Risk-Group-Specific Dose Finding Based on an Average Toxicity Score – Supplementary Materials

B. Nebiyou Bekele[1], Yisheng Li[2], Yuan Ji[3],

1. Corresponding author; Department of Biostatistics,

University of Texas M.D. Anderson Cancer Center, Houston, TX 77030.

2. Department of Biostatistics

3. Department of Bioinformatics and Computational Biology

February 17, 2009

1 Computation

1.1 Algorithm

The steps involved in implementing this method may seem somewhat daunting. As such in this section we provide guidance on the required computational steps for implementation of the method for dose-finding:

1) For each dose/risk-group combination, calculate the posterior Dirichlet distribution of $\mathbf{P}_{jh}$ and randomly draw one observation from this distribution.

2) For each dose/risk-group combination, calculate the ATS, $\psi(\mathbf{s}, j, h)$, using the randomly drawn observation from the previous step.

3) Apply the MLSA to obtain partially ordered ATS's. Repeat these steps a large number of times (say, more than 1,000).

1.2 Minimum Lower Sets Algorithm Details

To use this algorithm, we first define a partial order on all dose/risk-group combinations and then obtain the associated lower, upper and level sets, which are required for the implementation of the algorithm. Let $C = \{(j,h) : j = 1,\ldots,J_h,\ h = 1,\ldots,M\}$, the set of all dose/risk-group combinations. Define an order $\preceq$ on C as follows:

$$(j_1,h_1) \preceq (j_2,h_2) \text{ if } j_1 \le j_2 \text{ and } h_1 \le h_2. \tag{1}$$

It is straightforward to verify that C is a partially-ordered set (Robertson et al., 1988). We derive all possible non-empty lower, upper and level sets in C based on the definitions in Robertson et al. (1988). The classes of these sets, denoted as $\mathcal{L}, \mathcal{U}$, and $\mathcal{B}$, respectively, are given below:

1. Lower sets: For $0 \le j_h \le J_h, h = 1,\ldots,M$, define $L_{j_1\ldots j_M} = \{(j,h) : h \in [1,\ldots,M],\ 1 \le j \le j_h\}$. Then $\mathcal{L} = \{L_{j_1\ldots j_M} : 1 \le j_1 \le J_1; \forall h \in [2,\ldots,M],\ 0 \le j_h \le J_h; \forall h \in [1,\ldots,M-1],\ j_h \ge j_{h+1}\}$.

2. Upper sets: For $1 \le j_h \le J_h{+}1, h = 1,\ldots,M$, let $U_{j_1\ldots j_M} = \{(j,h) : h \in [1,\ldots,M],\ j_h \le j \le J_h\}$. Then $\mathcal{U} = \{U_{j_1\ldots j_M} : \forall h \in [1,\ldots,M],\ 1 \le j_h \le J_h + 1$, with at least one $j_h \le J_h; \forall h \in [1,\ldots,M-1],\ j_h \ge j_{h+1}\}$.

3. Level sets: For $1 \le i_h \le J_h + 1$ and $0 \le j_h \le J_h,\ h = 1,\ldots,M$, let $B_{i_1\ldots i_M;j_1\ldots j_M} = \{(j,h) : h \in [1,\ldots,M],\ i_h \le j \le j_h\}$. Then $\mathcal{B} = \{B_{i_1\ldots i_M;j_1\ldots j_M} : \forall h \in [1,\ldots,M], 1 \le i_h \le J_h + 1,\ 0 \le j_h \le J_h,;\ \forall h \in [1,\ldots,M-1],\ i_h \ge i_{h+1}$, and $j_h \ge j_{h+1}$; and there is at least one pair of (i_h, j_h) such that $i_h \le j_h\}$.

We now describe the MLSA for isotonizing the unconstrained posterior samples of the average toxicity scores. We define

$$\mathrm{Av}(B) = \frac{\sum_{(j,h)\in B} \psi(\mathbf{s}, j, h) \times \tau_{j,h}}{\sum_{(j,h)\in B} \tau_{j,h}}, \tag{2}$$

as a weighted average of $\psi(\mathbf{s}, j, h)$ over the nonempty subset $B \subseteq C$, where $\tau_{j,h}$ is the inverse variance of the estimated marginal posterior distribution of the unconstrained $\psi(\mathbf{s}, j, h)$. In implementing the MLSA, we first select a lower set $L_1 \subseteq C$ such that $\mathrm{Av}(L_1) \leq \mathrm{Av}(L)$ for all $L \in \mathcal{L}$, i.e., L_1 is a lower set of minimum average relative to all lower sets in $\mathcal{L}$. Once we obtain this lower set of minimum average, we then proceed to derive the largest (with respect to number of elements) lower set with this same minimum average. It can be shown that the union of all lower sets of minimum average is the largest lower set of minimum average. Let L_1' denote this lower set. Also, $L_1' \equiv B_1$ (i.e., L_1' is both a lower and level set). This level set is the set on which $\psi^*(\mathbf{s}, j, h)$ as a function on C assumes its smallest value:

$$\psi^*(\mathbf{s}, j, h) = \mathrm{Av}(B_1) = \min\{\mathrm{Av}(L) : \ L \in \mathcal{L}\}, \ (j,h) \in B_1.$$

Now consider the average of level sets of the form $L \cap L_1^c$ − the level sets consisting of lower sets with L_1 omitted. Select again the largest among these level sets of minimum average and denote it by $B_2 = L_2 \cap L_1^c$. Thus, B_2 is the set on which $\psi^*(\mathbf{s}, j, h)$ assumes its next smallest value:

$$\psi^*(\mathbf{s}, j, h) = \mathrm{Av}(B_2), \ (j,h) \in B_2.$$

The process is continued until C is exhausted, and the isotonic regression $\psi^*(\mathbf{s}, j, h)$ is obtained.

1.3 Target average toxicity score

Similar to a phase I dose-finding design in which the investigators define a target toxicity rate, when using our proposed design one must define a target ATS, denoted by ψ_T, and attempt to find the highest dose within each risk group with an ATS closest to the target ψ_T. One elicits ψ_T by having the PIs evaluate a set of hypothetical toxicity outcomes. Let m be the total number of hypothetical cohorts of size c. The outcomes of these m hypothetical cohorts may be denoted as $\mathbf{y}_1^* = (y_{1,1}^*, \ldots, y_{1,c}^*), \ldots, \mathbf{y}_m^* = (y_{m,1}^*, \ldots, y_{m,c}^*)$, where $y_{i,j}^*$ (for $i = 1, \ldots m$, $j = 1, \ldots, c$) denotes the toxicity score of the jth hypothetical patient in the ith hypothetical cohort. For each of these outcomes the PIs provided a judgment on whether observing that outcome would result in *escalating* to the next higher dose, *staying* at the current dose, or *de-escalating* to an appropriate lower dose.

Given the process described above, we let V denote the set of cohorts in which the PIs' decision is to stay. For a cohort $v \in V$, we calculated the quantity

$$\overline{\mathrm{TS}}_v \;=\; \frac{1}{c} \sum_{j=1}^{c} \sum_{k=1}^{K} s_k I(y_{v,j}^* = y_k) \;.$$

The target ATS, ψ_T, can be defined as an appropriate function of the elicited $\overline{\mathrm{TS}}_v^{\,*}$ values for $v \in V$ (e.g., mean, median, min, max). For the MM trial, we used

$$\psi_T \;= \min_{v \in V} (\overline{\mathrm{TS}}_v^{\,*}).$$

2 Additional simulation results

Tables 1-16 present additional simulation results evaluating sensitivity to the prior (Tables 1-4), changes in the spacing of the elicited toxicity weights (Table 5), a set of cases in which the elicited toxicity scores are restricted to take on values of either 0 or 1 (equivalent to

the binary toxicity case) (Tables 6-8), a set of 3 additional simulation scenarios comparing the proposed method to the modified CRM method (Tables 9-10), scenarios comparing the proposed method and the M-CRM to a case in which the middle and lower risk groups are allowed to escalate to higher doses (Tables 11-12), cases in which the posterior mean toxicity score is isotonized rather than incorporating the isotonic regression as part of the Monte Carlo Simulation (Tables 13-14) and lastly, cases in which concurrent phase I trials in each risk group are conducted using our proposed Dirichlet model and a non-isotonized version of the Average Toxicity Scores (Tables 15-16).

2.1 Sensitivity to Prior

The prior used in the manuscript follows a Dirichlet distribution at the jth dose of risk group h, with the parameters given by $\boldsymbol{\alpha}_{jh} = (0.604, 0.178, 0.089, 0.071, 0.058)$. This results in an expected toxicity score of 0.20. This prior is a unit prior since the Dirichlet parameters sum to 1. Let us call this prior the *base* prior. We performed additional simulation studies examining the impact of the prior on the ultimate conclusions. First we define an *informative* prior as a Dirichlet distribution with parameters $4 \times \boldsymbol{\alpha}_{jh}$, and a *non-informative* prior as a Dirichlet with $0.25 \times \boldsymbol{\alpha}_{jh}$. The expected toxicity score is 0.20 for all three priors. The difference between the three priors is that the *base* prior assumes one patient's worth of information at each dose/risk-group combination, the *informative* prior assumes four patients worth of information, and the *non-informative* assumes prior information equivalent to that of one quarter of a patient. From our new simulation results (Tables 1-2) using each of the three priors, we observed that the escalation proceeds from least aggressive to most aggressive in the order of *non-informative*, *base*, and *informative* priors, respectively. The aggressive escalation for the *informative* prior is because *a priori* each dose/risk-group combination is

assumed safe with a mean prior ATS 0.20 and a relatively large prior effective sample size of four patients. As the prior belief weakens, the escalation associated with the *base* and *non-informative* priors becomes less aggressive.

Lastly, we present two additional priors demonstrating the effects of calibrating the *base* prior to be more *aggressive* or more *conservative*. The *aggressive* and *conservative* priors assume the same effective sample size as the *base* prior. For the *aggressive* prior the Dirichlet parameters were set to $(0.628, 0.322, 0.025, 0.0125, 0.0125)$, resulting in an expected toxicity score of 0.1125. As expected this prior more aggressively escalates than the *base* since it assumes that the doses are safer *a priori*. For the *conservative* prior the hyper-parameters were set to $(0.444, 0.222, 0.112, 0.111, 0.111)$, corresponding to an expected toxicity score 0.30. As expected this prior is more conservative (less aggressive in dose escalation) than the *base* prior and makes it hard to escalate even when the observed data indicate a dose is safe. Therefore, it has a tendency to "stick" to lower non-toxic doses.

2.2 Effects of Changes in the Spacing of the Elicited Toxicity Weights

We conducted an additional simulation study to explore the impact of changes the elicited toxicity scores. In this study we changed the toxicity score s_2 from 0.50 to 0.425 and s_3 from 0.75 to 0.85. These changes resulted in similar true ATS's and the same target ATS as was originally elicited from the investigator. Compare Table 1 of the manuscript and Table 5 of this document to see the effects of these changes on the true ATS's. As expected, the simulation results based on these two sets of elicited toxicity scores are similar. These simulation results suggest that as long as the true and Target ATS's are similar to there original counterparts, the conclusions will not be sensitive to the toxicity scores.

2.3 Comparison to Binary Toxicity Cases

Interestingly, we note that binary toxicities are a special case of our general method. For example, we obtain a binary toxicity by assigning an elicited toxicity score of 0 for the No toxicity category and a 1 to all other toxicities (the example given by the reviewer). The elicited toxicity scores for the binary case given above is numerically expressed as $\boldsymbol{s} = (s_0, ..., s_4) = (0, 1, 1, 1, 1)$. Since we defined $\psi(\mathbf{s}, j, h) = \sum_{k=1}^{K} I(s_k = 1) \cdot p_{kjh}$, the average toxicity score is exactly equal to the sum of the probabilities of observing each category of toxicity. Based on this connection between the binary case and our method, we have now added three additional simulation studies. In these simulation studies we assumed that Binary toxicity is defined as:

- $\boldsymbol{s} = (0, 1, 1, 1, 1)$;
- $\boldsymbol{s} = (0, 0, 1, 1, 1)$; and
- $\boldsymbol{s} = (0, 0, 0, 1, 1)$.

Note that for each of these new toxicity scores it is necessary to recalibrate the Dirichlet prior so that the new prior corresponds to an expected prior ATS of approximately 0.25. For example, if we used the *base* prior we would have an expected prior ATS of 0.397 for the case in which we define $\boldsymbol{s} = (0, 1, 1, 1, 1)$. This would be extremely conservative and lead to almost no escalation.

Therefore, for $\boldsymbol{s} = (0, 1, 1, 1, 1)$ we defined an alternative Dirichlet prior with parameters $\boldsymbol{\alpha}_{jh} = (0.757, 0.061, 0.061, 0.061, 0.061)$. This prior has a more reasonable expected prior ATS of 0.24. Similarly, for $\boldsymbol{s} = (0, 0, 1, 1, 1)$ we defined the Dirichlet prior with parameters $\boldsymbol{\alpha}_{jh} = (0.607, 0.143, 0.083, 0.083, 0.083)$. This prior has an expected prior ATS of 0.25. Lastly for $\boldsymbol{s} = (0, 0, 0, 1, 1)$ we used a Dirichlet prior with parameters $\boldsymbol{\alpha}_{jh} = (0.51, 0.12, 0.12, 0.125, 0.125)$. Again, for this case the expected prior ATS is 0.25.

Each of the binary toxicity definitions results in very different operating characteristics when compared to the original setup in the paper where $\boldsymbol{s} = (0, .25, .50, .75, 1)$. To see why this is the case note that changing the elicited scores automatically changes the true ATS's of the doses in any given scenario. We can see how each of these toxicity score configurations affects the true ATS's by comparing the true ATS's in the manuscript to the ATS's in Tables 6-8 of this document. Clearly, the ATS's are very different in each case and consequently affect the simulation results. These results show that when a clinician wants to characterize toxicity by differentiating between classes of toxicity, defining a binary toxicity results in trial conduct and simulation performance that are dissimilar from those under the ordinal framework. Thus, the ordinal toxicity framework may better capture the investigators preferences regarding toxicity severity.

2.4 Additional Simulation Scenarios

We have now added 3 additional scenarios (Tables 9-10) to show the reader how the method performs under other scenarios. We compared the proposed method to the modified CRM method. As in the original scenarios presented in the manuscript, in these 3 scenarios the proposed method outperforms the M-CRM method.

2.5 Simulation Scenarios Under Varying Assumptions

We ran a set of scenarios in which we removed the restriction which did not allow patients in the highest and middle risk groups to be assigned to the higher doses (Table 10-12). Again as with in the original scenarios presented in the manuscript, the proposed method outperformed the M-CRM method.

2.6 Modified ATS methods

We ran a set of scenarios in which:

1) the posterior mean toxicity score is isotonized rather than incorporating the isotonic regression as part of the Monte Carlo Simulation (Table 13-14). For Table 13 we use the same prior as that used in the *base* prior. Because the results in Table 13 seemed slightly aggressive (especially for scenario 6). We also ran a case (Table 14) we use a less informative and slightly more conservative prior ($\boldsymbol{\alpha}_{jh} = (0.125, 0.0613, 0.022, 0.022, 0.019)$).; Based on these two simulation studies, we obtained slightly less desirable operating characteristics when compared to the original method summarized in Table 1 of the manuscript. This suggests that there may be a disadvantage to transforming the unconstrained posterior means in small sample settings such as phase I/II oncology trials (the reason for this is unclear but we observed the same phenomena in unpublished simulation studies related to Li et al. [2008]). Furthermore, unlike our currently proposed decision rules, rules based on the posterior mean isotonization do not take into account the variability in the parameter estimates. Thus, there is no natural way to set up decision rules for early stopping due to excessive toxicity (for these simulations we used the posterior probability calculated from the non-isotonized version of the ATS).

2) concurrent phase I trials in each risk group are run using our proposed Dirichlet model and Average Toxicity Score (Tables 15-16). For this third set of simulations, we based dose escalation/de-escalation on the non-isotonized version ATS. At the end of the study we then ran the MLSA to make final dose selections. For Table 15 we use the same prior as that used in the *base* prior. Because the results in Table 15 seemed aggressive (especially for scenario 6). We also ran a case (Table 16) we use a less informative and slightly more conservative prior ($\boldsymbol{\alpha}_{jh} = (0.125, 0.0613, 0.022, 0.022, 0.019)$). While this simple version of the method works relatively well in some scenarios (Table 16), this method allowed patients with higher

risk profiles to be assigned higher doses than patients with lower risk profiles. Although this event would not result in inappropriate final dose selection, as a practical matter, such an outcome (even transient) could not be tolerated from a patient safety perspective. In addition, when no doses are safe within a given risk group (scenarios 5 and 6) the simple method tends to enroll more patients with higher risk profiles.

2.7 Minimum Lower Sets Algorithm Computer Program

The purpose of the two R functions below are to provide the reader with code for calculating matrix ordered transformed variates using the Minimum Lower Sets Algorithm (Robertson et al., 1988). The first function is the main program and the second function is called by the main program. This code can be used for up to a maximum of three risk groups. It can be easily modified if the investigator desires to evaluate more than three risk groups.

```
# MAIN MLSA Program
# pstr.mu:      vector of one unconstrained posterior sample;
# sampvar1:     vector of posterior variance of unconstrained posterior
#               samples value of a component is set to arbitrarily large if
#               the corresponding dose/risk-group combination is untried;
# h:            vector of total number of doses across risk groups here
#               h[i] are all equal;
# pstr.transmu: vector of isotonically transformed posterior sample;

function mlsa(pstr.mu, sampvar1, h, pstr.transmu) {

    L1l <- 1
    L2l <- 1
    L3l <- 1
    minL1 <- h[1]-1
    minL2 <- h[2]-1
    minL3 <- h[3]-1
    while ((minL1 < h[1]) | (minL2 < h[2]) | (minL3 < h[3])) {
```

```
    count <- 0
    for (L1r in (L1l-1) : h[1]) {
         for (L2r in (L2l-1) : min(L1r,h[2])) {
             for (L3r in (L3l-1):min(L2r,h[3])) {
                 if ((L1r > L1l-1) | (L2r > L2l-1) | (L3r > L3l-1)) {
                    cur.avg <- 0.d0
                    cur.avg <- avg.multi.grp(pstr.mu, sampvar1,L1l, L1r,
                               L2l, L2r, L3l, L3r)
                    if (count == 0) {
                        min.avg <- cur.avg
                        minL1 <- L1r
                        minL2 <- L2r
                        minL3 <- L3r
                        count <- 1
                    }
                    else if (cur.avg <= min.avg) {
                        min.avg <- cur.avg
                        minL1 <- L1r
                        minL2 <- L2r
                        minL3 <- L3r
                   }
               }
           }
       }
    }

    for (i in L1l: minL1) {
        pstr.transmu[1,i] <- min.avg
    }
    for (j in L2l: minL2) {
        pstr.transmu[2,j] <- min.avg
    }
    for (k in L3l: minL3) {
       pstr.transmu[3,k] <- min.avg
    }
    L1l <- minL1 + 1
    L2l <- minL2 + 1
    L3l <- minL3 + 1

   }

list(pstr.transmu)
```

```
}



# This function is used by the main MLSA program to calculate the
#weighted averages used for isotonization;

# pstr.mu:      vector of one unconstrained posterior sample;
# sampvar1:     vector of posterior variance of unconstrained
#               posterior samples value of a component is set to
#               arbitrarily large if the corresponding dose/risk-group
#               combination is untried;
# (Li,Ui), i = 1,2,3: lower and upper bounds of intervals of doses across
#                     risk groups as part of the lower set specification at
#                     least one Li < Ui to be expected;

function avg.multi.grp(pstr.mu, sampvar1, L1, U1, L2, U2, L3, U3) {

        sum.mu1 <- 0
        sum.mu2 <- 0
        sum.mu3 <- 0
        sum.w1  <- 0
        sum.w2  <- 0
        sum.w3  <- 0

        if (L1 <= U1) {
           for (I in L1 :U1){
                sum.mu1 <- sum.mu1+pstr.mu[1,i]/sampvar1[1,i]
                sum.w1  <-sum.w1+1/sampvar1[1,i]
           }
        }

        if (L2 <= U2) {
            for (I in L2:U2){
                sum.mu2 <- sum.mu2+ pstr.mu[2,i]/sampvar1[2,i]
                sum.w2  <- sum.w2+ 1/sampvar1[2,i]
            }
        }

        if (L3 <= U3) {
            for (I in L3:U3) {
                sum.mu3 <- sum.mu3+pstr.mu[3,i]/sampvar1[3,i]
```

```
            sum.w3  <- sum.w3+1/sampvar1[3,i]
        }
    }

    weightedAvg <- (sum.mu1 + sum.mu2 + sum.mu3)/(sum.w1 + sum.w2 + sum.w3)

    list(weightedAvg)

}
```

REFERENCES

Li Y, Bekele BN, Ji Y, Cook JD. (2008) Dose–schedule finding in phase I/II clinical trials using a Bayesian isotonic transformation. *Statistics in Medicine* **27**,4895-913.

Robertson, T., Wright, F.T. and Dykstra, R. (1988) *Order restricted statistical inference.* 12-26. New York: Wiley

Table 1: Simulation results for proposed design assuming an informative prior. For each scenario we provide the true ATS by risk group (italicized numbers in table) and report the correct selection probabilities and the average number of patients per dose for each risk group.

Selection percentage for each dose/risk-group combination (average # of patients)						
		Dose				
Scen	Risk Grp	1	2	3	4	None
1	1	*0.07*	*0.11*	*0.18*	*0.25*	0
		0.0 (3.0)	**0.8 (3.2)**	**21.3 (5.2)**	**77.9 (9.5)**	
	2	*0.11*	*0.18*	*0.25*	—	0
		1.1 (3.4)	**23.7 (5.3)**	**75.2 (9.3)**	— (—)	
	3	*0.14*	*0.21*	—	—	2.8
		18.1 (5.1)	**79.1 (6.7)**	—(—)	—(—)	
2	1	*0.01*	*0.01*	*0.03*	*0.38*	0
		0.0 (3.0)	**0.0 (3.0)**	**54.7 (5.4)**	**45.3 (9.6)**	
	2	*0.01*	*0.01*	*0.38*	—	0
		0.0 (3.0)	**54.3 (5.5)**	**45.7 (9.5)**	— (—)	
	3	*0.01*	*0.38*	—	—	0
		43.5 (4.3)	**56.5 (7.7)**	—(—)	—(—)	
3	1	*0.08*	*0.1*	*0.25*	*0.35*	0
		0.0 (3.0)	**7.7 (3.8)**	**58.3 (8.3)**	**34.0 (5.9)**	
	2	*0.1*	*0.21*	*0.35*	—	0.2
		1.2 (3.3)	**63.4 (7.5)**	**35.2 (7.1)**	— (—)	
	3	*0.1*	*0.25*	—	—	0.8
		19.6 (4.8)	**79.6 (7.1)**	—(—)	—(—)	
4	1	*0.06*	*0.17*	*0.24*	*0.31*	0
		0.7 (3.1)	**10.0 (4.8)**	**46.3 (7.3)**	**43.0 (5.7)**	
	2	*0.17*	*0.27*	*0.31*	—	1.9
		18.7 (5.4)	**48.5 (7.7)**	**30.9 (4.6)**	— (—)	
	3	*0.25*	*0.31*	—	—	17.5
		47.5 (7.3)	**35.0 (3.7)**	—(—)	—(—)	
5	1	*0.21*	*0.38*	*0.4*	*0.4*	1.6
		58.7 (9.0)	**32.3 (9.8)**	**6.2 (1.7)**	**1.2 (0.3)**	
	2	*0.24*	*0.38*	*0.4*	—	14.3
		74.7 (11.8)	**10.5 (4.3)**	**0.5 (0.3)**	— (—)	
	3	*0.36*	*0.38*	—	—	66.8
		31.9 (7.3)	**1.3 (0.7)**	—(—)	—(—)	
6	1	*0.47*	*0.49*	*0.51*	*0.52*	91
		8.2 (8.5)	**0.7 (1.2)**	**0.1 (0.1)**	**0.0 (0.0)**	
	2	*0.48*	*0.5*	*0.52*	—	99.6
		0.3 (4.7)	**0.1 (0.3)**	**0.0 (0.0)**	— (—)	
	3	*0.49*	*0.51*	—	—	99.9
		0.1 (3.0)	**0.0 (0.1)**	—(—)	—(—)	

Table 2: Simulation results for proposed design assuming an non-informative prior. For each scenario we provide the true ATS by risk group (italicized numbers in table) and report the correct selection probabilities and the average number of patients per dose for each risk group.

Selection percentage for each dose/risk-group combination (average # of patients)						
		Dose				
Scen	Risk Grp	1	2	3	4	None
1	1	*0.07*	*0.11*	*0.18*	*0.25*	0
		0.3 (3.1)	**6.1 (3.9)**	**29.4 (5.7)**	**64.2 (8.4)**	
	2	*0.11*	*0.18*	*0.25*	—	0.8
		8.7 (4.4)	**36.4 (6.4)**	**54.1 (7.1)**	— (—)	
	3	*0.14*	*0.21*	—	—	5.4
		33.6 (6.0)	**61.0 (5.5)**	—(—)	—(—)	
2	1	*0.01*	*0.01*	*0.03*	*0.38*	0
		0.0 (3.0)	**0.0 (3.0)**	**75.9 (8.4)**	**24.1 (6.6)**	
	2	*0.01*	*0.01*	*0.38*	—	0
		0.1 (3.0)	**77.5 (9.0)**	**22.4 (6.0)**	— (—)	
	3	*0.01*	*0.38*	—	—	0
		75.6 (7.2)	**24.4 (4.8)**	—(—)	—(—)	
3	1	*0.08*	*0.1*	*0.25*	*0.35*	0.1
		0.9 (3.2)	**20.1 (5.2)**	**52.9 (7.9)**	**26.0 (4.7)**	
	2	*0.1*	*0.21*	*0.35*	—	1.1
		11.7 (4.6)	**65.0 (8.4)**	**22.2 (4.8)**	— (—)	
	3	*0.1*	*0.25*	—	—	3.6
		39.3 (6.0)	**57.1 (5.7)**	—(—)	—(—)	
4	1	*0.06*	*0.17*	*0.24*	*0.31*	0.2
		3.5 (3.6)	**22.4 (6.2)**	**46.9 (7.0)**	**27.0 (4.2)**	
	2	*0.17*	*0.27*	*0.31*	—	4.8
		35.9 (7.8)	**43.5 (6.8)**	**15.8 (2.8)**	— (—)	
	3	*0.25*	*0.31*	—	—	24.2
		60.8 (8.0)	**15.0 (2.3)**	—(—)	—(—)	
5	1	*0.21*	*0.38*	*0.4*	*0.4*	4.8
		72.6 (12.2)	**19.2 (6.8)**	**3.0 (1.1)**	**0.4 (0.2)**	
	2	*0.24*	*0.38*	*0.4*	—	27.5
		68.5 (11.9)	**3.5 (2.3)**	**0.5 (0.3)**	— (—)	
	3	*0.36*	*0.38*	—	—	76.9
		22.8 (6.0)	**0.3 (0.4)**	—(—)	—(—)	
6	1	*0.47*	*0.49*	*0.51*	*0.52*	94.2
		5.3 (6.9)	**0.5 (0.9)**	**0.0 (0.1)**	**0.0 (0.0)**	
	2	*0.48*	*0.5*	*0.52*	—	99.5
		0.5 (3.7)	**0.0 (0.3)**	**0.0 (0.0)**	— (—)	
	3	*0.49*	*0.51*	—	—	100
		0.0 (2.6)	**0.0 (0.1)**	—(—)	—(—)	

Table 3: Simulation results for proposed design assuming an aggressive prior. For each scenario we provide the true ATS by risk group (italicized numbers in table) and report the correct selection probabilities and the average number of patients per dose for each risk group.

Selection percentage for each dose/risk-group combination (average # of patients)						
		Dose				
Scen	Risk Grp	1	2	3	4	None
1	1	*0.07*	*0.11*	*0.18*	*0.25*	0
		0.0 (3.0)	**1.6 (3.3)**	**24.5 (5.1)**	**73.9 (9.6)**	
	2	*0.11*	*0.18*	*0.25*	—	0
		2.5 (3.5)	**25.9 (5.2)**	**71.6 (9.4)**	— (—)	
	3	*0.14*	*0.21*	—	—	1.5
		18.4 (4.9)	**80.1 (7.0)**	—(—)	—(—)	
2	1	*0.01*	*0.01*	*0.03*	*0.38*	0
		0.0 (3.0)	**0.0 (3.0)**	**66.0 (6.8)**	**34.0 (8.2)**	
	2	*0.01*	*0.01*	*0.38*	—	0
		0.0 (3.0)	**64.7 (7.0)**	**35.3 (8.0)**	— (—)	
	3	*0.01*	*0.38*	—	—	0
		56.8 (5.4)	**43.2 (6.6)**	—(—)	—(—)	
3	1	*0.08*	*0.1*	*0.25*	*0.35*	0
		0.1 (3.0)	**10.9 (4.1)**	**55.4 (7.5)**	**33.6 (6.3)**	
	2	*0.1*	*0.21*	*0.35*	—	0
		4.5 (3.6)	**58.4 (7.4)**	**37.1 (7.0)**	— (—)	
	3	*0.1*	*0.25*	—	—	0.2
		25.0 (4.7)	**74.8 (7.2)**	—(—)	—(—)	
4	1	*0.06*	*0.17*	*0.24*	*0.31*	0
		1.3 (3.2)	**14.6 (5.0)**	**40.2 (6.7)**	**43.9 (6.1)**	
	2	*0.17*	*0.27*	*0.31*	—	0.4
		22.3 (5.8)	**48.1 (7.4)**	**29.2 (4.7)**	— (—)	
	3	*0.25*	*0.31*	—	—	12.9
		55.3 (7.6)	**31.8 (3.7)**	—(—)	—(—)	
5	1	*0.21*	*0.38*	*0.4*	*0.4*	0.4
		64.0 (9.7)	**26.5 (8.4)**	**7.1 (2.2)**	**2.0 (0.6)**	
	2	*0.24*	*0.38*	*0.4*	—	9.7
		79.7 (12.5)	**9.6 (4.1)**	**1.0 (0.6)**	— (—)	
	3	*0.36*	*0.38*	—	—	58.6
		39.5 (8.0)	**1.9 (1.0)**	—(—)	—(—)	
6	1	*0.47*	*0.49*	*0.51*	*0.52*	88.1
		11.0 (9.1)	**0.7 (1.7)**	**0.2 (0.2)**	**0.0 (0.0)**	
	2	*0.48*	*0.5*	*0.52*	—	98.9
		1.0 (5.6)	**0.1 (0.5)**	**0.0 (0.0)**	— (—)	
	3	*0.49*	*0.51*	—	—	99.9
		0.1 (3.7)	**0.0 (0.1)**	—(—)	—(—)	

Table 4: Simulation results for proposed design assuming an conservative prior. For each scenario we provide the true ATS by risk group (italicized numbers in table) and report the correct selection probabilities and the average number of patients per dose for each risk group.

Selection percentage for each dose/risk-group combination (average # of patients)						
		Dose				
Scen	Risk Grp	1	2	3	4	None
1	1	*0.07*	*0.11*	*0.18*	*0.25*	0
		0.1 (3.1)	**5.5 (4.0)**	**35.0 (6.4)**	**59.4 (7.6)**	
	2	*0.11*	*0.18*	*0.25*	—	0.1
		6.0 (4.2)	**43.1 (7.0)**	**50.8 (6.8)**	— (—)	
	3	*0.14*	*0.21*	—	—	4.2
		32.4 (6.0)	**63.4 (5.7)**	—(—)	—(—)	
2	1	*0.01*	*0.01*	*0.03*	*0.38*	0
		0.0 (3.0)	**0.0 (3.0)**	**75.0 (8.1)**	**25.0 (6.9)**	
	2	*0.01*	*0.01*	*0.38*	—	0
		0.0 (3.0)	**74.1 (8.2)**	**25.9 (6.8)**	— (—)	
	3	*0.01*	*0.38*	—	—	0
		66.9 (6.2)	**33.1 (5.8)**	—(—)	—(—)	
3	1	*0.08*	*0.1*	*0.25*	*0.35*	0.1
		0.6 (3.2)	**21.7 (5.5)**	**57.1 (8.3)**	**20.5 (4.1)**	
	2	*0.1*	*0.21*	*0.35*	—	0.2
		9.9 (4.5)	**71.2 (9.3)**	**18.7 (4.2)**	— (—)	
	3	*0.1*	*0.25*	—	—	0.7
		46.3 (6.4)	**53.0 (5.5)**	—(—)	—(—)	
4	1	*0.06*	*0.17*	*0.24*	*0.31*	0
		4.0 (3.7)	**23.8 (6.6)**	**47.6 (7.3)**	**24.6 (3.4)**	
	2	*0.17*	*0.27*	*0.31*	—	1.5
		38.9 (8.2)	**44.1 (7.1)**	**15.5 (2.5)**	— (—)	
	3	*0.25*	*0.31*	—	—	21.2
		63.5 (8.5)	**15.3 (2.2)**	—(—)	—(—)	
5	1	*0.21*	*0.38*	*0.4*	*0.4*	2.4
		77.5 (13.0)	**17.4 (6.7)**	**2.4 (0.8)**	**0.3 (0.1)**	
	2	*0.24*	*0.38*	*0.4*	—	17.3
		79.8 (14.2)	**2.8 (1.9)**	**0.1 (0.0)**	— (—)	
	3	*0.36*	*0.38*	—	—	73.4
		26.6 (7.2)	**0.0 (0.2)**	—(—)	—(—)	
6	1	*0.47*	*0.49*	*0.51*	*0.52*	94.9
		5.0 (7.9)	**0.1 (0.4)**	**0.0 (0.0)**	**0.0 (0.0)**	
	2	*0.48*	*0.5*	*0.52*	—	99.8
		0.2 (4.1)	**0.0 (0.1)**	**0.0 (0.0)**	— (—)	
	3	*0.49*	*0.51*	—	—	100
		0.0 (2.7)	**0.0 (0.0)**	—(—)	—(—)	

Table 5: Simulation results for proposed design in which the spacing of the elicited toxicity weights were modified so as to not be equally spaced. The new weights are $(0.25, 0.425, 0.85, 1)$. Note that these new weights result in slightly modified ATS. For each scenario we provide the true ATS by risk group (italicized numbers in table) and report the correct selection probabilities and the average number of patients per dose for each risk group.

Selection percentage for each dose/risk-group combination (average # of patients)						
		Dose				
Scen	Risk Grp	1	2	3	4	None
1	1	*0.07*	*0.1*	*0.18*	*0.25*	0
		0.0 (3.0)	**3.3 (3.5)**	**25.4 (5.4)**	**71.3 (9.1)**	
	2	*0.1*	*0.18*	*0.25*	—	0.1
		2.7 (3.6)	**32.9 (6.0)**	**64.3 (8.4)**	— (—)	
	3	*0.14*	*0.21*	—	—	3
		26.3 (5.4)	**70.7 (6.4)**	—(—)	—(—)	
2	1	*0*	*0.01*	*0.02*	*0.37*	0
		0.0 (3.0)	**0.0 (3.0)**	**66.9 (7.2)**	**33.1 (7.8)**	
	2	*0*	*0.01*	*0.37*	—	0
		0.0 (3.0)	**64.9 (7.2)**	**35.1 (7.8)**	— (—)	
	3	*0*	*0.37*	—	—	0
		56.8 (5.5)	**43.2 (6.5)**	—(—)	—(—)	
3	1	*0.08*	*0.1*	*0.25*	*0.36*	0
		0.2 (3.1)	**15.6 (4.6)**	**57.7 (8.0)**	**26.5 (5.3)**	
	2	*0.1*	*0.21*	*0.36*	—	0
		7.0 (3.9)	**65.5 (8.3)**	**27.5 (5.8)**	— (—)	
	3	*0.1*	*0.25*	—	—	0.3
		32.8 (5.4)	**66.9 (6.6)**	—(—)	—(—)	
4	1	*0.05*	*0.16*	*0.23*	*0.3*	0
		1.5 (3.3)	**14.5 (5.3)**	**40.9 (6.6)**	**43.1 (5.9)**	
	2	*0.16*	*0.26*	*0.3*	—	1
		24.4 (6.1)	**47.0 (7.5)**	**27.6 (4.3)**	— (—)	
	3	*0.25*	*0.3*	—	—	15.1
		55.2 (7.7)	**29.7 (3.4)**	—(—)	—(—)	
5	1	*0.21*	*0.37*	*0.4*	*0.4*	0.8
		65.3 (10.6)	**26.4 (8.2)**	**6.0 (1.7)**	**1.5 (0.4)**	
	2	*0.23*	*0.37*	*0.4*	—	7
		81.6 (12.7)	**10.4 (4.0)**	**1.0 (0.5)**	— (—)	
	3	*0.35*	*0.37*	—	—	55.5
		42.9 (7.9)	**1.6 (0.8)**	—(—)	—(—)	
6	1	*0.47*	*0.47*	*0.5*	*0.52*	88.6
		10.6 (9.4)	**0.8 (1.0)**	**0.0 (0.1)**	**0.0 (0.0)**	
	2	*0.48*	*0.48*	*0.51*	—	98.9
		1.0 (5.2)	**0.1 (0.2)**	**0.0 (0.0)**	— (—)	
	3	*0.49*	*0.49*	—	—	99.7
		0.3 (3.3)	**0.0 (0.1)**	—(—)	—(—)	

Table 6: Simulation results for proposed design assuming $\boldsymbol{s} = (0, 1, 1, 1, 1)$. For each scenario we provide the true ATS by risk group (italicized numbers in table) and report the correct selection probabilities and the average number of patients per dose for each risk group.

Selection percentage for each dose/risk-group combination (average # of patients)						
		Dose				
Scen	Risk Grp	1	2	3	4	None
1	1	*0.17*	*0.25*	*0.38*	*0.44*	0.3
		15.0 (5.2)	**49.5 (8.7)**	**26.6 (5.5)**	**8.6 (1.6)**	
	2	*0.25*	*0.38*	*0.44*	—	6.3
		60.4 (10.1)	**30.4 (6.2)**	**2.9 (1.1)**	— (—)	
	3	*0.25*	*0.44*	—	—	24
		69.2 (8.6)	**6.8 (2.0)**	—(—)	—(—)	
2	1	*0.01*	*0.02*	*0.05*	*0.76*	0
		0.0 (3.0)	**0.0 (3.0)**	**98.1 (10.6)**	**1.9 (4.3)**	
	2	*0.01*	*0.02*	*0.76*	—	0
		0.0 (3.0)	**98.5 (11.0)**	**1.5 (3.9)**	— (—)	
	3	*0.01*	*0.76*	—	—	0
		96.6 (8.3)	**3.4 (3.7)**	—(—)	—(—)	
3	1	*0.22*	*0.21*	*0.5*	*0.6*	0.6
		12.2 (4.9)	**72.0 (10.4)**	**14.1 (5.2)**	**1.1 (0.5)**	
	2	*0.21*	*0.44*	*0.6*	—	3.8
		69.2 (10.0)	**26.2 (6.7)**	**0.8 (0.7)**	— (—)	
	3	*0.21*	*0.44*	—	—	13.5
		80.1 (8.8)	**6.4 (2.3)**	—(—)	—(—)	
4	1	*0.12*	*0.36*	*0.48*	*0.65*	0.6
		42.5 (8.3)	**46.5 (9.7)**	**9.6 (2.7)**	**0.8 (0.2)**	
	2	*0.36*	*0.48*	*0.65*	—	32.5
		62.2 (12.6)	**5.3 (2.5)**	**0.0 (0.2)**	— (—)	
	3	*0.44*	*0.65*	—	—	78.3
		21.6 (7.0)	**0.1 (0.4)**	—(—)	—(—)	
5	1	*0.44*	*0.76*	*0.76*	*0.76*	65.3
		34.5 (13.3)	**0.2 (1.3)**	**0.0 (0.0)**	**0.0 (0.0)**	
	2	*0.57*	*0.76*	*0.76*	—	97.8
		2.2 (5.9)	**0.0 (0.2)**	**0.0 (0.0)**	— (—)	
	3	*0.65*	*0.76*	—	—	99.8
		0.2 (3.3)	**0.0 (0.0)**	—(—)	—(—)	
6	1	*0.83*	*0.88*	*0.88*	*0.88*	100
		0.0 (4.4)	**0.0 (0.0)**	**0.0 (0.0)**	**0.0 (0.0)**	
	2	*0.84*	*0.89*	*0.89*	—	100
		0.0 (2.6)	**0.0 (0.0)**	**0.0 (0.0)**	— (—)	
	3	*0.85*	*0.9*	—	—	100
		0.0 (2.0)	**0.0 (0.0)**	—(—)	—(—)	

Table 7: Simulation results for proposed design assuming $\boldsymbol{s} = (0, 0, 1, 1, 1)$. For each scenario we provide the true ATS by risk group (italicized numbers in table) and report the correct selection probabilities and the average number of patients per dose for each risk group.

Selection percentage for each dose/risk-group combination (average # of patients)						
		Dose				
Scen	Risk Grp	1	2	3	4	None
1	1	*0.05*	*0.1*	*0.2*	*0.3*	0
		0.2 (3.1)	**8.4 (4.3)**	**41.2 (6.9)**	**50.2 (6.7)**	
	2	*0.1*	*0.2*	*0.3*	—	0.4
		8.1 (4.5)	**50.2 (7.5)**	**41.3 (6.0)**	— (—)	
	3	*0.15*	*0.25*	—	—	3.1
		44.7 (6.8)	**52.2 (4.9)**	—(—)	—(—)	
2	1	*0.01*	*0.01*	*0.04*	*0.5*	0
		0.0 (3.0)	**0.0 (3.0)**	**73.9 (7.8)**	**26.1 (7.2)**	
	2	*0.01*	*0.01*	*0.5*	—	0
		0.0 (3.0)	**74.8 (8.1)**	**25.2 (6.9)**	— (—)	
	3	*0.01*	*0.5*	—	—	0
		73.6 (6.3)	**26.4 (5.7)**	—(—)	—(—)	
3	1	*0.08*	*0.13*	*0.31*	*0.43*	0
		0.7 (3.2)	**28.1 (5.8)**	**54.8 (8.7)**	**16.4 (3.3)**	
	2	*0.13*	*0.25*	*0.43*	—	0.3
		17.9 (5.4)	**69.5 (8.9)**	**12.3 (3.6)**	— (—)	
	3	*0.13*	*0.3*	—	—	2.1
		57.3 (7.2)	**40.6 (4.6)**	—(—)	—(—)	
4	1	*0.08*	*0.24*	*0.32*	*0.43*	0
		9.3 (4.5)	**43.0 (8.4)**	**37.0 (6.0)**	**10.7 (2.0)**	
	2	*0.24*	*0.32*	*0.43*	—	7.3
		50.6 (9.6)	**36.1 (6.0)**	**6.0 (1.6)**	— (—)	
	3	*0.3*	*0.43*	—	—	32.1
		61.2 (8.4)	**6.7 (1.6)**	—(—)	—(—)	
5	1	*0.25*	*0.5*	*0.5*	*0.5*	5.9
		74.4 (13.4)	**16.5 (5.9)**	**3.0 (1.0)**	**0.2 (0.1)**	
	2	*0.35*	*0.5*	*0.5*	—	41.1
		56.9 (12.5)	**1.9 (1.7)**	**0.1 (0.1)**	— (—)	
	3	*0.5*	*0.5*	—	—	88.3
		11.6 (6.0)	**0.1 (0.4)**	—(—)	—(—)	
6	1	*0.57*	*0.6*	*0.6*	*0.62*	88.6
		9.9 (9.2)	**1.4 (1.1)**	**0.1 (0.1)**	**0.0 (0.0)**	
	2	*0.58*	*0.61*	*0.61*	—	99
		1.0 (5.0)	**0.0 (0.3)**	**0.0 (0.0)**	— (—)	
	3	*0.59*	*0.62*	—	—	99.9
		0.1 (3.2)	**0.0 (0.1)**	—(—)	—(—)	

Table 8: Simulation results for proposed design assuming $\boldsymbol{s} = (0, 0, 0, 1, 1)$. For each scenario we provide the true ATS by risk group (italicized numbers in table) and report the correct selection probabilities and the average number of patients per dose for each risk group.

Selection percentage for each dose/risk-group combination (average # of patients)						
		Dose				
Scen	Risk Grp	1	2	3	4	None
1	1	*0.03*	*0.05*	*0.11*	*0.18*	0
		0.0 (3.0)	**0.5 (3.2)**	**9.8 (4.3)**	**89.7 (10.4)**	
	2	*0.05*	*0.11*	*0.18*	—	0
		1.6 (3.4)	**15.6 (4.8)**	**82.8 (9.7)**	— (—)	
	3	*0.1*	*0.13*	—	—	1
		19.0 (5.2)	**80.0 (6.7)**	—(—)	—(—)	
2	1	*0*	*0*	*0.01*	*0.2*	0
		0.0 (3.0)	**0.0 (3.0)**	**9.3 (3.8)**	**90.7 (11.2)**	
	2	*0*	*0*	*0.2*	—	0
		0.0 (3.0)	**11.2 (3.9)**	**88.8 (11.1)**	— (—)	
	3	*0*	*0.2*	—	—	0
		13.9 (3.6)	**86.1 (8.4)**	—(—)	—(—)	
3	1	*0.02*	*0.06*	*0.17*	*0.28*	0
		0.0 (3.0)	**2.5 (3.5)**	**34.2 (6.2)**	**63.3 (8.3)**	
	2	*0.06*	*0.13*	*0.28*	—	0
		1.2 (3.4)	**38.2 (6.4)**	**60.6 (8.2)**	— (—)	
	3	*0.06*	*0.18*	—	—	0.2
		18.9 (4.7)	**80.9 (7.2)**	—(—)	—(—)	
4	1	*0.02*	*0.07*	*0.1*	*0.13*	0
		0.0 (3.0)	**0.3 (3.4)**	**7.1 (4.2)**	**92.6 (10.5)**	
	2	*0.07*	*0.17*	*0.13*	—	0.1
		5.9 (4.1)	**16.9 (5.5)**	**77.1 (8.4)**	— (—)	
	3	*0.18*	*0.13*	—	—	3.9
		36.6 (6.7)	**59.5 (5.0)**	—(—)	—(—)	
5	1	*0.13*	*0.2*	*0.25*	*0.25*	0.1
		2.8 (3.6)	**16.5 (5.4)**	**32.7 (6.2)**	**47.9 (5.8)**	
	2	*0.05*	*0.2*	*0.25*	—	0.1
		10.2 (4.7)	**42.3 (7.0)**	**47.4 (6.3)**	— (—)	
	3	*0.2*	*0.2*	—	—	6.5
		45.1 (7.3)	**48.4 (4.4)**	—(—)	—(—)	
6	1	*0.32*	*0.28*	*0.35*	*0.37*	10.6
		26.6 (8.2)	**30.0 (6.0)**	**22.0 (4.0)**	**10.8 (1.6)**	
	2	*0.33*	*0.29*	*0.36*	—	31
		37.7 (9.4)	**24.5 (4.1)**	**6.8 (1.5)**	— (—)	
	3	*0.34*	*0.3*	—	—	55.3
		33.0 (6.7)	**11.7 (1.7)**	—(—)	—(—)	

Table 9: Simulation results for proposed design for 3 additional scenarios. For each scenario we provide the true ATS by risk group (italicized numbers in table) and report the correct selection probabilities and the average number of patients per dose for each risk group.

Selection percentage for each dose/risk-group combination (average # of patients)						
		Dose				
Scen	Risk Grp	1	2	3	4	None
1	1	*0.07*	*0.11*	*0.18*	*0.25*	0
		0.0 (3.0)	**3.3 (3.4)**	**28.6 (5.9)**	**68.1 (8.7)**	
	2	*0.01*	*0.01*	*0.38*	—	0
		0.0 (3.0)	**72.1 (7.8)**	**27.9 (7.2)**	— (—)	
	3	*0.14*	*0.21*	—	—	1.7
		21.7 (5.1)	**76.6 (6.8)**	—(—)	—(—)	
2	1	*0.06*	*0.11*	*0.24*	*0.39*	0.1
		0.6 (3.4)	**14.4 (5.9)**	**60.6 (8.5)**	**24.3 (3.2)**	
	2	*0.48*	*0.5*	*0.52*	—	92.2
		7.4 (7.7)	**0.4 (0.6)**	**0.0 (0.0)**	— (—)	
	3	*0.49*	*0.51*	—	—	99.1
		0.9 (4.0)	**0.0 (0.1)**	—(—)	—(—)	
3	1	*0.01*	*0.01*	*0.03*	*0.38*	0
		0.0 (3.0)	**0.0 (3.0)**	**67.6 (7.2)**	**32.4 (7.8)**	
	2	*0.01*	*0.01*	*0.38*	—	0
		0.0 (3.0)	**68.2 (7.4)**	**31.8 (7.6)**	— (—)	
	3	*0.49*	*0.51*	—	—	90.5
		8.8 (6.1)	**0.7 (0.3)**	—(—)	—(—)	

Table 10: Simulation results using the M-CRM for dose finding with risk groups for 3 additional scenarios. For each scenario we provide the true ATS by risk group (italicized numbers in table) and report the correct selection probabilities and the average number of patients per dose for each risk group.

Selection percentage for each dose/risk-group combination (average # of patients)						
		Dose				
Scen	Risk Grp	1	2	3	4	None
1	1	*0.07*	*0.11*	*0.18*	*0.25*	0.4
		0.0 (3.5)	**10.8 (5.4)**	**42.3 (7.1)**	**46.5 (4.9)**	
	2	*0.01*	*0.01*	*0.38*	—	0
		0.0 (3.0)	**21.8 (4.5)**	**78.2 (10.5)**	— (—)	
	3	*0.14*	*0.21*	—	—	7.7
		11.1 (4.5)	**81.2 (6.7)**	—(—)	—(—)	
2	1	*0.06*	*0.11*	*0.24*	*0.39*	0.1
		0.1 (3.3)	**13.4 (6.1)**	**69.1 (8.8)**	**17.3 (2.7)**	
	2	*0.48*	*0.5*	*0.52*	—	92.4
		7.2 (6.7)	**0.4 (0.4)**	**0.0 (0.0)**	— (—)	
	3	*0.49*	*0.51*	—	—	88
		12.0 (5.3)	**0.0 (0.3)**	—(—)	—(—)	
3	1	*0.01*	*0.01*	*0.03*	*0.38*	0
		0.0 (3.0)	**0.0 (3.0)**	**28.2 (4.6)**	**71.8 (10.4)**	
	2	*0.01*	*0.01*	*0.38*	—	0
		0.0 (3.0)	**22.1 (4.5)**	**77.9 (10.5)**	— (—)	
	3	*0.49*	*0.51*	—	—	87.4
		12.5 (5.2)	**0.1 (0.3)**	—(—)	—(—)	

Table 11: Simulation results for proposed design assuming that the patients in the highest 2 risk groups are allowed to escalate to higher doses. For each scenario we provide the true ATS by risk group (italicized numbers in table) and report the correct selection probabilities and the average number of patients per dose for each risk group.

Selection percentage for each dose/risk-group combination (average # of patients)						
		Dose				
Scen	Risk Grp	1	2	3	4	None
1	1	*0.07*	*0.11*	*0.18*	*0.25*	0
		0.1 (3.0)	**3.7 (3.5)**	**24.4 (4.4)**	**71.8 (6.1)**	
	2	*0.11*	*0.18*	*0.25*	*0.38*	0.2
		3.4 (3.6)	**25.6 (5.4)**	**55.1 (5.7)**	**15.7 (2.3)**	
	3	*0.14*	*0.21*	*0.27*	*0.38*	2.3
		20.5 (6.5)	**46.5 (6.5)**	**28.7 (3.1)**	**2.0 (0.6)**	
2	1	*0.01*	*0.01*	*0.03*	*0.38*	0
		0.0 (3.0)	**0.0 (3.0)**	**58.3 (5.0)**	**41.7 (5.9)**	
	2	*0.01*	*0.01*	*0.38*	*0.4*	0
		0.0 (3.0)	**66.0 (6.8)**	**32.9 (6.7)**	**1.1 (0.5)**	
	3	*0.01*	*0.38*	*0.41*	*0.45*	0
		73.1 (8.7)	**26.1 (7.8)**	**0.8 (0.5)**	**0.0 (0.0)**	
3	1	*0.08*	*0.1*	*0.25*	*0.35*	0
		0.3 (3.1)	**12.3 (4.0)**	**56.8 (6.1)**	**30.6 (3.8)**	
	2	*0.1*	*0.21*	*0.35*	*0.38*	0.2
		6.6 (3.8)	**56.4 (7.3)**	**32.4 (5.0)**	**4.4 (0.8)**	
	3	*0.1*	*0.25*	*0.38*	*0.4*	0.9
		30.3 (6.8)	**62.0 (8.3)**	**6.6 (1.7)**	**0.2 (0.1)**	
4	1	*0.06*	*0.17*	*0.24*	*0.31*	0
		2.4 (3.3)	**19.3 (5.1)**	**47.0 (5.7)**	**31.3 (2.9)**	
	2	*0.17*	*0.27*	*0.31*	*0.38*	1.4
		28.7 (6.3)	**46.7 (7.3)**	**21.1 (2.9)**	**2.1 (0.4)**	
	3	*0.25*	*0.31*	*0.38*	*0.4*	18.5
		59.8 (10.9)	**20.4 (3.7)**	**1.3 (0.5)**	**0.0 (0.0)**	
5	1	*0.21*	*0.38*	*0.4*	*0.4*	0.9
		62.4 (9.0)	**28.9 (6.7)**	**6.4 (1.1)**	**1.4 (0.2)**	
	2	*0.24*	*0.38*	*0.4*	*0.4*	11.5
		81.3 (12.9)	**6.2 (2.8)**	**0.9 (0.3)**	**0.1 (0.0)**	
	3	*0.36*	*0.38*	*0.4*	*0.4*	71.9
		27.8 (9.5)	**0.3 (0.5)**	**0.0 (0.0)**	**0.0 (0.0)**	
6	1	*0.47*	*0.49*	*0.51*	*0.52*	82.6
		15.9 (8.3)	**1.3 (0.8)**	**0.2 (0.1)**	**0.0 (0.0)**	
	2	*0.48*	*0.5*	*0.52*	*0.53*	98.4
		1.4 (4.8)	**0.1 (0.2)**	**0.1 (0.0)**	**0.0 (0.0)**	
	3	*0.49*	*0.51*	*0.53*	*0.54*	100
		0.0 (3.2)	**0.0 (0.0)**	**0.0 (0.0)**	**0.0 (0.0)**	

Table 12: Simulation results using the M-CRM for dose finding with risk groups assuming that the patients in the highest 2 risk groups are not allowed to escalate to higher doses. For each scenario we provide the true ATS by risk group (italicized numbers in table) and report the correct selection probabilities and the average number of patients per dose for each risk group.

Selection percentage for each dose/risk-group combination (average # of patients)						
		Dose				
Scen	Risk Grp	1	2	3	4	None
1	1	*0.07*	*0.11*	*0.18*	*0.25*	0
		0.6 (3.5)	**13.4 (4.9)**	**41.4 (5.3)**	**44.6 (3.2)**	
	2	*0.11*	*0.18*	*0.25*	*0.38*	1.8
		6.4 (4.6)	**33.3 (6.1)**	**38.3 (4.1)**	**20.2 (1.9)**	
	3	*0.14*	*0.21*	*0.27*	*0.38*	7
		9.2 (5.2)	**25.4 (4.6)**	**34.1 (3.9)**	**24.3 (2.3)**	
2	1	*0.01*	*0.01*	*0.03*	*0.38*	0
		0.0 (3.0)	**0.0 (3.0)**	**21.7 (3.9)**	**78.3 (7.1)**	
	2	*0.01*	*0.01*	*0.38*	*0.4*	0
		0.0 (3.0)	**20.9 (4.1)**	**62.3 (7.0)**	**16.8 (2.9)**	
	3	*0.01*	*0.38*	*0.41*	*0.45*	0.1
		26.3 (5.6)	**63.4 (8.1)**	**8.9 (2.9)**	**1.3 (0.4)**	
3	1	*0.08*	*0.1*	*0.25*	*0.35*	0
		0.4 (3.6)	**18.7 (5.6)**	**62.7 (6.2)**	**18.2 (1.6)**	
	2	*0.1*	*0.21*	*0.35*	*0.38*	1.9
		6.0 (4.6)	**52.7 (7.1)**	**30.9 (3.9)**	**8.5 (1.0)**	
	3	*0.1*	*0.25*	*0.38*	*0.4*	3
		15.1 (5.3)	**44.1 (6.0)**	**30.3 (4.2)**	**7.5 (1.0)**	
4	1	*0.06*	*0.17*	*0.24*	*0.31*	0
		1.8 (3.6)	**34.4 (7.0)**	**43.3 (4.9)**	**20.5 (1.5)**	
	2	*0.17*	*0.27*	*0.31*	*0.38*	5.2
		32.9 (7.9)	**38.2 (5.8)**	**17.5 (2.1)**	**6.2 (0.6)**	
	3	*0.25*	*0.31*	*0.38*	*0.4*	22.8
		39.0 (8.3)	**30.0 (4.0)**	**7.2 (1.7)**	**1.0 (0.2)**	
5	1	*0.21*	*0.38*	*0.4*	*0.4*	6.5
		60.9 (10.4)	**30.4 (5.6)**	**2.0 (0.3)**	**0.2 (0.0)**	
	2	*0.24*	*0.38*	*0.4*	*0.4*	6.7
		74.6 (12.1)	**17.1 (3.8)**	**1.5 (0.3)**	**0.1 (0.0)**	
	3	*0.36*	*0.38*	*0.4*	*0.4*	55.9
		37.7 (9.2)	**4.8 (1.5)**	**1.4 (0.3)**	**0.2 (0.0)**	
6	1	*0.47*	*0.49*	*0.51*	*0.52*	87.3
		12.4 (7.4)	**0.2 (0.4)**	**0.1 (0.0)**	**0.0 (0.0)**	
	2	*0.48*	*0.5*	*0.52*	*0.53*	89.1
		10.6 (6.6)	**0.3 (0.4)**	**0.0 (0.0)**	**0.0 (0.0)**	
	3	*0.49*	*0.51*	*0.53*	*0.54*	91.7
		8.2 (6.0)	**0.1 (0.3)**	**0.0 (0.0)**	**0.0 (0.0)**	

Table 13: Simulation results for proposed design using isotonization of the mean toxicity score. In this case we used the same prior as in the **base-case**. For each scenario we provide the true ATS by risk group (italicized numbers in table) and report the correct selection probabilities and the average number of patients per dose for each risk group.

Selection percentage for each dose/risk-group combination (average # of patients)						
		Dose				
Scen	Risk Grp	1	2	3	4	None
1	1	*0.07*	*0.11*	*0.18*	*0.25*	0
		0.0 (3.0)	**3.3 (3.4)**	**29.1 (4.6)**	**67.6 (10.0)**	
	2	*0.11*	*0.18*	*0.25*	—	0.6
		2.0 (3.2)	**35.3 (5.2)**	**62.1 (9.5)**	— (—)	
	3	*0.14*	*0.21*	—	—	2
		15.0 (3.0)	**83.0 (8.9)**	—(—)	—(—)	
2	1	*0.01*	*0.01*	*0.03*	*0.38*	0
		0.0 (3.0)	**11.1 (3.0)**	**55.6 (5.7)**	**33.3 (9.3)**	
	2	*0.01*	*0.01*	*0.38*	—	0
		14.0 (3.0)	**54.5 (5.6)**	**31.5 (9.4)**	— (—)	
	3	*0.01*	*0.38*	—	—	10.3
		48.3 (4.4)	**41.4 (7.6)**	—(—)	—(—)	
3	1	*0.08*	*0.1*	*0.25*	*0.35*	0.1
		0.3 (3.1)	**15.1 (4.2)**	**52.4 (7.0)**	**32.1 (6.8)**	
	2	*0.1*	*0.21*	*0.35*	—	0.5
		6.8 (3.5)	**59.4 (7.6)**	**33.3 (6.9)**	— (—)	
	3	*0.1*	*0.25*	—	—	1.2
		28.3 (4.1)	**70.5 (7.9)**	—(—)	—(—)	
4	1	*0.06*	*0.17*	*0.24*	*0.31*	0
		1.0 (3.2)	**14.6 (4.4)**	**40.4 (5.6)**	**44.0 (7.8)**	
	2	*0.17*	*0.27*	*0.31*	—	2.3
		21.0 (4.8)	**40.5 (6.1)**	**36.2 (6.8)**	— (—)	
	3	*0.25*	*0.31*	—	—	9.9
		39.3 (4.7)	**50.8 (6.9)**	—(—)	—(—)	
5	1	*0.21*	*0.38*	*0.4*	*0.4*	4.3
		48.5 (8.3)	**24.0 (6.4)**	**16.1 (3.4)**	**7.1 (2.3)**	
	2	*0.24*	*0.38*	*0.4*	—	15.2
		65.7 (9.9)	**13.3 (4.1)**	**5.8 (2.5)**	— (—)	
	3	*0.36*	*0.38*	—	—	43.1
		35.6 (5.6)	**21.3 (4.7)**	—(—)	—(—)	
6	1	*0.47*	*0.49*	*0.51*	*0.52*	77.9
		12.0 (6.8)	**7.2 (3.4)**	**2.3 (1.1)**	**0.6 (0.5)**	
	2	*0.48*	*0.5*	*0.52*	—	94.5
		2.9 (4.5)	**1.7 (1.5)**	**0.9 (0.9)**	— (—)	
	3	*0.49*	*0.51*	—	—	98.3
		0.8 (3.0)	**0.9 (1.9)**	—(—)	—(—)	

Table 14: Simulation results for proposed design using isotonization of the mean toxicity score and a more conservative yet less informative prior with $\boldsymbol{\alpha}_{jh} = (0.125, 0.0613, 0.022, 0.022, 0.019)$ (Expected Prior ATS of 0.25). For each scenario we provide the true ATS by risk group (italicized numbers in table) and report the correct selection probabilities and the average number of patients per dose for each risk group.

Selection percentage for each dose/risk-group combination (average # of patients)						
		Dose				
Scen	Risk Grp	1	2	3	4	None
1	1	*0.07*	*0.11*	*0.18*	*0.25*	0.1
		0.7 (2.9)	**7.0 (3.4)**	**32.1 (5.2)**	**60.1 (9.4)**	
	2	*0.11*	*0.18*	*0.25*	—	1.8
		5.2 (3.2)	**38.1 (5.8)**	**54.9 (8.8)**	— (—)	
	3	*0.14*	*0.21*	—	—	3.9
		18.4 (3.6)	**77.7 (8.2)**	—(—)	—(—)	
2	1	*0.01*	*0.01*	*0.03*	*0.38*	0.5
		4.3 (2.9)	**10.0 (2.5)**	**59.4 (6.4)**	**25.8 (9.3)**	
	2	*0.01*	*0.01*	*0.38*	—	6.3
		11.1 (2.4)	**54.0 (6.0)**	**28.6 (9.6)**	— (—)	
	3	*0.01*	*0.38*	—	—	12.2
		50.6 (4.7)	**37.2 (7.3)**	—(—)	—(—)	
3	1	*0.08*	*0.1*	*0.25*	*0.35*	0.5
		2.3 (2.9)	**19.9 (4.5)**	**49.5 (7.5)**	**27.8 (6.0)**	
	2	*0.1*	*0.21*	*0.35*	—	3.2
		9.7 (3.5)	**59.5 (8.2)**	**27.6 (6.0)**	— (—)	
	3	*0.1*	*0.25*	—	—	2.6
		32.0 (4.5)	**65.4 (7.3)**	—(—)	—(—)	
4	1	*0.06*	*0.17*	*0.24*	*0.31*	0.1
		3.0 (3.2)	**21.2 (4.8)**	**40.5 (6.2)**	**35.2 (6.8)**	
	2	*0.17*	*0.27*	*0.31*	—	5.8
		25.8 (5.4)	**37.5 (6.1)**	**30.9 (6.0)**	— (—)	
	3	*0.25*	*0.31*	—	—	16.3
		39.0 (5.0)	**44.7 (6.3)**	—(—)	—(—)	
5	1	*0.21*	*0.38*	*0.4*	*0.4*	8.8
		54.0 (9.1)	**21.7 (6.2)**	**9.9 (2.9)**	**5.6 (1.7)**	
	2	*0.24*	*0.38*	*0.4*	—	26
		61.0 (9.8)	**10.7 (3.8)**	**2.3 (1.5)**	— (—)	
	3	*0.36*	*0.38*	—	—	56.5
		27.4 (5.0)	**16.1 (3.9)**	—(—)	—(—)	
6	1	*0.47*	*0.49*	*0.51*	*0.52*	86.2
		8.2 (6.0)	**4.0 (2.2)**	**1.4 (0.6)**	**0.2 (0.2)**	
	2	*0.48*	*0.5*	*0.52*	—	98.8
		0.7 (3.4)	**0.3 (0.8)**	**0.2 (0.3)**	— (—)	
	3	*0.49*	*0.51*	—	—	99.9
		0.1 (2.2)	**0.0 (1.1)**	—(—)	—(—)	

Table 15: Simulation results for proposed design using no isotonization for dose-escalation decisions. In this case we used the same prior as in the **base-case**. At the end of the trial we use an isotonization algorithm to make the final dose selection decision. For each scenario we provide the true ATS by risk group (italicized numbers in table) and report the correct selection probabilities and the average number of patients per dose for each risk group.

Selection percentage for each dose/risk-group combination (average # of patients)						
		Dose				
Scen	Risk Grp	1	2	3	4	None
1	1	*0.07*	*0.11*	*0.18*	*0.25*	0.2
		0.1 (3.5)	**2.5 (4.1)**	**33.6 (6.0)**	**63.6 (7.3)**	
	2	*0.11*	*0.18*	*0.25*	—	1.1
		1.8 (4.3)	**34.2 (6.0)**	**62.9 (7.6)**	— (—)	
	3	*0.14*	*0.21*	—	—	1.7
		19.8 (5.1)	**78.5 (6.8)**	—(—)	—(—)	
2	1	*0.01*	*0.01*	*0.03*	*0.38*	0
		0.0 (3.0)	**0.0 (3.0)**	**52.2 (6.6)**	**47.8 (8.4)**	
	2	*0.01*	*0.01*	*0.38*	—	0
		0.0 (3.0)	**50.0 (6.4)**	**50.0 (8.6)**	— (—)	
	3	*0.01*	*0.38*	—	—	0
		40.2 (5.0)	**59.8 (7.0)**	—(—)	—(—)	
3	1	*0.08*	*0.1*	*0.25*	*0.35*	0.1
		0.1 (3.6)	**8.3 (5.0)**	**64.1 (8.0)**	**27.4 (4.3)**	
	2	*0.1*	*0.21*	*0.35*	—	0.4
		4.8 (4.8)	**59.9 (7.3)**	**34.9 (5.9)**	— (—)	
	3	*0.1*	*0.25*	—	—	0.5
		24.3 (4.6)	**75.2 (7.3)**	—(—)	—(—)	
4	1	*0.06*	*0.17*	*0.24*	*0.31*	0
		0.4 (3.6)	**19.8 (6.7)**	**46.7 (6.3)**	**33.1 (4.4)**	
	2	*0.17*	*0.27*	*0.31*	—	3.5
		22.3 (7.0)	**50.8 (6.7)**	**23.4 (3.8)**	— (—)	
	3	*0.25*	*0.31*	—	—	10.1
		56.6 (7.0)	**33.3 (4.4)**	—(—)	—(—)	
5	1	*0.21*	*0.38*	*0.4*	*0.4*	6
		57.8 (11.8)	**32.1 (7.0)**	**3.3 (1.1)**	**0.8 (0.2)**	
	2	*0.24*	*0.38*	*0.4*	—	10.2
		69.8 (11.2)	**17.2 (4.9)**	**2.8 (0.8)**	— (—)	
	3	*0.36*	*0.38*	—	—	45.6
		51.0 (7.7)	**3.4 (1.8)**	—(—)	—(—)	
6	1	*0.47*	*0.49*	*0.51*	*0.52*	93.4
		6.4 (7.8)	**0.2 (0.6)**	**0.0 (0.0)**	**0.0 (0.0)**	
	2	*0.48*	*0.5*	*0.52*	—	94
		5.6 (7.0)	**0.3 (0.4)**	**0.1 (0.0)**	— (—)	
	3	*0.49*	*0.51*	—	—	84.6
		14.9 (6.2)	**0.5 (0.4)**	—(—)	—(—)	

Table 16: Simulation results for proposed design using no isotonization for dose-escalation decisions and a more conservative yet less informative prior with $\boldsymbol{\alpha}_{jh} = (0.125, 0.0613, 0.022, 0.022, 0.019)$ (Expected Prior ATS of 0.25). At the end of the trial we use an isotonization algorithm to make the final dose selection decision. For each scenario we provide the true ATS by risk group (italicized numbers in table) and report the correct selection probabilities and the average number of patients per dose for each risk group.

Selection percentage for each dose/risk-group combination (average # of patients)						
		Dose				
Scen	Risk Grp	1	2	3	4	None
1	1	*0.07*	*0.11*	*0.18*	*0.25*	0
		0.3 (3.7)	**5.8 (4.7)**	**37.9 (6.3)**	**56.0 (6.3)**	
	2	*0.11*	*0.18*	*0.25*	—	1.5
		4.2 (4.8)	**38.1 (6.3)**	**56.2 (6.7)**	— (—)	
	3	*0.14*	*0.21*	—	—	3.1
		25.1 (5.2)	**71.8 (6.6)**	—(—)	—(—)	
2	1	*0.01*	*0.01*	*0.03*	*0.38*	0
		0.0 (3.0)	**0.0 (3.1)**	**72.1 (7.7)**	**27.9 (7.2)**	
	2	*0.01*	*0.01*	*0.38*	—	0
		0.0 (3.0)	**72.0 (7.4)**	**28.0 (7.5)**	— (—)	
	3	*0.01*	*0.38*	—	—	0
		66.7 (5.8)	**33.3 (6.2)**	—(—)	—(—)	
3	1	*0.08*	*0.1*	*0.25*	*0.35*	0.5
		0.3 (3.9)	**18.3 (5.9)**	**58.0 (7.7)**	**22.9 (3.4)**	
	2	*0.1*	*0.21*	*0.35*	—	1.4
		7.1 (5.2)	**66.8 (7.7)**	**24.7 (5.0)**	— (—)	
	3	*0.1*	*0.25*	—	—	1
		27.6 (4.9)	**71.4 (7.0)**	—(—)	—(—)	
4	1	*0.06*	*0.17*	*0.24*	*0.31*	0.2
		3.8 (4.3)	**28.4 (7.6)**	**44.6 (6.2)**	**23.0 (2.9)**	
	2	*0.17*	*0.27*	*0.31*	—	4.7
		29.7 (8.0)	**46.2 (6.5)**	**19.4 (2.9)**	— (—)	
	3	*0.25*	*0.31*	—	—	16.1
		61.3 (7.4)	**22.6 (3.6)**	—(—)	—(—)	
5	1	*0.21*	*0.38*	*0.4*	*0.4*	9.4
		67.2 (13.1)	**20.1 (5.6)**	**3.0 (0.8)**	**0.3 (0.1)**	
	2	*0.24*	*0.38*	*0.4*	—	21.8
		63.9 (11.2)	**11.3 (3.7)**	**3.0 (0.4)**	— (—)	
	3	*0.36*	*0.38*	—	—	57.7
		38.2 (7.5)	**4.1 (1.1)**	—(—)	—(—)	
6	1	*0.47*	*0.49*	*0.51*	*0.52*	95.2
		4.7 (6.7)	**0.0 (0.4)**	**0.1 (0.0)**	**0.0 (0.0)**	
	2	*0.48*	*0.5*	*0.52*	—	94.8
		5.0 (6.1)	**0.2 (0.3)**	**0.0 (0.0)**	— (—)	
	3	*0.49*	*0.51*	—	—	89.9
		9.8 (5.3)	**0.3 (0.2)**	—(—)	—(—)	