Performance, Accuracy and Safety of New Designs

Alexia Iasonos

Which design to use in practice?

Model based vs non model based

- Literature Review
    1. O' Quigley, Pepe, Fisher (1990)
    2. Goodman et al. (1995)- Group inclusions
    3. Shen, O'Quigley (1996) : Asymptotic distribution
    4. O'Quigley, Shen (1996): MLE – Frequentist approach
    5. Moller (1995) Comparisons
    6. Babb et al.(1998) Escalation with Overdose Control
    7. Heyd, Carlin (1999) Adaptive Design improvements
    8. Leung, Wang (2002) Extensions
    9. Rosenberger, Haines (2002) Competing designs
    10. Cheung YK, Chappell R (2002): Model sensitivity
    11. Garrett- Mayer (2007): Review
    12. Iasonos et al (2008): Comparative Review
    13. ….

CRM with modifications

- Start at lowest dose or dose closest to the target
- Skip dose levels or not (restrict escalation)
- Wait for all patients' responses or not
- CRM with fixed sample or with stopping rules

The modified versions do not change the operating characteristics.

Simulations: Methods

How do we compare methods based on simulations?

We know the true DLT rates at each level

Every patient has the same probability to experience a DLT

Design alone determines the dose allocation

	True Toxicity Rates (*pi*)				
Levels	D_1	D_2	D_3	D_4	D_5
Rates	.03	.05	.10	0.18	0.22

Simulate many trials and report on average

- How many trials recommended which level?
- Where patients were assigned/treated?

Comparisons of the methods via 1000 simulated trials

1. CRM starts at the dose level whose toxicity rate is closest to the target toxicity level, and restricts to no more than one dose level increase at a time.(O'Quigley et al., 1990)

2. CRM-I starts at the lowest dose level and prevents escalation of more than one dose level. It allows for one patient with incomplete DLT information (delayed response).

3. E-CRM starts at the lowest dose and uses an arbitrary starting plan for the dose allocation at the beginning of the trial until a DLT is observed. The method then switches to the CRM algorithm. (Moller, 1995)

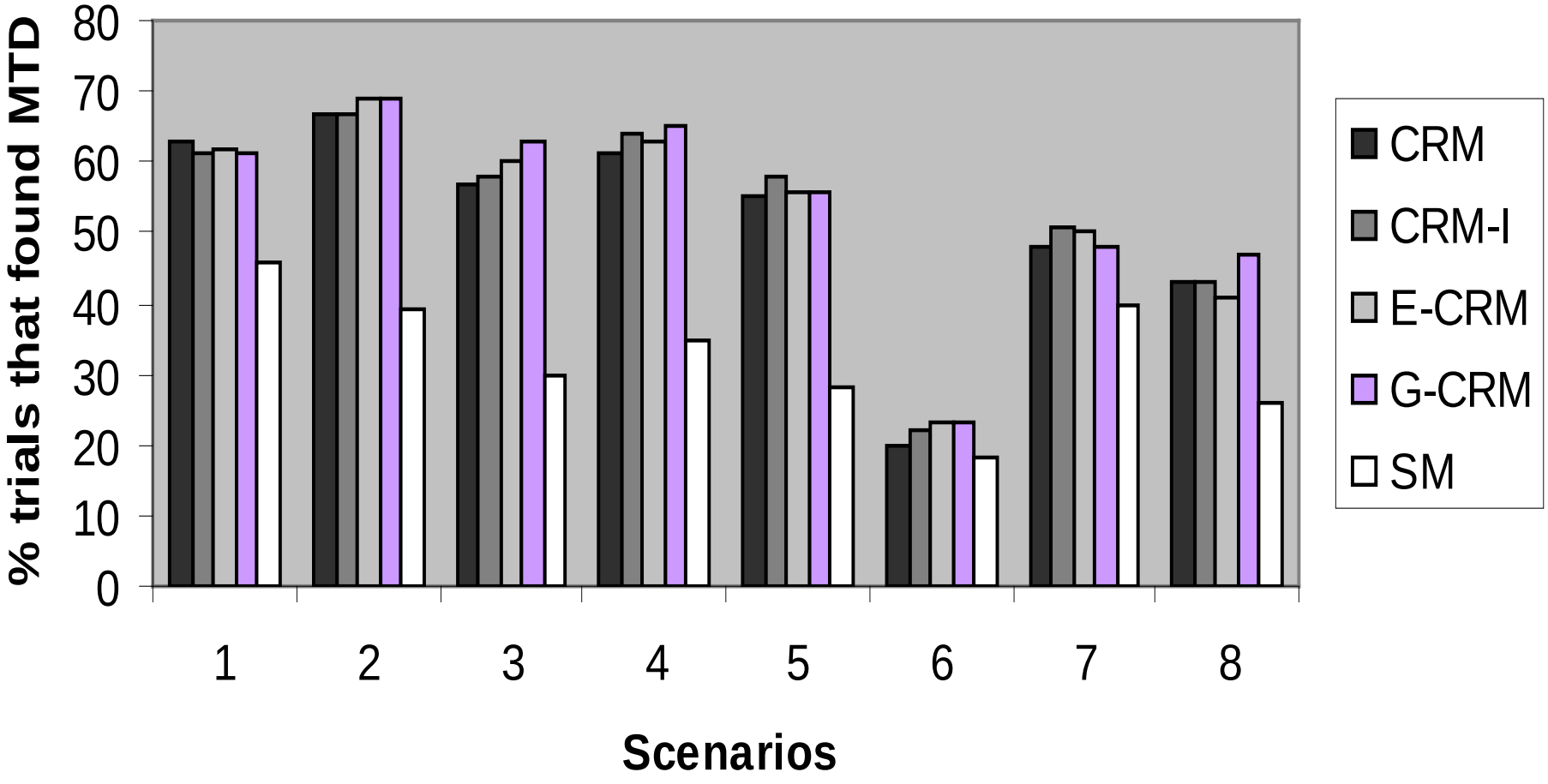
4. G-CRM starts at the lowest dose, treats patients in cohorts of 3 per dose level, and requires at least 6 patients to be treated at the MTD before the study completion. (Goodman et al, 1995)

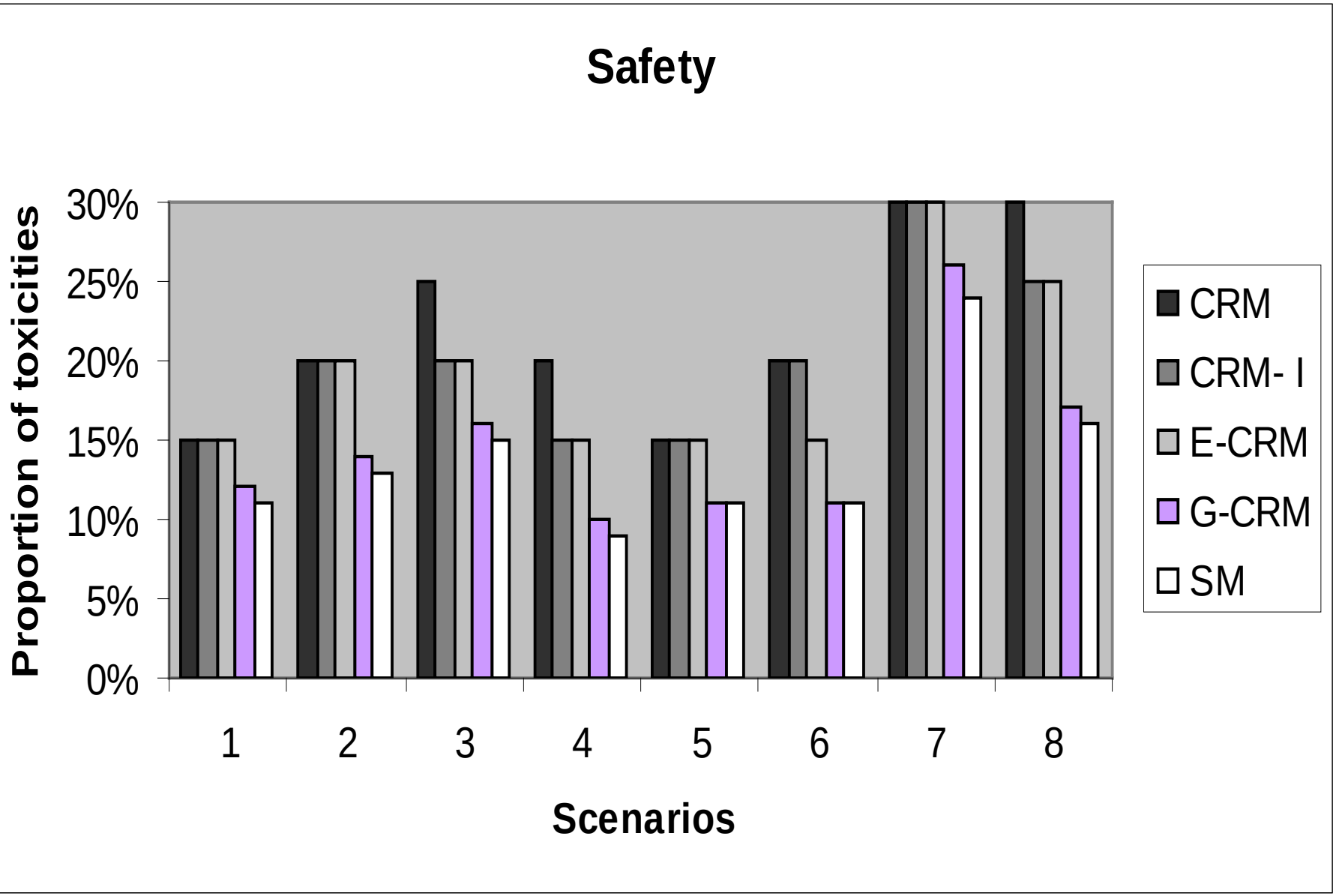
5. SM

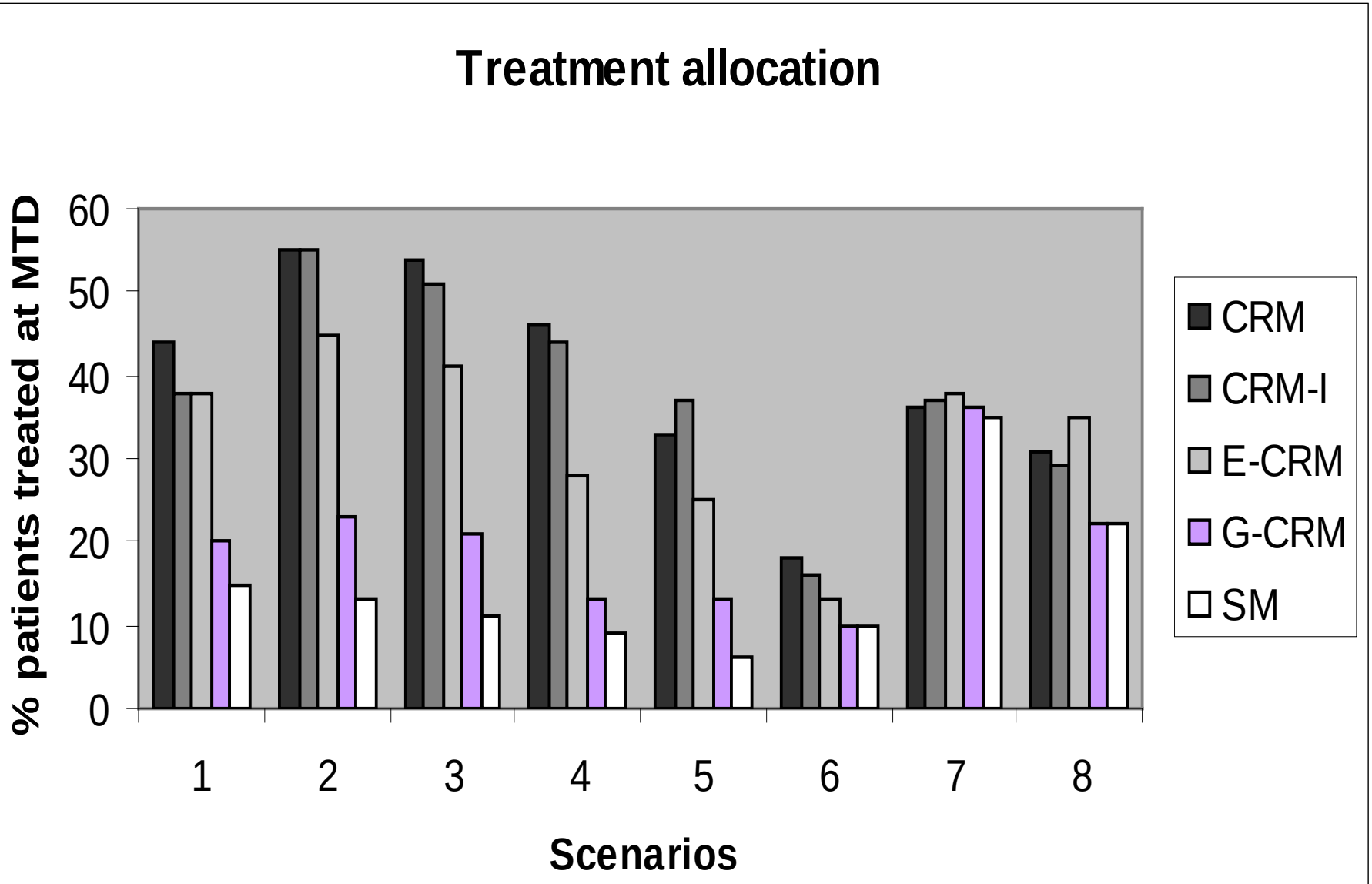
Scenario	True Toxicity Rates							
	D_1	D_2	D_3	D_4	D_5			
S1	0.03	0.05	0.10	0.18	0.22			
S2	0.06	0.09	0.13	0.16	0.25			
S3	0.06	0.10	0.15	0.19	0.28			
	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8
S4	0.0001	0.0025	0.02	0.06	0.09	0.12	0.16	0.25
S5	0.035	0.04	0.06	0.08	0.11	0.15	0.19	0.24
S6	0.0005	0.004	0.03	0.06	0.10	0.19	0.24	0.28
S7	0.1	0.22	0.39	0.50	0.55	0.59	0.63	0.71
S8	0.003	0.01	0.09	0.25	0.31	0.36	0.42	0.56

Simulations: Results

Accuracy







Conclusions

- CRM is more accurate
- Reaches MTD faster
- Treats more patients at or around MTD
- It is as safe (overdose control etc)
- Does not require larger sample – sample size depends on number of levels

Sample Size

- 15 for 3 levels
- 20-25 for 5-8 levels
- Since we reach the MTD faster, we could test more dose levels or treat more pateints near the MTD

Summary

- It is not complex
- It can handle more than just the simplest clinical problem
  - Binary outcome
  - Time to event outcome or ordinal outcome (grades)
  - Drug combinations,
  - Patient heterogeneity,
  - Bridging studies

Which design to use in practice?

Model based vs non model based

- Literature Review
    1. O' Quigley, Pepe, Fisher (1990)
    2. Goodman et al. (1995)- Group inclusions
    3. Shen, O'Quigley (1996) : Asymptotic distribution
    4. O'Quigley, Shen (1996): MLE – Frequentist approach
    5. Moller (1995) Comparisons
    6. Babb et al.(1998) Escalation with Overdose Control
    7. Heyd, Carlin (1999) Adaptive Design improvements
    8. Leung, Wang (2002) Extensions
    9. Rosenberger, Haines (2002) Competing designs
    10. Cheung YK, Chappell R (2002): Model sensitivity
    11. Garrett- Mayer (2007): Review
    12. Iasonos et al (2008): Comparative Review
    13. ….

Original CRM with binary response is often sufficient

Case studies

Motivation

- Model based designs such as CRM have been reviewed extensively via simulations
- Simulations tell you how close you are to the truth
- Average behavior vs within trial allocation
- Some examples in the literature of counterintuitive dose escalations

Aims

- To understand and avoid counterintuitive dose escalations or mistakes in the design set-up in practice

- Convince clinicians of the operating characteristics of these designs usign real trials

Systematic Review

- Review the literature to find published, completed Phase I trials in last 10 years that used these designs
- Model based designs:
  - Continual Reassessment Method
  - CRM, EWOC-CRM, TITE-CRM.

Systematic Review
Iasonos and O'Quigley, JCO 2013

- 53 trials (Jan 2003 to Sept 2013)
- Quantitative Review:
  - safety, patients to dose allocation
- Qualitative Review:
  - are they flexible
  - what is the clinical question
  - how do they deal with different schedules,
  - patient populations, drug combinations

Study Design: characteristics

- Enrolled 35 patients,
- Evaluable for DLT 25 patients
- 25 months, tested 5 dose levels,
- targeted an acceptable toxicity rate of 26% (range 10-33%).
- DLT timeframe 38 days (median= 28 days),
- 53% single agent regimen;
- 45% combination regimen.

Safety – based on review

- 19% DLT rate (target varied from 15-33%)

Treated below	Treated at MTD	Treated above
41%	39%	19%

- 75% treated within MTD +/- level

- 4 pts treated above MTD (19% of 20)

Qualitative Review

Supplemental material:

Supplemental Table 1: Qualitative review of trials: patient population, DLT definition, trial design and model parameters

Study: Trial Author, Year Aim Number of groups/schedules	Single agent (S) or combination regimen (C)	DLT definition (Endpoint); DLT timeframe N: number of enrolled patients (evaluable) Levels: number of dose levels	Type of Design Model Parameters	Comment
Thornton KA, 2013[28] **Aim:** MTD of Temsirolimus and liposomal doxorubicin for patients with soft tissue sarcoma **No of groups/schedules:** 2 groups - children and adults	C	**DLT:** Fatal toxicity gets score 1; reversible Gr 4 toxicity as 0.5; reversible Gr 3 toxicity as 0.25 **DLT timeframe:** 56 days N=15 **Levels:** 5	CRM with graded toxicities [1, 6] Rate=20% Cohort size: 3 Stopping rule: 6 patients at current level and no change in MTD Starting level 3	
Harvey RD, 2013[29] **Aim:** MTD of Bortezomib (B) and sunitinib (Sun) in patients with sold tumors **No of groups/schedules:** 1	C	**DLT:** Gr 4 neutropenia, anemia or thrombocytopenia, Gr 4 fatigue or 2 point decline in ECOG, Gr ≥3 gastrointestinal, any Gr ≥3 AE **DLT timeframe:** 42 days (6 weeks) N=31 (30) **Levels:** 7	CRM –EWOC (2 stage) [71] Alpha parameter varies Rate=33% Cohort size: 1 Escalation restricted: EWOC	The study escalated each drug. First it found the MTD of Sun with fixed B and then B was increased with fixed Sun. The toxicity ordering is not clear thus patients treated above MTD cannot be reported.
Ben-Josef E, 2012[30] **Aim:** MTD of Radiation + fixed dose gemcitabine in patients with pancreatic cancer **No of groups/schedules:** 1	C	**DLT:** Gastrointestinal toxicity Gr ≥3, neutropenic fever, deterioration in performance status to ≥3 **DLT timeframe:** Day 1-126 (13 weeks) N=51 **Levels:** 6	TITE-CRM (2 p logistic) [72] Rate=25%; Cohort size: 1 Escalation restricted: 1 level	

Design Type – modifications of CRM

TYPE OF DESIGN	N=53
CRM (O'Quigley 1990)	23/53
TITE CRM (Cheung 2002)	8/53
CRM with continuous dosing (Piantadosi 1998)	9/53
EWOC (Babb, Rogatko et al, 1999)	12/53
Lower grades (Goodman 1995)	1/35

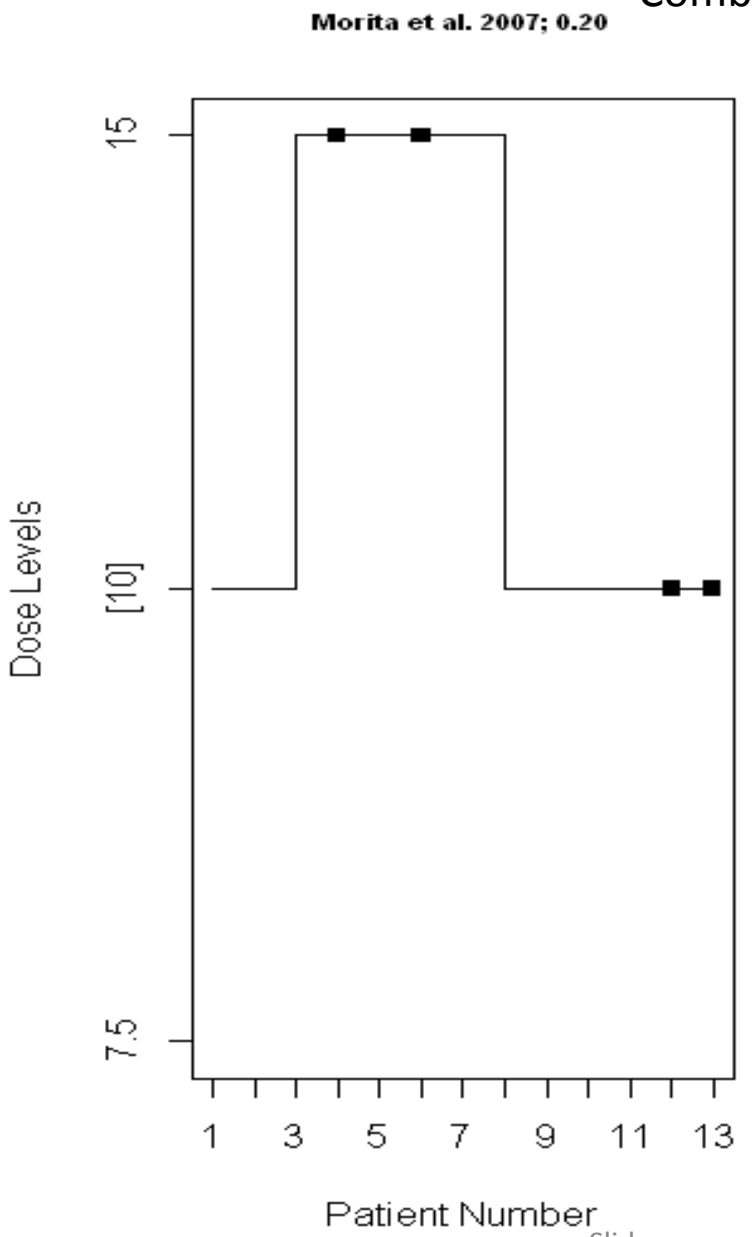
- TITE CRM: deals with late on set toxicities (radiotherapy, targeted agents with late on-set toxicities)
- EWOC: escalation with overdose control (Chu P, Lin Y, Shih WJ 2009)
- Accelerated 1st stage (doubling dose- 1 pt per cohort; gr 2 AE)

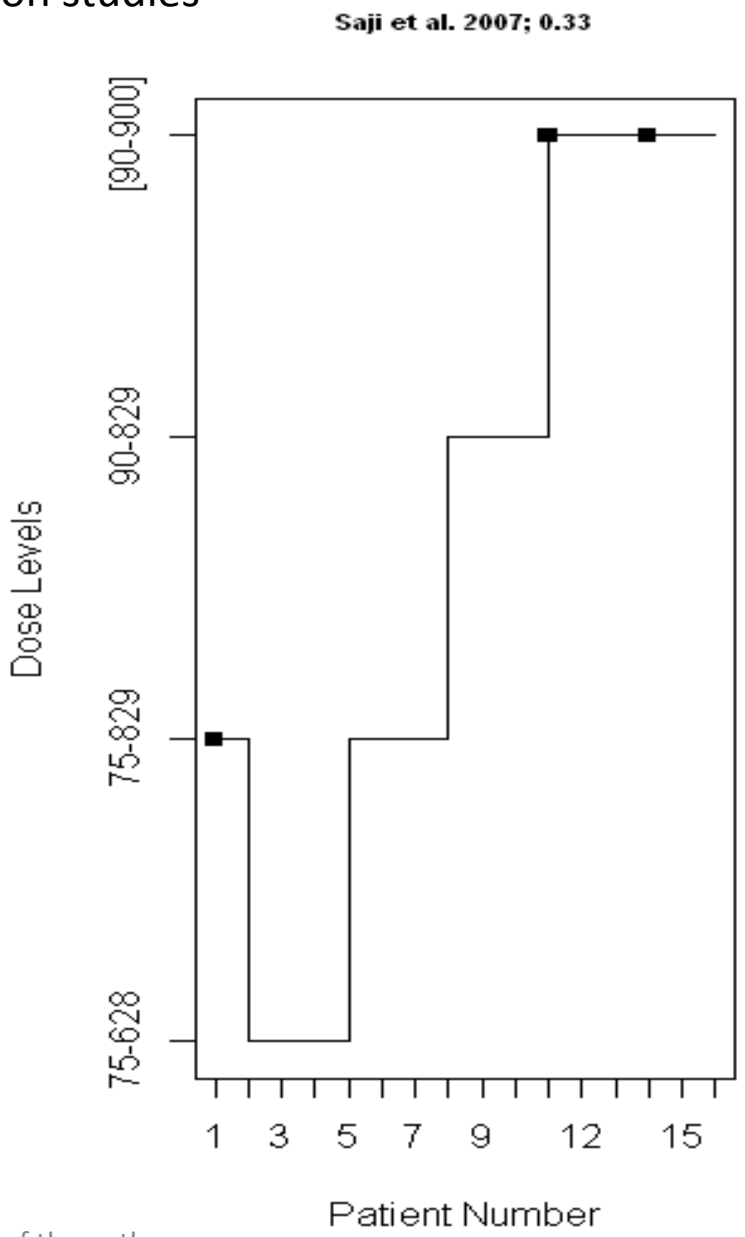
Iasonos and O'Quigley, JCO 2013

Specific trials:

Can escalations be counterintuitive or unsafe?

Combination studies

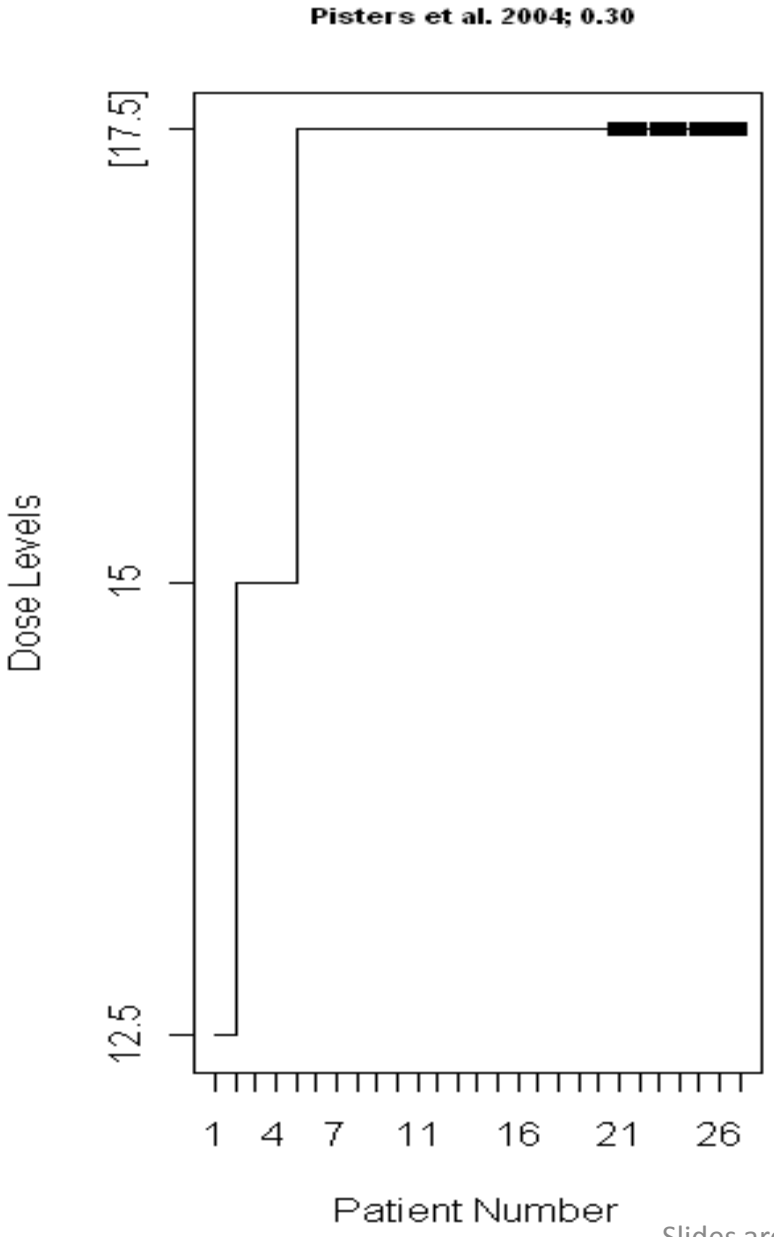


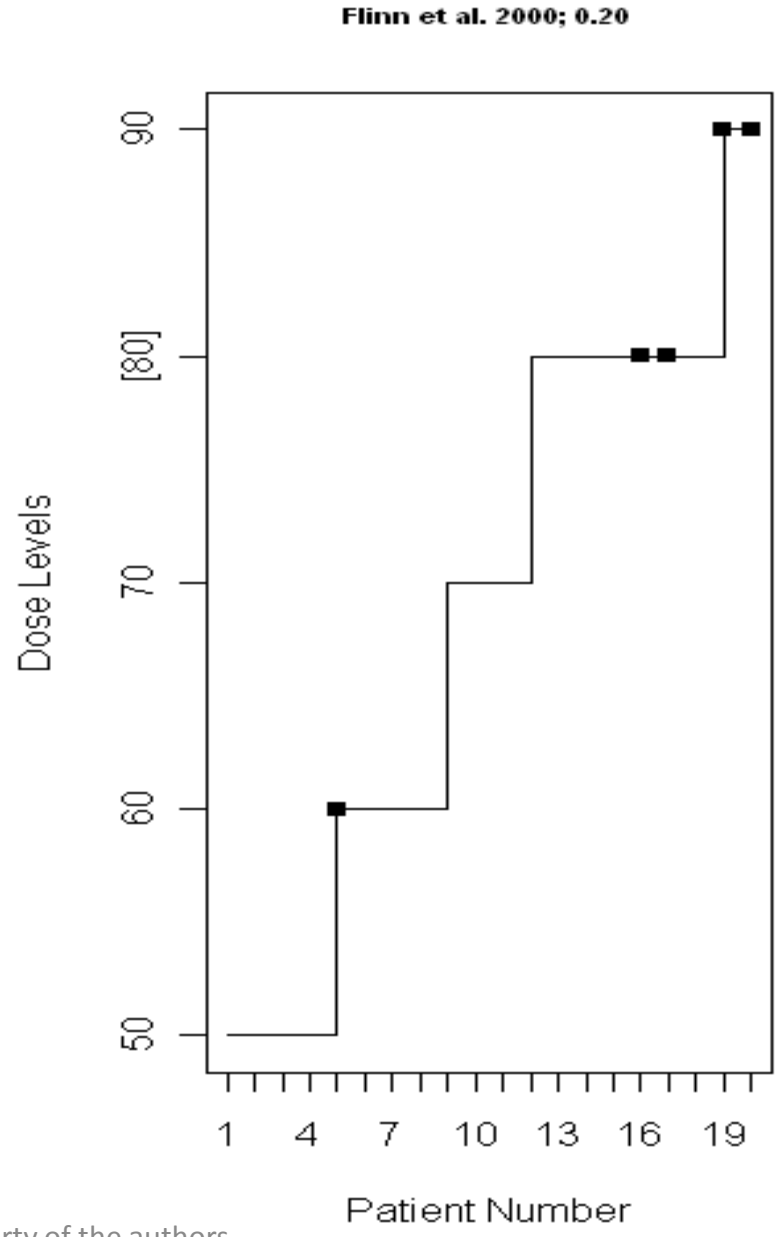


Observations:

- Each study targets a different threshold for acceptable toxicity
  - 30-33%
  - 20-25%

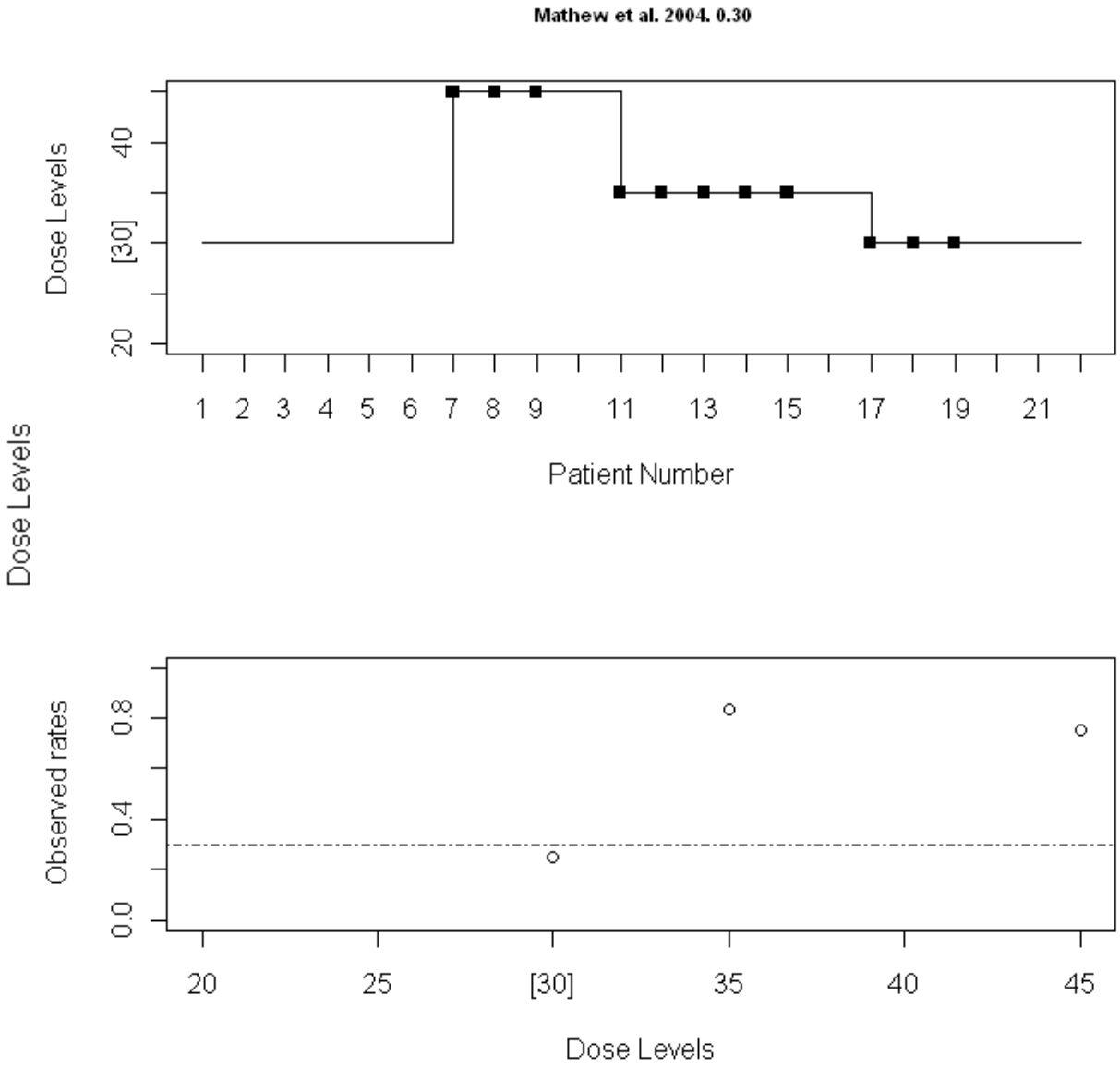
- Implications of the threshold
- Efficient in terms of sample size
- Explored drug combinations with 16 pts





3 Trials where things went wrong

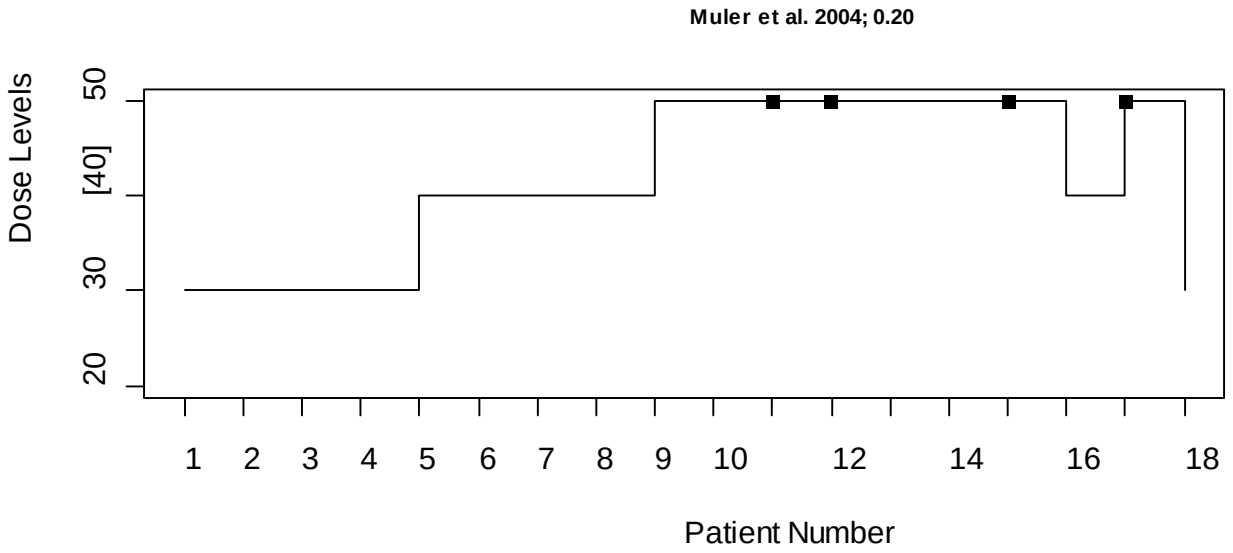
- imatinib and docetaxel in prostate cancer patients, where 8 out of 10 patients experienced a DLT above the MTD. [Mathew, P., et al., *Platelet-derived growth factor receptor inhibitor imatinib mesylate and docetaxel: a modular phase I trial in androgen-independent prostate cancer.* J Clin Oncol, 2004. **22**(16): p. 3323-9.]
- dose escalation study of cisplatin with gemcitabine in pancreatic cancer, 50% (4/8) of patients treated above the MTD experienced DLTs. [Muler, J.H., et al., *Phase I trial using a time-to-event continual reassessment strategy for dose escalation of cisplatin combined with gemcitabine and radiation therapy in pancreatic cancer.* J Clin Oncol, 2004. **22**(2): p. 238-43.]
- A CRM trial where the recommendation was to escalate after observing 2 DLTs out of 2 patients treated at a level [Neuenschwander, B., M. Branson, and T. Gsponer, *Critical aspects of the Bayesian approach to phase I cancer trials.* Stat Med, 2008. **27**(13): p. 2420-39.]

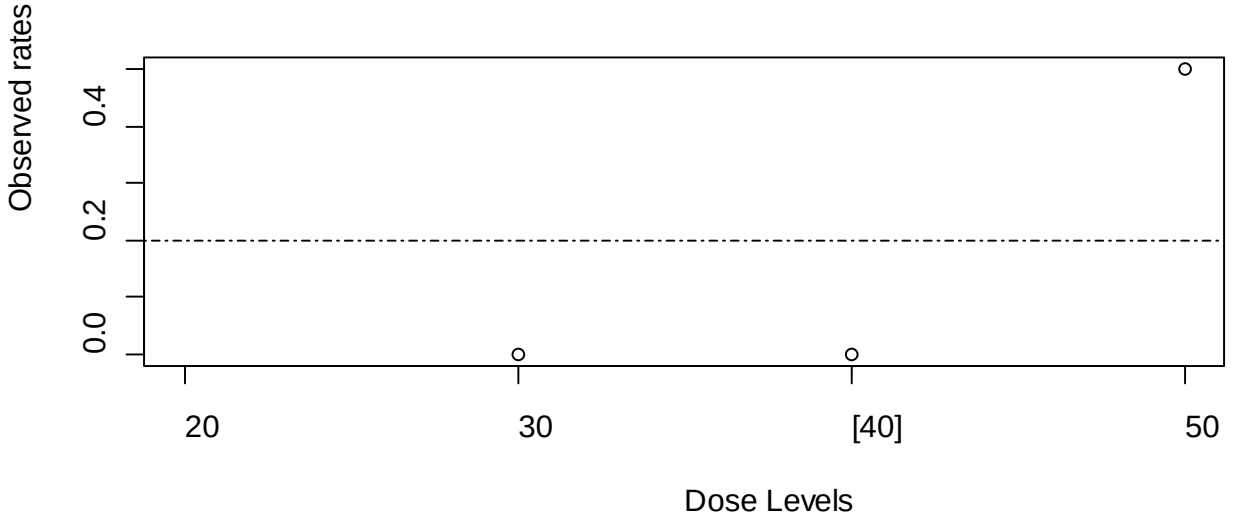


Points to consider

- Cohort size – 6 can be problematic
- Advantage of CRM is to react to current data; Outcome adaptive designs
- Accrual Rate vs DLT timeframe
- The faster the accrual rate, the higher the number of inevaluable patients for toxicity

- Increments in actual dose escalation is arbitrary
- Skipping dose levels





Muler et al 2004 study

- 19 patients enrolled in 15 months, DLT timeframe 9 weeks.
- While waiting for 9 week interval; what to do when a new patient is ready to enroll?

  Wait – better estimate of MTD at each step; conservative; longer trial (27 mos with 3+3)

  Treat- more aggressive ; shorter trial (15 mos)

Points to consider

- Observed rates of 0 and 100% or 0 and 50%
- Target rate 20 (1/5) or 33% (1/3); which dose is closer?

- Use model-recommendation as a GUIDE – auxiliary information
    - Toxicity onset, duration, reversible, patient's status
    - Toxicity observed in subsequent cycles or lower grade

Neuenschwander, B. 2008 Stat Med

- Data 0/16 at levels 1-4, 2/2 at level 7, CRM recommends Dose 9.
- Model skeleton $P(DLT)=p_i^{\exp(a)}$
- Prior $N(0,1.34^2)$

- Prior weight in parameter space (exp or Normal prior) vs probability scale (DLT rate); 56% on d_{10}
- Pseudo data showed that the prior corresponds to 2/17 DLTs at level 7 further supporting d_7 as safe.
- Likelihood vs Bayesian (coherent if model has been followed from trial onset; problems at the seam)

Table I. Posterior summaries for probabilities of DLT (CRM).

	Doses									
	1	2.5	5	10	15	20	25	30	40	50
No. of patients	3	4	5	4	—	—	2	—	—	—
No. of DLTs	0	0	0	0	—	—	2	—	—	—
(*A*) *Posterior summaries* (*original skeleton*)										
Skeleton (CRM)	0.010	0.015	0.020	0.025	0.030	0.040	0.050	0.100	0.170	0.300
Mean	0.069	0.085	0.099	0.111	0.123	0.144	0.163	0.242	**0.330**	0.465
Std. dev.	0.055	0.062	0.068	0.072	0.076	0.082	0.087	0.101	0.109	0.108
(*B*) *Posterior summaries* (*equidistant skeleton*)										
Skeleton (CRM)	0.063	0.125	0.188	0.250	0.313	0.375	0.438	0.500	0.563	0.625
Mean	0.024	0.054	0.090	0.130	0.176	0.226	**0.281**	0.341	0.405	0.475
Std. dev.	0.030	0.051	0.069	0.084	0.097	0.107	0.115	0.119	0.120	0.117

Columns in boldface highlight the recommended dose for the next cohort.

Supplemental Table 3: Case Study, Trial 3 as described in Supplemental Appendix A.1

Patient	Dose Treated (mg) (Level)	DLT	Actual Dose recommendation (mg) (Level) Initial curve	Updated DLT rate at dose 50mg	Correct Dose recommendation (mg) (Level) Correct curve	DLT
Prior/ Initial			50 (L10)	30	**1 (L1)**	
1	1 (L1)	No	50 (L10)	21	**1 (L1)**	No
2	1 (L1)	No	50 (L10)	17	**5 (L3)**	No
3	1 (L1)	No	50 (L10)	15	**15 (L5)**	No
4	2.5 (L2)	No	50 (L10)	13	**20 (L6)**	No
5	2.5 (L2)	No	50 (L10)	11	**25 (L7)**	Yes
6	2.5 (L2)	No	50 (L10)	10	**15 (L5)**	No
7	2.5 (L2)	No	50 (L10)	9	**20 (L6)**	No
8	5 (L3)	No	50 (L10)	8	**20 (L6)**	No
9	5 (3L)	No	50 (L10)	8	**25 (L7)**	Yes
10	5 (L3)	No	50 (L10)	7	**20 (L6)**	No
11	5 (L3)	No	50 (L10)	7	**20 (L6)**	No
12	5 (L3)	No	50 (L10)	6	**25 (L7)**	Yes
13	10 (L4)	No	50 (L10)	6	**20 (L6)**	No
14	10 (L4)	No	50 (L10)	6	**20 (L6)**	No
15	10 (L4)	No	50 (L10)	6	**20 (L6)**	No
16	10 (L4)	No	50 (L10)	5	**25 (L7)**	Yes
17	25 (L7)	Yes	40 (L9)	38	**20 (L6)**	No
18	25 (L7)	Yes	40 (L9)	47	**20 (L6)**	No
Recommended dose			**40 (L9)**		**25 (L7)**	

Footnote: Initial curve assigns 30% DLT rate at level 10 (50mg) and very low rates at all remaining levels (rates for each respective dose level: 0.01,0.015,0.02,0.025,0.03,0.04,0.05,0.1,0.17,0.3);

Correct curve assigns 30% initial DLT rate (prior to seeing the data) at dose 1 so that experimentation starts at dose 1. Rates under correct curve for each dose level: 0.30, 0.40, 0.48, 0.56, 0.64, 0.72, 0.80, 0.88, 0.92, 0.99,)

Points to consider

- Prior can be informative
- 2 parameter logistic model is not more flexible – non identifiable; performs worse even when the data are generated by 2 parm model; can get stuck
- One source of information – point mass at MTD cannot fit two parameters (slope and intercept)
- Iasonos et al 2016, Stats Med

Conclusions

- These applications and review of case studies confirmed results of simulated trials reported in the statistical literature
- The method is safe and efficient (patient to treatment allocation)
- Method is rigid once parameters are selected but it is flexible to deal with clinical problems
  through the choice of tuning parameters
- Care is needed for selecting design parameters: prior