An information-theoretic Phase I/II design for molecularly targeted agents that does not require an assumption of monotonicity. Supplementary Materials.

Pavel Mozgunov and Thomas Jaki

Medical and Pharmaceutical Statistics Research Unit, Department of Mathematics and Statistics, Lancaster University, Lancaster, UK.

1. Application of WE design to a single-agent trial

The proposed design can be applied to a wide range of Phase I/II clinical trials. While the performance of the WE design is demonstrated in the context of the motivating trial, it can be also applied to a single agent dose-finding trial for which several model-based designs were recently proposed (see e.g. Wages and Tait, 2015; Riviere et al., 2016). Here we show the comparison of the proposed design to the currently used and provide a step-by-step algorithm how the parameters of the proposed design can be calibrated.

1.1. Simulation setting

We consider $M = 6$ doses and $N = 60$ patients. The dose-toxicity relationship is known to be a non-decreasing function, but a clinician expects either a plateau or an umbrella shape for the dose-efficacy curve. A toxicity is evaluated after three weeks while an efficacy outcome is evaluated after six weeks. To conduct the trial in a timely manner, the next cohort of patients is allocated after the toxicity data for previous cohort are available. The upper toxicity and the lowest efficacy bounds are $\phi = 0.35$ and $\psi = 0.20$. The goal is to study the ability of the WE design to identify the *optimal* and *correct* doses. A dose is called *optimal* if it is safe, has maximal efficacy and minimal toxicity while a safe dose with maximum efficacy (irrespective of it also having lowest toxicity) is called *correct*.

We consider 14 scenarios that were used for the motivating trial simulations: eight plateau scenarios (1-8) suggested by Riviere et al. (2016), 4 umbrella shaped scenarios (9-12) studied in Wages and Tait (2015) and two scenarios with no correct doses (13-14, due to inefficacy and toxicity, respectively) − see Figure 4 in the main paper.

In the analysis we focus on (i) the proportion of optimal/correct recommendations, (ii) the average number of toxic responses, (iii) the average number of efficacy responses. The study is performed using `R` (R Core Team, 2015) and 10,000 replications for each scenario. We compare the characteristics with the 'MTA' design proposed by Riviere et al. (2016) and the 'WT' design developed by Wages and Tait (2015). Parameters of the designs are chosen as in the original proposals with an exception of using cohort size $c = 3$ and 80% confidence intervals for stopping rules for the WT design.

1.2. Design specification

As before, we use a target toxicity of $\gamma_t = 0.01$ and a target efficacy of $\gamma_e = 0.99$. Due to the known toxicity ordering, the design is restricted to satisfy the coherence principals with $q = 1$. While we consider both non-randomized and randomized versions of the WE design to study an allocation rule impact, the design specification for the non-randomized WE design is provided only.

1.2.1. Prior

Parameters $\beta_{t,i} = \beta_{e,i} = 1$ of the prior Beta distribution in (8) are chosen for all dose levels $i = 1, \dots, M$ to emphasize a limited available information. Parameters $\nu_{t,i}$ and $\nu_{e,i}$ (which coincide the prior probabilities of toxicity and efficacy for $\beta_{t,i} = \beta_{e,i} = 1$) are specified such that the WE design leads to accurate optimal dose recommendation in various different scenarios. The prior values of $\nu_{t,i}$ and $\nu_{e,i}$ are calibrated over scenarios 1-8 with different locations of the optimal and correct doses. There are two restrictions on the prior parameters: the escalation should start at the first dose and no dose skipping is allowed. To restrict number of possible parameters to be calibrated over, we assume that prior efficacy and toxicity probability increases linearly as $\nu_{t,i} = start_t + w_t \times i$ and $\nu_{e,i} = start_e + w_e \times i$. Then, we search for the values of $start_t, start_e, w_t, w_e$ such that the geometric mean of the proportion of optimal selection over all scenarios is maximised.

Prior vectors of toxicity probabilities $\hat{\mathbf{p}}_t^{(0)} = [0.05, 0.14, 0.23, 0.32, 0.41, 0.50]^{\mathrm{T}}$ and efficacy probabilities $\hat{\mathbf{p}}_e^{(0)} = [0.55, 0.58, 0.61, 0.64, 0.67, 0.70]^{\mathrm{T}}$ are subsequently used for the non-randomized WE design.

Similarly, vectors of prior toxicity $\hat{\mathbf{p}}_t^{(0)} = [0.25, 0.35, 0.45, 0.55, 0.65, 0.75]^{\mathrm{T}}$ and $\hat{\mathbf{p}}_e^{(0)} = [0.65, 0.69, 0.73, 0.77, 0.81, 0.85]^{\mathrm{T}}$ efficacy probabilities are used for the randomized WE(R) design. It was found that the randomised WE(R) design is more robust to the choice of the prior parameter than non-randomised WE.

1.2.2. Safety constraint

To set the time-varying safety constraint, we use $\zeta_N = 0.30$ and calibrate ϕ^*, r_t using the highly toxic scenario 14 and the flat scenario 6. These two scenarios are chosen to represent the trade-off in the safety constraint. The proportion of correct recommendations (terminations) and mean number of patients involved in a trial for different parameters values are given in Figure 1. The mean number of patients in scenario 6 does not vary a lot and the corresponding graph is not shown. In scenario 6 the highest proportion of the optimal recommendations corresponds to the least strict safety constraint (right bottom corner), but only 35% of trials in scenario 14 are then terminated. At the same time, the most strict rule (left top corner) results in 100% of terminations in scenario 14, but only in 5% of correct recommendation in scenario 6. Parameters $r_t = 0.0125$ and $\phi^* = 0.4$ are chosen for subsequent study as a reasonable trade-off. The same parameters of the safety constraints are used for the randomized design.

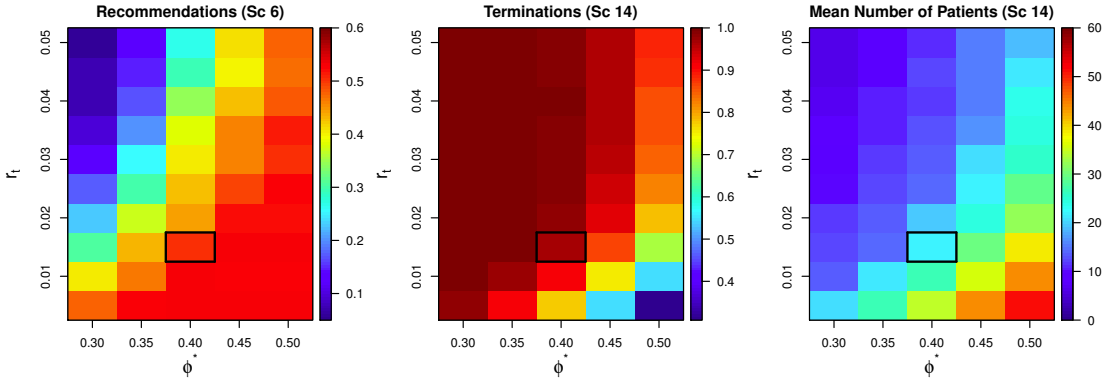

Fig. 1. Safety constraint parameters calibration: $\phi^* \in (0.3, 0.5)$, $r_t \in (0, 0.05)$ in scenarios 6 and 14. The proportion of correct recommendations (terminations) and the mean number of patients in a trial (scenario 14). The final choice is marked by a black frame. Results are based on 10^4 replications.

1.2.3. Futility constraint

We calibrate the futility constraint by fixing $\xi_N = 0.50$ and tuning ψ^* and r_e using two opposite scenarios - 2 and 13. In scenario 2 all doses have the same efficacy probability. In scenario 13 there are no correct doses as all efficacious doses have unacceptable toxicity. The proportion of correct recommendations (terminations) and the mean number of patients are given in Figure 2. Since the mean number of patients in scenario 2 does

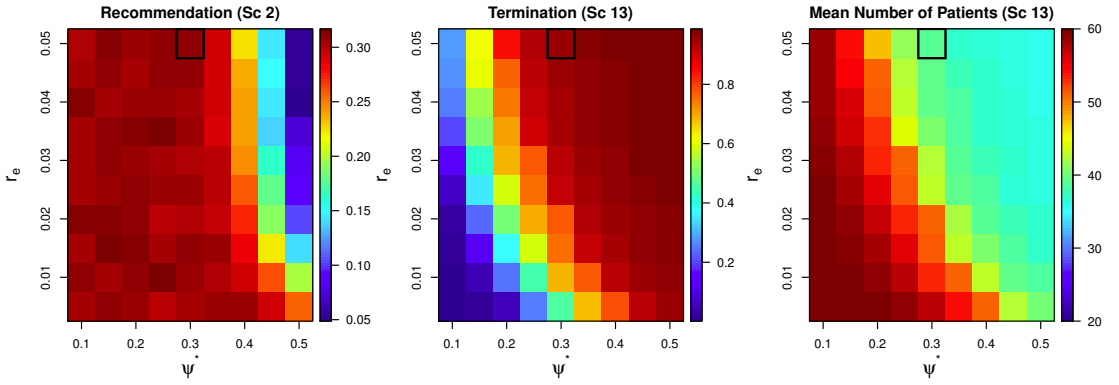

Fig. 2. Futility constraint parameters calibration: $\psi^* \in (0.1, 0.5)$ and $r_e \in (0, 0.05)$ in scenarios 2 and 13. The proportion of correct recommendations (terminations) and the mean number of patients in a trial (scenario 13). The final choice is in the black frame. Results are based on 10^4 replications.

not vary this graph is not shown. A stricter constraint is favourable in scenario 13 and less favourable in scenario 2 while the opposite is true for less strict constraints. Subsequently, parameters $\psi^* = 0.3$ and $r_e = 0.05$ are used for both non-randomized and randomized designs.

Table 1. Proportion of optimal and correct dose recommendations in scenarios 1-8 and 13-14 using $N = 60$ patients and $M = 6$ doses for all considered designs. Figures corresponding to the highest performance in each scenario are in **bold**. The results are based on 10^4 replicated trials.

Scenario	1	2	3	4	5	6	7	8	13	14
Proportion of the optimal dose recommendation										
WE	58.8	30.3	**65.8**	**71.1**	**60.3**	53.5	**60.0**	37.8	**95.2**	**96.9**
WE (R)	**72.0**	35.0	51.0	**69.9**	54.5	56.7	48.5	36.4	**93.2**	**97.4**
MTA	57.0	**60.2**	48.4	53.7	55.3	55.9	37.9	**43.0**	91.9	91.0
WT	19.6	41.9	29.3	25.3	27.0	**65.2**	27.1	26.1	91.5	90.4
Proportion of the correct dose recommendation										
WE	60.3	90.0	87.5	79.8	**89.7**	53.5	77.8	**91.5**	-	-
WE (R)	89.3	**97.0**	**92.8**	**89.1**	**87.5**	56.7	88.3	**91.1**	-	-
MTA	94.1	**96.6**	83.8	82.3	80.6	55.9	89.9	77.4	-	-
WT	**98.1**	**97.6**	**93.6**	86.5	80.0	**65.2**	**93.3**	81.1	-	-

1.3. Operating characteristics

The results of the comparison in scenarios 1-8 with plateau dose-efficacy relation and in scenarios 13-14 with no correct doses are summarized in Table 1 and Table 2. Each figure in Table 1 corresponds to proportions of optimal or correct dose recommendations. The detailed results, such as the selection proportions and mean number of patients on each dose, are given in Table 5 and Table 7 in the Appendix.

With respect to the *optimal* dose recommendation, both versions of the proposed design perform comparably or better than model-based designs in the majority of scenarios. The WE design without randomization leads to a considerable improvement in scenarios 3-5, 7 and 13-14 and outperform the best model-based alternative by up to 20%. While the randomized WE(R) shows the comparable to the best model-based alternative performance in scenarios 3,5 and 6, it also results in more accurate optimal dose recommendations in scenarios 1, 4, 7, 13-14. However, both WE and WE(R) are outperformed by MTA in scenarios 2 and 8 in which dose-toxicity and dose-efficacy curves are flat in the neighbourhood of the optimal dose. While MTA recommends the lowest dose by default, such small differences in toxicity and efficacy probabilities are difficult to find the small sample size. At the same time, the absence of a parametric model is not found to be a problem in any other cases. WT design outperforms all other designs in scenario 6 with the optimal dose being the highest safe one. Generally, WT is less conservative as it favours safe doses with higher toxicities that results in a low proportion of optimal recommendations if the optimal dose is not the highest safe one (see also Table 5 and Table 7).

Considering the proportion of the *correct* recommendations, WT outperforms MTA in all scenarios and has the best performance among all alternatives in scenarios 1-3 and 6-7. In the rest of scenarios WT has either comparable or worse performance than the randomized WE(R). Comparing WE and WE(R), the randomized design is more robust in the correct recommendations with a largest difference in scenario 1. Here, the chosen prior would not escalate to dose 6 once the optimal is already find at dose 5 if no randomization is used.

In terms of toxicities we find that the non-randomized WE design results in con-

Table 2. Mean number of toxicity and efficacy responses in scenarios $1-8$ and $13-14$ using $N=60$ patients and $M=6$ doses for all considered designs. The results are based on 10^4 replicated trials.

Scenario	1	2	3	4	5	6	7	8	13	14
Toxicity responses										
WE	**3.1**	**6.0**	**2.8**	**4.5**	**7.9**	**11.0**	**5.9**	**5.6**	**11.7**	**11.0**
WE (R)	4.0	6.9	4.3	6.0	8.7	**10.8**	6.9	6.9	13.0	**10.9**
MTA	5.5	8.1	6.0	10.0	12.1	13.2	9.6	9.3	**11.0**	**11.5**
WT	6.8	6.7	7.3	13.2	13.5	14.7	10.0	9.1	**11.2**	12.1
Efficacy responses										
WE	28.5	**24.0**	33.0	30.0	29.8	19.4	34.6	**28.9**	6.0	**9.1**
WE (R)	27.4	**24.0**	**34.5**	32.2	29.8	19.2	36.5	**29.2**	**7.1**	**8.9**
MTA	38.0	**24.0**	**34.6**	35.0	29.4	21.4	**39.1**	**29.3**	6.2	**9.6**
WT	**41.5**	**24.0**	**35.5**	**37.0**	**32.3**	**24.4**	**39.9**	**29.2**	5.4	**9.7**

Table 3. Proportion of optimal dose recommendations, mean number of toxicity and efficacy responses in scenarios 9-12 using $N=60$ patients and $M=6$ doses for all considered designs. The results are based on 10^4 replicated trials.

Scenario	9	10	11	12	9	10	11	12	9	10	11	12
	Optimal recommendation				Toxicity responses				Efficacy responses			
WE	**54.7**	**55.9**	46.5	80.1	**4.5**	**5.7**	**10.0**	**1.8**	**29.7**	**25.6**	27.4	35.4
WE (R)	**56.7**	**56.2**	47.9	70.9	5.5	6.7	**10.0**	3.3	28.4	24.6	27.1	32.7
MTA	20.3	35.3	46.0	**96.1**	5.0	6.3	12.8	2.6	26.3	23.9	28.7	35.5
WT	50.1	49.3	**56.9**	75.9	5.5	5.9	12.2	2.4	27.9	24.8	**29.4**	**36.3**

siderably lower number of toxicities in almost all scenarios with the largest difference observed in scenario 4. As the WT approach is less conservative, it results in a greater number of toxicities, but also leads to the highest average number of efficacies in all scenarios. In contrast, the cost of the WE's lowest number of toxicities is a smaller number of efficacies. In scenarios 13 and 14 with no optimal and correct doses all alternatives result in nearly the same average number of toxicities and efficacies.

The results of the comparison in scenarios 9-12 with an umbrella shaped dose-efficacy relationship and only one correct dose are given in Table 3. Overall, WE designs have more robust optimal dose identification in non-monotonic scenarios. The WE design with no randomization outperforms MTA by up to 35% and WT by up to 6%. WT has the highest proportion of the optimal dose recommendations in scenarios 11 with nearly 10% difference with the non-randomized WE. The MTA design is more conservative and recommends d_1 with the highest probability that results in the best performance in scenario 12, but poor performance in other cases. The non-randomized WE is favourable compared to the randomized version due to the single correct dose in each scenario. The average number of toxicities of the WE design is again the safest alternative. In contrast to the scenarios with plateau, it can now also result in a larger number of efficacy responses (e.g. in scenario 9) due to the non-monotonic shape of the dose-efficacy curve.

Overall, the proposed approaches have better or comparable operating characteristics in 9 out of 14 considered scenarios even with less information used in a trial. Comparing two assignment rule of the WE design, the non-randomized WE is always less accurate

in terms of the correct dose identification. As the result, the WE design without randomization should be preferred if only one correct dose is expected or a clinician is cautious about toxicity profile, while the randomized WE is a robust choice if multiple *correct* doses are expected.

1.4. Early efficacy data

In the setting above, it is assumed that it takes twice as long to observe the efficacy outcome than the toxicity endpoint. It is, however, possible that an efficacy (or lack of efficacy) can be observed at the time of the interim analysis for some of the patients. As the proposed design includes all available information, it can also accommodate earlier efficacy (no efficacy) data. This section we study how the operating characteristics of the non-randomised WE design are affected if a certain proportion of 'no efficacy' responses can be observed earlier.

The setting above remains unchanged with the following exception: if the patient has observed no DLT and will have 'no efficacy ', it is assumed that the outcome can be observed at the time of toxicity evaluation with probability π. If observed earlier, the WE design uses this information for the next patient allocation. We consider two cases: $\pi = 0$ (the original setting) and $\pi = 1/2$ (half of 'no efficacies' can be seen earlier). The results are given in Table 4.

As expected, the availability of some of the efficacy information earlier leads to a less conservative design that allows more rapid escalation. Earlier 'no efficacy' data even in half of the patients lead to more ethical patient allocation. This can be seen by increased numbers of efficacies almost in all scenarios with the cost of reasonable increase in the average number of toxicity responses. The largest increase can be seen in scenario 1 where the average number of efficacy response increase by nearly 7, while toxicity increases only by 1. The information about earlier efficacy also improves the proportion of optimal recommendations in the scenarios where the target dose is high - by 6% in scenario 1 and by 4% in scenario 6. As the design being less conservative it favours higher doses among correct ones. This decreases the proportion of optimal recommendations in scenario 3, 5, 7 and 12 by 3-7%. At the same time, the proportion of correct recommendations is either unchanged (scenario 5 and 8) or increased by at least 5% (all the rest plateau scenarios). This confirms that the WE design in the setting with no earlier efficacy information is more conservative, but the difference in correct selection is relatively small.

Table 4. Operating characteristics of WE in scenarios 1-12 with no early efficacy data available ($\pi = 0$) and with half 'no efficacy' outcomes ($\pi = 1/2$) available at the time of toxicity evaluation: recommendation proportions, mean number of toxicity (T) and efficacy (E) responses . The optimal dose is in **bold** and correct doses are underlined. Results are based on 10^4 replications.

WE	d_1	d_2	d_3	d_4	d_5	d_6	T	E
				Scenario 1				
	(.005;.01)	(.01;.10)	(.02;.30)	(.05;.50)	**(.10;.80)**	(.15;.80)		
$\pi = 0$	0.0	0.1	2.3	37.4	**58.8**	1.5	3.1	28.5
$\pi = 1/2$	0.0	0.0	1.5	23.5	**64.9**	10.0	4.3	34.8
				Scenario 2				
	(.01;.40)	(.04;.40)	(.10;.40)	(.25;.40)	(.50;.40)	(.70;.40)		
$\pi = 0$	**30.3**	26.3	20.8	12.5	6.3	3.7	6.0	24.0
$\pi = 1/2$	**33.2**	27.1	21.9	12.2	3.7	1.6	7.2	24.0
				Scenario 3				
	(.01;.25)	(.02;.45)	**(.05;.65)**	(.10;.65)	(.20;.65)	(.30;.65)		
$\pi = 0$	0.6	12.6	**65.8**	18.2	2.8	0.1	2.8	33.0
$\pi = 1/2$	0.9	9.4	**57.7**	24.0	6.8	1.1	3.7	34.8
				Scenario 4				
	(.01;.05)	(.02;.25)	(.05;.45)	**(.10;.70)**	(.25;.70)	(.50;.70)		
$\pi = 0$	0.0	0.5	19.6	**71.1**	8.5	0.3	4.5	30.0
$\pi = 1/2$	0.0	0.9	14.6	**68.0**	15.8	0.7	5.9	33.9
				Scenario 5				
	(.01;.10)	(.05;.35)	**(.15;.60)**	(.20;.60)	(.45;.60)	(.60;.60)		
$\pi = 0$	0.1	6.2	**60.3**	28.9	3.6	0.8	7.9	29.8
$\pi = 1/2$	0.1	6.3	**56.8**	32.8	3.3	0.7	9.2	31.5
				Scenario 6				
	(.01;.05)	(.05;.10)	(.10;.20)	(.20;.35)	**(.30;.55)**	(.50;.55)		
$\pi = 0$	0.4	0.8	3.7	18.9	**53.5**	18.9	11.0	19.4
$\pi = 1/2$	1.4	1.4	4.9	22.2	**57.2**	10.0	13.1	22.2
				Scenario 7				
	(.02;.30)	(.07;.50)	**(.13;.70)**	(.17;.73)	(.25;.76)	(.30;.77)		
$\pi = 0$	0.5	21.4	**60.0**	16.2	1.7	0.0	5.9	34.6
$\pi = 1/2$	0.9	16.4	**53.4**	23.2	5.5	0.6	6.9	37.0
				Scenario 8				
	(.03;.30)	**(.06;.50)**	(.10;.52)	(.20;.54)	(.40;.55)	(.50;.55)		
$\pi = 0$	3.2	**37.8**	34.7	19.1	4.2	1.0	5.6	28.9
$\pi = 1/2$	3.0	**37.8**	34.0	20.1	4.0	1.1	6.9	29.5
				Scenario 9				
	(.01;.30)	(.05;.50)	**(.10;.60)**	(.15;.40)	(.20;.25)	(.25;.15)		
$\pi = 0$	3.0	34.7	**54.7**	5.8	1.3	0.5	4.4	29.7
$\pi = 1/2$	3.8	34.2	**54.6**	6.	1.1	0.2	5.1	29.7
				Scenario 10				
	(.02;.38)	**(.06;.50)**	(.12;.40)	(.30;.30)	(.40;.25)	(.50;.20)		
$\pi = 0$	18.4	**55.9**	15.9	3.9	3.2	2.7	5.7	25.6
$\pi = 1/2$	20.2	**57.1**	17.2	2.8	1.6	1.0	6.3	25.5
				Scenario 11				
	(.03;.25)	(.09;.35)	(.16;.48)	**(.28;.65)**	(.42;.52)	(.56;.39)		
$\pi = 0$	2.2	9.8	30.4	**46.5**	8.6	2.8	10.0	27.4
$\pi = 1/2$	3.5	10.8	31.7	**45.5**	6.3	1.8	11.4	28.8
				Scenario 12				
	(.02;.68)	(.05;.56)	(.07;.49)	(.09;.40)	(.11;.33)	(.13;.26)		
$\pi = 0$	80.1	14.5	3.9	1.1	0.3	0.1	1.8	35.4
$\pi = 1/2$	**74.1**	17.5	5.8	1.8	0.6	0.1	2.2	36.8

References

R Core Team (2015) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: `https://www.R-project.org/`.

Riviere, M.-K., Yuan, Y., Jourdan, J.-H., Dubois, F. and Zohar, S. (2016) Phase i/ii dose-finding design for molecularly targeted agent: Plateau determination using adaptive randomization. *Statistical Methods in Medical Research*, 0962280216631763.

Wages, N. A. and Tait, C. (2015) Seamless phase i/ii adaptive design for oncology trials of molecularly targeted agents. *Journal of Biopharmaceutical Statistics*, **25**, 903–920.

A. Appendix

Table 5. Operating charsteristics of WE, WE(R), MTA and WT design in scenarios 1-5: recommendation proportions, mean number of patients assigned to a dose (in brackets), termination proportion (Term), mean number of toxicity (T) and efficacy (E) respones. The optimal dose is in **bold** and correct doses are underlined. Results are based on 10^4 replications.

	d_1	d_2	d_3	d_4	d_5	d_6	Term	T	E
				Scenario 1					
	(.005;.01)	(.01;.10)	(.02;.30)	(.05;.50)	**(.10;.80)**	(.15;.80)			
WE	0.0	0.1	2.3	37.4	58.8	1.5	0.0	3.1	28.5
	(6.1)	(6.3)	(9.5)	(20.5)	(17.4)	(0.3)			
WE(R)	0.0	0.2	1.0	9.5	72.0	17.3	0.0	4.0	27.4
	(5.1)	(5.0)	(8.2)	(13.8)	(21.7)	(6.2)			
MTA	0.0	0.0	0.7	4.5	57.0	37.1	0.7	5.5	38.0
	(3.3)	(3.7)	(4.8)	(7.7)	(21.5)	(19.0)			
WT	0.0	0.1	0.4	1.3	19.6	78.5	0.1	6.8	41.5
	(3.9)	(1.6)	(2.4)	(3.6)	(11.4)	(37.1)			
				Scenario 2					
	(.01;.40)	(.04;.40)	(.10;.40)	(.25;.40)	(.50;.40)	(.70;.40)			
WE	30.3	26.3	20.8	12.5	6.3	3.7	0.2	6.0	24.0
	(18.3)	(17.0)	(13.7)	(7.9)	(2.7)	(0.4)			
WE(R)	35.0	29.0	21.5	11.5	2.8	0.2	0.1	6.9	24.0
	(15.0)	(15.7)	(15.7)	(10.0)	(3.3)	(0.4)			
MTA	60.2	20.3	8.8	7.3	2.6	0.3	0.6	8.1	24.0
	(18.8)	(13.0)	(10.7)	(10.7)	(5.8)	(0.8)			
WT	41.9	24.5	16.9	14.2	2.4	0.0	0.0	6.7	24.0
	(23.1)	(13.0)	(10.7)	(8.6)	(3.4)	(1.3)			
				Scenario 3					
	(.01;.25)	(.02;.45)	**(.05;.65)**	(.10;.65)	(.20;.65)	(.30;.65)			
WE	0.6	12.6	65.8	18.2	2.8	0.1	0.0	2.8	33.0
	(7.2)	(15.9)	(29.2)	(6.7)	(0.9)	(0.0)			
WE(R)	0.7	6.6	51.0	30.5	10.6	1.2	0.0	4.3	34.5
	(6.4)	(9.6)	(21.3)	(16.1)	(5.8)	(0.9)			
MTA	2.0	14.3	48.4	19.2	9.8	6.4	0.0	6.0	34.6
	(6.3)	(9.8)	(15.5)	(12.9)	(10.4)	(5.1)			
WT	1.4	4.9	29.3	29.9	22.6	11.8	0.0	7.3	35.5
	(6.1)	(5.0)	(13.8)	(14.8)	(12.0)	(8.3)			
				Scenario 4					
	(.01;.05)	(.02;.25)	(.05;.45)	**(.10;.70)**	(.25;.70)	(.50;.70)			
WE	0.0	0.5	19.6	71.1	8.5	0.3	0.0	4.5	30.0
	(6.2)	(7.9)	(17.8)	(25.2)	(2.9)	(0.1)			
WE(R)	0.0	1.3	9.2	69.9	18.5	1.0	0.0	6.0	32.2
	(5.4)	(6.9)	(13.4)	(24.1)	(9.2)	(0.8)			
MTA	0.0	0.7	8.2	53.7	28.6	8.5	0.4	10.0	35.0
	(3.8)	(5.0)	(9.1)	(19.0)	(15.9)	(7.0)			
WT	0.0	0.4	2.1	25.3	61.2	10.8	0.2	13.2	37.0
	(4.5)	(2.4)	(4.0)	(13.5)	(24.7)	(10.9)			
				Scenario 5					
	(.01;.10)	(.05;.35)	**(.15;.60)**	(.20;.60)	(.45;.60)	(.60;.60)			
WE	0.1	6.2	60.3	28.9	3.6	0.8	0.1	7.9	29.8
	(6.4)	(12.3)	(29.4)	(10.5)	(1.3)	(0.1)			
WE(R)	0.1	7.3	54.5	35.4	4.2	0.2	0.3	8.7	29.8
	(6.1)	(12.9)	(23.0)	(14.8)	(3.0)	(0.2)			
MTA	0.0	8.5	55.3	25.3	9.7	1.2	0.1	12.1	29.4
	(5.0)	(8.7)	(18.5)	(15.6)	(10.1)	(2.0)			
WT	0.1	2.7	27.0	53.0	16.9	0.2	0.1	13.5	32.3
	(5.2)	(4.4)	(14.3)	(22.3)	(11.0)	(2.8)			

Table 6. Operating charsteristics of WE, WE(R), MTA and WT design in scenarios 6-10: recommendation proportions, mean number of patients assigned to a dose (in brackets), termination proportion (Term), mean number of toxicity (T) and efficacy (E) respones. The optimal dose is in **bold** and correct doses are underlined. Results are based on 10^4 replications.

	d_1	d_2	d_3	d_4	d_5	d_6	Term	T	E
				Scenario 6					
	(.01;.05)	(.05;.10)	(.10;.20)	(.20;.35)	**(.30;.55)**	(.50;.55)			
WE	0.4	0.8	3.7	18.9	53.5	18.9	4.7	11.0	19.4
	(6.6)	(7.4)	(10.0)	(15.7)	(16.8)	(2.8)			
WE(R)	0.4	0.8	4.8	25.4	56.7	7.1	4.4	10.8	19.2
	(6.3)	(7.7)	(11.1)	(16.3)	(15.2)	(2.9)			
MTA	0.1	0.7	4.5	17.0	55.9	13.7	8.3	13.2	21.4
	(4.5)	(5.5)	(7.9)	(12.4)	(19.0)	(7.8)			
WT	0.2	0.9	3.8	21.4	65.2	5.8	2.7	14.7	24.4
	(5.0)	(3.0)	(5.3)	(13.0)	(25.1)	(7.9)			
				Scenario 7					
	(.02;.30)	(.07;.50)	**(.13;.70)**	(.17;.73)	(.25;.76)	(.30;.77)			
WE	0.5	21.4	60.0	16.2	1.7	0.0	0.0	5.9	34.6
	(8.7)	(20.6)	(25.3)	(5.1)	(0.4)	(0.0)			
WE(R)	0.8	10.9	48.5	31.9	7.2	0.6	0.0	6.9	36.5
	(8.6)	(13.8)	(22.0)	(12.5)	(2.9)	(0.3)			
MTA	1.4	8.7	37.9	24.5	16.4	11.1	0.0	9.6	39.1
	(6.2)	(8.9)	(14.6)	(14.0)	(11.3)	(5.1)			
WT	1.6	5.2	27.1	29.8	24.7	11.7	0.0	10.0	39.9
	(6.7)	(5.4)	(13.7)	(14.8)	(12.1)	(7.3)			
				Scenario 8					
	(.03;.30)	**(.06;.50)**	(.10;.52)	(.20;.54)	(.40;.55)	(.50;.55)			
WE	3.2	37.8	34.7	19.1	4.2	1.0	0.0	5.6	28.9
	(9.4)	(23.9)	(17.8)	(7.1)	(1.5)	(0.1)			
WE(R)	3.9	36.4	33.8	20.9	4.4	0.4	0.0	6.9	29.2
	(9.2)	(16.9)	(18.7)	(11.7)	(3.2)	(0.3)			
MTA	12.8	43.0	21.7	12.7	8.2	1.7	0.1	9.3	29.3
	(10.1)	(14.3)	(12.8)	(12.3)	(8.5)	(2.0)			
WT	7.1	26.1	27.0	28.0	11.3	0.4	0.0	9.1	29.2
	(9.9)	(13.0)	(13.9)	(13.8)	(7.1)	(2.2)			
				Scenario 9					
	(.01;.30)	(.05;.50)	**(.10;.60)**	(.15;.40)	(.20;.25)	(.25;.15)			
WE	3.0	34.7	54.7	5.8	1.3	0.5	0.0	4.4	29.7
	(8.9)	(23.1)	(22.5)	(3.9)	(1.4)	(0.4)			
WE(R)	4.1	31.4	56.7	6.5	1.1	0.1	0.0	5.5	28.4
	(8.5)	(15.6)	(22.0)	(9.1)	(3.6)	(1.2)			
MTA	24.2	54.7	20.3	0.6	0.0	0.0	0.2	5.0	26.3
	(13.2)	(18.3)	(14.1)	(8.5)	(4.7)	(1.2)			
WT	7.3	30.9	50.1	8.8	2.1	0.7	0.2	5.5	27.9
	(9.8)	(15.4)	(22.1)	(7.1)	(3.3)	(2.2)			
				Scenario 10					
	(.02;.38)	**(.06;.50)**	(.12;.40)	(.30;.30)	(.40;.25)	(.50;.20)			
WE	18.4	55.9	15.9	3.9	3.2	2.7	0.1	5.7	25.6
	(14.7)	(27.4)	(11.0)	(4.3)	(2.1)	(0.5)			
WE(R)	22.7	56.2	17.0	2.5	1.2	0.2	0.1	6.7	24.6
	(13.4)	(21.4)	(14.9)	(6.6)	(3.0)	(0.6)			
MTA	60.4	35.3	2.8	0.4	0.1	0.1	1.0	6.3	23.9
	(20.2)	(17.6)	(10.1)	(8.0)	(3.3)	(0.5)			
WT	29.0	49.3	15.8	4.9	0.9	0.1	0.1	5.9	24.8
	(19.5)	(21.6)	(10.5)	(4.8)	(2.3)	(1.2)			

Table 7. Operating charsteristics of WE, WE(R), MTA and WT design in scenarios 11-14: recommendation proportions, mean number of patients assigned to a dose (in brackets), termination proportion (Term), mean number of toxicity (T) and efficacy (E) respones. The optimal dose is in **bold** and correct doses are underlined. Results are based on 10^4 replications.

	d_1	d_2	d_3	d_4	d_5	d_6	Term	T	E
				Scenario 11					
	(.03;.25)	(.09;.35)	(.16;.48)	**(.28;.65)**	(.42;.52)	(.56;.39)			
WE	2.2	9.8	30.4	46.5	8.6	2.8	0.2	10.0	27.4
	(9.3)	(13.5)	(19.4)	(15.1)	(2.4)	(0.3)			
WE(R)	3.6	12.0	30.9	47.9	5.3	0.5	0.2	10.0	27.1
	(10.2)	(13.9)	(18.0)	(14.5)	(3.0)	(0.2)			
MTA	6.7	14.0	27.0	46.0	5.6	0.3	0.3	12.8	28.7
	(8.1)	(10.2)	(14.1)	(17.0)	(8.7)	(1.7)			
WT	6.9	9.4	23.2	56.9	3.5	0.0	0.1	12.2	29.4
	(9.9)	(7.9)	(13.8)	(22.2)	(4.6)	(1.6)			
				Scenario 12					
	(.02;.68)	(.05;.56)	(.07;.49)	(.09;.40)	(.11;.33)	(.13;.26)			
WE	80.1	14.5	3.9	1.1	0.3	0.1	0.0	1.8	35.4
	(44.8)	(10.0)	(3.4)	(1.1)	(0.5)	(0.1)			
WE(R)	70.9	18.7	7.1	2.4	0.9	0.2	0.0	3.3	32.7
	(20.5)	(15.1)	(11.8)	(7.0)	(3.6)	(1.9)			
MTA	96.1	3.2	0.6	0.1	0.0	0.0	0.0	2.6	35.5
	(34.6)	(10.1)	(6.4)	(4.8)	(2.8)	(1.4)			
WT	75.9	17.1	5.1	1.2	0.6	0.2	0.0	2.4	36.3
	(37.0)	(11.0)	(5.4)	(3.0)	(2.1)	(1.4)			
				Scenario 13					
	(.05;.01)	(.10;.02)	(.25;.05)	(.55;.35)	(.70;.55)	(.90;.70)			
WE	0.0	0.0	0.2	2.9	0.6	1.1	95.2	11.7	6.0
	(6.7)	(7.6)	(10.9)	(12.2)	(1.6)	(0.1)			
WE(R)	0.1	0.0	0.2	4.4	1.1	0.2	93.9	13.0	7.1
	(6.9)	(7.7)	(10.6)	(12.1)	(3.2)	(0.4)			
MTA	0.0	0.0	2.3	5.8	0.0	0.0	91.9	11.0	6.2
	(5.8)	(5.9)	(7.7)	(11.0)	(2.7)	(0.3)			
WT	0.0	0.1	5.5	3.1	0.0	0.0	91.5	11.2	5.4
	(6.3)	(6.5)	(14.8)	(7.9)	(1.5)	(1.2)			
				Scenario 14					
	(.50;.40)	(.60;.55)	(.69;.65)	(.76;.65)	(.82;.65)	(.89;.65)			
WE	2.2	0.2	0.1	0.3	0.2	0.0	96.9	11.0	9.1
	(17.6)	(2.8)	(0.6)	(0.1)	(0.0)	(0.0)			
WE(R)	1.9	0.5	0.3	0.2	0.0	0.0	97.4	10.9	8.9
	(17.6)	(2.4)	(0.8)	(0.1)	(0.0)	(0.0)			
MTA	8.8	0.2	0.0	0.0	0.0	0.0	91.0	11.5	9.6
	(16.0)	(4.3)	(1.1)	(0.2)	(0.0)	(0.0)			
WT	9.5	0.0	0.0	0.0	0.0	0.0	90.4	12.1	9.7
	(23.0)	(0.4)	(0.1)	(0.1)	(0.1)	(0.1)			