Intermediary Grade Toxicities in Phase I Clinical Trials: How Much Information Is There?

Alexia Iasonos, PhD
Memorial Sloan Kettering Cancer Center, New York

Joint work with John O'Quigley,

University of Virginia, September 2012

Lower grades toxicities in dose-finding algorithms

- Toxicities are graded in a scale of 0-5:

  1-2 mild/moderate; 3-4 severe; 5 death

- Binary response: dose limiting toxicities

  DLT: yes/no

- The frequency of DLT underlies most Phase I designs.

Should the response be dichotomous or polytomous?

Statisticians

- Additional information in Grades 0-4
- Non-DLTs are an indication for a future DLT
- Outcome Ordinal response has to be more beneficial
- Models for ordinal response
- Modify /extend existing models to accommodate ordinal response

Clinicians

- Lower grades may add noise
- Non-DLTs resolve and are not worrisome; information that is not useful
- Non-DLTs are not an indication for a future DLT
- MTD (max tolerated dose) is the dose where an acceptable number of DLTs is observed
- DLTs alone guide us to the MTD

Literature Review

- Summary measures of “toxicity burden'' score
  (Bekele, Thall 2004)
  Form of a linear combination of weighted tox resulting in a single, continuous or quasi-continuous outcome.

- Combining various toxicities to a summary measure / continuous response (Yuan et al. 2007; Ivanova and Kim 2009; Lee and Cheung 2010, Chen et al 2010)

- Fitting models for ordinal response - outcome is any toxicity grade in the scale of 1-5 (Ivanova ,2006; Van Meter, 2011; proportional odds model).

Objective

- Maintain the response as binary: DLT (yes/no)
- Refine our estimate of *Prob*(DLT) by using intermediary grades explicitly
- To what extent **auxiliary information** on lower and intermediary grades can help us in estimating the MTD

Problem I

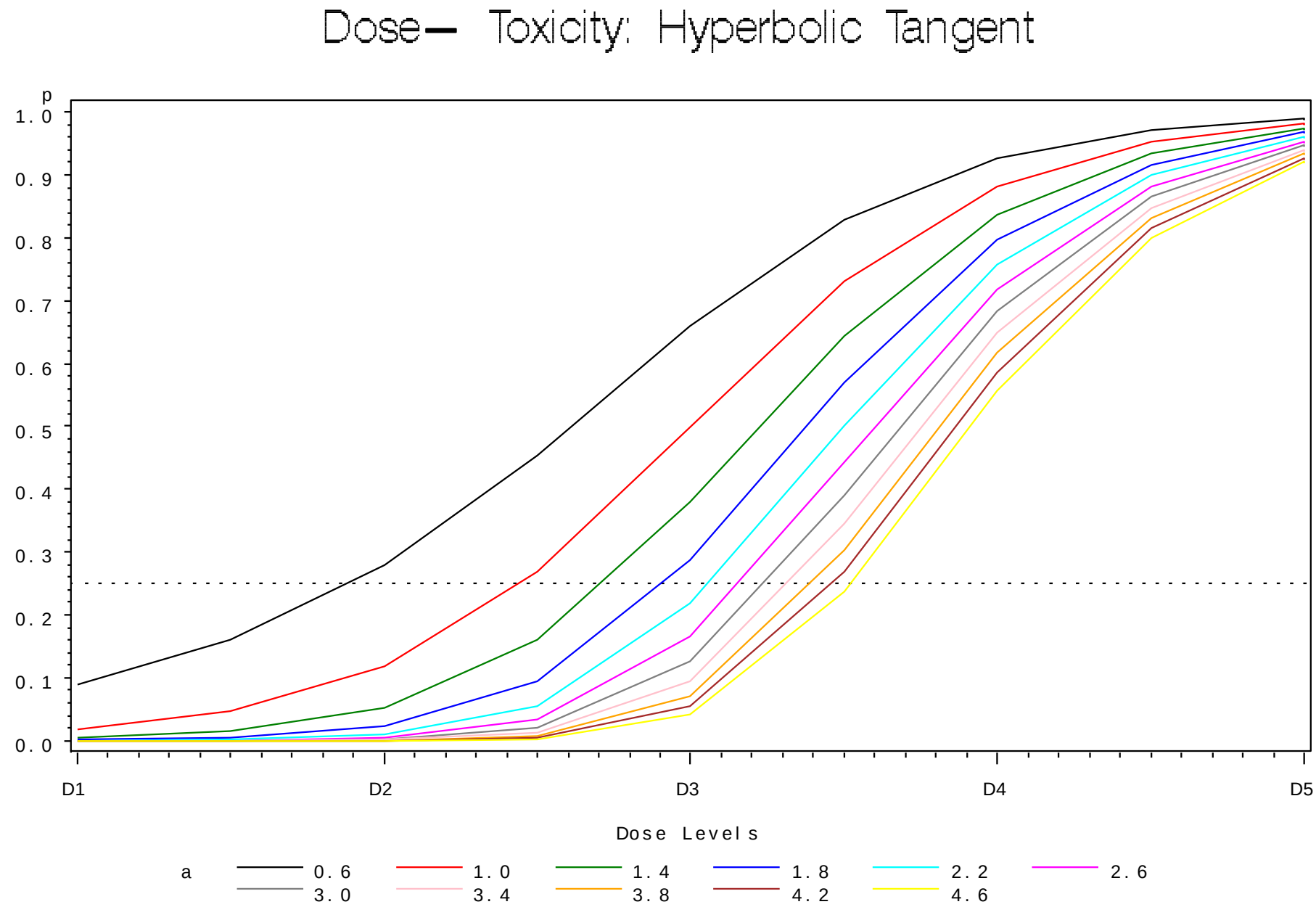
- Trinary response
- A patient can have either
  - No toxicity
  - Mild/moderate
  - Severe

Problem II

- Model the effect of intermediary grades on DLT
- Conditional probabilities
- A patient can have both
  - mild/moderate
  - severe

Background: Model based designs

- Model based designs
- Continual Reassessment Method (CRM)


Dose– Toxicity: Hyperbolic Tangent
p
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
D1
D2
D3
D4
D5
Dose Levels
a
0.6
1.0
1.4
1.8
2.2
2.6
3.0
3.4
3.8
4.2
4.6

CRM background
CRM (O' Quigley et al. 1990)

Step 1

Define ***k*** pre-specified discrete dose levels, d_i

Dose – toxicity curve: p = Prob (DLT at dose d_i)

- Hyperbolic tangent: $p = [e^{d} / (e^{d} + e^{-d})]^{\alpha}$
- Logistic: $p = e^{3+\alpha * d} / (1+ e^{3+\alpha * d})$
- Power model: $p = d^{\alpha}$

CRM: Bayesian Framework

$$\psi(d_i, a) = \frac{\exp(c + ad_i)}{1 + \exp(c + ad_i)}, g(a) = \exp(-a), \textit{prior}$$

$$F_j = \{(d_1, y_1), \ldots (d_j, y_j)\}, \textit{data}$$

$$f(a, F_{j+1}) = \frac{g(a)\prod_{l=1}^{j} \psi(d_l, a)^{y_l}[1 - \psi(d_l, a)]^{1-y_l}}{\int_0^{\infty} g(u)\prod_{l=1}^{j} \psi(d_l, u)^{y_l}[1 - \psi(d_l, u)]^{1-y_l} du}, \textit{posterior}$$

$$a(j) = \int a f(a, F_{j+1}) da, \textit{mean}$$

$$\hat{p}_i(j) = \psi(d_i, a(j)), i = 1, \ldots, k$$

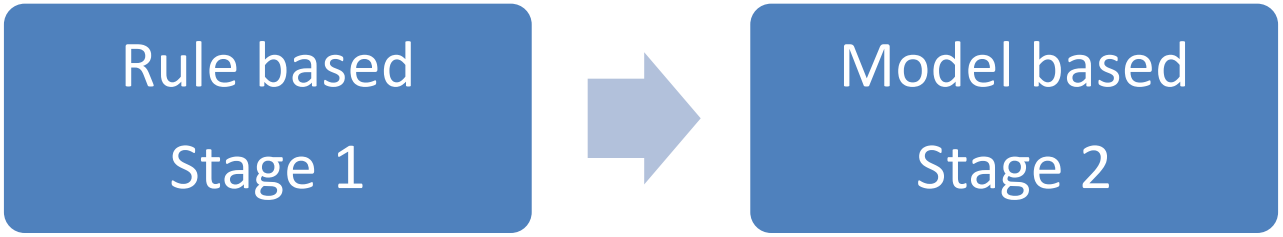
CRM: Likelihood Approach

Derivative of the log likelihood after we have observed *J* patients

Data (x_j, y_j)

$$L_j(a) = \sum_{j=1}^{J} y_j \frac{\psi'}{\psi}(x_j, a) + (1 - y_j) \frac{-\psi'}{(1-\psi)}(x_j, a)$$

Two stage Designs?



Clinicians , FDA, Statisticians

Advantages both clinically and in model estimation

Prior less informative

Heterogeneity allows for Likelihood estimation

Where should we use lower or intermediary grades?

- During the first stage only (part of a rule)
  - Clinically appealing
  - Database infrastructure is unchanged
- During the second stage only (part of the model)
- Both
- Neither

Problem I (trinary)

- A patient can have one of three mutually exclusive outcomes:
  - No toxicity (none)
  - Intermediary grade (mild-moderate)
  - Severe grade (DLT)

- Model the rates of severe toxicity (DLT) conditional on what we know in terms of the other rate
- Sum of rates is 1

Problem I
Notation: Data (x_j, Y_j)

- Y_j : outcome :
  - 0, corresponding to NCI grade 0 (none)
  - 1, corresponding to NCI grade 1-2 (mild/moderate),
  - 2, corresponding to NCI grade 3-4 (severe/ DLT)

$d_i, i = 1,\ldots,k$

x_j : visited dose level for patient j

Comparison of Designs

- no grades , 3+3 → CRM (binary)
- 1st stage grades → CRM (binary)
- both stages – 1 par model
- both stages – 2 par model

Likelihood approach for model estimation

$$L_j = L_{j-1}\varphi(x_j, Y_j)$$

where $\varphi(x_j, Y_j)$ is the contribution of the jth patient to the likelihood given by:

$$\varphi(x_j, Y_j) = p_0^{Y_{j0}} p_1^{Y_{j1}} (1 - p_0 - p_1)^{1 - Y_{j0} - Y_{j1}}$$

$$p_0 = P(Y = 0) = 1 - \psi_2(x_j, a, b) = 1 - (x_j{}^a)^b$$

$$p_1 = P(Y = 1) = \psi_2(x_j, a, b) - \psi_1(x_j, a) = (x_j{}^a)^b - x_j{}^a$$

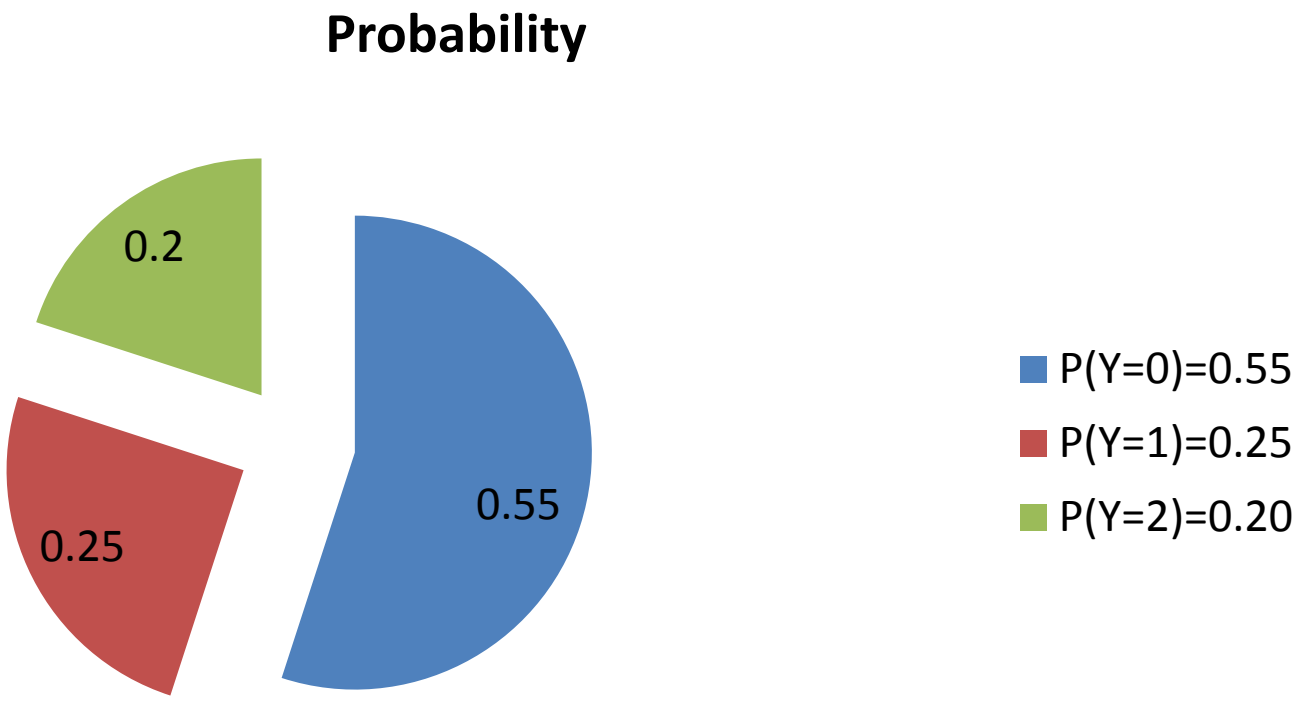
$$p_2 = P(Y = 2) = \psi_1(x_j, a) = x_j{}^a \text{ and}$$

$$P(Y = 1 \text{ or } Y = 2) = \psi_2(x_j, a, b)$$

$$P(Y=2)=x_i^{\,a};P(Y=0)=1-(x_i^{\,a})^b$$

$$Example: P(Y=0)=1-(0.2)^{0.5}=1-0.45=0.55$$

$$P(Y=2\,|\,Y=2\ \ or\ 1)=(x_j^{\,a})^{1-b}$$



A. Design CRM:

B. Two-stage CRM: 1st stage is based on 3+3 escalation

B. Design CRMG(1,0):

Two stage CRM; grades used in first stage only

Dose-escalation occurs when:

Sum of toxicities ≤ 2 escalate;

Sum of toxicities > 2, then use CRM: binary

Likelihood approach in 2nd stage

C) CRMG (1,1) (*b* is known):

Use a model to relate the probability of a higher grade toxicity given information on the lower grade toxicity.

1) *b* is known and set to a constant value,
2) *b* is known but our assigned value is wrong

D) CRM G (1,2) (*b* is unknown) :

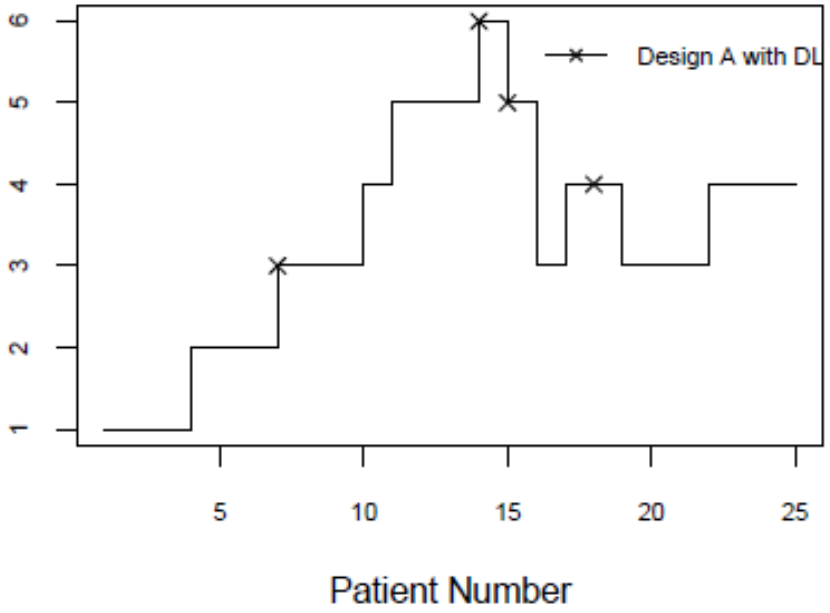
Same as CRMG (1,1)

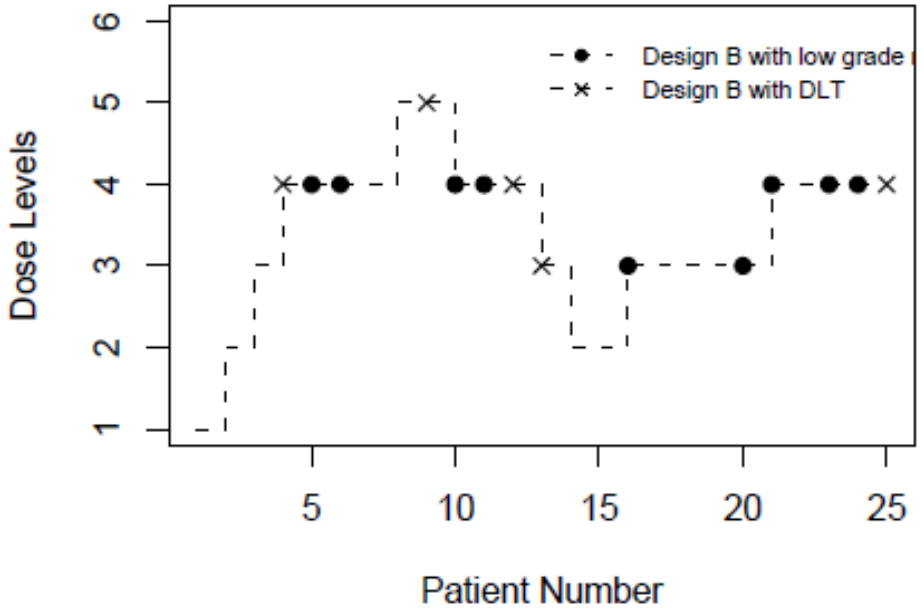
Parameter *b* to be unknown and to be estimated, simultaneously with the parameter *a,* by maximizing the likelihood.

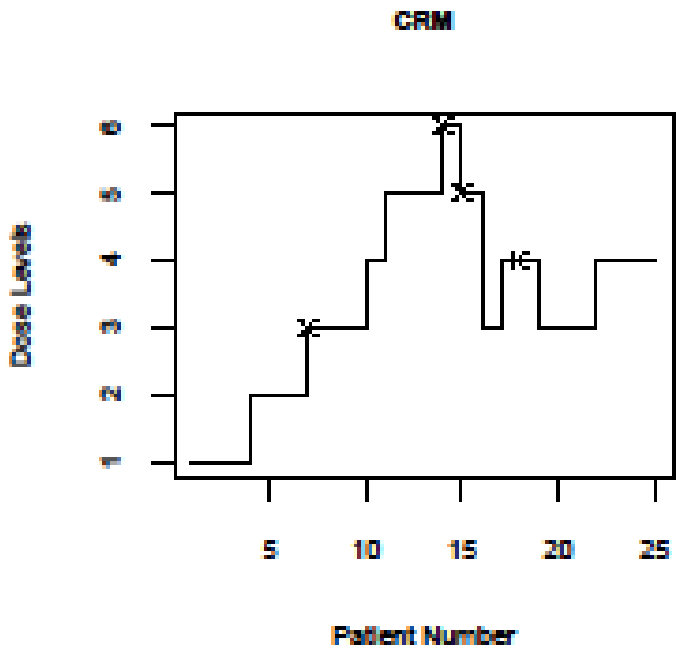
An example: trial history

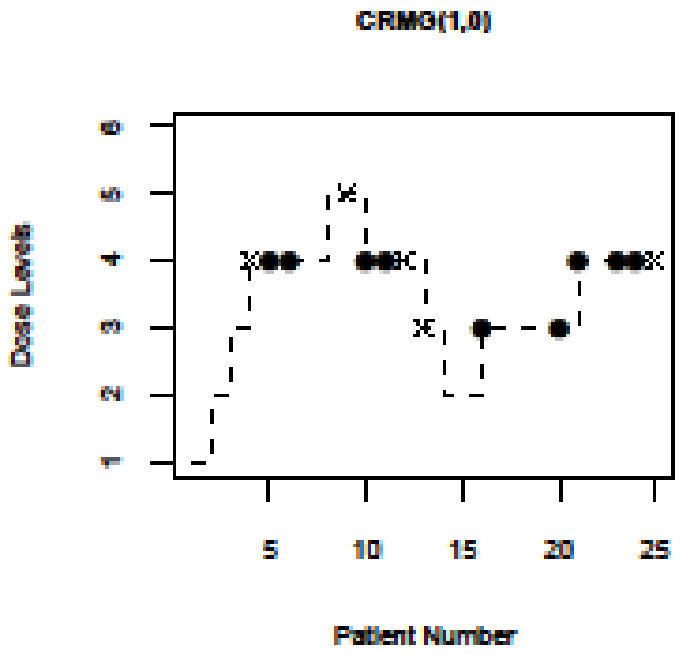
True DLT rates

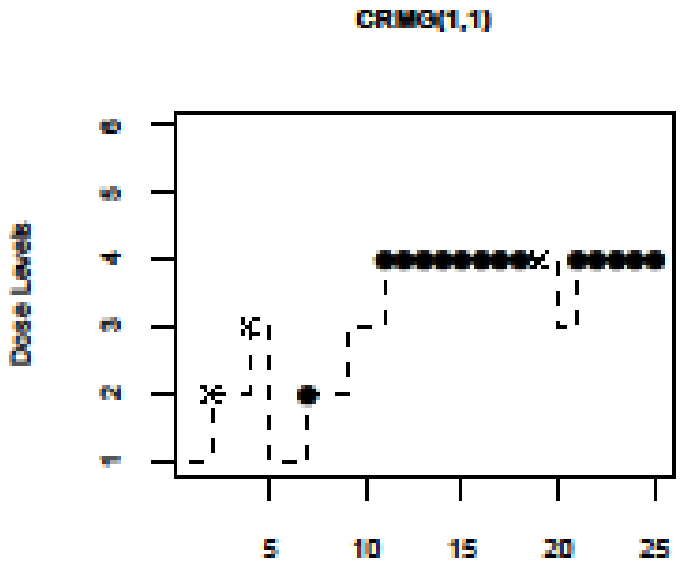
0.05, 0.1, 0.15, 0.2, 0.3, 0.4 (MTD= dose 4, target=0.2)

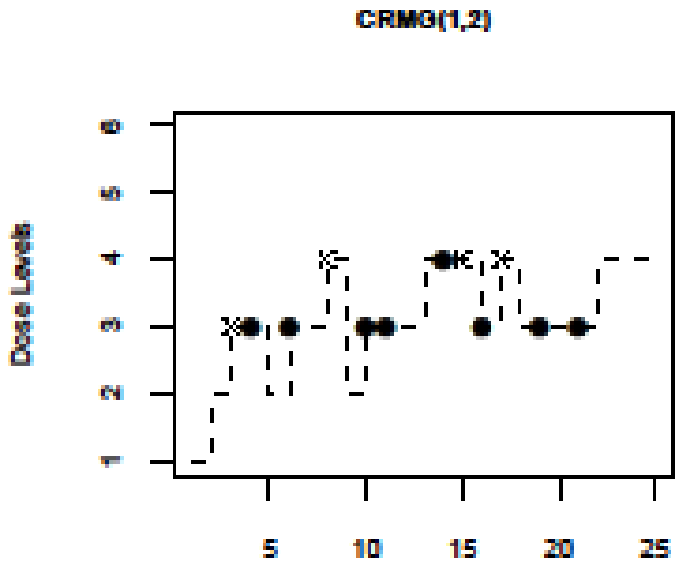





CRM
Dose Levels
Patient Number


CRMG(1,0)
Dose Levels
Patient Number


CRMG(1,1)
Dose Levels


CRMG(1,2)
Dose Levels

Simulation Study

- We simulated 1000 trials testing six dose levels with a fixed sample size of 25 patients; varied the number of dose levels from 4 to 6.

- Skeleton values = (0.05,0.1,0.15,0.2,0.25,0.3) representing the standardized units for actual dose levels.

- The target rate of acceptable toxicity at the MTD varied between 0.2 and 0.3. $\theta = 0.2$.

- The MTD was selected as the level with estimated P(DLT) closer to the target rate of 0.2.

Various scenarios for true toxicity rates

For the data generation, a, b are known

Once P (Y = 2) was known,

P (Y = 0) = 1−P(Y = 2)^b, where b=0.32. The value of b = 0.32 was chosen so that the P(Y = 0) = P(Y = 1) = 0.4 at the MTD.

For the analysis part, we used the previous working models and we evaluated cases when b is known with various values of b, such as

b = 0.32, 0.5, 0.4, 0.25, 0.20, which correspond to P(Y=0) at MTD of 0.4, 0.55, 0.47, 0.33, 0.27

True Toxicity rates for simulations

SCENARIO 1	SCENARIO 2	SCENARIO 3	SCENARIO 4
True Toxicity Rates: P(Y=2) | P(Y=1) | P(Y=0)			
.10 | .38 | .52	.05 | .33 | .62	.01 | .22 | .77	.01 | .22 | .77
.15 | .39 | .46	.10 | .38 | .52	.04 | .32 | .64	.05 | .33 | .62
.20 | .40 | .40	.15 | .39 | .46	.08 | .37 | .55	.07 | .36 | .57
.30 | .38 | .32	**.20 | .40 | .40**	.12 | .39 | .49	.10 | .38 | .52
.45 | .32 | .23	.30 | .38 | .32	**.20 | .40 | .40**	.12 | .39 | .49
.50 | .30 | .20	.40 | .35 | .35	.30 | .38 | .32	**.20 | .40 | .40**

Designs were compared in terms of:

- Accuracy (percent of trials)
- Patient allocation (percent of patients)
- Safety (median DLTs)
- Precision (95% confidence interval for the predicted probability of DLT at the MTD)

95% CI estimation for the Probability (DLT) at MTD

A 95% CI for the probability of DLT at the MTD can be estimated by normal approximation when the response is binary [O'Quigley et al. 2002, Biostatistics]

For designs with 2 parms the variance of a is approximated numerically and obtained via the inverse of the information matrix when maximizing the likelihood at the $(n + 1)^{th}$ assignment

Results: Sensitivity analysis

- Comparing designs with *b* known
  - equal to the true rate=0.32 vs
  - *b* is known but assumed a wrong value (sensitivity analysis with b=0.5, 0.4, 0.25, 0.20)

- Even if *b* is assumed a wrong value, accuracy is very close to when *b* is assumed the correct value (1-2%), except when *b* far away from the truth (depending on scenario).

Iasonos et al.
Clin Trials 2011

Table 4: Estimated Probability of Dose selection

Dose Level	d_1	d_2	d_3	d_4	d_5	d_6	NF
Scenario 1							
CRM	0.142	0.277	**0.299**	0.211	0.054	0.017	
CRMG(1,0)	0.154	0.300	**0.320**	0.182	0.036	0.008	
CRMG(1,1)	0.122	0.315	**0.345**	0.171	0.042	0.005	
CRMG(1,2)	0.144	0.323	**0.329**	0.164	0.032	0.008	
Scenario 2							
CRM	0.022	0.132	0.237	**0.300**	0.213	0.096	
CRMG(1,0)	0.021	0.148	0.277	**0.312**	0.183	0.059	
CRMG(1,1)	0.005	0.127	0.291	**0.325**	0.192	0.060	
CRMG(1,2)	0.020	0.132	0.281	**0.328**	0.164	0.075	
Scenario 3							
CRM	0.001	0.011	0.065	0.228	**0.320**	0.368	0.007
CRMG(1,0)	0	0.015	0.059	0.286	**0.319**	0.321	
CRMG(1,1)	0	0.003	0.055	0.233	**0.358**	0.351	
CRMG(1,2)	0	0.016	0.057	0.266	**0.318**	0.343	
Scenario 4							
CRM	0.001	0.007	0.039	0.095	0.187	**0.651**	0.020
CRMG(1,0)	0.008	0.007	0.021	0.094	0.209	**0.661**	
CRMG(1,1)	0	0.001	0.017	0.071	0.203	**0.708**	
CRMG(1,2)	0	0.005	0.024	0.095	0.189	**0.678**	0.009

NF= not found

Table 5: Patient allocation: proportion of patients treated at each level

Dose Level	d_1	d_2	d_3	d_4	d_5	d_6
Scenario 1						
CRM	0.31	0.26	**0.21**	0.14	0.05	0.03
CRMG(1,0)	0.25	0.24	**0.23**	0.16	0.07	0.05
CRMG(1,1)	0.23	0.25	**0.24**	0.16	0.07	0.04
CRMG(1,2)	0.24	0.25	**0.23**	0.16	0.07	0.04
Scenario 2						
CRM	0.19	0.21	0.22	**0.18**	0.12	0.09
CRMG(1,0)	0.12	0.17	0.21	**0.20**	0.15	0.15
CRMG(1,1)	0.10	0.16	0.21	**0.23**	0.16	0.13
CRMG(1,2)	0.11	0.17	0.22	**0.21**	0.15	0.13
Scenario 3						
CRM	0.13	0.14	0.16	0.18	**0.17**	0.21
CRMG(1,0)	0.05	0.07	0.13	0.20	**0.21**	0.34
CRMG(1,1)	0.05	0.06	0.11	0.19	**0.23**	0.35
CRMG(1,2)	0.05	0.07	0.12	0.20	**0.21**	0.34
Scenario 4						
CRM	0.13	0.14	0.14	0.15	0.14	**0.30**
CRMG(1,0)	0.05	0.06	0.09	0.11	0.16	**0.53**
CRMG(1,1)	0.05	0.06	0.08	0.12	0.16	**0.53**
CRMG(1,2)	0.05	0.06	0.09	0.12	0.16	**0.52**

Results: accuracy

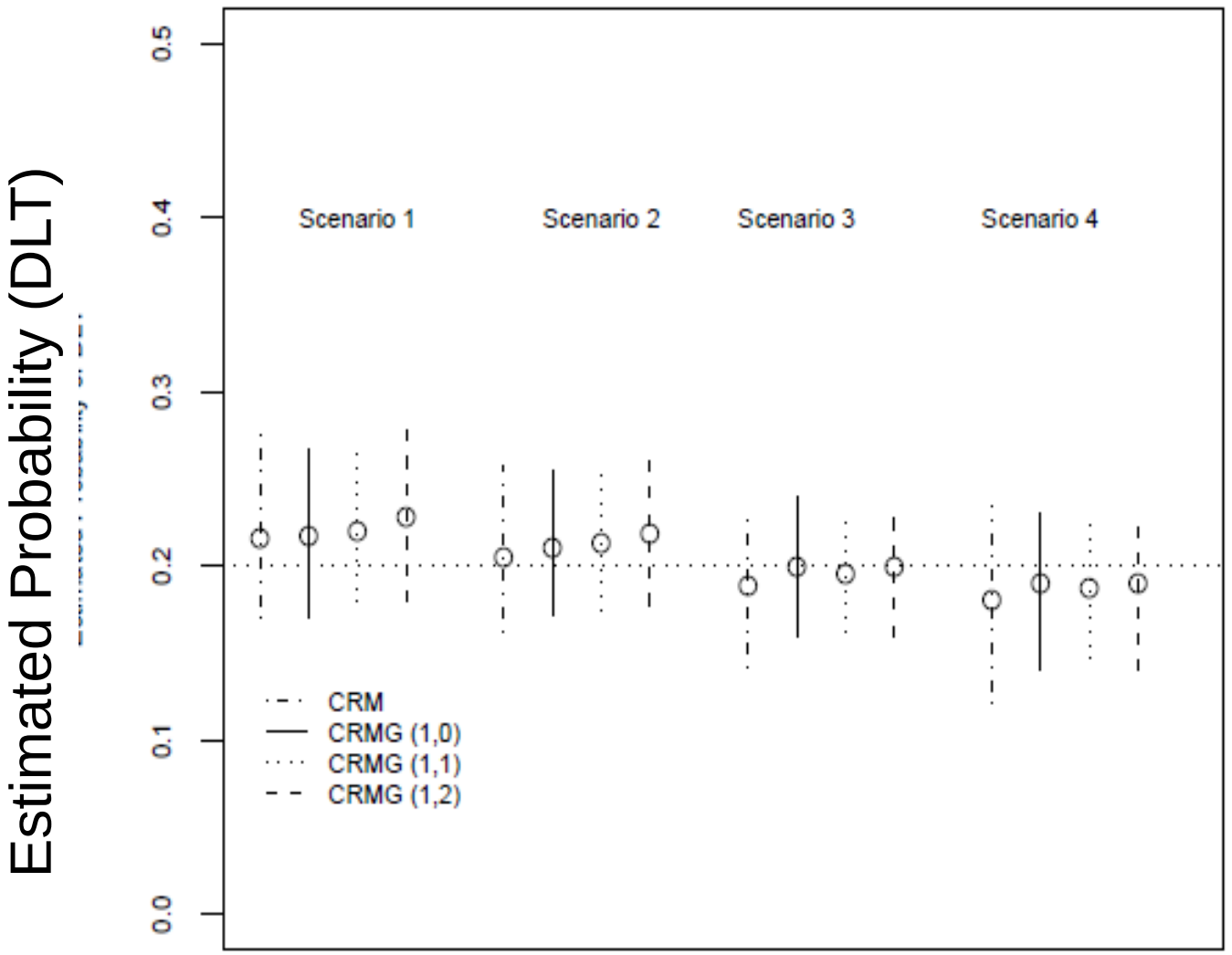
1. There are small gains (few % points) in accuracy and pt allocation in designs that utilize grades

2. Even when models are known explicitly and are not prone to misspecification, performance does not improve much

3. Most of the gains appear to result from using grades in 1st stage

Results : pt/trt allocation

- Using grades in stage 1 provides a compromise between aggressive/rapid dose escalation vs conservative
- CRM with some use of grades increases pt allocation at MTD by 5%
- Depending on the scenario, moderate tox may give the green light for the method to stay there, especially if *b* is underestimated


Estimated Probability (DLT)
0.0
0.1
0.2
0.3
0.4
0.5
Scenario 1
Scenario 2
Scenario 3
Scenario 4
CRM
CRMG (1,0)
CRMG (1,1)
CRMG (1,2)

95% Confidence Interval (CI) width of the dose-toxicity model parameter $\hat{a}$

	Median 95% CI width	$\hat{a}$	$Var(\hat{a})$
Scenario 1			
CRM	0.371	0.807	0.046
CRMG(1,0)	0.364	0.794	0.042
CRMG(1,1)	0.321	0.794	0.034
CRMG(1,2)	0.358	0.787	0.040
Scenario 2			
CRM	0.344	0.986	0.060
CRMG(1,0)	0.330	0.943	0.051
CRMG(1,1)	0.319	0.958	0.050
CRMG(1,2)	0.358	0.945	0.058
Scenario 3			
CRM	0.325	1.204	0.080
CRMG(1,0)	0.300	1.151	0.065
CRMG(1,1)	0.316	1.178	0.074
CRMG(1,2)	0.356	1.165	0.089
Scenario 4			
CRM	0.312	1.421	0.107
CRMG(1,0)	0.276	1.393	0.087
CRMG(1,1)	0.304	1.390	0.100
CRMG(1,2)	0.350	1.388	0.128

Results : precision (CI estimation)

- Design where *b* is known provides the smallest variance for the parameter of interest.

- Having an additional parameter to estimate (ie when the models are not known, are estimated from the data) might provide noise

Problem II

- A patient can have different types and grades of toxicities leading to multiple outcomes
- Model the rates of DLT conditional on the presence or absence of intermediary grades
  - P(severe myelosuppression| grade 2 rash)
- If intermediary grades are predictive of future DLTs, then incorporating this information should improve the estimation of the MTD

Problem II
Conditional Probabilities

- Different question:

  Model the effect of intermediary grades on DLT

  P(DLT| presence tox of intermediary grade) >

  P(DLT| absence tox intermediary grade)

Working models for conditional probabilities

Y =1 presence of DLT (binary)

W=1 presence of intermediary grade toxicity (binary)

$$\psi_1(d_i, a) = P(Y = 1|W = 1) = \alpha_i^a$$

$$\psi_2(d_i, a, D) = P(Y = 1|W = 0) = \alpha_{i-D}^a, where D = 0, 1, 2$$

$$\phi(d_i, b) = P(W = 1) = \beta_i^b$$

P(W=1)	.09	.18	.30	.38	.53	.66
P(Y=1|W=1)	.04	.08	.11	.22	.27	.36
P(Y=1|W=0)	.01	.04	.08	.11	.22	.27

Log Likelihood

$$
\begin{aligned}
log\mathcal{L}_j(a,b,D) = & \sum_{l=1}^{j} \quad y_l(1-w_l)log(\psi_2(x_l,a,D)(1-\phi(x_l,b))) \\
& + \quad (y_l * w_l)log(\psi_1(x_l,a)\phi(x_l,b)) \\
& + \quad (1-y_l)(1-w_l)log((1-\psi_2(x_l,a,D))(1-\phi(x_l,b))) \\
& + \quad (1-y_l)w_l log((1-\psi_1(x_l,a,D))\phi(x_l,b))
\end{aligned}
$$

Simulations

- Two stage Designs
  - 1st stage is based on an algorithm until heterogeneity
  - 2nd stage: dose allocation depends on estimated

  P(DLT) closer to an acceptable toxicity rate;

  P(DLT) depends on:
    - original LCRM (binary DLT)
    - Based on the following:

$$\hat{P}(Y = 1|d_i) = \psi_1(x_l, a)\phi\,(d_i, b) + \psi_2\,(d_i, a\ , D)(1 - \phi\,(d_i, b))$$

Simulations cont.

Parameters

- 1000 trials
- N=25 or 50 (~8 pts stage 1)
- 4 scenarios; true value of D=0, 1, 2
- MTD varied from level 3,4, 5,6
- Acceptable target rate: 0.20 - 0.25

Each scenario is described by 3 probabilities

- P(W=1)
- P(Y=1|W=1)
- P(Y=1|W=0)

→ P(Y=1)= P(DLT)

P(W=1)	.09	.18	.30	.38	.53	.66
P(Y=1|W=1)	.04	.08	.11	.22	.27	.36
P(Y=1|W=0)	.01	.04	.08	.11	.22	.27
P(Y=1)=	**.03**	**.07**	**.12**	**.18**	**.27**	**.36**

Simulation true rates

Scenario 1						
P(Y=1|W=0)	0.03	0.07	0.12	**0.16**	0.29	0.34
P(Y=1|W=1)	0.07	0.13	0.16	**0.29**	0.34	0.44
P(W=1)	0.17	0.28	0.40	**0.49**	0.62	0.74
P(Y=1)=	0.04	0.08	0.14	**0.22**	0.32	0.41
Scenario 2						
P(Y=1|W=0)	0.10	0.15	**0.22**	0.32	0.36	0.50
P(Y=1|W=1)	0.22	0.32	**0.36**	0.50	0.55	0.63
P(W=1)	0.02	0.06	**0.14**	0.21	0.35	0.51
P(Y=1)=	0.10	0.16	**0.24**	0.35	0.43	0.57
Scenario 3						
P(Y=1|W=0)	0.01	0.04	0.08	0.11	**0.22**	0.26
P(Y=1|W=1)	0.04	0.08	0.11	0.22	**0.27**	0.36
P(W=1)	0.09	0.18	0.30	0.38	**0.53**	0.66
P(Y=1)=	0.03	0.07	0.12	0.18	**0.28**	0.36
Scenario 4						
P(Y=1|W=0)	0.01	0.02	0.03	0.09	0.13	**0.21**
P(Y=1|W=1)	0.01	0.02	0.03	0.09	0.13	**0.21**
P(W=1)	0.04	0.10	0.19	0.27	0.42	**0.57**
P(Y=1)=	0.01	0.02	0.03	0.09	0.13	**0.21**

Table 1: Proportion of Dose selection

Dose Level	d_1	d_2	d_3	d_4	d_5	d_6
Scenario 1						
CRM (no grades)	0.01	0.09	0.32	**0.38**	0.16	0.04
CRM (grades)	0.01	0.10	0.32	**0.40**	0.14	0.04
Scenario 2						
CRM (no grades)	0.09	0.22	**0.39**	0.23	0.04	0.01
CRM (grades)	0.09	0.29	**0.39**	0.19	0.03	0.00
Scenario 3						
CRM (no grades)	0.00	0.01	0.09	0.29	**0.32**	0.29
CRM (grades)	0.00	0.01	0.11	0.30	**0.33**	0.24
Scenario 4						
CRM (no grades)	0.01	0.00	0.02	0.10	0.21	**0.66**
CRM (grades)	0.01	0.00	0.02	0.12	0.26	**0.60**

Why there is no improvement in accuracy?

- Depending on how far you start from MTD:
  - vicinity of MTD (DLT drive dose-escalation)
  - much higher (DLT will de-escalate)
  - much lower (moderate tox will guide us faster to MTD - more valuable when grades are used in first stage)

Conclusions
Ongoing Research

- As the trial progresses, we do learn some information on the rate of DLT vs rate of moderate toxicities

- The rate at which we learn this, and the orthogonality of this info with identifying the MTD, means we are never precise enough to sharpen our inference concerning the identification of MTD alone

- When some info is assumed to be known about the true rates linking the occurrence of lower vs higher grades tox

Bayesian Approach

- Bayesian framework (N. Wages, M. Conaway)

- Assume priors for the two parameters (some prior knowledge for b) and use the updated posterior distribution for the next dose-assignment

THANK YOU

iasonosa@mskcc.org

Acknowledgements:

Iasonos A, Zohar S., O'Quigley J.
Clinical Trials; Aug 2011 8(4):370-9.

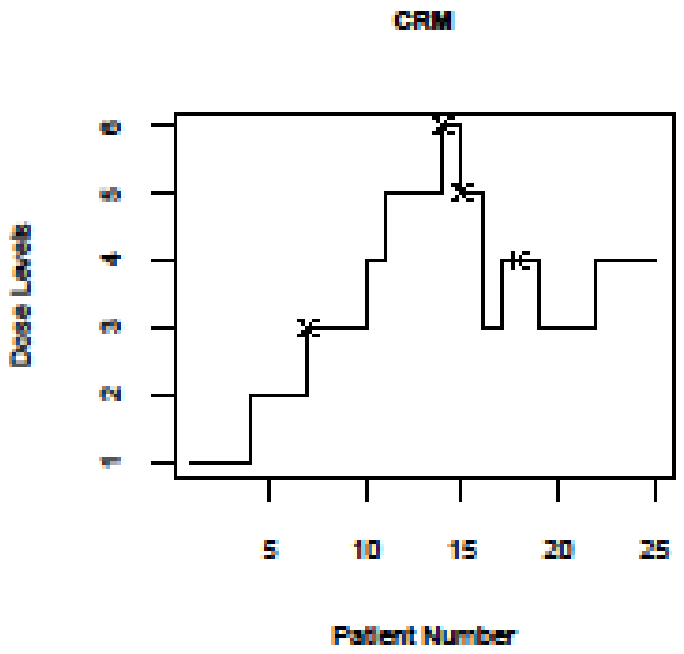
Partially funded by NCI: 1R01CA142859
(PI M. Conaway)
N. Wages and M. Conaway

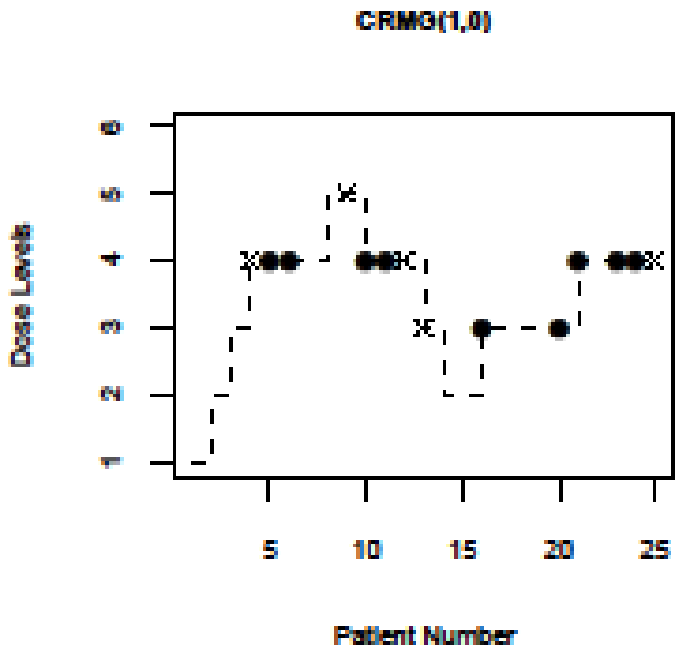
Extra slides

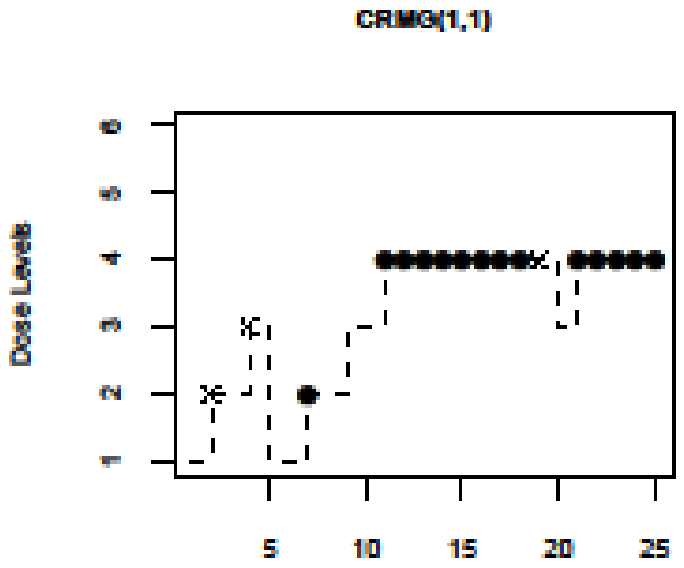
How much information is there?

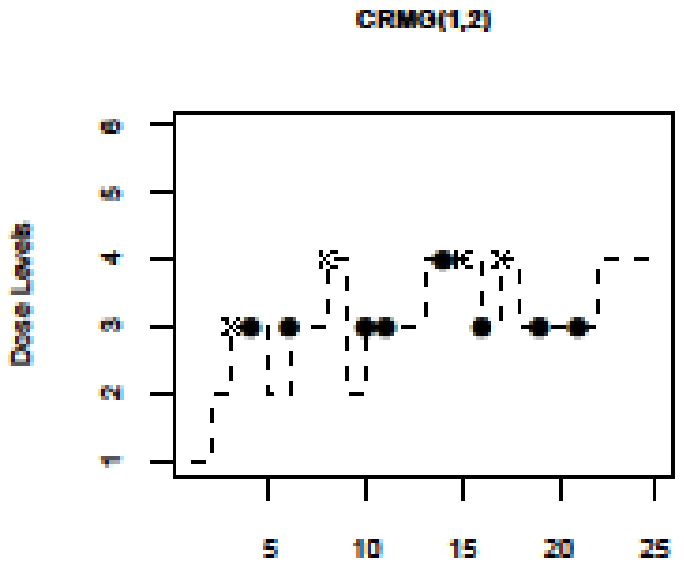
- Different types, grades of toxicities
- Toxicities are not exchangeable
- Time onset, reversible, attribution, other factors
- Clinicians use this information to decide MTD
- MTD can be different than RP2D

- Comparison of Designs must utilize same amount of information


CRM
Dose Levels
Patient Number


CRMG(1,0)
Dose Levels
Patient Number


CRMG(1,1)
Dose Levels


CRMG(1,2)
Dose Levels

Example cont.

dose	1	2	3	4	5	6
CRM	0/3	0/3	1/7	1/7	1/4	1/1
GCRM	0/1	0/3	1/7	3/12	1/2	0
			18/25 stage 1: 3+3 21/25 stage 1: grades			

Smaller spread, smaller oscillation