Bayesian Model Averaging CRM in Phase I Clinical Trials

Ying Yuan
Department of Biostatistics
U. T. M. D. Anderson Cancer Center
Houston, TX

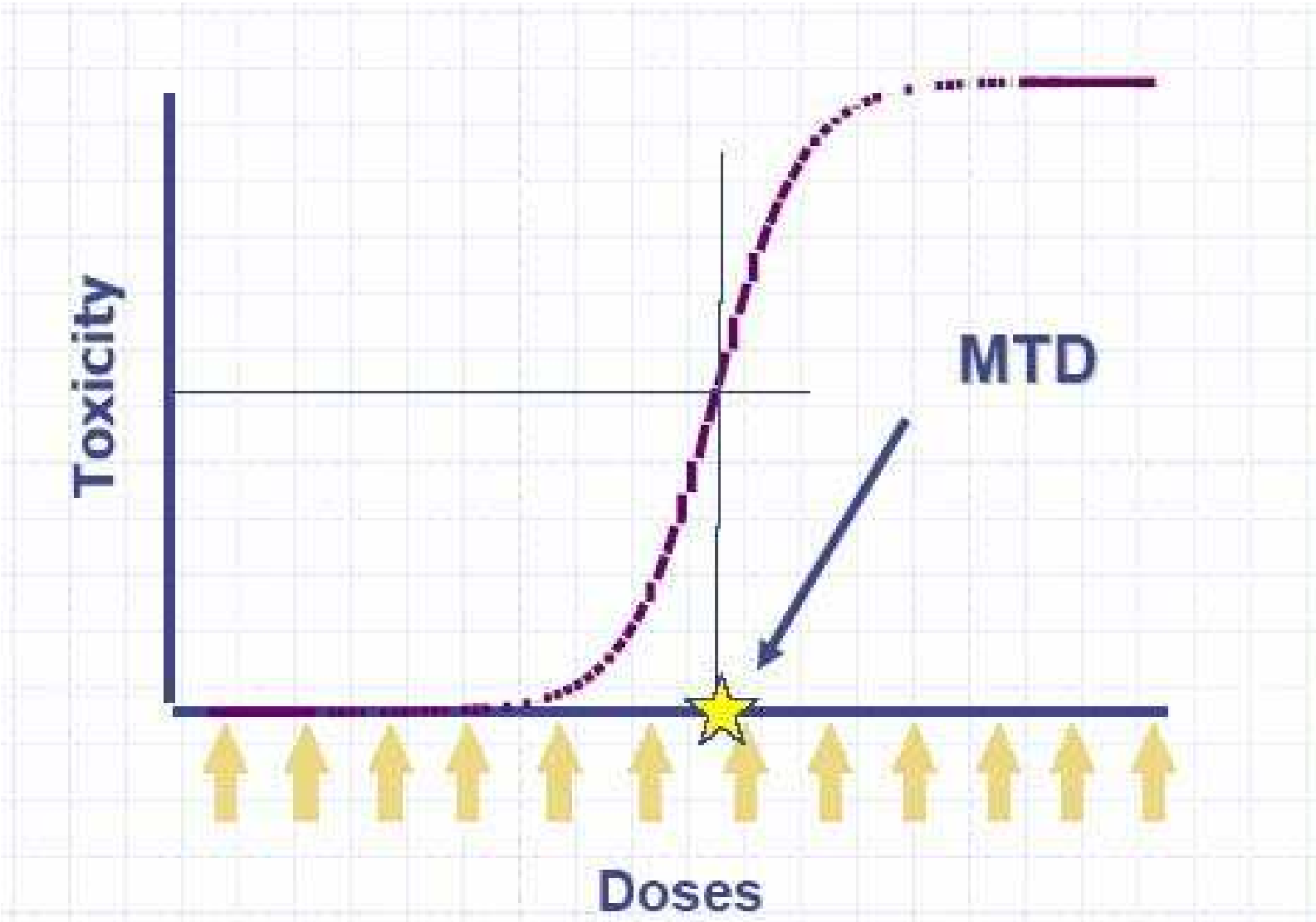
Joint work with Guosheng Yin

Outline

- Background and motivation
- Continual reassessment method (CRM)
- Bayesian Model Averaging (BMA)
- **BMA-CRM**
- Simulations
- Software and concluding remarks

Phase I Trials

- In conventional phase I trials, the primary objective is often to find the maximum tolerated dose (MTD).
- A sequence of doses is screened in order to find the target dose associated with the maximum level of tolerable toxicity.
- **Typically, we assume that toxicity monotonically increases with the dose.**


Toxicity
MTD
Doses

Dose-finding Methods

- "3+3" design (Storer, 1989)
- Continual reassessment method (O'Quigley et al., 1990)
- Decision theoretic approach (Whitehead and Brunier, 1995)
- Random walk rule (Durham et al., 1997)
- Dose escalation with overdose control (Babb et al., 1998)
- Many methods have been proposed for phase I trials, see Chevret (2006) for comprehensive reviews.

Continual Reassessment Method (CRM)

- CRM assumes a parametric function between the true toxicity probabilities and prespecified probabilities (O'Quigley et al., 1990), e.g.,

  $$\text{pr}(\text{toxicity at } d_j) = \pi_j(\alpha) = p_j^{\exp(\alpha)} \quad \text{for } j = 1, \cdots, J.$$

  where α is an unknown parameter.

- Based on observed data, the toxicity curve is continuous updated to direct the dose escalation and selection.

Likelihood and Posterior

- At dose level j, y_j out of n_j subjects experienced dose-limiting toxicities (DLT).
- The likelihood function is

$$L(D|\alpha) = \prod_{j=1}^{J} \{p_j^{\exp(\alpha)}\}^{y_j} \{1 - p_j^{\exp(\alpha)}\}^{n_j - y_j}.$$

- Posterior mean of the dose toxicity probability

$$\hat{\pi}_j = \int p_j^{\exp(\alpha)} \frac{L(D|\alpha) f(\alpha)}{\int L(D|\alpha) f(\alpha) d\alpha} d\alpha,$$

where $f(\alpha)$ is a prior distribution for the parameter $\alpha \sim N(0, \sigma^2)$.

CRM Decision Rule

- A new cohort of patients is assigned to dose level j^* such that

$$j^* = \text{argmin}_{j\in(1,\ldots,J)}|\hat{\pi}_j - \phi|.$$

- The trial continues until the exhaustion of the total sample size, and then the dose with a posterior toxicity probability closest to ϕ is selected as the MTD.
- A stopping rule: if pr(toxicity rate at $d_1 > \phi|D) > 0.9$, the trial is terminated for safety.

Refined CRM

- Faries (1994) and Goodman, Zahurak and Piantadosi (1995) developed practical improvements: assigning more than one subject to each dose and limiting dose escalation by one dose level.
- Møller (1995) used a preliminary up-and-down design in order to reach the neighborhood of the target dose during a successive escalation.
- Piantadosi, Fisher and Grossman (1998) used a simple dose-toxicity model to guide data interpolation.

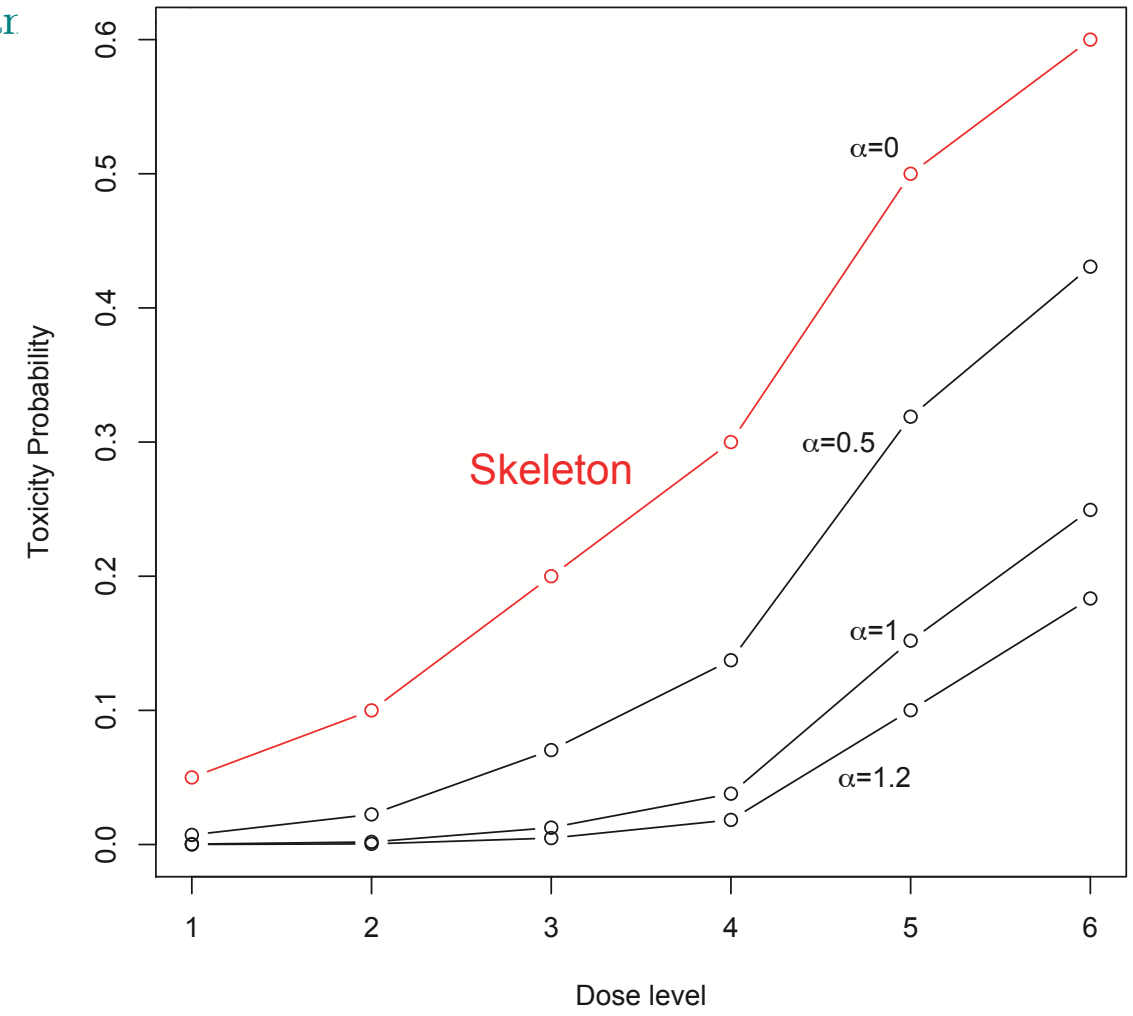
- Heyd and Carlin (1999) allowed the trial to stop earlier when the width of the posterior 95% probability interval for the MTD becomes sufficiently narrow.
- Leung and Wang (2002) used decision theory to optimize the number of patients allocated to the highest dose with toxicity not exceeding the tolerable level.
- Braun (2002) extended the CRM to model bivariate competing outcomes.
- For a comprehensive introduction, see the tutorial by Garrett-Mayer (2006).

Skeleton of CRM

- The CRM requires prespecification of prior toxicity probabilities for the doses, i.e., p_j's

$$\pi_j(\alpha) = p_j^{\exp(\alpha)} \text{ for } j = 1, \cdots, J.$$

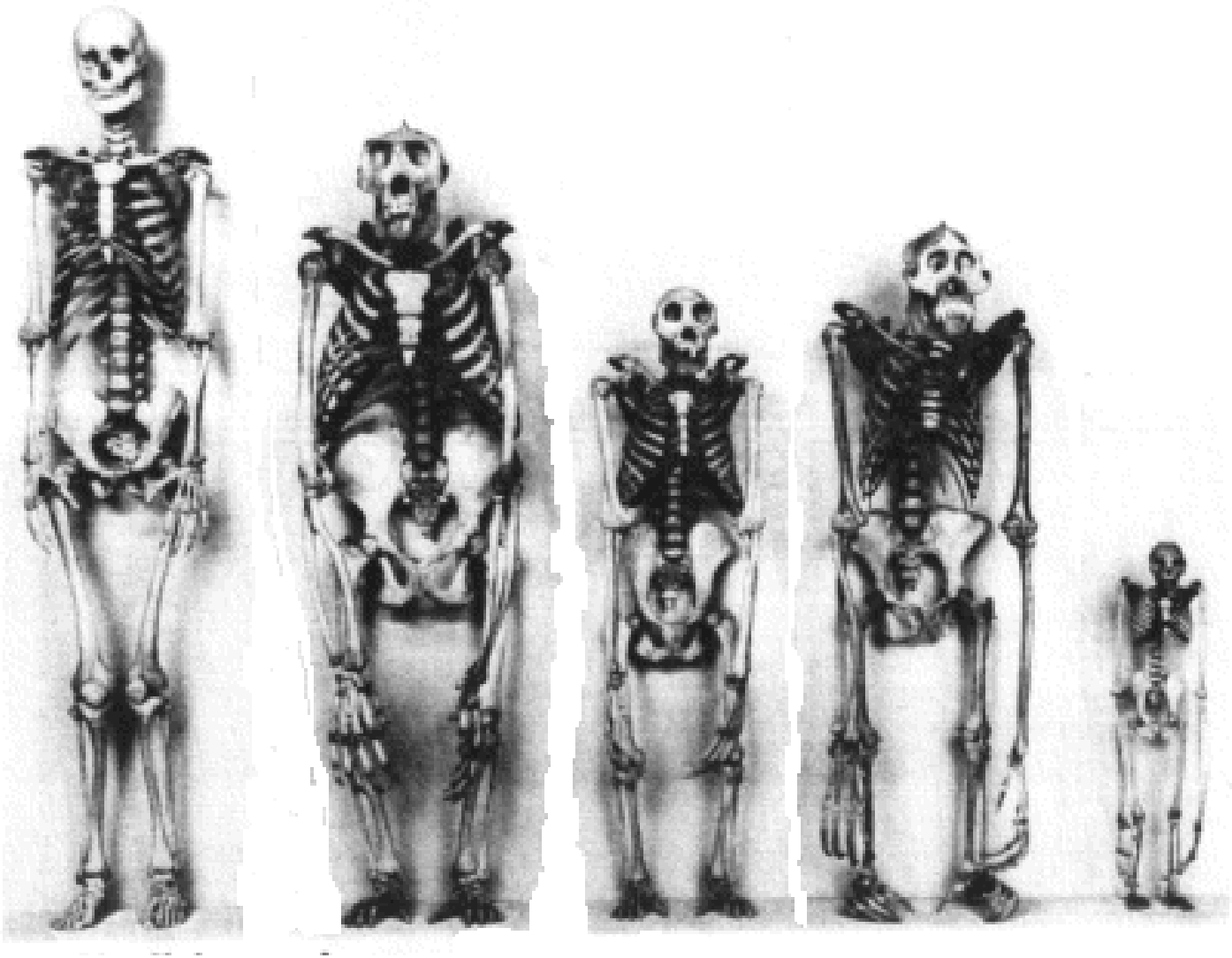
- We call $\{p_j\}$ the "**skeleton**" of the CRM as it forms the baseline (or prior) structure of the dose-toxicity curve.



- Idealy, we want to choose a set of $\{p_j\}$, which can reflect the true dose-toxicity relationship by a certain value of α.

Limitations of CRM

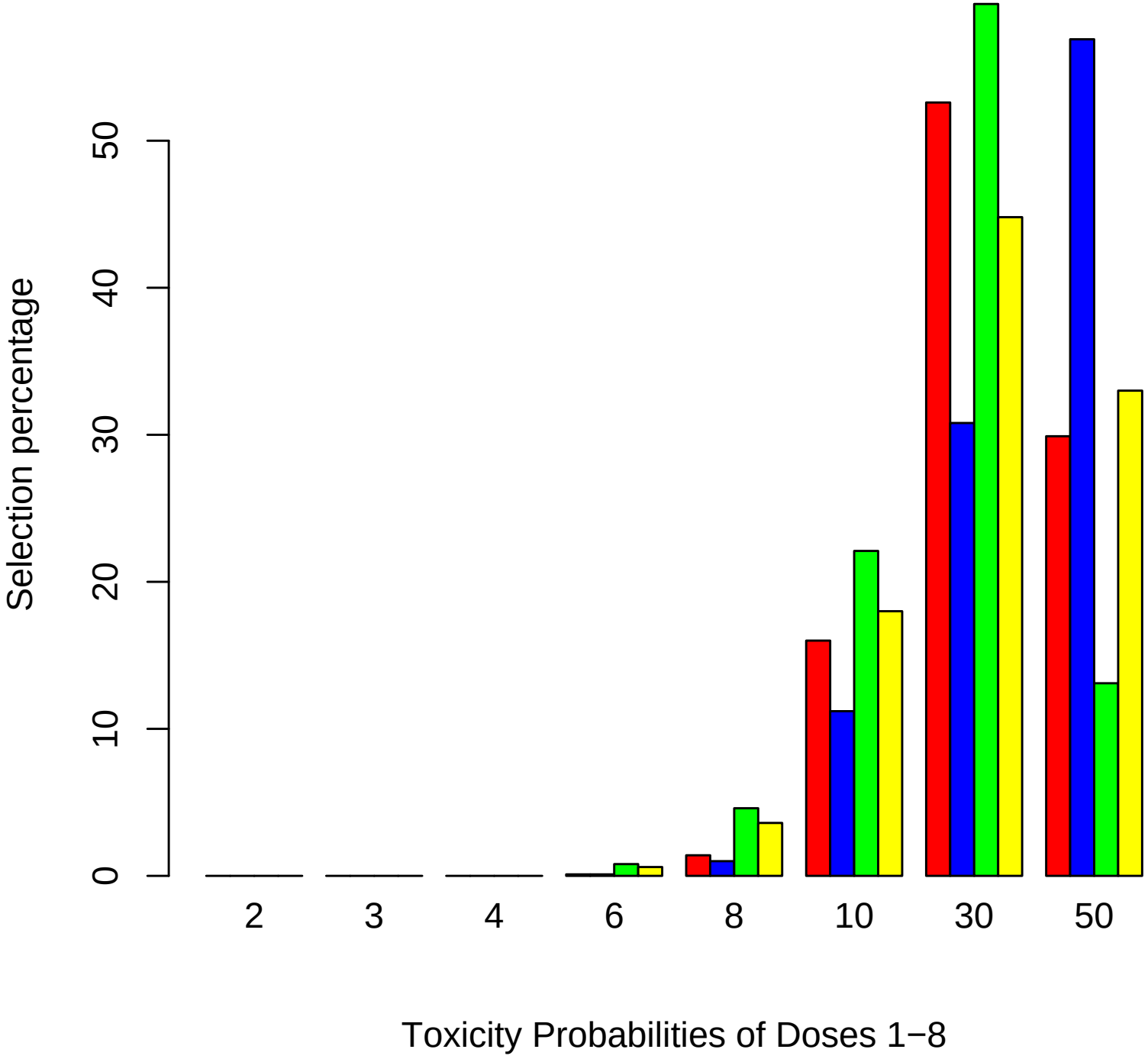
- Unfortunately, the prespecification of the skeleton can be arbitrary and very subjective.
- No information to justify whether a specific skeleton is reasonable because the underlying true toxicity probabilities are unknown.
- Different skeletons can lead to very different operating characteristics.
- Cheung and Chappell (2002) proposed a simple technique to evaluate the sensitivity of the CRM, which requires knowing the true dose-toxicity profile.

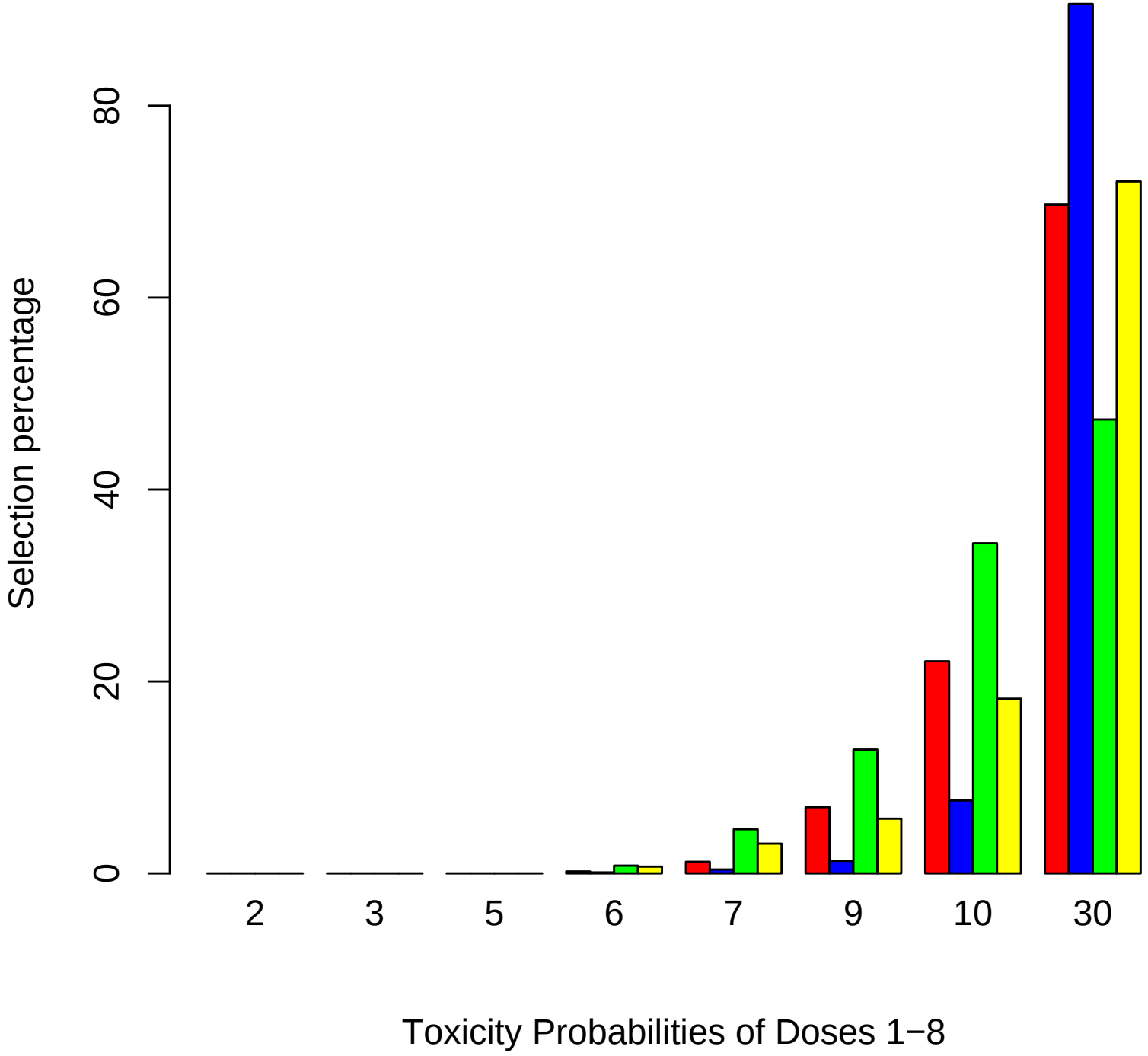


Sensitivity to skeleton

- An example: eight dose levels with a target $\phi = 30\%$:
- Two (true) toxicity scenarios:
  (0.02, 0.03, 0.04, 0.06, 0.08, 0.10 0.30, 0.50);
  (0.02, 0.03, 0.05, 0.06, 0.07, 0.09, 0.10, 0.30)
- Four skeletons $(p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8) =$

$$= \begin{cases} (0.02, 0.06, 0.08, 0.12, 0.20, 0.30, 0.40, 0.50), & 1 \\ (0.01, 0.05, 0.09, 0.14, 0.18, 0.22, 0.26, 0.30), & 2 \\ (0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80), & 3 \\ (0.20, 0.30, 0.40, 0.50, 0.60, 0.65, 0.70, 0.75), & 4. \end{cases}$$


Selection percentage
0
10
20
30
40
50
2
3
4
6
8
10
30
50
Toxicity Probabilities of Doses 1−8


Selection percentage
0
20
40
60
80
2
3
5
6
7
9
10
30
Toxicity Probabilities of Doses 1−8

Multiple Skeletons

- To avoid subjectivity in the specification of the skeleton, we propose prespecifying multiple skeletons, each representing a set of prior estimates of the toxicity probabilities.
- We view each skeleton as corresponding to a CRM model with a different set of p_j's.
- To accommodate the uncertainty in the specification of these skeletons, we take a Bayesian model averaging (BMA) approach to average $\hat{\pi}_j$ across the CRM models to obtain the BMA estimate of the toxicity probability for dose level j.

- BMA is known to provide a better predictive performance than any single model (Raftery, Madigan and Hoeting, 1997; and Hoeting et al., 1999).
- We incorporate the uncertainty in the prespecification of the toxicity probabilities into the estimation procedure such that the potential estimation bias caused by a misspecification of the p_j's can be averaged out.

BMA

- Let $(M_1, \ldots, M_K)$ be the models corresponding to each set of prior guesses of the toxicity probabilities $\{(p_{11}, \ldots, p_{1J}), \ldots, (p_{K1}, \ldots, p_{KJ})\}$.
- Model M_k $(k = 1, \ldots, K)$ in the CRM is given by

$$\pi_{kj}(\alpha_k) = p_{kj}^{\exp(\alpha_k)}, \quad j = 1, \ldots, J,$$

  which is based on the kth skeleton $(p_{k1}, \ldots, p_{kJ})$.
- Let $\text{pr}(M_k)$ be the prior probability that model M_k is the true model.

Posterior Model Probability

- The likelihood function under model M_k is

$$L(D|\alpha_k, M_k) = \prod_{j=1}^{J} \{p_{kj}^{\exp(\alpha_k)}\}^{y_j} \{1 - p_{kj}^{\exp(\alpha_k)}\}^{n_j - y_j}.$$

- The posterior model probability for M_k is given by

$$\text{pr}(M_k|D) = \frac{L(D|M_k)\text{pr}(M_k)}{\sum_{i=1}^{K} L(D|M_i)\text{pr}(M_i)}$$

where $L(D|M_k)$ is the marginal likelihood under M_k,

$$L(D|M_k) = \int L(D|\alpha_k, M_k) f(\alpha_k|M_k) d\alpha_k.$$

BMA Estimate

- The BMA estimate for the toxicity probability is

$$\bar{\pi}_j = \sum_{k=1}^{K} \hat{\pi}_{kj} \text{pr}(M_k|D), \quad j = 1, \ldots, J,$$

where $\hat{\pi}_{kj}$ is the posterior mean of the toxicity probability of dose level j under model M_k, i.e.,

$$\hat{\pi}_{kj} = \int p_{kj}^{\exp(\alpha_k)} \frac{L(D|\alpha_k, M_k) f(\alpha_k|M_k)}{\int L(D|\alpha_k, M_k) f(\alpha_k|M_k) d\alpha_k} d\alpha_k.$$

- By assigning $\hat{\pi}_{kj}$ a weight of $\text{pr}(M_k|D)$, the BMA method automatically favors the best fitting model, thus $\bar{\pi}_j$ is close to the best estimate.

Dose-finding Algorithm

- Patients in the first cohort are treated at the lowest dose d_1, or the physician-specified dose.
- At the current dose level j^{curr}, we obtain the BMA estimates for the toxicity probabilities, $\bar{\pi}_j$ $(j = 1, \ldots, J)$, based on the cumulated data.
- We then find dose level j^* that has a toxicity probability closest to ϕ,

$$j^* = \text{argmin}_{j \in (1,\ldots,J)} |\bar{\pi}_j - \phi|.$$

  - If $j^{\text{curr}} > j^*$, we de-escalate the dose level to $j^{\text{curr}} - 1$;

  - if $j^{\text{curr}} < j^*$, we escalate the dose level to $j^{\text{curr}} + 1$;
  - otherwise, the dose stays at the same level as j^{curr} for the next cohort of patients.

- Once the maximum sample size is reached, the dose that has the toxicity probability closest to ϕ is selected as the MTD.

- We require an early termination of a trial if the lowest dose is too toxic,

$$\sum_{k=1}^{K} \text{pr}\{\pi_{k1}(\alpha_k) > \phi | M_k, D\} \text{pr}(M_k|D) > 90\%.$$

Occam's Window

- If the fit of a model is far worse than the best fitting model, it would be reasonable to exclude that model from the model averaging set.
- Only if model M_k satisfies

$$\frac{\text{pr}(M_k|D)}{\max_{i\in(1,\ldots,K)}\text{pr}(M_i|D)} > \delta,$$

model M_k is included in the model averaging set.

CRM Model Selection

- Model selection takes a different perspective in regression models.
- Among a set of competing models, we can simply select the best fitting model according to a suitable model selection criterion.
- A natural candidate for the model selection criterion is based on the posterior model probability.

Simulation Studies

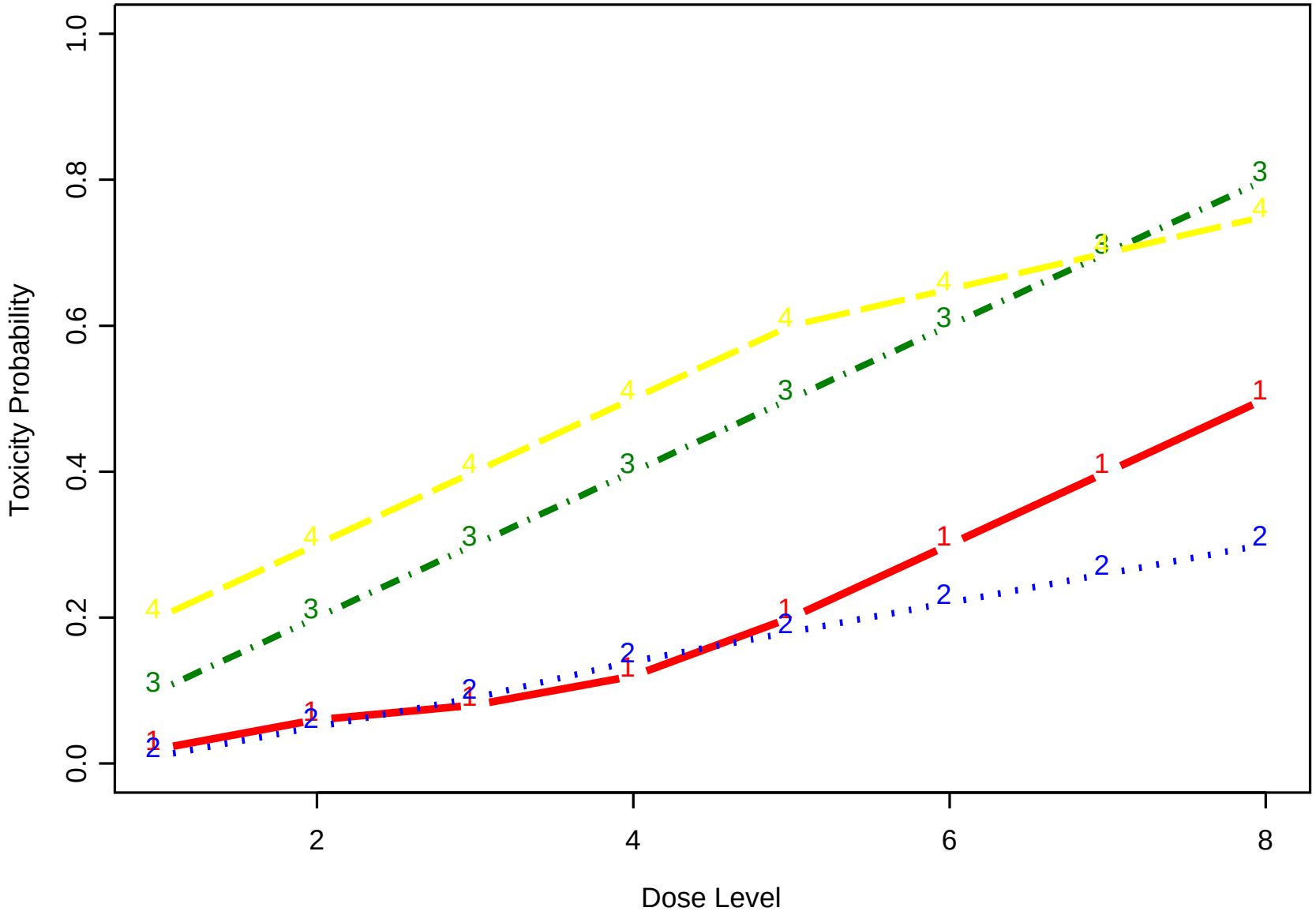
- We considered eight doses and assumed that toxicity monotonically increased with the dose.
- We prepared four sets of initial guesses of the toxicity probabilities:

$$(p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8)$$

$$= \begin{cases} (0.02, 0.06, 0.08, 0.12, 0.20, 0.30, 0.40, 0.50), & 1 \\ (0.01, 0.05, 0.09, 0.14, 0.18, 0.22, 0.26, 0.30), & 2 \\ (0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80), & 3 \\ (0.20, 0.30, 0.40, 0.50, 0.60, 0.65, 0.70, 0.75), & 4. \end{cases}$$

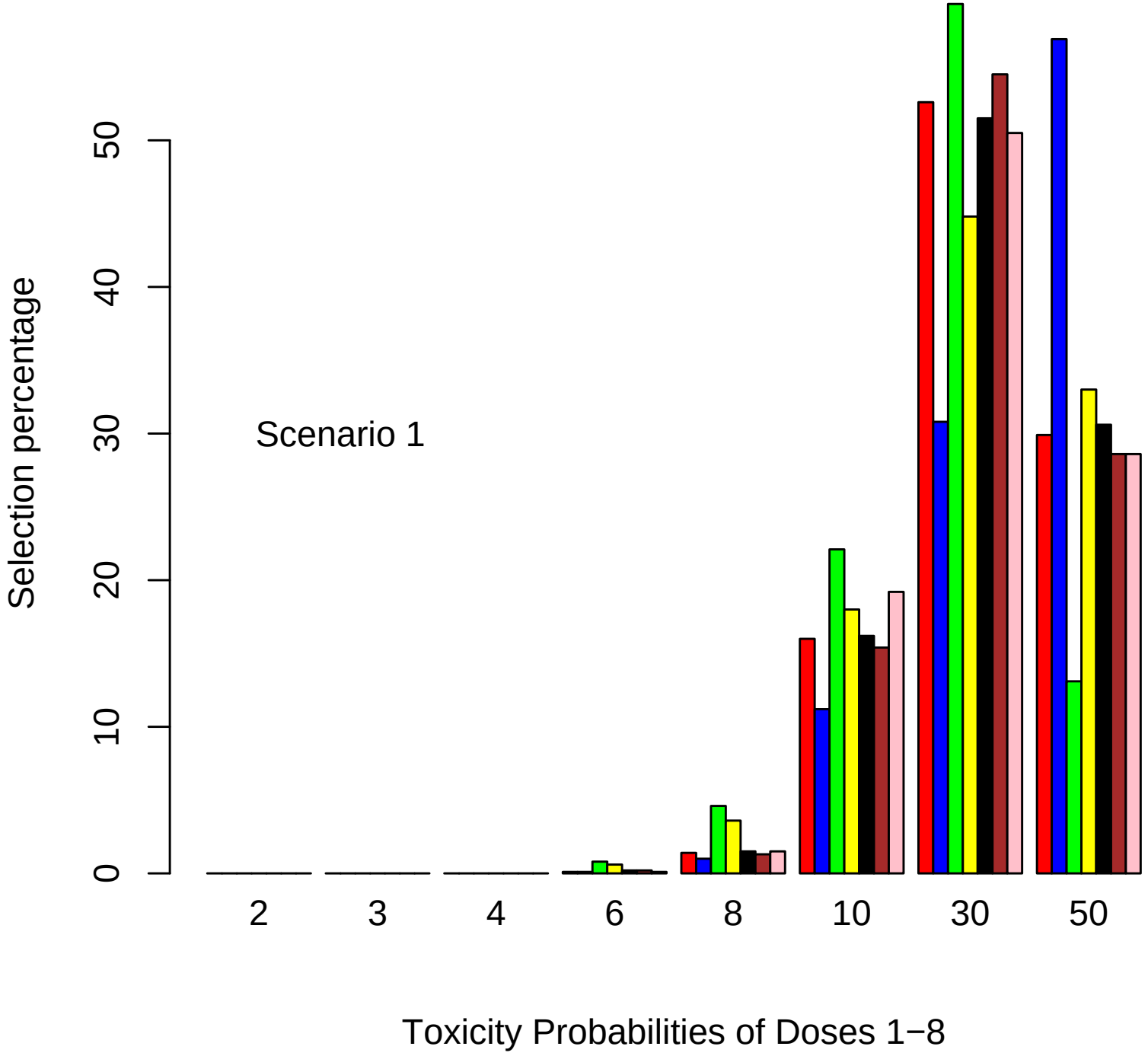
Skeletons

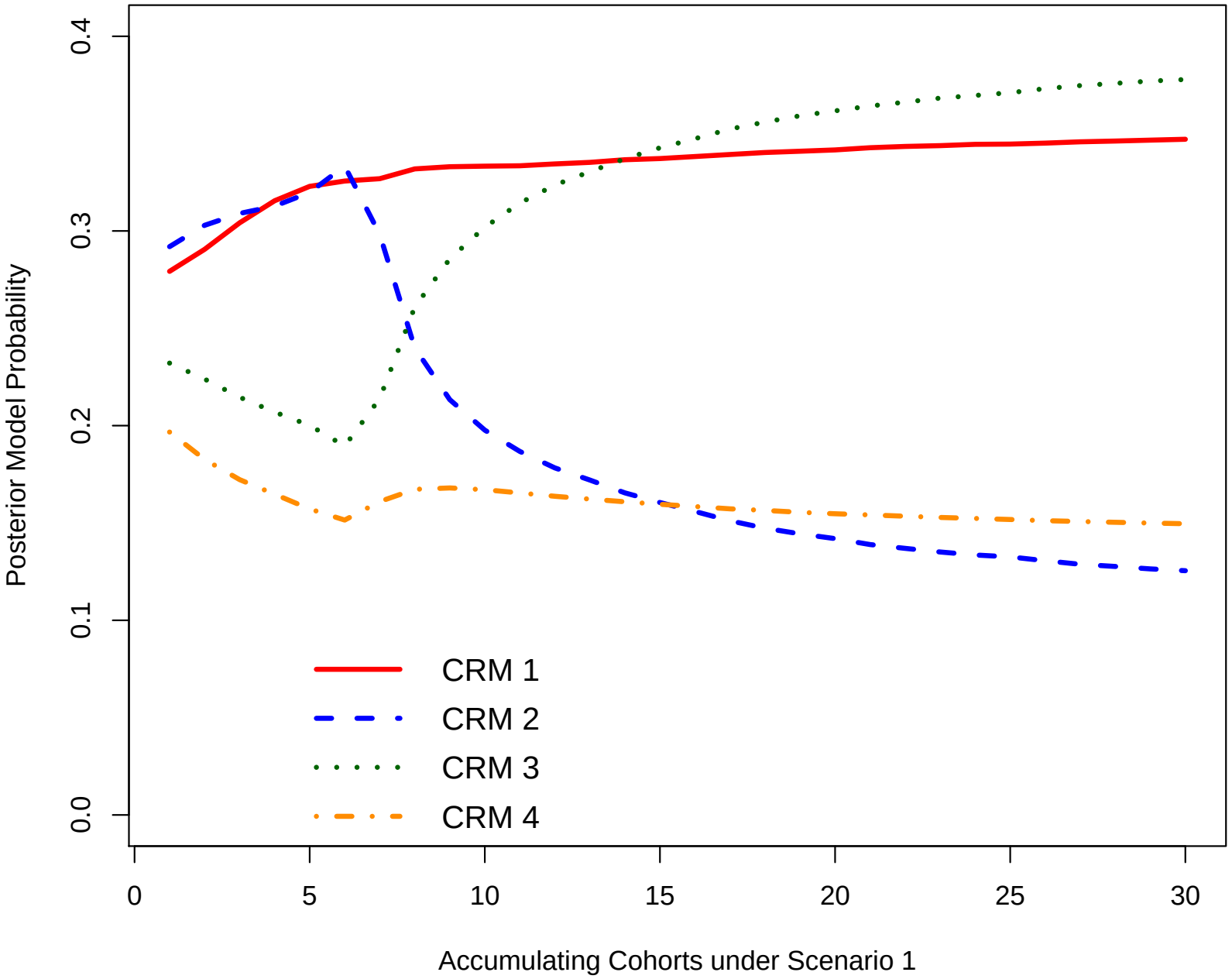
- Skeleton 1 is for the case in which toxicity increases slowly at the low doses but increases quickly at the high doses.
- Skeleton 2 is more concentrated at the low toxicity levels (toxicity probabilities ≤ 0.3).
- Skeleton 3 has the toxicity probabilities evenly spread over a range of 0.1 up to 0.8.
- Skeleton 4 starts at a relatively high toxicity probability of 0.2, and increases quickly at the low doses before leveling off at the high doses.

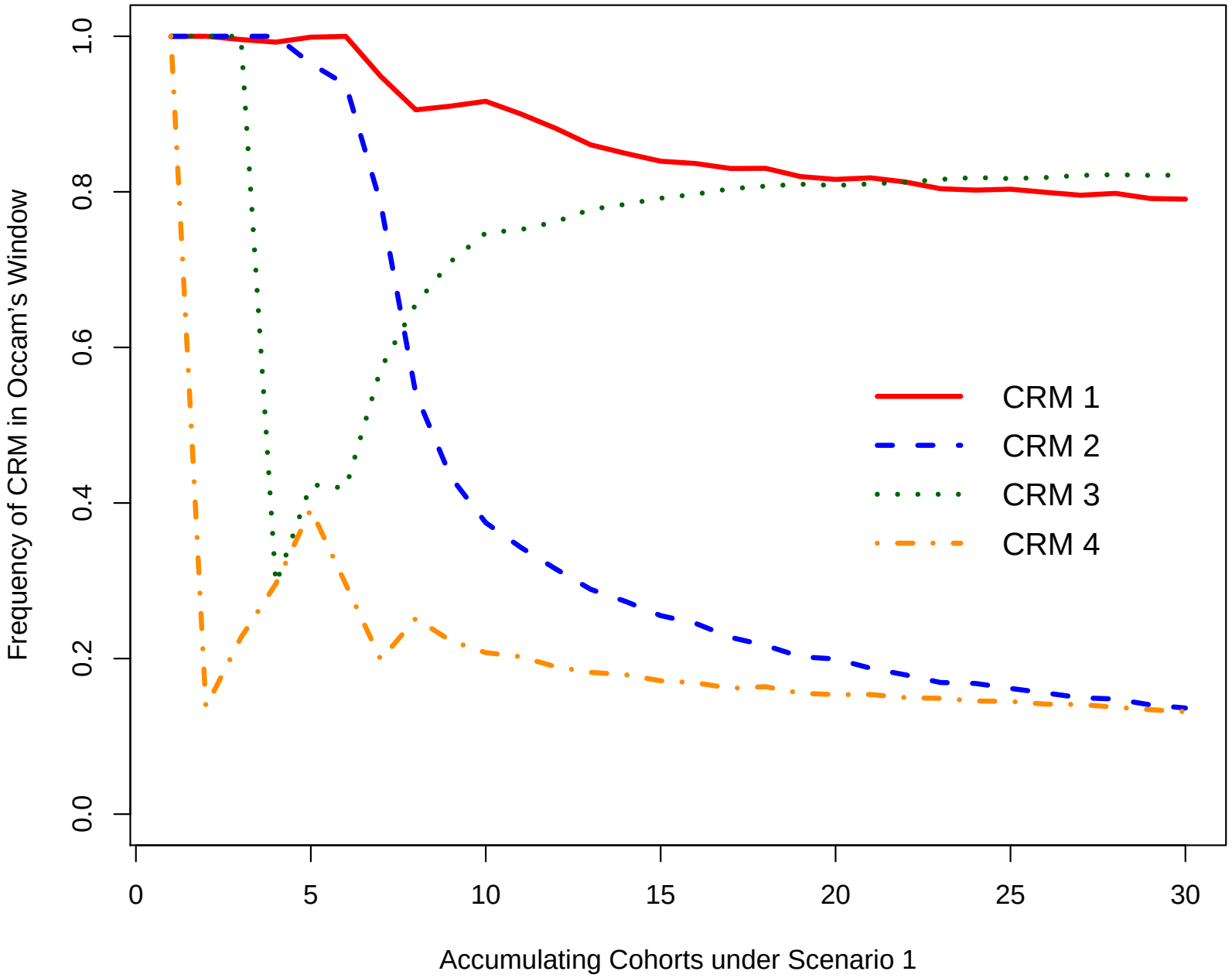

Toxicity Probability
0.0
0.2
0.4
0.6
0.8
1.0
2
4
6
8
Dose Level

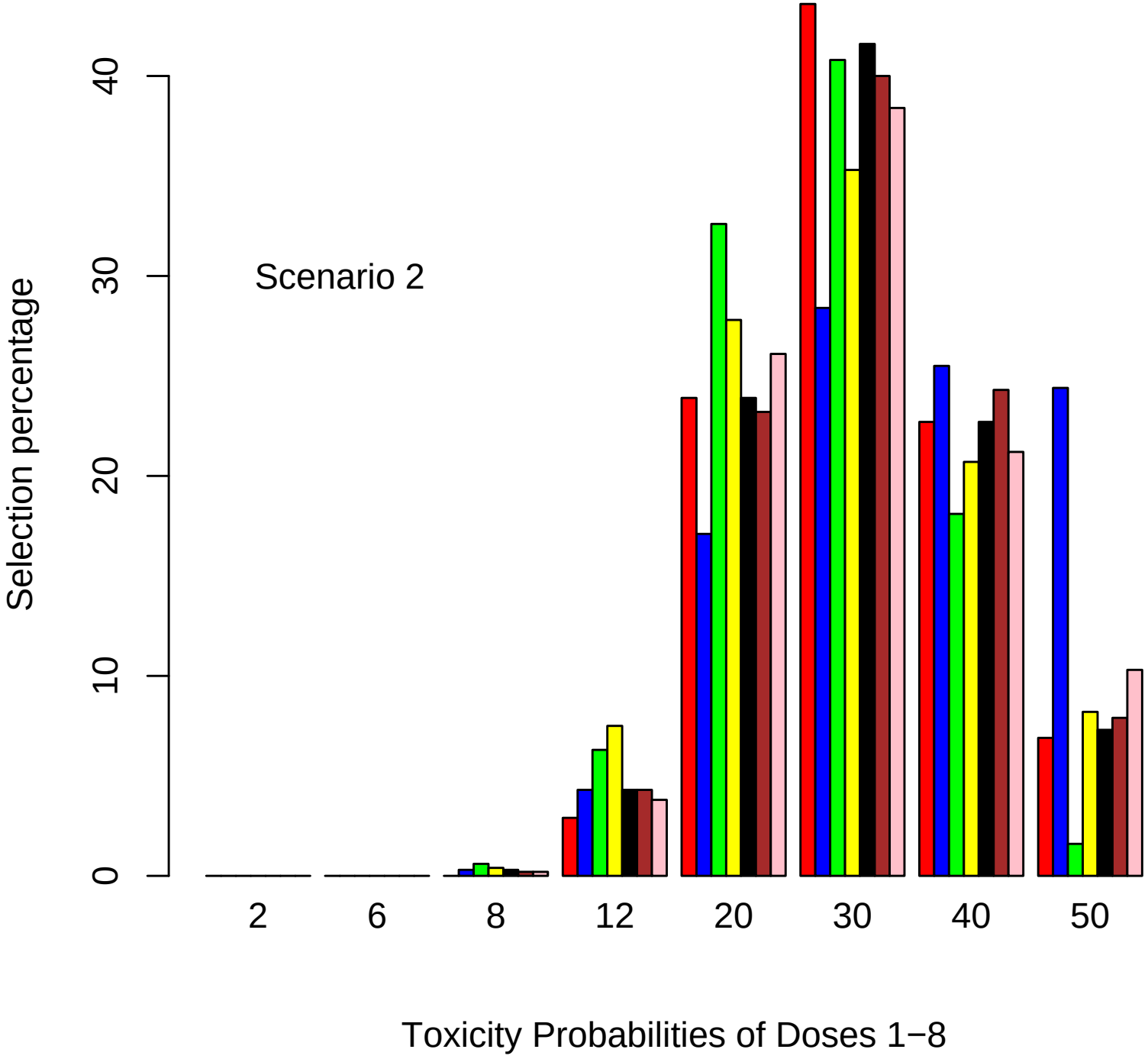
Simulation Setups

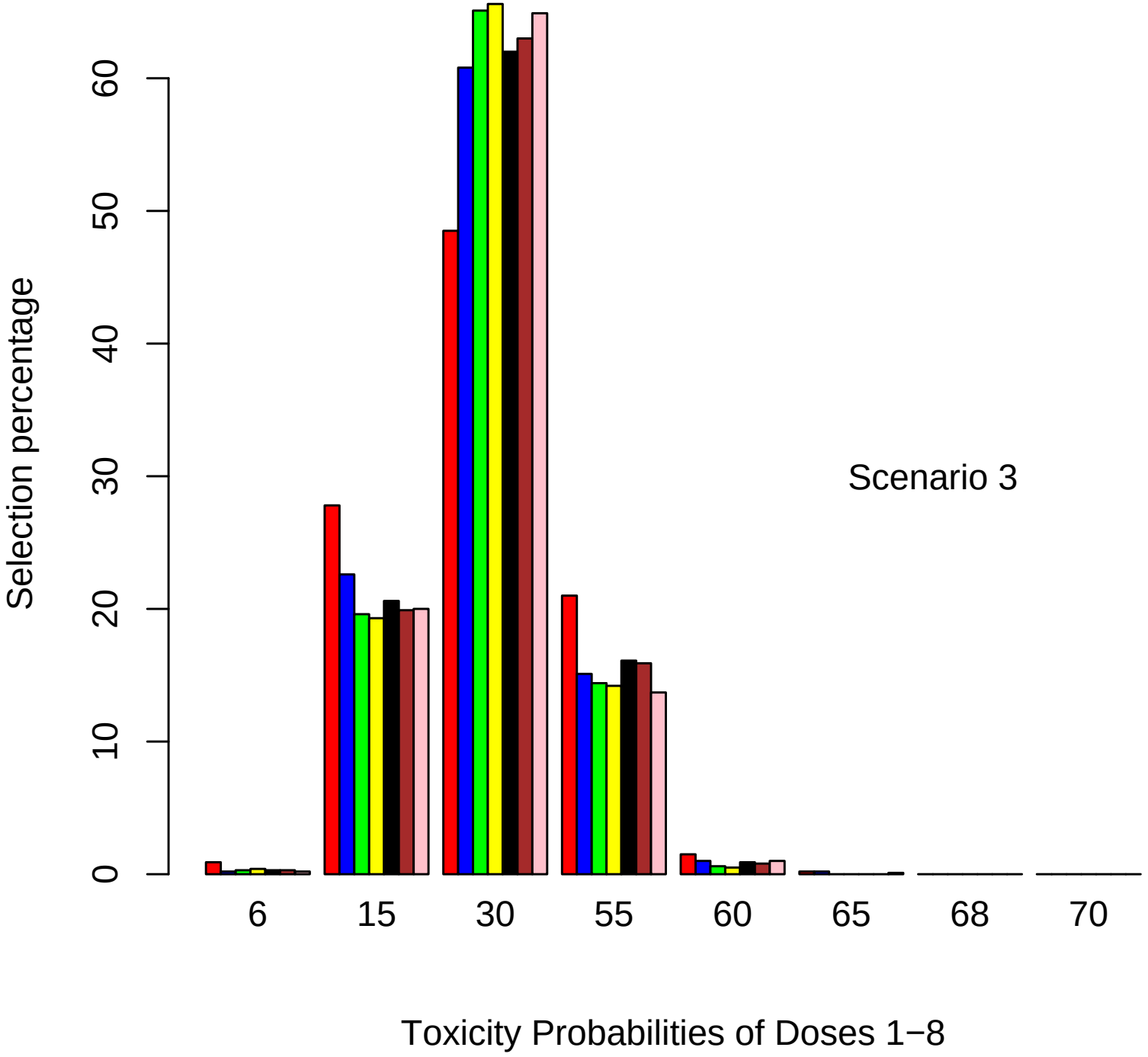
- The target toxic probability was $\phi = 30\%$.
- We took the prior distribution of $\alpha \sim N(0, 4)$, the prior model probability $\text{pr}(M_k) = 1/4$ for $k = 1, \ldots, 4$.
- We took the cohort size 3, and treated the first cohort of patients at the lowest dose level.
- The maximum sample size was 30, and for each scenario we carried out 10,000 simulated trials.

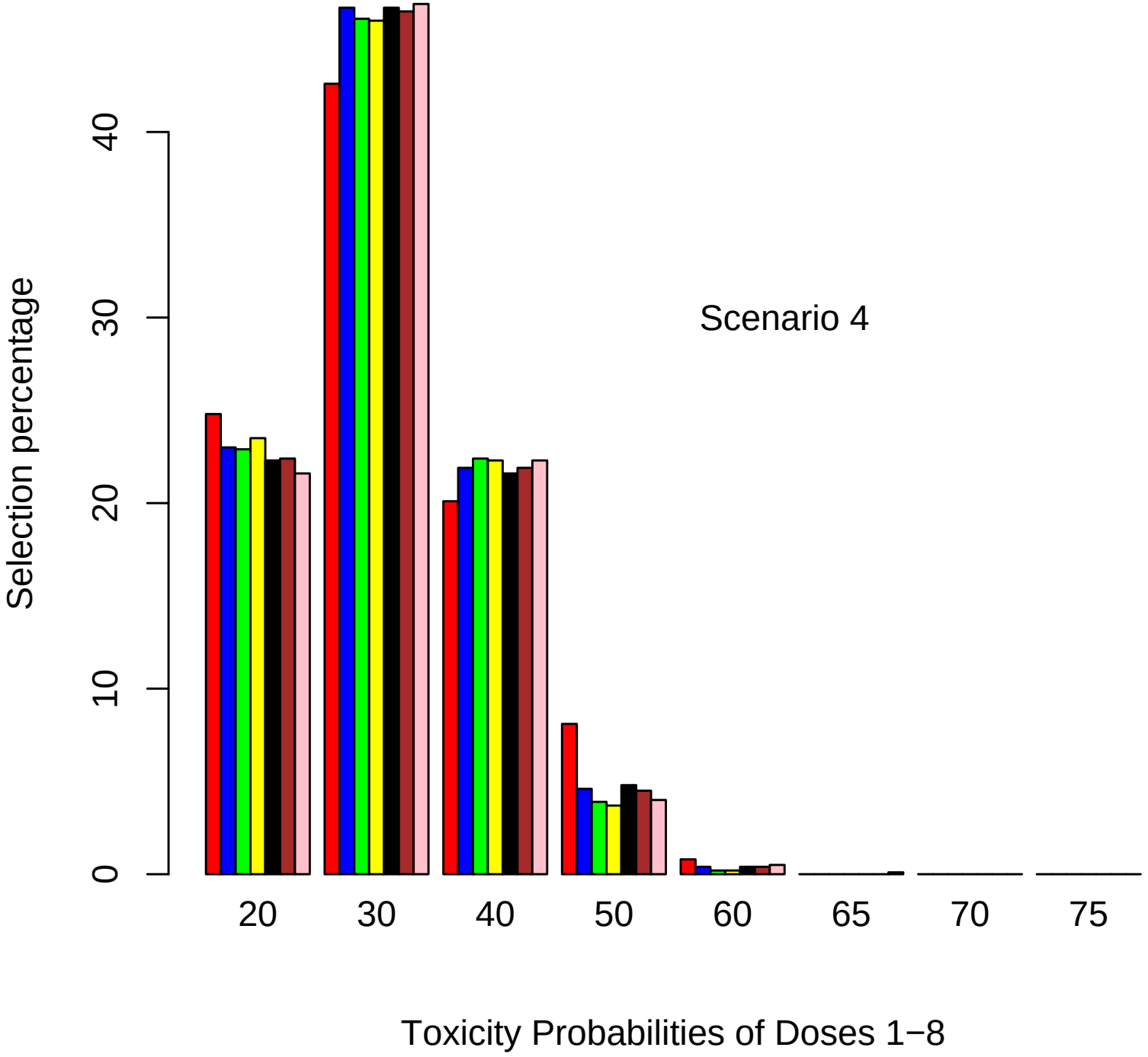

Selection percentage
0
10
20
30
40
50
Scenario 1
2
3
4
6
8
10
30
50
Toxicity Probabilities of Doses 1−8

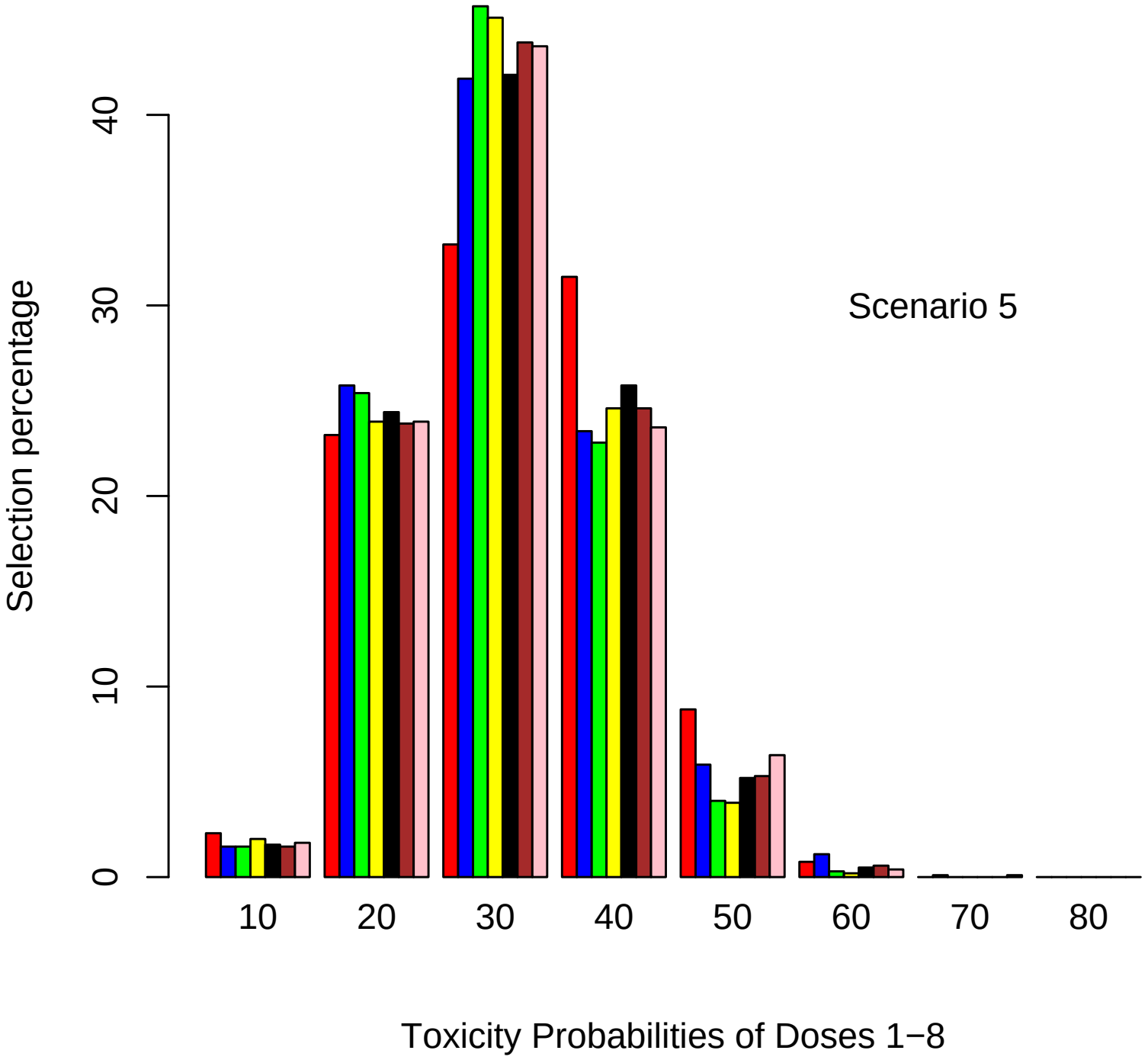

Posterior Model Probability
0.0
0.1
0.2
0.3
0.4
0
5
10
15
20
25
30
Accumulating Cohorts under Scenario 1
CRM 1
CRM 2
CRM 3
CRM 4


Frequency of CRM in Occam's Window
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
Accumulating Cohorts under Scenario 1
CRM 1
CRM 2
CRM 3
CRM 4


Scenario 2
Selection percentage
0
10
20
30
40
2
6
8
12
20
30
40
50
Toxicity Probabilities of Doses 1−8

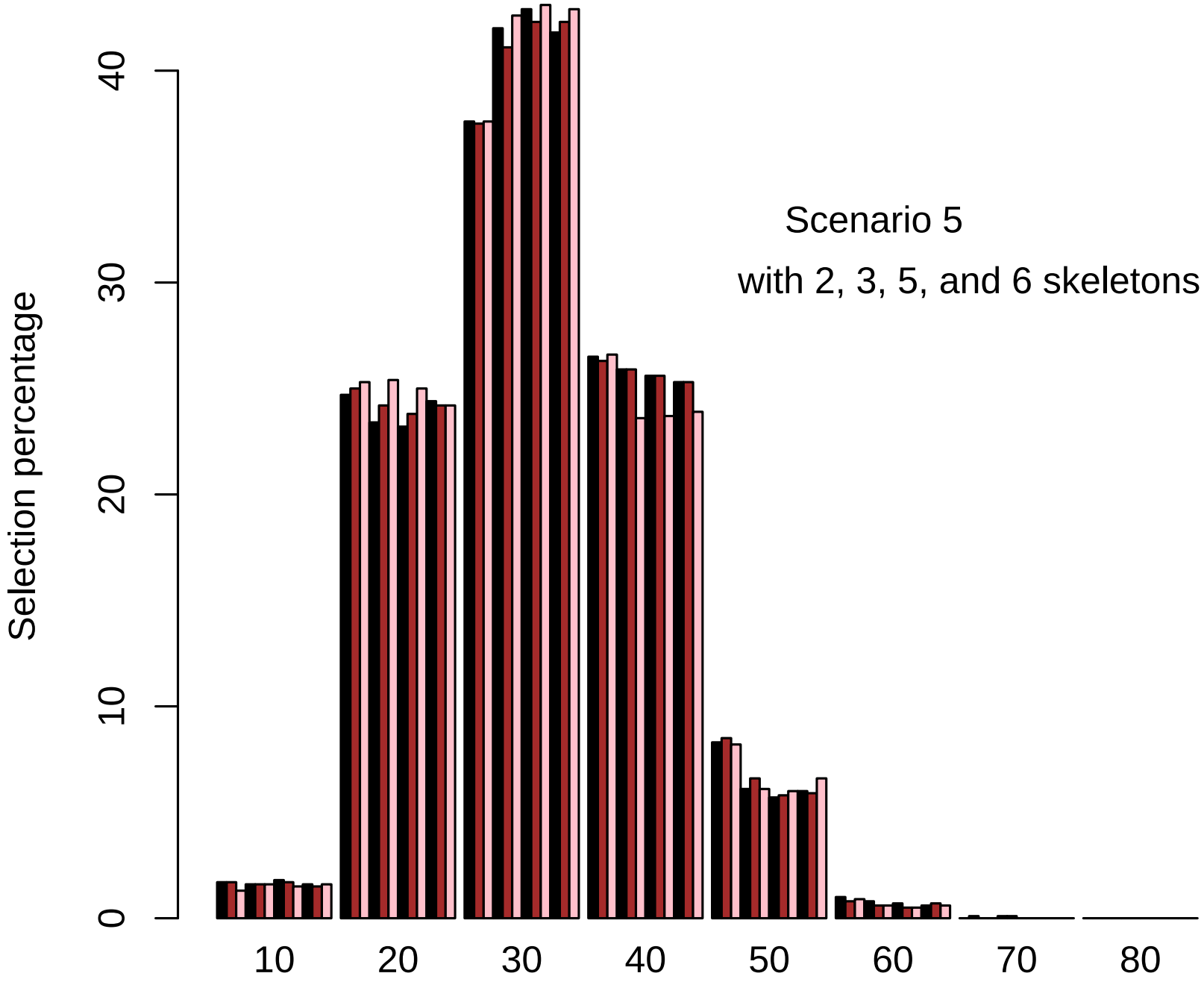

Selection percentage
0
10
20
30
40
50
60
Scenario 3
6
15
30
55
60
65
68
70
Toxicity Probabilities of Doses 1−8


Scenario 4
Selection percentage
0
10
20
30
40
20
30
40
50
60
65
70
75
Toxicity Probabilities of Doses 1−8


Scenario 5
Selection percentage
0
10
20
30
40
10
20
30
40
50
60
70
80
Toxicity Probabilities of Doses 1−8

Number of Skeletons?

- Under scenario 5 we present the selection percentage at each dose when using one, two, three, four, five and six skeletons.
- Skeleton 5 =
  (0.08, 0.15, 0.21, 0.29, 0.37, 0.44, 0.51, 0.58),
  Skeleton 6 =
  (0.05, 0.10, 0.20, 0.25, 0.30, 0.40, 0.47, 0.55).


Scenario 5
with 2, 3, 5, and 6 skeletons
Selection percentage
0
10
20
30
40
10
20
30
40
50
60
70
80
Toxicity Probabilities of Doses 1−8

- In practice, we recommend using three skeletons in the trial design.

Conclusion

- We have proposed a new dose-finding algorithm based on the Bayesian model averaging CRM, using multiple sets of prespecified toxicity probabilities.
- The performance of the proposed designs can be substantially improved over that of the original CRM, if the skeleton in the CRM happens to be very far off the true model.

- The BMA-CRM method is straightforward to implement and very easy to compute based on the Gaussian quadrature approximation or the Markov chain Monte Carlo procedure.
- It provides a nice compromise for the initial guesses of toxicity probabilities from different physicians.
- This Bayesian model averaging procedure dramatically improves the robustness of the CRM.

Software

- We have developed user-friendly software to conduct BMA-CRM for actual trials.
- Free download from *http://biostatistics.mdanderson.org/SoftwareDownload/SingleSoftware.aspx?Software_Id* = 81

Thank You!