J Pharmacokinet Pharmacodyn (2016) 43:235–242
DOI 10.1007/s10928-016-9466-0

REVIEW PAPER

# 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

**E. M. J. van Brummelen[1] · A. D. R. Huitema[2] · E. van Werkhoven[3] · J. H. Beijnen[2,4] · J. H. M. Schellens[1,4]**

Received: 27 August 2015 / Accepted: 17 February 2016 / Published online: 10 March 2016


**Abstract** Phase I studies with anticancer drugs are used to evaluate safety and tolerability and to choose a recommended phase II dose (RP2D). Traditionally, phase I trial designs are rule-based, but for several years there is a trend towards model-based designs. Simulations have shown that model-based designs perform better, faster and are safer to establish the RP2D than rule-based designs. However, the superiority of model-based designs has never been confirmed based on true trial performance in practice. To aid evidence-based decisions for designing phase I trials, we compared publications of model-based and rule-based phase I trials in oncology. We reviewed 172 trials that have been published in the last 2 years and assessed the following operating characteristics: efficiency (trial duration, population size, dose-levels), patient safety (dose-limiting toxicities (DLTs)) and treatment optimality (percentage of patients treated below and at or above the recommended phase 2 dose). Our results showed a non-significant but clinically relevant difference in trial duration. Model-based trials needed 10 months less than rule-based trials (26 versus 36 months; $p = 0.25$). Additionally, fewer patients were treated at dose-levels below the RP2D (31 % versus 40 %; $p = 0.73$) while safety was preserved (13 % DLTs versus 14 % DLTs). In this review, we provide evidence to encourage the use of model-based designs for future phase I studies, based on a median of 10 months of time gain, acceptable toxicity rates and minimization of suboptimal treatment.

**Keywords** Clinical trial design · Model-based trials · Rule-based trials · Dose-escalation trials · Trial duration · Trial performance

## Introduction

Phase I trials investigate the safety, tolerability, and preliminary efficacy of novel agents or combinations. In oncology, the primary goal of these trials is to determine the recommended dose in patients, known as the recommended phase II dose (RP2D), for use in a follow-up trial. It is commonly acknowledged that phase I trials should identify an accurate RP2D while minimizing sub-therapeutic treatment or toxic treatment. These operating characteristics depend on the trial design (i.e., escalation method), so careful consideration of the design is crucial.

Traditionally, dose-escalation has been conducted according to the 3 + 3 principle and its variants. In these rule-based designs, dose-levels are chosen according to a pre-specified rule or algorithm [1, 2]. Although the use of rule-based designs is still prevailing, model-based designs such as the continual reassessment method (CRM) gain popularity in clinical practice [1, 3–5]. In these designs, dose-levels are determined by estimating a model for the

**Electronic supplementary material** The online version of this article (doi:10.1007/s10928-016-9466-0) contains supplementary material, which is available to authorized users.

✉ J. H. M. Schellens
j.schellens@nki.nl

[1] Department of Clinical Pharmacology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

[2] Department of Pharmacy, The Netherlands Cancer Institute, Amsterdam, The Netherlands

[3] Department of Biometrics, The Netherlands Cancer Institute, Amsterdam, The Netherlands

[4] Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands

dose–toxicity relationship. Based on results from simulations, model-based designs are considered to have several advantages over classical rule-based designs, such as shorter trial duration [1], minimal suboptimal treatment [5] and a more accurate estimation of the RP2D [2, 5, 6]. Both the European Medicines Agency (EMA) and U.S. Food and Drug Administration (FDA) have recommended the use of model-based designs in order to improve phase I trial performance [7, 8]. However, the practical performance of rule-based and model-based trials has never been compared, hindering evidence-based decision making on trial design.

Our objective is to fill this gap in literature by providing a quantitative comparison of the performance of rule-based versus model-based oncological phase I trials based on a comprehensive systematic review of literature. We provide an overview of the theoretical and practical performance of rule-based and model-based phase I trials, which can be used for decision-making and future research. The main question to be addressed is whether model-based designs are indeed superior to classical rule-based designs.

## Theoretical

### Rule-based designs

Rule-based designs have been considered a safe and easy-to-implement approach to determine the RP2D [1]. The most commonly used rule-based designs are the 3 + 3 design and its variations including the 6 + 6 design, accelerated-titration and pharmacologically-guided-dose-escalation (PGDE) [1, 2]. The characteristic of rule-based designs is that dose escalation is guided by predefined rules based on the actual occurrence of dose-limiting-toxicities (DLTs) among patients at the last dose-level. In the 3 + 3 design for example, three patients are included in each cohort, with escalation to the next higher dose-level if no DLT occurs and expansion to six patients if one of the first three subjects develops a DLT. If upon expansion no additional DLTs are reported, the dose will be escalated and the same rule applies. Escalation stops if two or more out of six patients (or out of three) experience a DLT. The previous lower dose-level is then expanded to six patients. The dose at which at most one out of six patients experiences a DLT is considered the maximum tolerated dose or RP2D.

The 3 + 3 design has been generally accepted as a safe dose-escalation approach. The incidence of DLTs among all included patients (toxicity rate) in rule-based trials should be at most 33 %. At any dose-level, less than two out of six patients should experience a DLT. The same holds true for the final RP2D in rule-based trials [1].

Previously, it has been reported that rule-based trials need many dose-escalation steps to find the RP2D. This may result in excessive treatment at low (suboptimal) doses [1, 9], large population sizes [10] and long trial duration [1, 5]. In addition, the established RP2D may be too low, as shown by Zhou et al. who reported that the toxicity rate of the RP2D identified by rule-based trials may vary between 10 and 29 % [3, 11].

### Model-based designs

The first well known model-based design was introduced as the continual reassessment method (CRM) [12]. Variations on the CRM include the escalation with overdose control (EWOC) design and the time to event CRM (TITE-CRM) [1]. For a detailed description of these designs we refer to reviews by Jaki et al. [2] or Le Tourneau et al. [4].

In model-based designs, dose-escalation is guided by a model describing the dose-toxicity relationship. By repeatedly incorporating toxicity data from all explored doses including data from previous trials if they are available, an estimate of the toxicity rate at each dose-level is provided. In model-based designs, the RP2D is defined as the dose that induces toxicity at the pre-defined target toxicity rate (mostly set to 10-33 %) with an acceptable confidence interval according to the model [3].

Based on simulations, model-based designs are considered to establish the RP2D faster and more accurate while less patients are needed [1, 2, 5, 6, 9]. Additionally, it has been suggested that model-based designs are better when no expected RP2D can be pre-specified [1]. This is the case in many first-in-man trials and in drug-combination trials. Also, there are model-based designs that allow incorporation of pharmacodynamic endpoints next to toxicity endpoints which is considered beneficial for molecularly targeted agents [1, 5]. Yet, the implementation of model-based designs in practice seems to be difficult. Their use may be hindered by insufficient statistical expertise and lack of familiarity compared to rule-based designs.

In Table 1, an overview of the theoretical advantages and drawbacks of the different designs for phase I trials is provided. In this systematic review, we aim to provide the true advantages and drawbacks of model-based and rule-based trials based on their actual performance.

## Methods

### Search strategy

A search in PubMed was performed on June 28th 2014 to include all publications of phase I studies over the last

**Table 1** Comparison of the theoretical advantages and drawbacks of model-based and rule-based designs [1]

| | Model-based | Rule-based |
|---|---|---|
| Toxicity rate at RP2D | Target rate to be specified, generally between 10 and 33 % | Less than 33 % |
| Precision of RP2D | Provides a confidence interval around selected RP2D | Uncertain |
| Population size | Likely to be smaller than rule-based | Likely to be larger than model-based |
| Trial duration | Likely to be short | Likely to be long |
| Suboptimal dosing | Likely to be minimal | Likely to be high |
| Use of available information | All clinical information incorporated in model for dose-escalation and determining the RP2D | Only information of previous dose-level used during dose-escalation |
| Implementation | Statistical expertise needed | Easy to implement |
| Application | If no prior information on dose is available | If dose-levels can rationally be pre-specified |

*RP2D* recommended phase II dose

24 months addressing small-molecule targeted therapies and dose-escalation.

The following search terms were used: ((((((maximum tolerated dose[mesh] OR maximum tolerated dos*[tiab] OR dose escalation*[tiab] OR doses escalation*[tiab] OR drug administration schedule[mesh] OR drug dose–response relationship[mesh]))))) **AND** ((((molecular targeted therapy[mesh] OR targeted therap*[tiab] OR molecularly targeted therap*[tiab] OR inhibitor [tiab] OR small molecule*[tiab] OR tyrosine kinase*[tiab] OR kinase*[tiab] OR protein-tyrosine kinases[mesh])))))) **NOT** ((pediatric study[tiab] OR pediatric studies[tiab] OR pediatric[tiab] OR hormone therap*[tiab] OR hormonal therap*[tiab] OR radiotherap*[tiab] OR radio-therap*[tiab] OR cytotox*[tiab] OR children[tiab] OR virus[tiab] OR viral[tiab])) **AND** (phase I[tiab] OR phase 1[tiab] OR phase one[tiab] OR phase 1a[tiab] OR phase 1b[tiab] OR phase Ia[tiab] OR phase Ib[tiab]) **NOT** (expansion OR expansion phase)

*Limits: English, From: 2012/06/01 to 2014/06/01*

Search results were screened to include studies in which at least one small molecule targeted agent was escalated, either or not combined with fixed conventional chemotherapy/cytotoxic therapy. The following articles were excluded: paediatric studies, studies without dose-escalation/non-phase I studies, immunoglobulin therapies, gene therapy, vaccine/viral therapy, (combinations with) radiotherapy, non-oncologic applications and studies in which primary data were incomplete or inaccessible and early termination for reasons other than results on efficacy or tolerability. For each excluded trial, the principal reason for exclusion was recorded. Included articles were grouped by rule-based designs (key words: $3 + 3$ or variants, mFibonacci escalation, accelerated titration, PGDE) or model- based designs (key words: Bayesian model, CRM, EWOC, toxicity probability method, nonparametric up and down design).

## Endpoints

Study characteristics were recorded including the PubMed identification number, the number of schedules that were tested, the number of escalations, reports on trial delay or amendments, the use of intermediate dose-levels, the number of active agents used in the trial, route of administration and first-in-man administration. No formal review protocol was used.

Data on the endpoints as described in Table 2 were extracted in duplicate by the first author. The median and range were reported and compared between designs. Data on trial duration were obtained from ClinicalTrials.gov or the published article. If data on duration could not be retrieved authors were approached to complete data. From a random selection the authors were asked to report the trial duration in addition to published data to check for consistency. The overall toxicity rate was calculated for each trial based on the incidence of DLTs.

## Statistical analysis

Data on all endpoints were compared between model-based and rule-based trials using a Wilcoxon rank-sum test with continuity correction. The correlation between trial duration and the number of included patients was assessed using Spearman's rank correlation (rho). Statistical tests were performed in R [13]. Subgroup analyses were performed for first-in-man studies, combination studies and studies with oral administration only since it was expected that these factors may influence trial performance. Study characteristics were analyzed using Fisher's exact test for

**Table 2** Description of endpoints

| Endpoint | Description | Indicator for |
|---|---|---|
| Number of patients needed to establish RP2D | Number of patients receiving study treatment until the preliminary RP2D was identified and the cohort was expanded (i.e., for a 3 + 3 design when the cohort was expanded to more than 6 patients as formally needed to determine the RP2D) | Efficiency |
| Number of patients included | Number of patients enrolled/included | Efficiency |
| Number of escalations | As reported in publication | Efficiency |
| Ratio RP2D/starting dose | In case more than one schedule was tested, the first RP2D was divided by the initial starting dose. Starting dose refers to the starting point of dose-escalation (dose-level 1) | Efficiency |
| Trial duration | The time in months from start of the trial to data-closure as stated on ClinicalTrials.gov, the published article or as reported by the author | Efficiency |
| Patients treated below and at or above RP2D | Number of patients categorized per group that were treated below, and at or above RP2D as a percentage of the included population. The sum of patients treated below and at or above the RP2D may be lower than 100 % in case the included population was larger than the number of treated patients | Patient safety<br>(Sub)optimal treatment |
| Number of DLTs | As reported in publication | Patient safety |
| | Toxicity rate in percentage for each trial was calculated by DLTs/N included × 100 % | |

*RP2D* recommended phase II dose, *DLTs* dose limiting toxicities, *N* included population

categorical variables and Wilcoxon rank-sum test for continuous variables.

## Results

Of the 343 search results, 171 publications were excluded for reasons as specified in Fig. 1 which left 172 studies for inclusion. Study characteristics and results on the pre-defined endpoints are presented in Table 3. The complete overview of outcomes per included trial is available as Online Supplementary Material. Results on trial duration were obtained from 122 out of 172 trials, among these 23 were reported to us by the author. In total 68 authors were approached. Only the subgroup of first-in-man studies performed differently compared to the total dataset which will be discussed later.

### Performance of rule-based trials

Among 172 trials that were included in this review, 161 (94 %) used a rule-based design. All rule-based trials applied the 3 + 3 design or its variations and in 12 trials this was proceeded by an accelerated titration phase.

The median time to finish was 36 months for rule based trials with a median inclusion of 30 patients. Among the

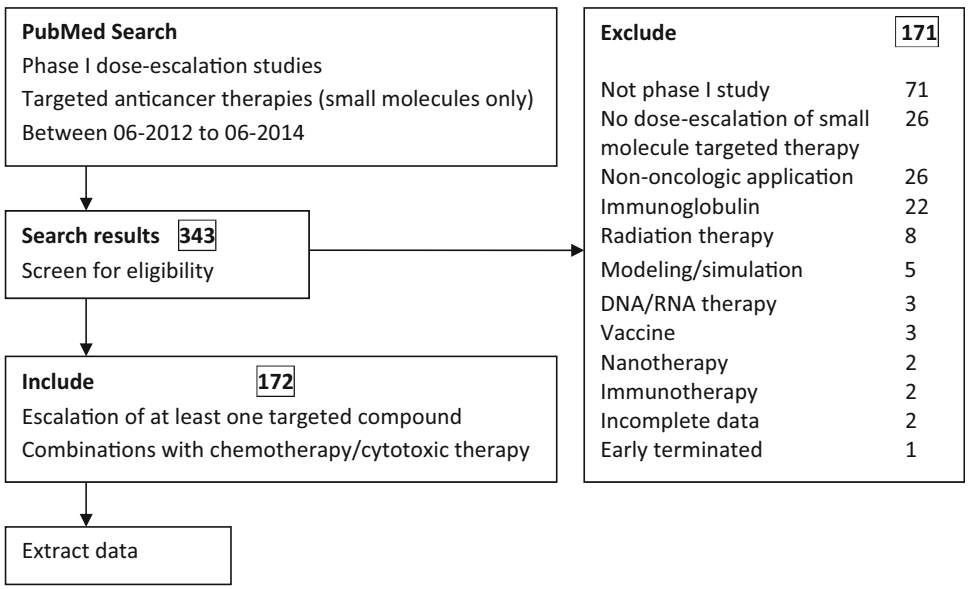


**Fig. 1** Study selection overview

**Table 3** Characteristics and performance of rule-based trials versus model-based trials

| Trials included | Rule-based 161 (100 %) | Model-based 11 (100 %) | *p* value |
|---|---|---|---|
| First in man (FIM) studies | 55 (34 %) | 3 (27 %) | 0.75* |
| Combination therapies (including 4 FIM studies) | 66 (41 %) | 3 (27 %) | 0.53* |
| Schedules tested | | | 0.87^ |
| 1 | 107 (66 %) | 7 (64 %) | |
| 2 | 37 (23 %) | 3 (27 %) | |
| ≥3 | 17 (11 %) | 1 (9.0 %) | |
| Administration route oral | 117 (73 %) | 9 (82 %) | 0.73* |
| Number of patients needed to determine RP2D[b] | 26 [8–147] | 34 [15–135] | 0.07^ |
| Number of patients[b] | 30 [8–206] | 56 [15–135] | 0.09^ |
| Patients treated below RP2D (% of included)[b] | 40 [0–100] | 31 [0–68] | 0.73^ |
| Patients treated at or above RP2D (% of included)[b] | 53 [0–100] | 60 [21–100] | 0.76^ |
| Number of DLTs in the trial[b] | 3 [0–18] | 5 [1–28] | 0.14^ |
| Number of escalations[b] | 4 [0–20] | 6 [1–12] | 0.55^ |
| Ratio RP2D/starting dose[b] | 3.0 [0–180] | 2.0 [1–40] | 0.96^ |
| Trial duration[a] in months[b] | 36 [8–90] | 26 [16–48] | 0.25^ |

*p* value significance level = 0.05

[a] Available for 113 (70 %) rule-based and 9 (82 %) model-based studies

[b] Values presented as: median [range]

* Obtained by Fisher's Exact Test; ^ Obtained by Wilcoxon rank-sum test

included population, 40 % was treated at doses below the RP2D which is potentially suboptimal and 53 % at the RP2D or above the RP2D which is potentially toxic. The starting dose was increased a median of 3.0 times in a median of 4 dose-escalations. For the dose-escalation part of the trials, a median of 26 patients was needed. In rule-based trials, the median number of DLTs was 3, which resulted in an incidence of DLTs (toxicity rate) of on average 14 %. This confirms that the toxicity rate in rule-based trials is indeed much lower than 33 %, as has been suggested by Zhou [11]. These numbers provide an indication of the performance of classical phase I trials but can only be interpreted when they are placed into perspective. Therefore, we provide a comparison to the performance of model-based trials.

## Performance of model-based trials

In line with previous reviews, only 11 out of 172 trials that have been published in the last 2 years (6 %) used a model-based design [3]. Of these, 7 used a Bayesian Logistic Regression Model with Overdose Control (BLRM-EWOC), the others used BLRM, TITE-CRM, toxicity probability method or non-parametric up and down design with bivariate isotonic regression. The median time to finish was 26 months for model-based trials, with a median inclusion of 56 patients. Among these patients, 31 % was treated below the RP2D and 60 % at or above the RP2D. This confirms that model-based trials tend to treat more patients at or above the RP2D, as simulations suggested. Hereby, suboptimal treatment can be reduced. In model-based trials, the starting dose was increased median 2.0 times in a median of 6 dose-escalations. For the dose-escalation part of the trial, a median of 34 patients was needed. There was a large difference between the included population, consisting of 56 patients, and the population that was used for dose-escalation, consisting of 34 patients. This implies inclusion of more patients at or around the preliminary established RP2D. This may explain why in model-based trials, more patients were treated at or above the RP2D compared to rule-based trials. In model-based trials, the median number of DLTs was 5 per trial, which is slightly more than in rule-based trials. However, the total incidence of DLTs (toxicity rate) of 13 % is comparable to the toxicity rate of 14 % in rule-based trials.

## Clinically relevant differences in trial duration and population size

Although we were not able to detect any significant differences in the operating characteristics for trial performance (Table 3), we observed pronounced differences in trial duration (36 vs. 26 months; $p = 0.25$) and population size (30 vs. 56 patients; $p = 0.09$), which are considered clinically relevant. Paradoxically, whereas the data on trial duration favor model-based trials, the data on population size seem to favor rule-based trials. In the next paragraphs, we will discuss possible explanations for these findings.

### Trial delay

The median time to finish data collection was 26 months (range: 16–48 months) for model based trials, whereas rule-based trials needed 36 months (range: 8–90 months). Despite the wide ranges in trial duration in both groups, a clear trend towards shorter duration of model-based trials is visible with a median difference of 10 months (Fig. 2).

Several factors could contribute to prolonged trial duration. These include the investigation of more dose-levels, a high ratio between the starting dose and the RP2D, inclusion of more patients, or a high incidence of DLTs during the trial. However, rule-based trials did not investigate more dose-levels, nor did they perform more escalations, report more DLTs or include more patients than model-based trials. Paradoxically, the number of included patients was even lower in rule-based trials. These parameters could not explain why rule-based trials needed more time to finish.

Therefore, we explored other factors that could contribute to the difference in trial duration. We found that in 43 % of the rule-based trials, trial delay was reported whereas this was the case in none of the model-based trials. Reasons for trial delay included amendments to the protocol (18 %) or introduction of intermediate dose-levels (39 %). Rule-based trials for which such a delay was reported lasted 41 months, whereas those for which no delay was reported lasted only 31 months ($p = 0.02$).

### More patients in first-in-man model-based trials

Although we expected that shorter trial duration of model-based trials would coincide with a smaller population size, this was not the case. Trial duration was weakly associated with sample size (Spearman's rho = 0.26, $p = 0.004$) and

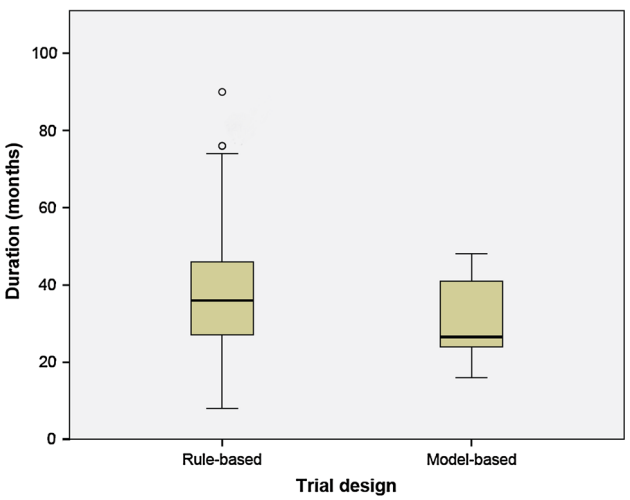


**Fig. 2** Trial duration in months for rule-based ad model-based trials defined as time from start to data-closure

while model-based trials took 10 months less to finish, they accrued 46 % more patients (30 for rule-based trials vs. 56 for model-based trials; $p = 0.09$).

It may be noticed that the presented data on population size in model-based trials show a discrepancy. Model-based trials included more patients than rule-based trials while only slightly more patients were needed to determine the RP2D. This could be explained by the inclusion of more patients at or around the preliminary established RP2D in model-based trials. Another possible explanation could come from the first-in-man studies, which may contribute to the larger population sizes in model-based trials. Within the first-in-man studies, model-based trials ($n = 3$) did not only include more patients (median 101 vs. 44; $p = 0.03$) but also needed significantly more patients to determine the preliminary RP2D (median 75 vs. 34; $p = 0.02$) compared to rule-based ($n = 55$) trials. Additionally, the first-in-man model-based trials had a remarkably high ratio between the RP2D and the starting dose (median 35.0 for model-based vs. 8.0 for rule-based; $p = 0.11$). In this subgroup, the median difference between trial duration of model-based trials and rule-based trials decreased to 4 months (31 vs. 35 months; $p = 0.9$) instead of 10 months. Although data are limited, this raises the question whether model-based trials are less suitable for first-in-man studies, which we will discuss below.

### Increased accuracy of RP2D

In the previous paragraph we discussed the finding that especially in first-in-man trials, model-based trials included more patients and that these trials lasted only marginally shorter. This can be interpreted in two ways. Firstly, it may indicate that the use of a model-based design leads to inferior performance of first-in-man trials, but this would be difficult to explain. Secondly, it may be an indication that where a rule-based design would stop too early, a model-based design could continue and establish a more accurate RP2D. Such a scenario would be likely in particular for first-in-man studies, where no clinical information is available in advance to predict the RP2D. As a result, model-based and rule-based designs may find very different RP2Ds in first-in-man trials. This was confirmed in practice in the first-in-man model-based study by Sessa et al. [14]. They reported that the RP2D they found was 2.5 times higher than the RP2D as it would have been defined by a rule-based approach. More data should be obtained to investigate if model-based trials indeed need more patients and if this results in a more accurate estimate of the RP2D.

### Quicker estimate of the RP2D in model-based trials

In line with the shorter trial duration, it seems that model-based trials are able to provide a first estimate of the RP2D

more swiftly. To establish the RP2D, model-based trials investigated more dose-levels (6 vs. 4; $p = 0.55$) and needed more patients (34 vs. 26; $p = 0.07$). Yet, the total trial duration was shorter.

Additionally, model-based trials approach the RP2D rather quick given the low percentage of patients that were treated at levels below the RP2D (31 % vs. 40 %; $p = 0.73$). Quicker dose-escalation does not seem to have compromised patient safety, since the toxicity rate of model-based trials (13 %) and rule-based trials (14 %) was comparable. It can thus be stated that the swift dose-escalation in model-based trials does not affect patient safety.

## Discussion and conclusion

Among the 172 included phase I studies that were published between 2012 and 2014 only 11 used a model-based design. Although we expected that more eligible model-based trials could be included over the last 2 years, the included number fits with previous reviews [3]. The relatively low number of model-based trials reflects current practice; a slight trend towards the use of model-based designs is noticed, but the use of rule-based designs is prevailing persuasively. Despite the low number, the data we report on model-based trials are comparable to the data that were reported by Iasonos et al. in a review based on 53 model-based designs and are therefore considered representative [3].

With the presented data we tried to answer the question whether or not there is evidence to prefer a model-based design over a rule-based design. Based on results from simulations, model-based designs have been considered superior to rule-based designs. We provided data on trial performance in practice to allow comparison of performance of model-based and rule-based trials. We found no statistical superiority of either rule-based trials or model-based trials. However, our data suggest that with model-based designs, the RP2D can be established more swiftly compared to rule-based trials. Additionally, we showed that patients are more likely to receive optimal and potentially effective doses in a model-based phase I trial without additional severe toxicity. This has been assumed before but was never truly compared to rule-based trials [1, 5]. A disadvantage of model-based trials is that more patients are needed overall, but this may be counterbalanced by a more accurate estimate of the RP2D.

The evidence we provide is limited by several factors. Firstly, we retrieved data on trial duration from three sources but these data were highly inconsistent. We searched ClinicalTrials.gov, extracted data on duration from the published articles and additionally we asked authors to report the duration of their trial. If available, data from ClinicalTrials.gov or the published article were used. Otherwise the authors' report was used. Ideally, trial duration should have been defined as the time from start of accrual to determination of the RP2D, but since these data were not available we have defined it as time from start to data closure.

Secondly, only few publications on model-based trials could be included. Although our search was broadened to increase the number of model-based trials, the proportion of model-based trials remained low. This should be considered when interpreting the results. Selective publication could have biased our results. However, there is no reason nor evidence to assume that either model-based or rule-based trials are more prone to selective publication.

Thirdly, the performance of individual trials can be affected by several factors that are not included in this analysis, such as the investigational product, speed of recruitment, the number of participating centers, financial and logistical issues. Heterogeneity of the included trials possibly contributes to the wide variation in the data and to non-significant differences between designs. Additionally, a crucial aspect of phase I trial performance is the accuracy of the RP2D. However, this is difficult to address because for many trials it is unknown what RP2D would have been found if another design was used. Despite these limitations, we encountered strong indications that establishment of the RP2D with model-based designs is quick and safe.

For future phase I trials we encourage the use of model-based designs in order to shorten clinical development of anticancer agents and to potentially increase patient benefit. Currently, dose-escalation trials use toxicity data (DLTs) as the primary endpoint. Since the introduction of targeted anticancer agents and immunotherapy, the use of toxicity data as the only endpoint has become doubtful. There is an increasing need for additional endpoints, such as pharmacodynamics, to support the optimal dose. Model-based designs allow using pharmacodynamic data next to toxicity data [1, 5, 10], whereas in rule-based designs the use of different outcome parameters is problematic. Current research on biomarker development and validation will hopefully facilitate incorporation of pharmacodynamic endpoints in dose-escalation studies.

Although previously FDA, EMA, and several reviewers already recommended the use of model-based designs [1, 2, 5, 6], the use is still uncommon. The implementation of model-based designs into daily practice may be hindered by the lack of familiarity with these designs and insufficient statistical expertise. We hope these obstacles may be overcome to improve the performance of dose-escalation trials. Pharmaceutical companies, patients, and society may benefit from the use of model-based trials given their potential to shorten clinical development of novel therapies.

**Compliance with Ethical Standards**

**Conflicts of interest** The authors declare no conflict of interest.

## References


1. Le Tourneau C, Lee JJ, Siu LL (2009) Dose escalation methods in phase I cancer clinical trials. J Natl Cancer Inst 101(10):708–720
2. Jaki T, Clive S, Weir CJ (2013) Principles of dose finding studies in cancer: a comparison of trial designs. Cancer Chemother Pharmacol 71(5):1107–1114
3. A Iasonos, J O'Quigley (2014) Adaptive dose-finding studies: a review of model-guided phase I clinical trials. J Clin Oncol 32(23):2505–2511
4. Le Tourneau C, Gan HK, Razak ARA, Paoletti X (2012) Efficiency of new dose escalation designs in dose-finding phase I trials of molecularly targeted agents. PLoS ONE 7(12):e51039
5. Harrington JA, Wheeler GM, Sweeting MJ, Mander AP, Jodrell DI (2013) Adaptive designs for dual-agent phase I dose-escalation studies. Nat Rev Clin Oncol 10(5):277–288
6. Mandrekar SJ (2014) Dose-finding trial designs for combination therapies in oncology. J Clin Oncol 32(2):65–67
7. European Medicines Agency Committe for Medicinal Products for Human Use, "Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design," 2007. [Online]. Available: http://www.ema.europa.eu/docs/
8. U.S. Food and Drug Administration (2010) Guidance for industry: adaptive design clinical trials for drugs and biologics 2010. [Online]. Available: http://www.fda.gov/downloads/Drugs/Guidances/ucm201790.pdf
9. Rogatko A, Schoeneck D, Jonas W, Tighiouart M, Khuri FR, Porter A (2007) Translation of innovative designs into phase I trials. J Clin Oncol 25(31):4982–4986
10. Eisenhauer EA, O'Dwyer PJ, Christian M, Humphrey JS (2000) Phase I clinical trial design in cancer drug development. J Clin Oncol 18(3):684–692
11. Zhou Y (2004) Choice of designs and doses for early phase trials. Fundam Clin Pharmacol 18(3):373–378
12. O'Quigley J, Pepe M, Fisher L (1990) Continual reassessment method: a practical design for phase 1 clinical trials in cancer. Biometrics 46(1):33–48
13. R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria., 2014. [Online]. Available: http://www.r-project.org/
14. Sessa C, Shapiro GI, Bhalla KN, Britten C, Jacks KS, Mita M, Papadimitrakopoulou V, Pluard T, Samuel TA, Akimov M, Quadt C, Fernandez-Ibarra C, Lu H, Bailey S, Chica S, Banerji U (2013) First-in-human phase I dose-escalation study of the HSP90 inhibitor AUY922 in patients with advanced solid tumors. Clin Cancer Res 19(13):3671–3680