# The CRM for ordinal and multivariate outcomes

Elizabeth Garrett-Mayer, PhD
Emily Van Meter
Hollings Cancer Center
Medical University of South Carolina

# Outline

- Part 1:  Ordinal toxicity model
- Part 2:  Efficacy and Toxicity model (very brief)

# The gist

- Most of our adaptive dose finding designs for toxicity (alone) in single agent settings use binary endpoint
- But, toxicity is scored by an ordinal rating scale
- Would be do a better job titrating if we incorporated ordinal nature?
- Example: Consider a dose where three grade 2 toxicities occur
  - standard binary approach:  3 patients without DLTs
  - ordinal approach:  3 patients that were pretty close to grade 3
  - would an ordinal approach make us less likely to overdose the next cohort of patients?

# The gist (continued)

- Some proposals for ordinal-like toxicity outcomes
  - Bekele and Thall (2004)
    - all toxicities not created equal and not independent
    - in addition to grading, types are considered
    - total toxicity burden (TTB)
    - uses Bayesian multivariate ordinal probit regression
  - Yuan et al. (2007)
    - uses severity weights to convert grades to numeric scores
    - acknowledges multiple toxicities per patient
    - continuous toxicity value then converted to binary score
    - quasi-Bernoulli approach
    - depends heavily on conversion metric
  - Wang (2000)
    - count both grades 3 and 4 DLTs
    - but give grade 3 a lower weight than 4.

# Proportional Odds Model*

- Common regression approach to ordinal data
- Relatively efficient:
  - shared slope across categories
  - ordered intercepts differentiate between categories

$$P(Y_i \geq j \mid X_j) = \frac{1}{1+\exp(-(\alpha_j + X_j\beta)}$$

for $j = 1,2,3,4$

$Y_i$ = highest toxicity grade for patient $i$

$X_i$ = "dose" for patient $i$

$\alpha_1 \geq \alpha_2 \geq \alpha_3 \geq \alpha_4$

*McCullagh, 1980

# Proportional Odds Model

- Considers grade 0 (no toxicity) through 4
- An easy modification from the logistic CRM with binary outcome
  - Goodman et al, (1995): one parameter
  - Piantadosi et al. (1998): two parameter
- May not perfectly model the full dose-toxicity model.
- But, that is not so critical: the real questions are
  - is it robust for dose finding when the underlying model is not a proportional odds model?
  - are fewer patients exposed to toxic dose levels?
  - do fewer patients have DLTs?

# Scenarios for expected improvements

- Toxic treatments
- Prior assumptions about toxicity underestimate the true dose-toxicity relationship

# Simulations

- Trials simulated using a multinomial distribution
- All simulations were conducted in R
- Per scenario
  - 2000 datasets for each simulation scheme
  - 50/50 and 70/30 prior weighting schemes
  - Cohort size and sample size combinations of 3/30, 2/20, and 3/21 respectively
- Assess design under 2 different priors
- 4 underlying true dose-response models (2 PO models and 2 that violate PO assumptions)

# Estimation Approach

- Extension of Piantadosi's "practical" CRM
  - Maximum likelihood estimation of two-parameter logistic regression model
- Pseudo-data is used
  - Piantadosi includes two pseudo-datapoints: one at high and one at low dose level
  - outcomes are probabilities of toxicity at high and low doses
- Unlike Piantadosi:
  - to initialize the POM and make it estimable, we need pseudo-data as a "pseudo-prior" in all categories
  - simulate "large" datasets to represent prior
  - BUT, give the observations each very low weight

# Prior Weighting Schemes

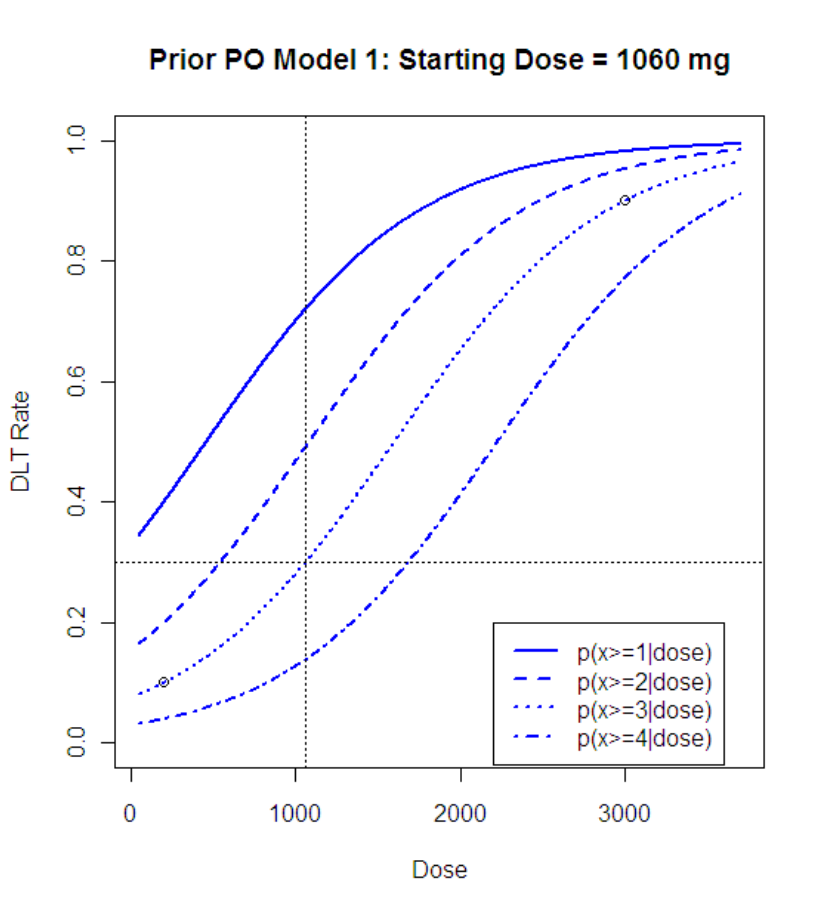
With each newly simulated cohort, the weighting scheme will update so that the weight of the prior clinical estimates continues to decrease with the addition of more simulated data

| Weights | Iteration | Prior Data Weight | Number of Prior Data | Simulated Data Weight | Number of Simulated Data | Percentage Simulated Data |
|---|---|---|---|---|---|---|
| 50-50 | 1 | 0.005000 | 100 | 0.166670 | 3 | 50.00 |
| | 2 | 0.003333 | 100 | 0.111111 | 6 | 66.67 |
| | 3 | 0.002500 | 100 | 0.083334 | 9 | 75.00 |
| | 4 | 0.002000 | 100 | 0.066667 | 12 | 80.00 |
| | 5 | 0.001667 | 100 | 0.055556 | 15 | 83.33 |
| | 6 | 0.001429 | 100 | 0.047619 | 18 | 85.71 |
| | 7 | 0.001250 | 100 | 0.041667 | 21 | 87.50 |
| | 8 | 0.001111 | 100 | 0.037034 | 24 | 88.88 |
| | 9 | 0.001000 | 100 | 0.033333 | 27 | 90.00 |
| | 10 | 0.000909 | 100 | 0.030303 | 30 | 90.91 |
| 70-30 | 1 | 0.003000 | 100 | 0.233333 | 3 | 70.00 |
| | 2 | 0.001765 | 100 | 0.137255 | 6 | 82.35 |
| | 3 | 0.001250 | 100 | 0.097222 | 9 | 87.50 |
| | 4 | 0.000968 | 100 | 0.075269 | 12 | 90.32 |
| | 5 | 0.000789 | 100 | 0.061403 | 15 | 92.10 |
| | 6 | 0.000667 | 100 | 0.051852 | 18 | 93.33 |
| | 7 | 0.000577 | 100 | 0.044872 | 21 | 94.23 |
| | 8 | 0.000508 | 100 | 0.034548 | 24 | 82.92 |
| | 9 | 0.000455 | 100 | 0.035353 | 27 | 95.45 |
| | 10 | 0.000411 | 100 | 0.031963 | 30 | 95.89 |

# “Prior” Models

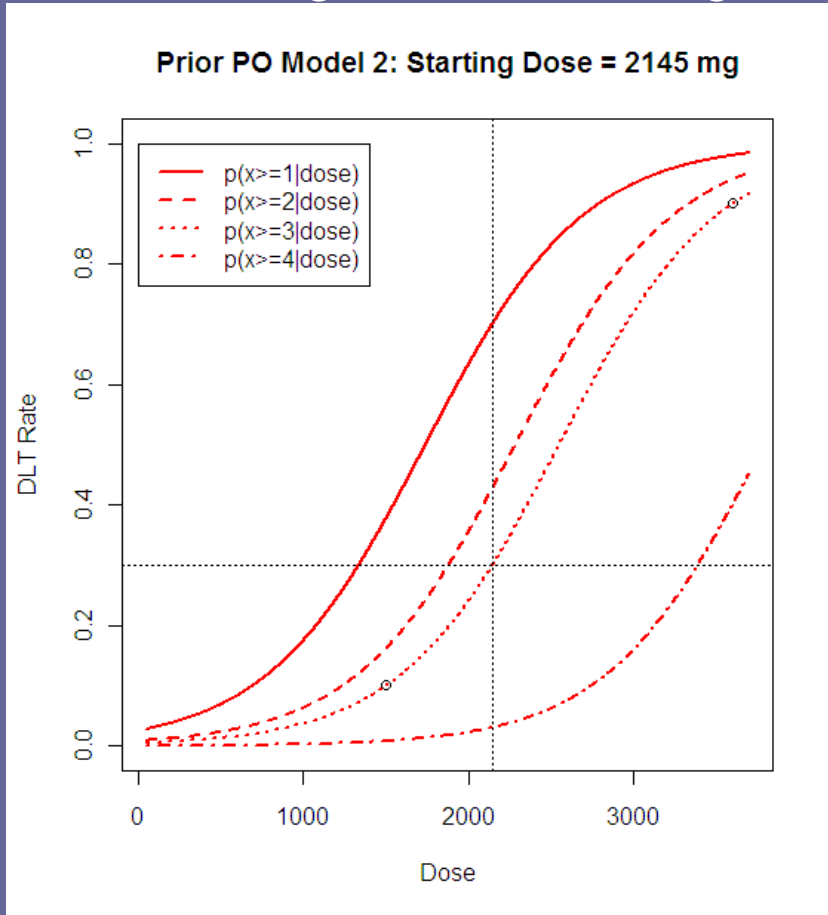
Prior #1:

- 10% DLT rate at 200 mg
- 90% DLT rate at 3000 mg
- Starting dose = 1060 mg




Prior #2:

- 10% DLT rate at 1500 mg
- 90% DLT rate at 3600 mg
- Starting dose = 2145 mg



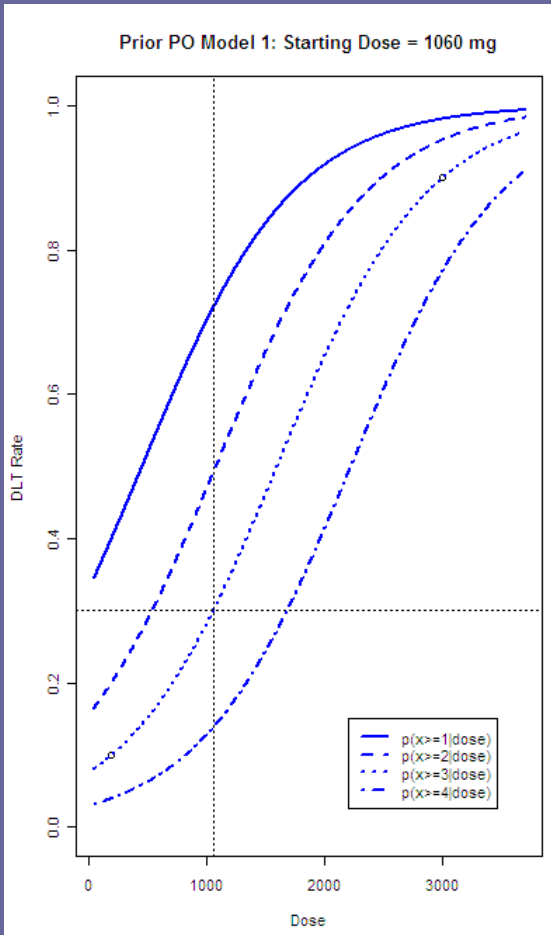
# Interesting results to consider

- Percentage of trials estimating final dose within 250 mg and 400 mg of the true MTD
- Percentage of trials estimating final dose at levels highly toxic (>40% DLT) or suboptimal (<20%DLT)
- Percentage of patients exposed to dose levels highly toxic (>40% DLT) or suboptimal (<20%DLT)
- Percentage of patients with a DLT or with a “non-DLT” grade 1 or 2 toxicity
- Nominal coverage of 95% confidence interval for the final estimated dose as a measure of accuracy (using the delta method)

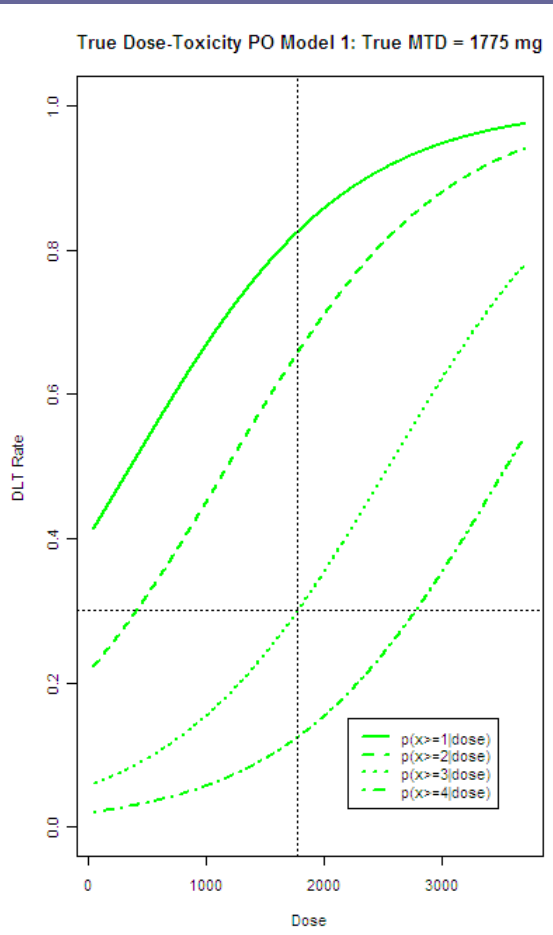
# Scenario A

Fit the true underlying dose-toxicity relationship to PO model #1, Prior #1, Target dose-limiting toxicity (DLT) rate = 30%

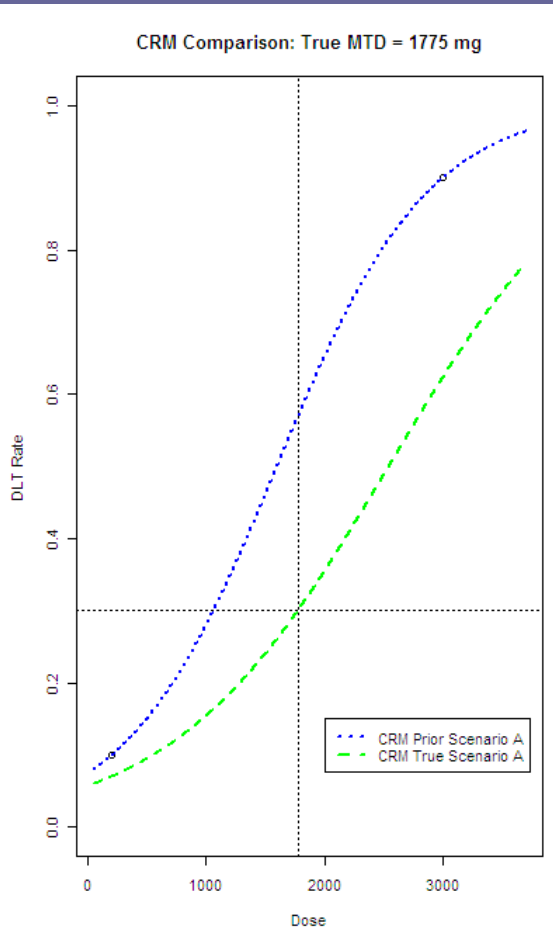
PRIOR




TRUE MODEL




BINARY MODEL



# Scenario A Simulation Results

*Scenario A - 30% Target DLT Rate, Actual MTD = 1775 mg*

| | PO* 50-50 | CRM 50-50 | PO 50-50 | CRM 50-50 | PO 50-50 | CRM 50-50 | PO 70-30 | CRM 70-30 | PO 70-30 | CRM 70-30 | PO 70-30 | CRM 70-30 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Total Sample Size | 30 | 30 | 20 | 20 | 21 | 21 | 30 | 30 | 20 | 20 | 21 | 21 |
| Patients per Cohort | 3 | 3 | 2 | 2 | 3 | 3 | 3 | 3 | 2 | 2 | 3 | 3 |
| Mean Dose | 1690 | 1655 | 1679 | 1678 | 1651 | 1618 | 1702 | 1666 | 1701 | 1676 | 1683 | 1646 |
| **% of CIs that include the True MTD** | **86.90** | **69.85** | **85.75** | **63.15** | **90.00** | **72.80** | **82.55** | **58.05** | **77.70** | **46.50** | **83.30** | **62.95** |
| % of trials with recommended dose within 250 mg of true dose | 54.05 | 53.25 | 47.95 | 45.20 | 47.70 | 46.70 | 49.70 | 48.80 | 40.65 | 41.55 | 45.90 | 43.70 |
| % of trials with recommended dose within 400 mg of true dose | 76.60 | 73.70 | 67.80 | 66.95 | 69.65 | 68.25 | 71.80 | 71.55 | 61.90 | 62.20 | 64.00 | 65.05 |
| % of trials with recommended dose at DLT rate of >40% | 6.35 | 6.15 | 10.20 | 10.20 | 7.85 | 7.05 | 9.50 | 7.30 | 14.65 | 13.65 | 12.30 | 9.60 |
| % of trials with recommended dose at DLT rate of <20% | 10.90 | 13.65 | 16.55 | 17.60 | 17.15 | 18.35 | 12.90 | 14.50 | 18.45 | 18.50 | 17.05 | 19.05 |
| Average % of patients treated at doses with >40% DLT rate | 6.07 | 4.47 | 8.62 | 8.35 | 5.04 | 3.84 | 8.05 | 6.54 | 11.43 | 9.54 | 6.59 | 5.27 |
| Average % of patients treated at doses <20% DLT rate | 30.43 | 30.92 | 32.69 | 32.52 | 37.79 | 38.40 | 31.12 | 31.25 | 33.73 | 33.38 | 37.72 | 37.01 |
| Average % of patients with DLT (grade 3 or 4) | 0.25 | 0.25 | 0.25 | 0.25 | 0.24 | 0.24 | 0.26 | 0.25 | 0.25 | 0.25 | 0.24 | 0.24 |
| Average % of patients with a non-DLT (grade 1 or 2) | 51.47 | NA | 50.55 | NA | 51.83 | NA | 51.11 | NA | 50.71 | NA | 51.44 | NA |

*Proportional Odds Model

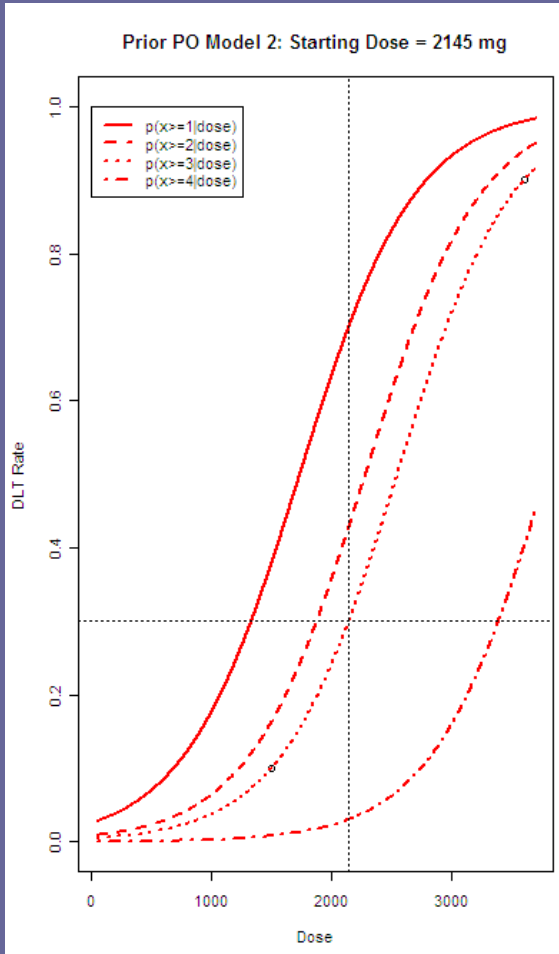
# Scenario B

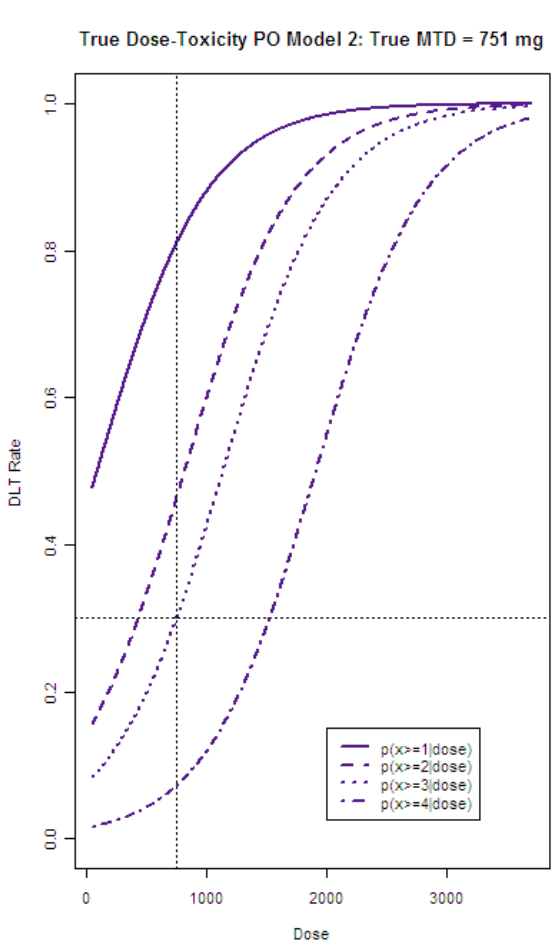
Fit the true underlying dose-toxicity relationship to PO model #2, Prior #2, Target DLT rate = 30%

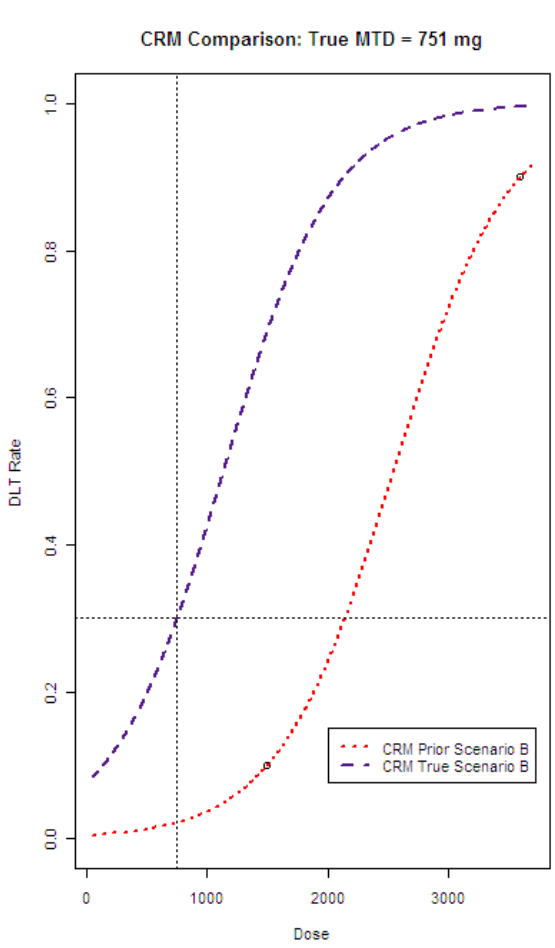
PRIOR

TRUE MODEL

BINARY MODEL



# Scenario B Simulation Results

*Scenario B - 30% Target DLT Rate, Actual MTD = 751 mg*

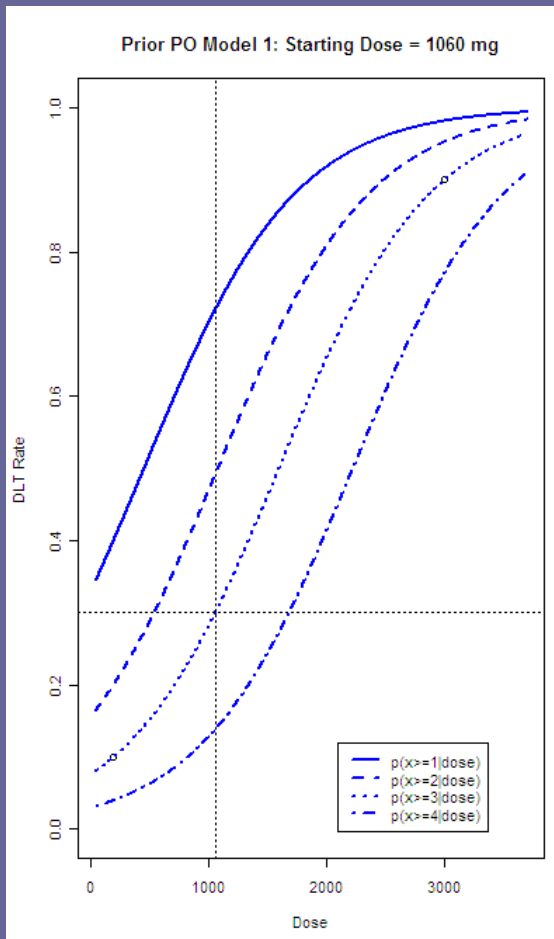
| | PO* 50-50 | CRM 50-50 | PO 50-50 | CRM 50-50 | PO 50-50 | CRM 50-50 | PO 70-30 | CRM 70-30 | PO 70-30 | CRM 70-30 | PO 70-30 | CRM 70-30 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Total Sample Size | 30 | 30 | 20 | 20 | 21 | 21 | 30 | 30 | 20 | 20 | 21 | 21 |
| Patients per Cohort | 3 | 3 | 2 | 2 | 3 | 3 | 3 | 3 | 2 | 2 | 3 | 3 |
| Mean Dose | 769 | 771 | 767 | 789 | 776 | 784 | 759 | 764 | 747 | 789 | 766 | 793 |
| **% of CIs that include the True MTD** | **99.10** | **93.65** | **98.15** | **89.85** | **98.55** | **92.40** | **97.00** | **88.95** | **94.85** | **84.70** | **96.55** | **87.55** |
| % of trials with recommended dose within 250 mg of true dose | 74.55 | 72.15 | 67.95 | 62.50 | 64.90 | 62.85 | 78.30 | 71.55 | 69.95 | 61.90 | 67.90 | 62.00 |
| % of trials with recommended dose within 400 mg of true dose | 94.20 | 92.00 | 88.00 | 82.15 | 85.80 | 84.60 | 95.75 | 90.70 | 88.10 | 83.45 | 88.20 | 81.25 |
| % of trials with recommended dose at DLT rate of >40% | 20.60 | 22.20 | 24.05 | 27.75 | 25.30 | 28.85 | 16.45 | 21.45 | 19.40 | 28.85 | 23.65 | 29.85 |
| % of trials with recommended dose at DLT rate of <20% | 11.60 | 12.95 | 14.85 | 15.15 | 15.65 | 15.25 | 10.50 | 13.35 | 15.55 | 15.20 | 15.60 | 16.25 |
| Average % of patients treated at doses with >40% DLT rate | 32.88 | 48.55 | 36.23 | 52.91 | 37.63 | 57.81 | 30.32 | 49.76 | 34.70 | 51.19 | 34.26 | 59.80 |
| Average % of patients treated at doses <20% DLT rate | 22.19 | 17.51 | 25.78 | 17.51 | 25.91 | 18.75 | 24.28 | 18.25 | 26.50 | 18.23 | 28.38 | 19.02 |
| Average % of patients with DLT (grade 3 or 4) | 36.41 | 42.67 | 37.30 | 44.40 | 38.47 | 47.11 | 35.52 | 44.06 | 37.17 | 44.91 | 37.40 | 48.74 |
| Average % of patients with a non-DLT (grade 1 or 2) | 43.26 | NA | 42.05 | NA | 41.04 | NA | 43.22 | NA | 41.56 | NA | 41.03 | NA |

*Proportional Odds Model

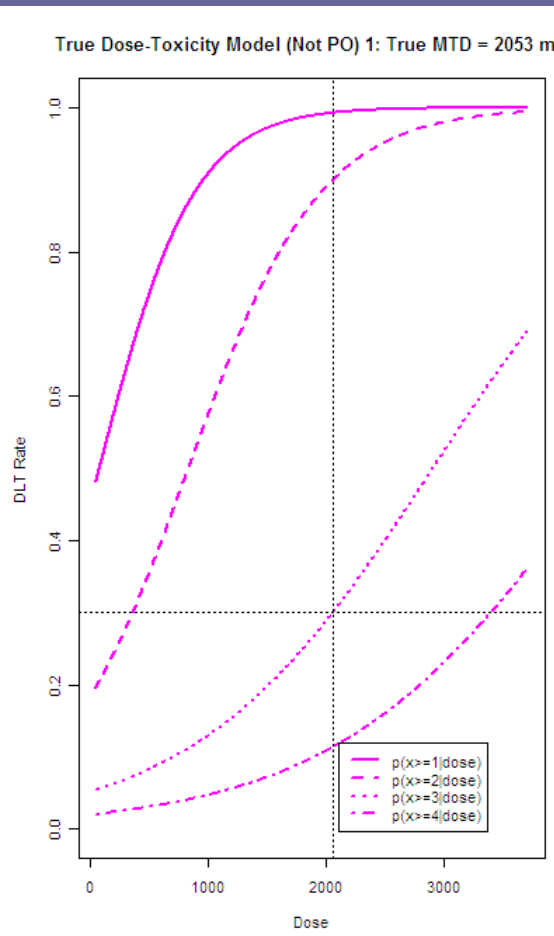
# Scenario C

Fit the true underlying dose-toxicity relationship to not a PO model #1, Prior #1, Target DLT rate = 30%

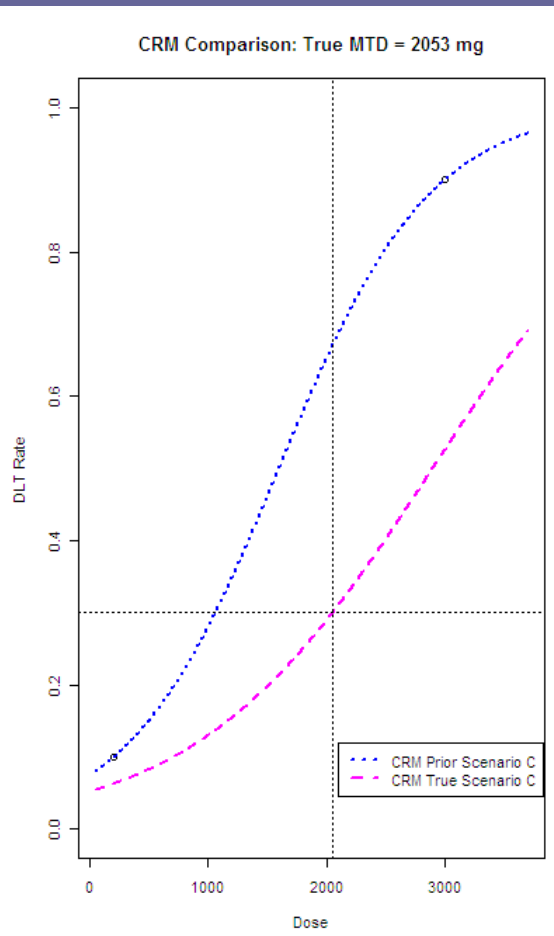
PRIOR




TRUE MODEL




BINARY MODEL



# Scenario C Simulation Results

*Scenario C - Underlying Dose-Toxicity Model Not PO 1, 30% Target DLT Rate, Actual MTD = 2053 mg*

| | PO* 50-50 | CRM 50-50 | PO 50-50 | CRM 50-50 | PO 50-50 | CRM 50-50 | PO 70-30 | CRM 70-30 | PO 70-30 | CRM 70-30 | PO 70-30 | CRM 70-30 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Total Sample Size | 30 | 30 | 20 | 20 | 21 | 21 | 30 | 30 | 20 | 20 | 21 | 21 |
| Patients per Cohort | 3 | 3 | 2 | 2 | 3 | 3 | 3 | 3 | 2 | 2 | 3 | 3 |
| Mean Dose | 1871 | 1833 | 1869 | 1858 | 1815 | 1804 | 1868 | 1874 | 1873 | 1838 | 1845 | 1809 |
| **% of CIs that include the True MTD** | **72.30** | **63.90** | **71.30** | **61.40** | **74.75** | **67.75** | **65.50** | **60.20** | **61.40** | **46.30** | **66.70** | **56.40** |
| % of trials with recommended dose within 250 mg of true dose | 49.00 | 42.70 | 43.25 | 40.75 | 42.55 | 41.10 | 44.85 | 44.25 | 38.15 | 35.45 | 37.70 | 38.95 |
| % of trials with recommended dose within 400 mg of true dose | 71.10 | 64.70 | 63.60 | 59.80 | 63.50 | 60.50 | 66.40 | 63.85 | 57.50 | 53.45 | 59.85 | 57.20 |
| % of trials with recommended dose at DLT rate of >40% | 2.50 | 3.45 | 6.45 | 7.50 | 3.20 | 4.10 | 4.95 | 6.05 | 9.50 | 10.10 | 7.15 | 5.30 |
| % of trials with recommended dose at DLT rate of <20% | 14.00 | 18.30 | 18.60 | 22.15 | 20.55 | 24.70 | 16.00 | 18.90 | 21.55 | 25.05 | 21.50 | 24.15 |
| Average % of patients treated at doses with >40% DLT rate | 1.97 | 2.12 | 4.02 | 4.58 | 1.45 | 1.57 | 3.39 | 3.42 | 6.05 | 5.89 | 2.55 | 1.90 |
| Average % of patients treated at doses <20% DLT rate | 40.21 | 43.72 | 44.91 | 46.37 | 50.86 | 52.16 | 42.47 | 42.79 | 47.32 | 49.16 | 51.86 | 52.14 |
| Average % of patients with DLT (grade 3 or 4) | 22.83 | 22.80 | 22.54 | 22.45 | 21.24 | 21.04 | 23.17 | 22.96 | 22.84 | 22.72 | 21.38 | 21.50 |
| Average % of patients with a non-DLT (grade 1 or 2) | 73.75 | NA | 73.27 | NA | 74.75 | NA | 73.37 | NA | 72.69 | NA | 74.39 | NA |

*Proportional Odds Model

# Conclusions

- Proportional odds model incorporating ordinal toxicity performs better or similarly to a comparable binary CRM
- When the prior underestimates the toxicity, gains are seen
  - fewer patients treated at toxic doses
  - fewer DLTs observed
- When the true underlying dose-toxicity relationship violates the proportional odds assumption, the ordinal design may still perform better than the CRM
  - many possible violation scenarios to consider
  - so far, seems to be the same or better
- The coverage of nominal 95% confidence intervals for the estimated dose is much closer to 95% for the ordinal designs versus the traditional CRM
  - 95% CIs are wider than in binary CRM
  - but, they are more realistic
- Easy to implement!  R library coming soon....

# Part II: safety and toxicity outcomes

- Motivating trial: Hidalgo et al. Dose finding study of rapamycin in adults with solid tumors.
  - desire to find dose which will inhibit S6K kinase activity in peripheral blood mononuclear cells
  - change in pre versus post expression of S6K kinase
  - Pharmacodynamic response: 80% inhibition
  - Still concern over toxicity: rapamycin known to be toxic agent
- Idea: Titrate to high level of PD response while constraining to escalation by safety

# Two regression model

- Efficacy Model:

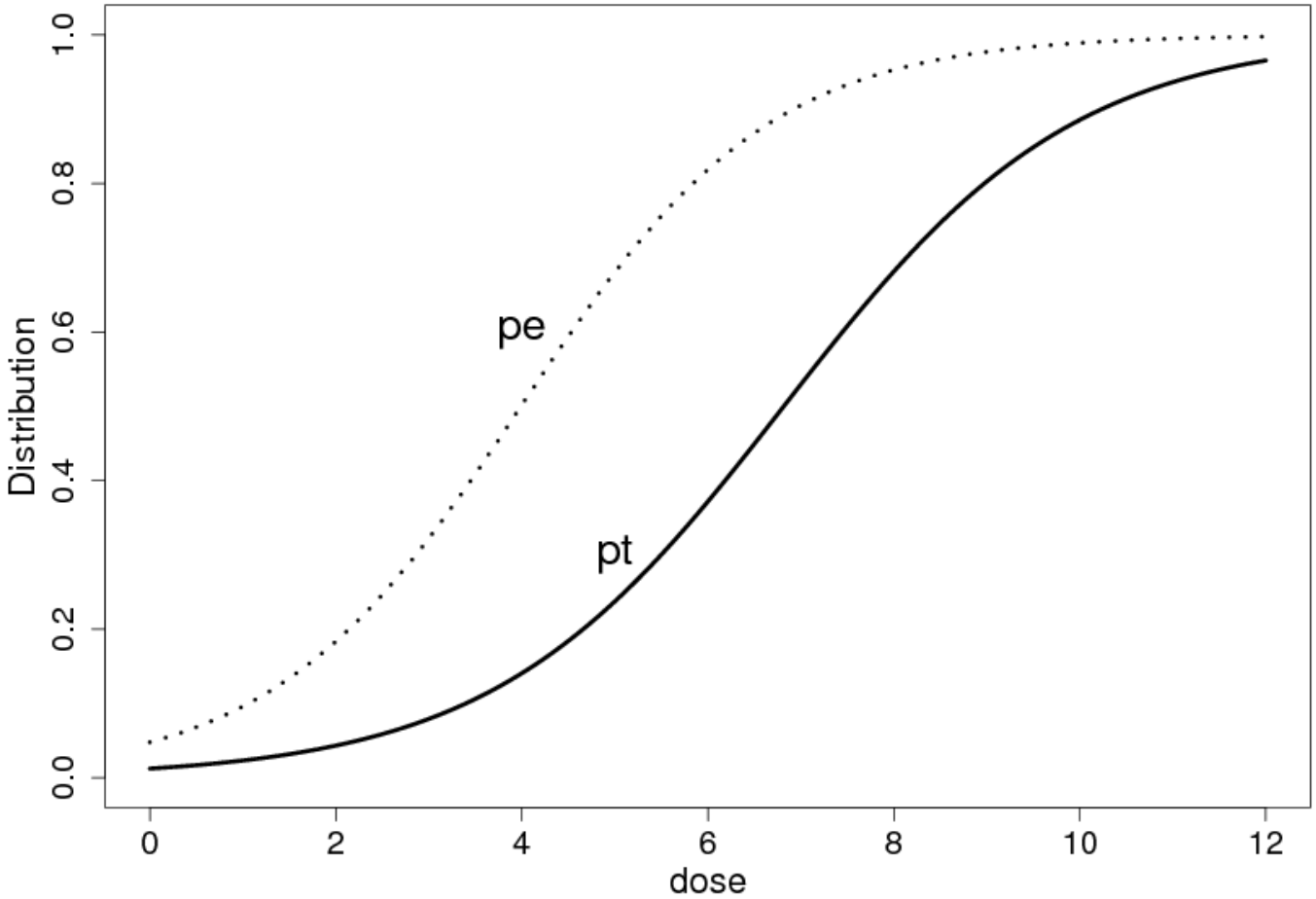
$$\text{logit}(p_e(d)) = -\beta(d - b_{50})$$

- Toxicity Model:

$$\text{logit}(p_t(d)) = -\alpha(d - a_{50})$$

- Using “two stage” CRM (Goodman et al. 1995; Moller, 1995, etc.)

# Two regression model (example)



# Competing model

- Published around the same time:  Continuation ratio model (Zhang et al. (2006), Mandrekar et al. (2007)). (earlier work by Thall and Russell (1998) and Fan and Chaloner)
- Continuation Ratio Model:
  - $\pi_2(d)$: monotone increasing function of dose ($d$)
  - $\pi_0(d)$: monotone non-increasing function of dose.
  - $\pi_1(d)$ is unimodal and can be either non-increasing or non-decreasing across a range of doses.

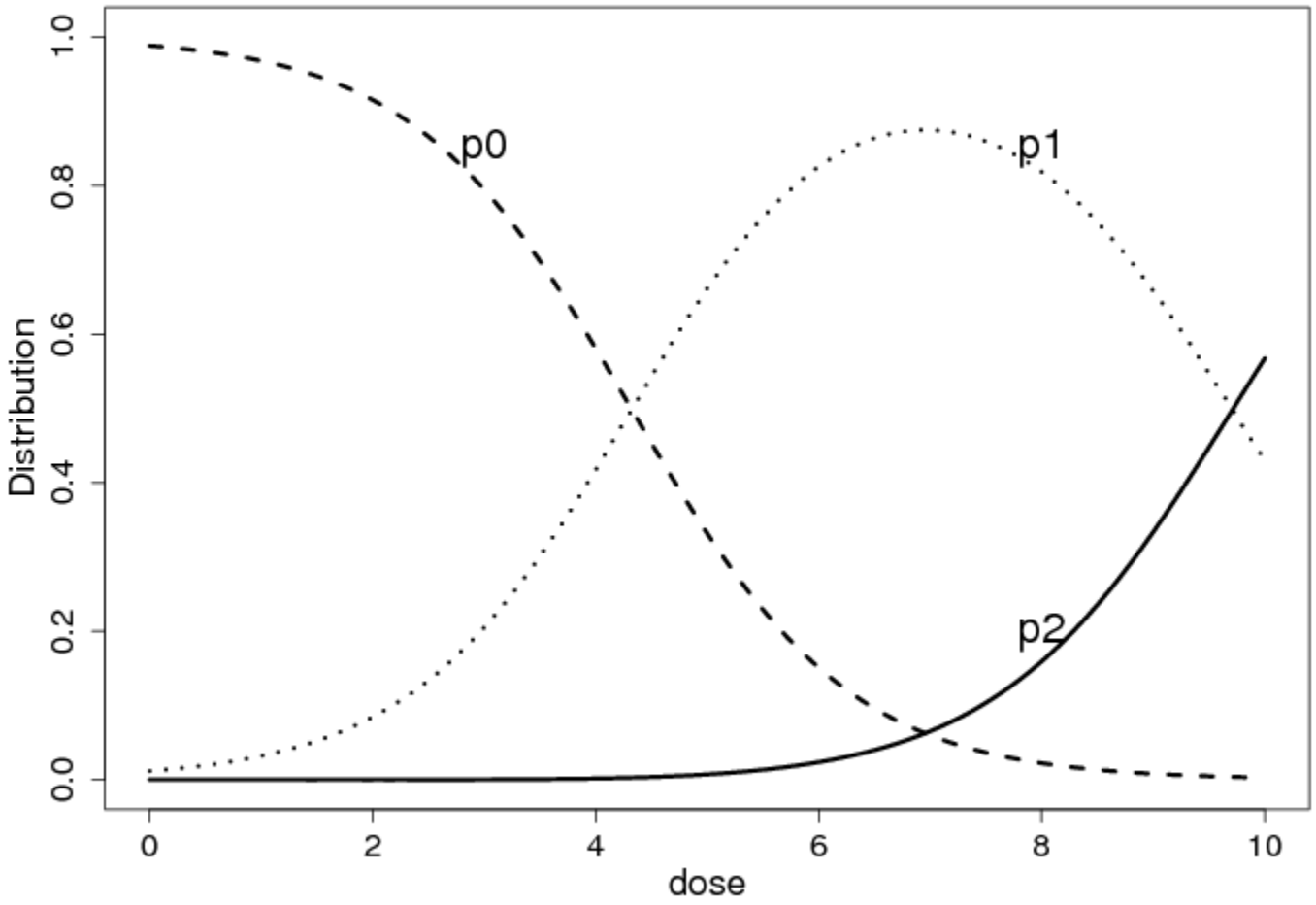
$$\log\left(\frac{\pi_1(d)}{\pi_0(d)}\right) = \gamma_{01} + \gamma_{11} d$$

$$\log\left(\frac{\pi_2(d)}{1-\pi_2(d)}\right) = \gamma_{02} + \gamma_{12} d$$

$$\sum_{r=0}^{2} \pi_r(d) = 1$$

$$\gamma_{01} > \gamma_{02}$$

# Continuation ratio model (example)



# Differences

- Zhang et al. choose the dose for the next cohort of patients based on:

$$\max_d(\hat{\pi}_1(d) - \lambda\hat{\pi}_2(d))$$

  - this allows a trade-off between toxicity and efficacy where $0 \leq \lambda \leq 1$.
- The two regression model assumes $\lambda = 0$
  - implies maximizing efficacy subject to the toxicity constraint is sufficient.
  - we choose $\lambda = 0$ *in simulations* to ensure comparability of results from the two approaches.

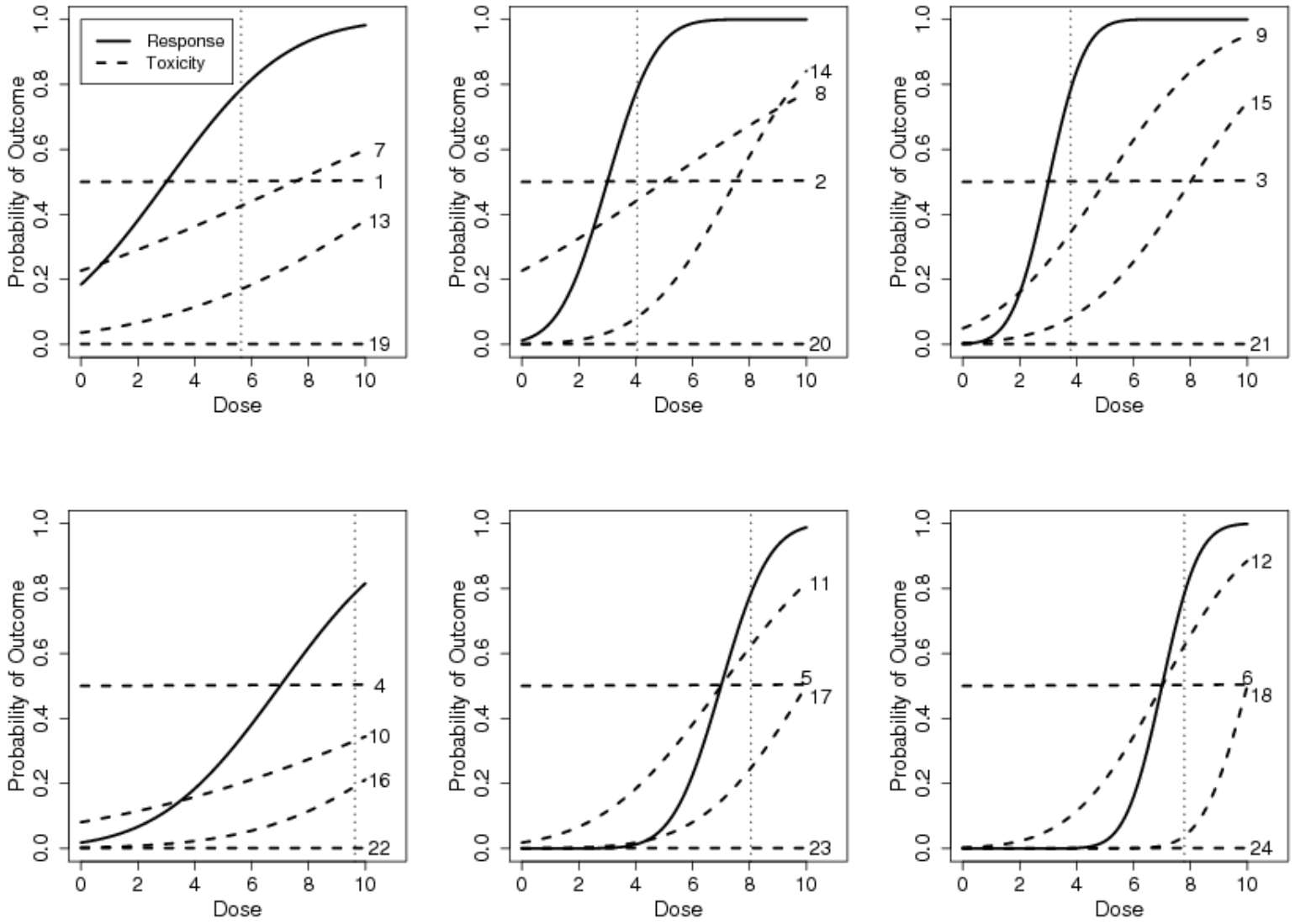
# Estimation

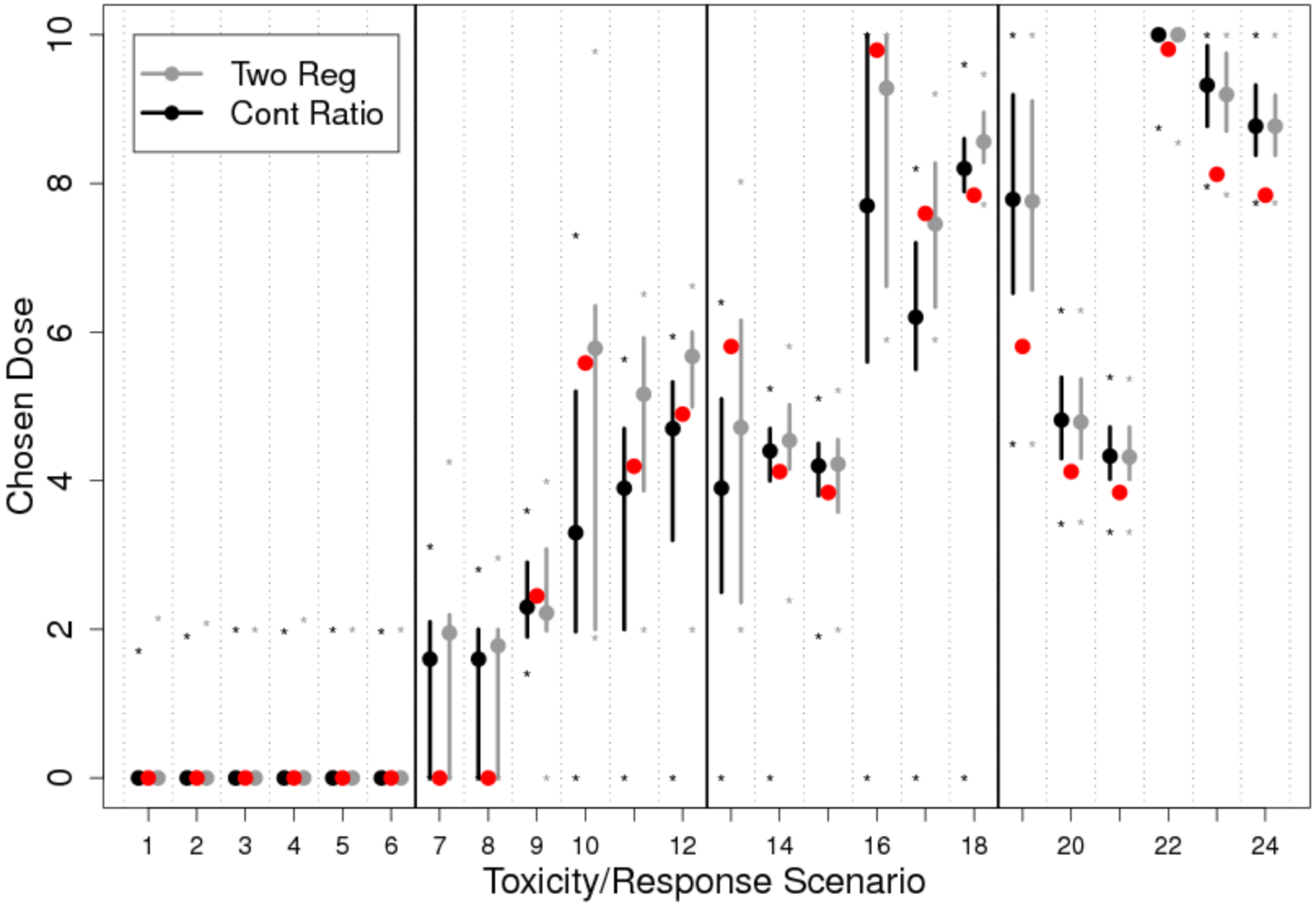
- Treat *k* patients at each dose until either response or toxicity or both occur
- For next patient, fit regression model(s)
- Choose dose for which
  - $p_e = c_e$
  - subject to the constraint: $p_t < c_t$
  - subject to no skipping doses
- Can use pseudo-data to stabilize estimation
  - e.g., if toxicities occur at first dose, then no variance in doses. model unestimable.

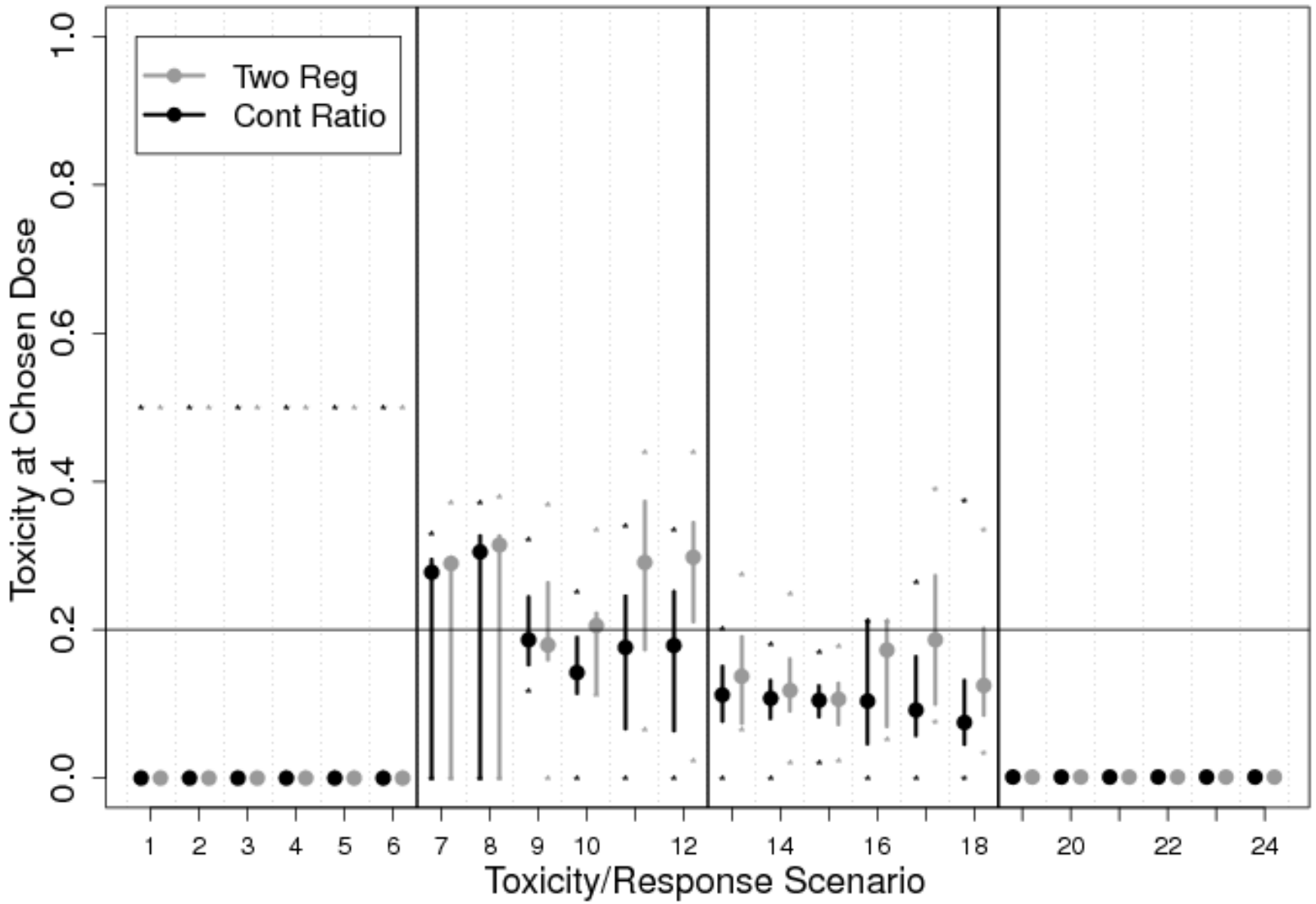
# Simulations

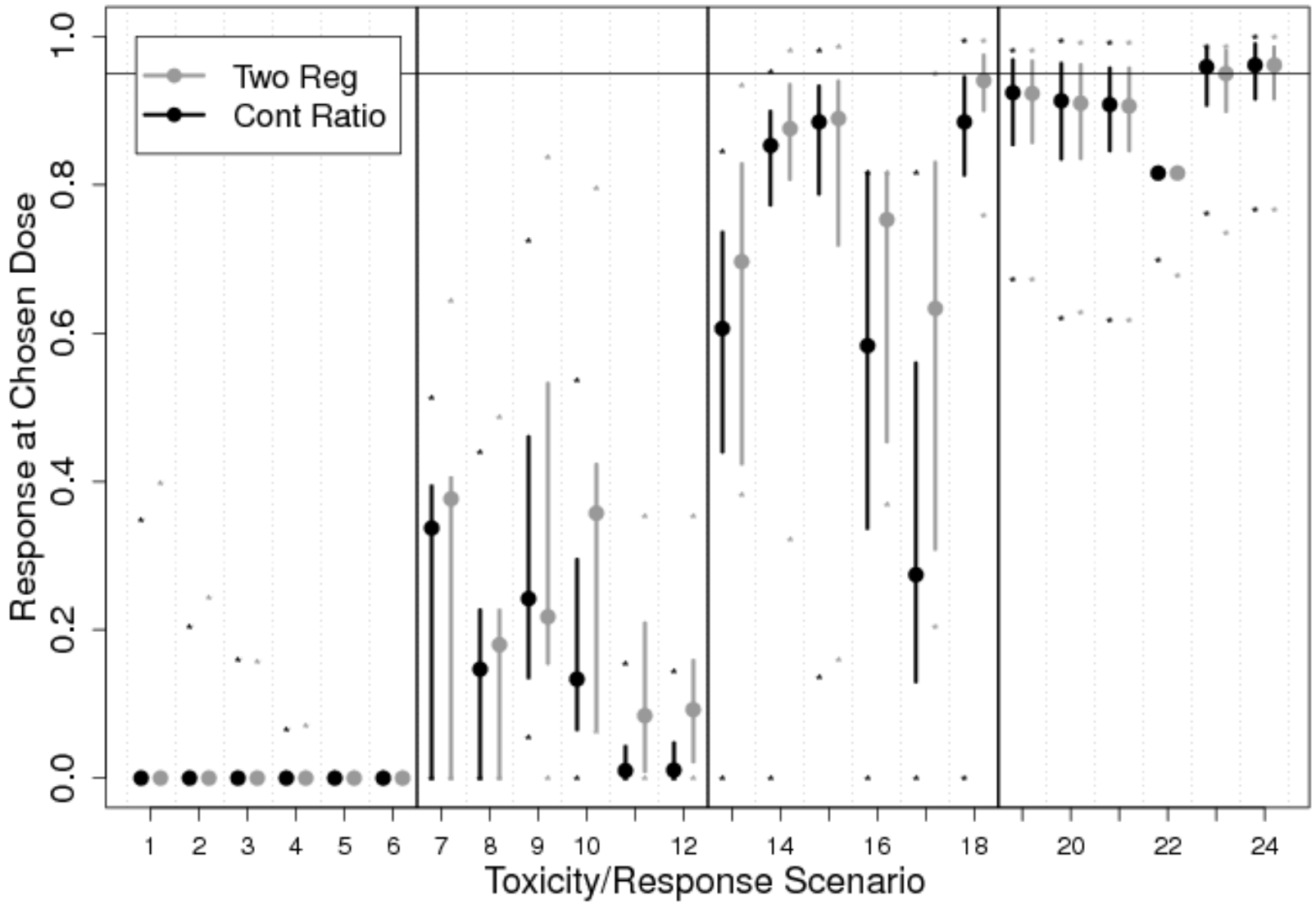
- 10000 trials per 24 toxicity/response scenarios
- Up to 39 patients per trial
  - cohort size 3
  - up to 13 cohorts
- Stopping rules
  - predicted dose < 0
  - once 10 patients have been treated at doses within 10% of predicted dose for next cohort.
- dose range: 1-10
  - start at dose 2
  - dose escalation limited to 4 increments
  - in the absence of toxicity:  doses 2, 6, 10 tried
- Continuation ratio approach (with $\lambda=0$) used as comparison

# 24 toxicity response scenarios considered




Two Reg
Cont Ratio
Chosen Dose
0
2
4
6
8
10
1
2
3
4
5
6
7
8
9
10
12
14
16
18
20
22
24
Toxicity/Response Scenario


Two Reg
Cont Ratio
Toxicity at Chosen Dose
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
6
7
8
9
10
12
14
16
18
20
22
24
Toxicity/Response Scenario


Two Reg
Cont Ratio
Response at Chosen Dose
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
6
7
8
9
10
12
14
16
18
20
22
24
Toxicity/Response Scenario

# Conclusions

- Two regression approach does as well as continuation ratio
- Comparable results:  computationally similar
- Benefits:
  - simple to fit in standard software
  - simple(r) to explain to clinical colleagues
  - extension of some of “accepted” CRMs in use
- Work to be considered:
  - vary scenarios (e.g., sample size, cohort size)
  - estimation approach (Bayesian better?)