Some Bayesian Methods for Clinical Trial Design and Analysis

Peter F. Thall

Biostatistics and Applied Mathematics Dept.
University of Texas
M.D. Anderson Cancer Center

BASS XII
Savannah, November 7-11, 2005

Outline

I. A Phase I/II Dose-Finding Design

II. Covariate-Adjusted Adaptive Randomization

III. Posterior Sensitivity Analyses of Confounded Treatment Effects

A Phase I/II Dose-Finding Design

- ***Patient Outcome*** **= {Response, Toxicity}**

- **The physician(s) specify**

  - **A Lower Limit** p_R^* **on** π_R **= P(Res)**
  - **An Upper Limit** p_T^* **on** π_T **= P(Tox)**
  - **Three equally desirable(**π_R**,** π_T**) targets which are used to construct an Efficacy-Toxicity Trade-off Contour**

Thall and Cook, ***Biometrics*** **Sept. 2004**

Dose Acceptability Criteria

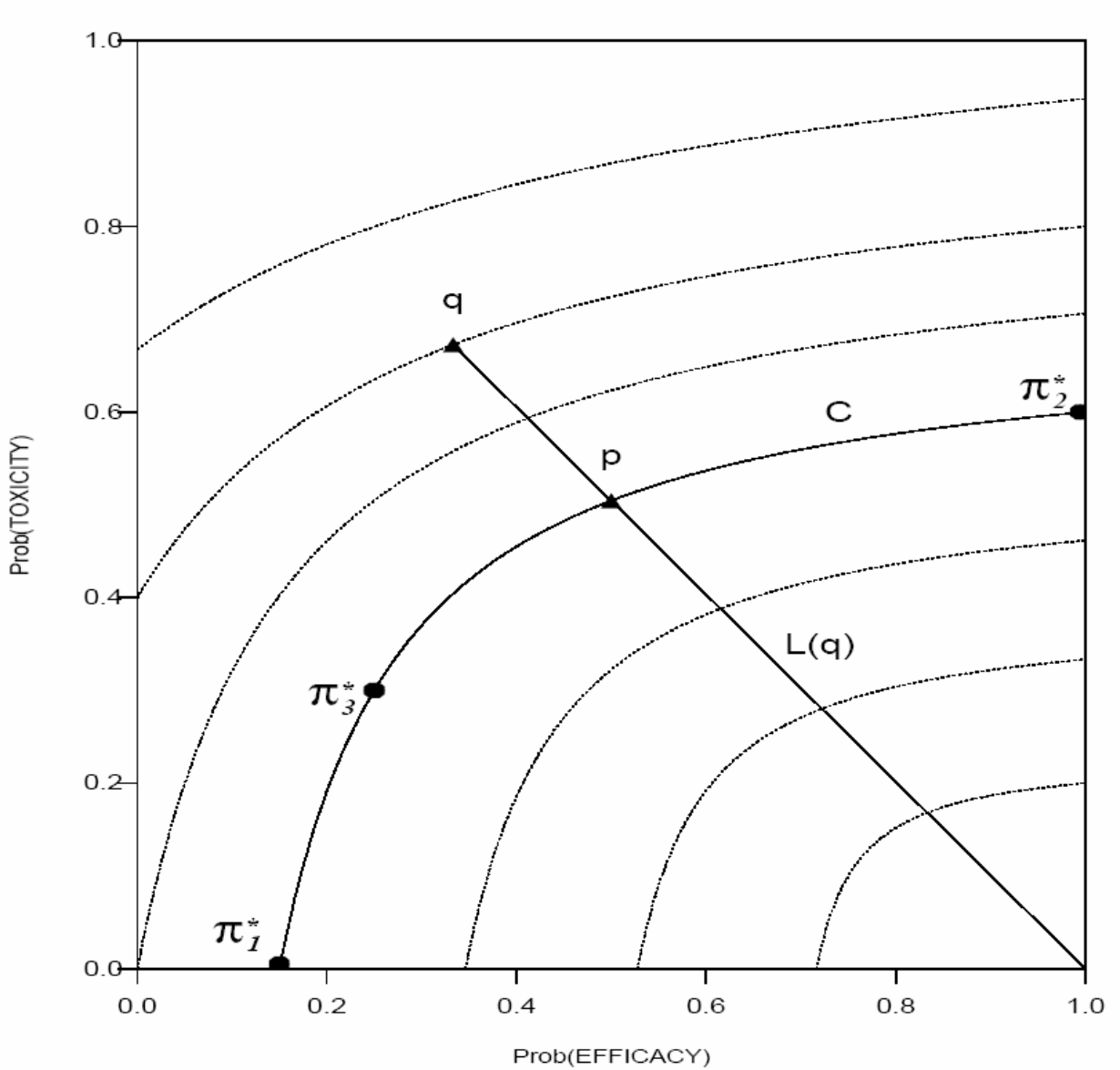
Given current data, a dose x is Acceptable if

$$\textbf{Pr}\{\ \pi_E(x,\theta) > p_E^{*} \mid \textit{data}\ \} > .90$$

$$\textbf{Pr}\{\ \pi_T(x,\theta) < p_T^{*} \mid \textit{data}\ \} > .90$$

(other numerical upper cutoffs may be used)

Efficacy-Toxicity Trade-Off Contours



Comparing Doses

Given current data ***D,*** for each dose x

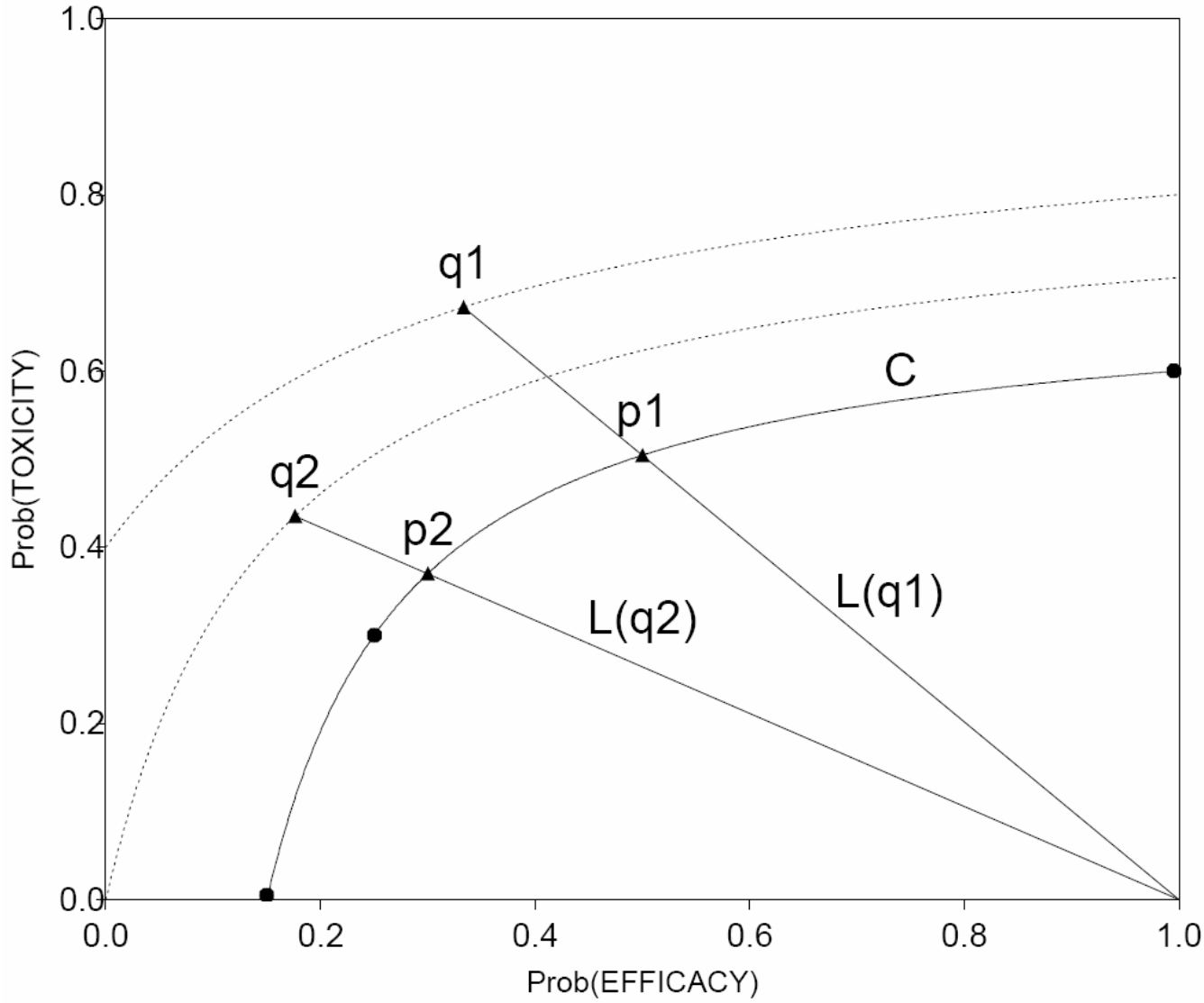
$\delta(x,D)$ = ***desirability*** of x ***is*** the desirability of

$$(q_E, q_T) = \left(E\{ \pi_E(x,\theta) \mid D\},\ E\{ \pi_T(x,\theta) \mid D\} \right)$$

Compare x_1 to x_2 using $\delta(x_1,D)$ and $\delta(x_2,D)$

$E\{\pi(x_1,\theta)|\ D\} = q_1$ and $E\{\pi(x_2,\theta)|\ D\} = q_2$

➔ $\delta(x_2, D) > \delta(x_1, D)$



Trial Conduct

1) Physician chooses the starting dose
2) Dose ***x*** is ***acceptable*** if

- ***x* has acceptable $\pi_E(x,\theta)$ & $\pi_T(x,\theta)$ or**
- ***x* is the lowest untried dose & has acceptable $\pi_T(x,\theta)$**

3) Treat each cohort at current most desirable dose
4) Do not skip untried doses
5) No dose acceptable ➔ **Stop the trial**
6) At the end of the trial, select the most desirable dose

Treating Acute Ischemic Stroke

Rx = Fixed dose abciximab

+ 0.0, 2.5, 5.0, 7.5, or 10.0 U reteplase

N_{max} = 72 patients, cohort size = 3, first cohort treated at 0.0 U reteplase

Conducted by *National Institute of Neurological Diseases & Stroke*, NIH, USA (S. Warach, P.I.)

Treating Acute Ischemic Stroke

Tox = Intra-cranial bleeding, death, or other severe AE, within 48 hrs.

Res = Reperfusion at 24 hrs. w/o Tox

➔ Outcome is Res, Tox, or Neither

p_T^* = .10 Upper Limit on π_T

p_E^* = .50 Lower Limit on π_R

Treating Steroid-Refractory GVHD

Patients with steroid-refractory GVHD after allotx from an HLA-matched donor

Rx = .25, .50, .75, or 1.00 mg/m^2 Pentostatin

N_{max} = 36 patients, cohort size = 3, treat the first cohort at 25 mg/m^2

Conducted at *M.D. Anderson Cancer Center* (D. Couriel, P.I.)

Treating Steroid-Refractory GVHD

Tox = {Unresolved infection or death}

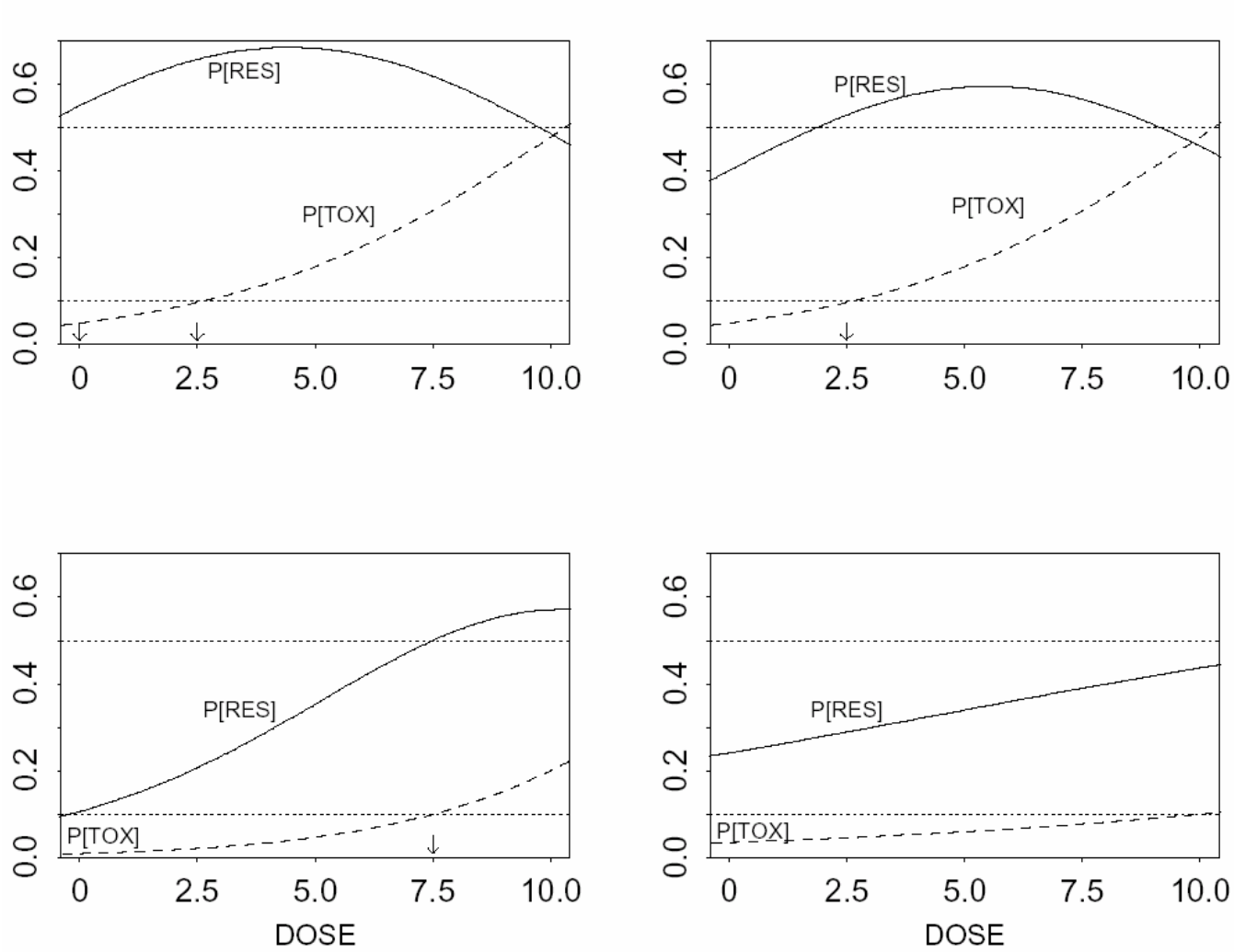
Res = { ≥ 1 grade drop in GVHD}
both within 2 weeks

Tox and **Res** may both occur ➔
Outcome is bivariate binary

p_T^* = .40 Upper Limit on P(Tox)

p_E^* = .20 Lower Limit on P(Res)

Some Possible Dose-Outcome Curves



Probability Models

- Trinary Outcomes:

  4-parameter continuation ratio model

- Bivariate Binary Outcomes:

  6-parameter odds ratio model

Probability Model: Trinary Outcomes

$$\text{logit}\ \pi_T(x,\theta) = \mu_T + x\,\beta_T$$

$$\text{logit}\ \pi_E(x,\theta)/\{1-\pi_T(x,\theta)\} = \mu_E + x\,\beta_E$$

$$\theta = (\mu_T\,,\ \beta_T\,,\ \mu_E\,,\ \beta_E\,)$$

Probability Model: Bivariate Binary Outcomes

$$\text{logit}\ \pi_T(x,\theta) = \mu_T + x\beta_T$$

$$\text{logit}\ \pi_E(x,\theta) = \mu_E + x\beta_{E,1} + x^2\beta_{E,2}$$

$$\pi_{a,b} = \pi_E^{a}(1-\pi_E)^{1-a}\,\pi_T^{b}(1-\pi_T)^{1-b} +$$

$$(-1)^{a+b}\,\pi_E(1-\pi_E)\pi_T(1-\pi_T)(e^{\psi}-1)/(e^{\psi}+1)$$

$$\theta = (\mu_T\,,\ \beta_T\,,\ \mu_E\,,\ \beta_{E,1}\,,\ \beta_{E,2}\,,\psi)$$

Establishing Priors

Each component θ_r of θ is normally distributed,

$$\theta_r \sim N(\mu_r, \sigma_r) \rightarrow$$

$\xi = (\mu_1, \sigma_1, \mu_2, \sigma_2, \ldots, \mu_p, \sigma_p)$ = hyperparameters

For each x_j , elicit

$m_{E,j}$ = prior mean and $s_{E,j}$ = prior sd of $\pi_E(x_j,\theta)$

$m_{T,j}$ = prior mean and $s_{T,j}$ = prior sd of $\pi_T(x_j,\theta)$

Establishing Priors

Find the vector ξ that minimizes

$$h(\boldsymbol{\xi}) = \sum_{y=E,T} \sum_{1 \le j \le J} \left[\{m_{y,j}(\boldsymbol{\xi}) - \hat{m}_{y,j}\}^2 + \{s_{y,j}(\boldsymbol{\xi}) - \hat{s}_{y,j}\}^2 \right]$$

$$+ \; c \sum_{1 \le j < k \le J} (\tilde{\sigma}_j - \tilde{\sigma}_k)^2$$

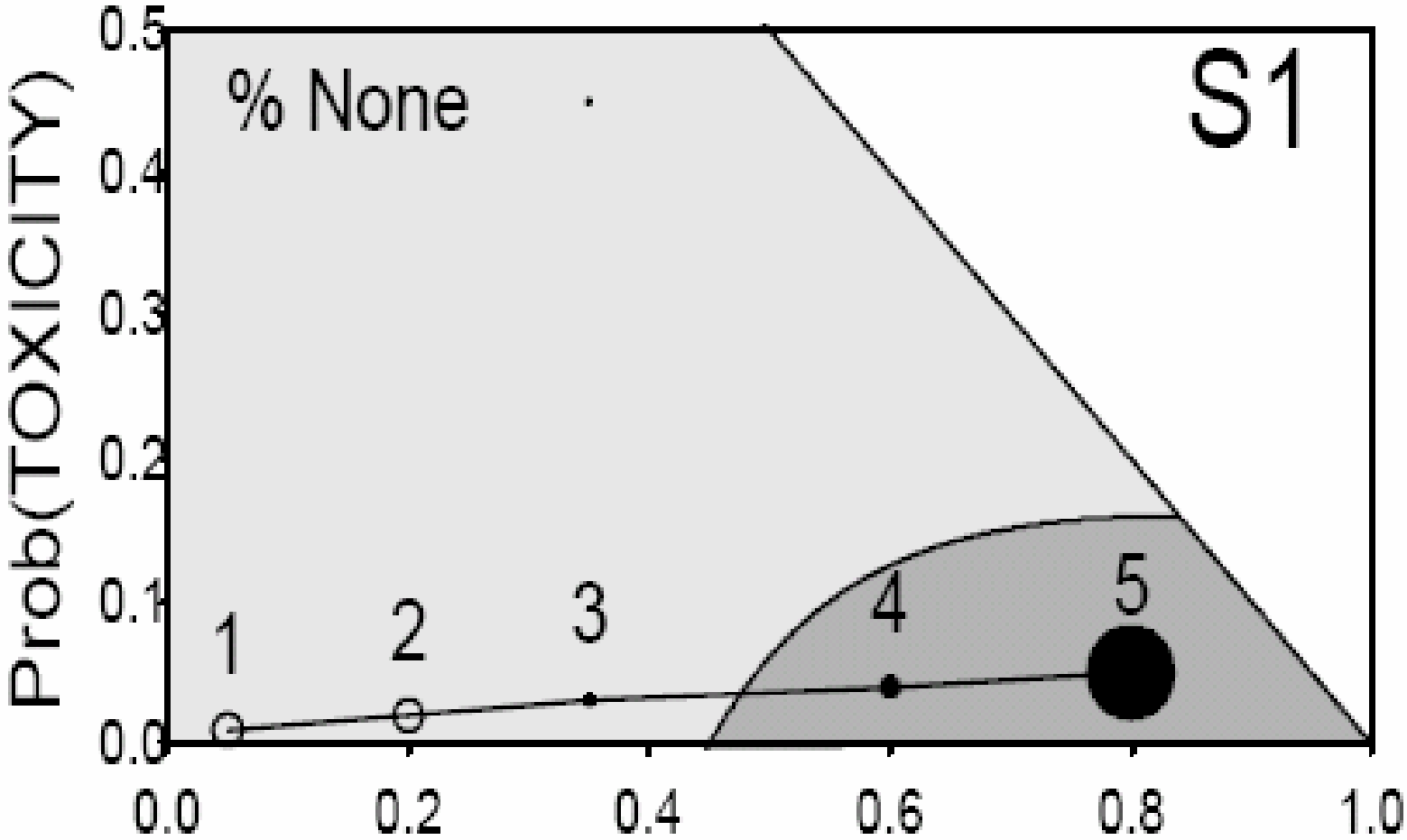
Computing Posteriors

Numerical integration w.r.t θ of

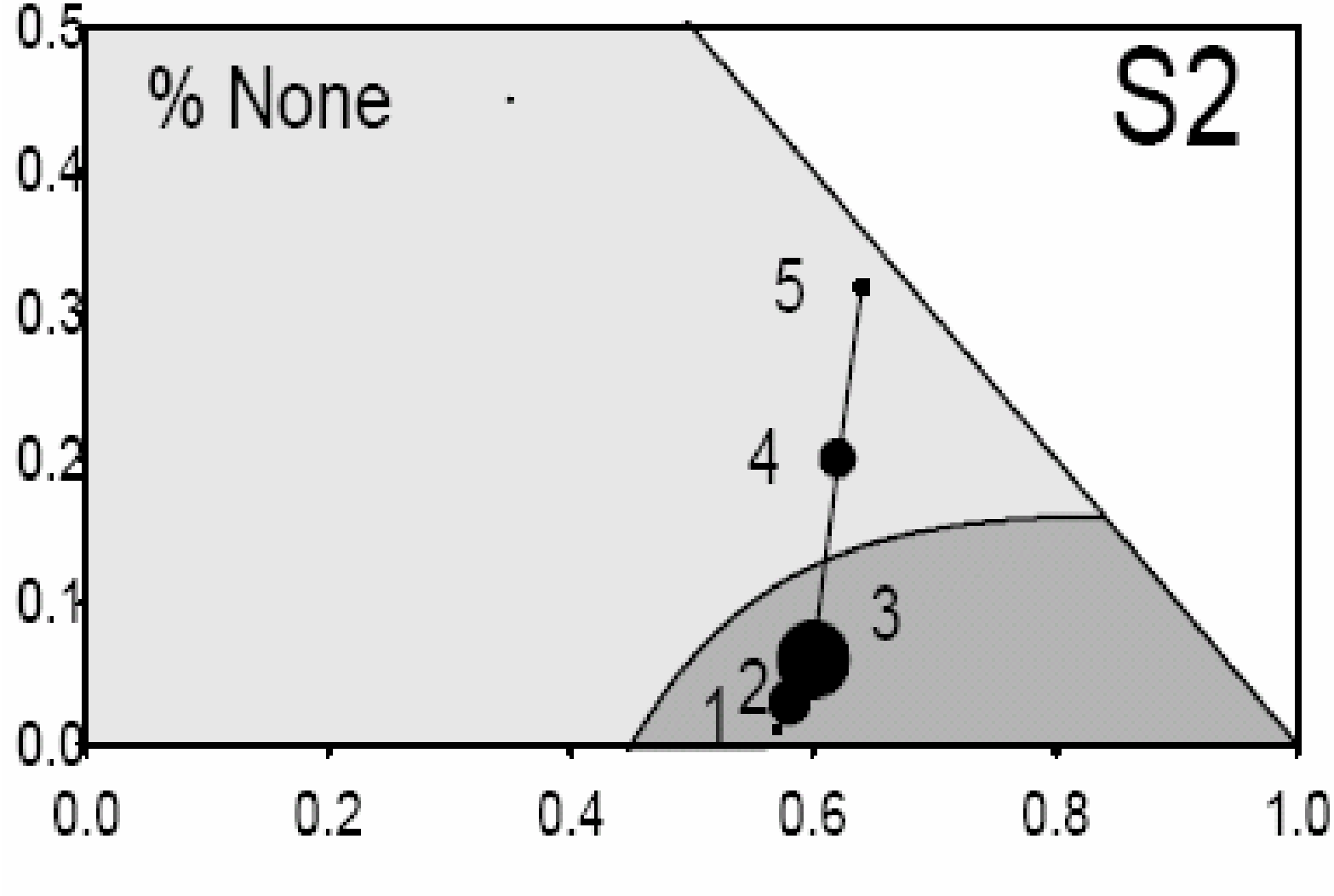
$$f(\theta) = Lik(D_n \mid \theta) \times Prior(\theta \mid \xi)$$

using defensive importance sampling

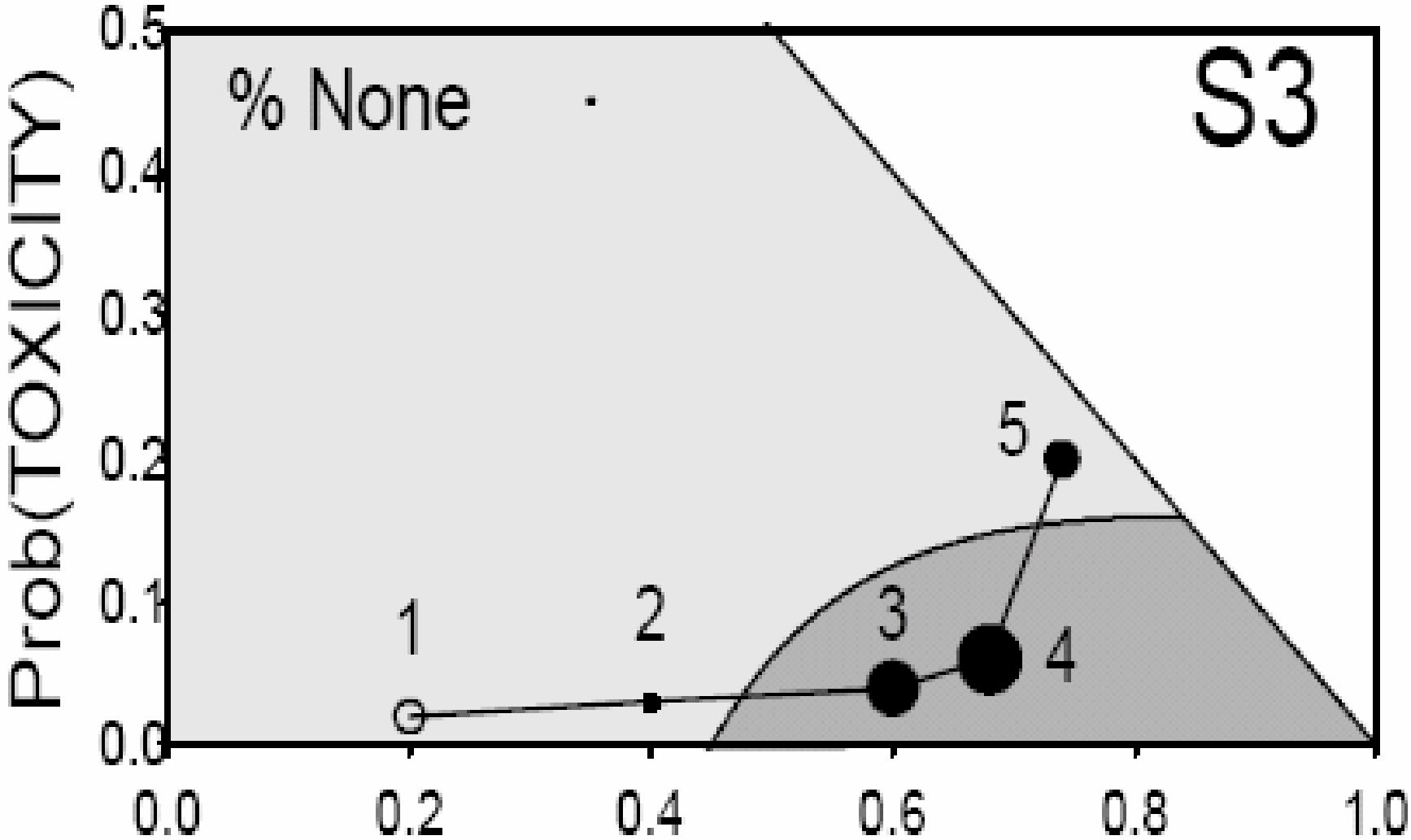
Simulation Results for the Stroke Trial

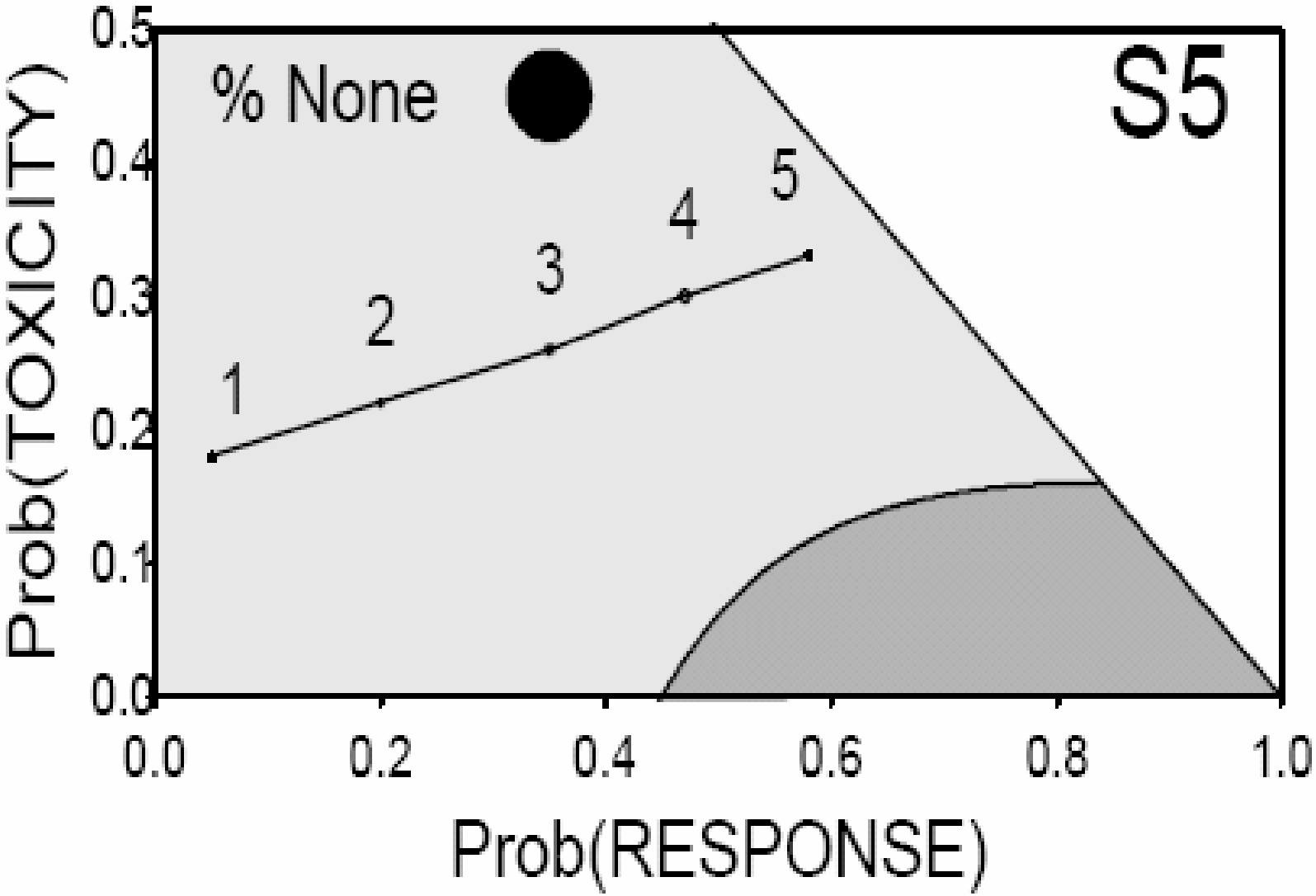


Prob(Response)

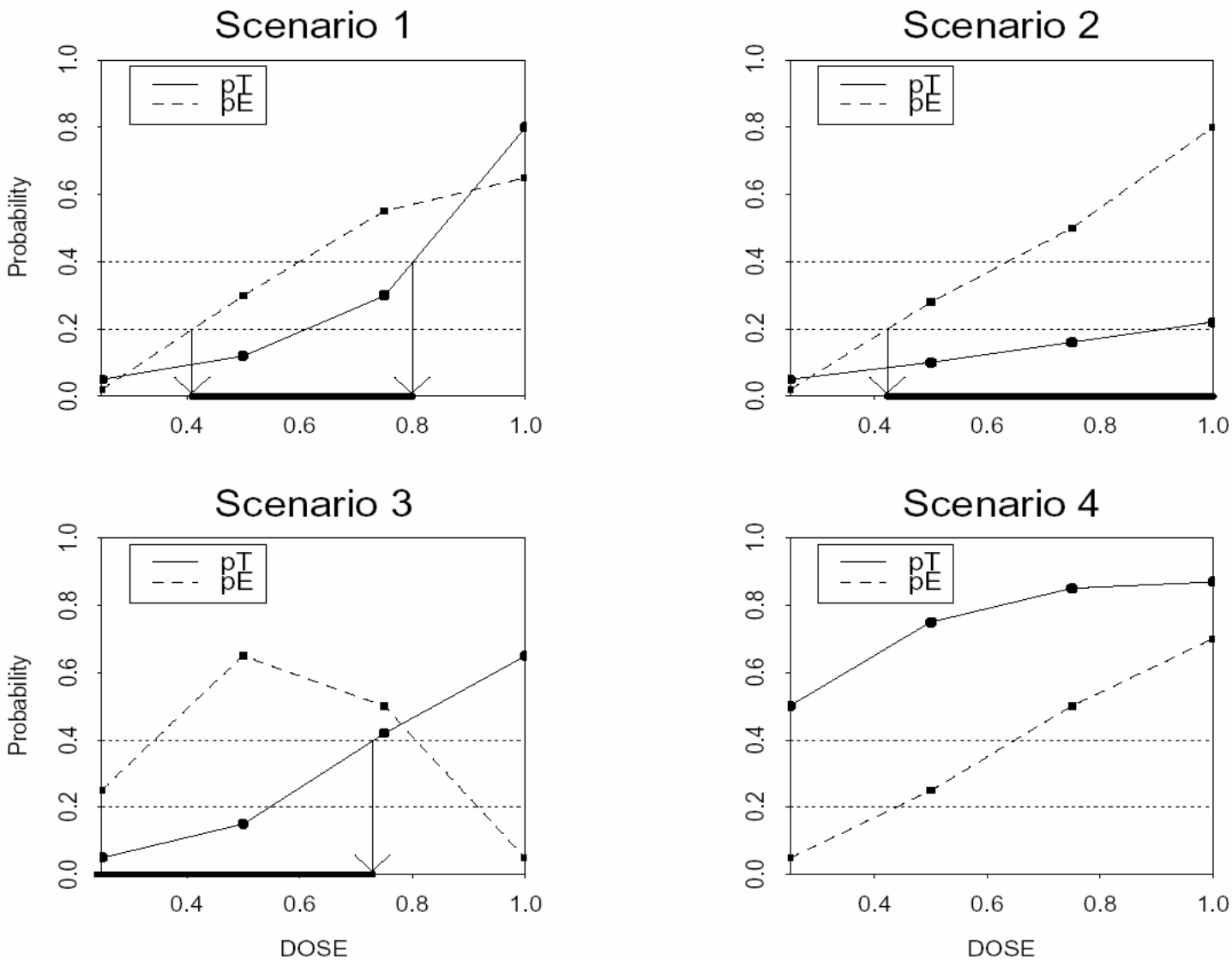


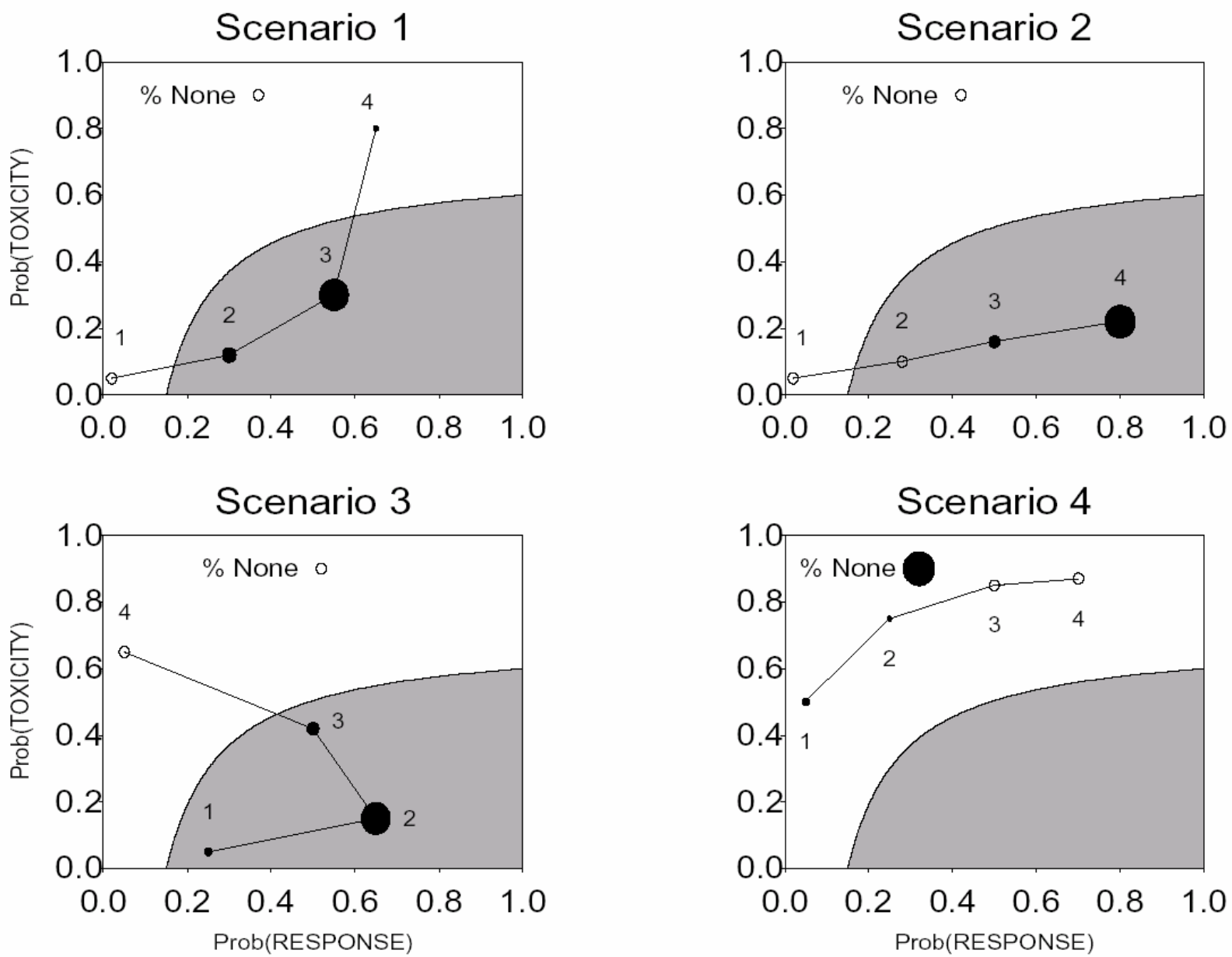
Prob(Response)


% None
S3
Prob(TOXICITY)
0.5
0.4
0.3
0.2
0.1
0.0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
Prob(Response)


% None
S5
1
2
3
4
5
0.5
0.4
0.3
0.2
0.1
0.0
Prob(TOXICITY)
0.0
0.2
0.4
0.6
0.8
1.0
Prob(RESPONSE)

Simulation Scenarios for the Pentostatin Trial




Scenario 1
% None
Prob(TOXICITY)
1.0
0.8
0.6
0.4
0.2
0.0
0.0 0.2 0.4 0.6 0.8 1.0
1
2
3
4
Scenario 2
% None
1
2
3
4
Scenario 3
% None
Prob(TOXICITY)
Prob(RESPONSE)
4
3
1
2
Scenario 4
% None
1
2
3
4
Prob(RESPONSE)

A Cohort-by-Cohort Illustration

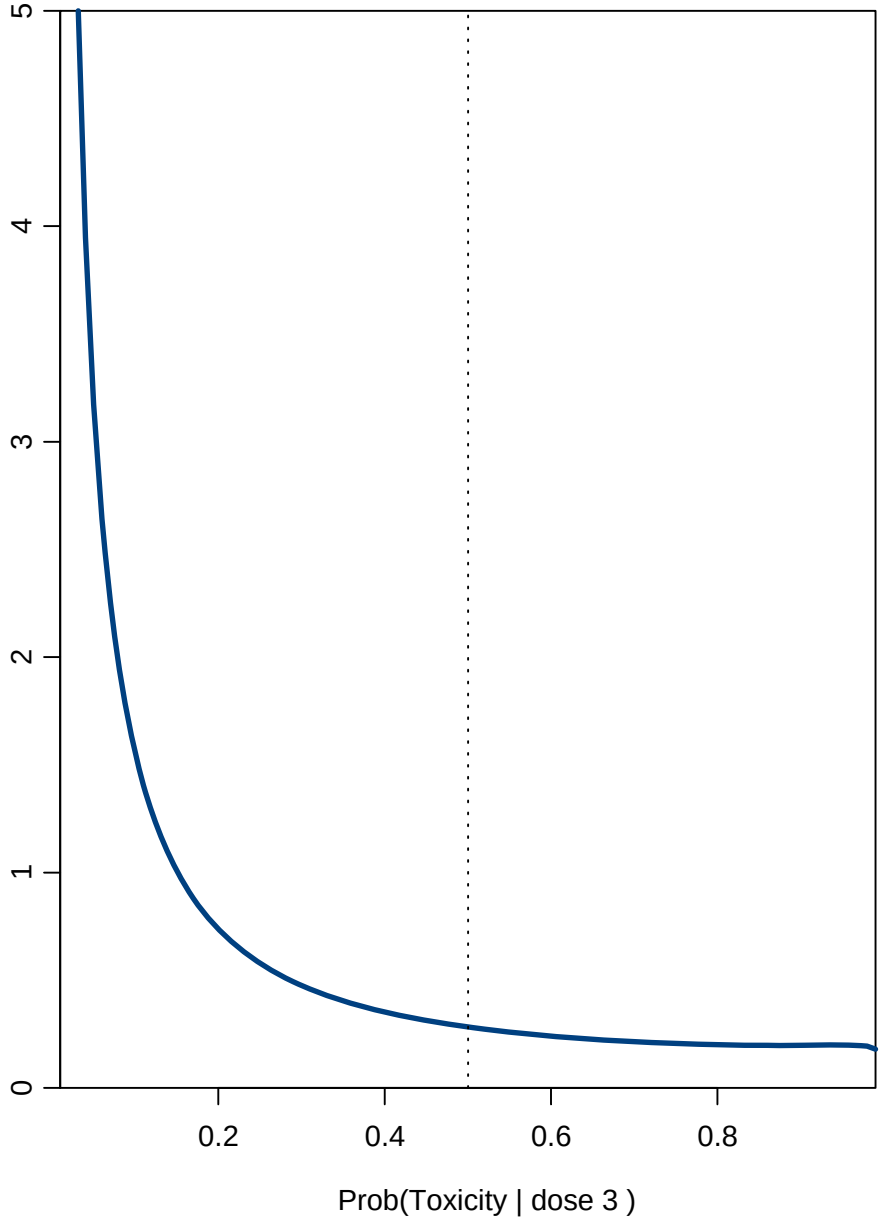
AML patients relapsed within 6 mos of CR

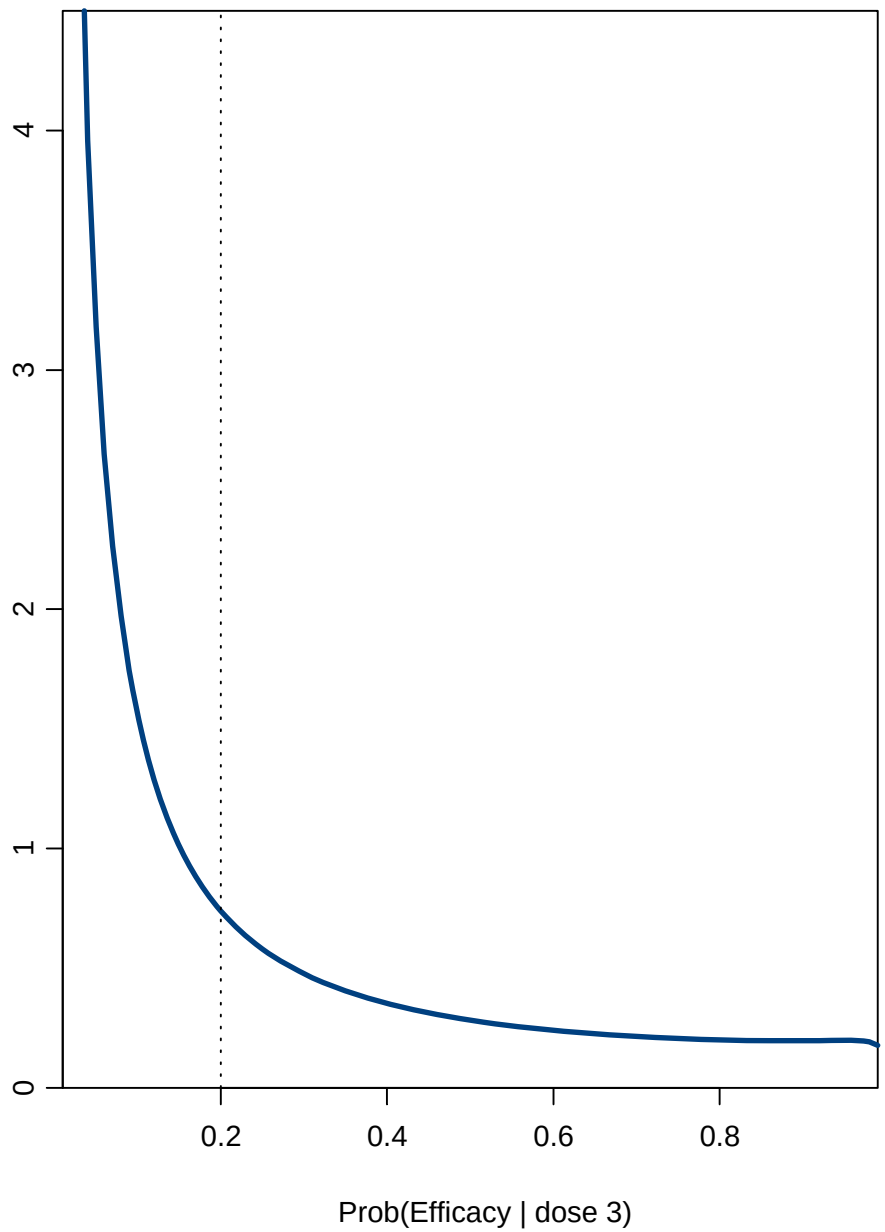
Rx = Fixed dose ara-C + one of 4 doses of XIAP, an anti-sense biological agent

Res = Alive & in CR at day35
Tox = Gr. 4 symptomatic tox within 35 days

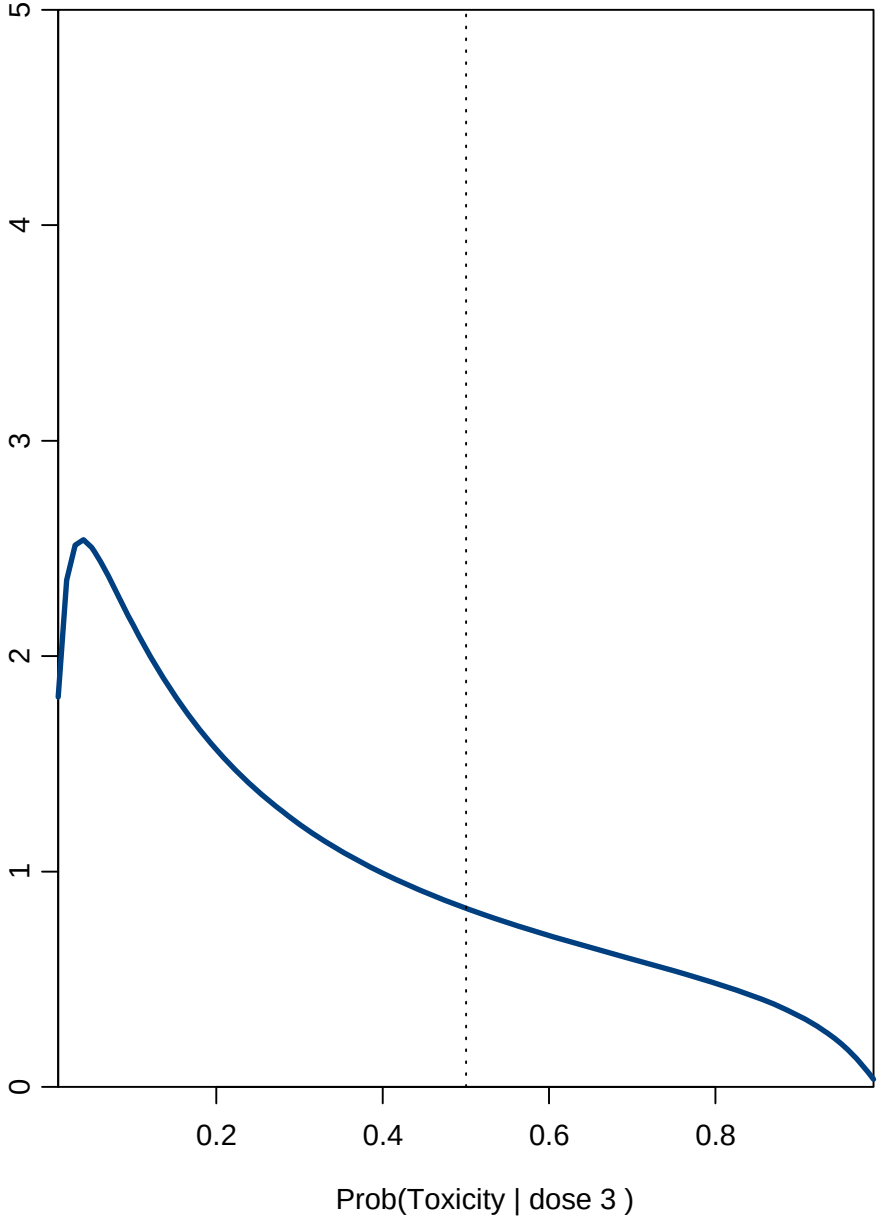
- **N_{max} = 36, cohort size = 3**
- **p_T^* = .50 for π_T, p_E^* = .20 for π_E**
- **Target pairs (.20, 0), (.60, .40), (1.00, .50)**

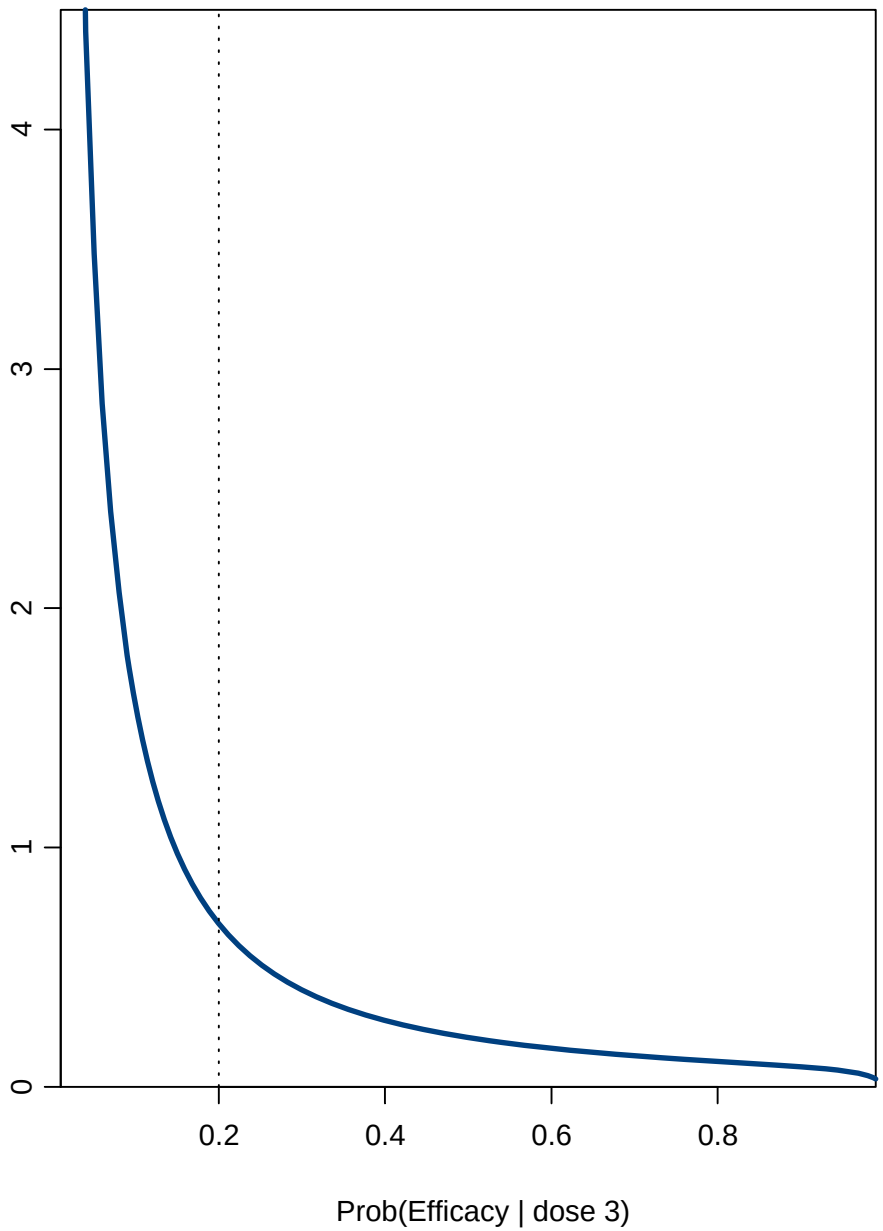
Prior Distributions



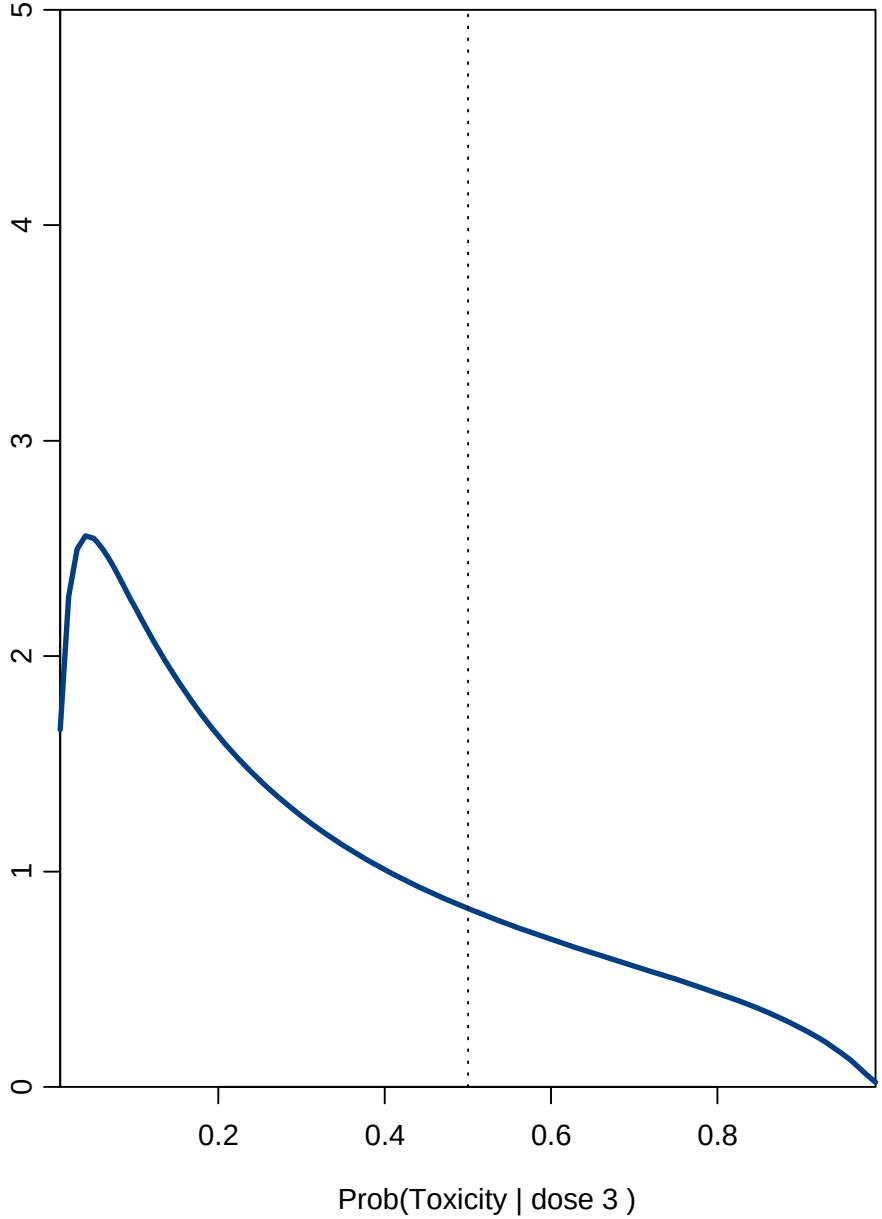


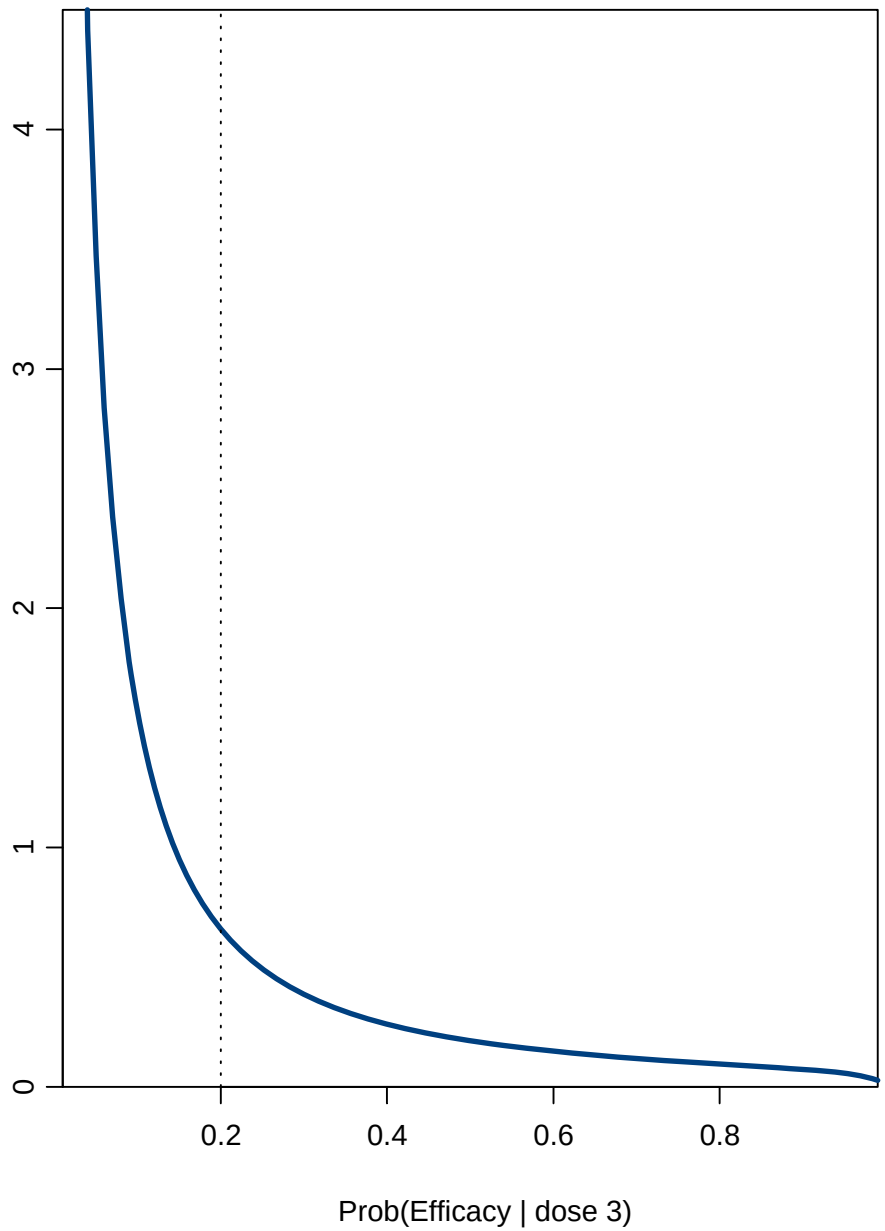
3 Patients


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose 3)

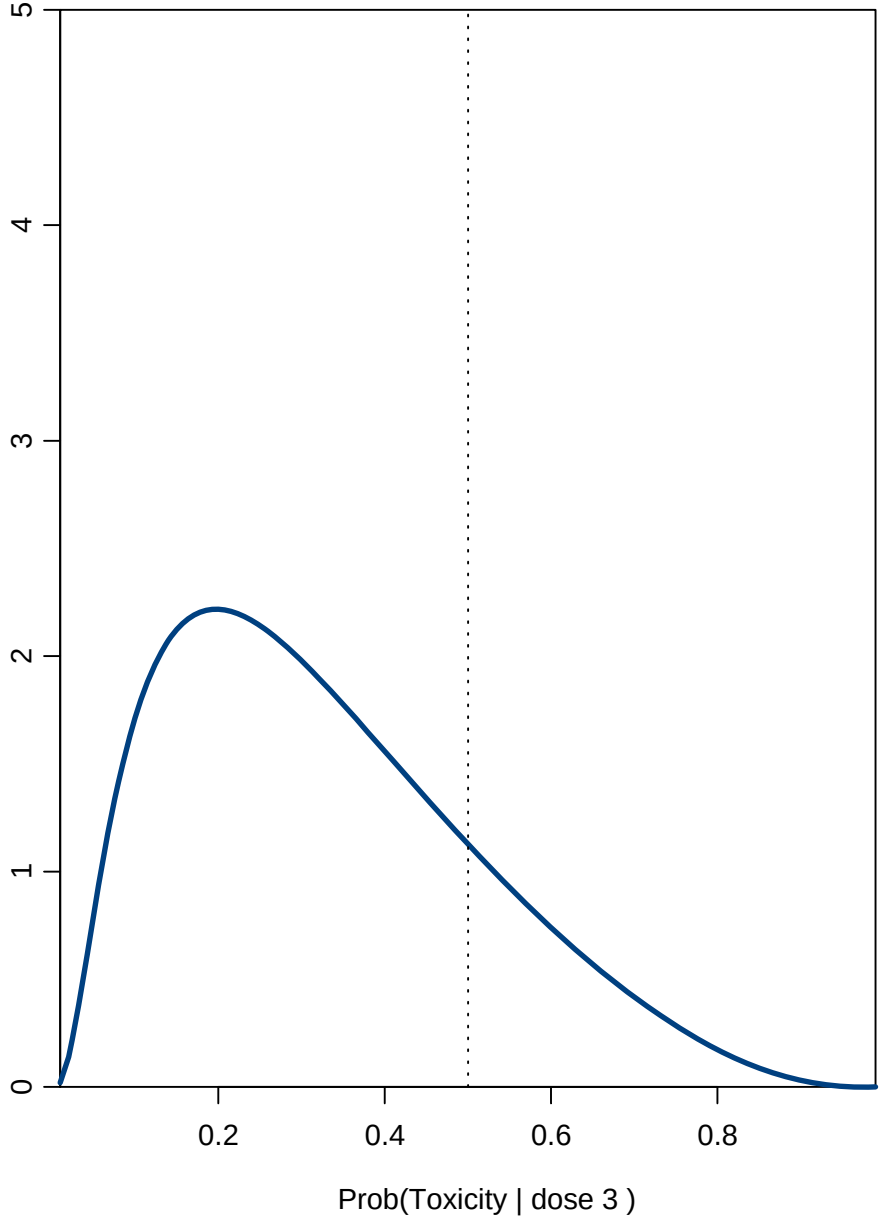

0
1
2
3
4
0.2
0.4
0.6
0.8
Prob(Efficacy | dose 3)

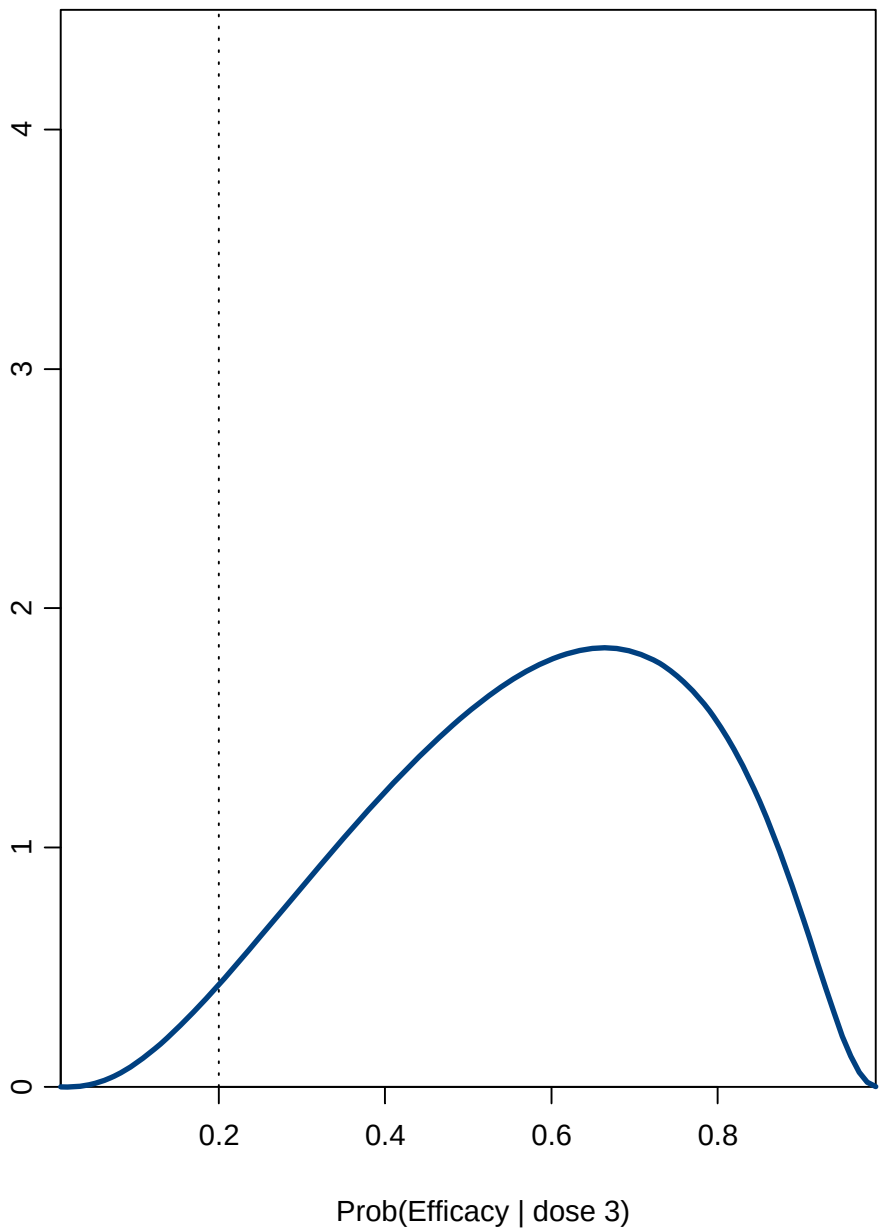
6 Patients


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose 3)

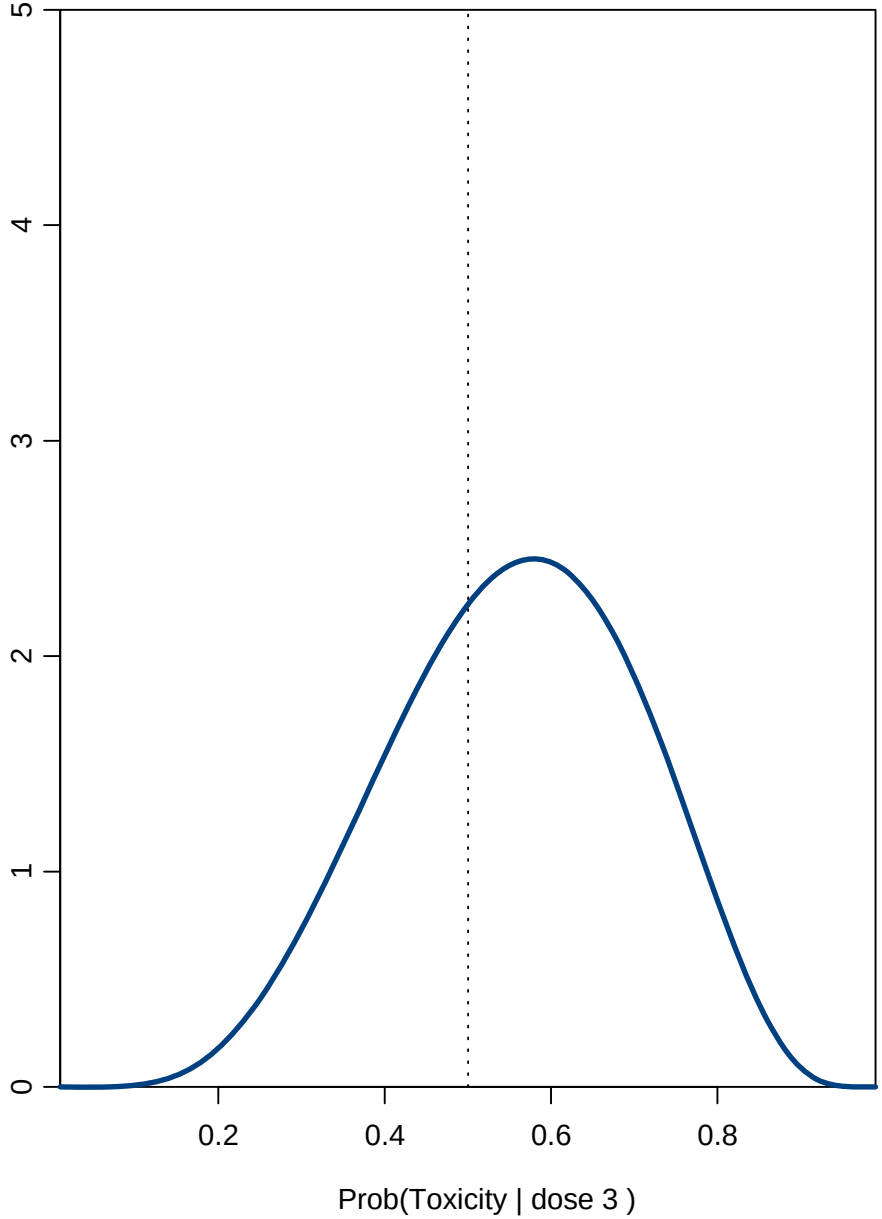

0
1
2
3
4
0.2
0.4
0.6
0.8
Prob(Efficacy | dose 3)

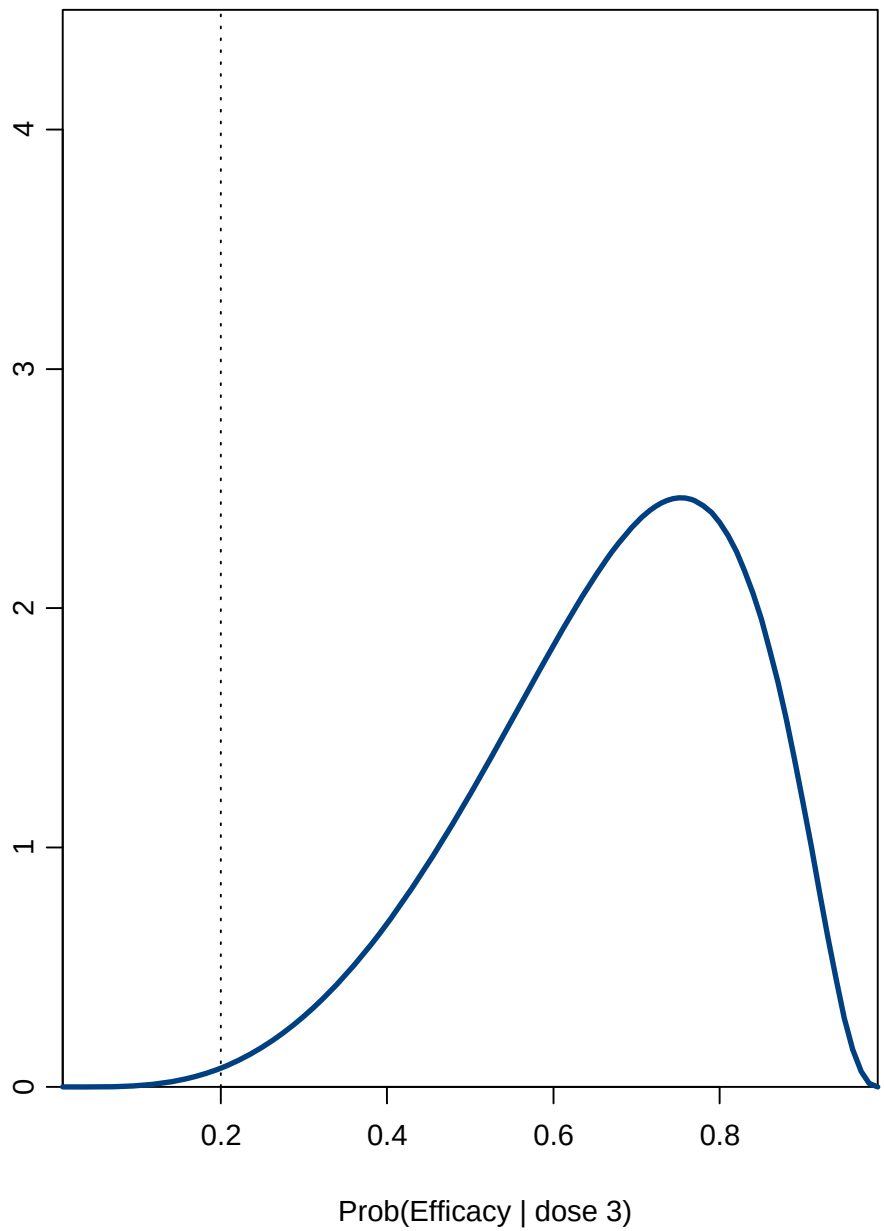
9 Patients


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose 3)

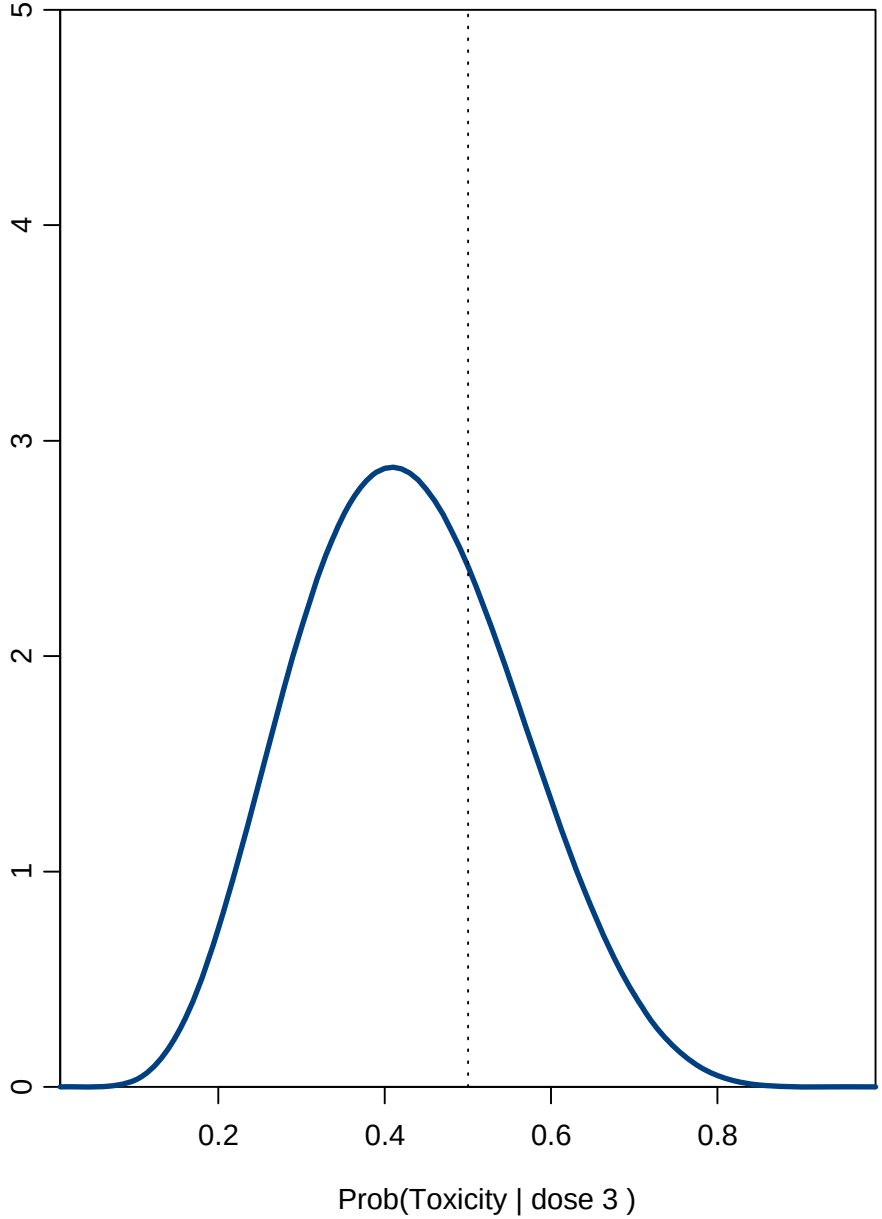

0
1
2
3
4
0.2
0.4
0.6
0.8
Prob(Efficacy | dose 3)

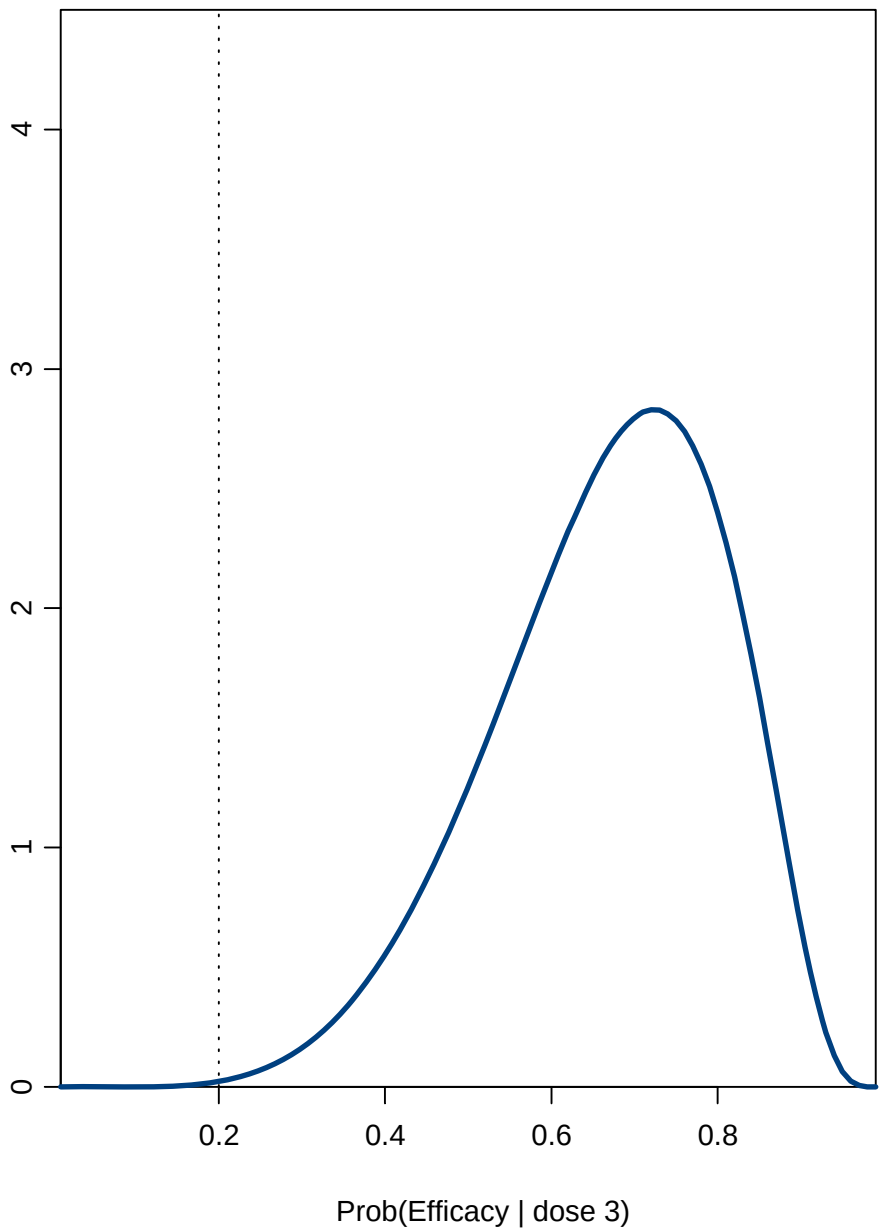
12 Patients


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose 3)

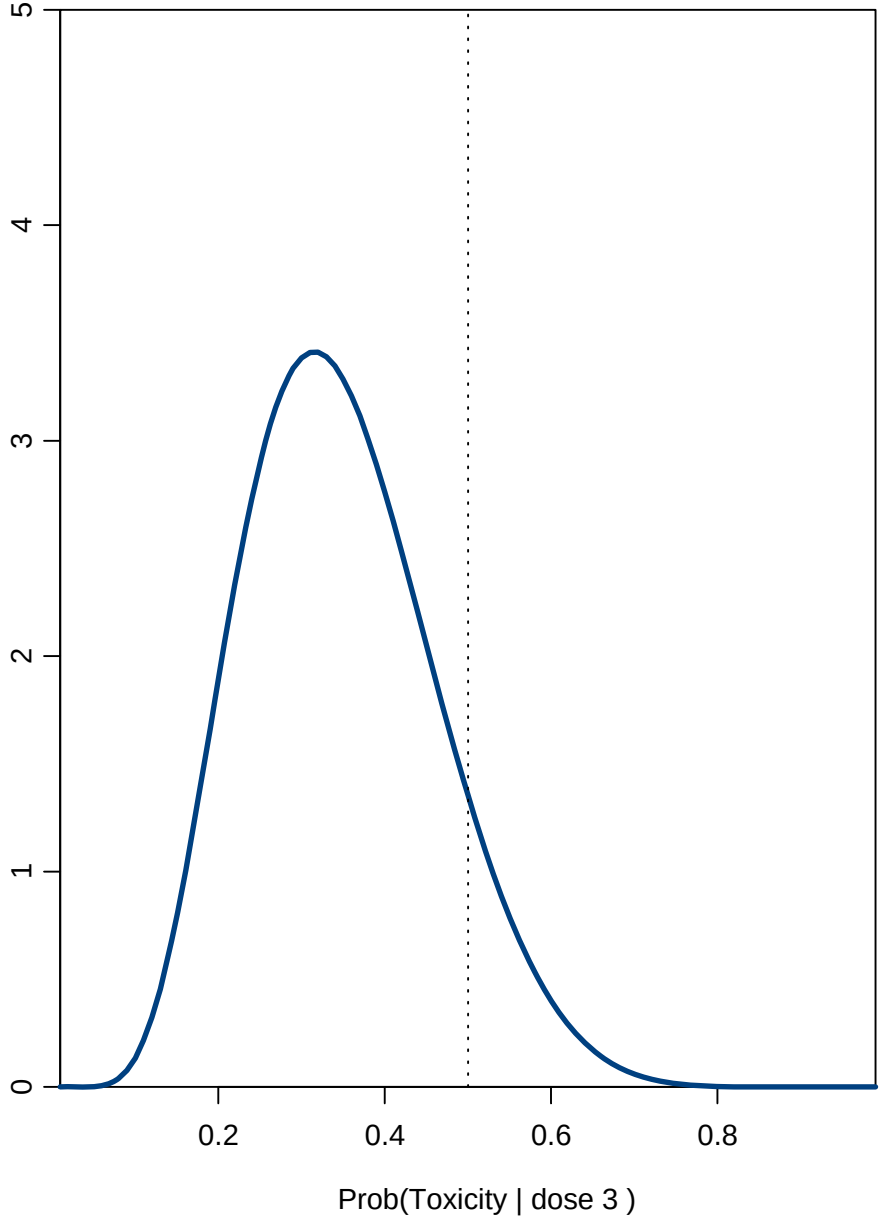

0
1
2
3
4
0.2
0.4
0.6
0.8
Prob(Efficacy | dose 3)

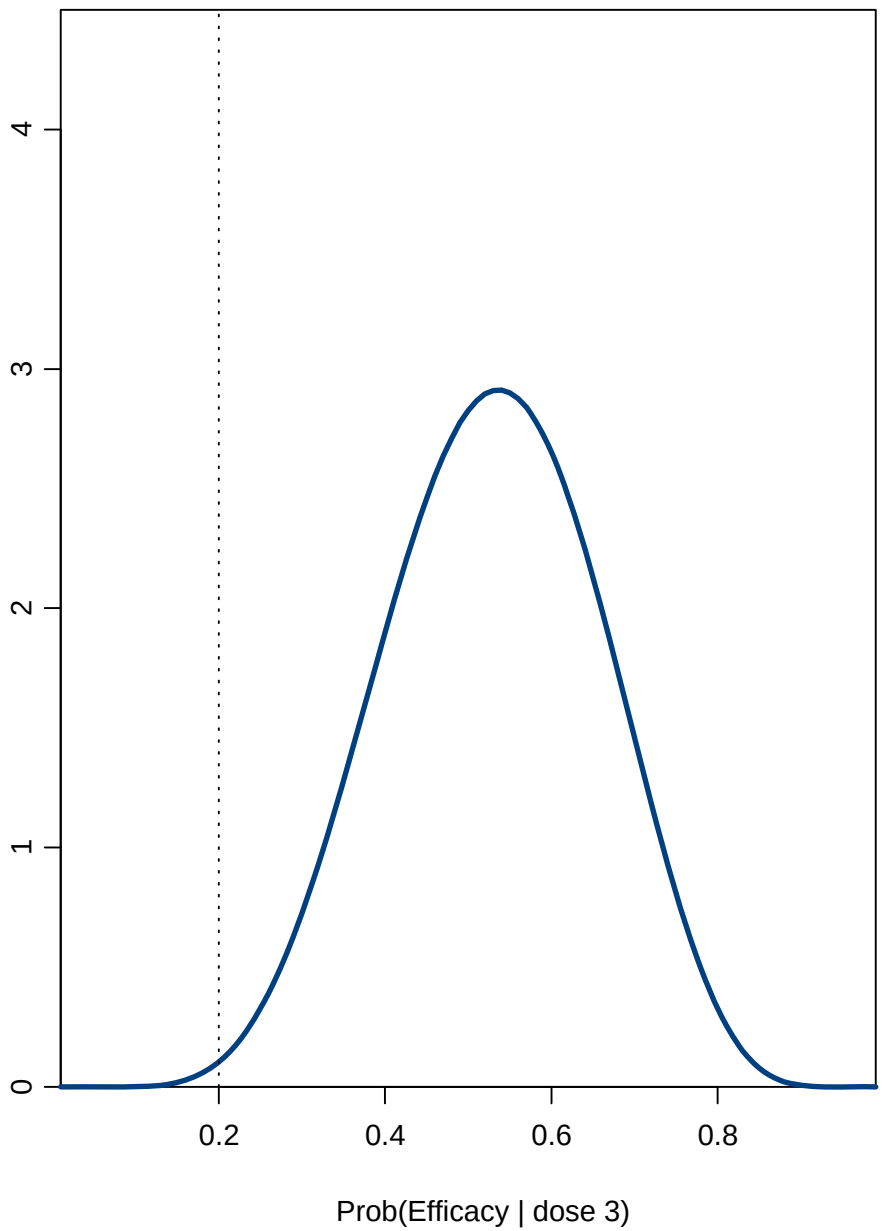
15 Patients


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose 3)

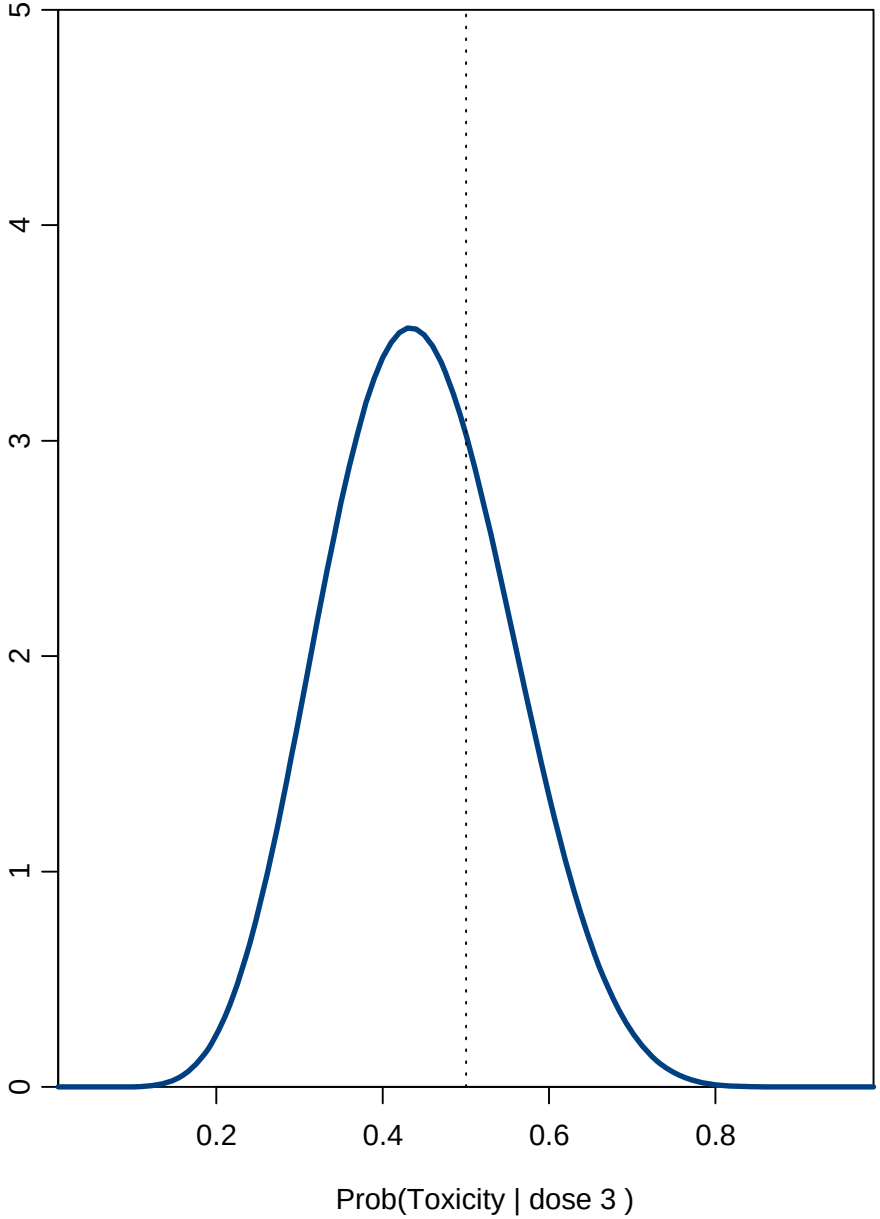

0
1
2
3
4
0.2
0.4
0.6
0.8
Prob(Efficacy | dose 3)

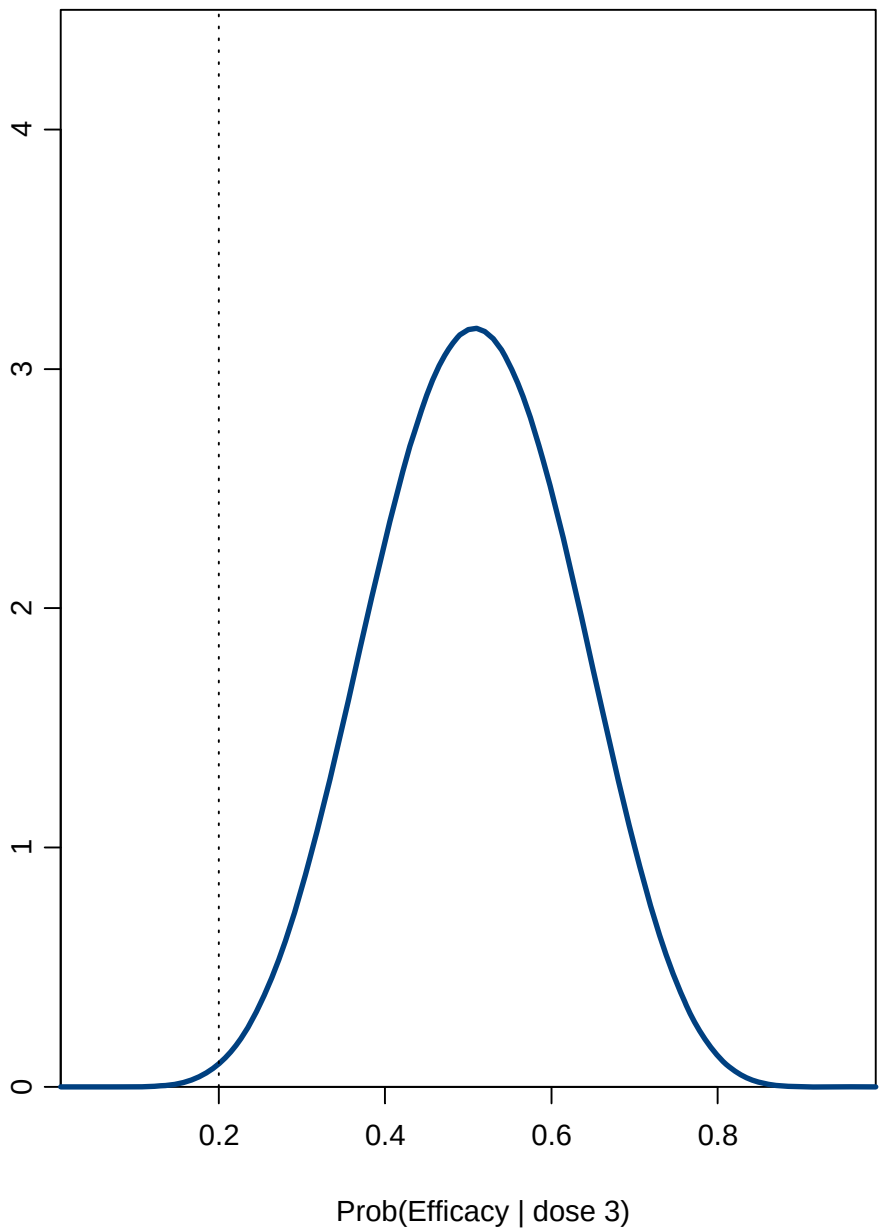
18 Patients


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose 3)

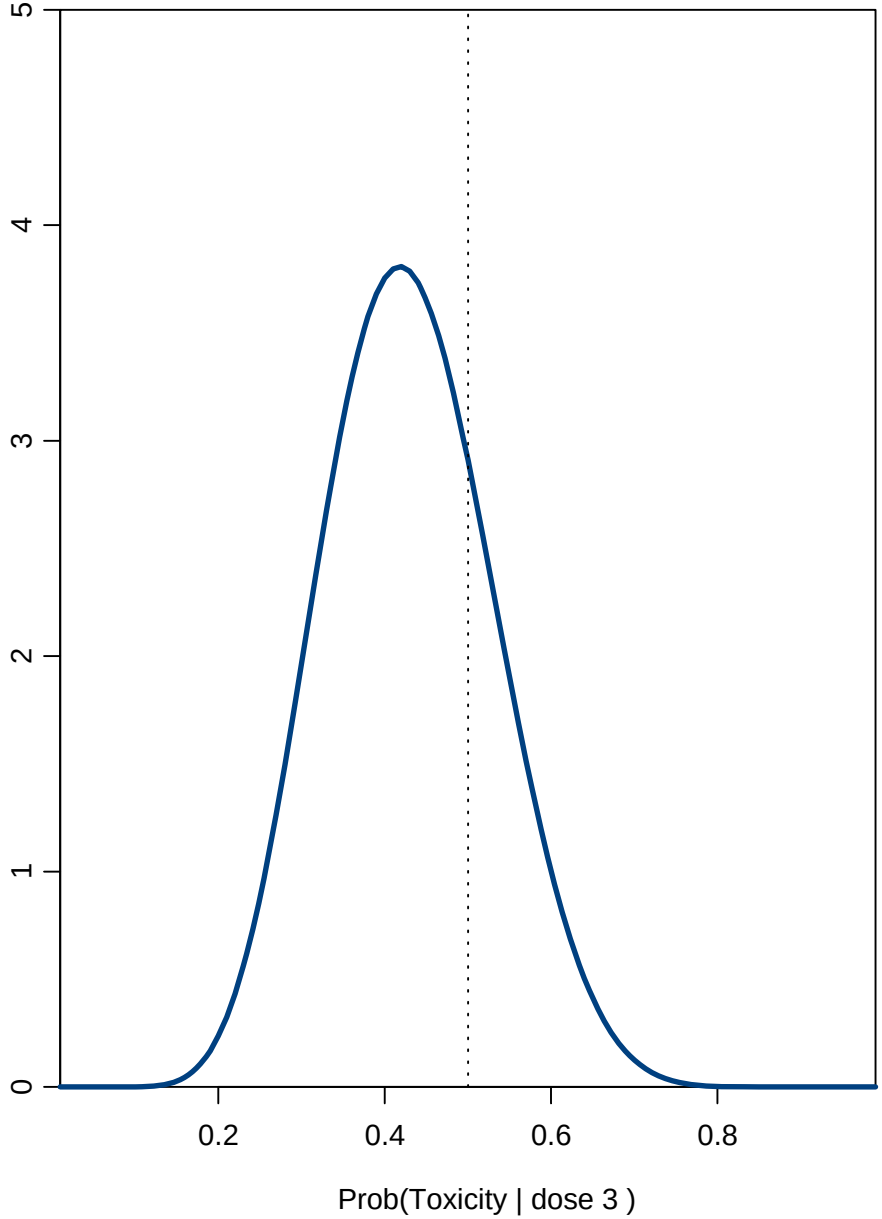

0
1
2
3
4
0.2
0.4
0.6
0.8
Prob(Efficacy | dose 3)

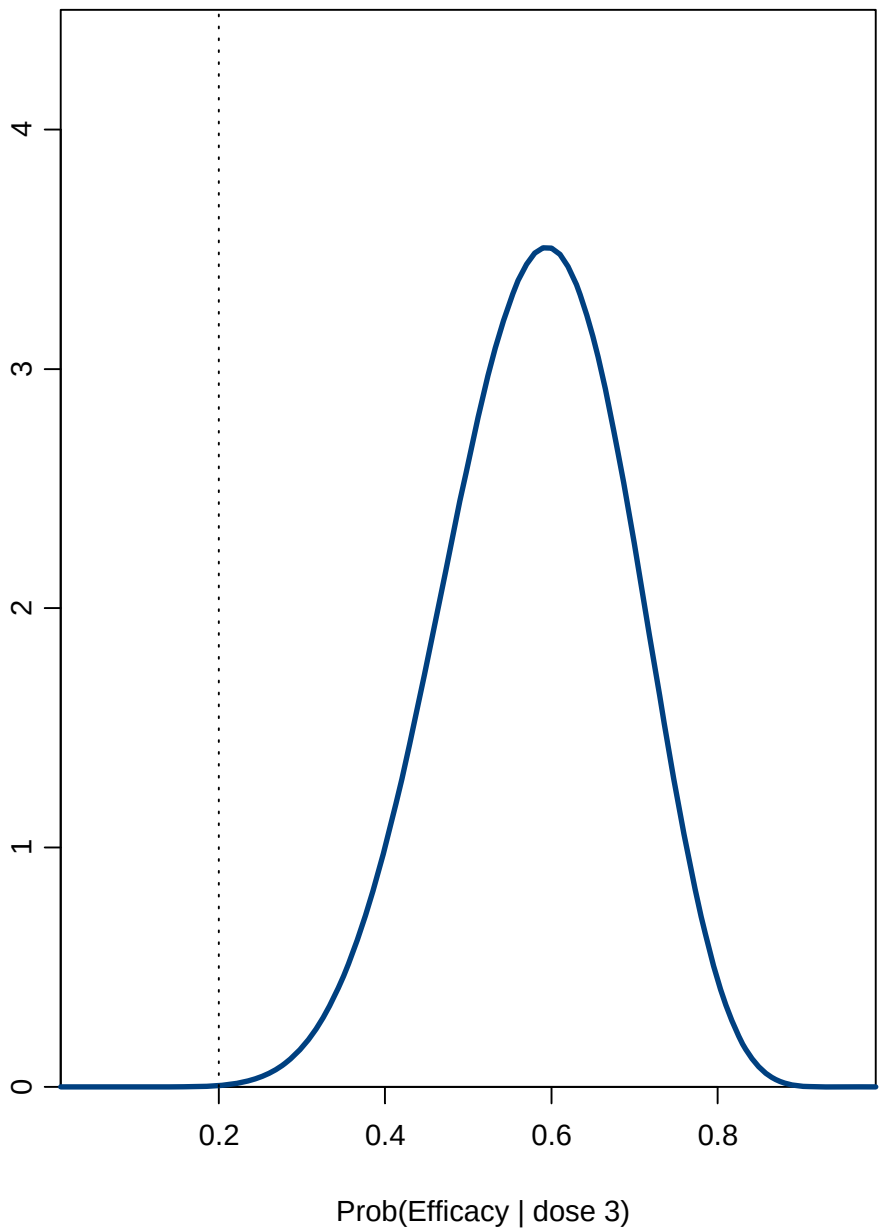
21 Patients


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose 3)

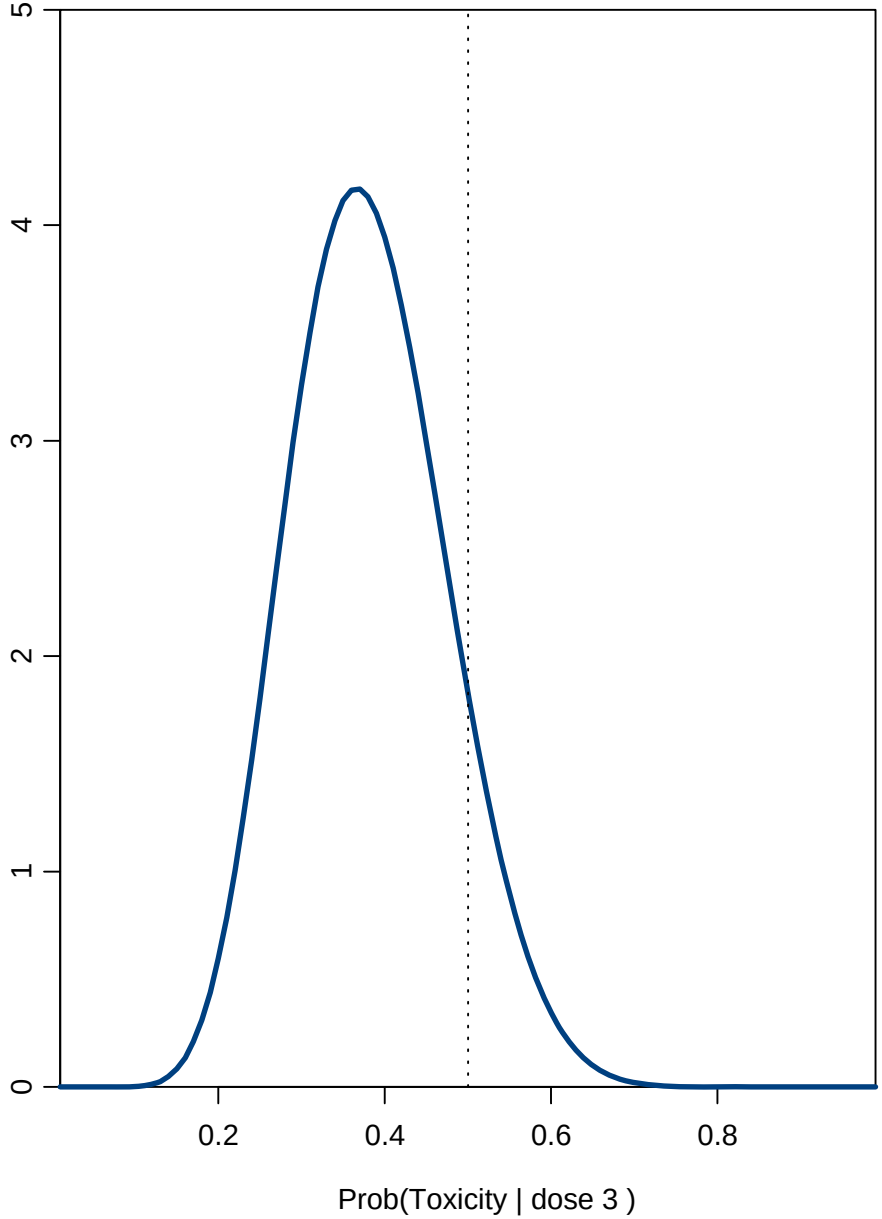

0
1
2
3
4
0.2
0.4
0.6
0.8
Prob(Efficacy | dose 3)

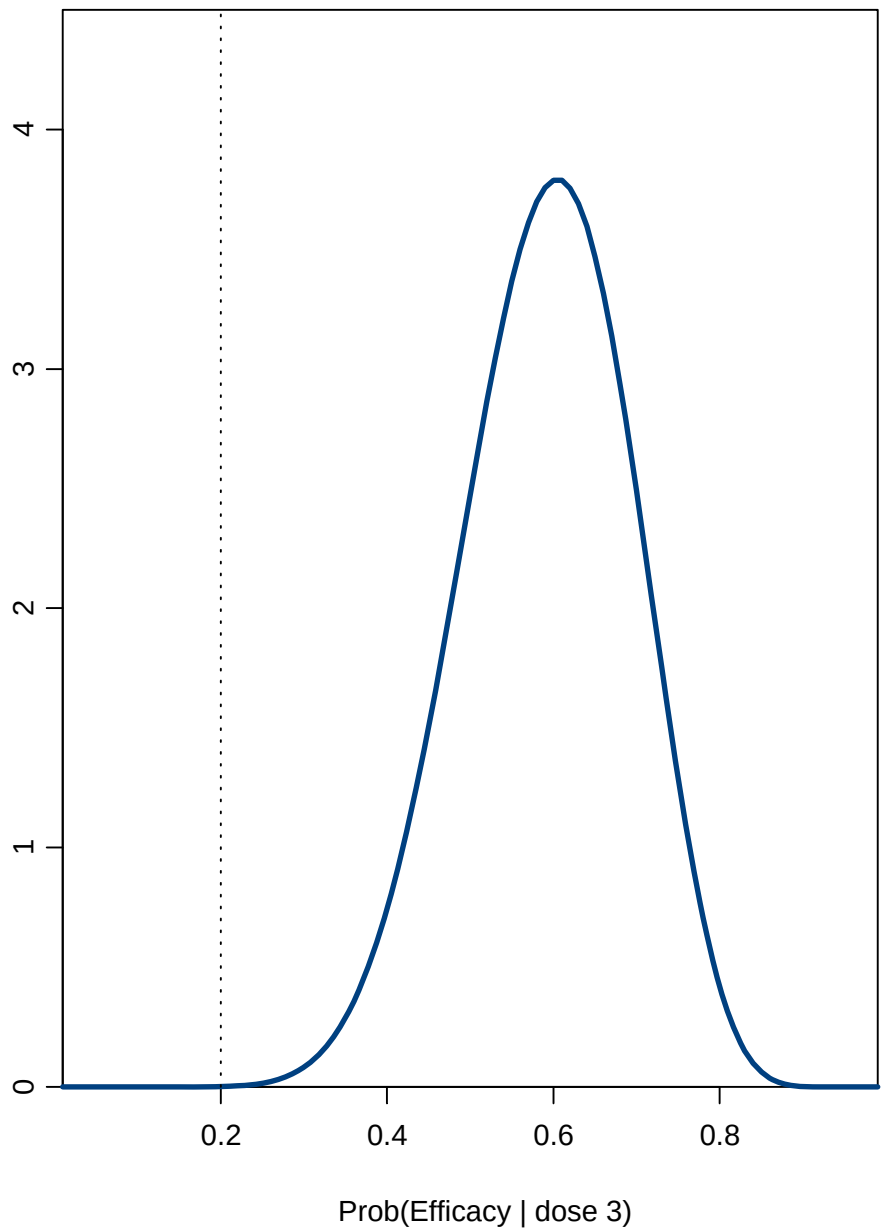
24 Patients


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose 3)

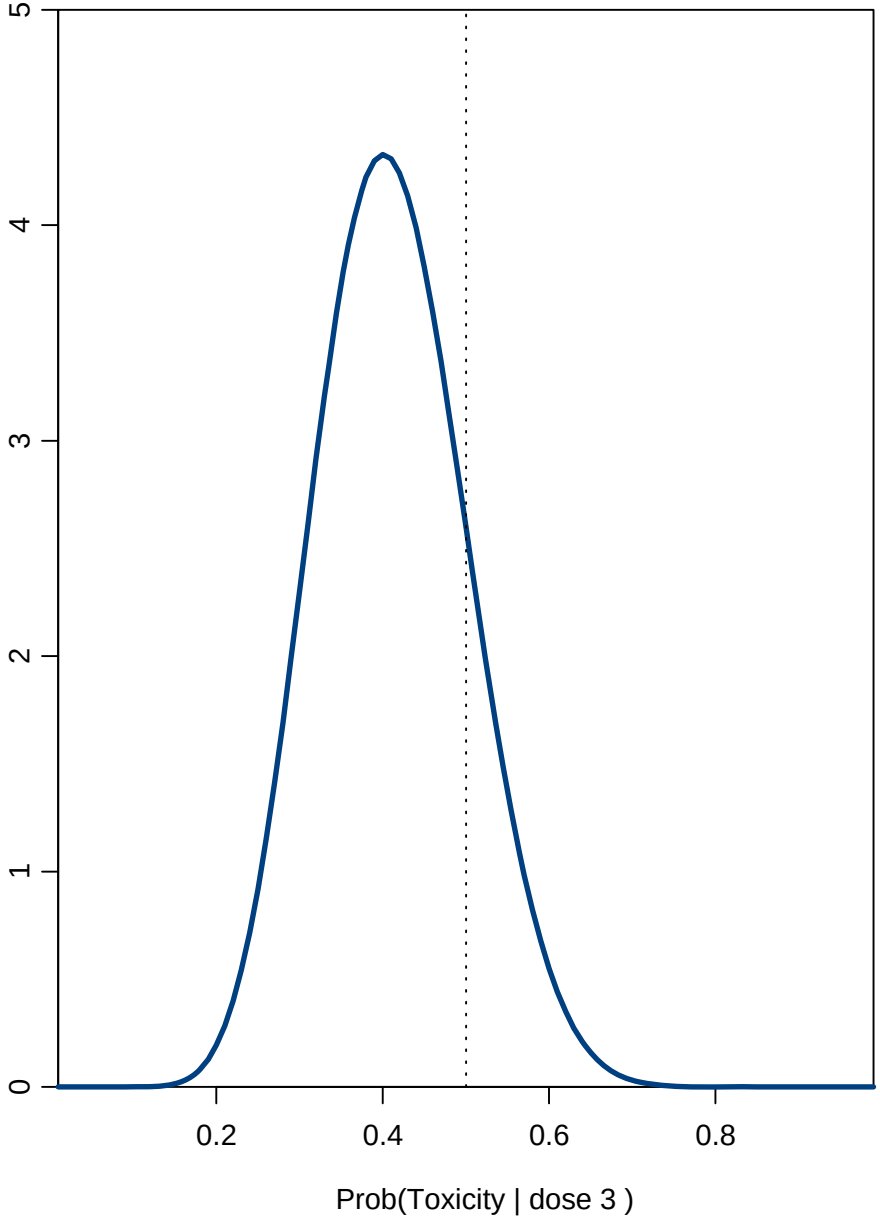

0
1
2
3
4
0.2
0.4
0.6
0.8
Prob(Efficacy | dose 3)

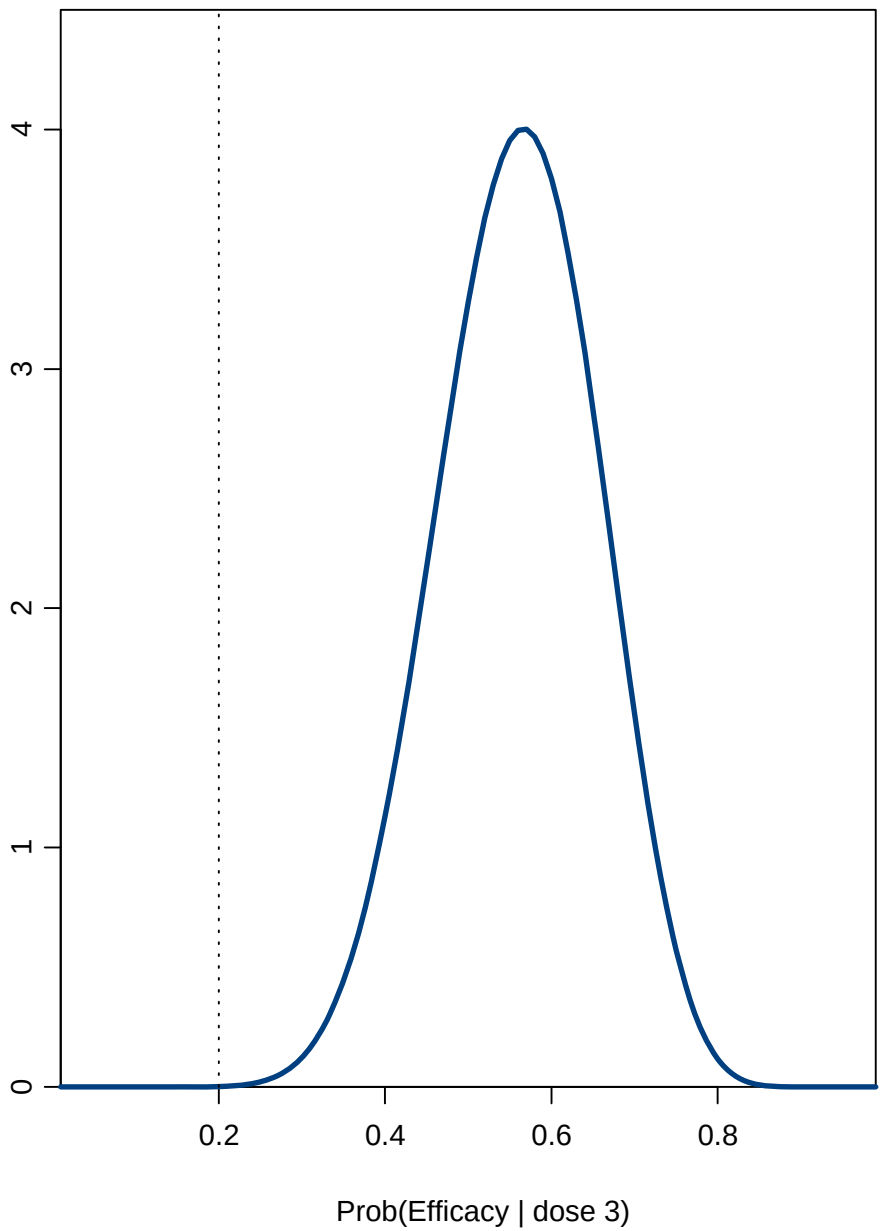
27 Patients


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose 3)

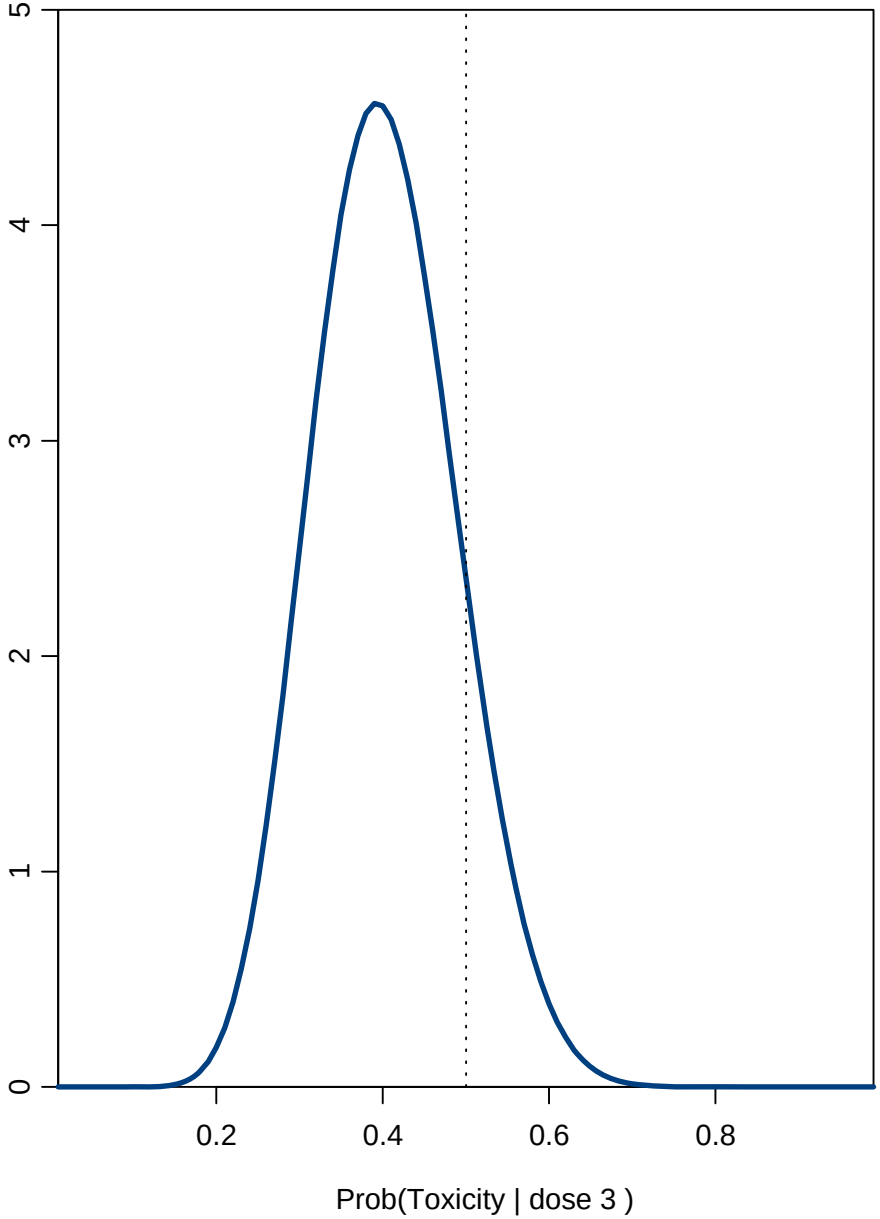

0
1
2
3
4
0.2
0.4
0.6
0.8
Prob(Efficacy | dose 3)

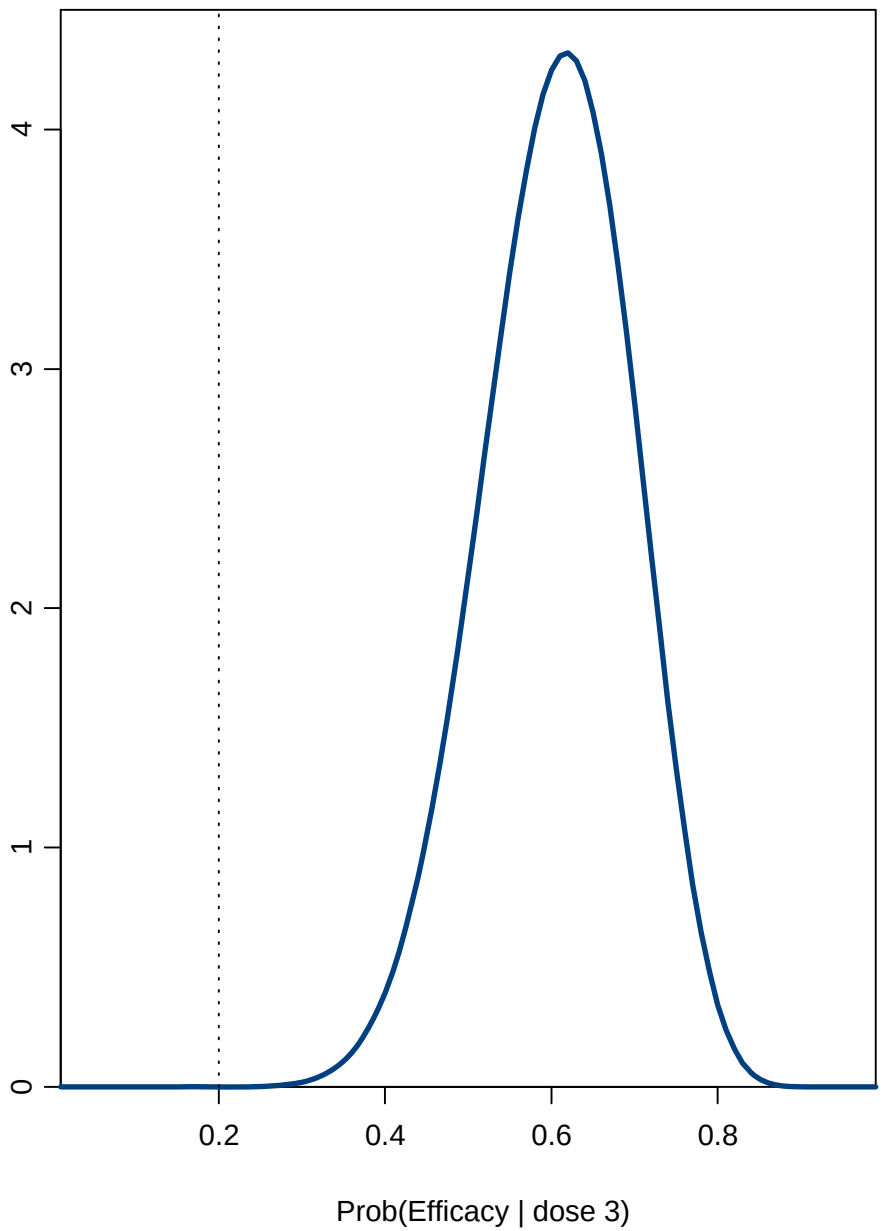
30 Patients


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose 3)

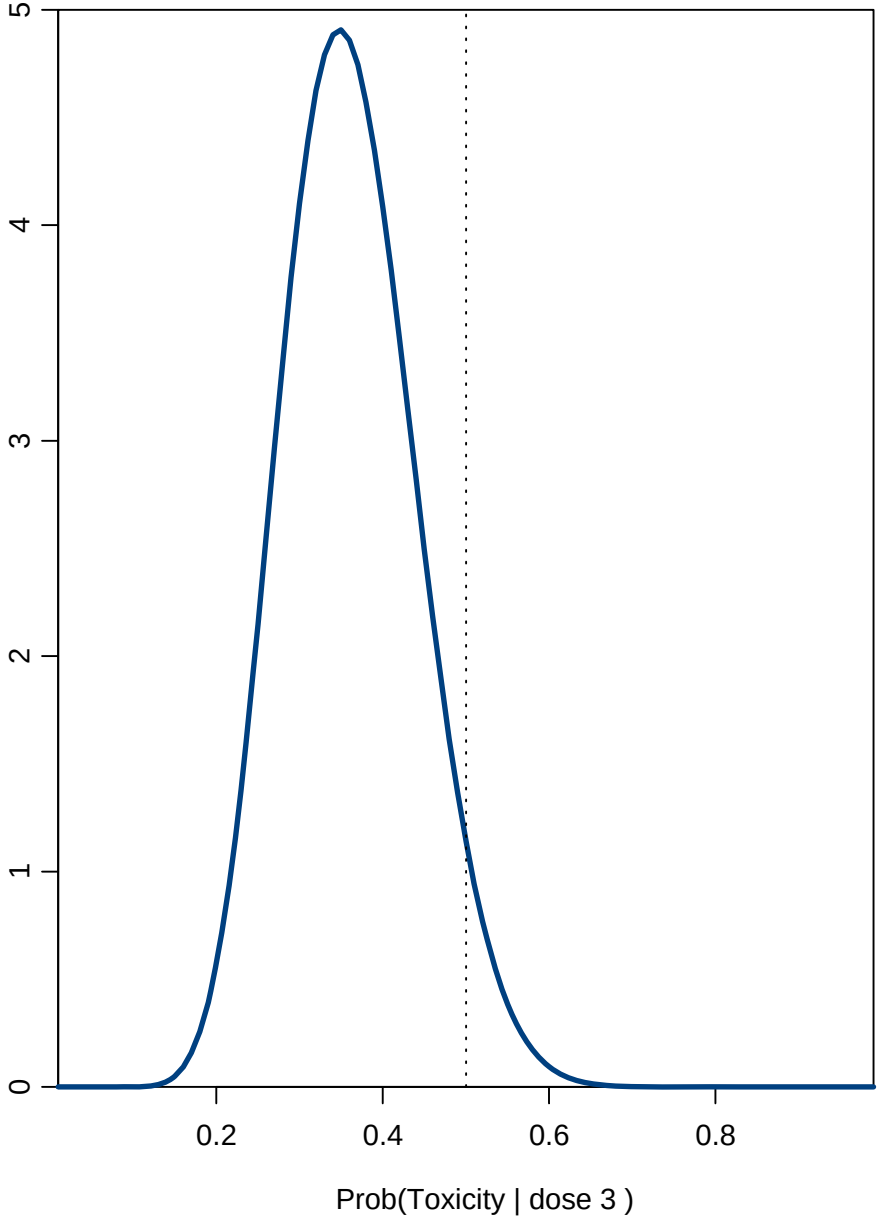

0
1
2
3
4
0.2
0.4
0.6
0.8
Prob(Efficacy | dose 3)

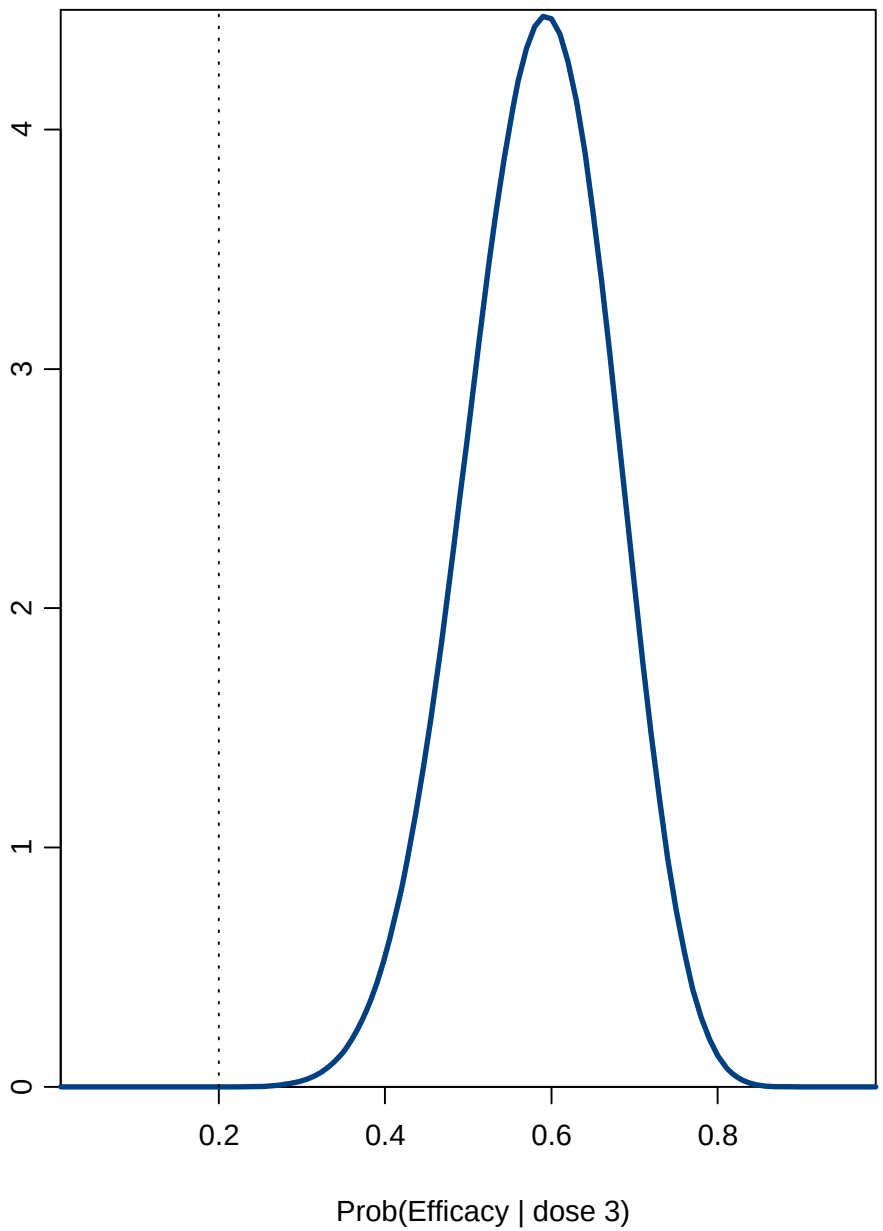
33 Patients


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose 3)

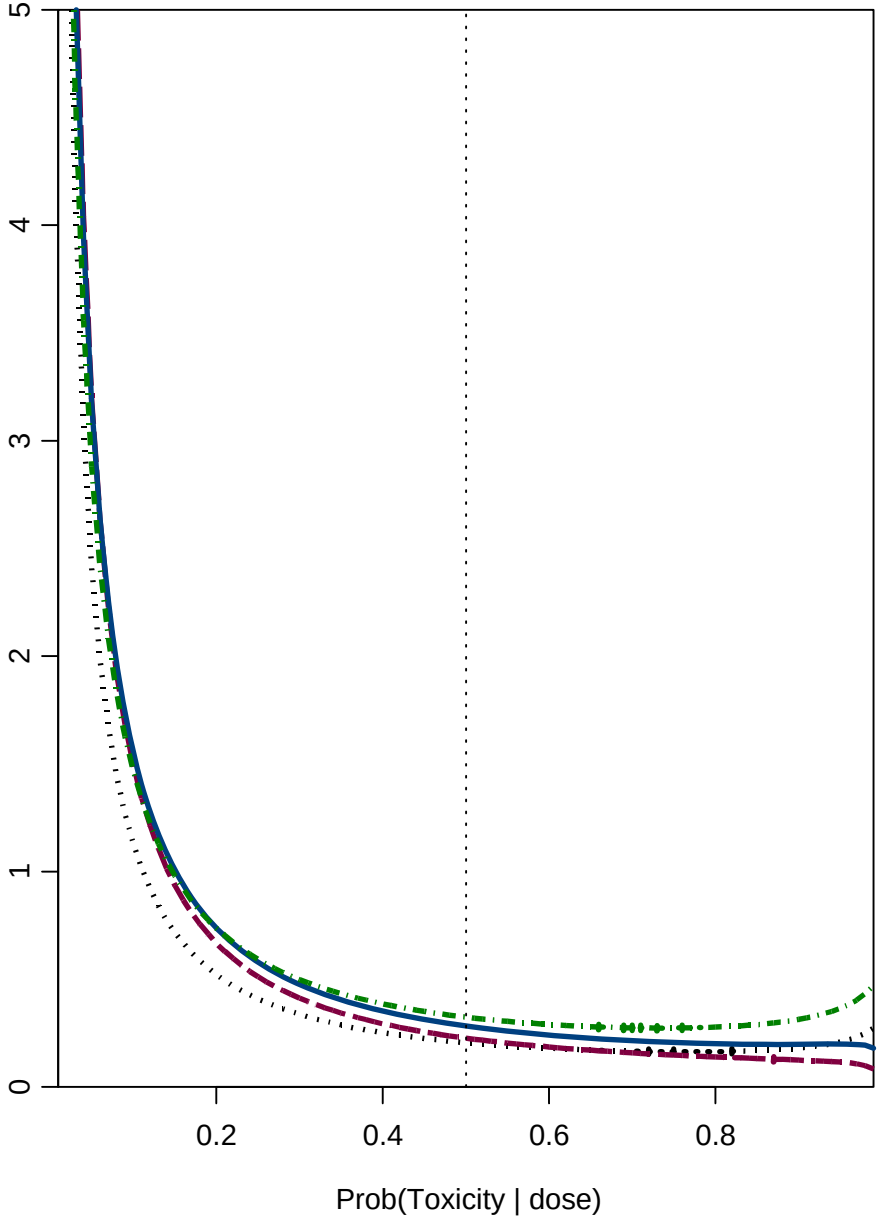

0
1
2
3
4
0.2
0.4
0.6
0.8
Prob(Efficacy | dose 3)

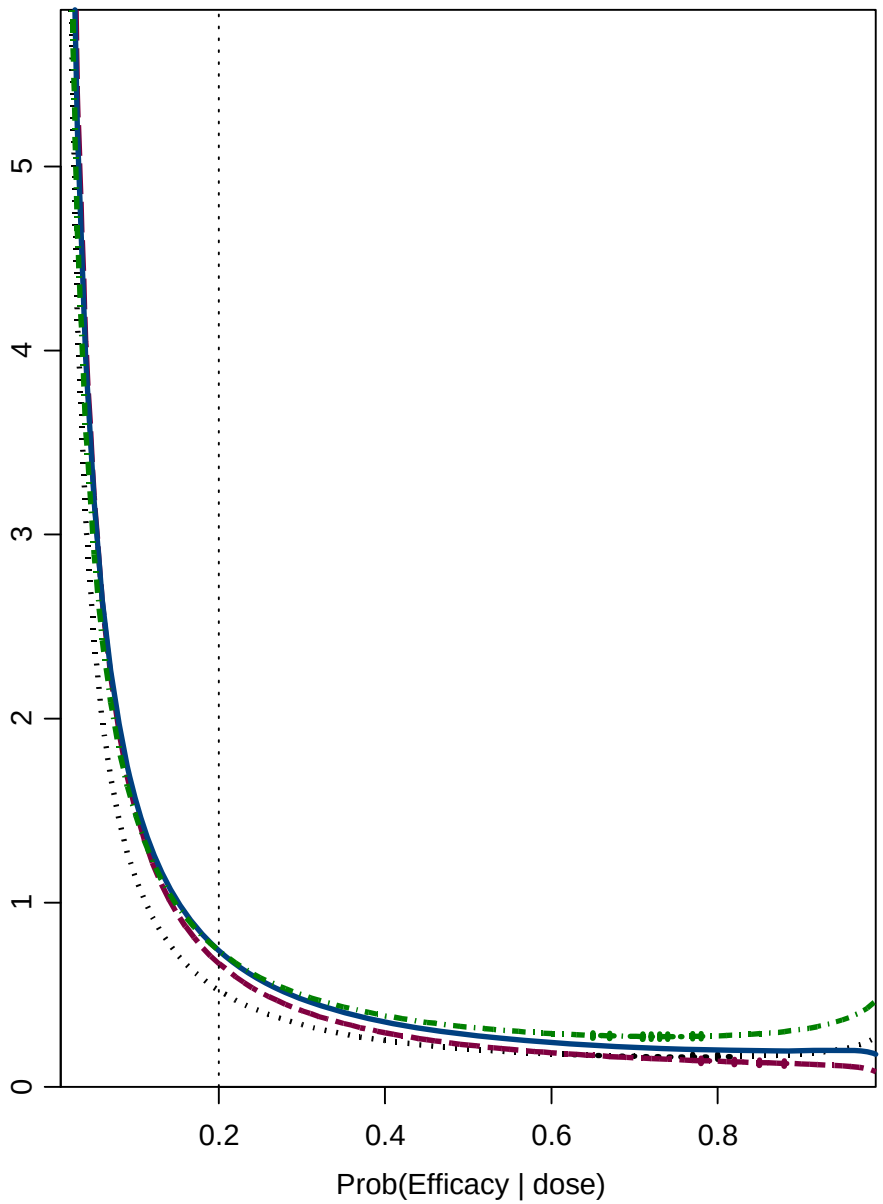
36 Patients

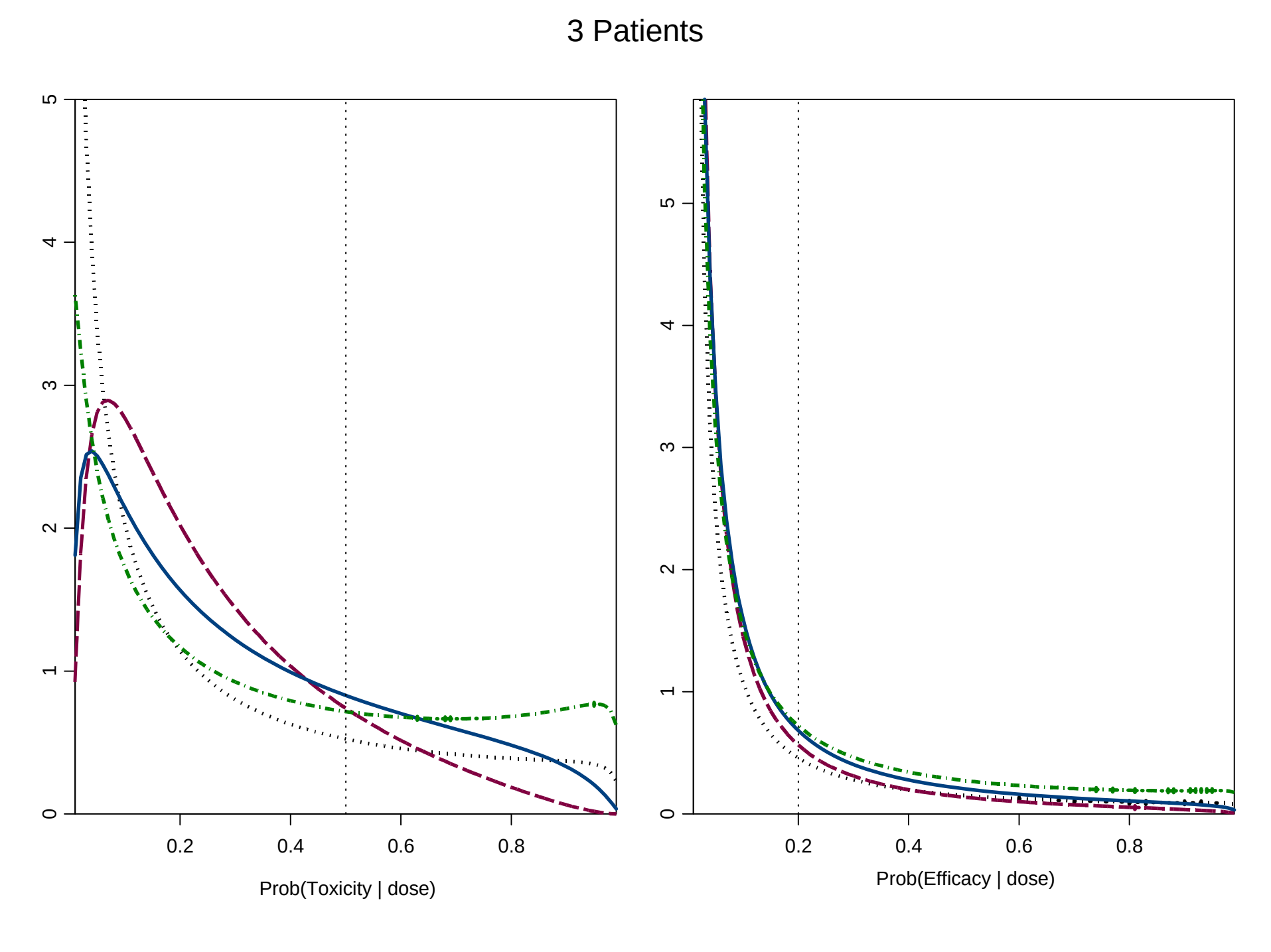

0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose 3)


0
1
2
3
4
0.2
0.4
0.6
0.8
Prob(Efficacy | dose 3)

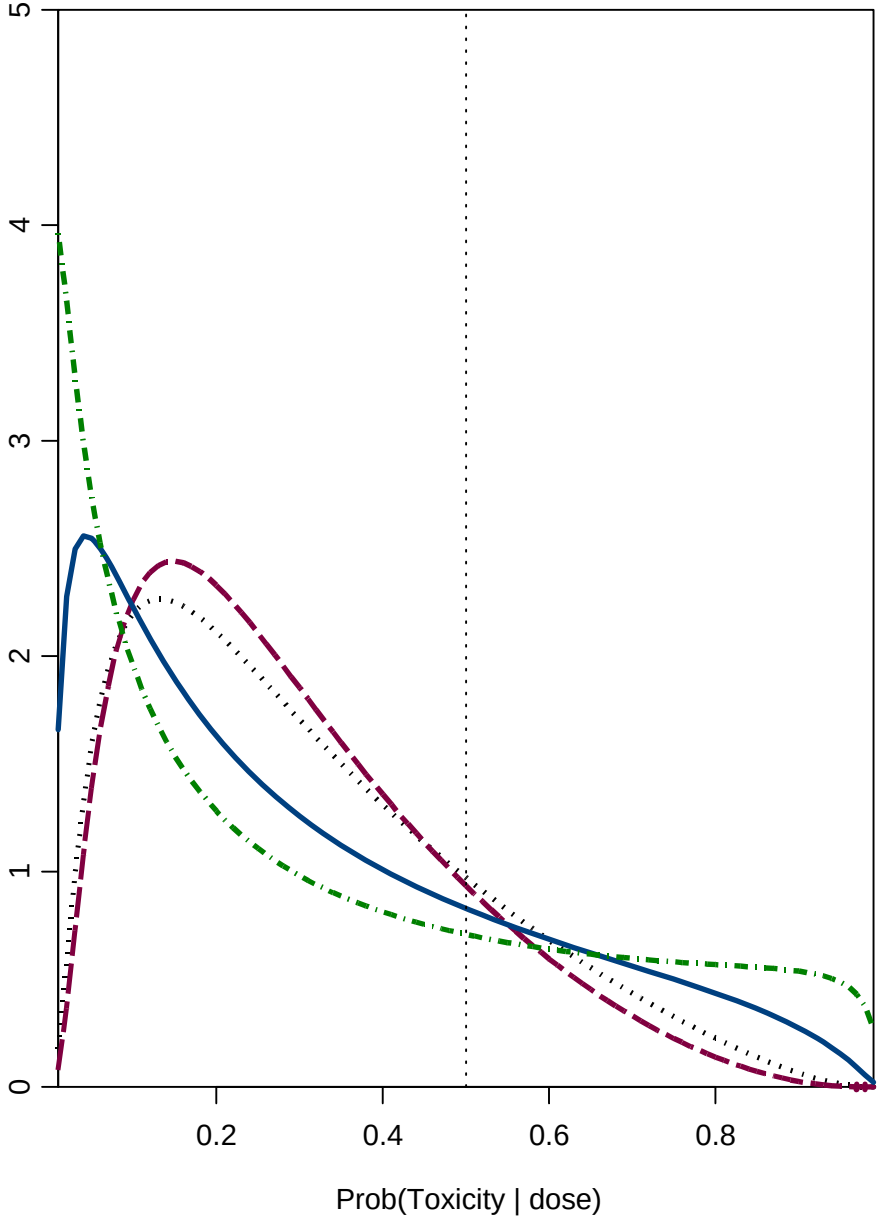
Prior Distributions

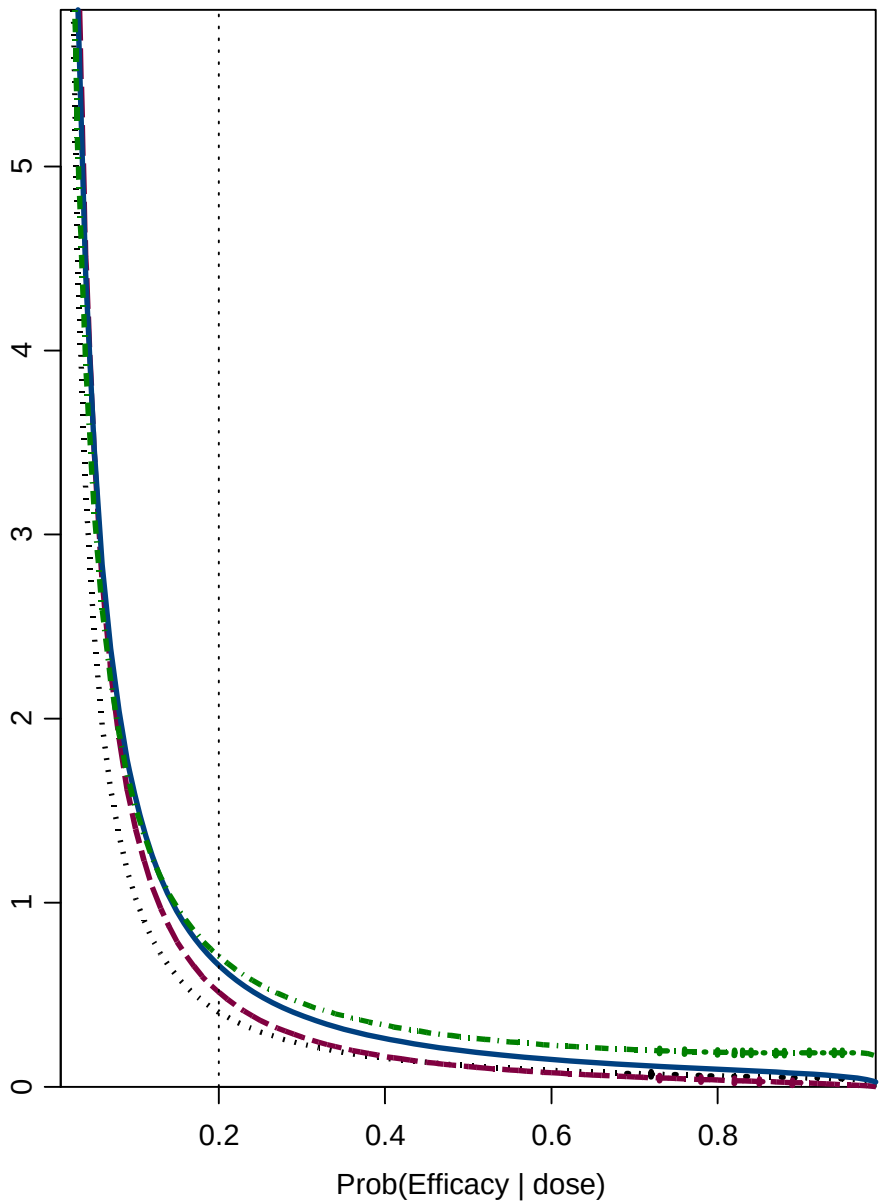




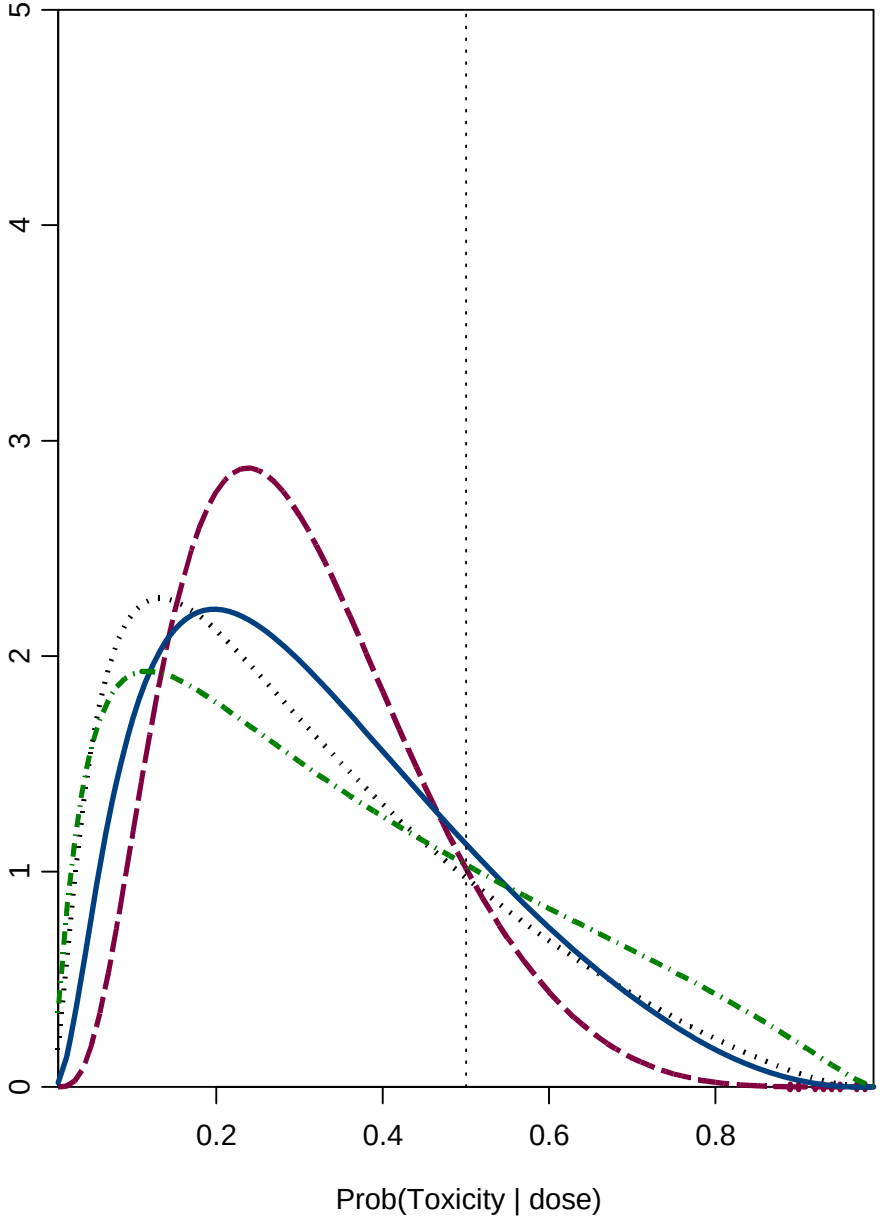

3 Patients
0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose)
Prob(Efficacy | dose)

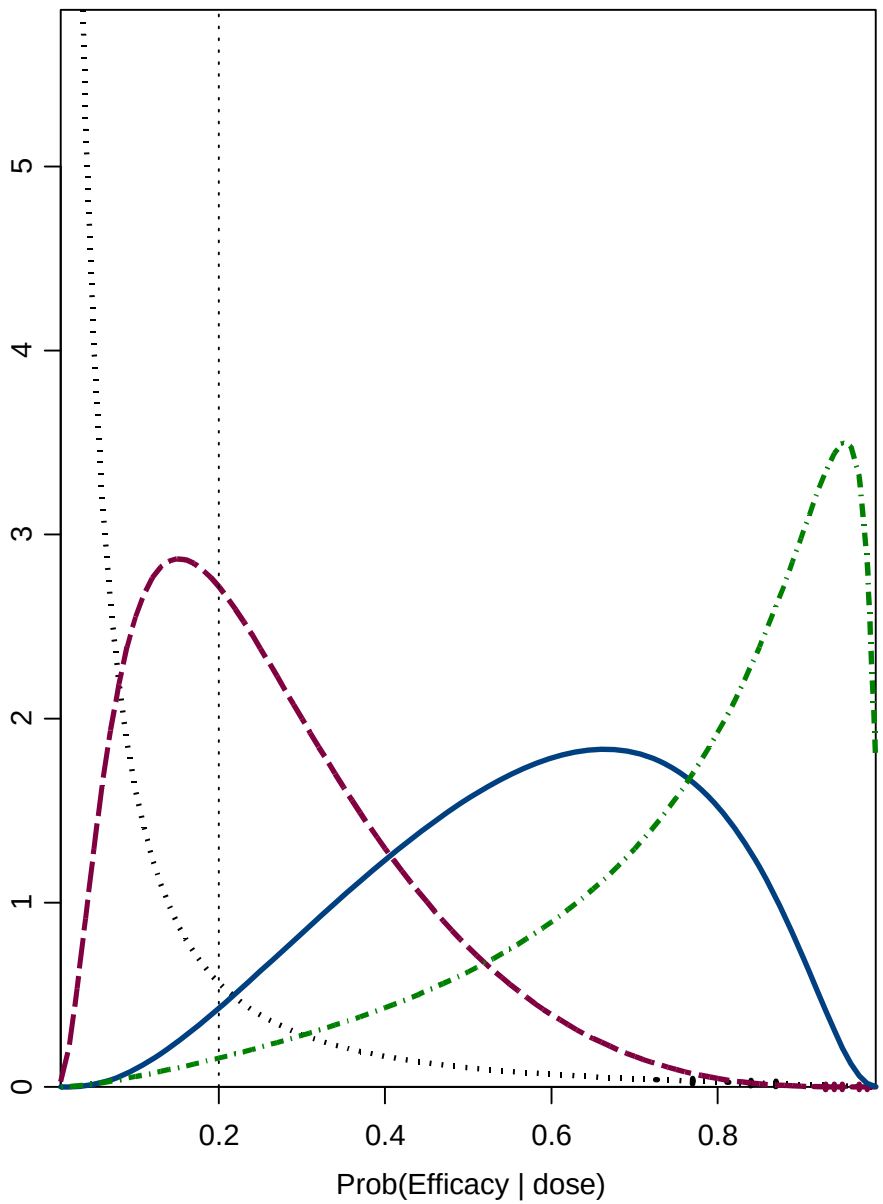
6 Patients


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose)

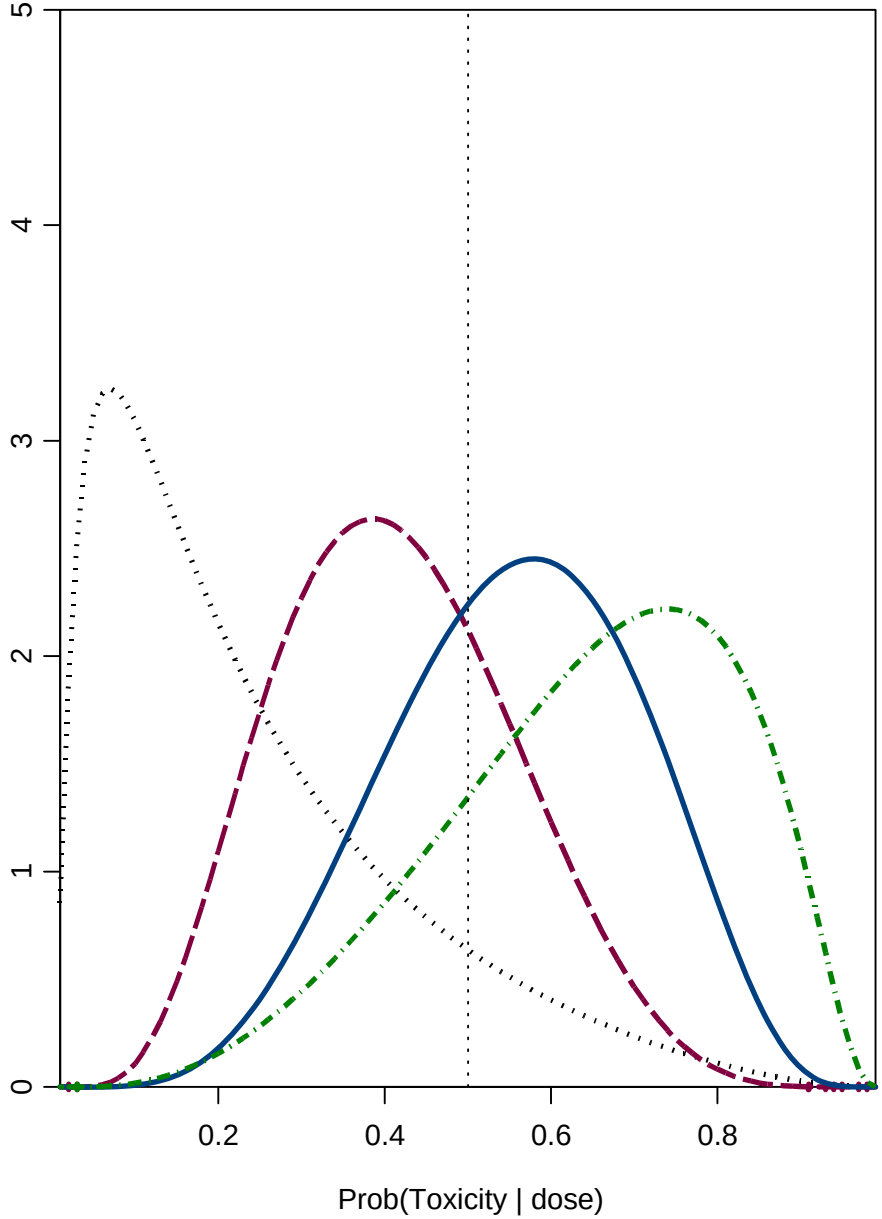

0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Efficacy | dose)

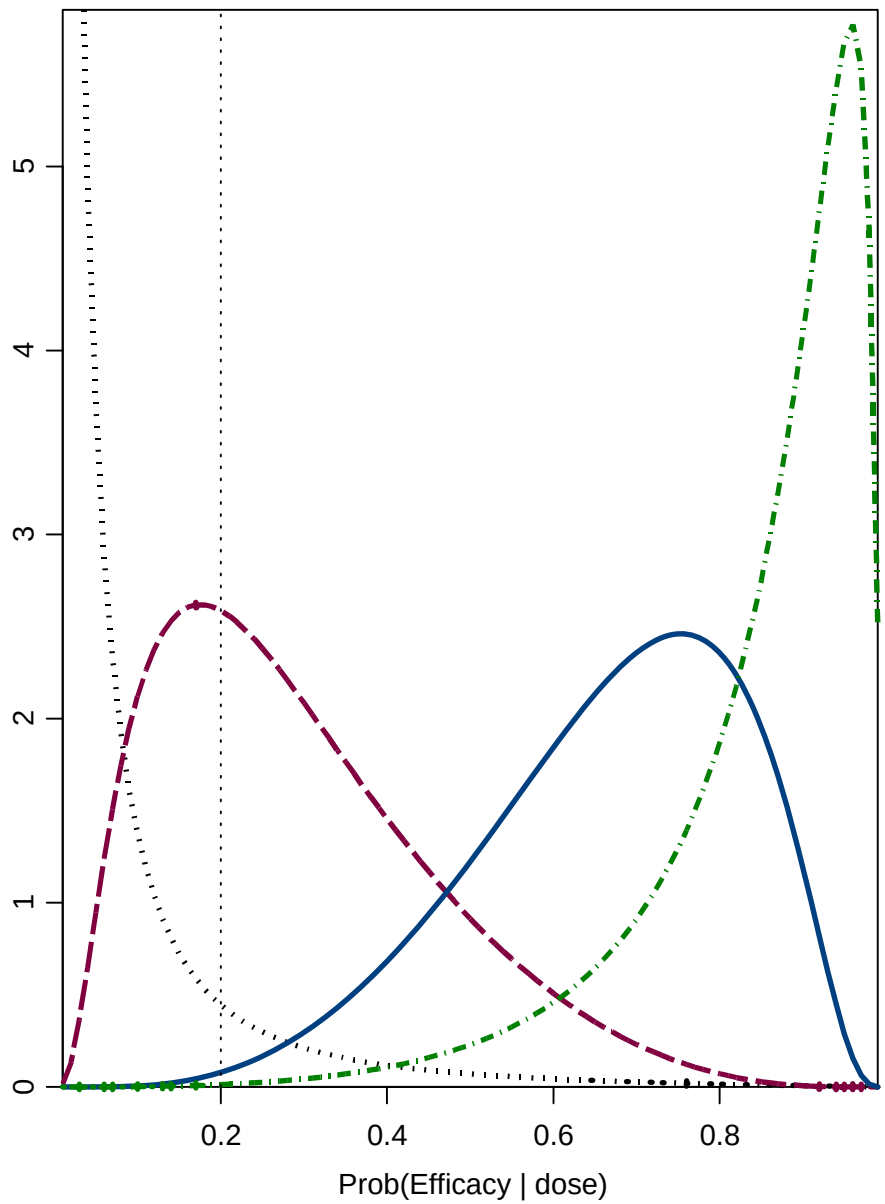
9 Patients


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose)

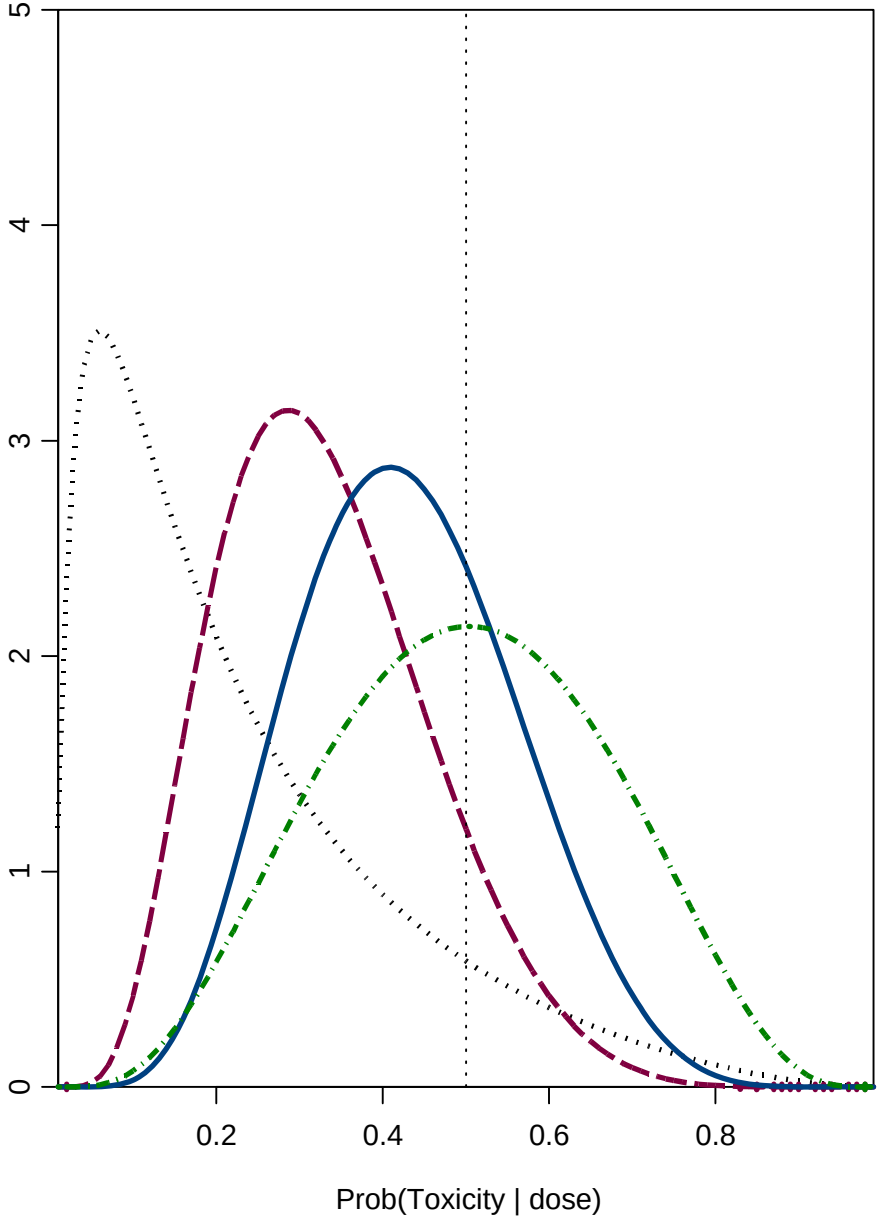

0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Efficacy | dose)

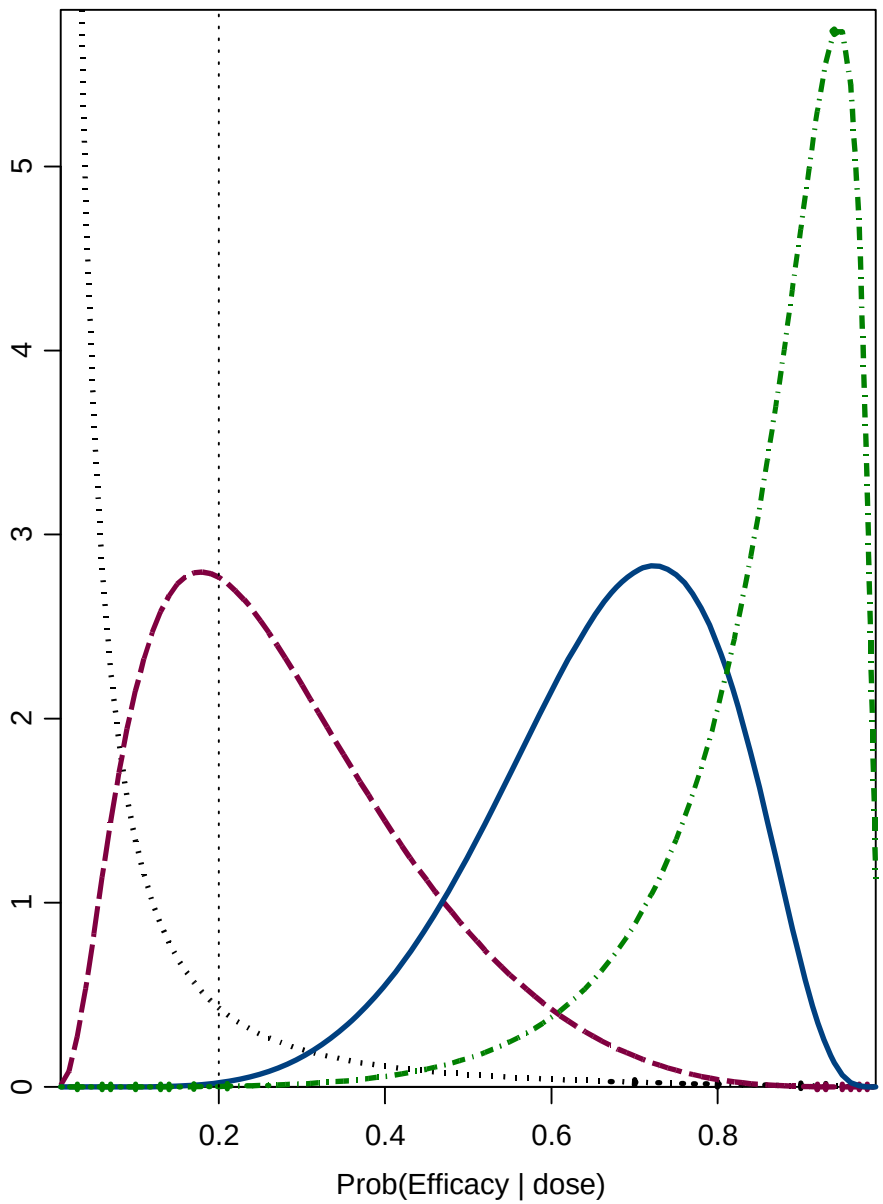
12 Patients


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose)

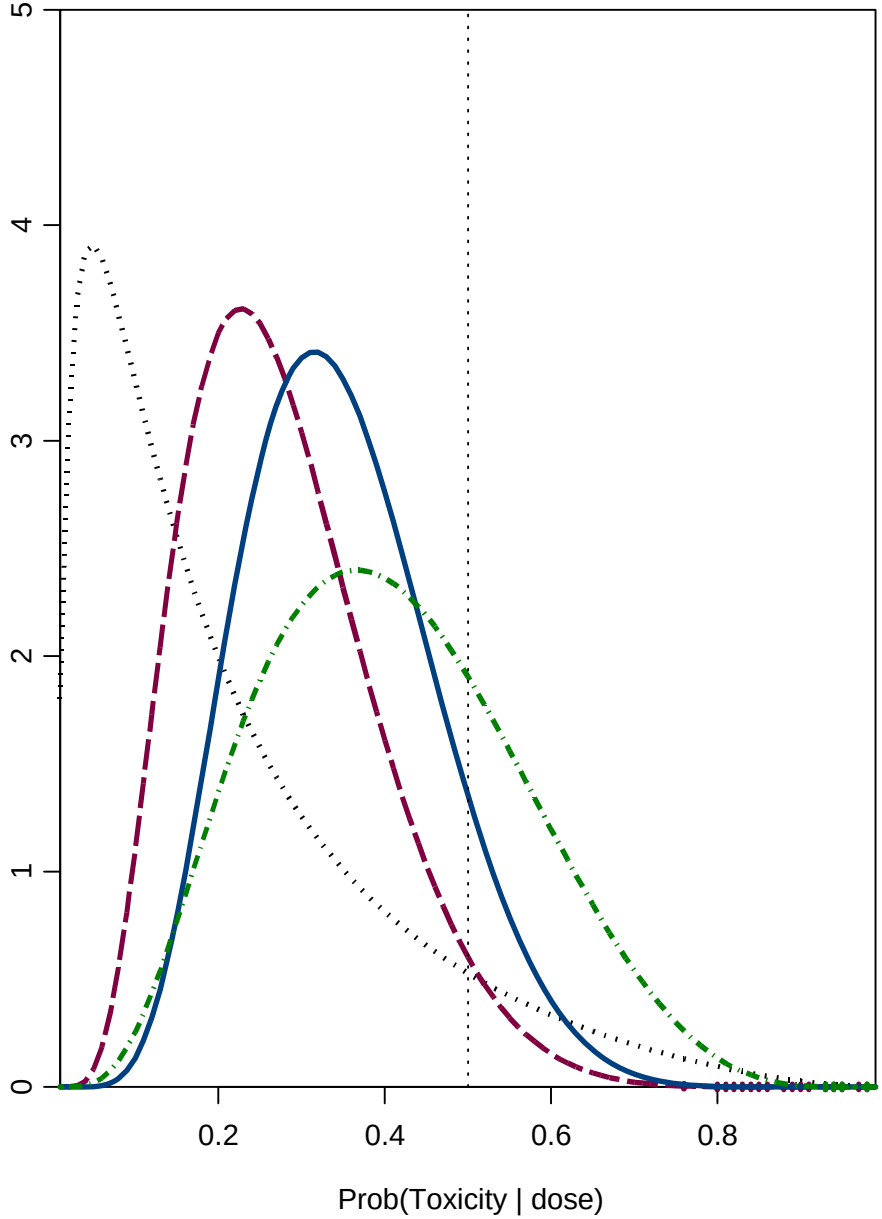

0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Efficacy | dose)

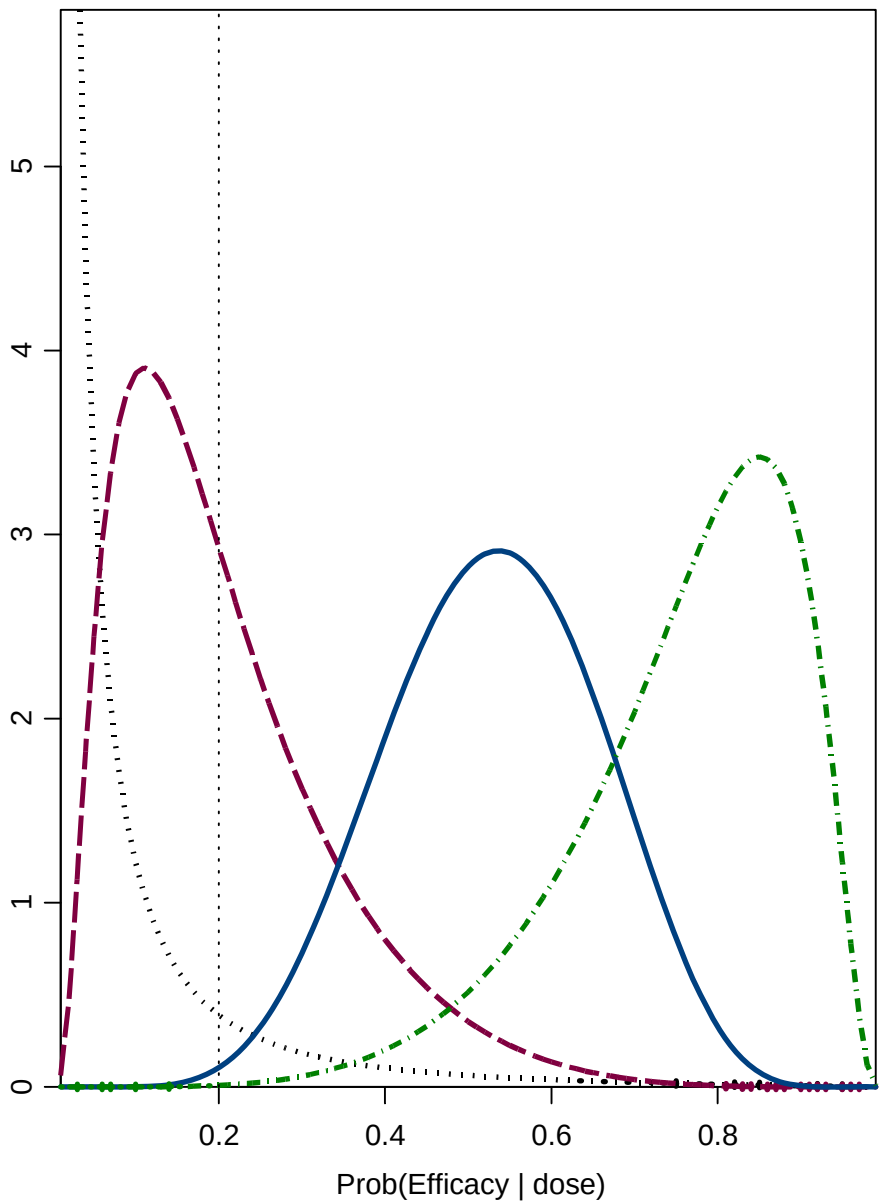
15 Patients


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose)

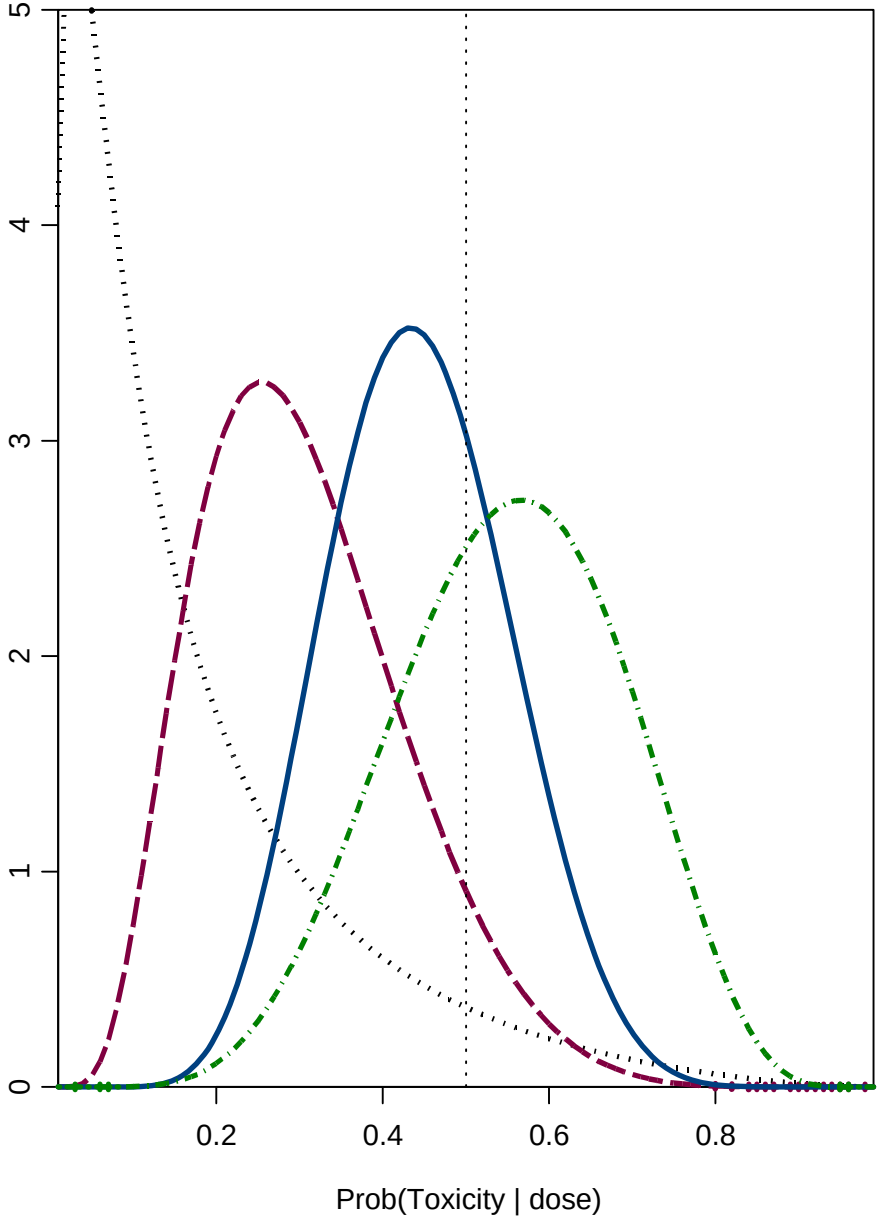

0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Efficacy | dose)

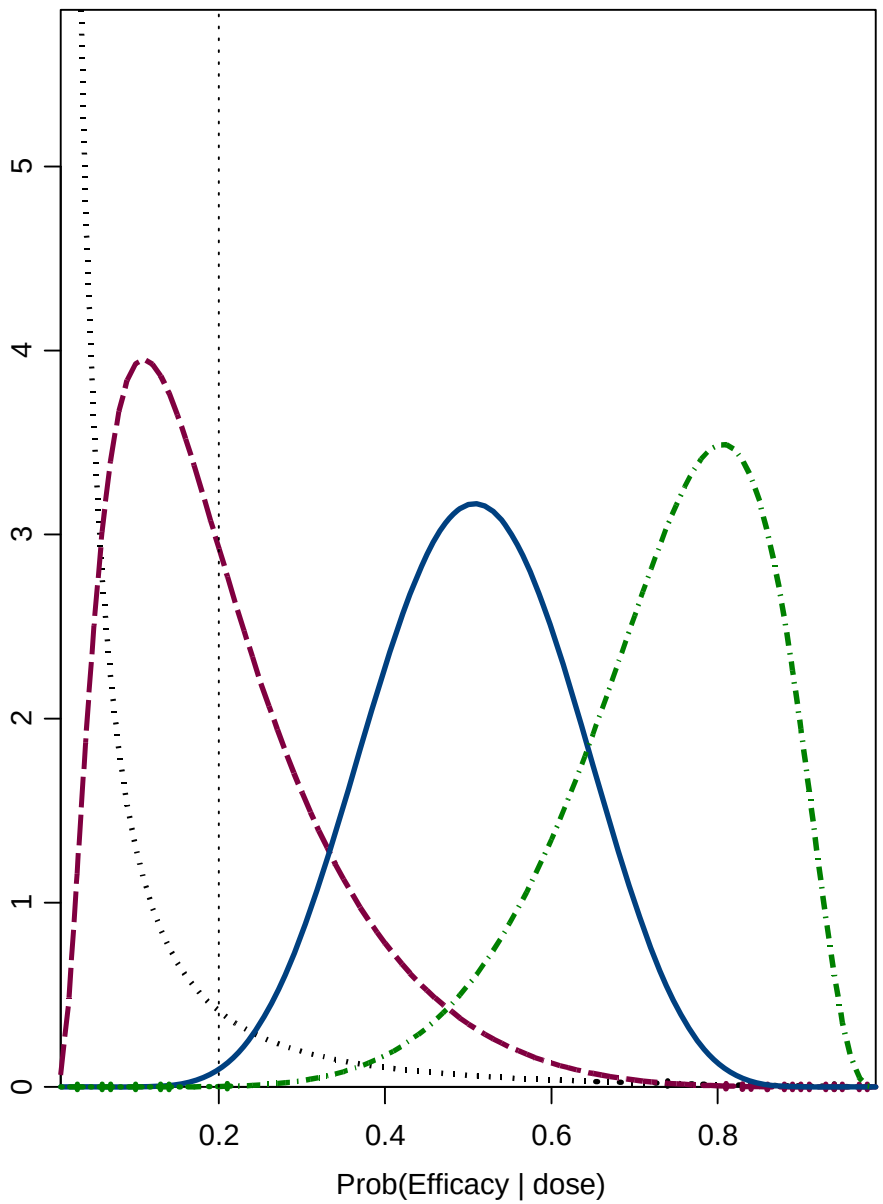
18 Patients


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose)

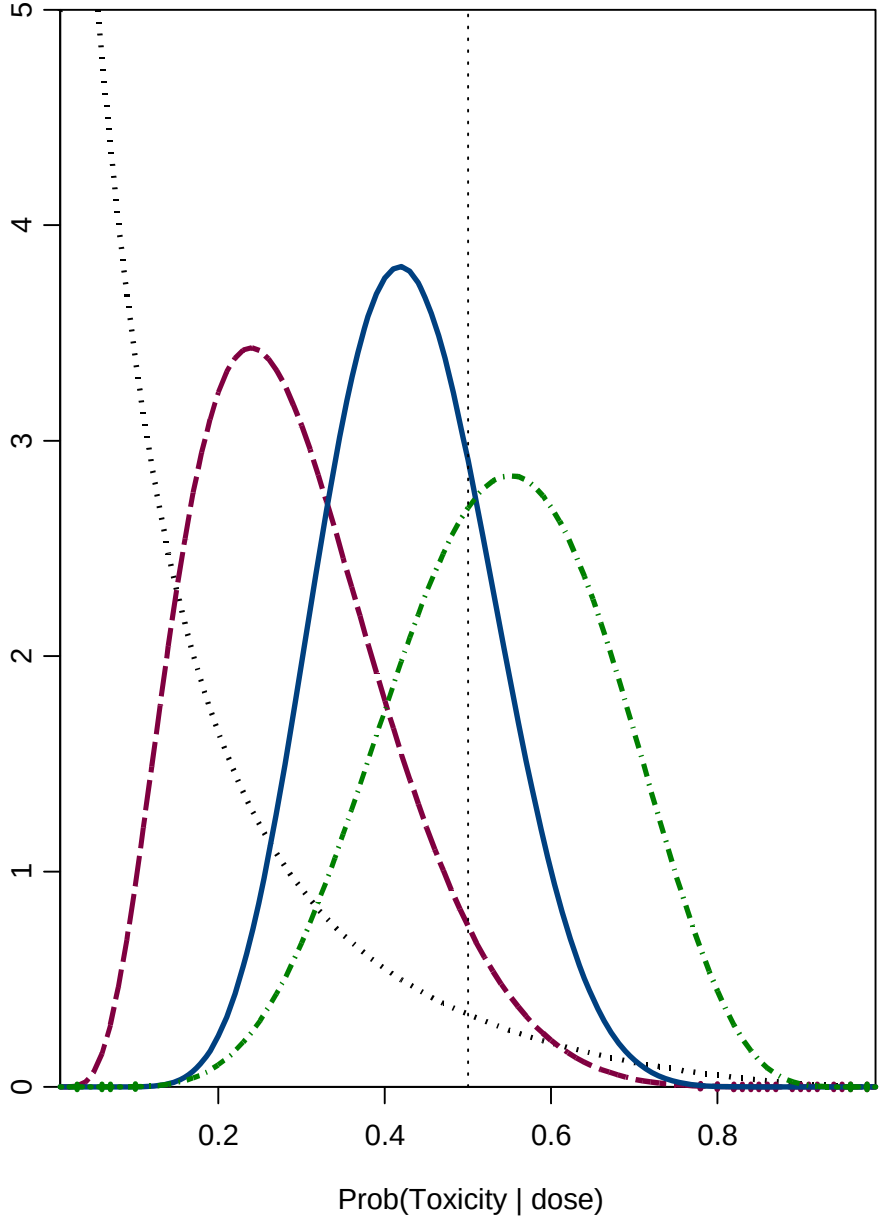

0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Efficacy | dose)

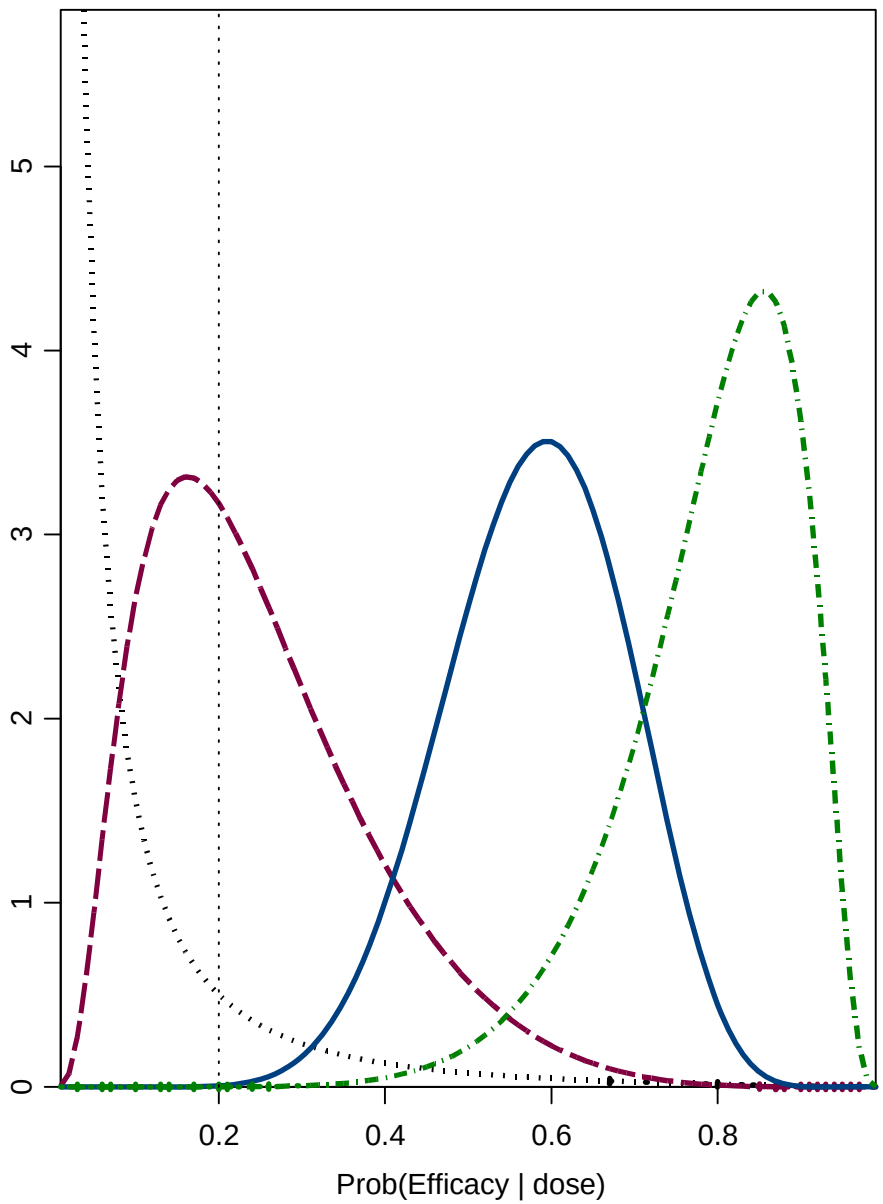
21 Patients


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose)

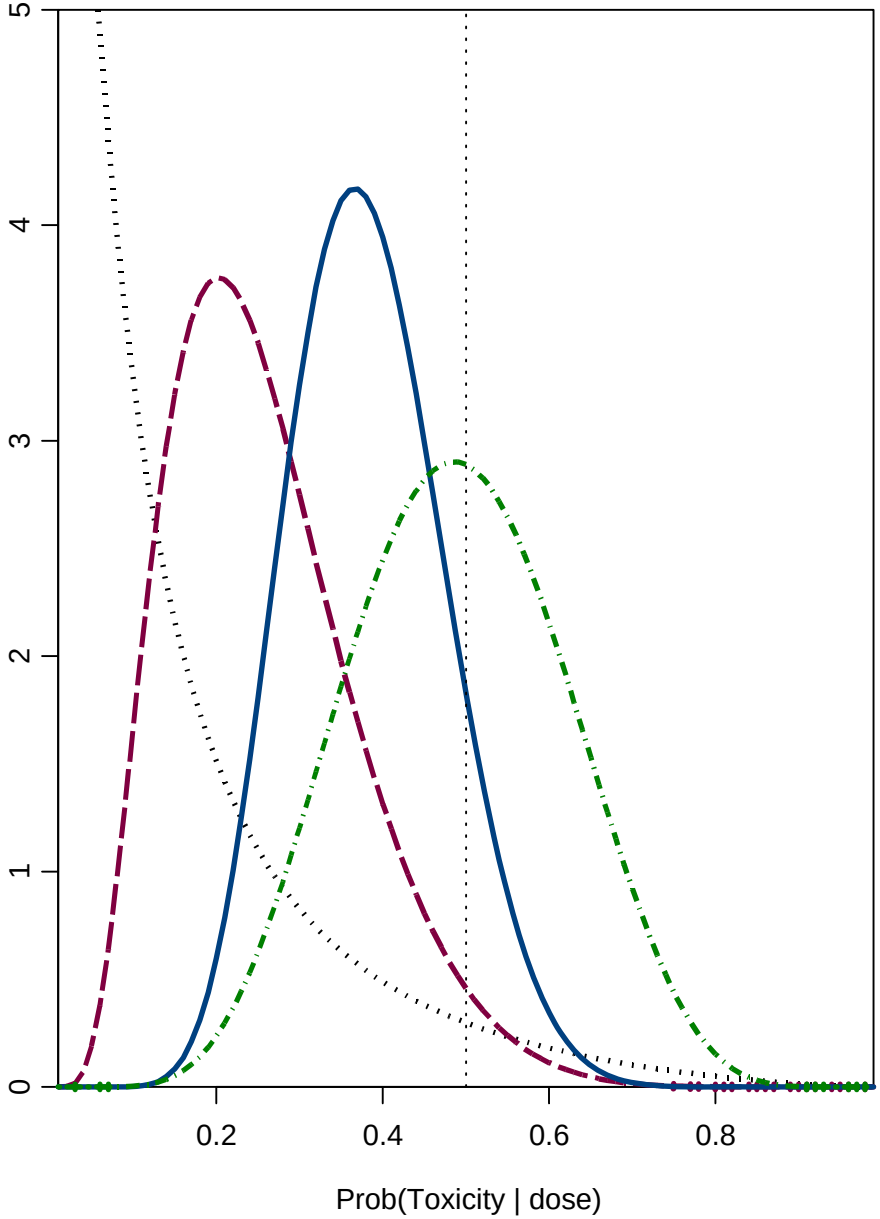

0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Efficacy | dose)

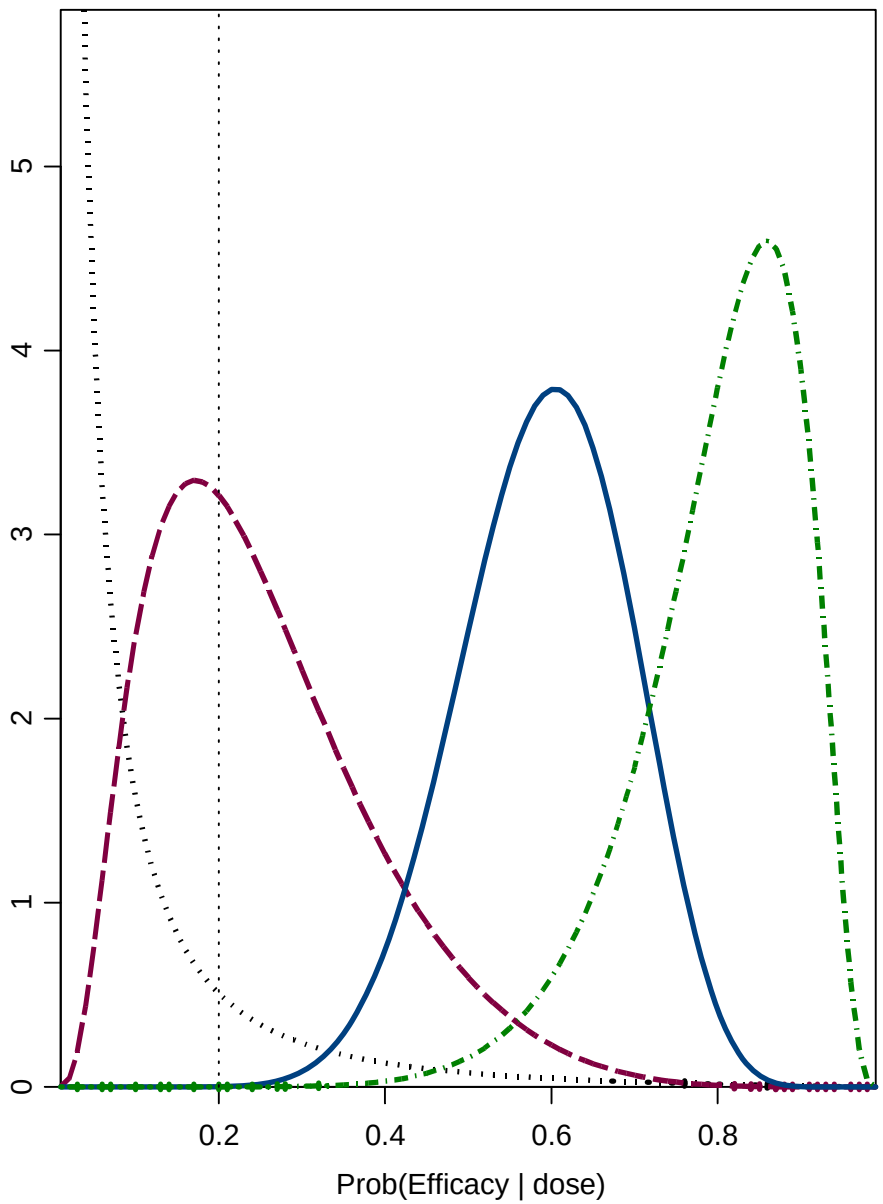
24 Patients


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose)

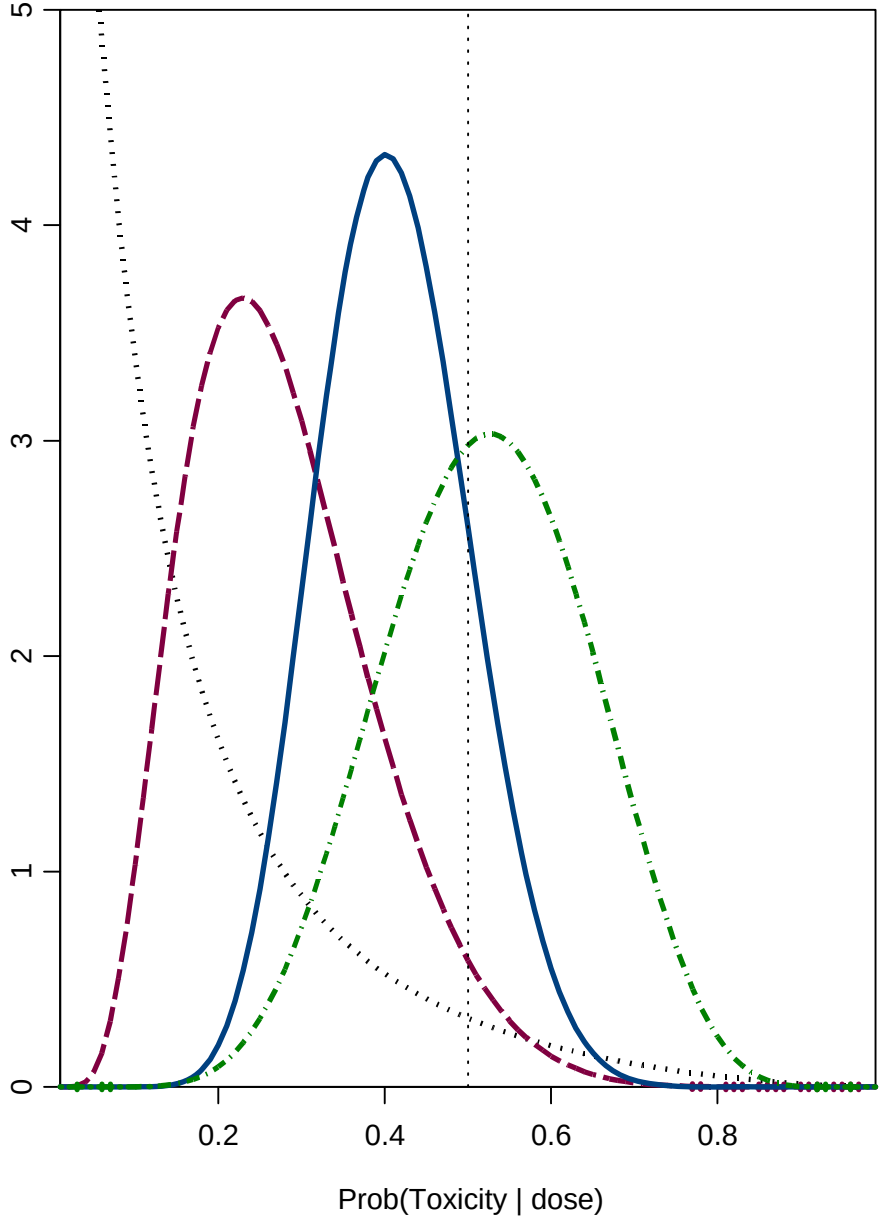

0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Efficacy | dose)

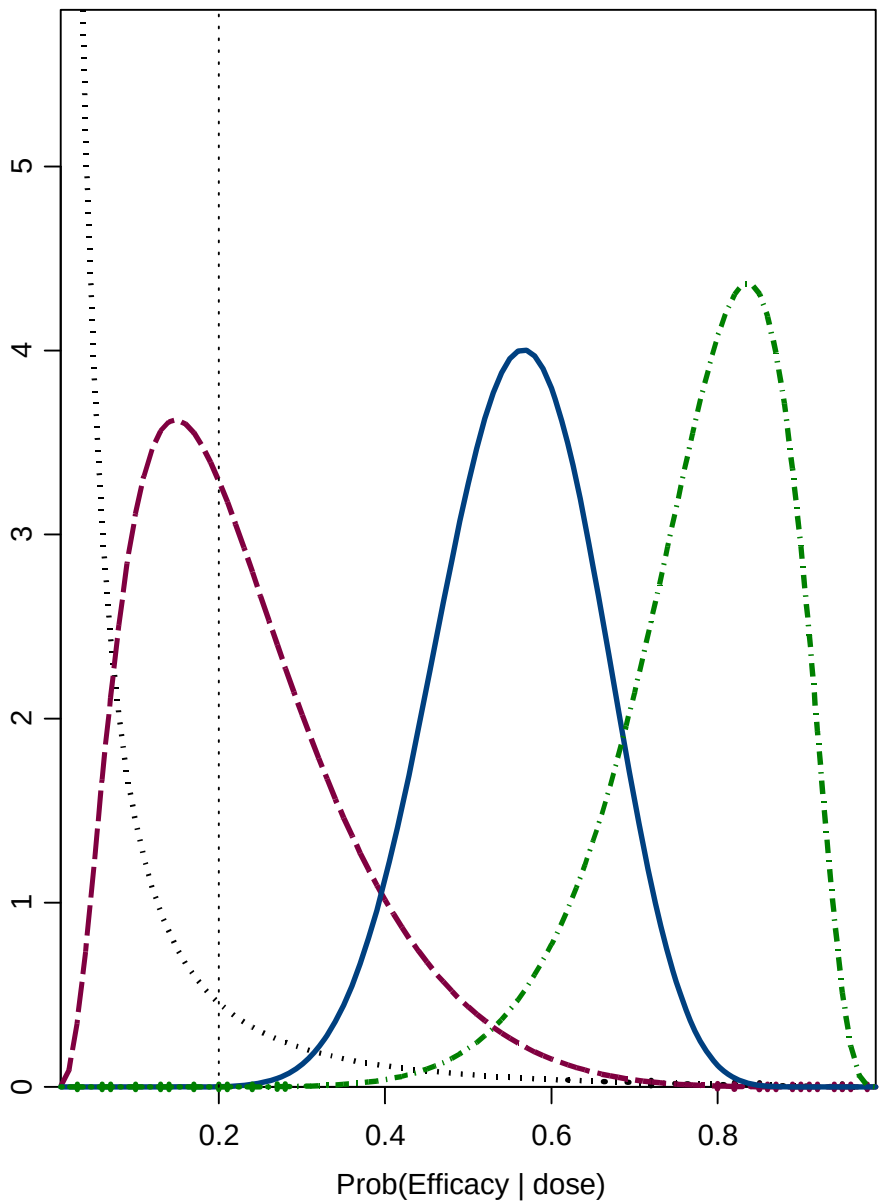
27 Patients


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose)

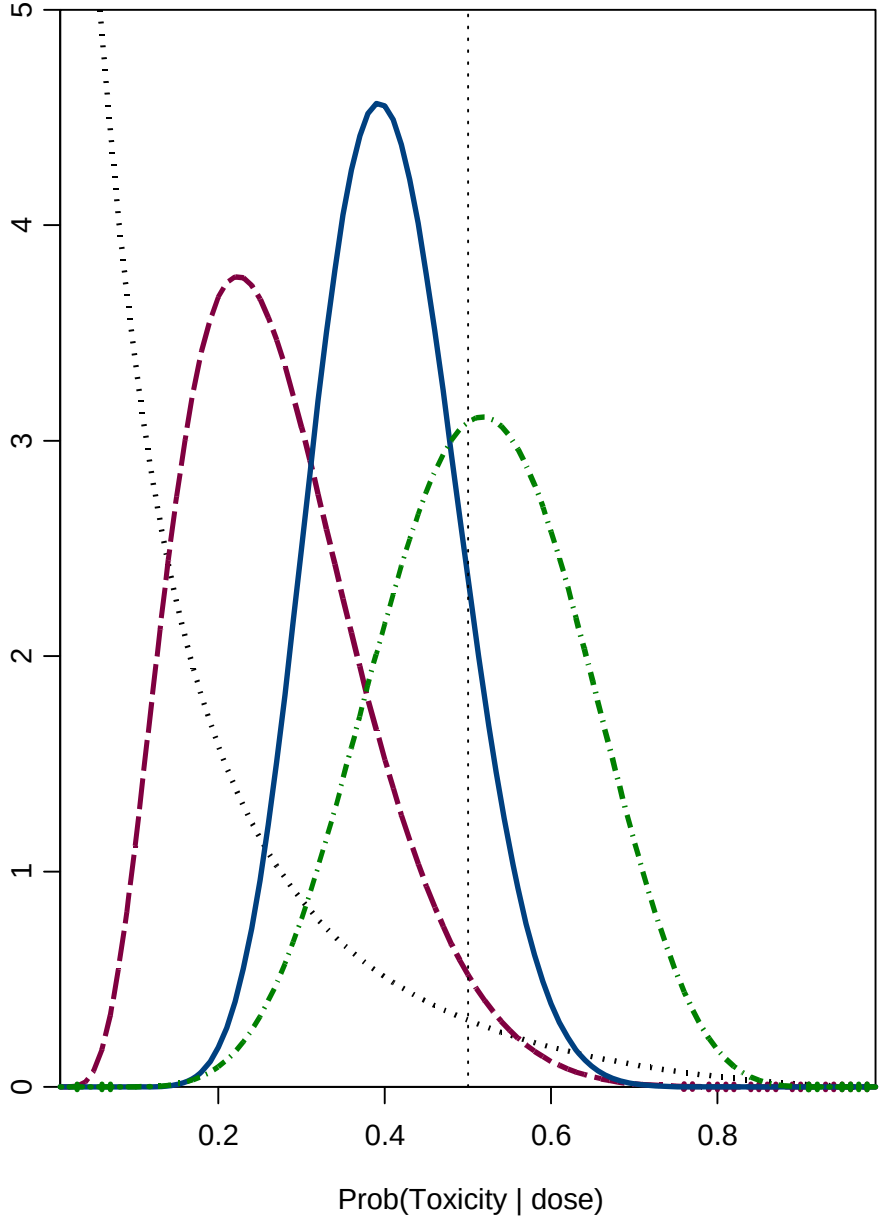

0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Efficacy | dose)

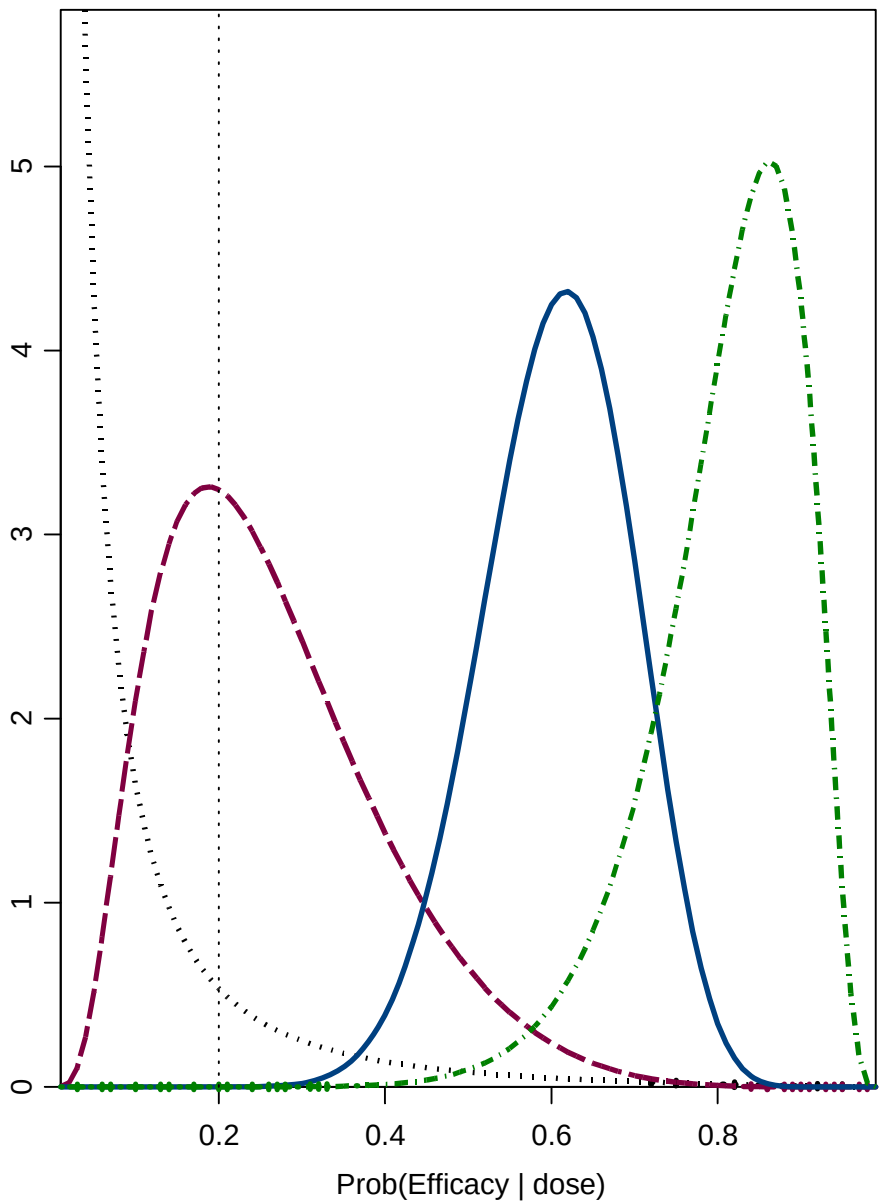
30 Patients


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose)

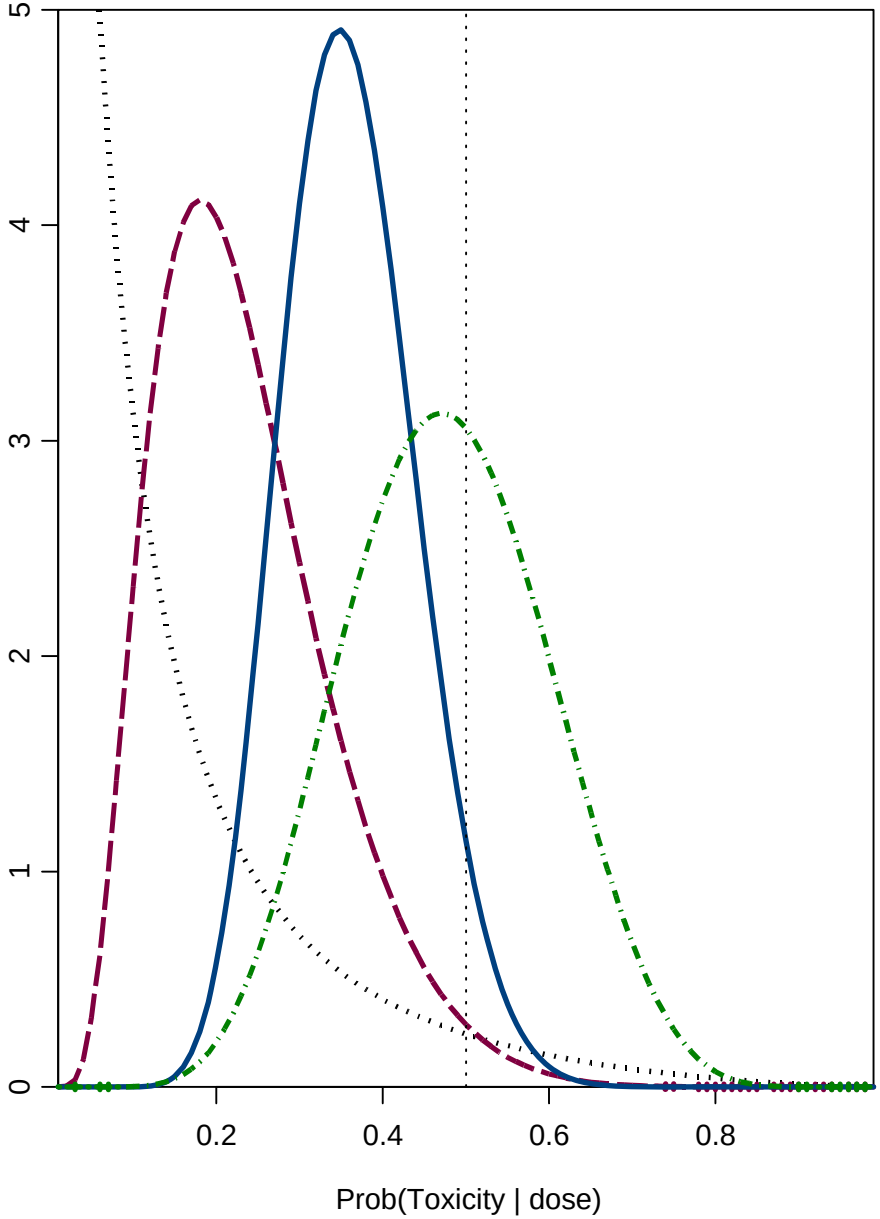

0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Efficacy | dose)

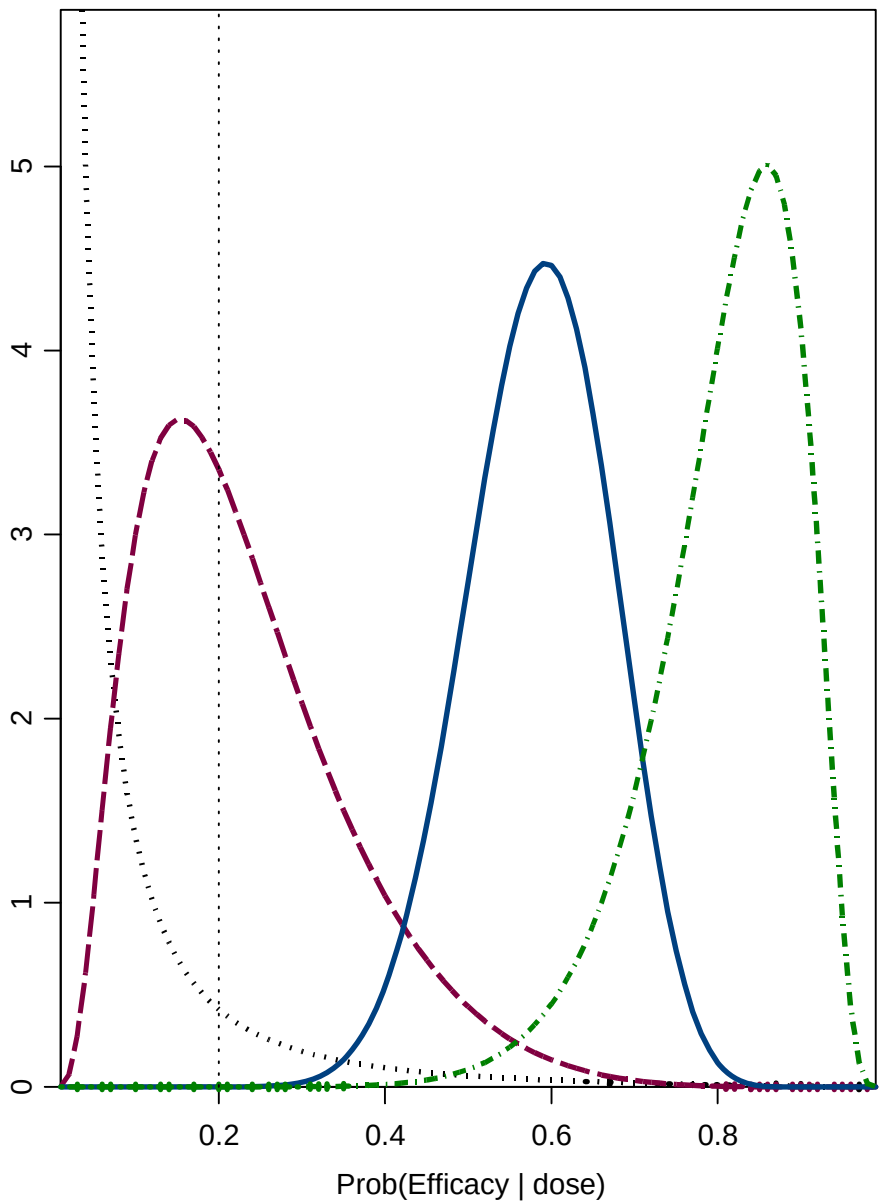
33 Patients


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose)


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Efficacy | dose)

36 Patients


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Toxicity | dose)


0
1
2
3
4
5
0.2
0.4
0.6
0.8
Prob(Efficacy | dose)

The Trade-Off-Based Algorithm reliably :

- **Finds Safe Doses having High Efficacy**
- **Stops if no dose is acceptable**
  (all doses are too toxic or inefficacious)

Computer code is freely available

Covariate-Adjusted Adaptive Randomization in a Sarcoma Trial with Multi-Stage Treatments

Motivation: **A Trial of**

1200 mg/m^2 Gemcitabine (G)

vs

900 mg/m^2 G + Docetaxel (G+D)

for

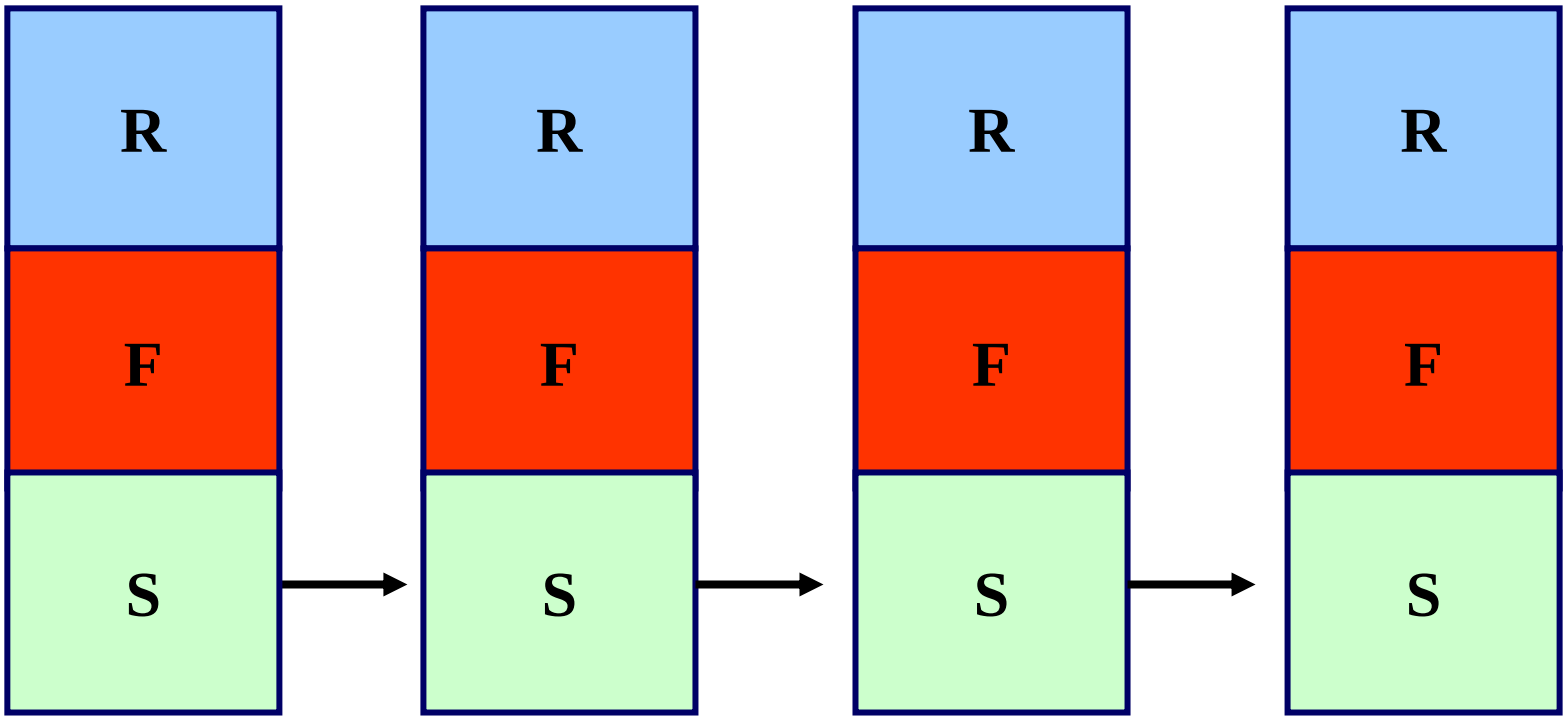
Unresectable Soft Tissue Sarcoma

- **Two Prognostic Covariates:**
  - **Leiomyosarcoma (LMS) vs any other sarcoma subtype**
  - **Prior Pelvic Radiation (PPR)**

PPR ➔ 25% lower doses of G and G+D ➔ Built-in treatment-covariate interaction

- **120 Patients**

- **Each patient evaluated at 6, 12, 18, 24 weeks**



After each of up to four 6-week stages of therapy :

R = Response (CR/PR)

F = Treatment Failure (Progression or Death)

S = Stable Disease

S at evaluations 1, 2, or 3 → Continue evaluation

Possible Overall Outcomes :

6 weeks	12 weeks	18 weeks	24 weeks	Overall
R				*R*
F				*F*
S	R			*R*
S	F			*F*
S	S	R		*R*
S	S	F		*F*
S	S	S	R	*R*
S	S	S	F	*F*
S	S	S	S	*S*

Probability Model

P*er-course* probabilities:

$\pi_{k,R}$**(T, Z, θ) = Prob{Response| T, Z, θ}**

$\pi_{k,F}$**(T, Z, θ) = Prob{Failure| T, Z, θ}**

in stage *k* = 1,2,3,4 with treatment T = G or G+D

for a patient with covariate **Z** = (Z_1, Z_2)

Linear Components

$$\eta_{k,j}(T, Z, \theta) = \mu_j + \alpha_j T + \gamma_{k,j} + \sum_{r=1,2}(\beta_{j,r} + \tau_{j,r} T)Z_r$$

= **Main Effect** (outcome j= R or F)

+ **Treatment Effect** ($T = \pm 1$)

+ **Stage Effect** (k=1, 2, 3, 4)

+ **Covariate Effects** (Z_1, $Z_2 = \pm 1$)

+ **Treatment x Covariate Interactions**

Generalized Logistic Model Probabilities

$$\pi_{k,y}(T, \mathbf{Z}, \boldsymbol{\theta}) = \frac{\exp\{\eta_{k,y}(T, \mathbf{Z}, \boldsymbol{\theta})\}}{1 + \exp\{\eta_{k,F}(T, \mathbf{Z}, \boldsymbol{\theta})\} + \exp\{\eta_{k,R}(T, \mathbf{Z}, \boldsymbol{\theta})\}}$$

For stage k = 1,2,3,4 and outcome y = R or F

Priors

Prior parameters were obtained by eliciting the means of π_R and π_F within each prognostic subgroup, and calibrating variances to ensure suitably "uninformative" priors

Adaptive Randomization Criterion

$\xi_{4,R}^{+}(T, Z, \theta)$ = **Pr(R within 4 stages|Z)**

$\xi_{4,F}^{+}(T, Z, \theta)$ = **Pr(F within 4 stages|Z)**

AR Criterion:

$$\zeta(T, Z, \theta) = \omega\, \xi_{4,R}^{+}(T, Z, \theta) + (1-\omega)\{1- \xi_{4,F}^{+}(T, Z, \theta)\}$$

using elicited weight $\omega = 1.0/(1.0+1.3) = 0.565$

Given the current data, randomize a patient

with covariates **Z** to **G+D** with probability

$$\nu(\mathbf{Z}, \textit{data}) = \Pr\{ \zeta(+1, \mathbf{Z}, \theta) > \zeta(-1, \mathbf{Z}, \theta) \mid \textit{data}\}$$

and to **G** with probability $1 - \nu(\mathbf{Z}, \textit{data})$

Equivalently, one may replace

$\zeta = \omega\, \xi_{4,R}^{+} + (1-\omega)\{1- \xi_{4,F}^{+}\}$ **with** $\xi_{4,R}^{+} - 1.3\xi_{4,F}^{+}$

Early Stopping Rule:

At any time during the trial, if

$$\nu(Z, data) > .99 \quad \text{or} \quad \nu(Z, data) < .01$$

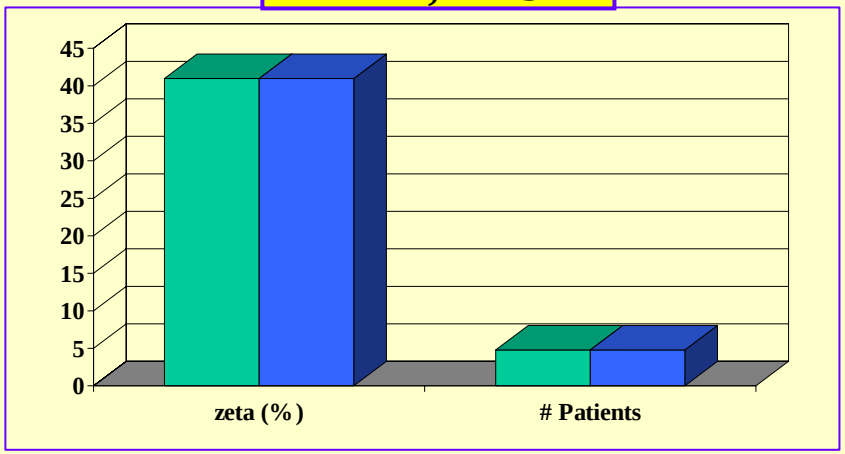
Stop the trial in subgroup **Z** and **Select** the superior treatment arm in that subgroup

Scenario 1 (Null Case)
Patients Treated

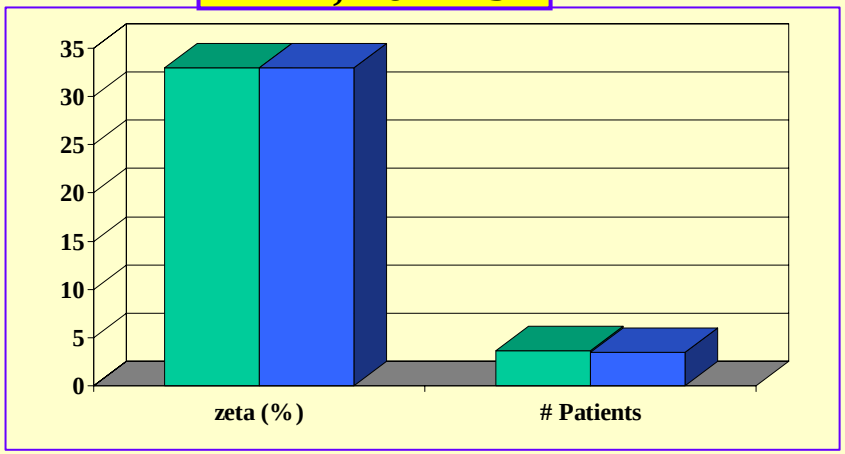
G

G+D

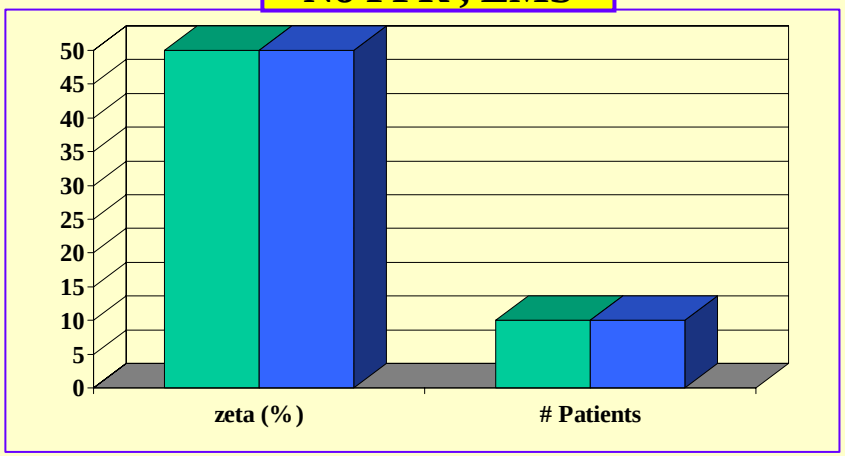
PPR , LMS



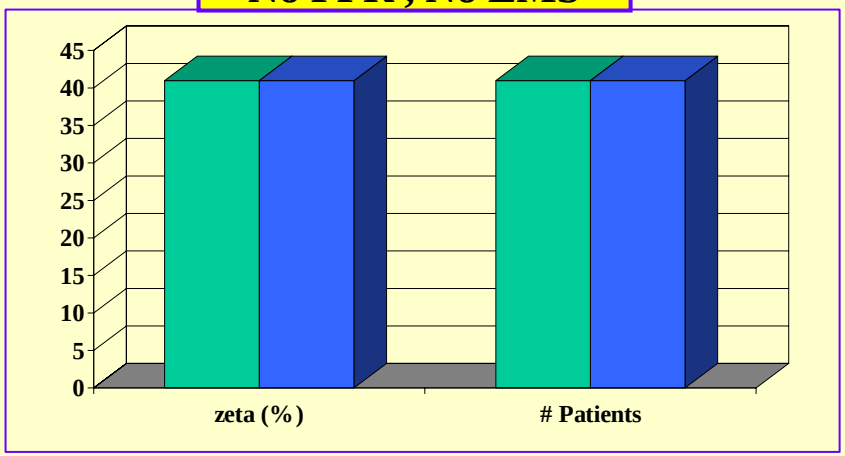
PPR , No LMS



No PPR , LMS



No PPR , No LMS

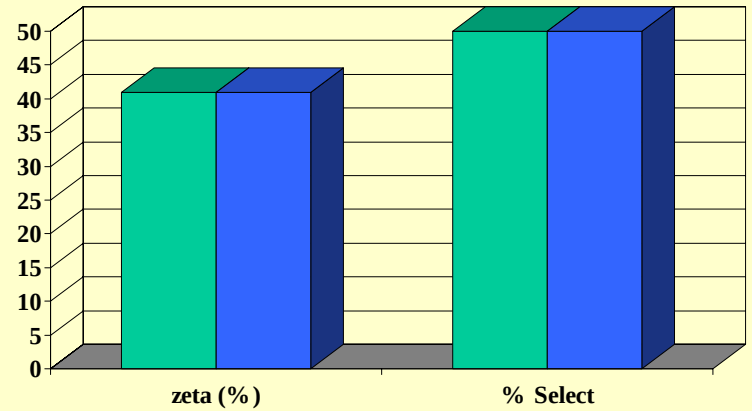


Scenario 1 (Null Case)
Treatment Selection Percentages

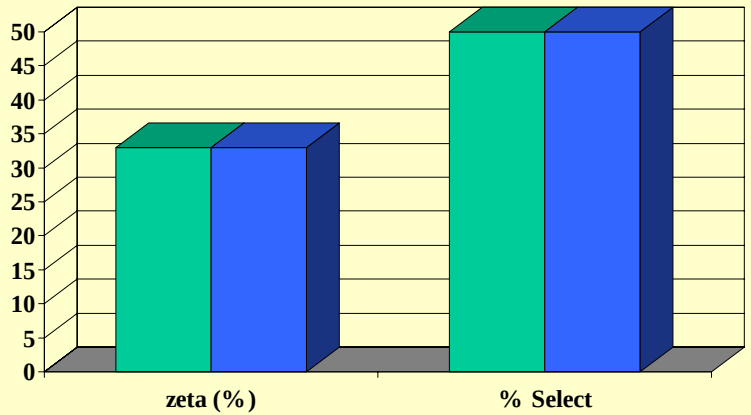
G

G+D

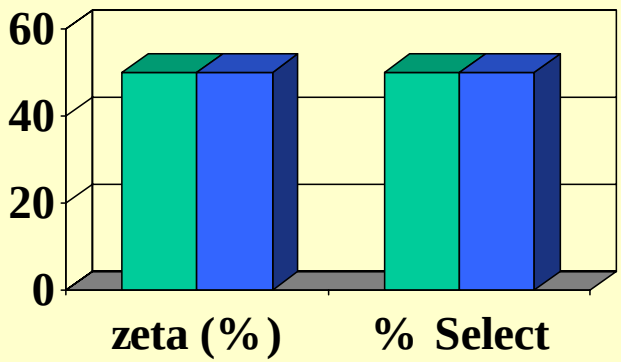
PPR , LMS



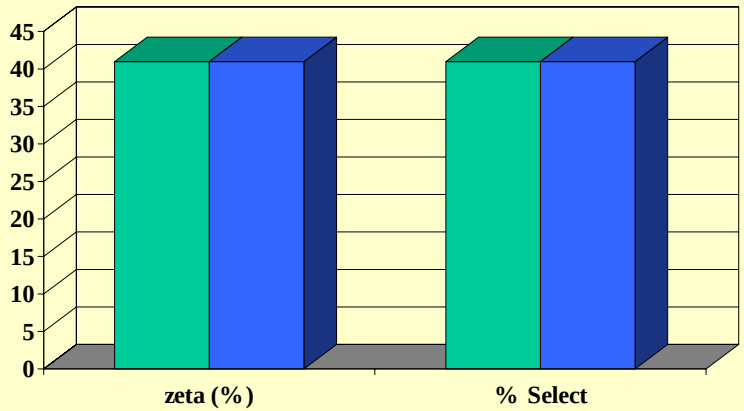
PPR , No LMS



No PPR , LMS



No PPR , No LMS

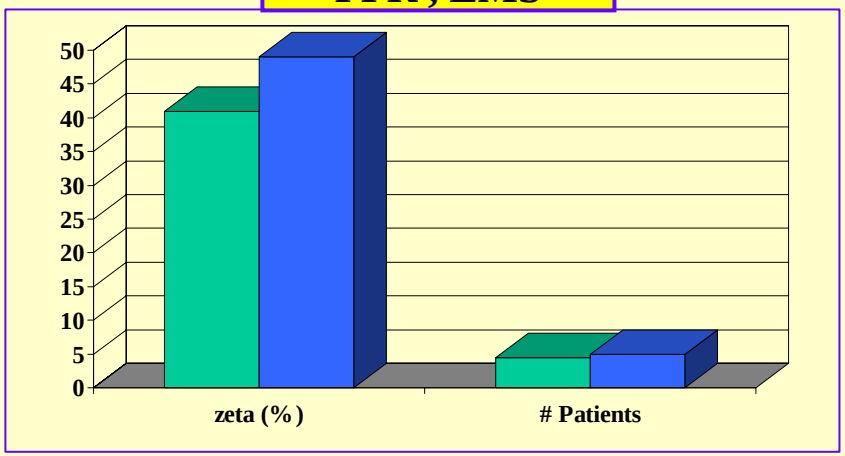


Scenario 2
Patients Treated

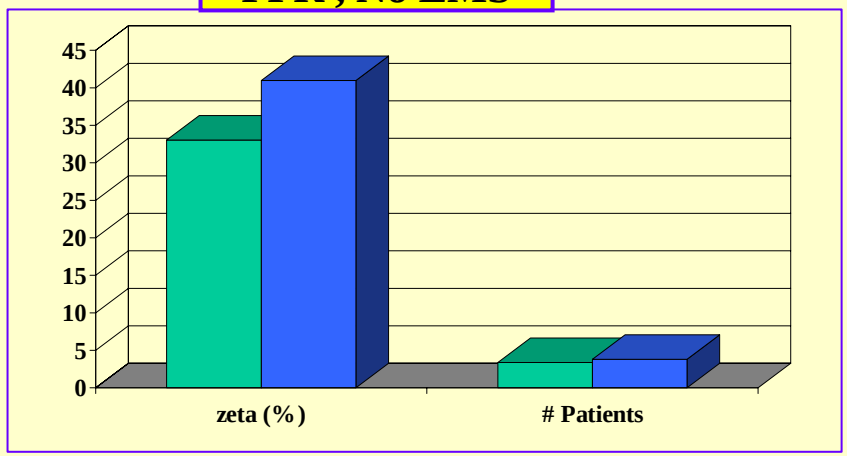
G

G+D

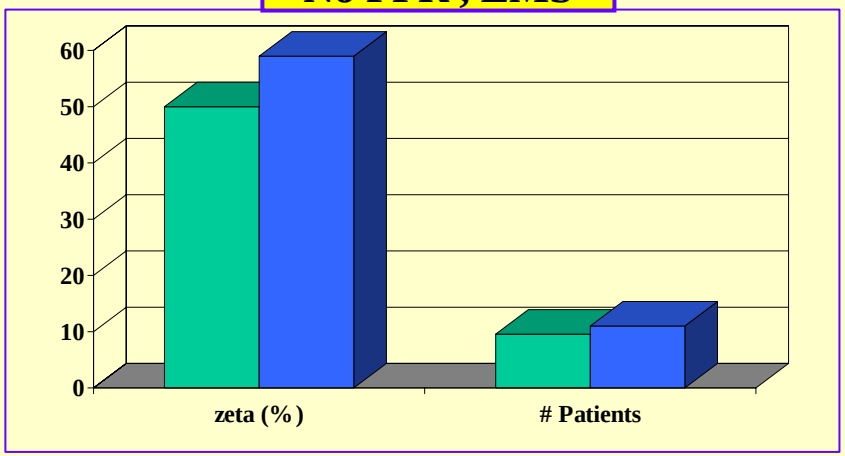
PPR , LMS



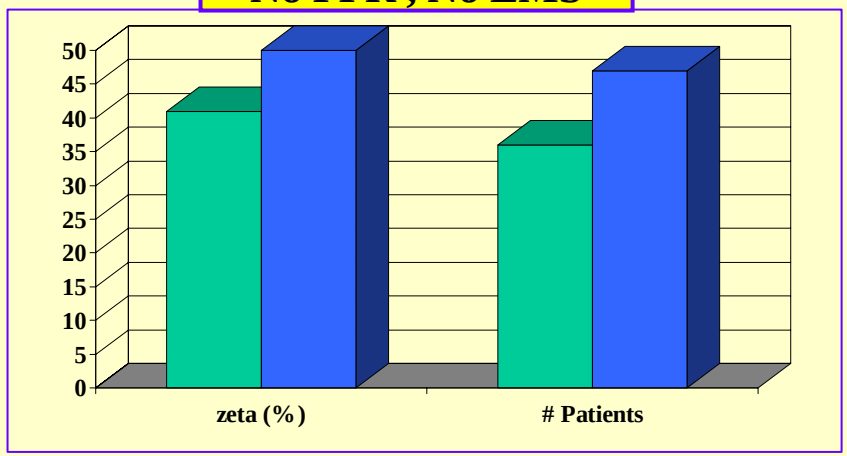
PPR , No LMS



No PPR , LMS



No PPR , No LMS

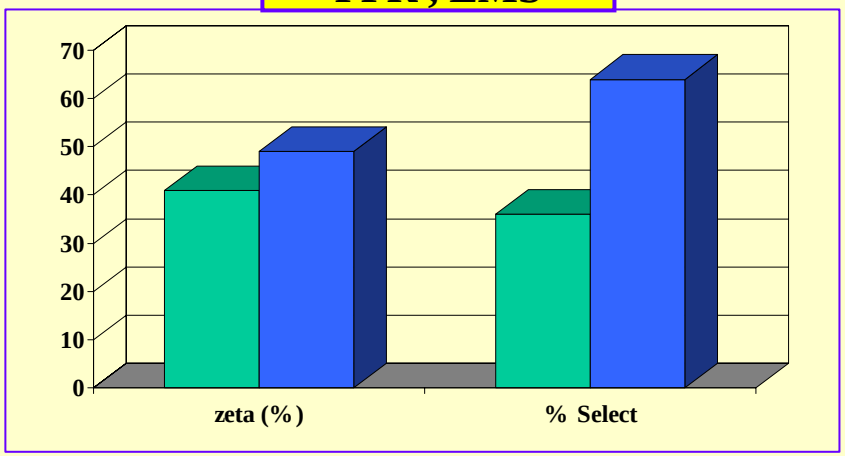


Scenario 2
Treatment Selection Percentages

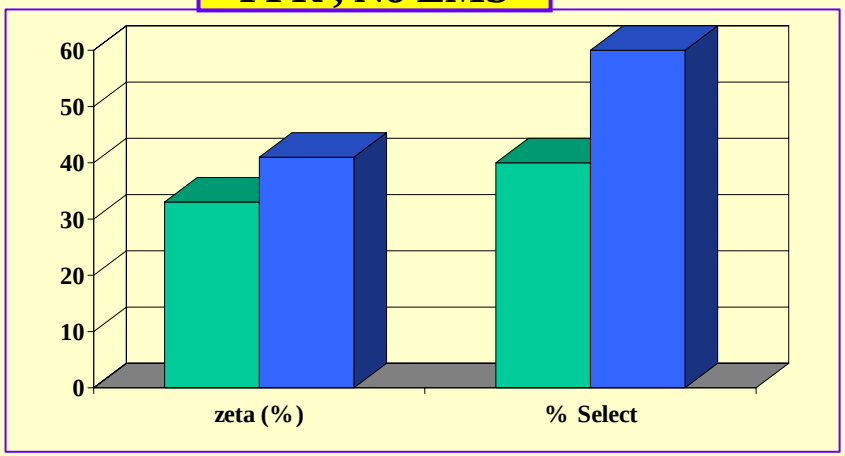
G

G+D

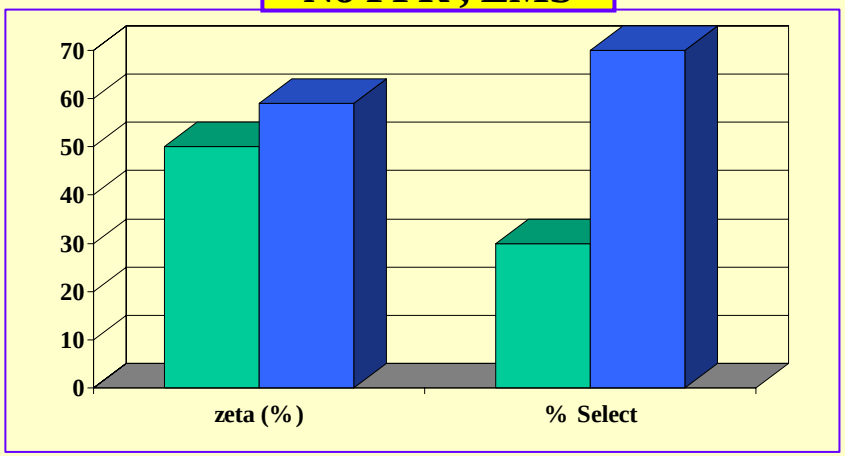
PPR , LMS



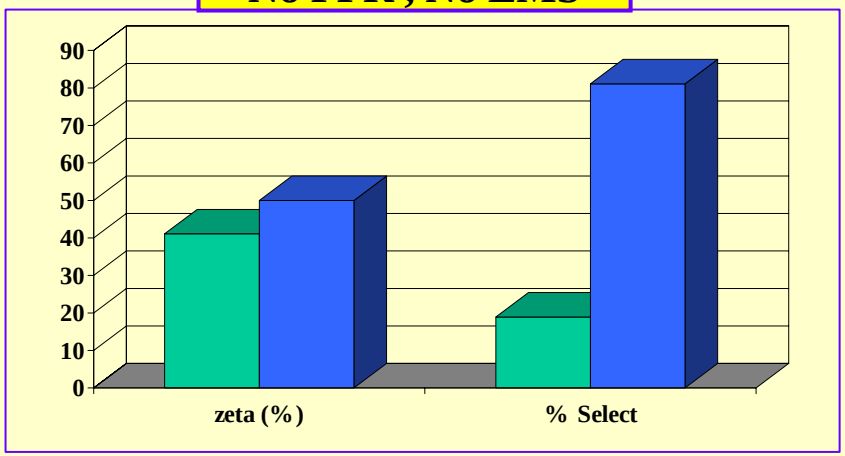
PPR , No LMS



No PPR , LMS



No PPR , No LMS

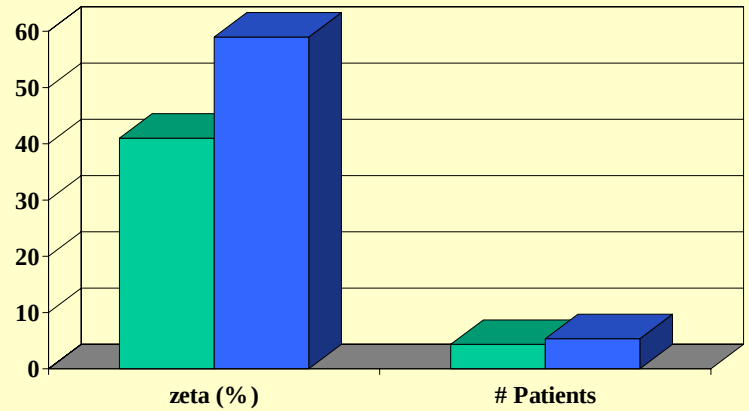


Scenario 3
Patients Treated

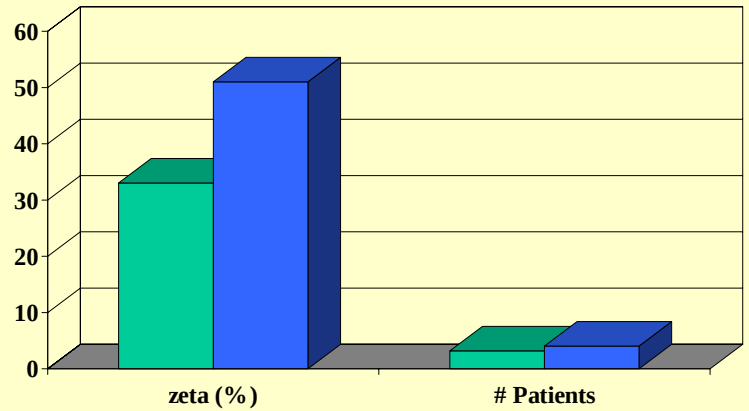
G

G+D

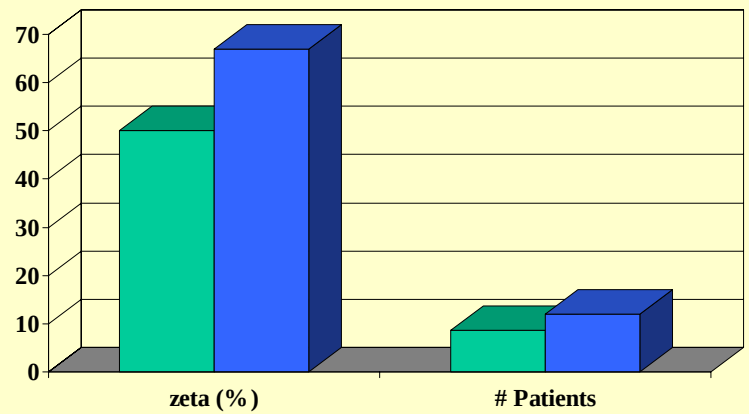
PPR , LMS



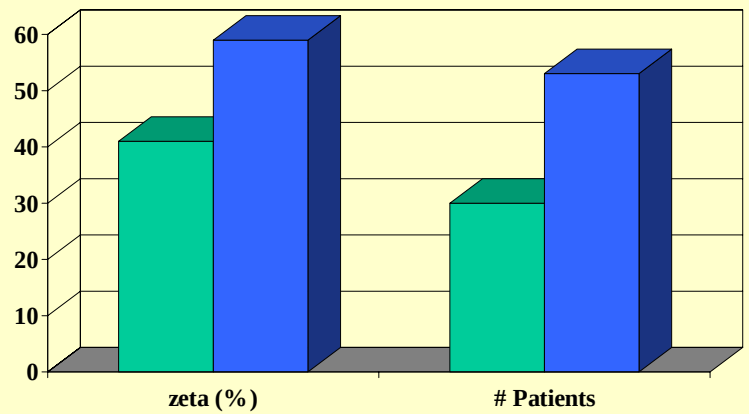
PPR , No LMS



No PPR , LMS



No PPR , No LMS

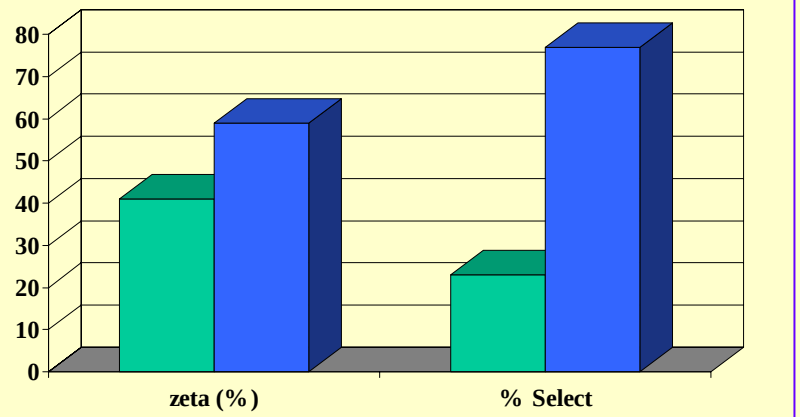


Scenario 3
Treatment Selection Percentages

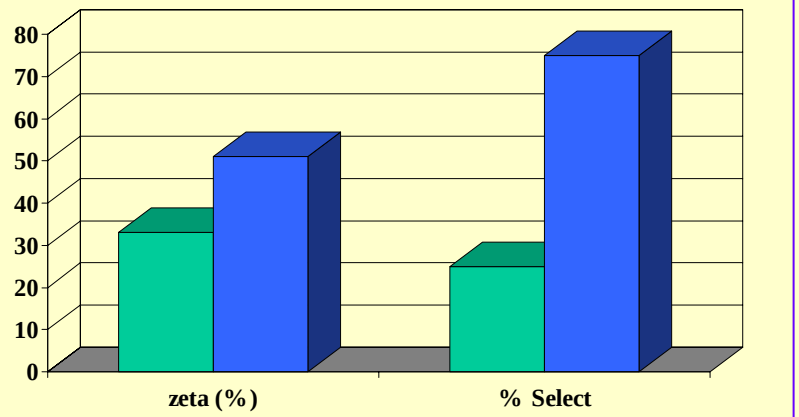
G

G+D

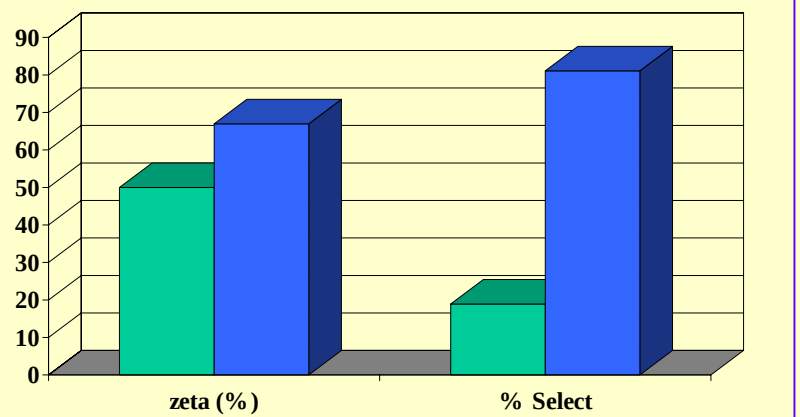
PPR , LMS



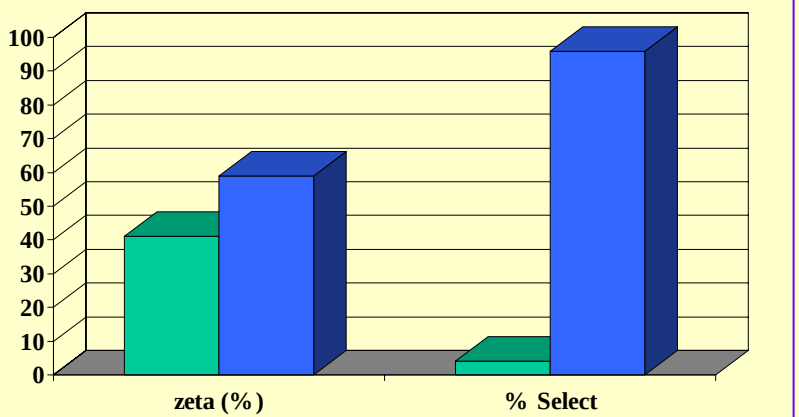
PPR , No LMS



No PPR , LMS



No PPR , No LMS

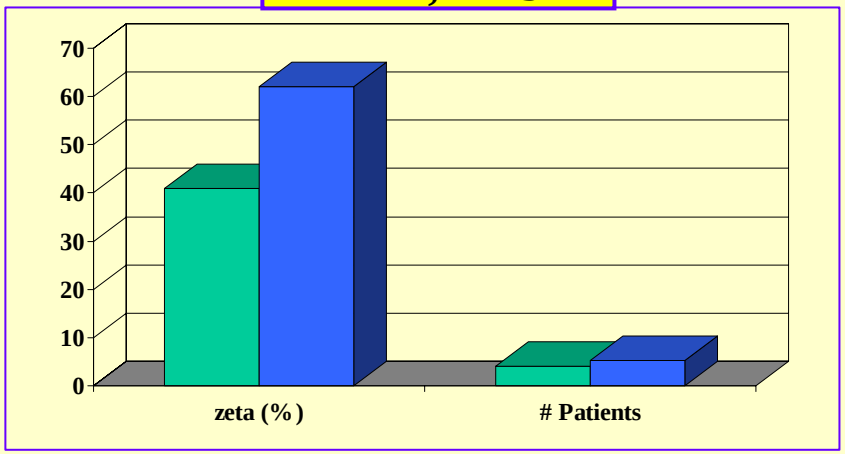


Scenario 4a (N=120, True Subgroup %)
Patients Treated

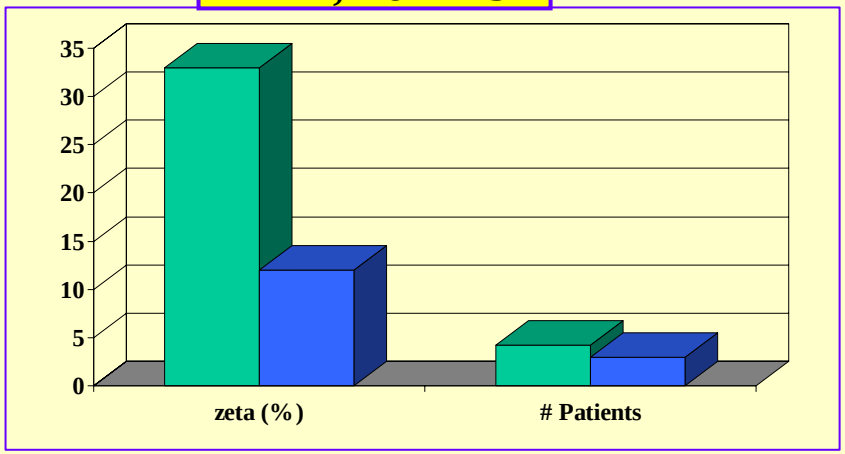
G

G+D

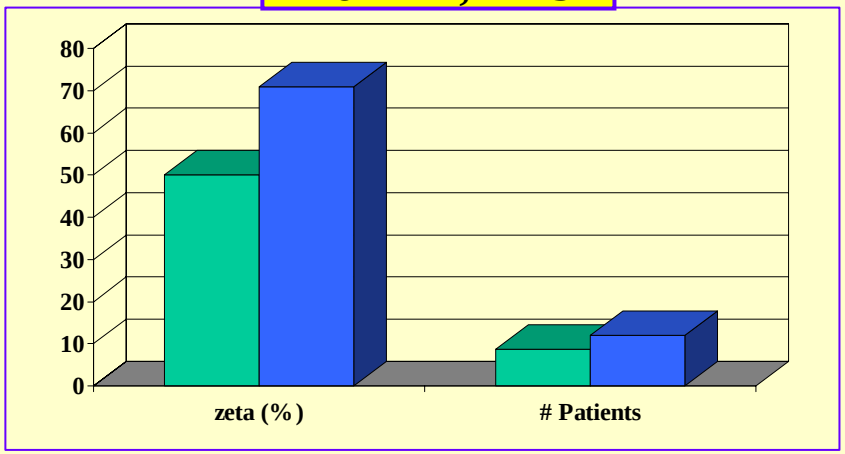
PPR , LMS



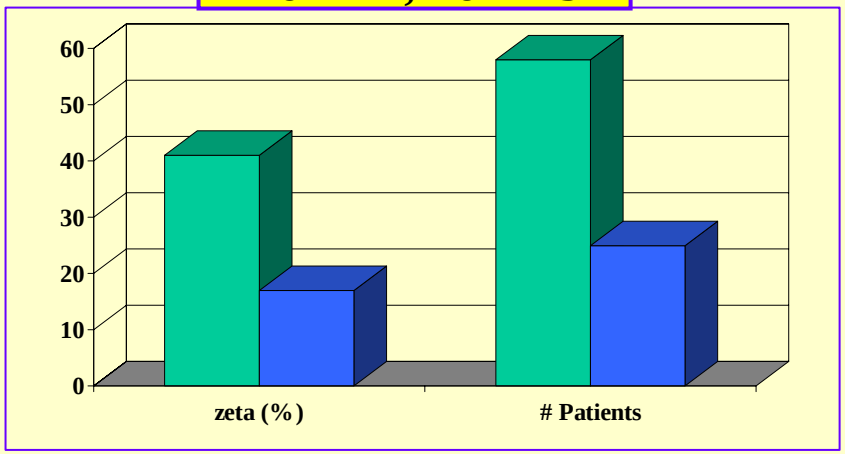
PPR , No LMS



No PPR , LMS



No PPR , No LMS

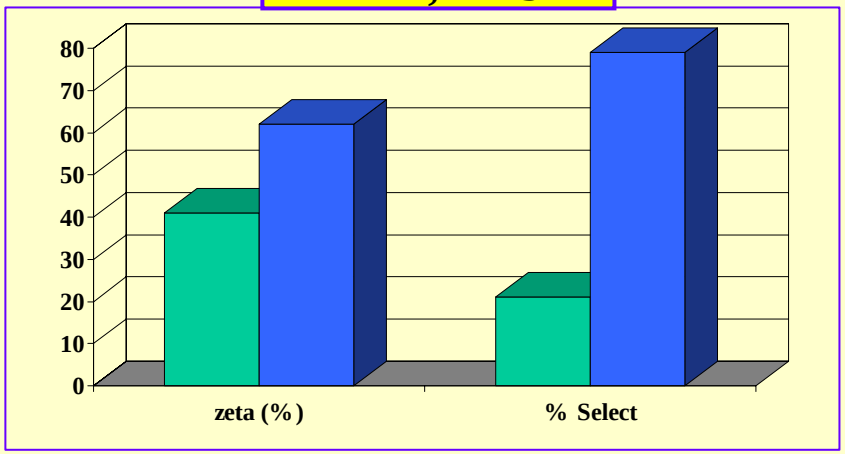


Scenario 4a (N=120, True Subgroup %)
Treatment Selection Percentages

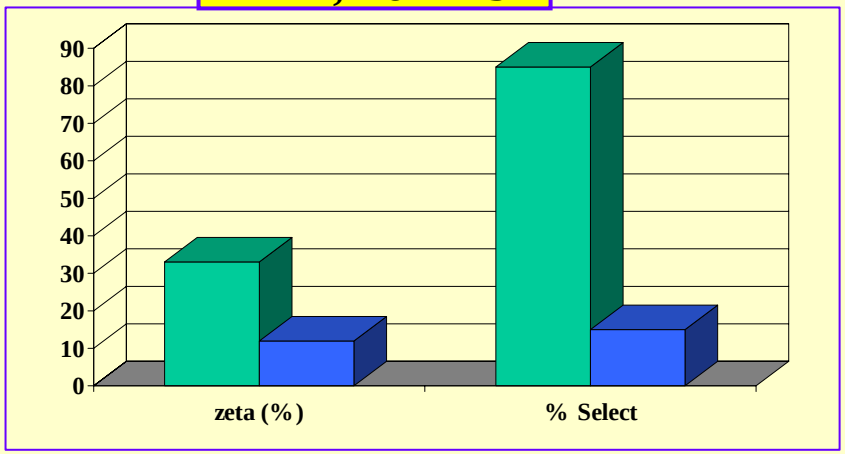
G

G+D

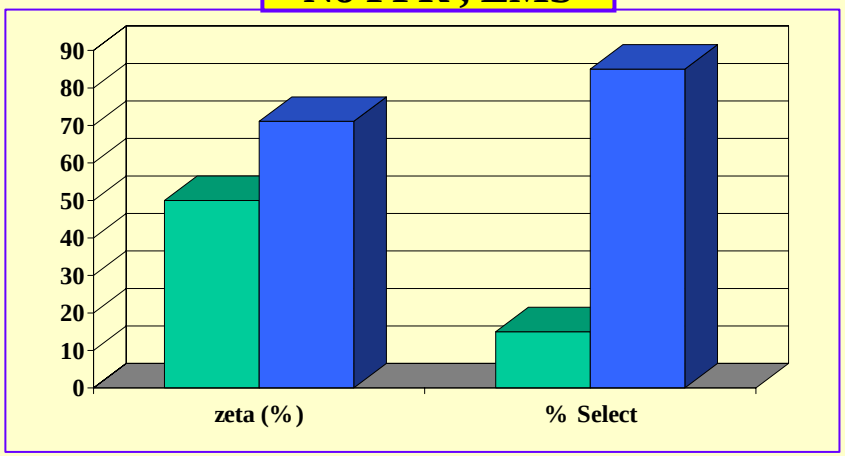
PPR , LMS



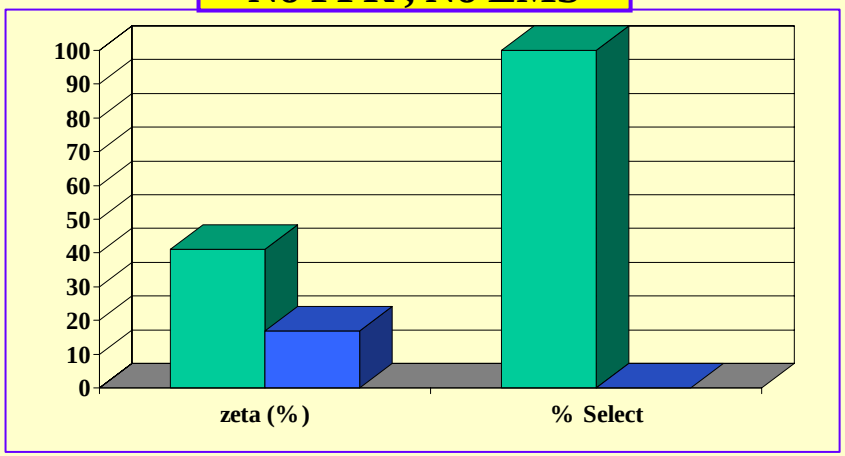
PPR , No LMS



No PPR , LMS



No PPR , No LMS

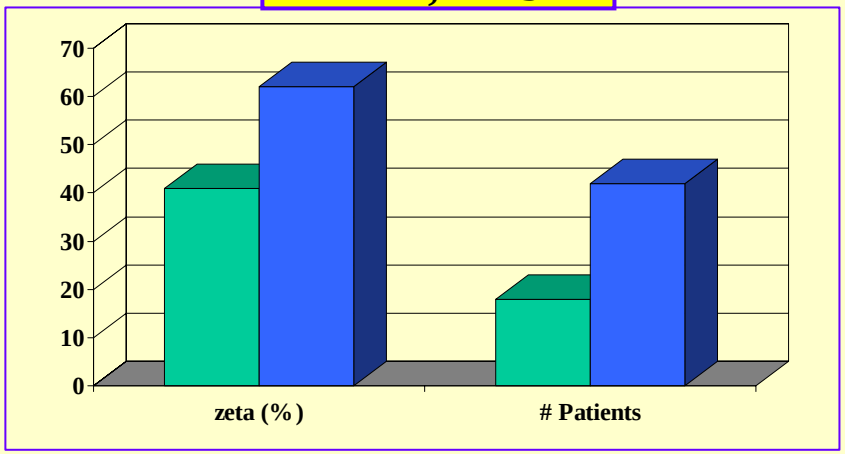


Scenario 4c (N=240, 25% Per Subgroup)
Patients Treated

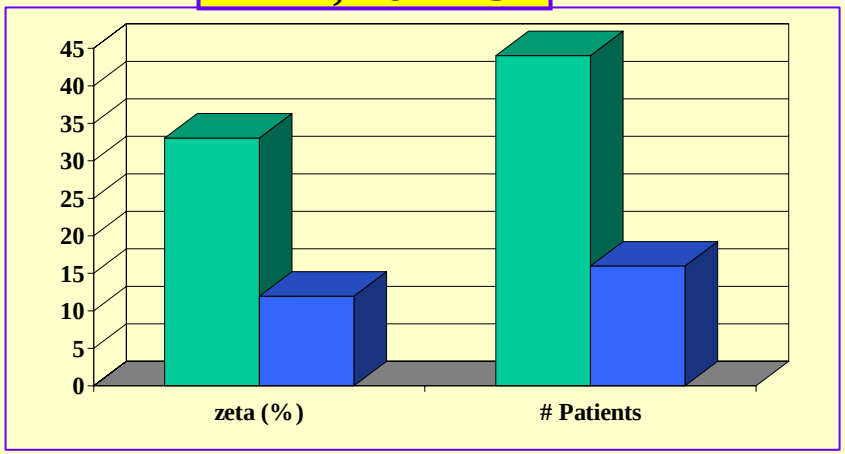
G

G+D

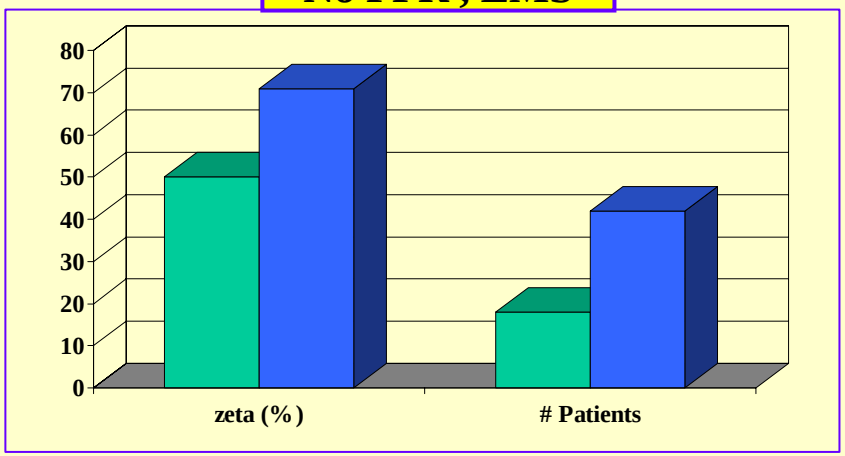
PPR , LMS



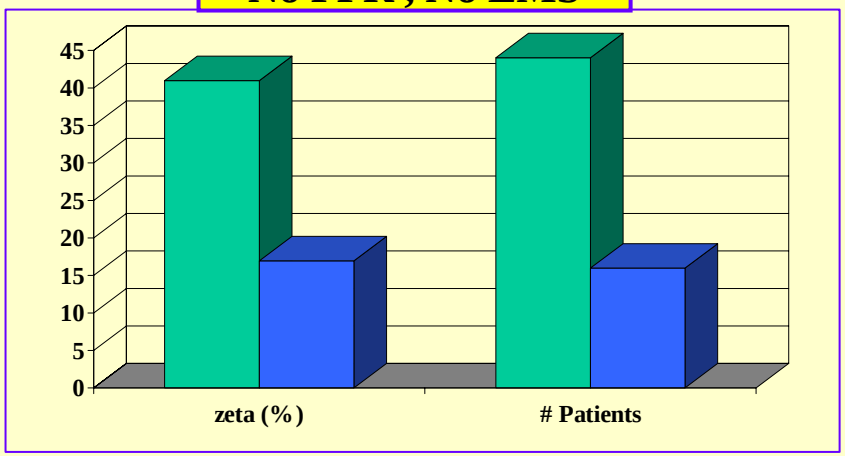
PPR , No LMS



No PPR , LMS



No PPR , No LMS

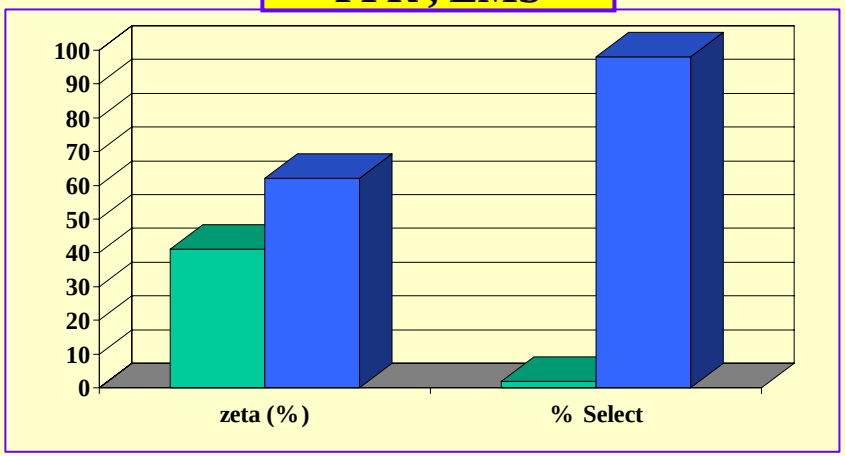


Scenario 4c (N=240, 25% Per Subgroup)
Treatment Selection Percentages

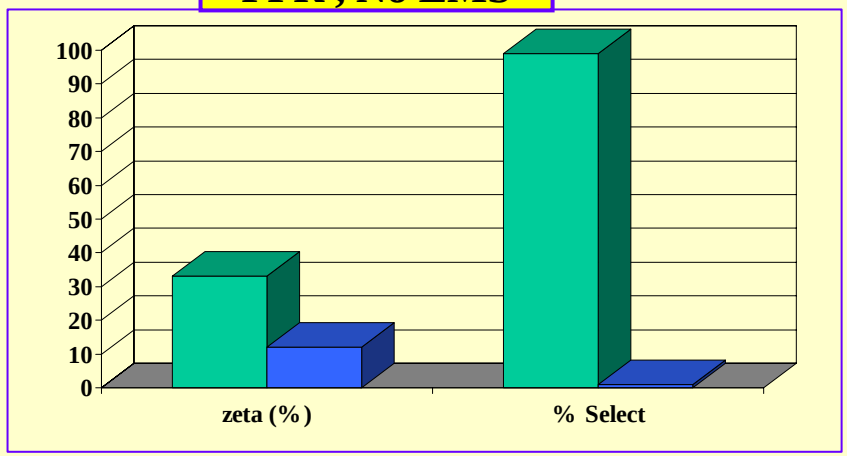
G

G+D

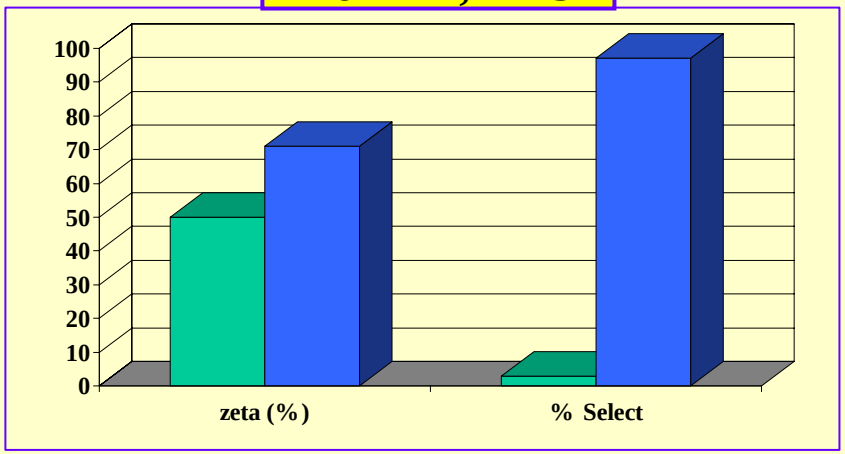
PPR , LMS



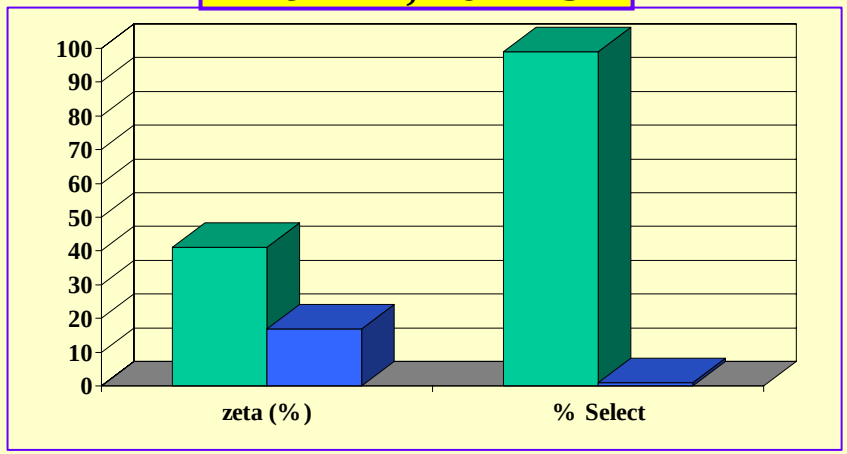
PPR , No LMS



No PPR , LMS



No PPR , No LMS



Sensitivity Analyses

- **Ignore Z** → **Treatment-covariate interactions are missed completely, & the Selection and AR imbalances may be *backwards* within subgroups**
- **Do AR separately within subgroups** → **Substantial loss in AR imbalance if treatment-covariate interactions present, because there is no borrowing strength**
- **Weighting: The AR method is *very insensitive to changes in* ω_R = .10 to .91**

Conclusions

% patients assigned to the better treatment &

% correct selection probability both with ↑ treatment effect & sample size

The method reliably detects treatment-covariate interactions

Complex & *computationally intensive* →

User Interface required for trial conduct

Bayesian Sensitivity Analyses of Confounded Treatment Effects in Survival Analysis

Why Randomize?

- A single-arm phase II clinical trial of combination chemotherapy A (n=44) for acute myelogenous leukemia (AML) was conducted at M.D. Anderson Cancer Center (MDACC) in 1995
- Combination chemotherapy B was studied subsequently at MDACC, as one arm of a four-arm randomized AML trial in 1996-98

Estey and Thall, Blood, 2003

Why Randomize?

- ***A Bayesian Weibull regression model*** was fit to data consisting of Survival times (T), Treatment indicators (I_B) and Covariates (**Z**):

  Weibull hazard(t) = $\phi\, t^{\phi-1} \exp(\mu+\gamma I_B+\beta \mathbf{Z})$

- **Z =** Performance status, Age,

  Treatment in a laminar airflow room (yes/no),

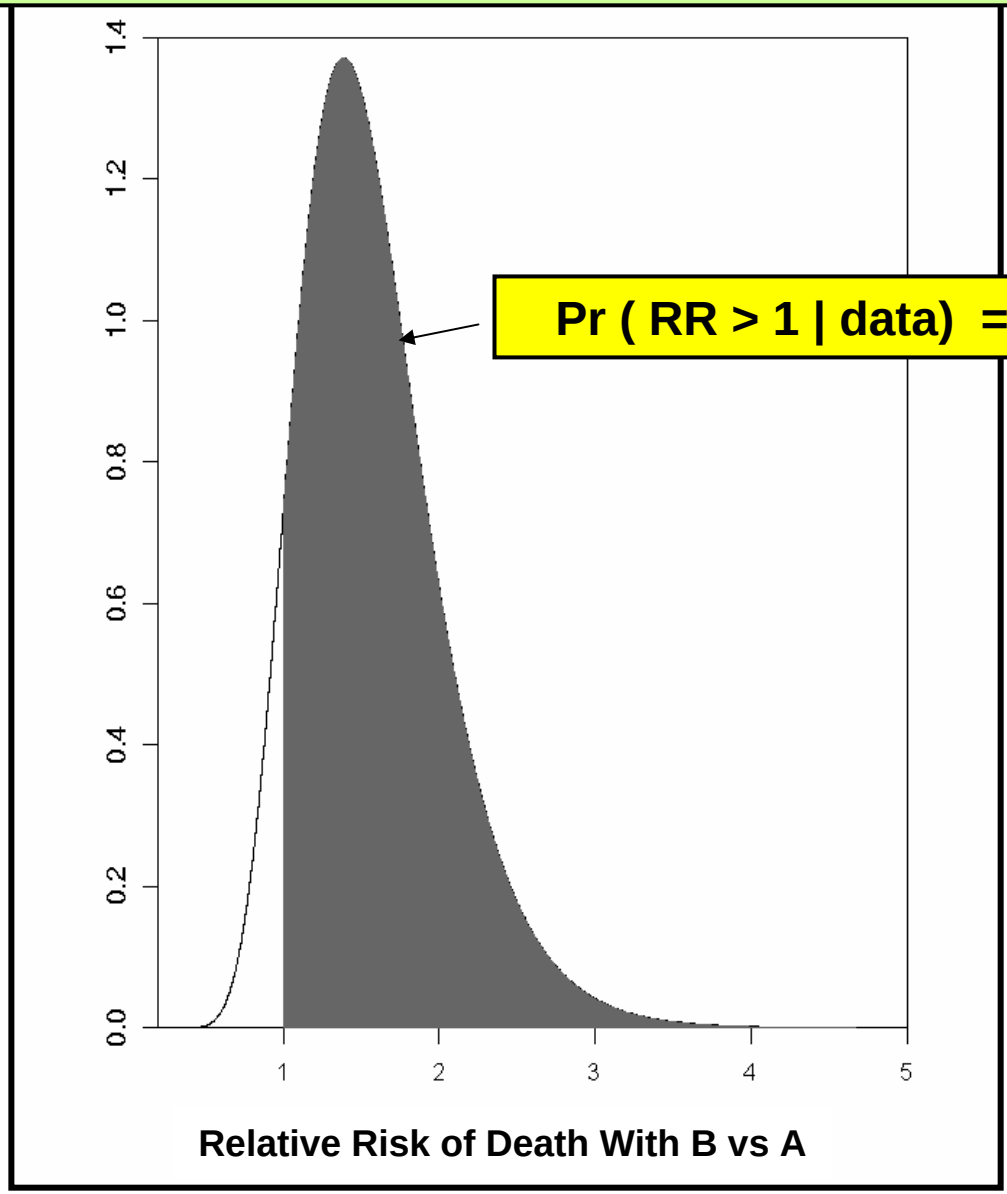
  Cytogenetic abnormality (3 categories)

- ***Very disperse priors were assumed:***

  $\mu, \gamma, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5 \sim$ iid N(0,1000)

  $\phi \sim$ Gamma(mean=1, var=1000)

Posterior of RR of Death With B versus A



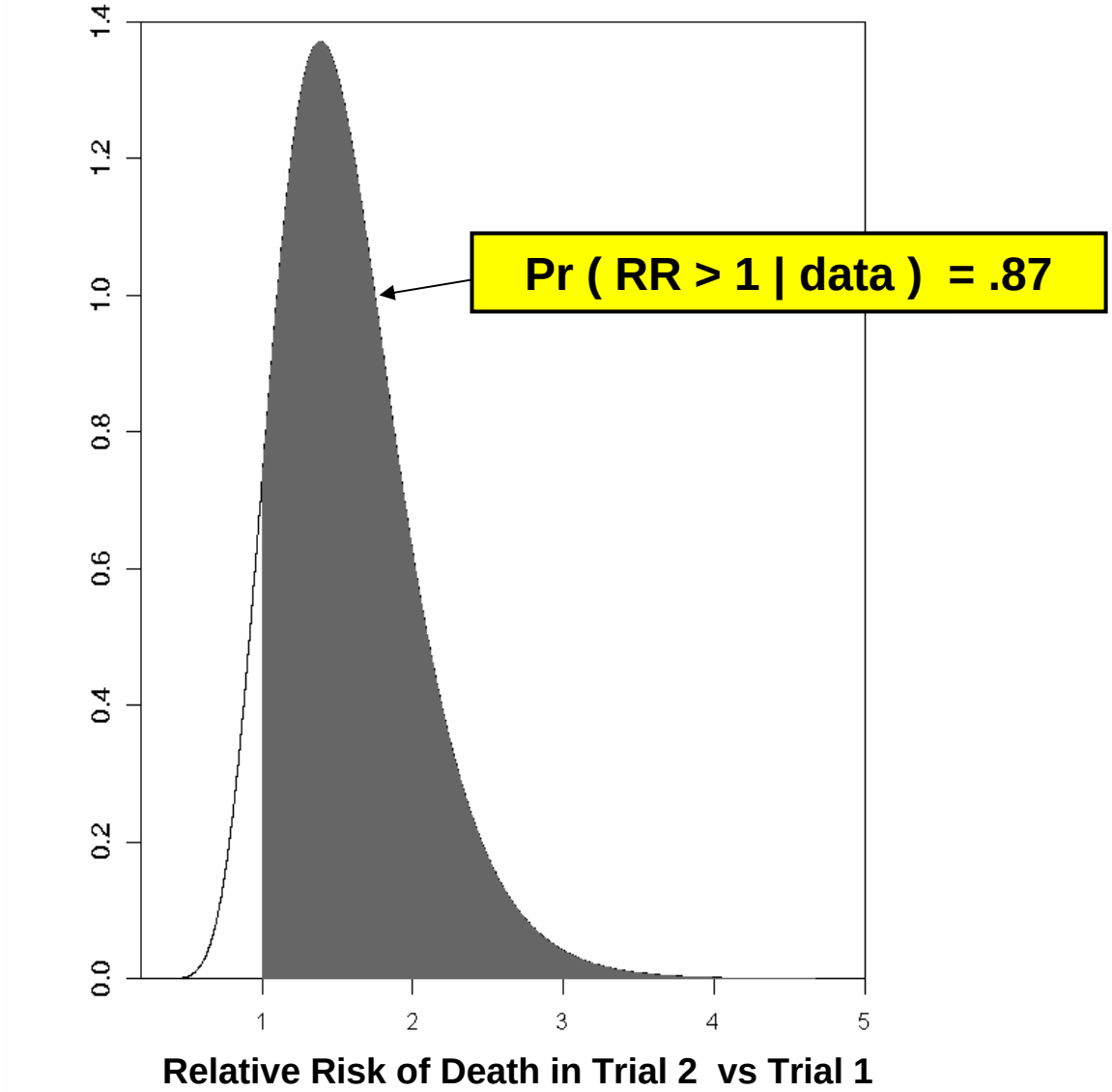
In fact, A and **B were the same treatment** !!

A = B = Fludarabine + idarubicin + ara-C + G-CSF + ATRA (FAIGA)

→

The observed RR was actually Between-Trial Effect !!

Posterior RR of Death in *Trial 2-vs-Trial 1*



The General Problem

Goal: Compare treatments, A and B, based on real-valued parameters, θ_A and θ_B

- Typically θ_A and θ_B are probabilities or hazards, possibly transformed and/or covariate-adjusted
- Comparative inferences are based on $\delta_\theta = \theta_A - \theta_B$

Problem: If the data arise from two separate studies of A and B, one can estimate

$\gamma_{A,1}$ = Effect of treatment A in study 1

$\gamma_{B,2}$ = Effect of treatment B in study 2 ➔

A usual estimator estimates *the confounded effect*

$\delta = \gamma_{A,1} - \gamma_{B,2}$ **, not** $\delta_\theta = \theta_A - \theta_B$

Applied Bayesian Subtraction

Overall = Treatment + Latent

→

Treatment = Overall – Latent

1. Estimate the **Overall** effect from the data
2. For several hypothetical **Latent Effects**, compute the **Treatment Effect** and

Pr{Treatment Effect > 0 | data, Latent Effect}

Applied Bayesian Subtraction

Assume that

$$\gamma_{A,1} = \theta_A + \lambda_1 \quad \text{and} \quad \gamma_{B,2} = \theta_B + \lambda_2$$

where λ_1 and λ_2 are study effects ➔

$$\delta = (\theta_A - \theta_B) + (\lambda_1 - \lambda_2) = \delta_\theta + \delta_\lambda \rightarrow$$

$$\delta_\theta = \delta - \delta_\lambda$$

Bayesian Computations

1) Given data D_A and D_B from the trials of A and B, compute the usual posterior, f($\delta \mid D_A, D_B$)

2) Hypothesize a trial effect distribution, $f^{(h)}(\delta_\lambda)$

3) Compute the hypothetical posterior of δ_θ :

$$f^{(h)}(\delta_\theta \mid D_A D_B) = f^{(h)}(\delta - \delta_\lambda \mid D_A D_B)$$

4) Use $f^{(h)}(\delta_\theta \mid D_A D_B)$ to make hypothesis-based inferences about δ_θ

$pr^{(h)}(\delta_\theta > 0 \mid D_A D_B)$, $E^{(h)}(\delta_\theta \mid D_A D_B)$, etc.

Bayesians are Sensitive!!

Usual Bayesian Sensitivity Analysis

- **Prior$_1(\theta)$ + Lik(*data*$|\ \theta$) → Posterior$_1(\theta|$*data*)**
- **Prior$_2(\theta)$ + Lik(*data*$|\ \theta$) → Posterior$_2(\theta|$*data*)**
- **Prior$_3(\theta)$ + Lik(*data*$|\ \theta$) → Posterior$_3(\theta|$*data*)**

Sensitivity to Hypothetical Trial Effects

- **Trial effect dist'n $f_1^{(h)}(\delta_\lambda)$, f($\delta|$*data*) → $f_1^{(h)}(\delta_\theta \mid$ *data*$)$**
- **Trial effect dist'n $f_2^{(h)}(\delta_\lambda)$, f($\delta|$*data*) → $f_2^{(h)}(\delta_\theta \mid$ *data*$)$**
- **Trial effect dist'n $f_3^{(h)}(\delta_\lambda)$, f($\delta|$*data*) → $f_3^{(h)}(\delta_\theta \mid$ *data*$)$**

Constructing Hypothetical Distributions

1) **Fix $\mathbf{var^{(h)}(\delta_\lambda)}$, vary $\mathbf{E^{(h)}(\delta_\lambda)}$ over a reasonable domain, and compute $\mathbf{f^{(h)}(\delta_\theta \mid D_A\, D_B)}$ as a function of $\mathbf{E^{(h)}(\delta_\lambda)}$**

or

2) **Use historical data D_H to obtain a finite set of reasonable $\mathbf{f^{(h)}(\delta_\lambda \mid D_H)}$, and compute $\mathbf{f^{(h)}(\delta_\theta \mid D_A, D_B, D_H)}$ for each**

Comparing gemtuzumab ozogamicin (GO, "Mylotarg") to idarubicin + ara-C (IA)

- A trial of IA in 31 AML/MDS patients was conducted at MDACC in 1991-92
- A trial of GO ± IL-11 in 31 AML/MDS patients was conducted at MDACC in 2000-2001
- Since IL-11 had no effect on survival, we will collapse the 2 arms of the GO trial and focus on the GO-vs-IA comparison

First GO-vs-IA Sensitivity Analysis

Covariates and Trial Effects

- **Zubrod performance status (PS)**

"Good" = [PS=0,1,2] vs "Poor" = [PS=3,4]

- **If patient treated in a laminar airflow room (LAR)**
- **Cytogenetic karyotype: normal (baseline), -5/-7 abnormality, or other abnormality**
- $\beta Z = \beta_0+\beta_1 Z_1 + \ldots + \beta_4 Z_4$
- $\delta\tau = \delta_2\tau_2+\ldots+\delta_6\tau_6 =$ **confounded treatment-trial effects vs. trial 1,** τ_j **= Indicator of trial j=2,…,6**

S(t|Z) = pr(T>t | Z,θ**) for t>0,**

T = survival time,

θ **= model parameter vector**

First GO-vs-IA Sensitivity Analysis

We considered three possible survival models:

Weibull: $\log[-\log\{S(t|\mathbf{Z})\}] = \beta\mathbf{Z} + \delta\tau + \phi \log(t)$

Log logistic: $-\log[S(t|\mathbf{Z})/\{1-S(t|\mathbf{Z})\}] = \beta\mathbf{Z} + \delta\tau + \phi \log(t)$

Lognormal : mean = $\beta\mathbf{Z} + \delta\tau$, with constant variance.

Maximized log likelihoods = $-137.0,\ -139.4,\ -141.7$ ➔

The Weibull gives a slightly better fit.

Fitted Weibull Model for the GO and IA Trials

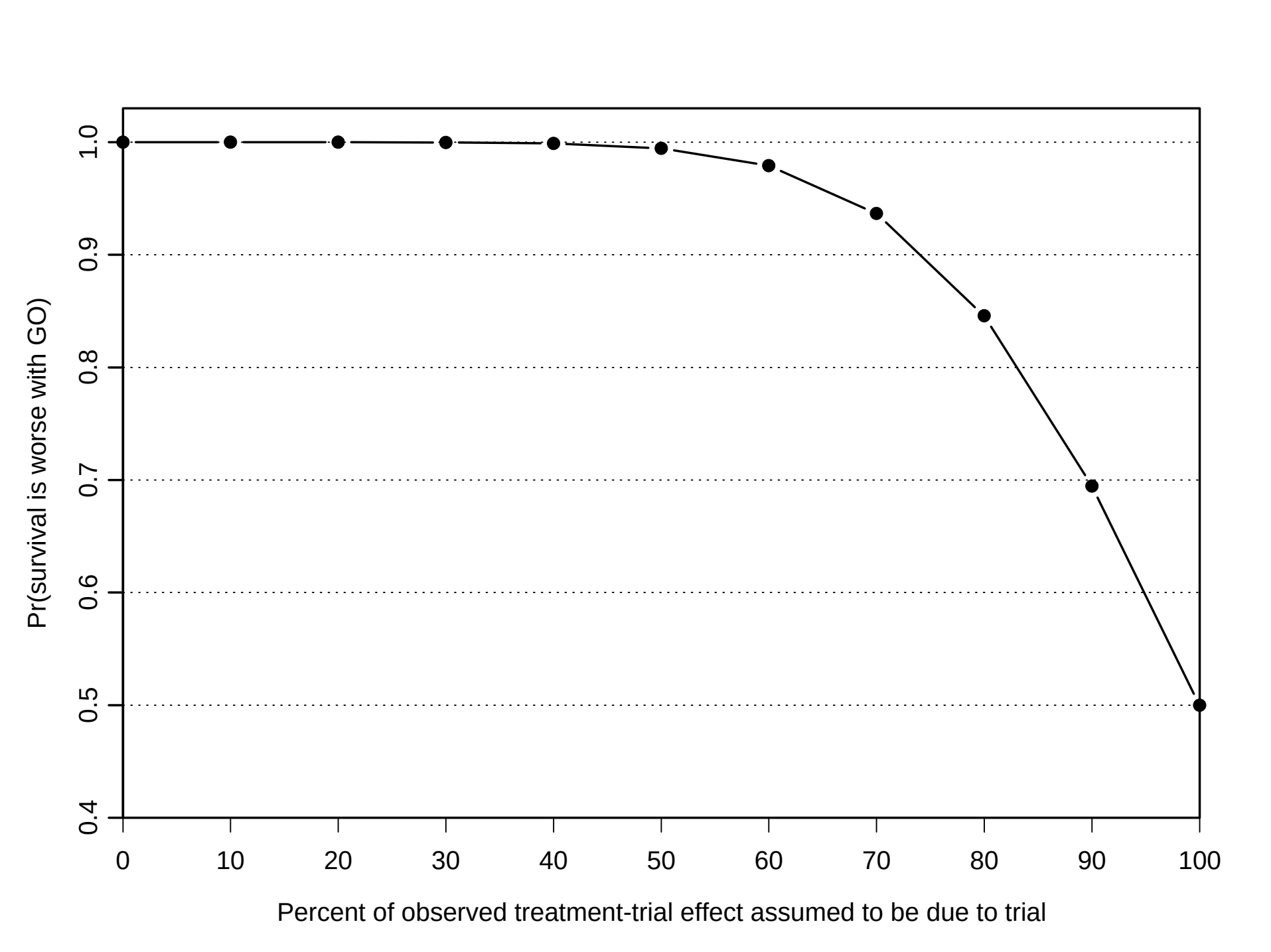
	Posterior Distribution	
Variable	**Mean (sd)**	**95% CI**
Intercept	-1.14 (.39)	(-1.93 -0.43)
PS=3,4	0.24 (.39)	(-0.55, 0.96)
Treatment in LAR	-0.43 (.34)	(-1.09, 0.22)
Cyto = –5/-7	1.52 (.43)	(0.64, 2.43)
Cyto = Other Abn.	1.22 (.40)	(0.46, 2.06)
GO Trial vs IA Trial	0.84 (.36)	(0.14, 1.53)
ϕ	0.83 (.09)	(0.66, 1.02)

First GO-vs-IA Sensitivity Analysis

1) Assume $\delta = \delta_{GO} + \delta_{\lambda}$ and
$var(\delta_{\lambda}) = ½\ var(\delta|data) = 0.065$

2) Vary $E(\delta_{\lambda})$ from 0 to $E(\delta\ |data) = 0.84$

3) Compute $pr^{(h)}(\delta_{\theta} > 0|data) =$
$pr^{(h)}(\ \delta - \delta_{\lambda} > 0|data)$ as a function of $E(\delta_{\lambda})$


Pr(survival is worse with GO)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
10
20
30
40
50
60
70
80
90
100
Percent of observed treatment-trial effect assumed to be due to trial

Second GO-vs-IA Sensitivity Analysis

Incorporate additional historical data

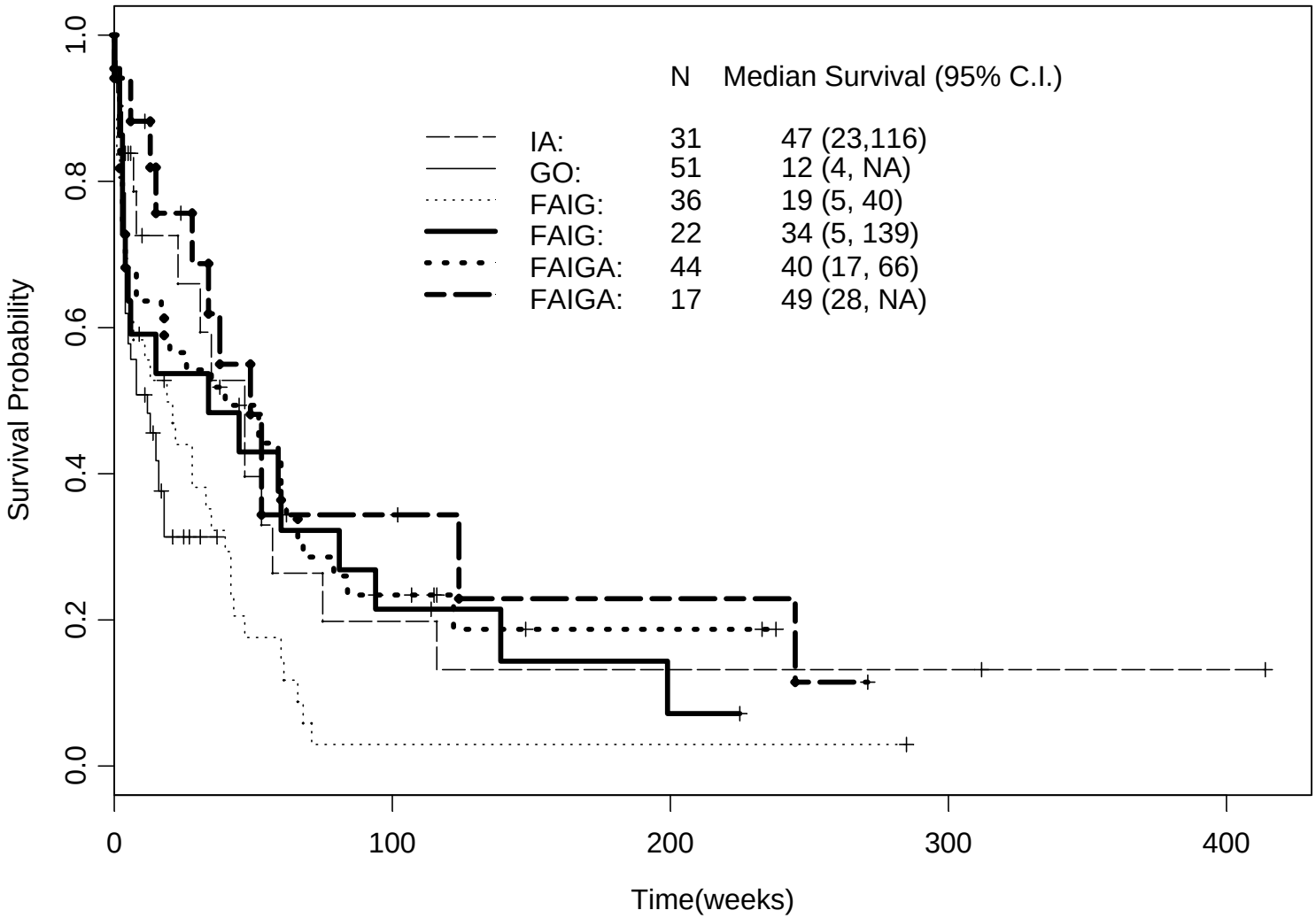
From four other trials:

- **Two trials of FAIG**
  (fludarabine + ara-C + idarubicin + G-CSF)
- **Two trials of FAIGA (FAIG+ATRA)**

Survival in the Six Trials

Trial	Treatment	# Deaths / # Patients	Median (95% Credible Interval)
1	IA	16 / 31	47 (20-105)
2	GO	29 /51	12 (7-93)
3	FAIG	34 / 36	14 (7-24)
4	FAIG	18 / 22	30 (13-63)
5	FAIGA	33 / 44	37 (20-64)
6	FAIGA	12 / 17	53 (18-128)

Kaplan-Meier Plots for the 6 Trials



Second GO-vs-IA Sensitivity Analysis

- **For j=2,…,6, denote τ_j= I(trial j) and τ_j = effect of the jth treatment-trial versus IA in trial 1**

- $\delta\tau = \delta_2\tau_2 + \ldots + \delta_6\tau_6$ = **linear term of confounded treatment-trial effects vs. trial 1**

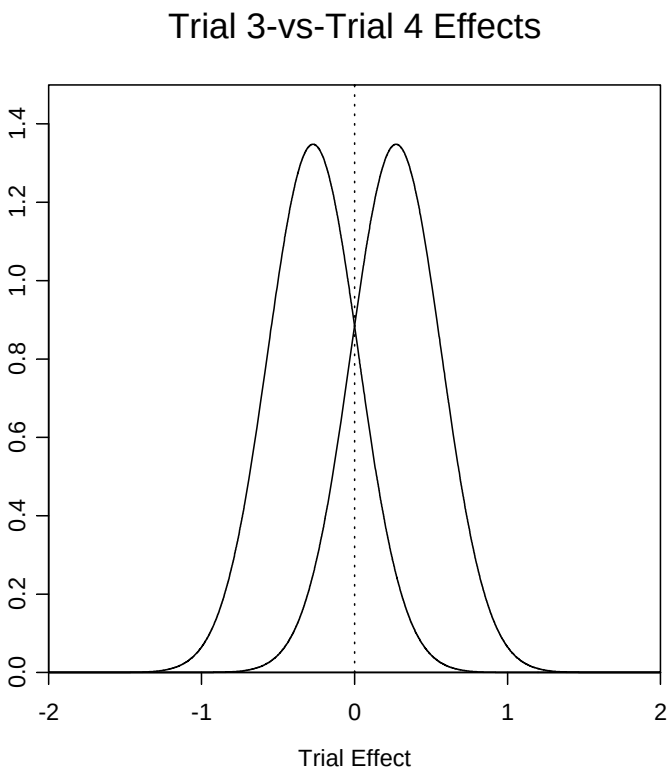
- **Fit the extended Weibull model**

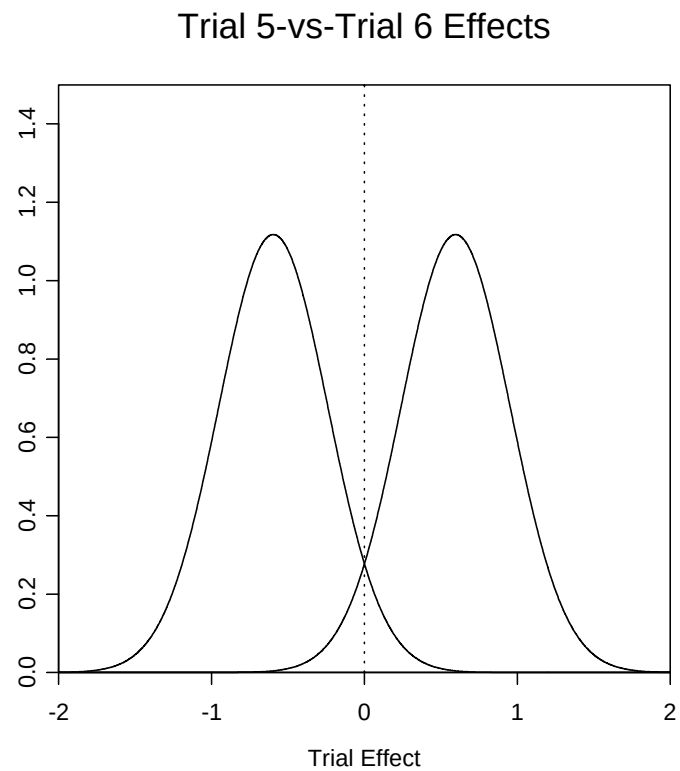
$$\log[-\log\{ S(t \mid Z, \tau)\}] = \beta Z + \delta\tau + \phi \log(t)$$

Fitted Weibull Model for All 6 Trials

	Posterior Distribution	
Variable	**Mean (sd)**	**95% CI**
Intercept	0.62 (.29)	(-1.21 -0.09)
PS=3,4	0.67 (.24)	(0.20, 1.14)
Treatment in LAR	-1.04 (.21)	(-1.45, -0.61)
Cyto = –5/-7	1.23 (.24)	(0.77, 1.71)
Cyto = Other Abn.	0.63 (.24)	(0.18, 1.09)
ϕ	0.71 (.05)	(0.63, 0.81)
δ_2 **(GO)**	0.84 (.33)	(0.20, 1.48)
δ_3 **(FAIG)**	0.74 (.33)	(0.10, 1.45)
δ_4 **(FAIG)**	0.47 (.38)	(-0.27, 1.23)
δ_5 **(FAIGA)**	0.39 (.34)	(-0.25, 1.08)
δ_6 **(FAIGA)**	-0.21 (.40)	(-0.96, 0.61)

Posterior Between-Trial Effects





FAIG studied in trials 3 and 4; FAIGA studied in trials 5 and 6

Additivity Assumptions

Trial	Treatment	Identifiable Effects	Assumed Effects
1	IA	$\delta_1 = 0$	
2	GO	δ_2	$\delta_{GO} + \delta_{\lambda,2}$
3	FAIG	δ_3	$\delta_{FAIG} + \delta_{\lambda,3}$
4	FAIG	δ_4	$\delta_{FAIG} + \delta_{\lambda,4}$
5	FAIGA	δ_5	$\delta_{FAIG} + \delta_{\lambda,5}$
6	FAIGA	δ_6	$\delta_{FAIG} + \delta_{\lambda,6}$

Computing Hypothetical $\delta_{\lambda,2} = \delta_{\lambda,2} - \delta_{\lambda,1}$

$\delta_3 = \delta_{FAIG} + \delta_{\lambda,3}$ and $\delta_4 = \delta_{FAIG} + \delta_{\lambda,4}$ ➔

$\delta_3 - \delta_4 = (\delta_{FAIG} + \delta_{\lambda,3}) - (\delta_{FAIG} + \delta_{\lambda,4}) = \delta_{\lambda,3} - \delta_{\lambda,4}$

$\delta_5 = \delta_{FAIGA} + \delta_{\lambda,5}$ and $\delta_6 = \delta_{FAIGA} + \delta_{\lambda,6}$ ➔

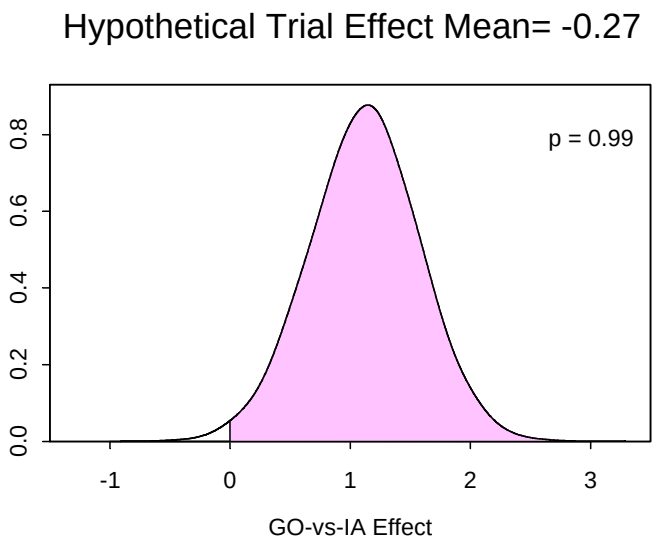
$\delta_5 - \delta_6 = (\delta_{FAIGA} + \delta_{\lambda,5}) - (\delta_{FAIGA} + \delta_{\lambda,6}) = \delta_{\lambda,5} - \delta_{\lambda,6}$

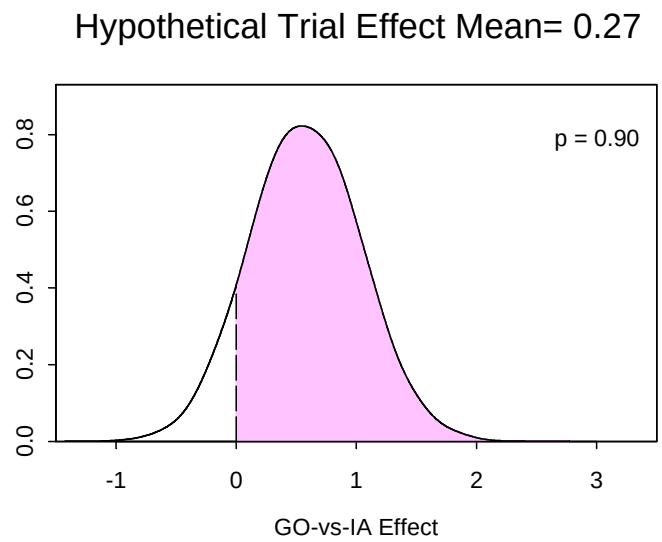
➔Use the actual between-trial effects
$\pm(\delta_3 - \delta_4)$ and $\pm(\delta_5 - \delta_6)$
as hypothetical $\delta_{\lambda,2} - \delta_{\lambda,1} = \delta_{\lambda,2} - 0 = \delta_{\lambda,2}$

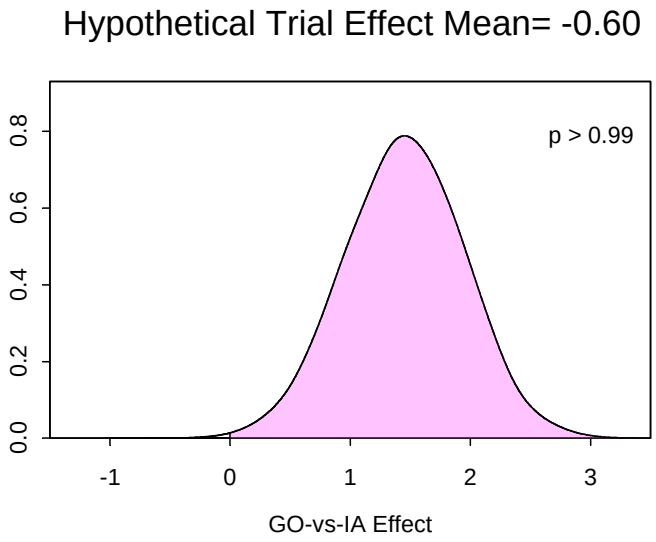
Second GO-vs-IA Sensitivity Analysis

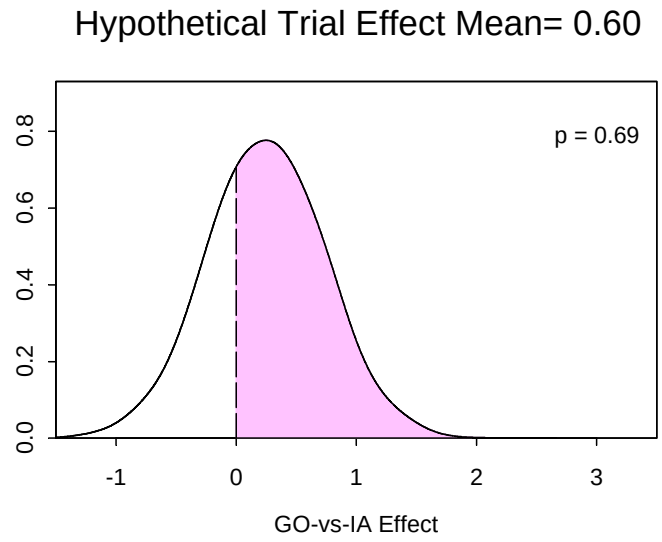
Hypothetical Effects		Hypothetical Posterior pr(GO is inferior to IA) = pr(δ_{GO}>0|data)
$\delta_{\lambda,2}$	δ_{GO}	
-0.27 (.29)	1.11 (.44)	0.99
0.27 (.29)	0.56 (.44)	0.90
-0.60 (.35)	1.44 (.50)	>0.99
0.60 (.35)	0.23 (.47)	0.69
0.84 (.35)	0.00 (.48)	0.50
1.45 (.35)	-0.61 (.48)	0.10
1.62 (.35)	-0.78 (.48)	0.05
1.94 (.35)	-1.10 (.48)	0.01

Four Hypothetical GO-vs-IA Effect Posteriors









What We Do Not Assume

Given the data

$f(\delta_{\lambda,2} - \delta_{\lambda,1}) = f(\delta_{\lambda,3} - \delta_{\lambda,4})$ or $f(\delta_{\lambda,5} - \delta_{\lambda,6})$

This assumption would imply that, once one between-trial effect distribution is available, thereafter one never needs to randomize.

➔ $f(\delta_{\lambda,3} - \delta_{\lambda,4})$ and $f(\delta_{\lambda,5} - \delta_{\lambda,6})$ are ***hypothetical*** versions of $f(\delta_{\lambda,2} - \delta_{\lambda,1})$

References

[1] Estey EH, Thall PF, Wang X, et al. Gemtuzumab ozogamicin with or without interleukin 2 in patients 65 years of age or older with untreated AML and high-risk MDS: comparison with idarubicin + continuous infusion high-dose cytosine arabinoside. ***Blood*****,** 99:4343-4349, 2002.

[2] Estey EH, Thall PF. New designs for phase 2 clinical trials. ***Blood***, 102:442-448, 2003.

[3] Thall PF, Cook JD. Dose-finding based on efficacy-toxicity trade-offs**.** ***Biometrics*****,** 60:684-693, 2004.

[4] Thall PF, Wathen JK. Covariate-adjusted adaptive randomization in a sarcoma trial with multi-stage treatments. ***Statistics in Medicine.*** **In press.**

[5] Thall PF and Wang X. Bayesian sensitivity analyses of confounded treatment effects. ***Handbook of Statistics in Clinical Oncology: 2nd Edition, Revised and Expanded***, (J.Crowley and D. Pauler, eds.) New York: Marcel-Dekker, 2005. **In press.**