

Research paper

2D (2 Dimensional) TEQR design for Determining the optimal Dose for safety and efficacy

Revathi Ananthkrishnan^{a,*}, Stephanie Green^b, Daniel Li^c, Michael LaValley^d

^a Celgene, 300 Connell Drive, Berkeley Heights, NJ, 07922, USA

^b New London, CT, 06320, USA

^c Juno Therapeutics, A Celgene Company, Seattle, WA, 98109, USA

^d Boston University, School of Public Health, Boston, MA, 02118, USA



ARTICLE INFO

Keywords:

Early phase oncology design
2D TEQR design
OBD isotonic and eff-tox designs
Optimal dose for safety and efficacy

ABSTRACT

Designs, such as the Eff-Tox, OBD (optimal biological dose), STEIN (simple efficacy toxicity interval), and TEPI (toxicity efficacy probability interval) designs, have been proposed to determine the optimal dose of a new oncology drug using both efficacy and toxicity. The goal of these designs is to select the optimal drug dose for further phase trials more accurately than dose finding designs that only consider toxicity, such as the 3 + 3, TEQR (toxicity equivalence range), mTPI (modified toxicity probability interval), and EWOC (escalation with overdose control) designs. We propose a new frequentist design for optimal dose selection, the 2D TEQR design, that is easier to understand and simpler to implement than the TEPI, Eff-Tox, STEIN and OBD designs, as it is based on the empirical or observed toxicity and efficacy rates and does not require specialized computations. We compare the performance of this new design with those of the TEPI, STEIN, Eff-Tox and OBD Isotonic designs. Although for the same sample size and cohort size, the frequentist 2D TEQR design is less accurate than the Bayesian TEPI design and also the STEIN design in selecting the optimal dose, the accuracy of optimal dose selection of the 2D TEQR design can be increased, in many cases, with a moderate increase in cohort size. The 2D TEQR design is as accurate as or more accurate than the Eff-Tox design in optimal dose selection, and better than the OBD Isotonic design, unless there is a clear peak in the true response rates, in which case the OBD Isotonic design performs better than the other designs.

1. Introduction

Several oncology dose finding designs, such as the CRM, mTPI, TEQR, EWOC, BOIN, Keyboard and i3+3 designs [1–9,32], have been proposed to select the maximum tolerated dose (MTD) of the study drug. These designs select the MTD more accurately than the 3 + 3 design, but use only the toxicity of the drug (dose limiting toxicities (DLTs)) in dose selection.

Designs incorporating both safety and efficacy, such as the Eff-Tox, STEIN, OBD and the TEPI designs [10–14] among others [8,15–25,33], aim to find a better dose than one determined by safety alone. The MTD selected in a Phase 1 oncology trial may not be the best dose for safety and efficacy, and examples demonstrate how critical it is to select the best dose to avoid poor outcomes in later phase trials and clinical use [34]. The drug pegylated liposomal doxorubicin, for treating platinum resistant ovarian cancer, was approved at 50 mg/m² administered every

28 days, but is commonly used at 40 mg/m² due to greater toxicity at the higher dose but equal efficacy at the lower dose [26]. Other examples where the approved dose and the standard clinical dose differ are provided by Schilsky [27]. Furthermore, adverse effects are assumed to increase with dose, but efficacy need not increase with dose, as seen in some immuno-oncology drugs [28,29]. Hence, it is critical to consider both safety and efficacy concurrently in a single reasonably sized trial to determine the best dose.

The mTPI is a Bayesian dose finding design that uses the unit probability mass (UPM)s, calculated for underdosing, overdosing and the target intervals from observed DLTs at the current dose level, to make the next dosing decision and determine the MTD. The TEQR design, its frequentist counterpart, uses the DLT rate observed at each dose level for dose finding decisions and determining the MTD. The TEPI is a 2D (2 dimensional) version of the mTPI design, which uses both toxicity and efficacy in determining the optimal drug dose. Here, we propose a non-

* Corresponding author.

E-mail addresses: revathia@gmail.com, rananthkrishnan@celgene.com (R. Ananthkrishnan).

<https://doi.org/10.1016/j.conctc.2019.100461>

Received 27 April 2019; Received in revised form 25 September 2019; Accepted 9 October 2019

Available online 12 October 2019

2451-8654/© 2019 The Author(s).

Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Bayesian rule-based (frequentist) counterpart of the TEPI design, the 2D TEQR design, which considers toxicity and efficacy to determine the optimal dose (see Fig. 1 for the relationship among the mTPI, TEQR, TEPI and 2D TEQR designs). This design is easier to understand and simpler to implement than the Bayesian TEPI design and requires no complex computation. The Eff-Tox and OBD Isotonic designs are Bayesian designs that consider the efficacy and toxicity outcomes in their dosing algorithm. The Eff-Tox design uses efficacy-toxicity tradeoff contours constructed using three efficacy-toxicity probability pairs considered to be equally desirable as determined by clinical judgement. The contours are used in the dosing decisions (escalate/stay at the same dose/de-escalate) as well as in the final dose selection. The OBD Isotonic design chooses as the optimal dose, the lowest dose with the highest efficacy probability among all the safe doses. The STEIN design, while basing its dosing decisions on the observed DLT and response rates, also incorporates a posterior probability calculation in its dosing algorithm. Isotonic regression is then applied to the observed toxicity rates and model averaging of unimodal isotonic regression(s) is performed on the observed efficacy rates. The final or optimal dose is selected using various utility functions in the STEIN, TEPI and 2D TEQR designs. We compare the operating characteristics of the proposed 2D TEQR design to those of the TEPI, STEIN, Eff-Tox and OBD Isotonic designs and investigate the effect of changing sample and cohort sizes on the accuracy of optimal dose selection of the proposed 2D TEQR design.

2. Design/methods

2.1. Dosing decisions/dosing algorithm (based on Tables 1 and 2)

Our design incorporating toxicity and efficacy in dose selection is a frequentist version of the TEPI design proposed by Li et al. [14]. In the TEQR design, only safety (DLTs) is considered, and the dosing decisions made during the trial are based on the observed DLT rates [4]. In our 2D version of the TEQR design, called the 2D TEQR design, efficacy (response) rates are also used, and the dosing decision table for our proposed 2D TEQR design is given in Table 1. For example, based on Table 1, if the empirical or observed toxicity (DLT) rate is low and the empirical or observed efficacy (response) rate is moderate at the current

Table 1

Dosing Decision Table for TEQR 2D Design based on the cumulative observed toxicity and efficacy rates at the current dose level subject to acceptable levels at the previous and next dose level(s) (i.e. Table 2 also needs to be considered when Table 1 says Escalate or De-escalate).

	Empirical Efficacy rate low (0, 0.2]	Empirical Efficacy rate moderate (0.2, 0.4]	Empirical Efficacy rate high (0.4, 0.6]	Empirical Efficacy rate superb (0.6, 1]
Empirical Toxicity rate low (0, 0.15]	E	E	E	E
Empirical Toxicity rate moderate (0.15, 0.33]	E	E	E	S
Empirical Toxicity rate high (0.33, 0.4]	D	S	S	S
Empirical Toxicity rate unacceptable (0.4, 1]	D	D	D	D

E, D and S denote “escalate dose”, “de-escalate dose”, and “stay at the same dose” respectively.

dose level, then we can escalate to the next higher dose level. However, additional rules are considered in our proposed design when Table 1 recommends escalating or de-escalating a dose level from the current dose, as shown in Table 2. These additional frequentist rules are similar in spirit to the two additional Bayesian rules (safety and futility) proposed and used in the TEPI design. They ensure that we consider dose levels that are neither too toxic nor inefficacious, and that our decision to escalate or de-escalate considers not only the observed cumulative toxicity and efficacy rates at the current dose but also the observed rates at the higher and lower dose(s). In the context of toxicity, note that the underlying toxicity rates are assumed to always increase with an increase in dose, but the underlying efficacy rates need not necessarily increase with an increase in dose. Also, these example rules in Table 2 used in our simulations can be modified for other trials/simulations. The

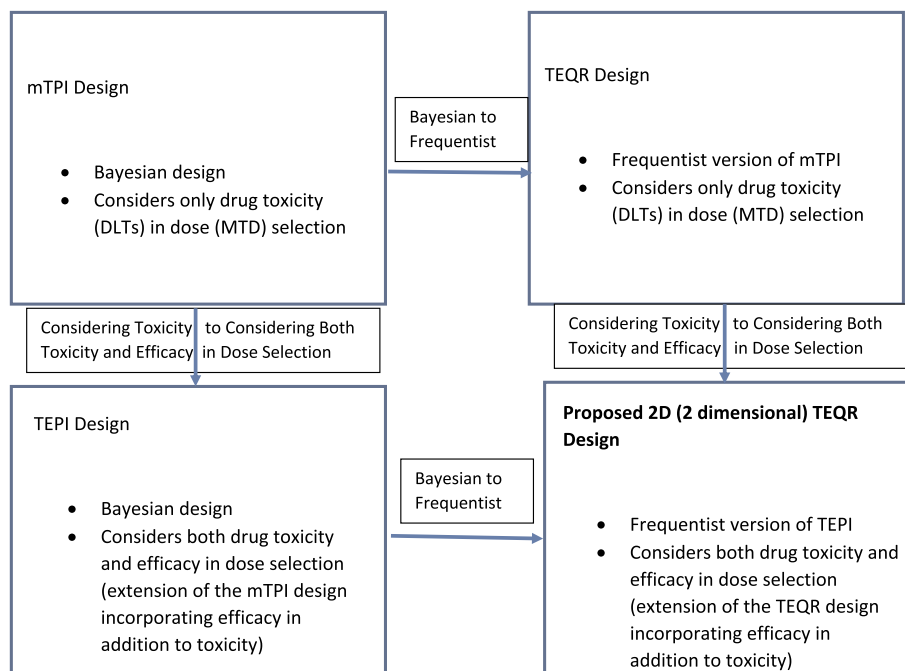


Fig. 1. Relation of mTPI, TEQR, TEPI and Proposed 2D TEQR Designs.

Table 2Additional Dosing Decisions to [Table 1](#) to be followed for the 2D TEQR design.

Table 1 says Escalate	
If toxicity rate < toxlim and efficacy rate \geq efflim at the next higher dose level, then	Escalate one dose level higher than the current one
If toxicity rate \geq toxlim at the next higher dose level, then irrespective of efficacy rate at the next higher dose level	Stay at the same dose level if efficacy rate \geq efflim at the same dose level, or if efficacy rate < efflim at the same dose level and the number of patients dosed at this dose level is $\leq 2 \times$ cohort size (this would be 6 patients for cohort size of 3 and 10 patients for cohort size of 5). Otherwise de-escalate one dose level lower than the current one
If toxicity rate < toxlim and efficacy rate < efflim at the next higher dose level, then	Escalate one dose level higher than the current one if the number of patients dosed at the next higher dose level is $\leq 2 \times$ cohort size Stay at the same dose level if the number of patients already dosed at the next higher dose level is $\geq 3 \times$ cohort size and efficacy rate \geq efflim at the same dose level, or if efficacy rate < efflim at the same dose level and the number of patients dosed at the same dose level is $\leq 2 \times$ cohort size. Otherwise de-escalate one dose level lower than the current one
Table 1 says De-Escalate	
If efficacy rate \geq efflim at the next lower dose level, then If efficacy rate < efflim at the next lower dose level, then	De-escalate to the next lower dose level than the current one De-escalate to the next lower dose level than the current one if the number of patients at the lower dose is $\leq 2 \times$ cohort size De-escalate one more dose level (i.e. 2 dose levels lower than the current one) if the number of patients at the next lower dose (1 dose level lower than the current one) is $\geq 3 \times$ cohort size
Table 1 says Stay	
	Stay at the same dose

Toxlim and efflim, the upper limit for toxicity and lower limit for efficacy, are parameters chosen by the physician and are respectively 0.51 and 0.1 in this paper.

intervals defining low, moderate, high and unacceptable toxicity as well as low, moderate, high and superb efficacy given in [Table 1](#) are example values. They can change for each trial and depend on the maximum accepted toxicity probability and the minimum accepted efficacy probability for that trial and study drug.

Thus, the dosing decisions of a 2D TEQR trial are based on [Tables 1 and 2](#), and these dosing decisions (escalate/stay at the same dose/de-escalate) stop when the maximum sample size is reached or when [Table 1](#) recommends de-escalating from dose level 1. A utility function is then used at the end of the trial to determine the optimal dose based on the toxicity and efficacy data from all the dose levels.

2.2. Selection of optimal dose

In our design, the optimal dose for efficacy and safety is selected at the end of the trial using the following utility function:

$$utility_i = r_i - c \cdot d_i$$

where, i represents the dose level, c is a constant that can vary between 0 and 1, r_i is the fraction of patients having a response and d_i is the fraction of patients with a DLT at dose level i . The dose level i providing the largest value of the utility function is the dose that is selected as optimal at the end of the trial. This kind of utility function was also employed in Ivanova et al. in their adaptive design to find the optimal dose, with application to a crossover dose finding study [30]. Note that $c = 1$ gives efficacy and toxicity equal weight, while a very small c (say $c = 0.1$) gives very little weight to toxicity. We also instituted boundary rules such that if the optimal dose selected has a toxicity (DLT) rate ≥ 0.51 or an efficacy (response) rate < 0.1 , then no dose is considered optimal – the values 0.51 and 0.1 were chosen as limits for toxicity and efficacy for this paper and can be changed for other cases.

3. An example of how the 2D TEQR might work in practice is shown in the appendix

3.1. Simulations

We compared the performance of our proposed 2D TEQR design with

that of its Bayesian counterpart, the TEPI design [14], as well as with that of the Eff-Tox design [10,11], that of the STEIN design [12] and that of the OBD Isotonic design [13], for different scenarios of true efficacy and toxicity rates. The scenarios considered in the main text for true toxicity and efficacy rates are based on those used in Li et al. to evaluate the TEPI design [14]; we also evaluated additional scenarios in the Appendix. For each scenario, we performed 10000 simulations of the 2D TEQR design using SAS. Each simulation stopped when the maximum sample size was reached, or if the design recommended de-escalating from dose level 1. The parameters used in the simulations of the 2D TEQR, TEPI, STEIN, Eff-Tox and OBD Isotonic designs are given in the Appendix.

For each scenario of true toxicity and efficacy rates in the main text, we summarize the average number of patients per dose level, and the percentage of times that each dose level was selected as the optimal dose. For our proposed 2D TEQR design, we selected the optimal dose in each simulation run using the utility function described above – the dose providing the largest value of the utility function in each simulated trial is the dose selected as optimal. If the optimal dose selected in a simulation run, has either a toxicity rate ≥ 0.51 or an efficacy rate < 0.1 , then no dose level was selected as optimal for that simulation run. The rates of 0.51 and 0.1 are chosen as limits for this paper and can be changed for other cases or trials, and would change the exact value for the accuracy of optimal dose selection from simulations.

We also summarize, for the different scenarios, the average percentages for patients treated at the optimal dose, those that were under-dosed, and those over-dosed. We observed the effect of changing the parameter c in the utility function on the accuracy of optimal dose selection. Finally, we observed the effect of changing the sample size and cohort size used on the accuracy of optimal dose selection.

4. Results

The results in [Tables 3 and 4](#) are obtained by using the parameters shown in Appendix [Table 1](#). In [Table 5](#), the value of c in the utility function is varied, while in [Table 6](#), the sample size and cohort size are varied.

The proposed frequentist 2D TEQR design is not as accurate as the Bayesian TEPI design in selecting the optimal dose, but is as accurate as

Table 3
Operating characteristics of the 2D TEQR, TEPI, eff-tox, OBD isotonic and STEIN designs.

Scenario	Dose Level	Toxicity	Efficacy	Percentage Each Dose is Selected as Optimal in Proposed 2D TEQR Design	Percentage Each Dose is Selected as Optimal in TEPI Design	Percentage Each Dose is Selected as Optimal in Eff-Tox Design	Percentage Each Dose is Selected as Optimal in OBD Isotonic Design	Percentage Each Dose is Selected as Optimal in STEIN Design	Average Number of Patients Treated at Each Dose in our TEQR 2D	Average Number of Patients Treated at Each Dose in TEPI Design	Average Number of Patients Treated at Each Dose in Eff-Tox Design	Average Number of Patients Treated at Each Dose in OBD Isotonic Design	Average Number of Patients Treated at Each Dose in STEIN Design
1	1	0.16	0.05	11.8	22.1	0	11.2	23.5	5.1	6.0	3.1	5.4	8.1
	2	0.2	0.1	7.7	17.9	0	20.2	32.8	4.7	5.7	3.1	6.3	8.1
	3	0.25	0.15	6.9	17.2	0	24.7	26.0	3.5	5.1	3.8	6.2	6.4
	4	0.3	0.18	3.4	7.5	8	43.9	9.4	1.9	4.3	5.3	9.2	3.7
2	1	0.15	0.8	66.5	83.9	66	37.9	89.1	9.0	9.1	17.8	9.9	24.4
	2	0.2	0.8	19.7	13.6	30	31.0	8.8	8.2	8.5	8.4	8.0	2.2
	3	0.25	0.8	5.8	2.1	3	18.9	0.8	4.9	5.6	0.7	5.3	0.2
	4	0.3	0.8	1.3	0.3	1	12.3	0.1	2.0	3.8	0.1	3.8	0.0
3	1	0.1	0.1	5.8	7.2	3	0.9	13.4	4.5	4.4	3.7	3.2	5.8
	2	0.2	0.7	65.0	88	42	95.8	84.2	9.9	12.3	11.8	19.1	19.4
	3	0.3	0.2	0.4	0.3	3	3.2	0.7	5.3	7.1	4.3	4.0	1.3
	4	0.7	0.1	0	0.1	2	0.1	0.0	1.1	2.3	2.0	0.6	0.4
4	1	0.15	0.43	49.3	53.9	19	25.1	40.1	6.6	6.1	7.3	8.2	11.6
	2	0.2	0.52	30.8	41.3	49	37.4	51.1	10	9.6	12.4	9.2	11.1
	3	0.4	0.5	4.5	3.6	22	21.9	7.1	5.5	9.0	5.3	5.3	3.6
	4	0.5	0.6	1.1	1.2	5	15.6	0.5	1.2	2.1	1.1	4.4	0.5
5	1	0.1	0.2	16.6	16.4	1	3.6	15.6	5.2	4.7	3.6	3.9	6.2
	2	0.2	0.6	47.5	65.4	48	40.5	66.1	8.6	8.6	12.4	9.4	15.4
	3	0.3	0.6	12.6	13.8	38	34.8	16.4	6.1	7.2	8.5	7.8	4.5
	4	0.4	0.6	2.6	1.0	10	21.1	1.4	2.2	4.9	2.0	5.9	0.7
6	1	0.5	0.4	30.7	33.9	16	45.3	16.9	6.2	14.9	7.9	10.1	13
	2	0.6	0.5	1.5	0.3	13	36.0	1.6	0.9	1.8	6.6	8.3	1
	3	0.7	0.6	0.1	0	2	15.9	0.1	0.05	0.1	1.2	5.3	0
	4	0.8	0.8	0	0	0	2.8	0.0	0	0	0.2	3.3	0

¹The maximum sample size is 27 and the cohort size is 3 for all the designs. The other parameters used for each design are given in the appendix.

²The dose level marked in bold is the optimal dose for each scenario.

No optimal dose is marked for Scenario 6 since even dose level 1 has a high toxicity rate.

No optimal dose is marked for Scenario 1 since dose level 4 has a low response rate.

³The percentages for dose selection for the proposed TEQR 2D design do not always add up to 100, since if the optimal dose selected in a simulation run (using the utility function) has a toxicity rate ≥ 0.51 or an efficacy rate of < 0.1 , then no dose level is selected as optimal for that simulation run. The percentages for dose selection for TEPI, STEIN, and Eff-Tox need not add to 100, but always do for OBD.

Table 4
Dosing at optimal dose, under-dosing and over-dosing in the 2D TEQR design and TEPI design.

Scenario	Percentage of Patients Dosed at Optimal Dose in the TEQR 2D Design	Percentage of Patients Under-Dosed in the TEQR 2D Design	Percentage of Patients Over-Dosed in the TEQR 2D Design	Percentage of Patients Dosed at Optimal Dose in the TEPI Design	Percentage of Patients Under-Dosed in the TEPI Design	Percentage of Patients Over-Dosed in the TEPI Design
2	40.6%	0%	59.4%	33.7%	0%	66.3%
3	37.5%	38.5%	24%	47.1%	16.9%	36.0%
4*	37.4% (DL1)	0% (DL1)	62.6% (DL1)	22.8% (DL1)	0% (DL1)	77.2% (DL1)
	37.1% (DL2)	37.4% (DL2)	25.5% (DL2)	35.8% (DL2)	22.8% (DL2)	41.4% (DL2)
5	32.3%	35.1%	32.6%	33.9%	18.5%	47.6%

For Scenario 4, the results are shown for both dose level 1 (DL1) and dose level 2 (DL2), based on Table 3, where dose level 2 is the true optimal dose but dose level 1 is chosen most frequently as the optimal dose by the 2D TEQR and TEPI designs.

Table 5
Effect of “c” in the utility function on accuracy of optimal dose selection in the 2D TEQR design.

Scenario	Accuracy of Optimal Dose Selection in the 2D TEQR Design when c = 1	Accuracy of Optimal Dose Selection in the 2D TEQR Design when c = 0.5	Accuracy of Optimal Dose Selection in the 2D TEQR Design when c = 0.1
2	66.5%	58.4%	49.9%
3	65.0%	65.9%	65.8%
4*	49.3% (DL1)	39.2% (DL1)	29.5% (DL1)
	30.8% (DL2)	32% (DL2)	27.4% (DL2)
5	47.5%	43.3%	35.6%

*For Scenario 4, the results are shown for both dose level 1 (DL1) and dose level 2 (DL2), based on Table 3, where dose level 2 is the true optimal dose but dose level 1 is chosen most frequently as the optimal dose by the 2D TEQR and TEPI designs.

or more accurate than the Eff-Tox design in selecting the optimal dose, for the scenarios examined (Table 3). However, the Eff-tox design does treat a larger percentage of patients at the optimal dose compared to the TEPI, 2D TEQR and OBD Isotonic designs in the scenarios presented in Table 3. The STEIN design is very comparable to the TEPI design in terms of accuracy of optimal dose selection and treats a larger percentage of patients at the optimal dose compared to the other designs for the scenarios in Table 3. The OBD Isotonic design does very well compared to the 2D TEQR design and the other designs in terms of optimal dose selection and patient allocation to the optimal dose when

Table 6
Effect of sample and cohort size on accuracy of optimal dose selection in the 2D TEQR design.

Scenario	Accuracy of Optimal Dose Selection in the 2D TEQR Design when sample size = 15 and cohort size = 3	Accuracy of Optimal Dose Selection in the 2D TEQR Design when sample size = 27 and cohort size = 3	Accuracy of Optimal Dose Selection in the 2D TEQR Design when sample size = 48 and cohort size = 3	Accuracy of Optimal Dose Selection in the 2D TEQR Design when sample size = 60 and cohort size = 3	Accuracy of Optimal Dose Selection in the 2D TEQR Design when sample size = 25 and cohort size = 5	Accuracy of Optimal Dose Selection in the 2D TEQR Design when sample size = 50 and cohort size = 5	Accuracy of Optimal Dose Selection in the 2D TEQR Design when sample size = 60 and cohort size = 5	Accuracy of Optimal Dose Selection in the 2D TEQR Design when sample size = 80 and cohort size = 5
2	68.6%	66.5%	65.7%	65.1%	66.6%	63.1%	62.5%	63.0%
3	63.7%	65.0%	65.8%	65.9%	82.2%	84.1%	83.7%	84.3%
4*	49.5% (DL1)	49.3% (DL1)	49.2% (DL1)	49.5% (DL1)	49.5% (DL1)	49.2% (DL1)	49.2% (DL1)	50.3% (DL1)
	30.2% (DL2)	30.8% (DL2)	32.3% (DL2)	33.3% (DL2)	37.4% (DL2)	40.3% (DL2)	41.0% (DL2)	40.8% (DL2)
5	46.3%	47.5%	48.0%	47.0%	54.1%	54.6%	55.0%	55.8%

*For Scenario 4, the results are shown for both dose level 1 (DL1) and dose level 2 (DL2), based on Table 3, where the true optimal dose is dose level 2 but dose level 1 is chosen most frequently as the optimal dose by the 2D TEQR and TEPI designs.

there is a clear peak in the true response rates (Scenario 3 of Table 3). Note that the OBD design picks an optimal dose for each simulation run – hence, for Scenario 6, it picks dose level 1 with the lowest toxicity most frequently (Table 3), although the underlying toxicity rate of dose level 1 of 0.5 is higher than the chosen upper bound for the toxicity rate of 0.4. The accuracy of optimal dose selection of the proposed 2D TEQR design can be increased, in many cases, with a moderate increase in cohort size (say cohort size 5; see Table 6 and further examples in the Appendix). The proposed 2D TEQR design puts a comparable percentage of patients at the optimal dose as the Bayesian TEPI design and overdoses a smaller percentage of patients but underdoses a larger percentage of patients, for the scenarios examined (Table 4). For the utility function used here, c = 1 appears to be the best choice in terms of selecting the optimal dose accurately for the proposed 2D TEQR design for the scenarios examined (Table 5 and Appendix scenarios). In each scenario (Table 5), the same dose level is chosen as the optimal dose for c = 1, c = 0.5 and c = 0.1, but is chosen most frequently for c = 1 in most cases (see Scenario 3 of Table 5 where this is not true).

In the scenarios we examined in the main text, where the true optimal dose is a low dose level (dose level 1 or 2), an increase in sample size (say from 27 to 48 for a cohort size of 3, or from 25 to 50 for a cohort size of 5) did not substantially increase the accuracy of optimal dose selection - in fact in Scenario 2 alone, the accuracy of optimal dose selection decreases with an increase in sample size since dose level 2, which has the same efficacy and only a slightly higher toxicity than dose level 1, is chosen more frequently as the sample size is increased (Table 6). However, if the true optimal dose is a higher dose (say dose level 4 or 5), then the accuracy of optimal dose selection is low for a small (inadequate) sample size; in this case, increasing the sample size (from say 25 to 50 for a cohort size of 5) does substantially increase the

accuracy of dose selection (see examples in the Appendix).

5. Discussion and conclusion

We propose a new frequentist design for optimal dose selection, the 2D TEQR design, that is easier to understand and simpler to implement than the TEPI, STEIN, Eff-Tox and OBD designs, as it is based on the observed toxicity and efficacy rates and does not require specialized computations. The TEPI, OBD Isotonic and Eff-Tox are Bayesian designs, which require the calculation of posterior probabilities. The STEIN design is model free like the 2D TEQR design and bases its dosing decisions (escalate/stay at the same dose/de-escalate) on the observed DLT and response rates. However, while using the observed DLT and response rates in its dosing algorithm, it also incorporates a posterior probability calculation – the dose level is retained if the observed DLT rate at the dose level is less than the upper critical value for the observed toxicity and the observed efficacy is greater than the critical value for response, but if there is no satisfactory dose, the admissible dose that has the largest posterior probability of efficacy is chosen [12].

We compare the performance of the proposed 2D TEQR design with those of the TEPI, STEIN, Eff-Tox and OBD Isotonic designs for accuracy of optimal dose selection and patient allocation to the optimal dose. Although the optimal dose selection criterion is not identical for these designs, the same dose level is recognized as the optimal dose level by all the designs for Scenarios 2–5 in Table 3. The frequentist 2D TEQR is not as accurate in selecting the optimal dose as the Bayesian TEPI design for the same sample and cohort size,¹ but it is as good as or better than the Eff-Tox design for the scenarios examined. However, the Eff-Tox design treats a larger percentage of patients at the optimal dose compared to the TEPI, 2D TEQR and OBD Isotonic designs for the scenarios examined in Table 3. The 2D TEQR design also selects the optimal dose more accurately than the OBD Isotonic design, unless there is a clear peak in the true response rates, in which case the OBD Isotonic design performs very well compared to the other designs in terms of optimal dose selection and patient allocation to the optimal dose. The STEIN design is very comparable to the TEPI design in terms of accuracy of optimal dose selection and treats a larger percentage of patients at the optimal dose compared to the other designs for the scenarios in Table 3. For many scenarios, the accuracy of optimal dose selection of the proposed 2D TEQR design can be increased for a similar sample size by increasing the cohort size to a moderate cohort size of say 5.

The exact value for the accuracy of optimal dose selection from simulations of the 2D TEQR design would also change with the utility function used. We use a utility function that is very simple to implement and easy to understand - we considered applying isotonic regression to the observed toxicity rates to ensure that they are non-decreasing before applying our simple utility function, but this did not change the accuracy of dose selection substantially in any of the scenarios examined. The accuracy of dose selection would likely be improved with a more complex utility function, similar to the one used in the TEPI design [14]. However, we did not explore other utility functions and systematically evaluate the effect of changing the utility function on the accuracy of optimal dose selection in the 2D TEQR design. For the simple utility function used in this paper, giving equal weight to efficacy and toxicity appears to be the best choice for accuracy of optimal dose selection for the scenarios we examined. Note also that using a utility function at the end of the trial is not the only method to select the final (optimal) dose. As stated in the TEPI paper, the authors could have also proposed to select the dose with the highest probability $\Pr(p_i < p_T, q_i > q_E + \delta | D)$ as the best dose, where δ is the expected increment over the minimum efficacy rate [14]. The proposed 2D TEQR design puts a comparable percentage

¹ Note the frequentist TEQR is also not as accurate as the mTEPI design in MTD dose selection for the same sample and cohort size, while using the same stopping rules [31].

of patients at the optimal dose as the TEPI design; it overdoses a smaller percentage of patients, but correspondingly underdoses a larger percentage of patients than the TEPI design.

The accuracy of optimal dose selection of the proposed 2D TEQR design does not increase substantially with an increase in sample size for the scenarios examined in the main text; a sample size of 27 for a cohort size of 3 seems to be adequate to determine the optimal dose in these cases. Note however, that for the scenarios examined in the main text, the true optimal dose is dose level 1 or 2, and in these cases, a large sample size may not be needed. As an example, for a sample size of 27 and cohort size of 3, there are 9 cohorts of patients available, which may be sufficient to determine the optimal dose in these cases where the true optimal dose is a lower dose level. However, if the true optimal dose is dose level 4 or 5 and the sample size is 18 patients with a cohort size of 3, then only 6 cohorts of 3 patients each are available, which may not be sufficient to determine the optimal dose accurately. In such cases, an increase in sample size from 18 to 36 would increase the accuracy of optimal dose selection substantially. The examples shown in the Appendix illustrate the above point - that if the true optimal dose is a higher dose level, then the accuracy of optimal dose selection is low for a small (inadequate) sample size and improves substantially if the sample size is increased. Thus, although in general, the accuracy of optimal dose selection increases with sample size, the percentage of increase in the accuracy of dose selection depends on the true toxicity and efficacy rates, the location of the true optimal dose (i.e. true optimal dose level), the efficacy and toxicity intervals chosen and the upper and lower limits chosen for toxicity and efficacy respectively. We find that this conclusion just described for the 2D TEQR design also holds for the corresponding Bayesian TEPI design regarding the effects of an increase in sample size on the accuracy of optimal dose selection. In the TEPI design too, the accuracy of optimal dose selection increases in many scenarios with a moderate increase in cohort size (simulation results with regard to sample and cohort size are not shown here for the TEPI design).

In conclusion, the proposed 2D TEQR design, a frequentist version of the TEPI design, is a simple design that is easy to understand and implement. Although for the same sample size and cohort size, it is less accurate than the Bayesian TEPI design in selecting the optimal dose, it is simpler to implement and its accuracy of optimal dose selection can be increased, in many cases, with a moderate increase in cohort size.

Acknowledgments

We thank the reviewers for their useful comments to improve the paper. We also thank Dr. Ruitao Lin for sharing his code for the STEIN design and answering questions on the STEIN design.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.conctc.2019.100461>.

References

- [1] J. O'Quigley, M. Pepe, M.L. Fisher, Continual reassessment method: a practical design for phase 1 clinical trials in cancer, *Biometrics* 46 (1990) 33–48.
- [2] E. Garrett-Mayer, The continual reassessment method for dose-finding studies: a tutorial, *Clin. Trials* 3 (1) (2006) 57–71.
- [3] Y. Ji, P. Liu, Y. Li, B.N. Bekele, A modified toxicity probability interval method for dose-finding trials, *Clin. Trials* 7 (6) (2010) 653–663.
- [4] M.S. Blanchard, J.A. Longmate, Toxicity equivalence range design (TEQR): a practical Phase I design, *Contemp. Clin. Trials* 32 (1) (2011) 114–121.
- [5] M. Tighiouart, A. Rogatko, Dose finding with escalation with overdose control (EWOC) in cancer clinical trials, *Stat. Sci.* 25 (2) (2010) 217–226.
- [6] S. Liu, Y. Yuan, Bayesian optimal interval designs for phase I clinical trials, *J. R. Stat. Soc. Ser. A C* 64 (2015) 507–523.
- [7] F. Yan, S.J. Mandrekar, Y. Yuan, Keyboard: a novel bayesian toxicity probability interval design for phase I clinical trials, *Clin. Cancer Res.* 23 (15) (2017 Aug 1) 3994–4003, <https://doi.org/10.1158/1078-0432.CCR-17-0220>. Epub 2017 May 25.

- [8] M. Liu, S.J. Wang, Y. Ji, The $i3+3$ design for phase I clinical trials, *J. Biopharm. Stat.* (2019 Jul 15) 1–11.
- [9] R. Ananthakrishnan, S. Green, M. Chang, G. Doros, J. Massaro, M. LaValley, Systematic comparison of the statistical operating characteristics of various Phase I oncology designs, *Contemp. Clin. Trials Commun.* 5 (2016 Nov 24) 34–48.
- [10] P.F. Thall, J.D. Cook, Dose-finding based on efficacy-toxicity trade-offs, *Biometrics* 60 (3) (2004) 684–693.
- [11] P.F. Thall, J.D. Cook, E. Estey, Adaptive dose selection using efficacy-toxicity trade-offs: illustrations and practical considerations, *J. Biopharm. Stat.* 16 (5) (2006) 623–638.
- [12] R. Lin, G. Yin, STEIN: a simple toxicity and efficacy interval design for seamless phase I/II clinical trials, *Stat. Med.* 36 (26) (2017) 4106–4120.
- [13] Y. Zang, J.J. Lee, Y. Yuan, Adaptive designs for identifying optimal biological dose for molecularly targeted agents, *Clin. Trials* 11 (3) (2014 Jun) 319–327.
- [14] D.H. Li, J.B. Whitmore, W. Guo, Y. Ji, Toxicity and efficacy probability interval design for phase I adoptive cell therapy dose-finding clinical trials, *Clin. Cancer Res.* 23 (1) (2017 Jan 1) 13–20, <https://doi.org/10.1158/1078-0432.CCR-16-1125>. Epub 2016 Oct 14.
- [15] H. Pan, P. Huang, Z. Wang, L. Wang, C. Li, J. Xia, A novel Bayesian seamless phase I/II design, *PLoS One* 8 (9) (2013 Sep 4), e73060.
- [16] H. Pan, F. Xie, P. Liu, J. Xia, Y. Ji, Clin Trials. A phase I/II seamless dose escalation/expansion with adaptive randomization scheme (SEARS), *Clin. Trials* 11 (1) (2014) 49–59.
- [17] A. Hoering, M. LeBlanc, J. Crowley, Seamless phase I–II trial design for assessing toxicity and efficacy for targeted agents, *Clin. Cancer Res.* 17 (4) (2011 Feb 15) 640–646.
- [18] A. Hoering, A. Mitchell, M. LeBlanc, J. Crowley, Early phase trial design for assessing several dose levels for toxicity and efficacy for targeted agents, *Clin. Trials* 10 (3) (2013) 422–429.
- [19] T.M. Braun, The bivariate continual reassessment method. extending the CRM to phase I trials of two competing outcomes, *Contr. Clin. Trials* 23 (3) (2002 Jun) 240–256.
- [20] G. Yin, Y. Li, Y. Ji, Bayesian dose-finding in phase I/II clinical trials using toxicity and efficacy odds ratios, *Biometrics* 62 (3) (2006 Sep) 777–784.
- [21] W. Zhang, D.J. Sargent, S. Mandrekar, An adaptive dose-finding design incorporating both toxicity and efficacy, *Stat. Med.* 25 (14) (2006 Jul 30) 2365–2383.
- [22] V. Dragalin, V.V. Fedorov, Y. Wu, Two-stage design for dose-finding that accounts for both efficacy and safety, *Stat. Med.* 27 (25) (2008 Nov 10) 5156–5176.
- [23] W. Guo, Y. Ni, Y. Ji, TEAMS: toxicity- and efficacy-based dose insertion design with adaptive model selection for phase I/II dose-escalation trials in oncology, *Stat. Biosci.* 7 (2) (2015 Oct) 432–459.
- [24] J. Lee, P.F. Thall, Y. Ji, P. Müller, Bayesian dose-finding in two treatment cycles based on the joint utility of efficacy and toxicity, *J. Am. Stat. Assoc.* 110 (510) (2015 Jun 1) 711–722.
- [25] K. Takeda, M. Taguri, S. Morita, BOIN-ET: Bayesian optimal interval design for dose finding based on both efficacy and toxicity outcomes, *Pharm. Stat.* 17 (4) (2018 Jul) 383–395.
- [26] M. Markman, Serious ethical dilemma of single-agent pegylated liposomal Doxorubicin employed as a control arm in ovarian cancer chemotherapy trials, *J. Clin. Oncol.* 28 (19) (2010 Jul 1) e319–320.
- [27] R.L. Schilsky, L. Minasian, D. Auclair, A. Rahman, R. Pazdur, Optimizing dosing of oncology drugs, in: Conference on Clinical Cancer Research, 2013.
- [28] H. Sato, A. Hirakawa, C. Hamada, An adaptive dose-finding method using a change-point model for molecularly targeted agents in phase I trials, *Stat. Med.* 35 (23) (2016) 4093–4109.
- [29] N.A. Wages, C. Tait, Seamless phase I/II adaptive design for oncology trials of molecularly targeted agents, *J. Biopharm. Stat.* 25 (5) (2015) 903–920.
- [30] A. Ivanova, K. Liu, E. Snyder, D. Snively, An adaptive design for identifying the dose with the best efficacy/tolerability profile with application to a crossover dose-finding study, *Stat. Med.* 28 (24) (2009 Oct 30) 2941–2951.
- [31] R. Ananthakrishnan, S. Green, D. Li, M. LaValley, Extensions of the mTPI and TEQR designs to include non-monotone efficacy in addition to toxicity for optimal dose determination for early phase immunotherapy oncology trials, *Contemp. Clin. Trials Commun.* 10 (2018 Jan 31) 62–76.
- [32] J.S. Babb, A. Rogatko, Patient specific dosing in a cancer phase I clinical trial, *Stat. Med.* 20 (14) (2001) 2079–2090.
- [33] S. Liu, V.E. Johnson, Biostatistics. A robust Bayesian dose-finding design for phase I/II clinical trials 17 (2) (2016 Apr) 249–263.
- [34] J.R. Sachs, K. Mayawala, S. Gadamsetty, S.P. Kang, D.P. de Alwis, Optimal dosing for targeted therapies in oncology: drug development cases leading by example, *Clin. Cancer Res.* 22 (6) (2016 Mar 15) 1318–1324.