Bayesian Hybrid Adaptive Designs for Clinical Trials

Peter F. Thall, PhD
Department of Biostatistics
M.D. Anderson Cancer Center

Cancer Outcomes Research Program and
Sheps Center for Health Services Research
University of North Carolina
September 13, 2013

Outline

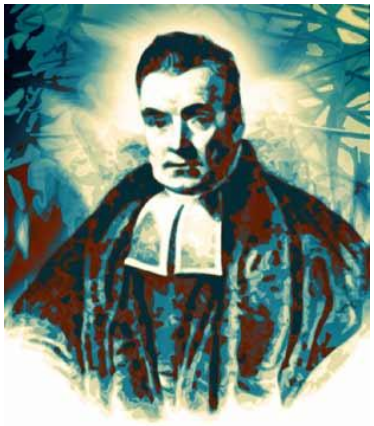
1. Some Bayesian concepts and examples

2. A phase II design based on a hierarchical model

3. Phase I-II utility-based dose-finding
   a. Radiation therapy for pediatric brain tumors
   b. Dose-finding in two cycles

4. Two-stage strategies for treating metastatic renal cancer: Re-randomization

Bayesian Statistics



Bayesians consider model parameters, θ,
to be *random* and give them "prior distributions."
θ = treatment effect, median survival, Pr(toxicity)

Bayes' Theorem prior(θ) + ***data*** $\longrightarrow$ posterior(θ | ***data***)

Start with prior(θ). Observe ***data***. Compute posterior(θ | ***data***). Use it for making inferences about θ, making decisions, and choosing actions.

Bayesian Learning If new data are obtained sequentially in a clinical trial, Bayes' Theorem may be applied repeatedly:
"Posterior" at each stage = "Prior" for the next stage.

→ **Bayesian** inference is based on ***prior + observed data***.
→ **Frequentist** inference is based on ***observed data + data that might have been observed.***

Some Advantages of Bayesian Statistics

1. Accounts naturally for multiple sources of variability:
   Patients, Covariates, Studies, Measurement error

2. Naturally incorporates historical data or expert opinion

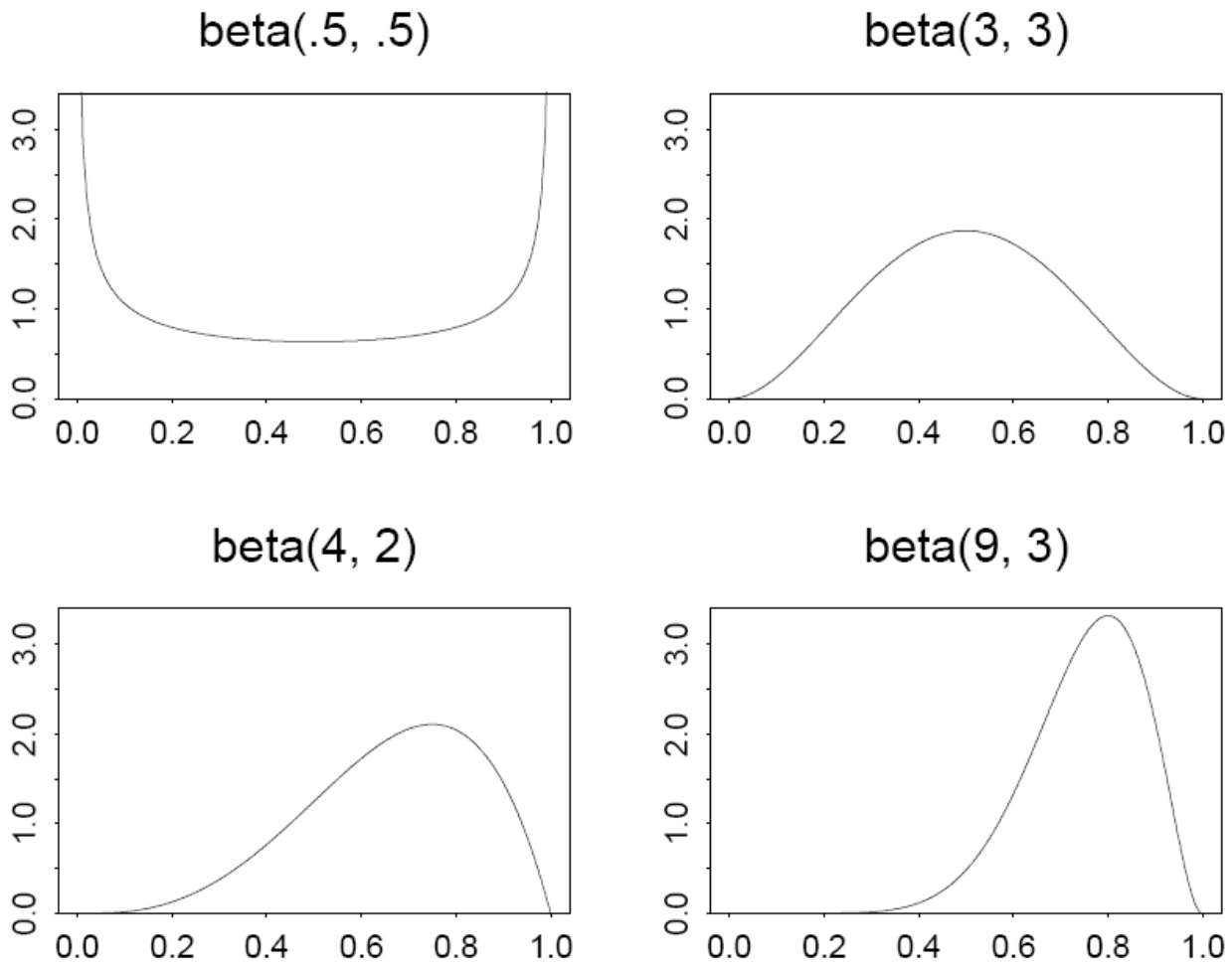
3. Hierarchical models provide a basis for combining data from multiple sources to do meta-analyses

4. Provides a coherent way to use accumulating data to make sequences of decisions

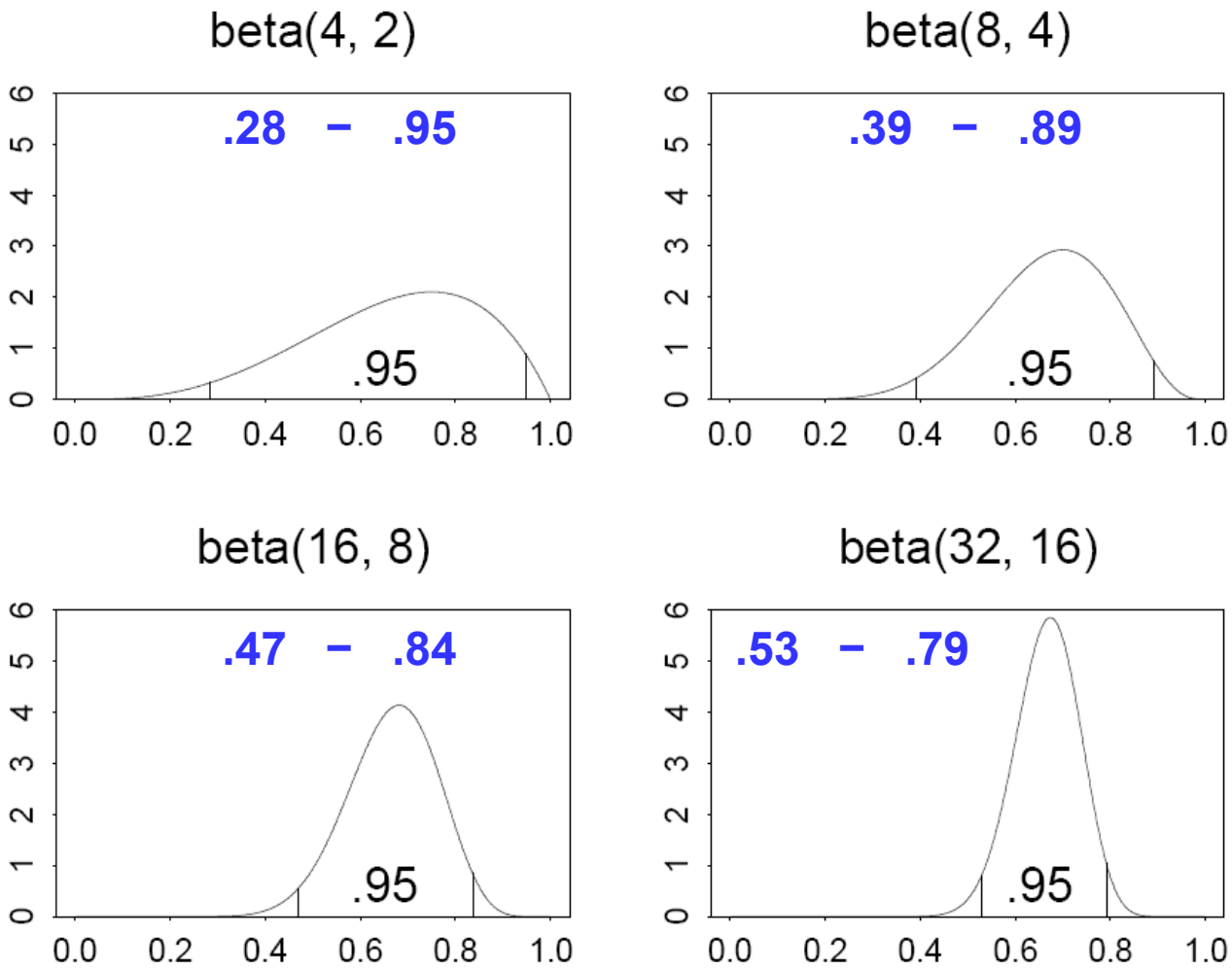
5. Posterior probabilities and credible intervals are easy to understand (unlike p-values and confidence intervals)

6. Plots of prior and posterior distributions illustrate knowledge

Some Distributions on θ = Pr(Response)



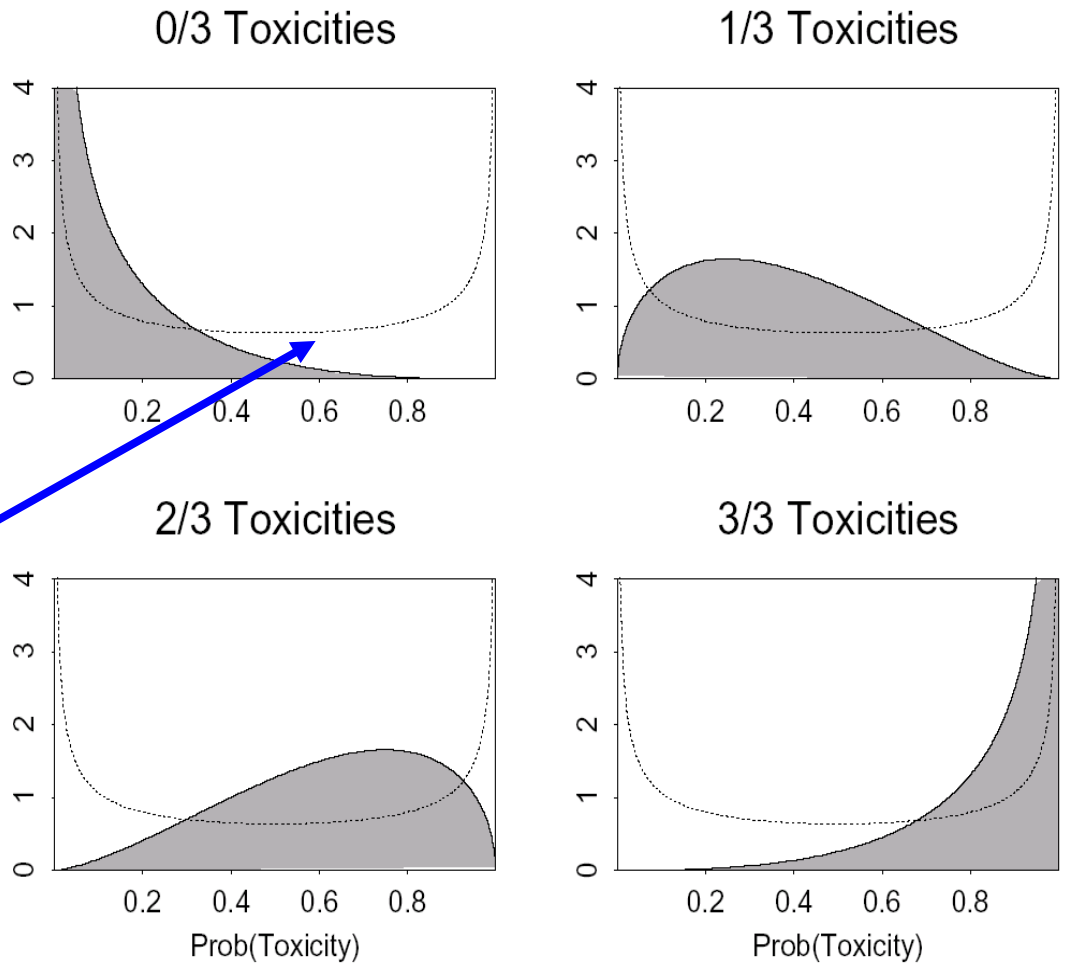
Bayesian Estimation: 95% Posterior Credible Intervals for θ under 4 different beta distributions, all with mean 2/3



[L, U] is a 95% Posterior CI if Prob[L < θ < U | data] = .95

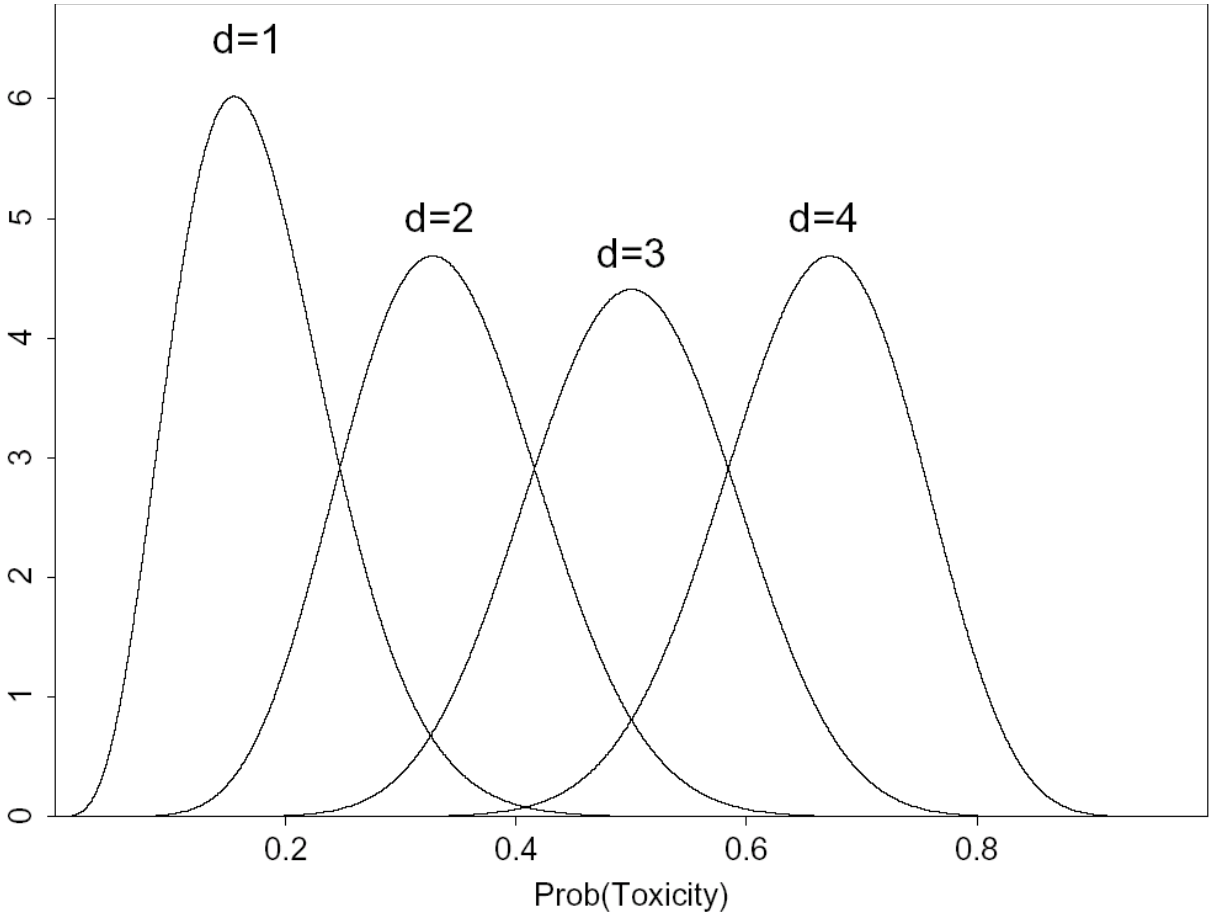
Posteriors based on the 4 possible binary toxicity data sets from 3 patients

Uninformative prior on θ = Pr(Toxicity), with effective prior sample size = 1



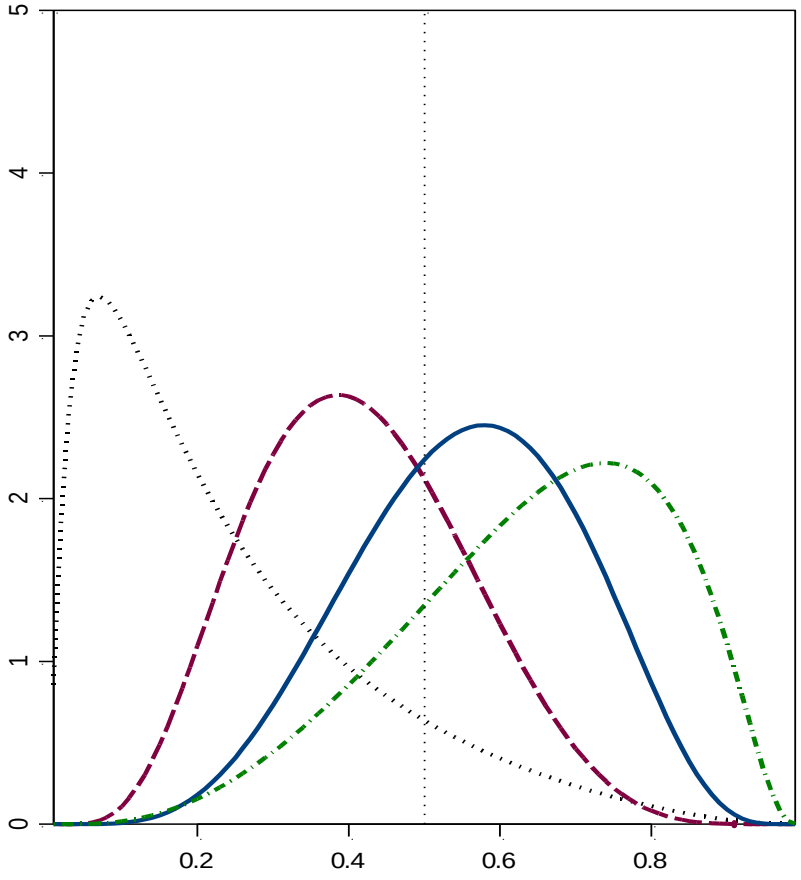
Borrowing Strength

Posterior distributions of Pr(Toxicity | *d* = *dose*, θ)
under an assumed dose-toxicity model

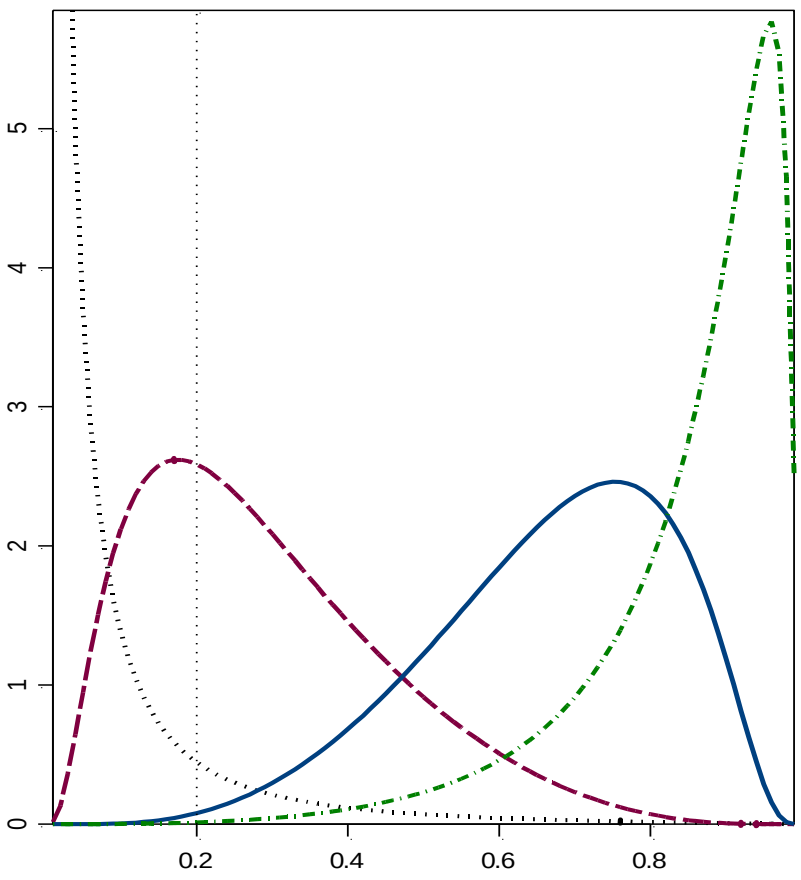


Two Outcomes

Posteriors of the probabilities $\pi_T(dose)$ of Toxicity and $\pi_E(dose)$ of Efficacy in a phase I-II clinical trial, based on data from 12 patients



Posteriors of $\pi_T(dose)$



Posteriors of $\pi_E(dose)$

Why Bayesian? A Very Simple Statistical Problem

θ = Pr(Toxicity) at a fixed dose of an experimental agent.

Observe X = [# toxicities] in 3 patients. **How to estimate θ ?**

Usual estimator: ***The sample proportion =***

[# toxicities] / [sample size] has four possible values:

0/3 = 0, 1/3, 2/3, 3/3 = 1 (0%, 33%, 67%, 100%)

But this estimator may not make sense :

→ "I estimate that the probability of toxicity equals 0" **says that you believe that toxicity is *impossible*.**

→ "I estimate that the probability of toxicity equals 1" **says you believe that toxicity is *certain*.**

The usual textook 95% ci for θ is **[0, 0] if X=0, and is [1, 1] if X=1**

Bayesian Estimator

θ = Pr(Toxicity) **is considered to be *random***.

Assume a non-informative *beta(.5, .5)* prior on θ

Given data X = # toxicities observed in 3 patients,

the posterior mean of θ is

(.50) ¼ + (X / 3) ¾

Prior Mean **Sample Mean**

Frequentist versus Bayesian Estimation

Number of Toxicities	Sample Mean	Posterior Mean of θ	Posterior 95% Credible Interval for θ
0	**0**	**.125**	.00015 — .54
1	**.333**	**.375**	.04 — .82
2	**.667**	**.625**	.18 — .96
3	**1**	**.875**	.46 — .9998

A Recommendation

When constructing a clinical trial design . . .

It is better to kill computer generated patients, rather than real ones, when calibrating design parameters.

Designing Bayesian Sequentially Adaptive Trials

1) <u>The Physician(s) Must Specify</u>

→ Disease, entry criteria

→ Treatments, doses, schedules, multi-stage regimes

→ Maximum N, trial duration, follow up, accrual rate

→ Information to establish a prior

→ **Utilities of clinical outcomes (or other criteria)**

→ Numerical limits for rules to protect patients.

E.g. an upper limit on Pr(toxicity)

2) <u>The Statistician Specifies</u> a **Bayesian Probability Model** for the clinical outcomes as functions of treatments (dose, schedule, etc.) and covariates, and a design

Designing Bayesian Sequentially Adaptive Trials

3) **Write a computer program, if necessary,** and **Simulate the Trial on a Computer** to calibrate design parameters and obtain good Operating Characteristics :

→ Sample Size, Pr(Select), Pr(Drop) for each treatment, dose, or regime

→ Pr(Stop the Trial Early). This should be large in cases where no treatment or dose is acceptable

4) **Iterate Steps 1 – 3** until a design that is ethically and scientifically acceptable is obtained

Hierarchical Models

Illustration: Design a phase II trial to evaluate

π = Pr(Tumor Response) with Imatinib

in 10 different sarcoma subtypes (Thall et al. 2003)

Approach 1 : Assume the subtypes have the same π and conduct one trial with one early stopping rule for futility.

But what if the subtypes have different Pr(Tumor Response) ?

Approach 2 : Assume the subtypes have different response probabilities, $\pi_1, \ldots, \pi_{10,}$ and conduct 10 trials, each with its own stopping rule. → But are the 10 subtypes really independent ?

Is conducting 10 trials feasible? What about rare subtypes ?


Kennels von


'99 11 6


ZABAVA

Bayesian Hierarchical Model

Response data in S_1

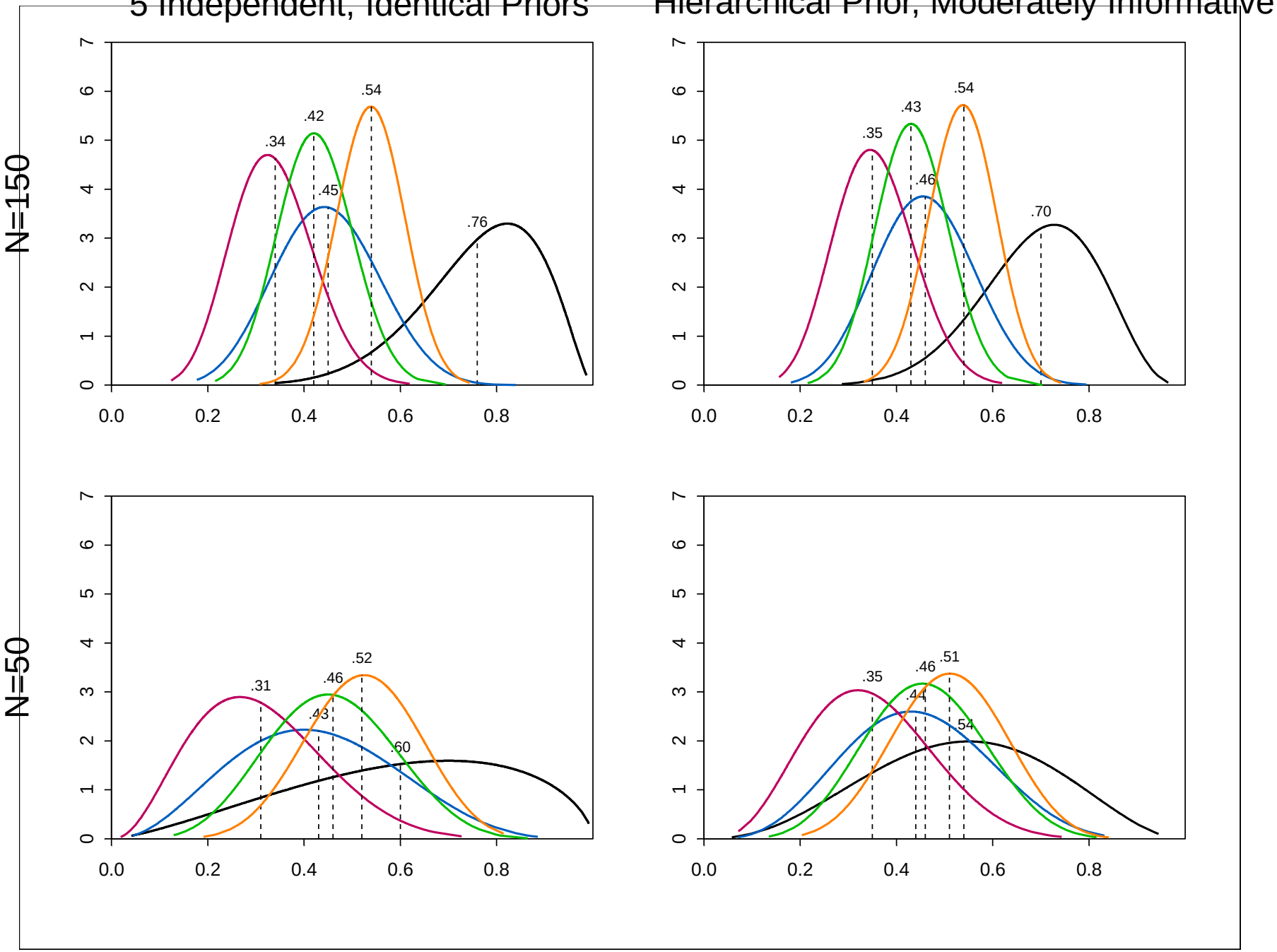
Response data in S_2

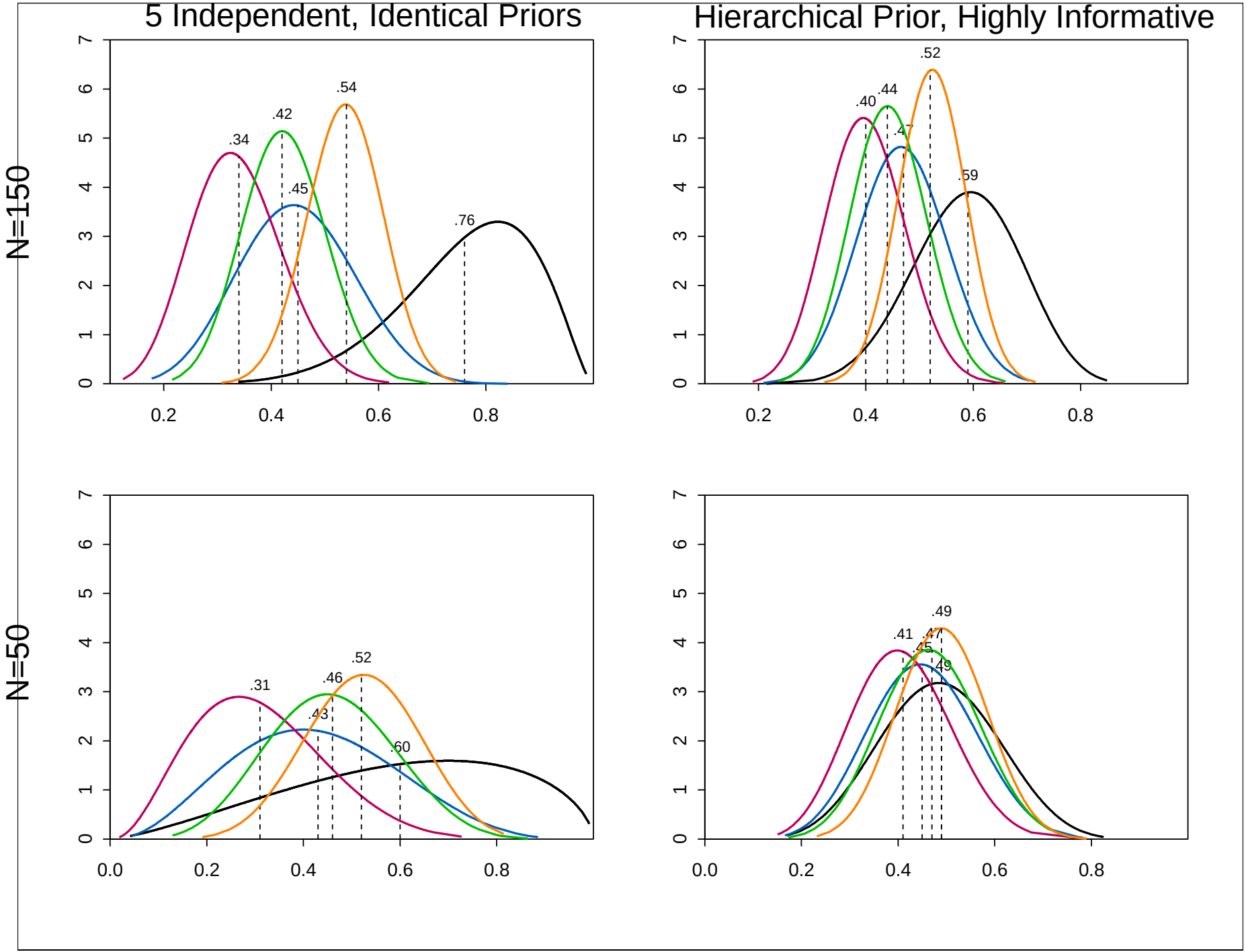
Response data in S_{10}

π_1 π_2 π_{10}

Prior on π_1 Prior on π_2 Prior on π_{10}

Hyper Prior


5 Independent, Identical Priors
Hierarchical Prior, Moderately Informative
N=150
N=50
.34
.42
.54
.45
.76
.35
.43
.54
.46
.70
.31
.46
.52
.43
.60
.35
.46
.51
.44
.54
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
6
7


5 Independent, Identical Priors
Hierarchical Prior, Highly Informative
N=150
N=50
.34
.42
.54
.45
.76
.40
.44
.52
.47
.59
.31
.46
.52
.43
.60
.41
.47
.49
.45
.49

Example of Within-Subtype Futility Stopping Rules for 7 disease subtypes in a phase II trial

Case	Outcomes	Decision	
		Non-Hierarchical	Hierarchical
1	3 subtypes with 0/9	Terminate	Terminate
	4 subtypes with 1/9	Continue	Continue
2	2 subtypes with 0/9	**Terminate**	**Continue**
	2 subtypes with 1/9	Continue	Continue
	3 subtypes with 2/9	Continue	Continue
3	3 subtypes with 0/15	Terminate	Terminate
	4 subtypes with 1/15	**Continue**	**Terminate**

Advantages of Bayesian Hierarchical Model

1) Allows data from each subtype to provide information about π_j's in all of the other subtypes ("Borrowing Strength")

2) Borrowing strength between subtypes reduces both false negative and false positive rates

3) Avoids the two undesirable approaches of conducting

→ 1 trial assuming 1 common π, ignoring the subtypes

→ 10 separate trials that ignore each others' data, and that probably are not feasible

The hyperprior must be calibrated to accurately reflect strength (informativeness) of prior belief or historical data.

A Dose-Finding Trial in Pediatric Brain Tumors

Diffuse Intrinsic Pontine Gliomas (DIPGs)

→ Very aggressive brain tumors.

→ **Median patient age = 5 years**

→ No treatment with substantive anti-disease activity exists, with **median survival < 1 year.**

→ Radiation Therapy (RT) is standard treatment, but it is **mainly palliative**.

→ The RT dose-toxicity and dose-efficacy profiles are not well understood.

Oncologists' Definition of **Toxicity Severity Levels**

	Mild	Moderate	High	Severe
CNS		Asymptomatic brain necrosis seen on MRI, not attributable to tumor progression	Brain necrosis, not attributable to tumor progression, requiring non-surgical therapy	**Death** attributed to radiation treatment **Edema** in reirradiated brain tissue **requiring surgical intervention** **Unilateral or total blindness attributed to radiation-related optic neuropathy**
Fatigue	Lasting <1 month	Lasting 1-3 months	Lasting > 3 mos	
Nausea / Vomiting	Controlled with antiemetics	Decreased appetite	Started during and up to 3 weeks after radiation treatment that cannot be controlled with antiemetics	
Headache	Headache that started during radiation treatment controlled with non-steroidal medications	Headache that started during radiation treatment requiring steroids	Headache that started during radiation treatment that cannot be controlled with medications	
Skin	Skin erythema in rad. field. Alopecia	Dry desquamation in radiation field	Moist desquamation in radiation field	

Outcomes in the Pediatric Brain Tumor Trial

Toxicity = Low, Moderate, High, or Severe

Efficacy = Total number of improvements in

(i) Clinical Symptoms

(ii) Radiographic Appearance of the Tumor

(iii) Quality of Life

→ Possible efficacy values = 0, 1, 2, or 3

(Toxicity, Efficacy) scored at day 42

4 x 4 = 16 possible (Toxicity, Efficacy) outcomes

Numerical **Consensus Utilities**

Elicited from A. Mahajan and H. Fontanilla, co-PIs

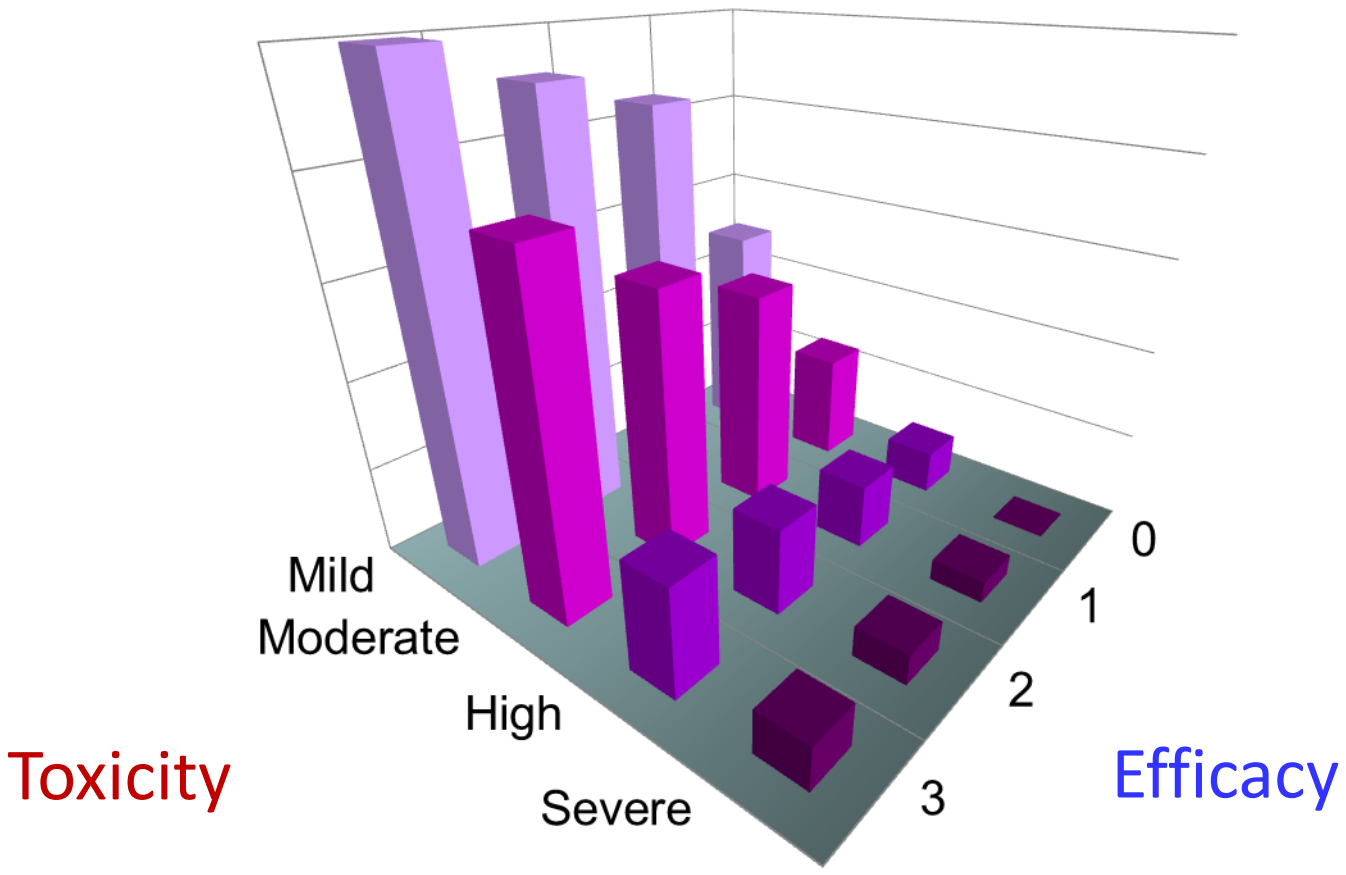
		Toxicity Severity			
		Low	**Moderate**	**High**	**Severe**
Efficacy Score	**0**	**50**	**25**	**10**	**0**
	1	**85**	**50**	**15**	**5**
	2	**92**	**60**	**20**	**7**
	3	**100**	**75**	**25**	**10**

U(Toxicity, Efficacy) are the basis for making decisions adaptively in the trial ("learn-as-you go"), currently ongoing at MD Anderson:

1) Decide which radiation does are acceptable

2) Choose the best dose for each successive cohort of 3 children

Joint Outcome Utilities



Why Bother With Utilities ?

		Toxicity Severity			
		Low	Moderate	High	Severe
Efficacy Score	0	50	25	10	0
	1	85	50	15	5
	2	92	60	20	7
	3	100	75	25	10

Question: If **Tox** = {Low, Moderate} is "acceptable" but {High, Severe} is "not acceptable" why not just use

DLT = {High, Severe} and apply a usual dose finding method (e.g. the "3+3" or "CRM") ?

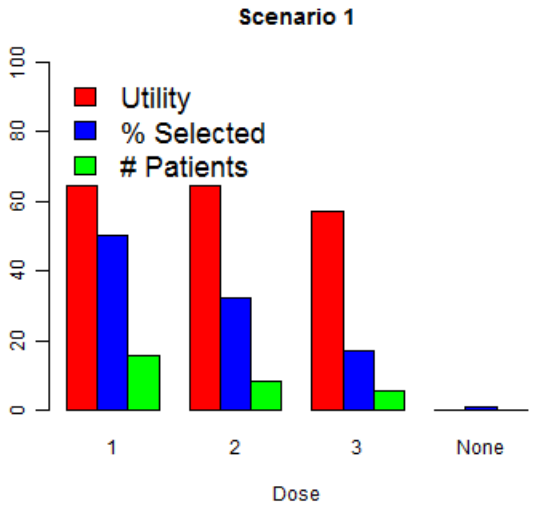
Answer: U(0,Moderate) = U(3, High) = 25 → Scoring these two outcomes as "No DLT" and "DLT" makes no sense!

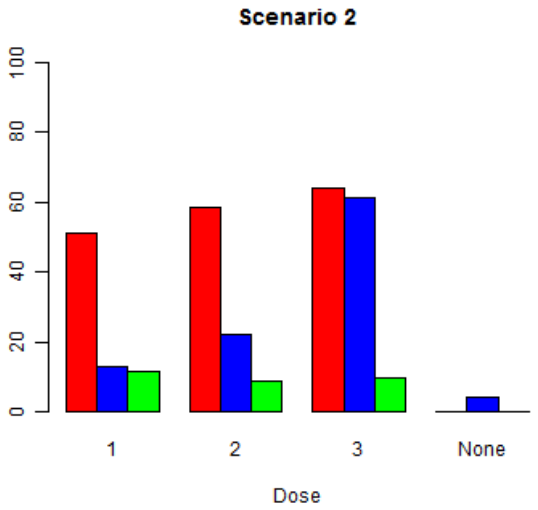
Conduct of the Radiation Therapy Trial

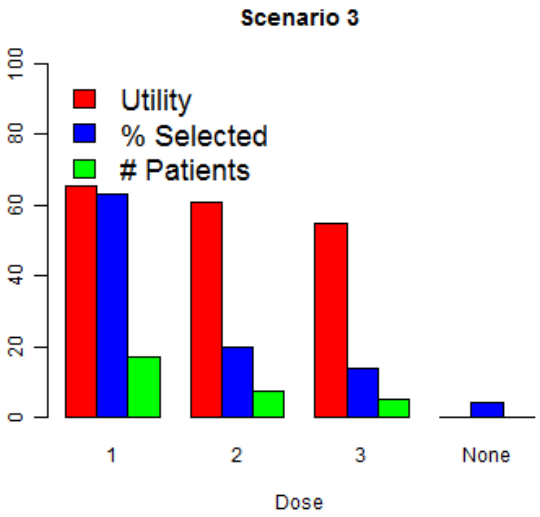
1) Accrual rate = 6 to 10 patients/year
2) N = 30 children maximum, cohorts of size 3
3) Treat the first cohort of 3 patients at the lowest dose, then apply the adaptive utility-based criterion.
4) Do not skip the middle dose when escalating.
5) A dose is unacceptably toxic if is it likely to have

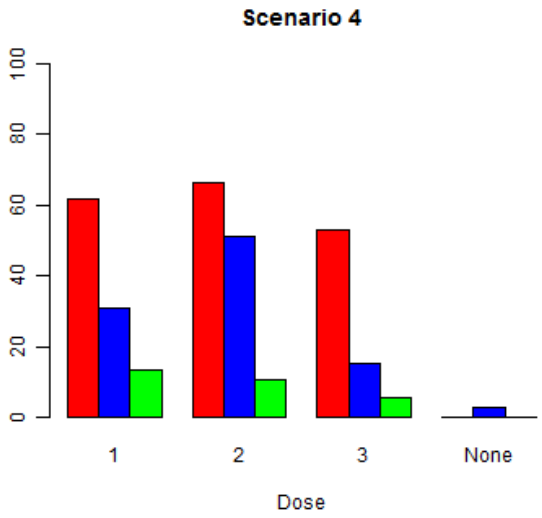
   Pr(High or Severe toxicity) > 10%

Computer Simulations: Operating Characteristics of RT Trial Design

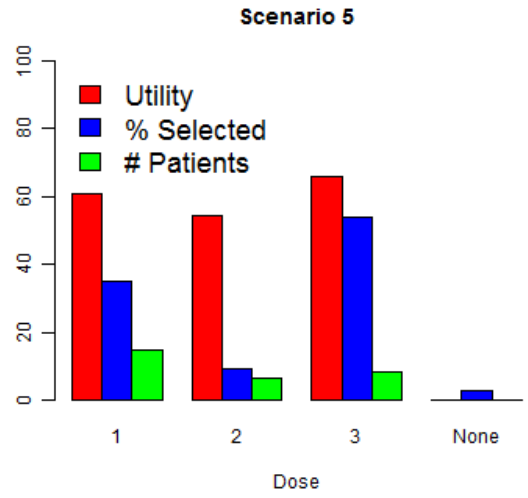


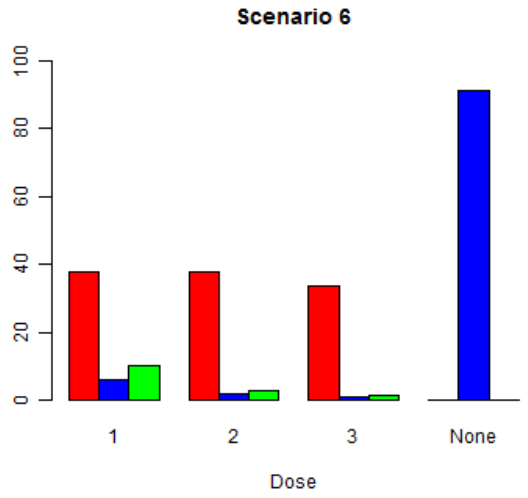


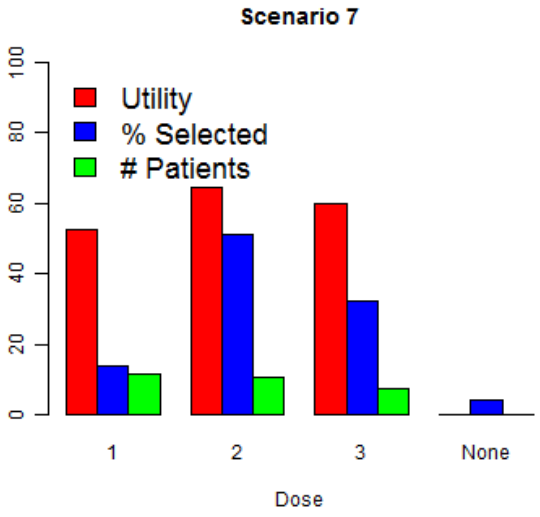


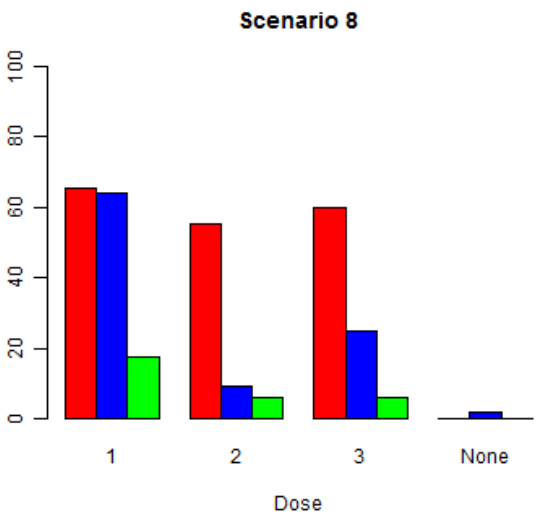


Computer Simulations: Operating Characteristics of RT Trial Design









Some Early Trial Results

By design, the first 3 patients were treated at BED level 1.

At BED level 1 n=4 patients had outcomes :

(Eff, Tox)	(0, Mod)	(2, Mod)	(1, Low)	(2, Low)
Utility	25	60	85	92

At BED level 2 : n=1 patient had outcome

(Eff, Tox) = (1, Low), for Utility = 85

Dose Finding Based on Efficacy and Toxicity in Two Treatment Cycles

Goal: Develop a practical phase I-II trial design to adaptively optimize each patient's doses in two cycles of therapy, using binary (**Toxicity** , **Efficacy**) in each cycle.

Methodology: Base cycle-specific actions on numerical utilities

1. **Actions** (a_1 a_2) **in each cycle** : Treat with the "optimal" dose, or possibly "Do not to treat (NT)"

2. **Bayesian hierarchical dose-outcome model**

3. **Safety**: Include dose acceptability rules

4. **Optimize** (a_1 a_2): Backward induction ←

using posterior means of a utility-based

objective function

Actions versus Doses

Bellman's Idea: First find a_2^{opt} by considering all possibilities, then work backwards to find a_1^{opt}, assuming that a_2^{opt} will be taken.

Finding (a_1^{opt}, a_2^{opt}) is not the same thing as optimizing doses separately in each cycle.

Example: $(d_1^{opt}, d_2^{opt}) = (3, 2)$ but $(a_1^{opt}, a_2^{opt}) = (3, a_2^{opt})$ where

$a_2^{opt}(d_1=3, \text{No Tox}_1, \text{Eff}_1) = 3$

$a_2^{opt}(d_1=3, \text{No Tox}_1, \text{No Eff}_1) = 4$

$a_2^{opt}(d_1=3, \text{Tox}_1, \text{Eff}_1) = 1$

$a_2^{opt}(d_1=3, \text{Tox}_1, \text{No Eff}_1) = \text{NT}$

Properties of the Hierarchical Model

Eff and Tox are each defined using latent (unobserved) continuous variables to facilitate computation

The model includes random patient effects

Pr(Eff) and Pr(Tox) each increase with dose

Numerical dose values are not used, just indices

d=1, 2, 3, 4, 5

Prior parameters were calibrated to have overall prior effective sample size < 2.0

Objective Functions

Utilities

	Eff=Yes	Eff=No
Tox = Yes	**65**	**0**
Tox = No	**100**	**35**

Cycle 2 Objective Function $q_2(a_2, d_1, Eff_1, Tox_1)$ =

Expected utility of action a_2 in cycle 2 if d_1 was given in cycle1 and the outcomes were (Eff_1, Tox_1)

Cycle 1 Objective Function

$q_1(d_1)$ = {Expected utility of giving d_1 in cycle 1}

+ .80 {Expected utility in cycle 2 if d_1 is given in cycle 1 and a_2^{opt} is taken in cycle 2 }

Additional Constraints

Because we do not completely trust our model

Dose Acceptability

d_1 is unacceptable if $E\{Utility(d_1)\} < 35 = U(0,0)$

d_2 is unacceptable if $E\{Utility(d_1, Eff_1, Tox_1\} < 35$

Safety Constraints (to reflect actual clinical practice)

1. In each cycle, do not skip an untried dose when escalating

2. Do not escalate in cycle 2 if TOX was observed in cycle 1

Adaptive Randomization

A Major Practical Problem:

"Greedy" algorithms that always optimize some criterion risk getting stuck at a suboptimal action.

A Practical Solution:

For each cycle, given the current data, first identify the acceptable doses / actions.

Adaptively Randomize among the doses /actions that have posterior expected payoff (objective function value) "close" to the maximum value.

We call the 2-cycle method **DTM2**

Trial Conduct

5 dose levels, 60 patients in 30 cohorts of size 2

1. Cohort 1 treated at d = 1 in cycle 1, their (Eff_1 ,Tox_1) observed, posterior is computed, and cycle 2 actions a_2 taken. When (Eff_2 ,Tox_2) observed from cycle 2, re-compute the q_1 and q_2
2. Cohort 2 enrolled after cohort 1 has been evaluated for cycle 1.
3. For cohorts 2, 3, ... , compute the optimal actions and use AR to choose the actions in each cycle.
4. Repeat steps 1 – 3 until trial stopped early, or N = 60

2-Cycle Comparators: 3+3 Methods

We compared the DTM2 design to **2-cycle extensions** of 3+3 algorithms and the continual reassessment method (CRM)

(3+3)a implicitly targets d with P(TOX | d) $\leq$ 0.17

(3+3)b implicitly targets d with P(TOX | d) $\leq$ 0.33

The **extended** (3+3) methods both choose d_2 as follows:

If TOX_1, then $d_2 = d_1 - 1$ (Tox in cycle 1 → de-escalate)

If **NO** TOX_1, then $d_2 = d_1$ (No Tox in cycle 1 → repeat d_1)

2-Cycle Comparators: CRM

Two-cycle extension of the CRM:

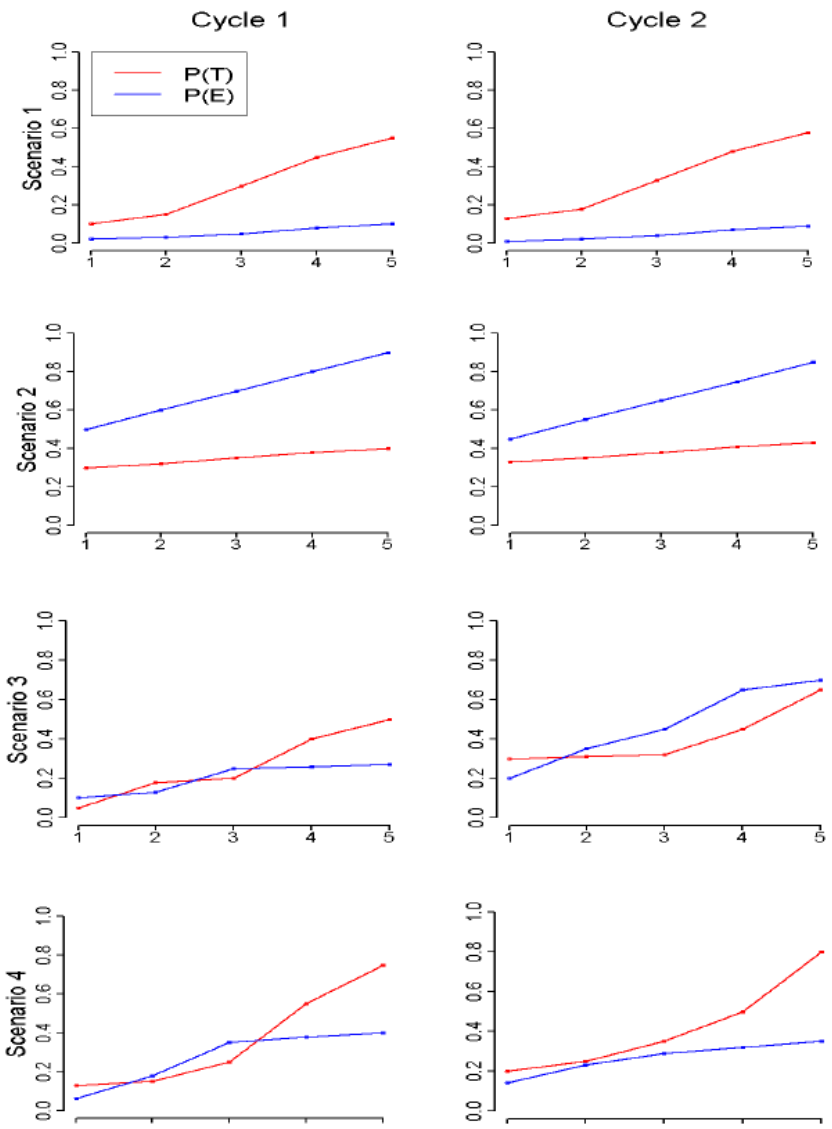
Cycle 1 : Choose d_1 with posterior mean Pr(TOX) closest to 0.30, the usual CRM, but impose the "do not skip an untried dose" rule.

Cycle 2 : Choose d_2 using the same adaptive rules as for the extended (3+3) methods.

Also, d_2 is unacceptable, given d_1, if it makes it likely that

Pr(at least one toxicity in two cycles | d_1,d_2) > .50.

Computer Simulation Scenarios



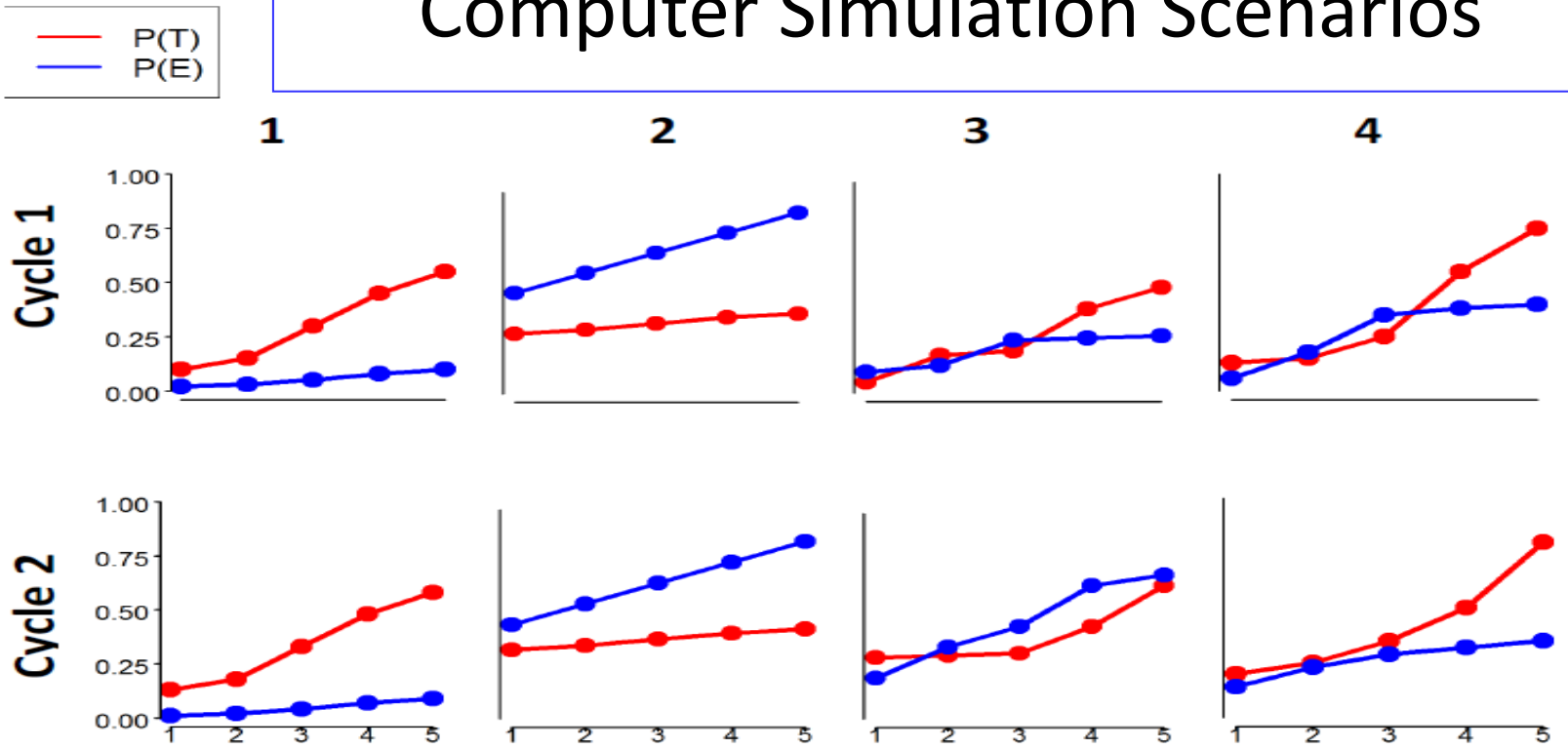
d=1,2 safe, d=3,4 toxic, no doses efficacious → Stop is best

Tox is OK, big payoff if escalate to higher doses

Complex, very important to account for cycle 1 outcomes in choosing a_2

Optimal doses are what 3+3 and CRM choose.

Computer Simulation Scenarios



d=1,2 safe,

d=3,4 toxic,

no doses

efficacious →

Stop is best

Tox OK,

big payoff

if escalate to higher doses

Complex, important to account for cycle 1 outcomes in choosing a_2

Optimal doses are what the 3+3 and CRM happen to choose.

Optimal Actions Under the 4 Scenarios

Scenario	a_1^{opt}	a_2^{opt}			
		(0,0)	(0, 1)	(1,0)	(1,1)
1	NT	NT	NT	NT	NT
2	5	5	5	4	4
3	3	4	4	2	2
4	3	3	3	NT	2

Percent Completed Trials

Scenarios	DTM2	(3+3)a	(3+3)b	Extended CRM
1	**2.3**	88.6	96.5	99.8
2	**99.4**	39.2	64.7	93.1
3	79.4	99.6	99.2	99.8
4	96.7	83.2	94.7	99.8

In Scenario 1, DTM2 correctly decides all doses are inefficacious and stops the trial 97.7% of the time. The other 3 methods all ignore the low efficacy and are very likely to continue.

Computer Simulations: Evaluation Criteria

$Q_{select}(a)$ = Expected total payoff, over 2 cycles, to a future patient treated with $a = (a_1, a_2)$

$\bar{U}$ = Mean total payoff, over 2 cycles, for the patents in the trial

Pr(TOX) = Empirical Pr(TOX) over both cycles

Pr(EFF) = Empirical Pr(EFF) over both cycles

Summary of Simulation Results

Scenario		DTM2	(3+3)a	(3+3)b	Extended CRM
2	$\bar{U}$	**136.35**	123.32	117.78	116.38
	Q_{select}	**135.76**	103.85	104.48	103.17
	Pr(Tox)	**0.39**	0.30	0.34	0.37
	Pr(Eff)	**0.72**	0.57	0.55	0.57
3	$\bar{U}$	**94.23**	85.72	85.49	89.29
	Q_{select}	**84.39**	77.97	80.13	78.29
	Pr(Tox)	0.38	0.27	0.28	0.30
	Pr(Eff)	0.38	0.26	0.27	0.33
4	$\bar{U}$	**75.84**	81.83	79.85	84.88
	Q_{select}	**69.49**	74.92	75.76	79.12
	Pr(Tox)	0.51	0.25	0.27	0.29
	Pr(Eff)	0.29	0.22	0.21	0.29

Two-Stage Treatment Strategies Based On Sequential Failure Times (Thall, et al., 2007)

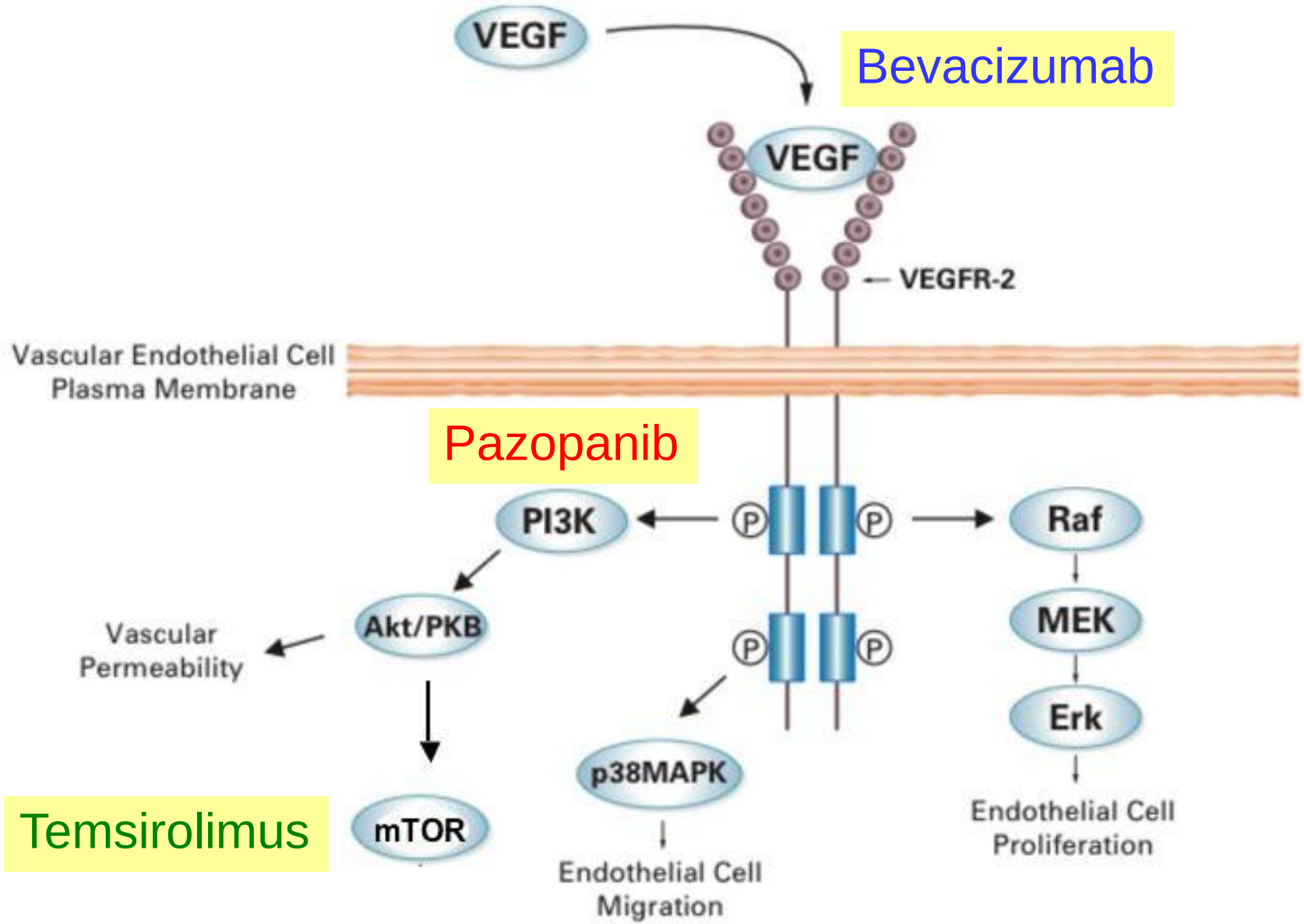
Motivating Application: ***The SPARC trial*** to treat patients with ***Metastatic Renal Cell Cancer*** (MRCC) who have not had previous systemic therapy (N. Tannir, PI)

Standard treatments are ineffective, **median(DFS) ≈ 8 mos**

Three "targeted" treatments studied in 240 MRCC patients, using 6 two-stage within-patient Dynamic Treatment Regimes (DTRs) of the form

(frontline agent, salvage agent at progression)

Three “Targeted Agents” for Metastatic Kidney Cancer



A Within-Patient Two-Stage Treatment Assignment Algorithm (***Dynamic Treatment Regime***)

SPARC Trial Treatments:

b = bevacizumab, s = sunitinib, t = temsirolimus

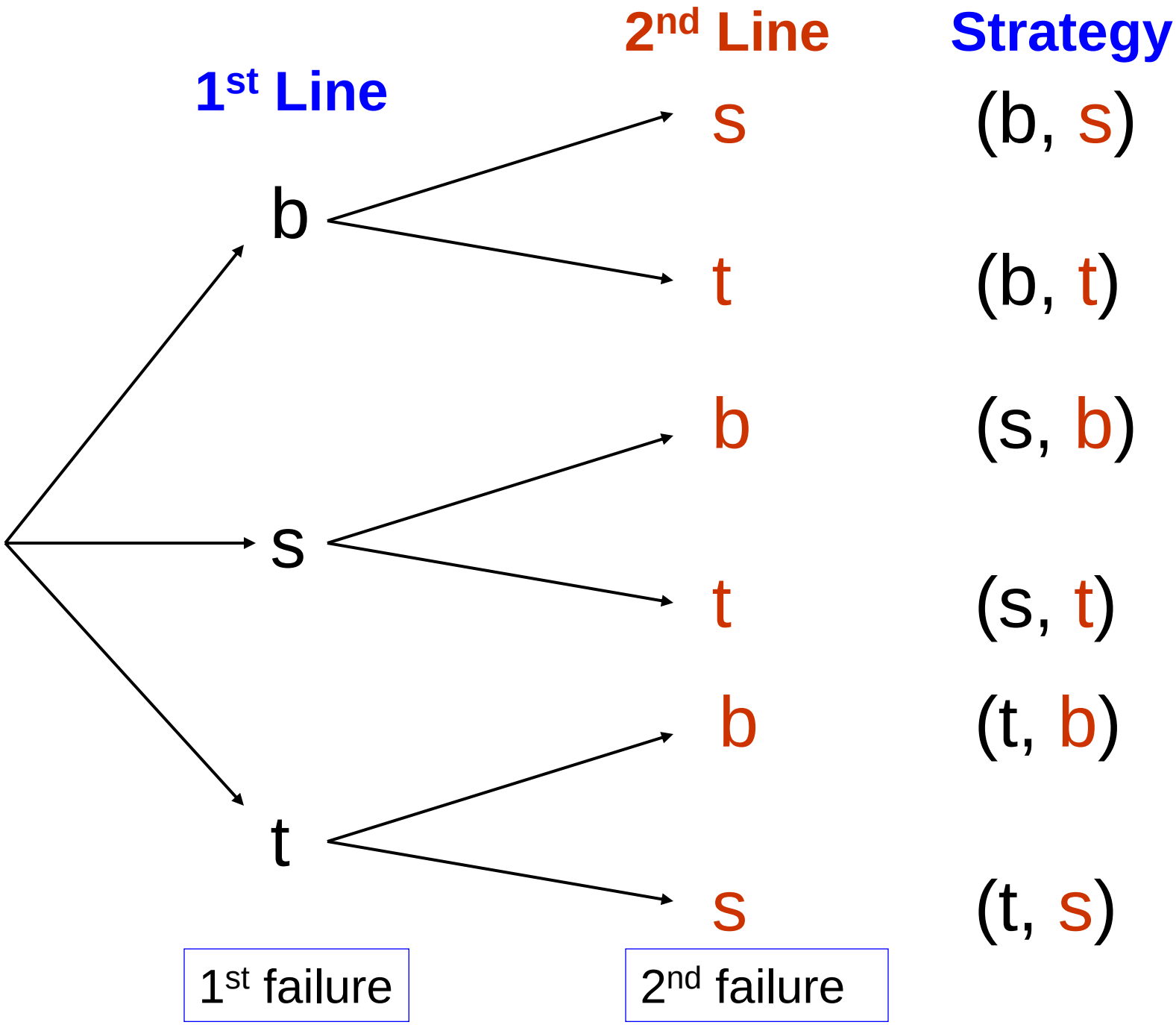
Stage1 of Therapy

At entry, randomize the patient fairly among { b, s, t }

Stage 2 of Therapy

If the stage 1 failure is disease progression and not discontinuation, re-randomize the patient fairly between the two treatments not received initially

"Switch-Away From a Loser"


1st Line
2nd Line
Strategy
b
s
t
s
t
b
t
b
s
(b, s)
(b, t)
(s, b)
(s, t)
(t, b)
(t, s)
1st failure
2nd failure

Clinical Outcomes

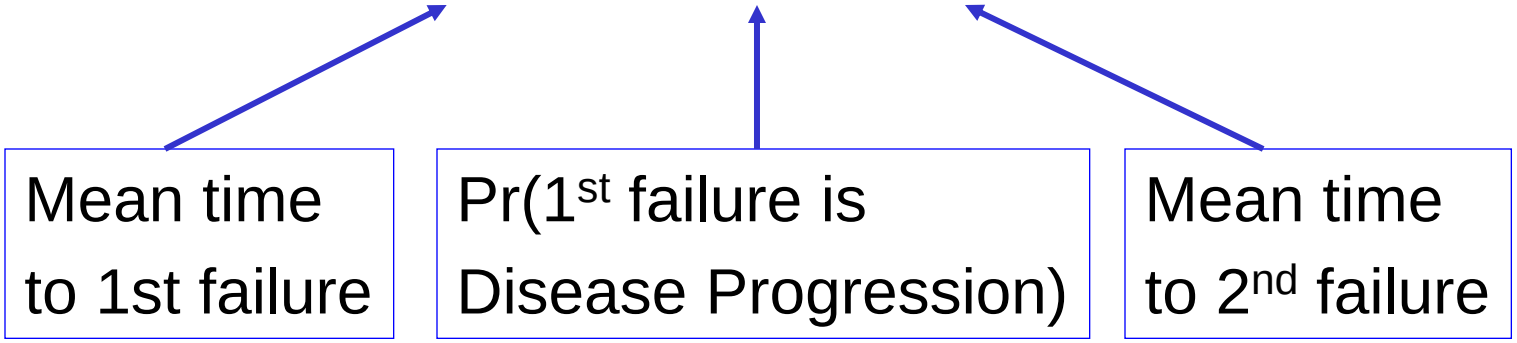
T_1 = Time to 1st treatment failure

Y = I [Patient continues to 2nd stage] = 0 or 1

T_2 = Time from 1st to 2nd treatment failure

$T = T_1 + Y\,T_2$ = Time to final treatment failure →

$$E(T) = E(T_1) + Pr(Y=1)\,E(T_2)$$



Advantages of Re-Randomization

Unbiased comparisons of the effects of the

two-stage treatment strategies on time to final failure :

(b, s) , (b, t), (s, b), (s, t) , (t, s) , (t, b)

The design accounts for the possibility that the effect of

b given after s may not be the same as the effect of

b given after t

Goal: Select the 2-stage strategy with largest mean time to final treatment failure

Complications

1) Because disease is evaluated repeatedly (by MRI or PET scan), either T_1 or $T_1 + T_2$ may be Interval Censored
2) There may be a delay between 1st failure and start of stage 2 therapy
3) T_1 may affect T_2
4) For metastatic renal cancer, the failure rate typically increases over time

A Tale of Four Designs

Design 1 (*February 21, 2006*)

***N*=240, 12 two-stage strategies, 16 patients per strategy**

Design 2 (*April 17, 2006*) Following "advice" from CTEP, NCI →

***N* = 240, 6 strategies, 32 patients per strategy**

Design 3 (*January 3, 2007*) CTEP no longer interested, but several Pharmas were now VERY interested →

***N* = 360, 6 new strategies, 48 patients per strategy**

Design 4 (*May 15, 2007*): **N=240, and a Bayesian Weeding Rule was added***:* When 120 patients are fully evaluated,

stop accrual to strategy (a,b) if

$$\Pr\{ m(a,b) < m(\text{best}) - 3 \text{ mos} \mid \text{data}\} > .90$$

In words: Drop a strategy if it is likely to have overall mean DFS more than 3 months smaller than the mean DFS of the best strategy

Computer Simulations

Simulation Scenarios were specified in terms of

$m_1(A) = \text{median}\ (T_1 \mid A)$

$m_2(A,B) = \text{median}\ \{ T_{2,2} \mid T_1 = 8, (A,B) \}$

Null values were $m_1 = 8$ and $m_2 = 3$ months

$m_1 = 12$ months → **Good frontline**

$m_2 = 6$ months → **Good salvage**

$m_2 = 9$ months → ***Very* good salvage**

Simulation Results w/o the weeding rule

		(b, s)	(b, t)	(s, b)	(s, t)	(t, b)	(t, s)
1	μ	15.7	15.7	15.7	15.7	15.7	15.7
	% select	15	17	17	18	17	16
2	μ	**19.4**	**19.4**	15.7	15.7	15.7	15.7
	% select	**52**	**48**	0	0	0	0
3	μ	15.7	**18.8**	15.7	**18.8**	15.7	15.7
	% select	0	**49**	0	**51**	0	0
4	μ	**19.4**	**23.3**	15.7	15.7	15.7	15.7
	% select	0	100	0	0	0	0
5	μ	15.7	**18.8**	15.7	**22.0**	15.7	15.7
	% select	0	**3**	0	**97**	0	0
6	μ	12.5	12.5	15.7	15.7	15.7	15.7
	% select	0	0	**28**	**25**	**25**	**23**

Sims with Weeding Rule (Scenario 5)

Acc rate		(b,s)	(b, t)	(s,b)	(s, t)	(t, b)	(t, s)
	μ	15.7	**18.8**	15.7	**22.0**	15.7	15.7
12	Pstop	.68	.24	.78	.01	.69	.70
	N	45	51	44	59	45	44
9	Pstop	.68	.25	.81	.01	.67	.71
	N	41	55	39	72	42	40
6	Pstop	.68	.22	.82	0	.68	.69
	N	37	59	34	84	37	36

General Conclusions

Utilities quantify trade-offs between adverse and efficacy events → An ethical basis for adaptive decision-making.

Bayesian Statistics provides a practical basis for design and conduct of complex clinical trials.

Computer Simulation is an essential tool for calibrating design parameters.

Accounting for Multiple Stages is much more realistic.

Major Caveat : Developing statistical models, methods, and computer programs is extremely time-consuming.

Selected Bibliography

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

Thall PF, Millikan R, Sung, H-G. Evaluating multiple treatment courses in clinical trials. *Statistics in Medicine*, 19: 1011-1028, 2000.

Thall PF, Wathen JK. Practical Bayesian adaptive randomization in clinical trials. *European J Cancer*, 43, 860-867, 2007.

Thall PF, Wathen JK. Bayesian methods to account for patient heterogeneity in phase II clinical trials. *Current Opinion in Oncology*, 20:407-411, 2008.

Thall PF, Wooten LH, Logothetis CJ, Millikan R, Tannir NM. Bayesian and frequentist two-stage treatment strategies based on sequential failure times subject to interval censoring. *Statistics in Medicine,* 26:4687-4702, 2007.

Thall PF, Nguyen HQ. Adaptive randomization to improve utility-based dose-finding with bivariate ordinal outcomes. *J Biopharmaceutical Statistics,* 22:785-801, 2012.

Thall PF, Sung H-G, Estey EH. Selecting therapeutic strategies based on efficacy and death in multi-course clinical trials. *J American Statistical Assoc,* 97:29-39, 2002.

Wang L, Rotnitzky A, Lin X, Millikan R, Thall PF. Evaluation of viable dynamic treatment regimes in a sequentially randomized trial of advanced prostate cancer. *J American Statistical Assoc,* 107:493-508, (with discussion, pages 509-517; rejoinder, pages 518-520), 2012.