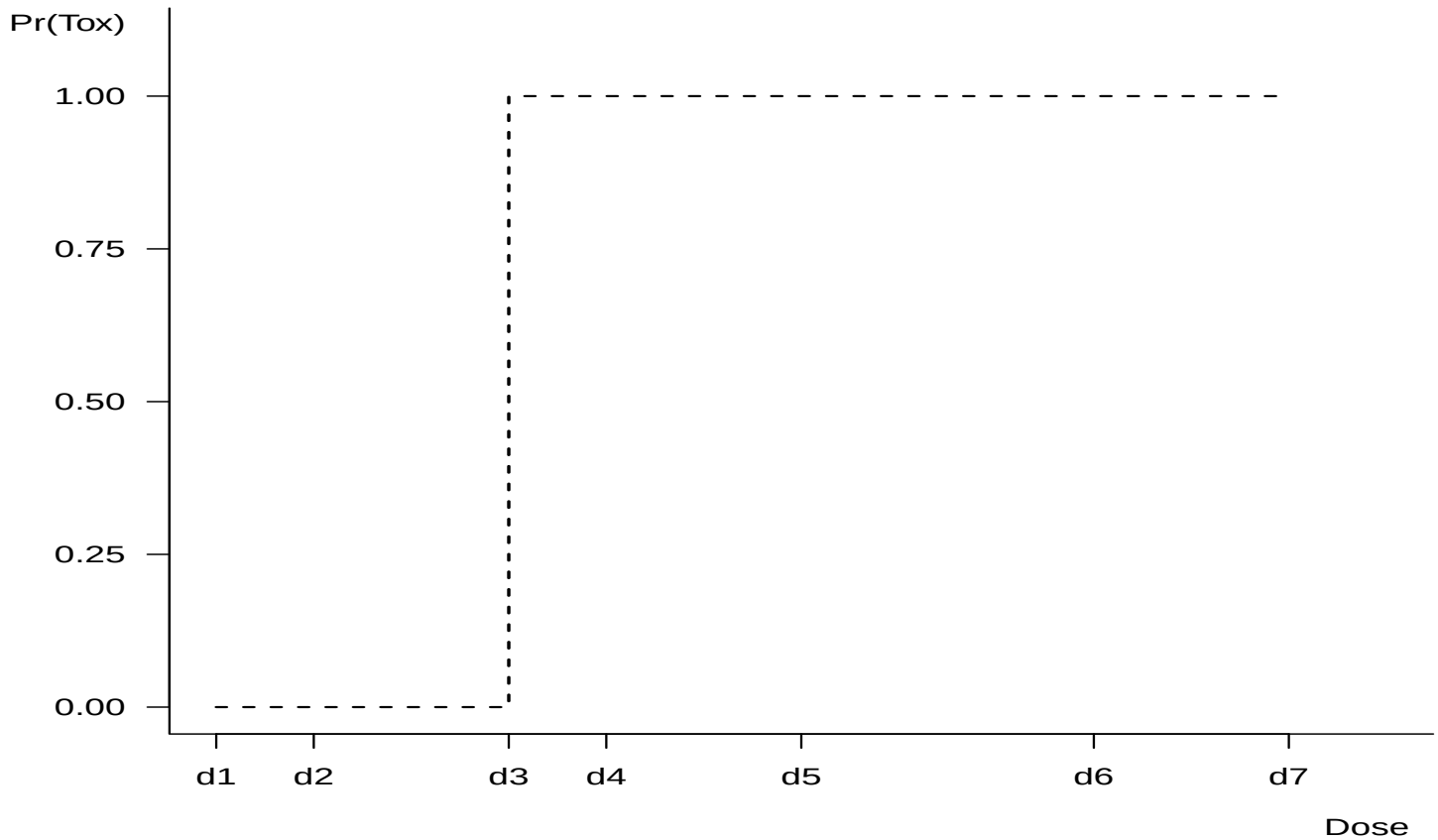
Dose finding methodology

John O'Quigley

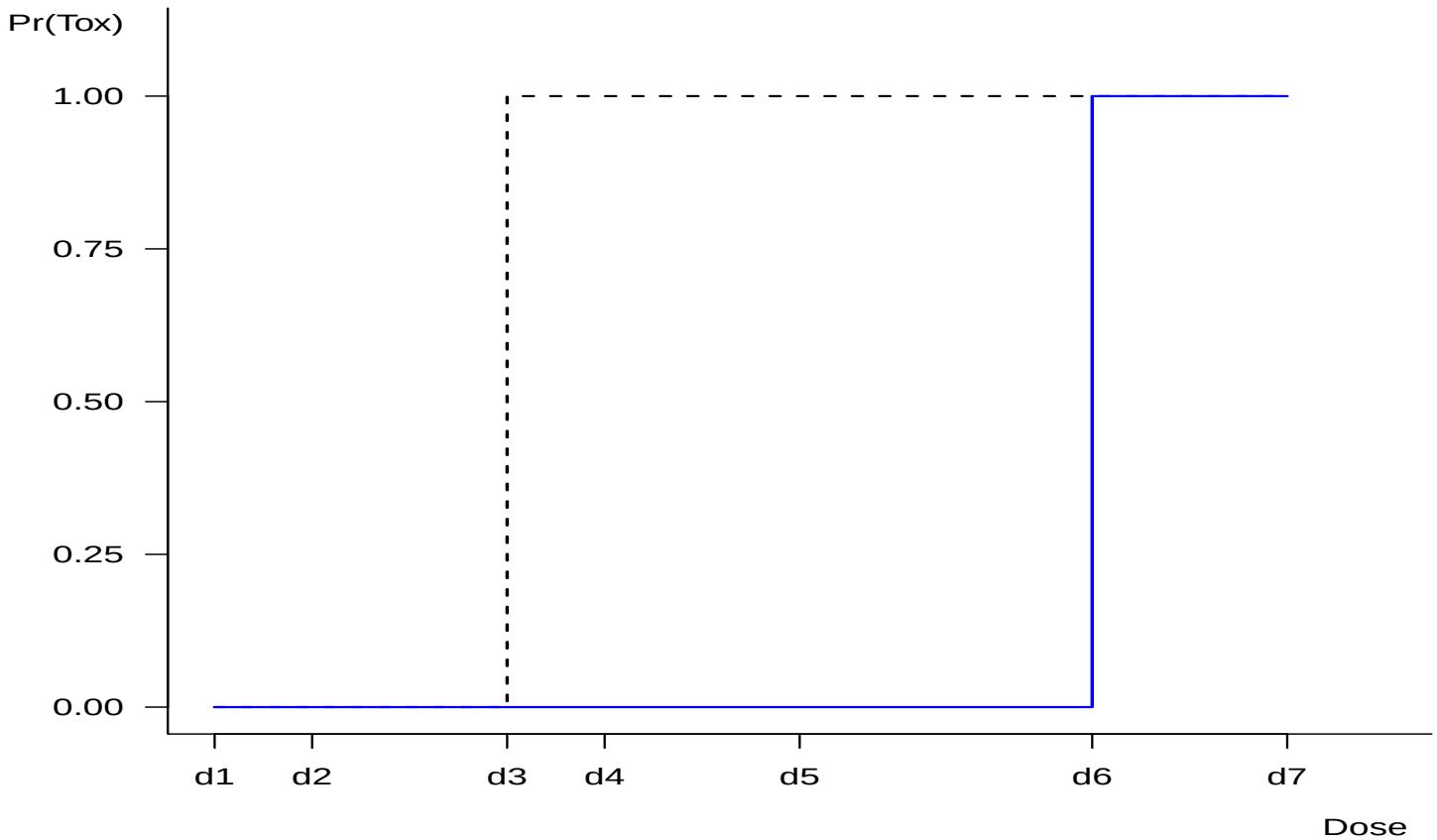
Laboratoire de Statistique Théorique et Appliquée

Université Paris VI

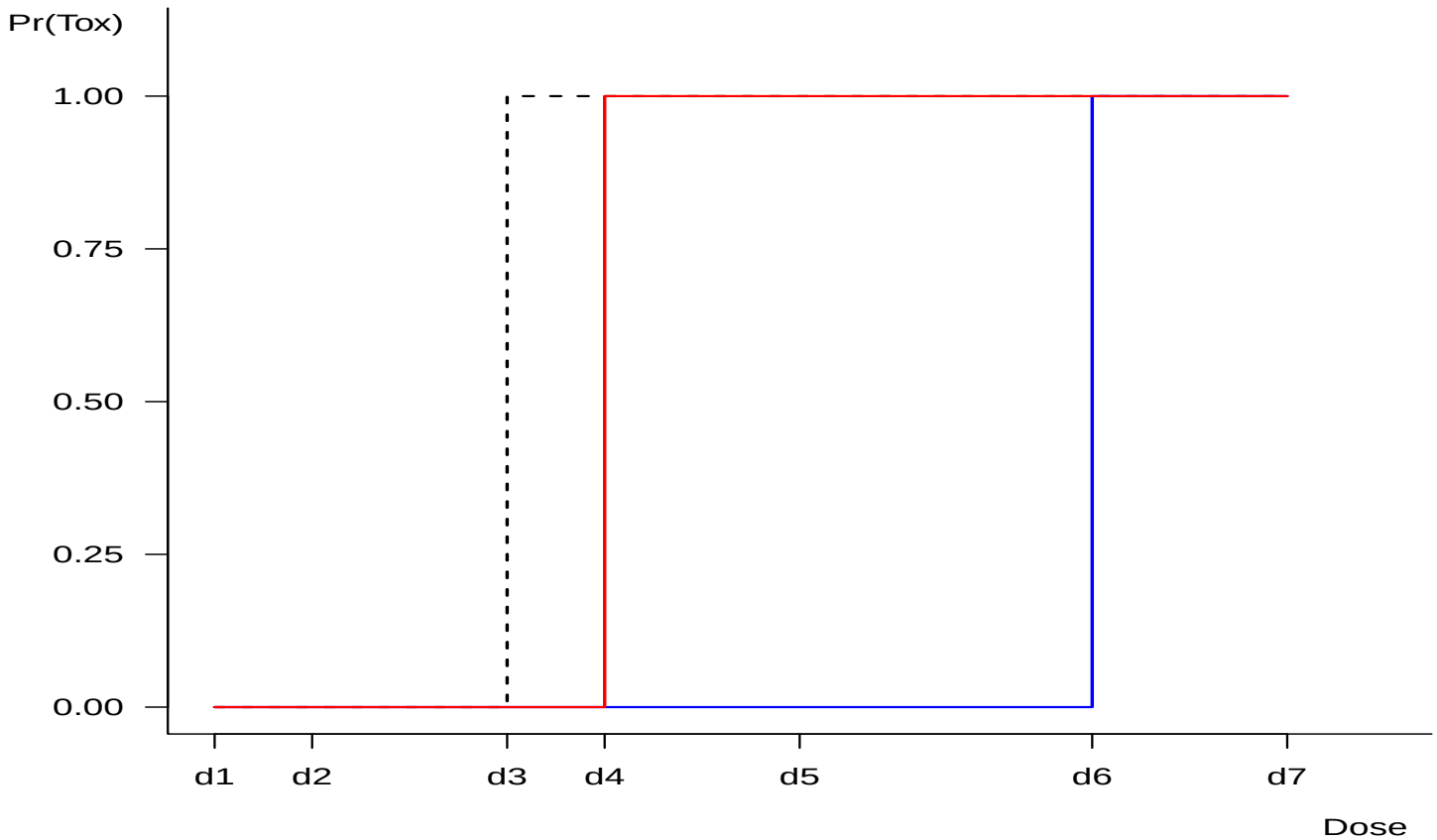
Model for cytotoxics (3 patients)



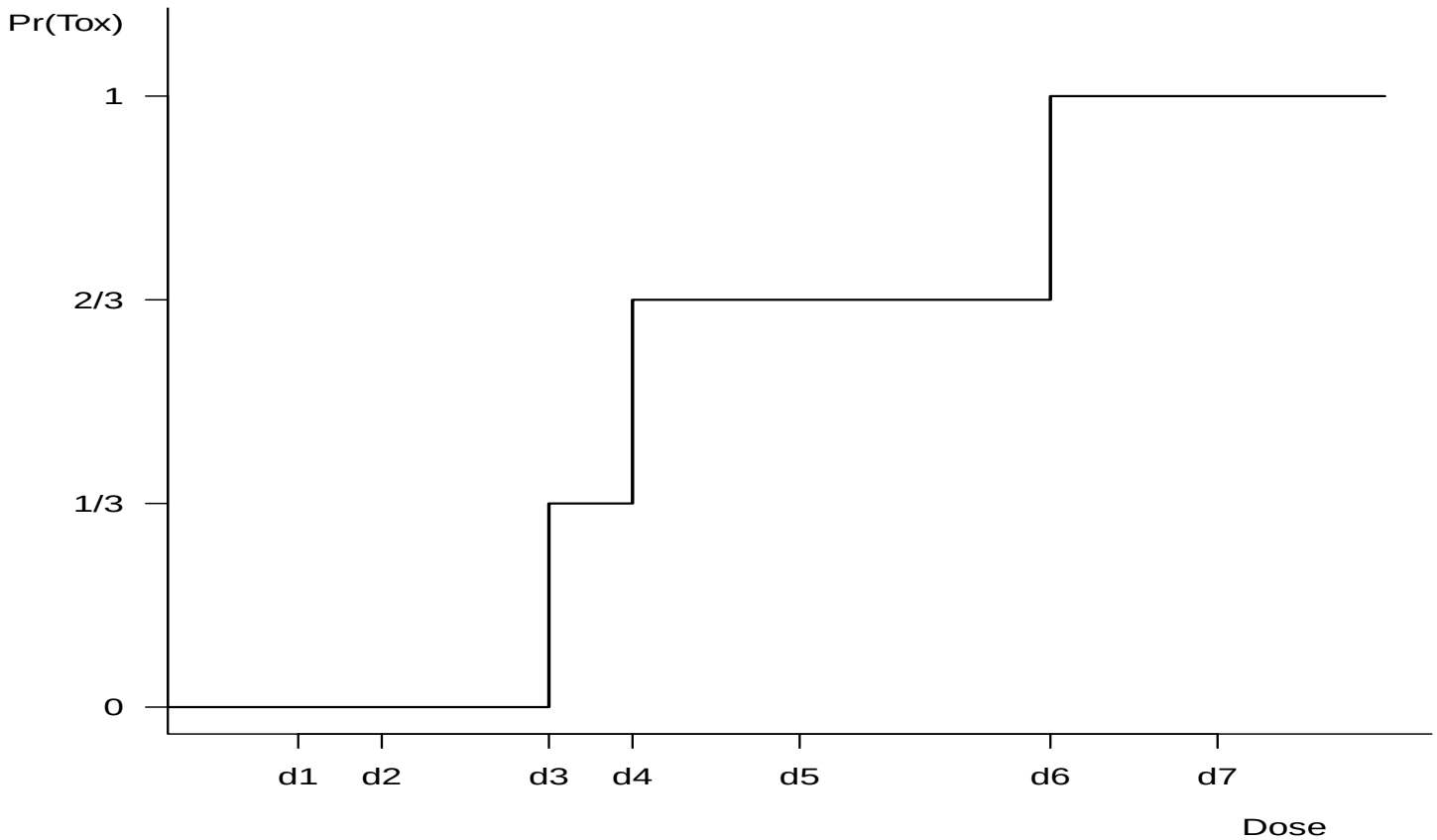
Model for cytotoxics (3 patients)



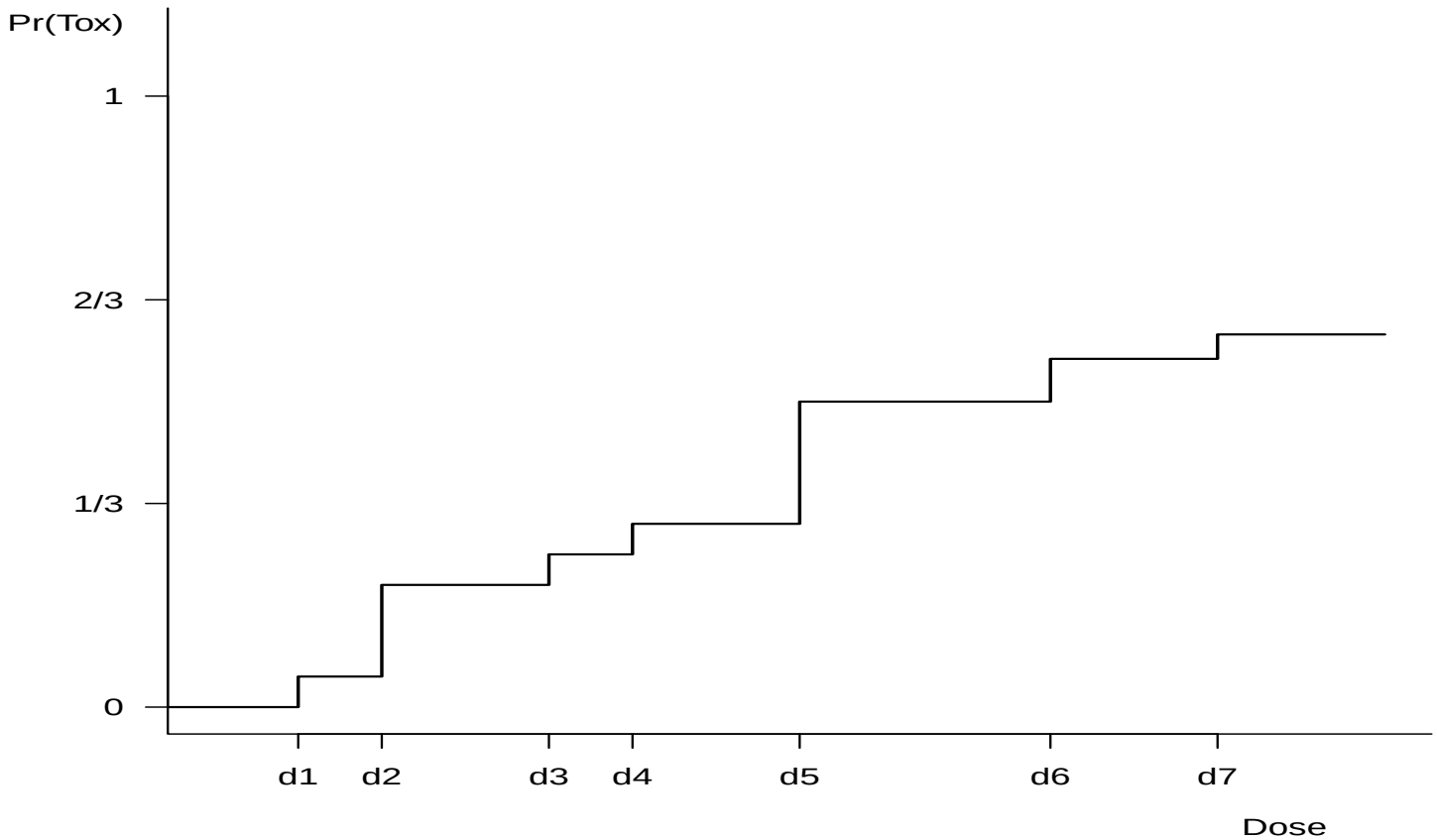
Model for cytotoxics (3 patients)



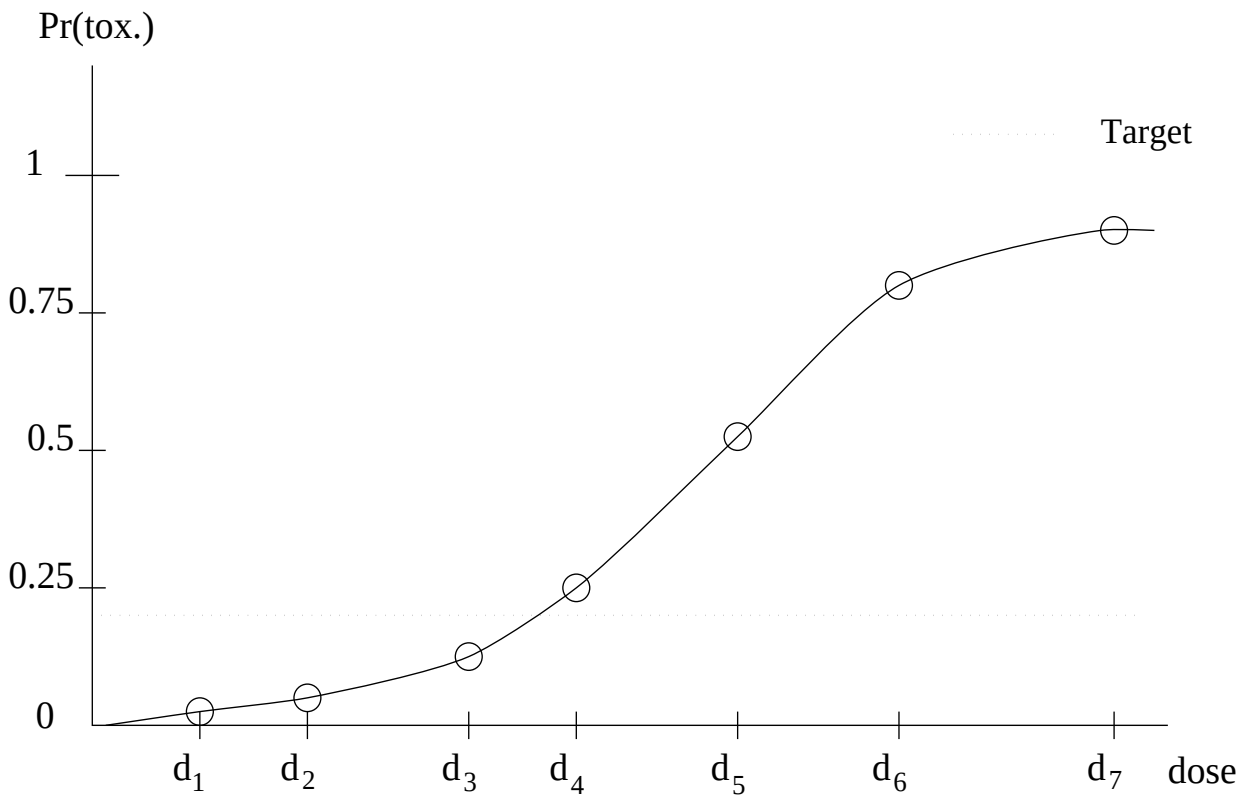
Empirical distribution for 3 patients



Model for cytotoxics (population)



Model for cytotoxics (population)



Ethical considerations for a Phase I trial

1. We **must** do the best for the treated patient. We cannot knowingly **undertreat** leaving no chance for therapeutic benefit. (Smith et al (1998) *J. Clin. Oncology*). We can not knowingly **overtreat**.
2. There is no "treatment versus experimentation dilemma" (Azriel et al 2011) .
3. There is no "future benefit", "current patient benefit" conflict.
4. We must abide by Helsinki Declaration

Ethical considerations for a Phase I trial

1. We **must** do the best for the treated patient. We cannot knowingly **undertreat** leaving no chance for therapeutic benefit. (Smith et al (1998) *J. Clin. Oncology*). We can not knowingly **overtreat**.
2. There is no "treatment versus experimentation dilemma" (Azriel et al 2011) .
3. There is no "future benefit", "current patient benefit" conflict.
4. We must abide by Helsinki Declaration

Ethical considerations for a Phase I trial

1. We **must** do the best for the treated patient. We cannot knowingly **undertreat** leaving no chance for therapeutic benefit. (Smith et al (1998) *J. Clin. Oncology*). We can not knowingly **overtreat**.
2. There is no "treatment versus experimentation dilemma" (Azriel et al 2011) .
3. There is no "future benefit", "current patient benefit" conflict.
4. We must abide by Helsinki Declaration

Ethical considerations for a Phase I trial

1. We **must** do the best for the treated patient. We cannot knowingly **undertreat** leaving no chance for therapeutic benefit. (Smith et al (1998) *J. Clin. Oncology*). We can not knowingly **overtreat**.
2. There is no "treatment versus experimentation dilemma" (Azriel et al 2011) .
3. There is no "future benefit", "current patient benefit" conflict.
4. We must abide by Helsinki Declaration

Ethical considerations for a Phase I trial

1. We do not want to “undertreat”
2. We do not want to “overtreat”, i.e. too much toxicity.
3. Use as few patients as possible (efficiency).

Ethical considerations for a Phase I trial

1. We do not want to “undertreat”
2. We do not want to “overtreat”, i.e. too much toxicity.
3. Use as few patients as possible (efficiency).

Ethical considerations for a Phase I trial

1. We do not want to “undertreat”
2. We do not want to “overtreat”, i.e. too much toxicity.
3. Use as few patients as possible (efficiency).

Up and Down Designs (Storer *Biometrics* (1989, 1993)

1. Random walk (no memory)
2. Decision rule uses only part of data.
3. Standard design is 3+3 design + stopping rule.
4. Fails all 3 ethical criteria;
    1. More patients under-treated than necessary.
    2. More patients over-treated than necessary.
    3. Poor (inefficient) estimate of MTD.

Up and Down Designs (Storer *Biometrics* (1989, 1993)

1. Random walk (no memory)
2. Decision rule uses only part of data.
3. Standard design is 3+3 design + stopping rule.
4. Fails all 3 ethical criteria;
    1. More patients under-treated than necessary.
    2. More patients over-treated than necessary.
    3. Poor (inefficient) estimate of MTD.

Up and Down Designs (Storer *Biometrics* (1989, 1993)

1. Random walk (no memory)
2. Decision rule uses only part of data.
3. Standard design is 3+3 design + stopping rule.
4. Fails all 3 ethical criteria;
    1. More patients under-treated than necessary.
    2. More patients over-treated than necessary.
    3. Poor (inefficient) estimate of MTD.

Up and Down Designs (Storer *Biometrics* (1989, 1993)

1. Random walk (no memory)
2. Decision rule uses only part of data.
3. Standard design is 3+3 design + stopping rule.
4. Fails all 3 ethical criteria;
    1. More patients under-treated than necessary.
    2. More patients over-treated than necessary.
    3. Poor (inefficient) estimate of MTD.

Continual Reassessment Method

Gasparini and Eisele (A curve-free method for Phase I clinical trials. *Biometrics* 2000) described CRM as:

1. An allocation rule to assign sequentially the incoming patients to one of the possible doses, with the intent of assigning doses ever closer to, and eventually recommending, the MTD.
2. A statistical procedure that updates the information on the probabilities of toxicity in light of the results obtained for the patients already observed

Same idea for MCRM (Faries 1994), GCRM (Goodman 1995, Heyd and Carlin 1998), RCRM, ECRM (Moller 1995).

Continual Reassessment Method

1. Select target θ (usually 1/4, 1/5 or 1/3).
2. $\Pr(Y_i = 1|d_j) = \psi(d_j, a) = \alpha_j^{\exp(a)}$
3. Calculate

$$\log L(a) = \sum_1 y_i \log \psi(x_i, a) + \sum_0 (1 - y_i) \log[1 - \psi(x_i, a)]$$

4. Allocate to dose $x_i \in \{d_1, ..., d_k\}$ where;

$$|\psi(x_i, \hat{a}) - \theta| \leq |\psi(d_j, \hat{a}) - \theta| \quad \forall d_j$$

Continual Reassessment Method

1. Select target θ (usually 1/4, 1/5 or 1/3).
2. $\Pr(Y_i = 1|d_j) = \psi(d_j, a) = \alpha_j^{\exp(a)}$
3. Calculate

$$\log L(a) = \sum_1 y_i \log \psi(x_i, a) + \sum_0 (1 - y_i) \log[1 - \psi(x_i, a)]$$

4. Allocate to dose $x_i \in \{d_1, ..., d_k\}$ where;

$$|\psi(x_i, \hat{a}) - \theta| \leq |\psi(d_j, \hat{a}) - \theta| \quad \forall d_j$$

Continual Reassessment Method

1. Select target θ (usually 1/4, 1/5 or 1/3).
2. $\Pr(Y_i = 1|d_j) = \psi(d_j, a) = \alpha_j^{\exp(a)}$
3. Calculate

$$\log L(a) = \sum_1 y_i \log \psi(x_i, a) + \sum_0 (1 - y_i) \log[1 - \psi(x_i, a)]$$

4. Allocate to dose $x_i \in \{d_1, ..., d_k\}$ where;

$$|\psi(x_i, \hat{a}) - \theta| \leq |\psi(d_j, \hat{a}) - \theta| \quad \forall d_j$$

Continual Reassessment Method

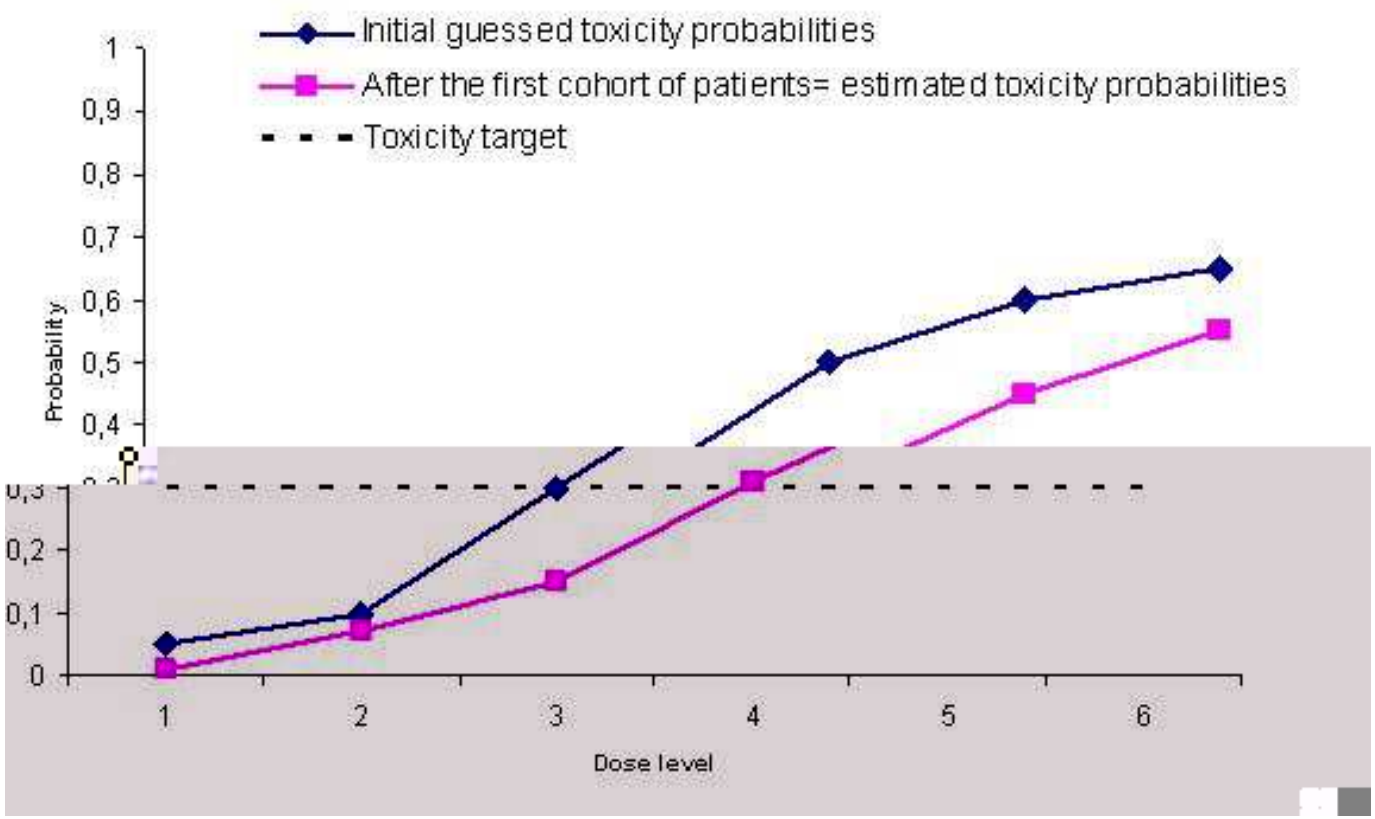
1. Select target θ (usually 1/4, 1/5 or 1/3).
2. $\Pr(Y_i = 1|d_j) = \psi(d_j, a) = \alpha_j^{\exp(a)}$
3. Calculate

$$\log L(a) = \sum_1 y_i \log \psi(x_i, a) + \sum_0 (1 - y_i) \log[1 - \psi(x_i, a)]$$

4. Allocate to dose $x_i \in \{d_1, ..., d_k\}$ where;

$$|\psi(x_i, \hat{a}) - \theta| \leq |\psi(d_j, \hat{a}) - \theta| \quad \forall d_j$$

Continual Reassessment Method

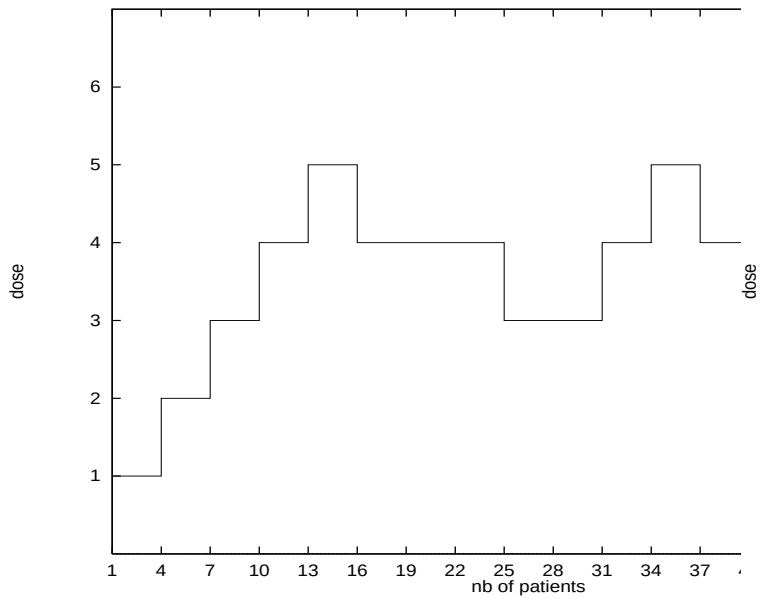


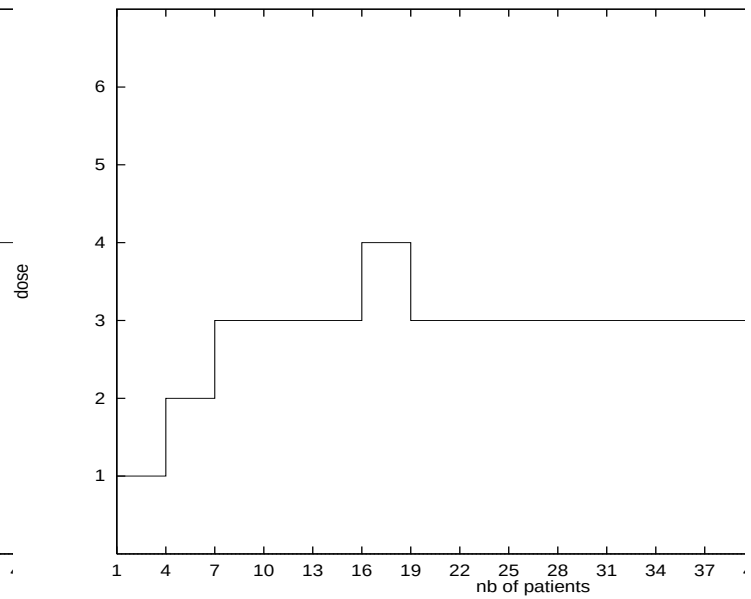
Bayesian and likelihood estimation

1	-0.69	d_1	0.2	0	–	d_1	–	0
2	-0.27	d_2	0.13	0	–	d_2	–	0
3	-0.03	d_3	0.21	0	–	d_3	–	0
4	0.23	d_3	0.13	0	–	d_4	–	0
5	0.40	d_4	0.21	0	–	d_5	–	1
6	0.61	d_4	0.14	1	0.34	d_4	0.23	0
7	0.08	d_3	0.18	0	0.52	d_4	0.17	0
8	0.17	d_3	0.15	0	0.64	d_4	0.14	0
9	0.25	d_4	0.26	0	0.74	d_5	0.29	0
10	0.35	d_4	0.23	0	0.87	d_5	0.24	1
11	0.42	d_4	0.20	0	0.58	d_4	0.15	0
12	0.50	d_4	0.18	1	0.65	d_4	0.13	1
13	0.26	d_4	0.26	0	0.35	d_4	0.23	0
14	0.32	d_4	0.23	0	0.42	d_4	0.20	0
15	0.37	d_4	0.22	0	0.47	d_4	0.19	0
16	0.42	d_4	0.20	0	0.52	d_4	0.17	0
17	0.47	$\boxed{d_4}$	0.19		0.56	$\boxed{d_4}$	0.16	

Behaviour of SM (no stopping rule) and CRM

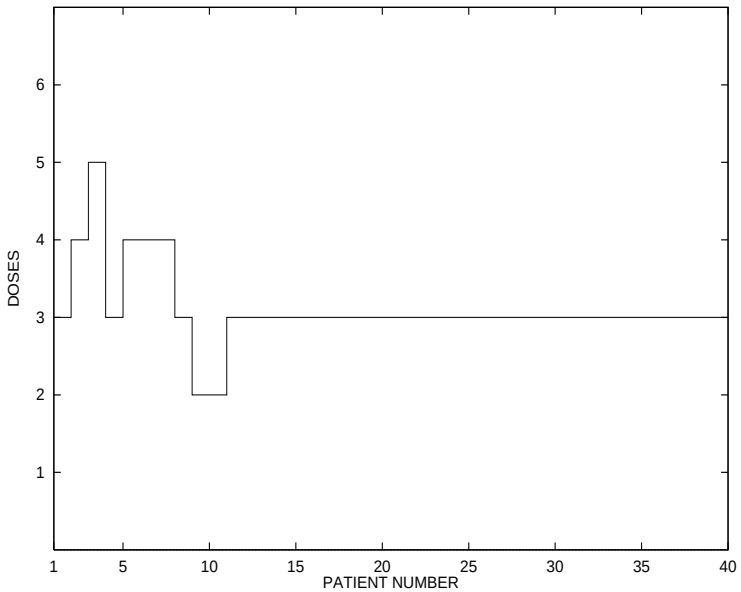
unknown probabilities at level i						
R_i	.04	.11	.23	.34	.42	.61



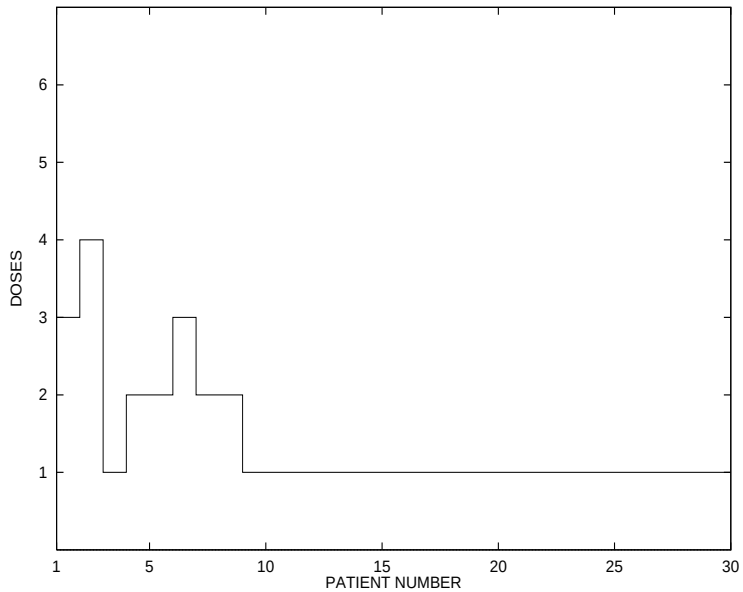


CRM examples, no stopping rule

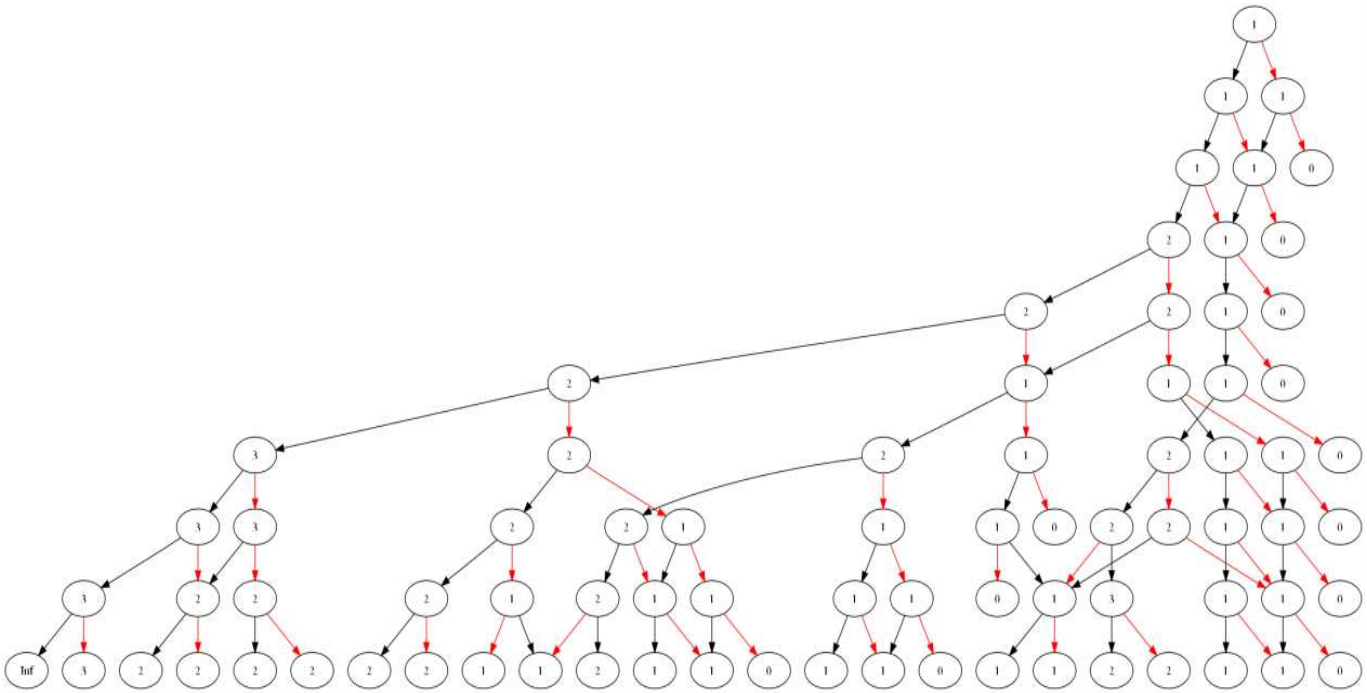
MTD = level 3



MTD = level 1



Potential sample paths



Model and inference (likelihood)

Using likelihood and letting

$$\Pr(Y_i = 1 | X_i = d_j) = (\alpha_j)^a$$

then the models

α_i	.81	.85	.89	.92	.95	.98
α_i	.01	.03	.09	.16	.35	.59

behave identically,

whereas the models

α_i	.05	.10	.20	.30	.50	.70
α_i	.05	.10	.20	.30	.40	.50

behave differently.

Model and inference (Bayes)

For distance measure use;

1. O'Quigley, Pepe, Fisher (1990) suggest $E\,\psi(d_j, a)$
2. O'Quigley, Pepe, Fisher (1990) suggest $\psi(d_j, E(a))$
3. Chu, Lin, Shih (2009) suggest $\psi_{1-\gamma}(d_j, a)$
4. Shih (1999) suggest $\gamma = 0.5$ corresponding to median.
5. Babb, Rogatko, Zacks (1998) suggest $\gamma = 0.75$
   This is known as EWOC.

1-parameter versus 2-parameter models

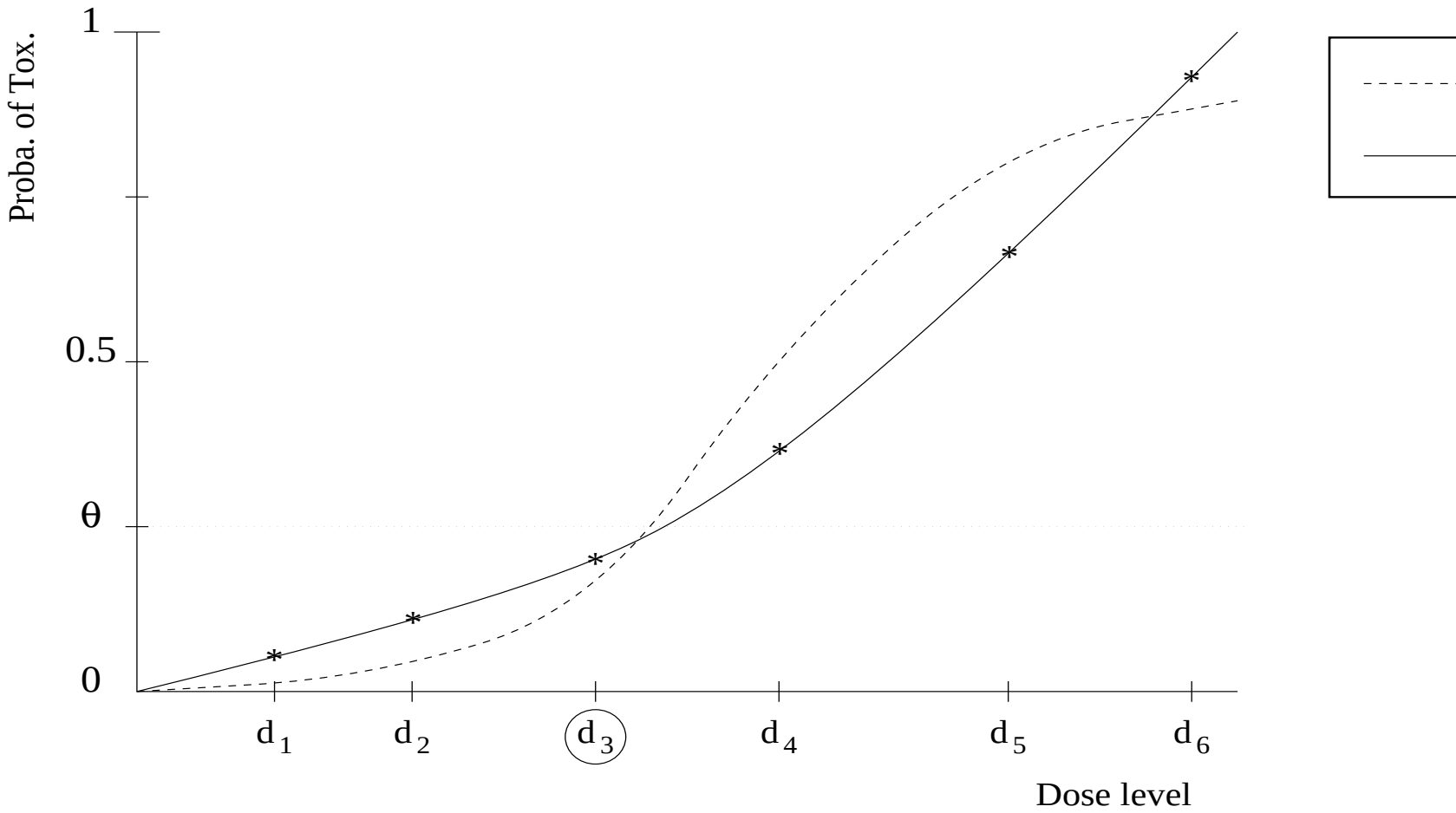
O'Quigley, Pepe and Fisher (1990) show that;

1. 2-parameter logistic model more noisy
2. Final recommendations less accurate

Table: 2-param logistic (O'Quigley, Pepe, Fisher 1990)

	Dose					
	1	2	3	4	5	6
$R(d_i)$	.06	.08	.12	.18	.40	.71
1-CRM	.00	.04	.23	.57	.15	.00
2-CRM	.01	.11	.16	.48	.19	.05

One parameter CRM models



Two parameter CRM (ADEPT, BLR)

- Two parameter CRM has weaker theoretical foundation
- 2CRM can be erratic, eg., first patient treated at level 1, suffers DLT, the recommendation is treat at level 6 (Shu 2008).
- ADEPT is 2CRM, using patient benefit as metric.
- BLR (Neuenschwander et al 2007) is also 2CRM

Two versus one parameter CRM models

1. $\hat{R}_j = \psi(d_j, \hat{a})$ may be too inflexible to work well for all j.
2. $\hat{R}_j \approx \sum Y_{ij}/n_j$ at recommended level.
3. $\hat{R}_j \xrightarrow{P} R_j$ and is fully efficient (Shen & O'Quigley, *Biometrika 96*)
4. $R_j = \psi(d_j, a, b)$ is over-parameterized, cannot identify a and b.

Simulations: Gerke and Siedentop (2008)

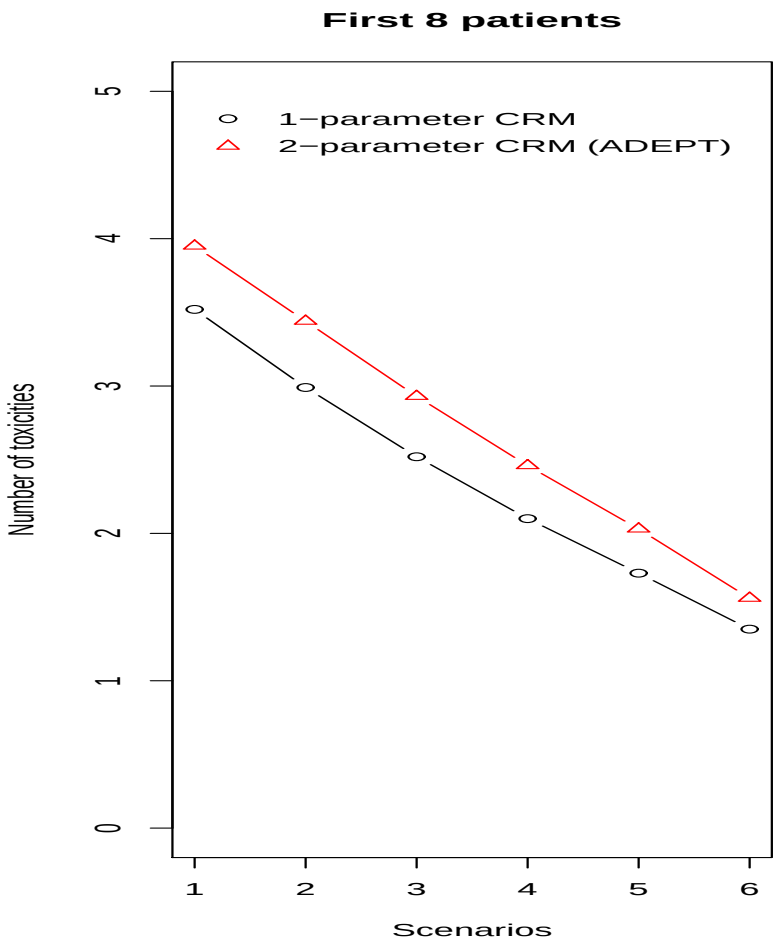
Table: Percentage of found MTD for the 3 scenarios from Gerke and Siedentop using fixed sample size. 3 scenarios, n=23, 18, 17.

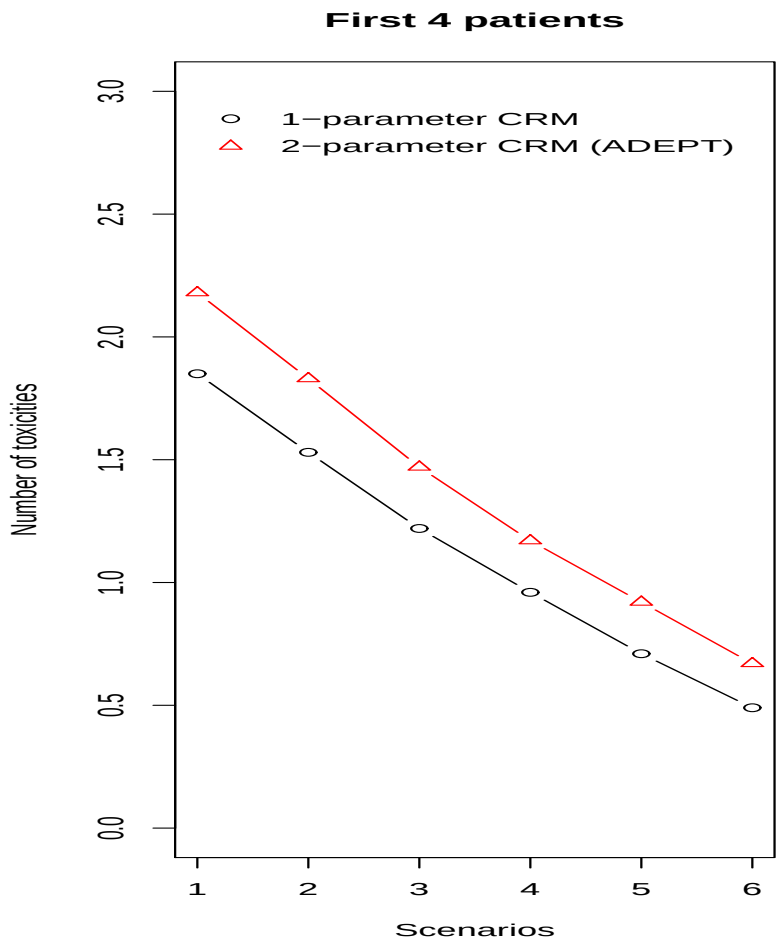
			Dose						
1	2	3	4	5	6	7	8	9	10
4.85	9.90	**34.95**	38.65	11.05	0.60	0.00	0.00	0.00	0.00
1.05	12.30	**39.50**	38.35	8.60	0.20	0.00	0.00	0.00	0.00
0.10	0.20	1.90	8.50	**38.25**	46.75	4.20	0.10	0.00	0.00
0.00	0.00	1.45	11.65	**42.90**	37.10	6.85	0.05	0.00	0.00
0.00	0.00	0.00	0.20	0.40	5.65	**44.35**	47.65	1.75	0.00
0.00	0.00	0.00	0.05	0.70	12.30	**43.45**	35.55	7.80	0.15

Table: The toxicity rate of six simulated scenarios

Scenario	Dose					
	1	2	3	4	5	6
1	**0.35**	0.45	0.55	0.70	0.80	0.95
2	0.25	**0.35**	0.45	0.55	0.70	0.80
3	0.15	0.25	**0.35**	0.45	0.55	0.70
4	0.10	0.15	0.25	**0.35**	0.45	0.55
5	0.05	0.10	0.15	0.25	**0.35**	0.45
6	0.02	0.05	0.10	0.15	0.25	**0.35**

One/two parameter CRM models: first patients





Novartis case study (Neuenschwander et al, Bailey and Neuenschwander 2008)

doses	1.0	2.5	5	10	15	20	25	30	40	50
# pats	3	4	5	4	0	0	2	-	-	-
# DLTs	0	0	0	0	-	-	2	-	-	-

Logical errors in Novartis case study

doses	1.0	2.5	5	10	15	20	25	30	40	50
E(prior)	.07	.08	.09	.11	.12	.14	.16	.24	**.33**	.46
E(post)	.02	.05	.09	.13	.18	.23	.28	**.34**	.41	.47

- CRM recommends **decrease** in 3 levels and **not** an increase.
- CRM is coherent (Cheung 2003).
- Bayesian methods require care.

Impact of skeleton/prior when using Bayes

doses	1.0	2.5	5	10	15	20	25	30	40	50
E(prior)	.07	.08	.09	.11	.12	.14	.16	.24	**.33**	.46
E(post)	.02	.05	.09	.13	.18	.23	.28	**.34**	.41	.47

doses	1.0	2.5	5	10	15	20	25	30	40	50
E(prior)	.00	.00	.00	.02	.12	**.30**	.50	.68	.80	.88
E(post)	.00	.00	.01	.05	.18	**.38**	.57	.73	.84	.90

doses	1.0	2.5	5	10	15	20	25	30	40	50
E(prior)	.00	.02	.12	**.30**	.50	.68	.80	.88	.93	.96
E(post)	.00	.00	.01	.08	**.23**	.44	.62	.76	.86	.92

Curve free designs (Gasparini, Eisele 2000)

1. $\theta_1 = 1 - R(d_1)$,

$$\theta_i = \frac{1 - R(d_i)}{1 - R(d_{i-1})}, \; i = 2, \ldots, k.$$

2. For each θ_i, $(i = 1, \ldots, k)$,

$$f(\theta_i) = B^{-1}(a_i, b_i)\theta_i^{a_i - 1}(1 - \theta_i)^{b_i - 1}$$

for parameters a_i and b_i and where $B(a, b)$ is the beta function. with parameters a and b.

3. $R(d_i) = 1 - \theta_1\theta_2 ... \theta_i$

4. O'Quigley (*Biometrics* 2005) shows Curve free $\equiv$ CRM.

EWOC designs (Babb, Rogatko, Zacks 1998)

1. Iterative updating same as CRM
2. Allocate to dose level d_j such that posterior probability of toxicity being greater than θ is α. BRZ choose $\alpha = 0.25$
3. Chu, Lin and Shih (2009) show that, when $\alpha = 0.5$, then EWOC $\equiv$ CRM.

Simple take home message

CRM, BLR, ADEPT EWOC, Curve-free

are all essentially equivalent.

How good can any design be?

Super-optimal designs:

1. Include zero patients in study: recommend level 2.
2. Include 5 patients at level 3. Recommend according to table:

Outcome	0/5	1/5	2/5	3/5	4/5	5/5
Recommendation	5	4	3	2	1	1

Super optimality

Super-optimality is common in the statistical literature, in particular for Bayesian designs.

Example for combinations, using partial orderings;

1. Yin and Yuan (2009) *Appl. Statist*, 211 - 224, show for 4×4 combinations, copula design finds MTD 52%.
2. PO-CRM (Wages et al, *Biometrics* 2011) finds MTD in 45%.
3. When ordering is known, CRM finds MTD 48%.
4. When ordering is known Optimal Design finds MTD 49%.

Optimal design benchmark

- Subject h experiences a toxicity at d_5.
- Subject j a non-toxicity at level d_3.

Doses	d_1	d_2	d_3	d_4	d_5	d_6
Observed Y_{hk}	X	X	X	X	1	1
Unobserved Y_{hk}	0	0	1	1	1	1
Observed Y_{jk}	0	0	0	X	X	X
Unobserved Y_{jk}	0	0	0	0	0	1

Consider;

Dose	d_1	d_2	d_3	d_4	d_5	d_6
$R_k = \Pr(Y_k = 1)$	0.05	0.11	0.22	0.35	0.45	0.60

Subject j	v_j	s_j	Toxicity at dose level					
			1	2	3	4	5	6
1	.53	6	0	0	0	0	0	1
2	.08	2	0	1	1	1	1	1
3	.29	4	0	0	0	1	1	1
4	.41	5	0	0	0	0	1	1
5	.79	-	0	0	0	0	0	0
6	.04	1	1	1	1	1	1	1
7	.87	-	0	0	0	0	0	0
8	.15	3	0	0	1	1	1	1
9	.63	-	0	0	0	0	0	0
10	.56	6	0	0	0	0	0	1
11	.32	4	0	0	0	1	1	1
12	.72	-	0	0	0	0	0	0
13	.20	3	0	0	1	1	1	1
14	.97	-	0	0	0	0	0	0
15	.52	6	0	0	0	0	0	1
16	.24	4	0	0	0	1	1	1
Frequencies		$\hat{R}_k$	0.06	0.13	0.25	0.44	0.50	0.69
		R_k	0.05	0.11	0.22	0.35	0.45	0.60

Summarizing results

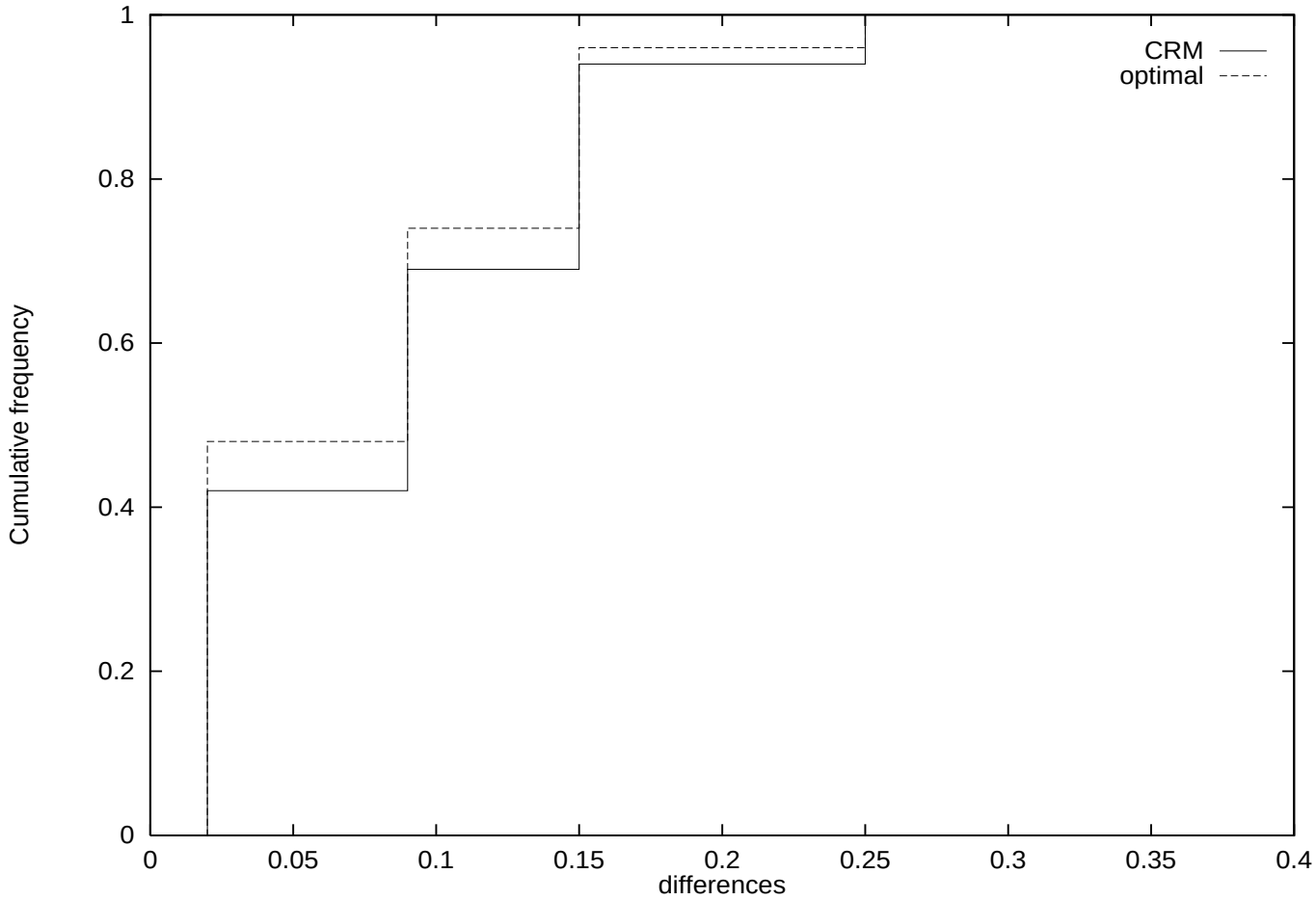
Relative performance by levels;

d_k	1	2	3	4	5	6
R_k	0.05	0.11	0.22	0.35	0.45	0.60
$p_k(16)$	0.05	0.26	0.42	0.21	0.06	0.0
$q_k(16)$	0.04	0.27	0.48	0.17	0.04	0.0

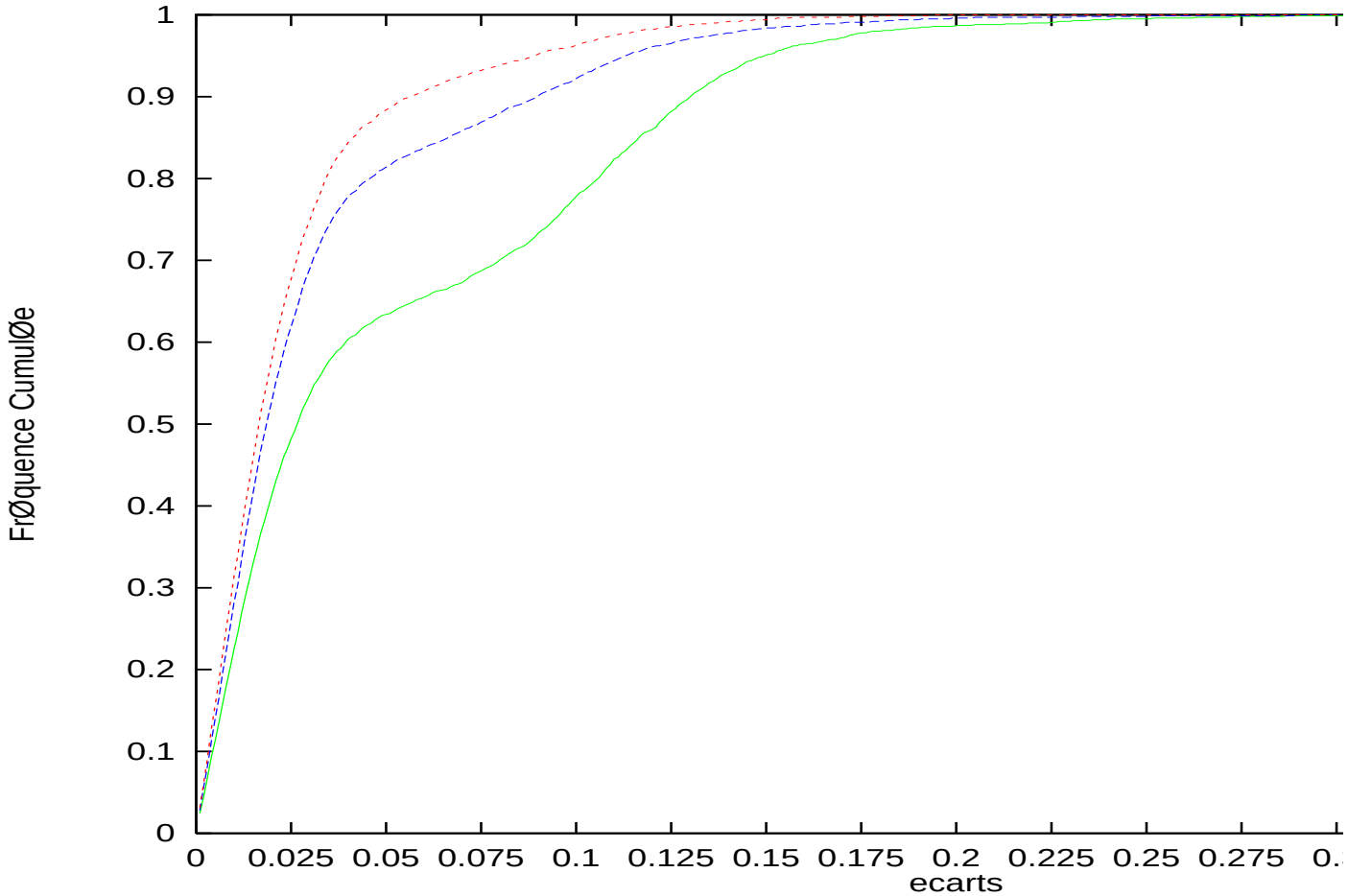
Relative performance by cumulative errors; Let $\alpha = 0.1$ be % simulations where $\Pr(Y = 1) \in (0.10, 0.30)$. This is 0.69 for CRM and 0.74 for optimal.

α	0.02	0.05	0.10	0.15	0.20
p_α	0.42	0.42	0.69	0.94	1.0
q_α	0.48	0.48	0.74	0.96	1.0

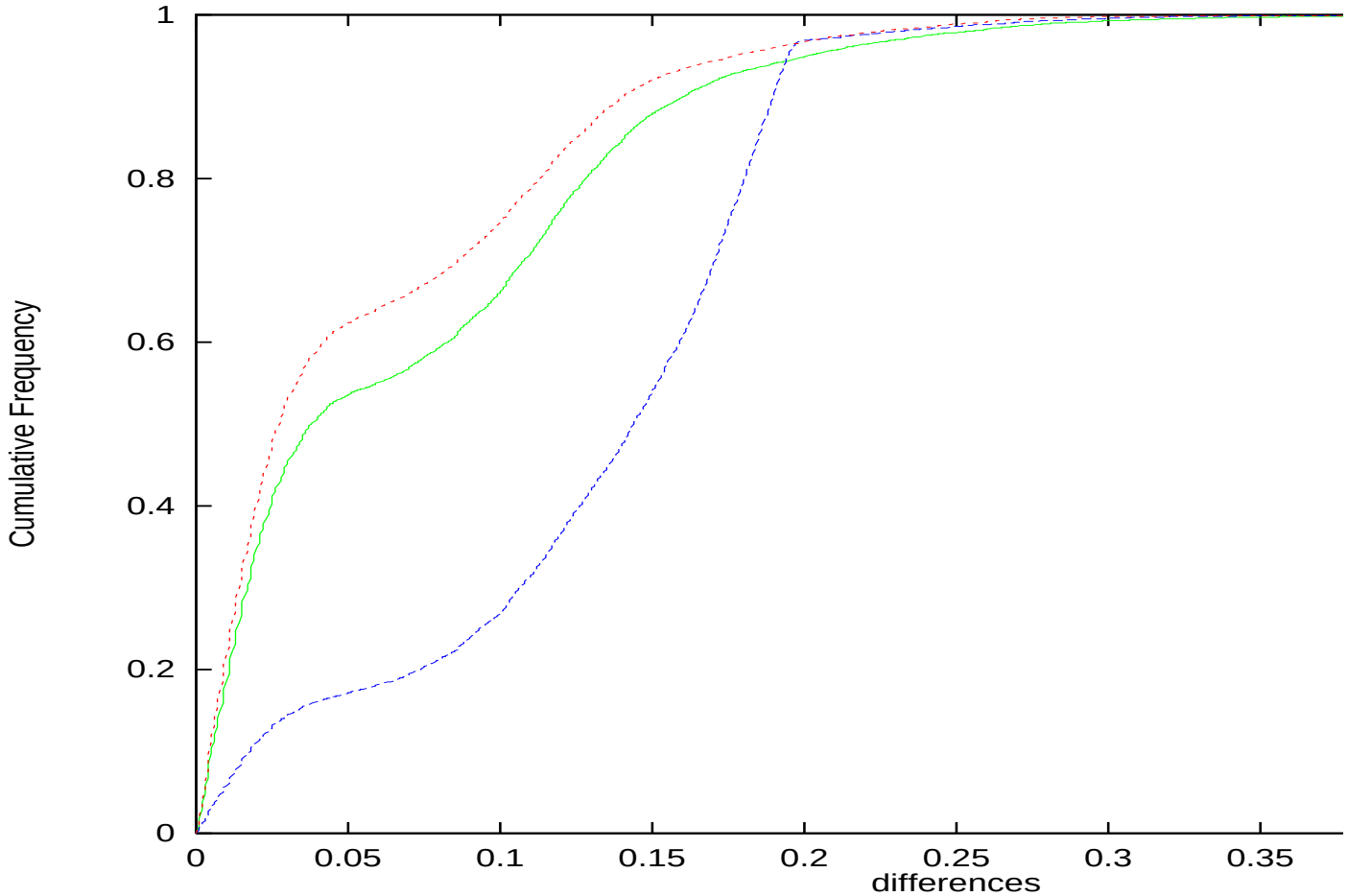
Graph of cumulative errors



Optimal .... , CRM1 .... , CRM2 ....



Optimal .... , CRM .... , 3+3 ....



Two stage designs (likelihood)

1. Likelihood is monotone unbounded until first observed toxicity.
2. First stage is largely arbitrary.
3. Different first stage algorithms lead to different operating characteristics.
4. First stage can incorporate information on grades.
5. Two stage designs accommodate many/open number of levels.

Two stage designs (likelihood)

1. Likelihood is monotone unbounded until first observed toxicity.
2. First stage is largely arbitrary.
3. Different first stage algorithms lead to different operating characteristics.
4. First stage can incorporate information on grades.
5. Two stage designs accommodate many/open number of levels.

Two stage designs (likelihood)

1. Likelihood is monotone unbounded until first observed toxicity.
2. First stage is largely arbitrary.
3. Different first stage algorithms lead to different operating characteristics.
4. First stage can incorporate information on grades.
5. Two stage designs accommodate many/open number of levels.

Two stage designs (likelihood)

1. Likelihood is monotone unbounded until first observed toxicity.
2. First stage is largely arbitrary.
3. Different first stage algorithms lead to different operating characteristics.
4. First stage can incorporate information on grades.
5. Two stage designs accommodate many/open number of levels.

Two stage designs (likelihood)

1. Likelihood is monotone unbounded until first observed toxicity.
2. First stage is largely arbitrary.
3. Different first stage algorithms lead to different operating characteristics.
4. First stage can incorporate information on grades.
5. Two stage designs accommodate many/open number of levels.

Design modifications

1. Patients can be included in groups, eg 3 at a time.
2. Grouping can be by design.
3. Overdose control.
4. Underdose control.
5. Joint underdose/overdose control.

Design modifications

1. Patients can be included in groups, eg 3 at a time.
2. Grouping can be by design.
3. Overdose control.
4. Underdose control.
5. Joint underdose/overdose control.

Design modifications

1. Patients can be included in groups, eg 3 at a time.
2. Grouping can be by design.
3. Overdose control.
4. Underdose control.
5. Joint underdose/overdose control.

Design modifications

1. Patients can be included in groups, eg 3 at a time.
2. Grouping can be by design.
3. Overdose control.
4. Underdose control.
5. Joint underdose/overdose control.

Design modifications

1. Patients can be included in groups, eg 3 at a time.
2. Grouping can be by design.
3. Overdose control.
4. Underdose control.
5. Joint underdose/overdose control.

Some initial escalation schemes

# pats	1	2	3	4	5	6	7	8	9	etc.
3+3	d_1	d_1	d_1	d_2	d_2	d_2	d_3	d_3	d_3	etc.
CRM(2S)	d_1	d_2	d_2	d_3	d_3	d_3	d_4	d_4	d_4	etc.
CRM(G)	d_1	d_2	d_3	d_3	d_4	d_4	d_5	d_5	d_5	etc.

Table: Example of initial escalation stage using acceleration.

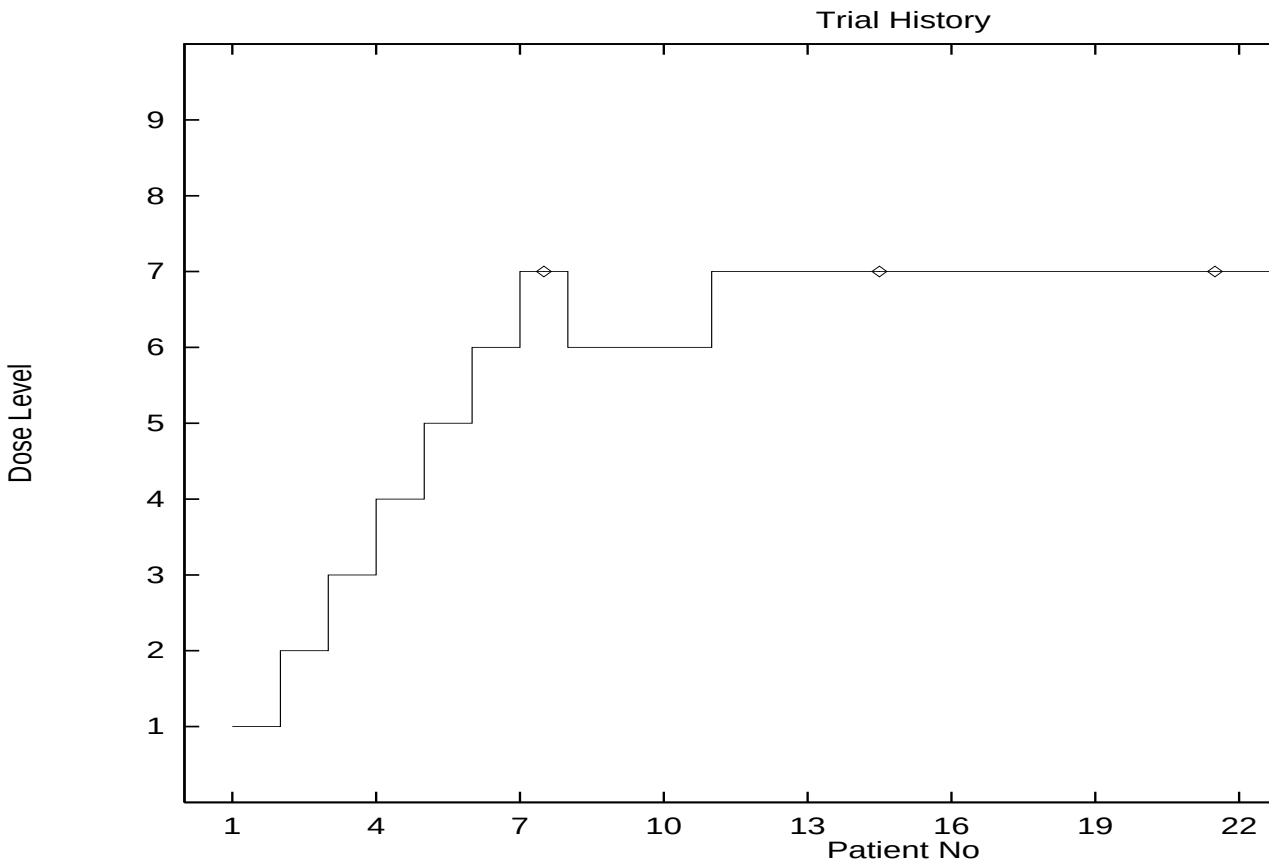
Rapid early escalation using grades

Severity	Degree of Toxicity
0	No toxicity
1	Mild toxicity (non dose-limiting)
2	Non-mild toxicity (non dose-limiting)
3	Severe toxicity (non dose-limiting)
4	Dose limiting toxicity

Table: Toxicity "grades" (severities) for trial.

The rule is to escalate providing $S(i)$ is less than 2. Furthermore, once we have included 3 patients at some level then escalation to higher levels only occurs if each cohort of 3 patients does not experience dose limiting toxicity.

Rapid escalation based on grades



More complex problems

- Two group problem (patient heterogeneity)
- Bridging studies
- Within patient escalation
- Recording errors and non-drug related DLTs
- Multi-drug problem, partial ordering
- Graded toxicities

More complex problems

- Two group problem (patient heterogeneity)
- Bridging studies
- Within patient escalation
- Recording errors and non-drug related DLTs
- Multi-drug problem, partial ordering
- Graded toxicities

More complex problems

- Two group problem (patient heterogeneity)
- Bridging studies
- Within patient escalation
- Recording errors and non-drug related DLTs
- Multi-drug problem, partial ordering
- Graded toxicities

More complex problems

- Two group problem (patient heterogeneity)
- Bridging studies
- Within patient escalation
- Recording errors and non-drug related DLTs
- Multi-drug problem, partial ordering
- Graded toxicities

More complex problems

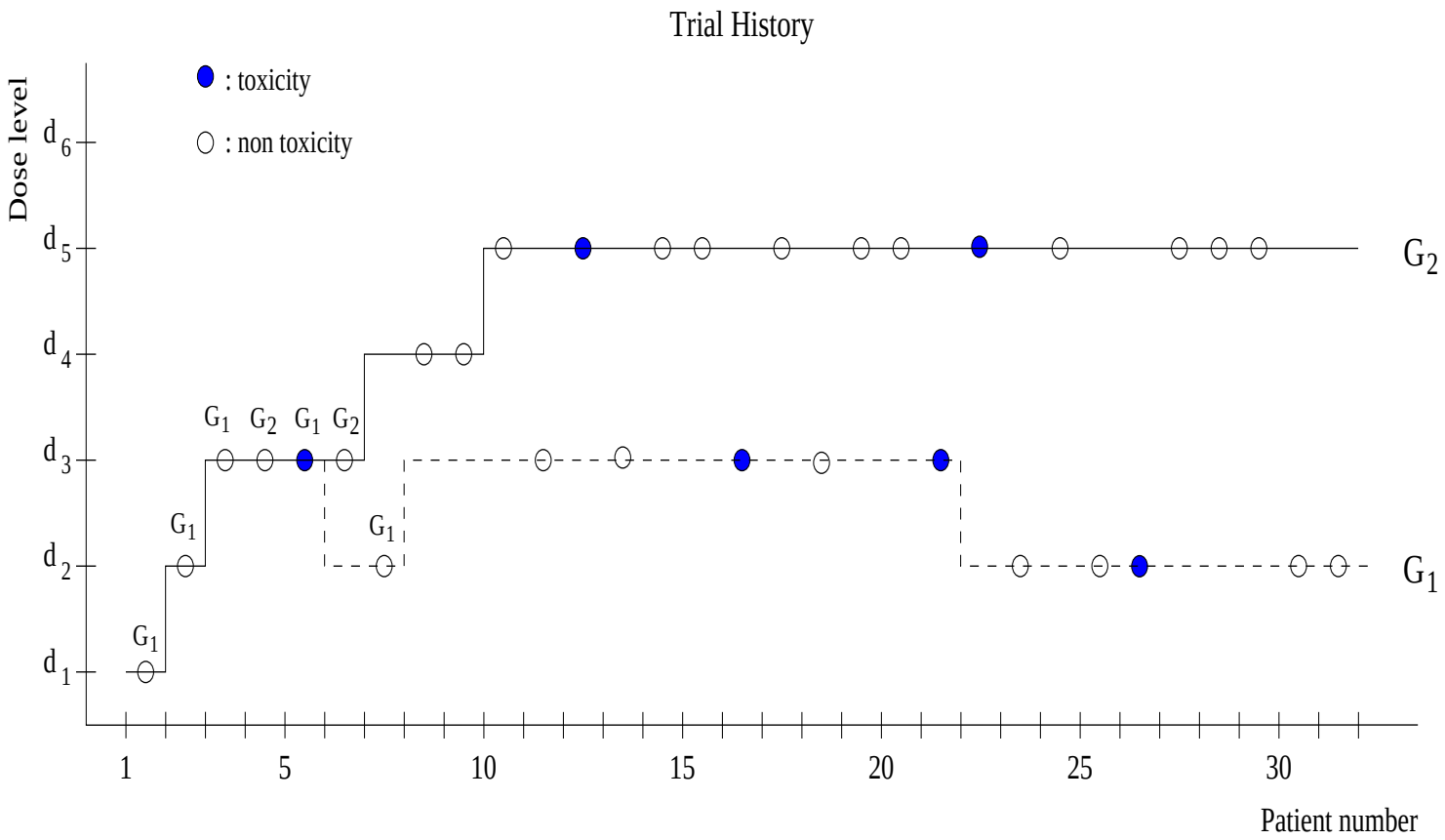
- Two group problem (patient heterogeneity)
- Bridging studies
- Within patient escalation
- Recording errors and non-drug related DLTs
- Multi-drug problem, partial ordering
- Graded toxicities

More complex problems

- Two group problem (patient heterogeneity)
- Bridging studies
- Within patient escalation
- Recording errors and non-drug related DLTs
- Multi-drug problem, partial ordering
- Graded toxicities

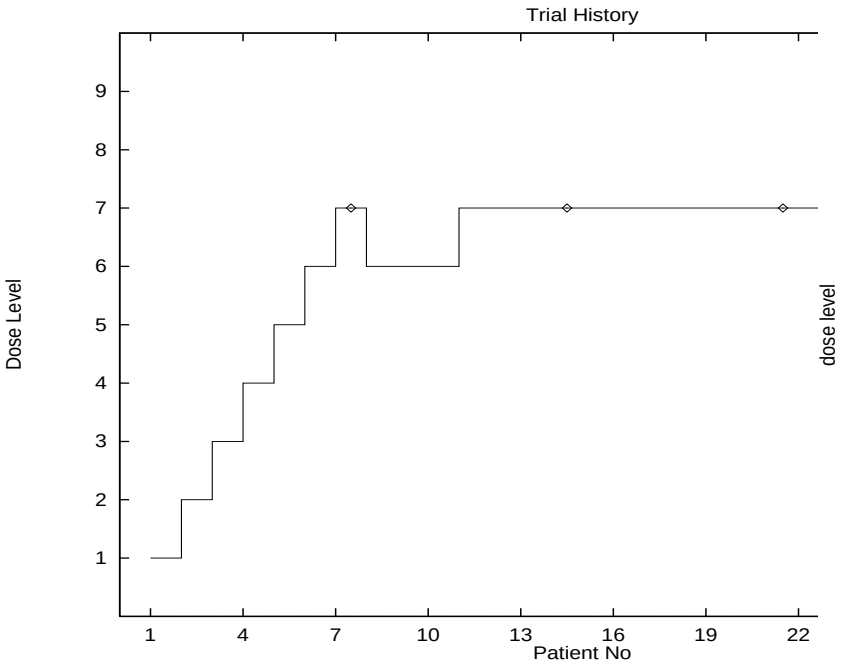
Two groups in a single trial

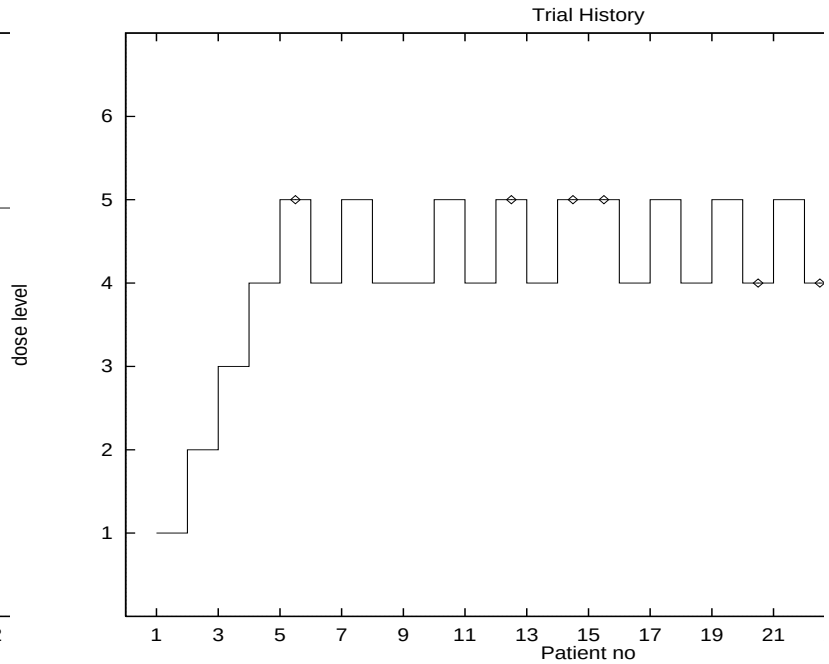
R_1	.02	.19	.31	.45	.51	.63
R_2	.03	.05	.11	.21	.39	.50



Modification of algorithm

1. Choose level d_j closest to target.
2. Choose level d_j according to some probability mechanism.





Within patient escalation

Patient	level 1	level 2	level 3	level 4	level 5
1	0	1	1		
2		0	1		
3			2		
4				2	3
5				1	4 (DLT)
6				?	

Table: *Acceleration information from graded toxicities. Entries are the grades.*

Phase I study of a combination

Table: Drug combinations used in Phase 1 trial of Samarium Lexidronam and Bortezomib DLT defined by as a grade 3+ neutropenia (Berenson et al. 2009)

	Drug Combination					
Agent	d_1	d_2	d_3	d_4	d_5	d_6
Sm (mCi/kg)	0.25	0.5	1.0	0.25	0.5	1.0
Bortezomib (mg/m^2)	1.0	1.0	1.0	1.3	1.3	1.3

- We index the models by M where M takes value M_h under the h^{th} possible ordering

$$M_1 : d_1 \rightarrow d_2 \rightarrow d_3 \rightarrow d_4 \rightarrow d_5 \rightarrow d_6$$
$$M_2 : d_1 \rightarrow d_2 \rightarrow d_4 \rightarrow d_3 \rightarrow d_5 \rightarrow d_6$$

Set of possible orders of toxicity probabilities

M	Simple Order								
M_1	$R(d_1)$	$\leq$	$R(d_2)$	$\leq$	$R(d_3)$	$\leq$	$R(d_4)$	$\leq$	$R(d_5)$
M_2	$R(d_1)$	$\leq$	$R(d_2)$	$\leq$	$R(d_4)$	$\leq$	$R(d_3)$	$\leq$	$R(d_5)$
M_3	$R(d_1)$	$\leq$	$R(d_2)$	$\leq$	$R(d_4)$	$\leq$	$R(d_5)$	$\leq$	$R(d_3)$
M_4	$R(d_1)$	$\leq$	$R(d_4)$	$\leq$	$R(d_2)$	$\leq$	$R(d_3)$	$\leq$	$R(d_5)$
M_5	$R(d_1)$	$\leq$	$R(d_4)$	$\leq$	$R(d_2)$	$\leq$	$R(d_5)$	$\leq$	$R(d_3)$

Model choice

- Likelihood $\mathcal{L}_{mj}(a)$ for model m after j patients is (proportional);

$$\sum_{\ell=1}^{j} y_\ell \log \psi_m(x_\ell, a) + \sum_{\ell=1}^{j} (1 - y_\ell) \log(1 - \psi_m(x_\ell, a))$$

- Obtain $\hat{a}_{mj}$
- Estimate probability of toxicity d_i via: $\hat{R}(d_i) = \psi_m(d_i, \hat{a}_{mj})\,, \quad (i = 1, \ldots, k)$.
- Given m, the dose to be given to the $(j+1)$ th patient, x_{j+1} is determined.
- Given $\quad_j$, posterior model probabilities are:

$$\pi(m|\quad_j) = \frac{\pi(m) \int_{-\infty}^{\infty} \exp\{\mathcal{L}_{mj}(u)\} g(u)\, du}{\sum_{m=1}^{M} \pi(m) \int_{-\infty}^{\infty} \exp\{\mathcal{L}_{mj}(u)\} g(u)\, du}$$

In some cases the $\pi(m|\quad_j)$ are only of very indirect interest,

Illustration

- $R = (0.04, 0.07, 0.20, 0.35, 0.55, 0.70)$.
- Target toxicity rate $\theta = 0.20$.
- The trial will treat $n = 24$ patients.
- For each ordering, we used the power model,

$$\psi_m(d_i, a) = \alpha_{mi}^{a}; \quad m = 1, \ldots, 5; \quad i = 1, \ldots, 6$$

Working Models

Table: Working model for five simple orders

M		Combinations					
		1	2	3	4	5	6
$m = 1$	1-2-3-4-5-6	0.01	0.07	0.20	0.38	0.56	0.71
$m = 2$	1-2-4-3-5-6	0.01	0.07	0.38	0.20	0.56	0.71
$m = 3$	1-2-4-5-3-6	0.01	0.07	0.56	0.20	0.38	0.71
$m = 4$	1-4-2-3-5-6	0.01	0.20	0.38	0.07	0.56	0.71
$m = 5$	1-4-2-5-3-6	0.01	0.20	0.56	0.07	0.38	0.71

Simulation results

Dose	d_1	d_2	d_3	d_4	d_5	d_6	n	tox
$R(d_i)$	0.26	**0.33**	0.51	0.62	0.78	0.86	-	-
Conaway et al.	0.35	**0.52**	0.11	0.02	0.00	0.00	21.3	8.5
POCRM	0.29	**0.50**	0.16	0.04	0.01	0.00	22.0	8.4
CRM	0.27	**0.49**	0.23	0.01	0.00	0.00	22.0	7.9
$R(d_i)$	0.12	0.21	**0.34**	0.50	0.66	0.79	-	-
Conaway et al.	0.07	0.29	**0.42**	0.21	0.01	0.00	25.6	9.0
POCRM	0.02	0.23	**0.55**	0.11	0.10	0.00	26.0	10.0
CRM	0.01	0.18	**0.63**	0.17	0.01	0.00	25.0	7.5
$R(d_i)$	0.04	0.07	0.20	**0.33**	0.55	0.70		-
Conaway et al.	0.00	0.02	0.38	**0.51**	0.08	0.02	28.5	8.8
POCRM	0.00	0.00	0.26	**0.50**	0.23	0.01	29.0	10.8
CRM	0.00	0.01	0.19	**0.67**	0.13	0.00	28.0	8.0
$R(d_i)$	0.01	0.04	0.05	0.17	**0.33**	0.67		-
Conaway et al.	0.00	0.00	0.06	0.25	**0.64**	0.05	29.0	7.8
POCRM	0.00	0.00	0.01	0.29	**0.61**	0.09	29.0	9.4
CRM	0.00	0.00	0.00	0.18	**0.76**	0.06	28.0	6.3
$R(d_i)$	0.01	0.02	0.05	0.15	0.20	**0.33**		-
Conaway et al.	0.00	0.00	0.01	0.04	0.37	**0.59**	26.2	5.8
POCRM	0.00	0.00	0.00	0.20	0.12	**0.68**	27.0	6.4
CRM	0.00	0.00	0.00	0.05	0.26	**0.69**	27.0	4.3

Most successful dose (MSD)

Example in HIV;

1. Treatment over long period.
2. Toxicity is inability to take treatment.
3. Observation window for efficacy comparable to toxicity.
4. Lack of efficacy as bad, possibly worse, than toxicity.

Introduce the following definitions;

1. $R(x_j) = \Pr(Y_j = 1 | X_j = x_j)$
2. $Q(x_j) = \Pr(V_j = 1 | X_j = x_j, Y_j = 0)$
3. $P(d_i) = Q(d_i)\{1 - R(d_i)\}$.

Models

Let; $R(x_j) = E(Y_j|x_j) = \psi(x_j, a)$;
$Q(x_j) = E(V_j|x_j, Y_j = 0) = \phi(x_j, b)$

$P(x_j) = \phi(x_j, b)\{1 - \psi(x_j, a)\}$ and $Q(x) = H\{R(x)\}$

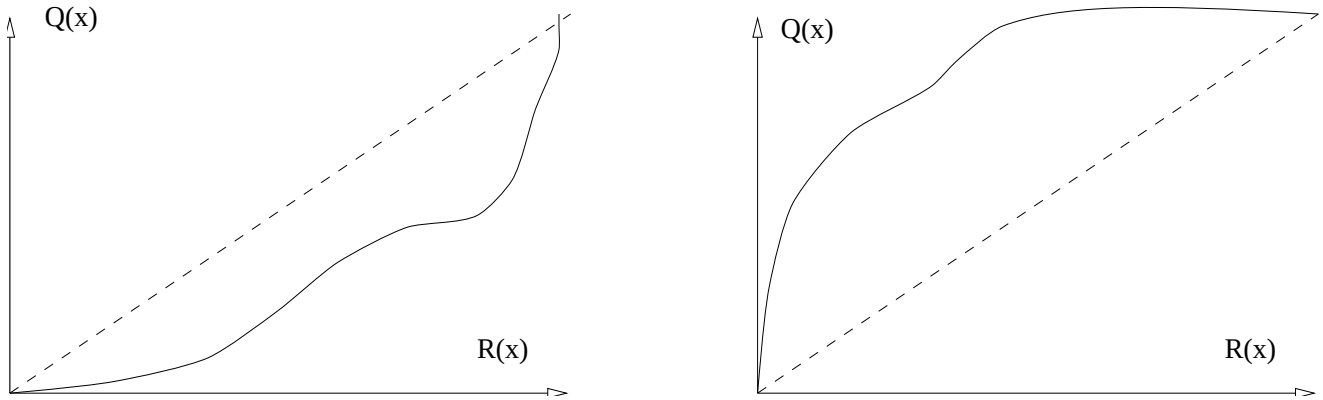


Figure: Possible relationships for $Q(x) = H\{R(x)\}$

Compromise structure

O'Quigley, Hughes and Fenton (*Biometrics* **57**, 1018-29) suggest;

1. Choose, say, $\theta = 0.1$
2. Use SPRT to test $H_0 : P \in (0, 0.7)$ versus $H_1 : P \in (0.7, 1.0)$
3. If SPRT chooses H_0 at d_i then remove levels $d_1, ..., d_i$, and, modify θ to $\theta +$.

Some simulated situations

	d_1	d_2	d_3	d_4	
R_k	0.06	0.15	0.25	0.30	
Q_k	0.21	0.82	0.80	0.71	Scheme 1
P_k	0.20	0.70	0.60	0.50	
R_k	0.15	0.30	0.40	0.50	
Q_k	0.82	0.71	0.83	0.80	Scheme 2
P_k	0.70	0.50	0.50	0.40	
R_k	0.00	0.05	0.15	0.30	
Q_k	0.10	0.32	0.82	0.71	Scheme 3
P_k	0.10	0.30	0.70	0.50	
R_k	0.00	0.00	0.10	0.15	
Q_k	0.20	0.30	0.56	0.82	Scheme 4
P_k	0.20	0.30	0.50	0.70	

Some simulated situations

	d_1	d_2	d_3	d_4	
% rec	0.00	0.97	0.03	0.00	Scheme 1
% alloc	0.23	0.75	0.02	0.00	$n = 24.9$
% rec	0.96	0.04	0.00	0.00	Scheme 2
% alloc	0.76	0.24	0.00	0.00	$n = 21.7$
% rec	0.00	0.01	0.93	0.06	Scheme 3
% alloc	0.06	0.44	0.44	0.06	$n = 37.6$
% rec	0.00	0.00	0.12	0.87	Scheme 4
% alloc	0.00	0.37	0.32	0.31	$n = 48.5$

Table: Recommendation and in-trial allocation for the 4 schemes