the maximum tolerated dose (~~MTD~~) of a novel drug or treatment. In practice this often means identifying a dose for which the probability of a patient developing a dose-limiting toxicity (~~DLT~~) is close to a prespecified target toxicity level (~~FTL~~), typically between 0.20 and 0.33 in cancer trials. Patients enter the study in cohorts of one or more (usually three), and for every new cohort a decision is made whether to stay at the current dose level, escalate or de-escalate the dose, or stop the study entirely. A statistical study design informs and guides this process, but the ultimate decision will always be based on clinical judgement.

It is good practice that early-phase dose escalation designs use a well-fitting statistical model to synthesise all available information (from prior knowledge and accumulating patient data) and deduce a recommendation for how to proceed with the study.^{1,2} The first model-based designs were developed in the 1990s, most prominently the continual reassessment

method³, and also Bayesian decision procedures,^{4,5} which are closely related to one another.⁶

The alternative to model-based designs are methods that rely on largely heuristic rules or algorithms, such as the 3+3 design. The only real virtue of these algorithms is their simplicity (at least when nothing unforeseen happens), but on the downside they are less likely to identify the

¹Centre for Trials Research, Cardiff University, UK

²Department of Mathematics & Statistics, Lancaster University, UK

³MRC Biostatistics Unit, University of Cambridge, UK

⁴Cancer Research UK & UCL Cancer Trials Centre, University College London, UK

⁵Cancer Research UK Clinical Trials Unit, University of Birmingham, UK

⁶Edinburgh Cancer Centre, Western General Hospital, UK

⁷Statistical Methodology, Novartis Pharma AG, Basel, Switzerland

Corresponding author:

Philip Pallmann, Centre for Trials Research, College of Biomedical & Life Sciences, Cardiff University, Neuadd Meirionnydd, Heath Park, Cardiff, CF14 4YS, UK.

Email: pallmannp@cardiff.ac.uk

correct **MTD maximum tolerated dose** than model-based designs, require on average more patients to reach a dose recommendation, have less flexibility to accommodate deviations from the pre-specified dose-escalation procedure, and lack any theoretical foundation.^{7–10} Despite these clear drawbacks, rule-based designs are still in wide use, while the uptake of model-based designs remains slow, especially in the public sector.^{11–13} Unfortunately, the latter appear to many as a black box requiring specialist statistical input—which they are not¹⁴. There is also a misconception that they are overly complicated because they require more planning than rule-based designs, such as the choice of a prior probability distribution (or ‘prior’ for short).

In this paper we introduce MoDEsT (**Model-based Dose Escalation Trials**), a free and easy-to-use web tool for designing and conducting single-agent dose-escalation studies guided by a Bayesian decision procedure to estimate the **MTD maximum tolerated dose**.¹⁵ This method is conceptually straightforward and statistically sound: it uses logistic regression to model the relationship between dose and risk of toxicity, ~~which many are familiar with~~, and allows the investigator to specify prior distributions for dose-toxicity model parameters through the means of ‘pseudo-observations.’ These pseudo-observations should represent our best guesses, prior to the start of the study, at the toxicity outcomes that would be recorded if hypothetical (i.e. pseudo-)patients were administered certain doses of the compound. Usually, priors are specified by stipulating pseudo-observations for the lowest and highest doses available for administration during the future study. Prior specification can be informed by a scientific understanding of the drug’s anticipated mechanism of action. Alternatively, the pseudo-observations can be set so as to ensure the procedure has favourable operating characteristics (so-called ‘operational priors’).

Onerous tasks that are currently a barrier to the use of model-based designs, like setting a prior, become easy with MoDEsT: it allows trying out different priors and immediately visualises the consequences in terms of the operating characteristics (e.g. probability of identifying the correct **MTD maximum tolerated dose**, expected number of patients required) of the procedure. This provides users with insight as to how dose recommendations come about, thus demystifying the model-based design.

In contrast to most other software for study design, MoDEsT is specifically aimed at **both clinical trialists and statisticians with no previous experience of model-based dose escalation who would default to the 3+3 design for simplicity**, although we believe it is also useful for statistical experts who already have a thorough understanding of model-based dose escalation. The intuitive point-and-click interface of MoDEsT encourages users to ~~play around with~~ **and** explore a variety design options, allows them to watch changes become effective in real time, and get a feel for the design’s performance in different clinically relevant scenarios. It facilitates the consideration and inclusion of efficient model-based dose escalation at the earliest stage of clinical trial design, **which should always be a collaborative effort between clinical and statistical experts. as no specialised statistical programming is required.**

MoDEsT was written in the R¹⁶ programming language and using the extension package `shiny`,¹⁷ which provides a framework for building interactive web applications. `shiny` is steadily gaining popularity in the context of methods for early-phase dose finding. Recent years have seen the development of `shiny` apps to

- design and run dose-escalation studies using the continual reassessment method,¹⁸
- design dual-agent dose-escalation studies,¹⁹
- compare the performances of various model- and rule-based designs,^{20,21}
- simulate the highest achievable (i.e. optimal benchmark) accuracy when selecting the maximum tolerated dose.²²

There are also a number of graphical user interfaces (GUIs) for dose-finding methods that are not based on `shiny`, such as ‘NextGen-DF’ (now called ‘U-Design’)²³, ‘Web-EWOC’²⁴, a plethora of tools provided by the MD Anderson Cancer Center (<https://biostatistics.mdanderson.org/software/download>), and commercial packages such as EAST Escalate (<https://www.cytel.com/software/east>) and FACTS (<https://www.berryconsultants.com/software>). To the best of our knowledge, none of them holds any functionality for the method outlined above. The only software implementation of this method that we are aware of is ‘Bayesian ADEPT’,^{25,26} which has been defunct for several years. Further distinct strengths of MoDEsT are:

1. it runs under any operating system;
2. it does not require any software package to be installed;
3. its point-and-click interface means no specialist software or programming skills are required;
4. it automatically generates PDF reports;
5. it is free to use.

Methods

Bayesian decision procedure

The Bayesian decision procedure implemented in MoDEsT is made up of four main components: 1) a logistic regression model, 2) prior information about the dose-toxicity relationship, 3) a gain function, and 4) a set of rules for (de-) escalating the dose and stopping the study. We describe each component ~~in more detail~~ **briefly below**; for a detailed exposition we refer to the original paper by Zhou & Whitehead.¹⁵

1) *Logistic model.* We assume the relationship between dose and risk of toxicity follows a logistic model

$$\log \left(\frac{P(\text{dose-limiting toxicity})}{1 - P(\text{dose-limiting toxicity})} \right) = \beta_0 + \beta_1 \log(\text{dose})$$

where the logit transformation of the probability (i.e. the log odds) of observing a **DLF dose-limiting toxicity** (left-hand side of ~~the Equation 1~~) is assumed to depend on the log-transformed dose in a linear fashion (right-hand side of ~~the Equation 1~~); see the illustration in Figure 1. We use toxicity data from study patients to estimate the values of the model parameters β_0 (intercept) and β_1 (slope).

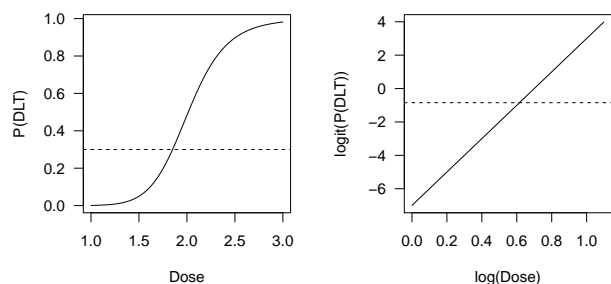


Figure 1. Example of an S-shaped dose-toxicity curve (left) and the corresponding straight line after transformation (right). The dotted horizontal line indicates a target toxicity level of 0.3, or $\text{logit}(0.3/0.7) = -0.847$ on the logit scale. $P(\text{DLT})$ is the probability of a dose-limiting toxicity.

2) *Prior information.* To get the Bayesian decision procedure started before any patient data are available, we need prior information on the dose-toxicity relationship. Guessing values of the model parameters β_0 and β_1 would be hard, so we prefer to formulate our prior beliefs about the toxicity rates of two distinct doses. The strength of these beliefs can be expressed in terms of their ‘effective sample sizes’²⁷: the information from, say, 3 pseudo-observations will be weighted as if they had been obtained from 3 real patients, and mathematically converted into so-called beta priors.¹⁵

3) *Gain function.* A gain function can be used to quantify, for each dose, the advantage of prescribing different dose levels to the next cohort of patients, where larger gains are to be preferred. Thus the gain function helps to determine which dose should be recommended for the next patient cohort. The ‘patient gain’ function would assign the dose currently thought to be closest to the target toxicity level (which is optimal from a current patient’s perspective) whereas the ‘variance gain’ function would choose the dose that will likely maximise learning about the dose-toxicity relationship (which is optimal from an investigator’s perspective and also from the perspective of future patients who will be treated beyond the current clinical trial).⁵ In practice the choice of gain function is unlikely to have a significant impact on the performance characteristics of the study design, but this can be explored in MoDEsT.

4) *Escalation and stopping rules.* Dose recommendations are determined primarily by the model and the gain function, but we may wish to apply additional restrictions such as:

- always start at the lowest dose;
- do not skip over any doses when escalating;
- do not escalate upon observing a toxicity in the current cohort.

Stopping recruitment to the study will be recommended once

- the maximum number of patients have been analysed;
- a pre-defined maximum number of consecutive patients receiving the same dose has been reached;
- a sufficiently accurate estimate of the **MTD maximum tolerated dose** has been obtained;
- no dose among those in the pre-specified set is deemed safe.

Implementation

We have created the web app MoDEsT written in the R¹⁶ programming language and using the extension package `shiny`,¹⁷ which provides a framework for building interactive web applications. `shiny` is steadily gaining popularity in the context of methods for early-phase dose finding. Recent years have seen the development of `shiny` apps to

- design and run dose-escalation studies using CRM,¹⁸
- design dual-agent dose-escalation studies,¹⁹
- compare the performances of various model- and rule-based designs,^{20,21}
- simulate the highest achievable (i.e. optimal benchmark) accuracy when selecting the MTD.²²

There are also a number of graphical user interfaces (GUIs) for dose-finding methods that are not based on `shiny`, such as ‘NextGen-DF’ (now called ‘U-Design’)²³, ‘Web-EWOC’²⁴, a plethora of tools provided by the MD Anderson Cancer Center (<https://biostatistics.mdanderson.org/softwaredownload>), and commercial packages such as EAST Escalate (<https://www.cytel.com/software/east>) and FACTS (<https://www.berryconsultants.com/software>). To the best of our knowledge, none of them holds any functionality for the method outlined above. The only software implementation of this method that we are aware of is ‘Bayesian ADEPT’,^{25,26} which has been defunct for several years. Further distinct strengths of MoDEsT are:

1. it runs under any operating system;
2. it does not require any software package to be installed;
3. its point-and-click interface means no specialist software or programming skills are required;
4. it automatically generates PDF reports;
5. it is free to use.

Results

The MoDEsT app

MoDEsT comes in two parts: a ‘Design’ module to investigate candidate design options and simulate their operating characteristics, and a ‘Conduct’ module to guide decision making throughout the study, incorporate accruing patient data into the model and provide summaries of the final dataset on completion of the study. Both modules are fully reactive i.e. changes made by the user become effective in real time. MoDEsT can be accessed online (<https://www.modest.lancaster.ac.uk> <https://medstats-lancs.shinyapps.io/design/> and <https://medstats-lancs.shinyapps.io/conduct/>) from any device with a web browser. For R users the app is also available in the add-on package `modest`.²⁸ We will keep maintaining both the web app and the R package (including bug fixes and possibly adding new options), hence the appearance and functionality of MoDEsT may change slightly as it evolves.

The ‘Design’ module. The ‘Design’ module takes as inputs the basic study parameters (maximum sample size, cohort size, dose levels, **FTL target toxicity level**, gain function), the pseudo-observations needed to specify prior distributions for parameters of the dose-toxicity model, ‘true’ values of model parameters for simulation of the Bayesian procedure in different scenarios, and additional escalation and stopping rules as detailed above; all these are conveniently specified via sliders, text and tick boxes (Figure 3). The app then creates graphical displays of the dose-toxicity curves, simulates an example of a study given the current specifications, and suggests a variety of scenarios* for use in a subsequent simulation study (Figure 4). For the scenario chosen MoDEsT assesses a variety of operating characteristics and presents the results in tables and graphics. On the basis of the inputs the app generates a CSV design file that can subsequently be fed into the ‘Conduct’ module. Additionally a report summarising the design, prior information, and simulation results can be downloaded in PDF format.

The ‘Conduct’ module. The ‘Conduct’ module requires the user to upload a design file (obtained from the ‘Design’ module) and supply (anonymised) patient data. The latter can either be uploaded as a CSV file (typically created with a text editor or spreadsheet software such as Microsoft Excel, OpenOffice/LibreOffice Calc, or Google Sheets), or entered manually via a spreadsheet interface. The app then produces graphical displays of the data, fits the logistic model, calculates the current estimate of the **MTD maximum tolerated dose**, and recommends either a dose for the next cohort or stopping the study in case a relevant criterion is fulfilled (Figure 5). A PDF report summarising the design, data, analysis, and recommendation is available for download. This can all be easily produced by the clinical study team for each dose review meeting so that dose recommendations based on statistical analyses of current and past patient data are used real-time alongside clinical opinion from the investigators to decide on dosing for the next cohort.

Getting help. While the app’s user guidance should be intuitive and most inputs and outputs self-explanatory, additional help may occasionally be required. The quickest way to learn more about an input element (e.g. slider, button, check box, text box) is by mousing over it, and a tooltip will appear. A full description/documentation of all of MoDEsT’s functionality along with a detailed explanation of all inputs and outputs is given in the help pages on the website. They are also included in the R package in the form of two HTML vignettes (<https://cran.r-project.org/web/packages/modest/vignettes/Design.html> and <https://cran.r-project.org/web/packages/modest/vignettes/Conduct.html>).

Example: a phase I study of quercetin

The workflow when designing and conducting a study with MoDEsT is best illustrated with a real data example. In the following we re-design and re-analyse (parts of) a dose-escalation study of a novel drug product in terminal cancer that originally used a 3+3-type design with several spontaneous modifications.

Dataset. Ferry *et al.* conducted a phase I study of the flavonoid quercetin in cancer patients suffering from a variety of forms of solid tumour no longer amenable to standard therapies.³⁰ They assessed 9 dose levels (60, 120, 200, 300, 420, 630, 945, 1400, 1700 mg/m²) with the aim of finding the **MTD maximum tolerated dose** under the premise that a 20% risk of renal toxicity (WHO grade ≥ 2) would be acceptable. A maximum of 18 patient cohorts of size 3 was to be recruited to the study.

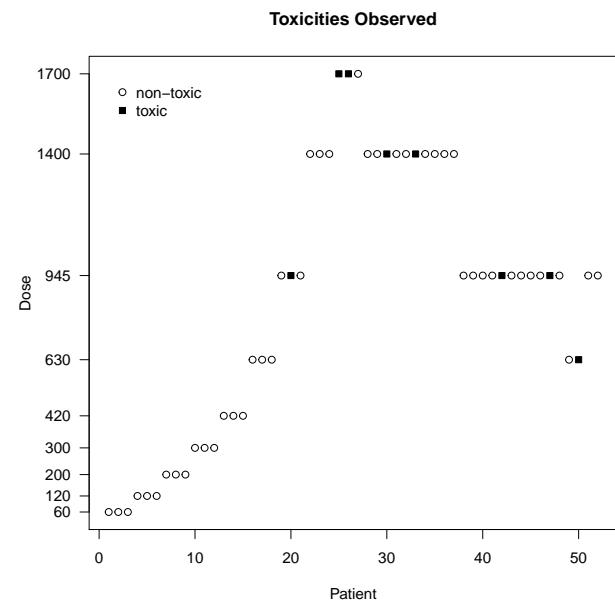


Figure 2. Overview of the quercetin study data.

Figure 2 provides an overview of the study data from a total of 52 patients. We note a number of oddities and decisions made against the rules of 3+3:

- The dose was escalated to 1400 mg/m² for the 8th cohort despite a **DLT dose-limiting toxicity** having been recorded for the 7th cohort at 945 mg/m².
- **DLTs Dose-limiting toxicities** occurred in both the 10th and 11th cohort at 1400 mg/m², and yet the dose was not de-escalated for the 12th cohort.
- No **DLTs dose-limiting toxicities** occurred in the 12th cohort, but the dose was de-escalated for the 13th cohort to 945 mg/m².
- The 12th cohort consisted of 4 patients.
- A **DLT dose-limiting toxicity** was recorded for the 17th cohort at 630 mg/m², still the dose was escalated to 945 mg/m² for the 18th cohort.
- The 16th, 17th, and 18th cohort each consisted of 2 patients only.

These deviations cannot be incorporated within the 3+3 design with its inherent inflexibility. This design no longer provides a relevant contribution to dose-escalation decisions, whereas the model-based procedure implemented in MoDEsT can easily handle cohorts of non-standard size and dose recommendations overruled by clinical judgement.

*The simulation scenarios are based on Table 1 of Zhou & Whitehead (2003)¹⁵ and Table 7.9 of Whitehead (2006).²⁹

Study design. We set the study design parameters (maximum sample size, cohort size, dose levels, **FTL target toxicity level**) as in Ferry *et al.*'s original study and use the patient gain function to drive dose-escalation recommendations (Figure 3). We specify a ‘pessimistic’ or ‘conservative’ prior distribution for parameters of the dose-toxicity relationship by specifying pseudo-observations consistent with the opinion that, a priori, we would expect to see 0.3 **DLTs dose-limiting toxicities** if 3 patients were treated with 60 mg/m² and 1.5 **DLTs dose-limiting toxicities** if 3 patients received 1700 mg/m², corresponding to anticipated **DLT dose-limiting toxicity risks** of 10% and 50%, respectively. Note that non-integer values of **DLTs dose-limiting toxicities** are acceptable when specifying prior distributions. For the assumed ‘true’ dose-toxicity model used to simulate **DLT dose-limiting toxicity** occurrences we choose the ‘true’ **DLT dose-limiting toxicity risks** on the lowest and highest doses to be 3% and 40%, respectively. One does not necessarily have to specify priors by considering **DLT dose-limiting toxicity risks** on the lowest and highest dose levels, but we do recommend choosing one dose at the lower and one at the upper end of the spectrum.

For the purpose of this example, we enforce starting at the lowest dose, not skipping over any doses when escalating, and not escalating when a **DLT dose-limiting toxicity** occurs in the current cohort. We consider trial designs which would recommend stopping the study for accuracy once 12 consecutive patients have received the same dose, or when the ratio of the upper and lower 95% credible limit around the estimated **MTD maximum tolerated dose** is 3 or less.

Figure 3. Input mask of the ‘Design’ module with specifications inspired by the quercetin study.

From these inputs MoDEsT generates a number of graphs and tables to summarise the operating characteristics of the stipulated design. We see that if our simulation model were indeed the true dose-toxicity curve, the **MTD maximum tolerated dose** would be estimated as 584 mg/m², but only 206 mg/m² under the much more cautious prior model (Figure 4, top left panel). Unsurprisingly, the 95% credible

band around the prior curve is extremely wide, as it is based on only 3 (pseudo-)observations.

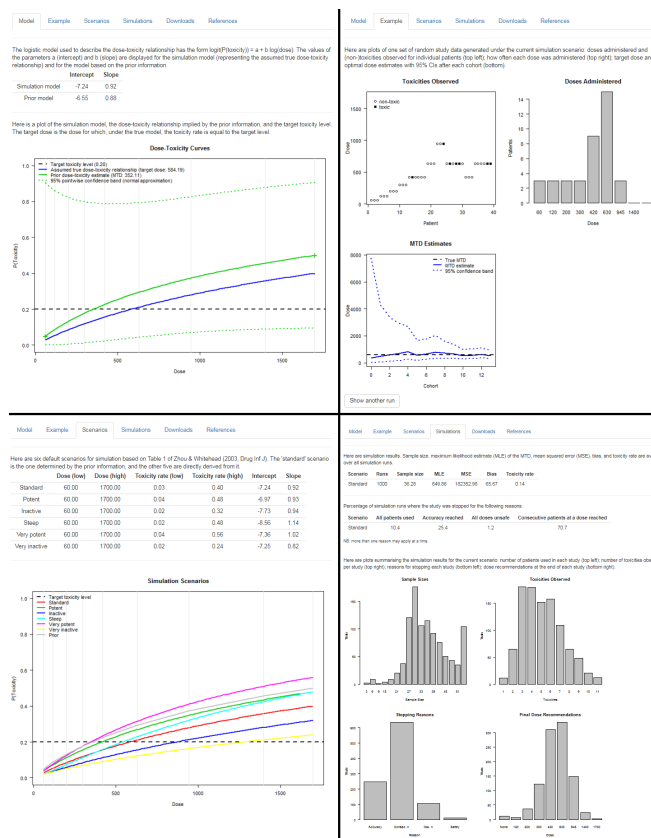


Figure 4. Top left: ‘Model’ tab of the ‘Design’ module displaying the prior and the assumed true dose-toxicity curve. Top right: ‘Example’ tab of the ‘Design’ module showing one simulated example dataset. Bottom left: ‘Scenarios’ tab of the ‘Design’ module giving an overview of six simulation scenarios. Bottom right: ‘Simulations’ tab of the ‘Design’ module summarising simulation results.

MoDEsT displays one simulated realisation of a study which proceeds according to the proposed Bayesian dose-escalation procedure and stopping rules, simulating patient outcomes setting **DLT dose-limiting toxicity risks** equal to values consistent with the current dose-toxicity simulation model. This is intended as an illustrative example of what the study *could* look like. A single simulated dataset, and the corresponding evolution of the dose-escalation trial, will not necessarily be representative of what would typically be observed if the true underlying dose-toxicity relationship was identical to the simulation model. However, by simulating a large number of trials and averaging across them, we can deduce what might happen on average. MoDEsT is designed to allow easy repeated simulation, which will be helpful in getting a feel for the variation of output that can occur.

The example trial shown in Figure 4 (top right panel) is stopped after 12 cohorts, when 12 consecutive patients have received the same dose (420 mg/m²), upon de-escalation from 630 mg/m². The 95% credible band for the **MTD maximum tolerated dose** becomes narrower over the course of the study and always contains the ‘true’ value of 584 mg/m².

To facilitate assessment of the design’s operating characteristics, MoDEsT automatically creates six simulation

scenarios that can be used in the simulation study: the standard scenario defined by the ‘true’ simulation model, and five additional scenarios derived from it that imply lower (‘inactive’) or higher (‘potent’) toxicity rates over the whole dose range or parts of it. They are summarised in a table and graph, alongside the prior for comparison (Figure 4, bottom left panel).

Figure 4 (bottom right panel) shows summary tables and graphs of 1000 simulations performed within seconds by the press of a button under the (anticipated) standard scenario. In this example the average sample size required was 36.28 patients (averaged over all 1000 simulated trials), the average of the maximum likelihood estimate (MLE) of the MTD maximum tolerated dose was 649.86 mg/m², with a large mean squared error (MSE) and notable bias, and an average toxicity rate of 14%, which is well below the targeted 20%. We see how many simulated trials were stopped for which reason(s); the sum of the percentages is greater than 100% because multiple stopping criteria can be fulfilled at the same time. We also get an overview of the sample sizes used, numbers of DLTs dose-limiting toxicities observed, reasons for stopping, and doses recommended as the MTD maximum tolerated dose across the 1000 simulated trials. These summary plots and tables produced instantly by MoDEsT will be sufficient for most users but a detailed account of all individual simulation runs can be downloaded as a CSV file.

Study conduct. For illustrative purposes we present and discuss here only the analyses following the 7th and the last cohort, respectively; in practice a similar analysis would be performed after every single patient cohort.

We upload the design file and a CSV containing the study data. MoDEsT generates tabular overviews of the design parameters and patient data, as well as plots such as the one in Figure 2. We recommend that users review this output to double-check the information and also whether data have been read in as intended i.e. correct columns were specified for the cohort, dose, and response variable.

In Ferry *et al.*'s study the dose was escalated from 945 to 1400 mg/m² for the 8th cohort, despite a DLT dose-limiting toxicity being observed in the 7th cohort (Figure 2). Re-analysing the data up to and including the 7th cohort with MoDEsT, we find that the Bayesian dose-escalation procedure recommends administering 945 mg/m² to the 8th patient cohort (Figure 5, left panel), which is in line with our the stipulated safety rule of not escalating when a DLT dose-limiting toxicity has been observed in the current cohort. Due to the amount of data accruing, the 95% credible band is much narrower than for the prior.

The dataset accrued upon completion of the study comprises 52 patients, which is less than the envisaged maximum sample size of 54, so we have to tick the box in MoDEsT to indicate that the study has been stopped (Figure 5, right panel). The final model-based estimate of the MTD maximum tolerated dose is 998 mg/m², but this is influenced by the very pessimistic prior. Removing the pseudo-observations used to formulate the prior yields a final (maximum likelihood) estimate of 1090 mg/m² and a marginally wider credible band. Both estimates lead to a recommendation of 945 mg/m² for the MTD maximum

tolerated dose, which is the same as in Ferry *et al.* However, had MoDEsT been used in the original study, the deviations from the protocol (such as smaller and larger cohorts) could easily have been accommodated, and the (clearly too high) dose of 1700 mg/m² would probably never[†] have been administered.

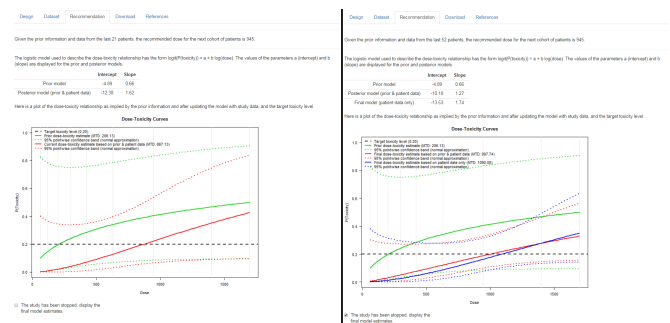


Figure 5. ‘Recommendation’ tab of the ‘Conduct’ module after the 7th cohort (left) and after the final analysis (right) of the quercetin study.

Discussion

Building trust in the utility, safety, and practical applicability of model-based dose-escalation designs is an essential step towards their wider acceptance within the clinical community. To assist this process we have created MoDEsT, a software tool that is straightforward to use even without any specialist knowledge of statistics or programming. We are positive it will convince trialists and statisticians that model-based methods are a feasible and worthwhile alternative to the 3+3 design and can be implemented with limited additional effort. Tasks that currently discourage many clinicians from using model-based designs, like having to set a prior, are made simple in MoDEsT. One of the main advantages of the software is that it allows investigators to input different (hypothetical or real) datasets ahead of time to see what dose recommendations the Bayesian dose-escalation procedure would generate, allowing them to develop some intuition as to how the procedure is working and how it would compare with their own intuition or algorithmic rules they might be more familiar with. We hope that by seeing the effects of changing design parameters in real time, trialists will become more confident in using model-based designs and that these will increasingly become the norm in early-phase dose-escalation studies.

Acknowledgements

The authors are grateful to Ian Wadsworth for providing R code, and to Michael Sweeting for advice during the early stages of the project, and two anonymous referees, the Associate Editor, and the Editor for a number of very helpful suggestions.

Declaration of conflicting interests

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

[†]Simulations with the final model of Figure 5 (right panel) as the ‘true’ simulation model show that it is extremely unlikely to ever visit this dose level.

Funding

This work was funded by MRC Network of Hubs for Trials Methodology Research project grant N78. CY was funded by Cancer Research UK grant C22436/A15958. LVH's contribution to the project was funded by Medical Research Council grant MR/J014079/1.

References

- Love SB, Brown S, Weir CJ et al. Embracing model-based designs for dose-finding trials. *Br J Cancer* 2017; 117(3): 332–339.
- Wheeler GM, Mander AP, Bedding A et al. How to design a dose-finding study using the continual reassessment method. *BMC Med Res Methodol* 2019; 18: 19.
- O'Quigley J, Pepe M and Fisher L. Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics* 1990; 46(1): 33–48.
- Whitehead J and Brunier H. Bayesian decision procedures for dose determining experiments. *Stat Med* 1995; 14(9–10): 885–893.
- Whitehead J and Williamson D. Bayesian decision procedures based on logistic regression models for dose-finding studies. *J Biopharm Stat* 1998; 8(3): 445–467.
- Shu J and O'Quigley J. Dose-escalation designs in oncology: ADEPT and the CRM. *Stat Med* 2008; 27(26): 5345–5353.
- Jaki T, Clive S and Weir CJ. Principles of dose finding studies in cancer: a comparison of trial designs. *Cancer Chemother Pharmacol* 2013; 71(5): 1107–1114.
- Paoletti X, Ezzalfani M and Le Tourneau C. Statistical controversies in clinical research: requiem for the 3 + 3 design for phase I trials. *Ann Oncol* 2015; 26(9): 1808–1812.
- Nie L, Rubin EH, Mehrotra N et al. Rendering the 3 + 3 design to rest: more efficient approaches to oncology dose-finding trials in the era of targeted therapy. *Clin Cancer Res* 2016; 22(11): 2623–2629.
- Adaptive Designs Working Group of the MRC Network of Hubs for Trials Methodology Research. A quick guide why not to use A+B designs, 2016. URL https://www.methodologyhubs.mrc.ac.uk/files/6814/6253/2385/A_quick_guide_why_not_to_use_AB_designs.pdf.
- Jaki T. Uptake of novel statistical methods for early-phase clinical studies in the UK public sector. *Clin Trials* 2013; 10(2): 344–346.
- van Brummelen EMJ, Huitema ADR, van Werkhoven E et al. The performance of model-based versus rule-based phase I clinical trials in oncology: a quantitative comparison of the performance of model-based versus rule-based phase I trials with molecularly targeted anticancer drugs over the last 2 years. *J Pharmacokinetic Pharmacodyn* 2016; 43(3): 235–242.
- Chiuzan C, Shtaynberger J, Manji GA et al. Dose-finding designs for trials of molecularly targeted agents and immunotherapies. *J Biopharm Stat* 2017; 27: 477–494.
- Yap C, Billingham LJ, Cheung YK et al. Dose transition pathways: the missing link between complex dose-finding designs and simple decision-making. *Clin Cancer Res* 2017; 23(24): 7440–7447.
- Zhou Y and Whitehead J. Practical implementation of Bayesian dose-escalation procedures. *Drug Inf J* 2003; 37(1): 45–59.
- R Core Team. R: a language and environment for statistical computing., 2017. URL <https://www.r-project.org>.
- Chang W, Cheng J, Allaire JJ et al. shiny: web application framework for R. R package version 1.0.4, 2017. URL <https://cran.r-project.org/package=shiny>.
- Wages NA and Petroni GR. A web tool for designing and conducting phase I trials using the continual reassessment method. *BMC Cancer* 2018; 18: 133. URL <https://uvatrapps.shinyapps.io/crmb>.
- Sweeting M. pipe.design: dual-agent dose escalation for phase I trials using the PIPE design. R package version 0.5, 2017. URL <https://cran.r-project.org/package=pipe.design>.
- Wheeler GM, Sweeting MJ and Mander AP. AplusB: a web application for investigating A + B designs for phase I cancer clinical trials. *PLOS One* 2016; 11(7): 1–17. URL <https://graham-wheeler.shinyapps.io/AplusB>.
- Chen SC and Shyr Y. A web application for optimal selection of adaptive designs in phase I oncology clinical trials. *Trials* 2017; 18(Suppl 1): P83. URL <https://cqs.mc.vanderbilt.edu/shiny/AdaptiveDesignS>.
- Wages NA and Varhegyi N. A web application for evaluating phase I methods using a non-parametric optimal benchmark. *Clin Trials* 2017; 14(5): 553–557. URL <https://uvatrapps.shinyapps.io/nonparbnch>.
- Yang S, Wang SJ and Ji Y. An integrated dose-finding tool for phase I trials in oncology. *Contemp Clin Trials* 2015; 45: 426–434. URL <https://udesign.laiyaconsulting.com>.
- Rogatko A, Tighiouart M and Bresee C. Web-EWOC: interactive web tool for designing and conducting dose finding trials in cancer: user guide, web-based portal v1.7, 2017. URL <https://biostatistics.csmc.edu/ewoc/ewocWeb.php>.
- Zhou Y and Whitehead J. Bayesian ADEPT: operating manual, 2002.
- Bedding A. Software review: Bayesian ADEPT assisted decision making in early phase trials. *Pharm Stat* 2003; 2(3): 219–221.
- Morita S, Thall PF and Müller P. Determining the effective sample size of a parametric prior. *Biometrics* 2008; 64(2): 595–602.
- Pallmann P and Wan F. modest: model-based dose-escalation trials. R package version 0.3-1, 2017. URL <https://cran.r-project.org/package=modest>.
- Whitehead J. Using Bayesian decision theory in dose-escalation studies. In Chevret S (ed.) *Statistical Methods for Dose-Finding Experiments*. New York, NY: Wiley, 2006.
- Ferry DR, Smith A, Malkhandi J et al. Phase I clinical trial of the flavonoid quercetin: pharmacokinetics and evidence for *in vivo* tyrosine kinase inhibition. *Clin Cancer Res* 1996; 2(4): 659–668.