Adaptive trial design in oncology clinical trials: The state of the art and some innovations in dose finding strategies

Elizabeth Garrett-Mayer

Emily Van Meter

September 28, 2009

What is an adaptive trial?

- *Adaptive clinical trials* are designed to use accumulating information to determine how to modify the trial as it progresses.
- Adaptive aspects of the trial are predetermined before the trial begins and are a part of the planned design
- Modifications in an adaptive trial must maintain the *validity* and *integrity* of the trial
  - Validity: providing correct inference (e.g., adjusted estimates of confidence intervals and p-values)
  - Integrity: providing convincing evidence to the scientific community

Level of adaptation

- Adaptive designs occur in stages
- At the end of each stage, a possible adaptation of the trial may occur.
- Examples:
  - Simon two-stage design: two stages
  - Continual reassessment method: Continuous staging, adaptation may occur after each cohort (of size ≤ 3) is enrolled or observed for outcomes.

Why oncology?

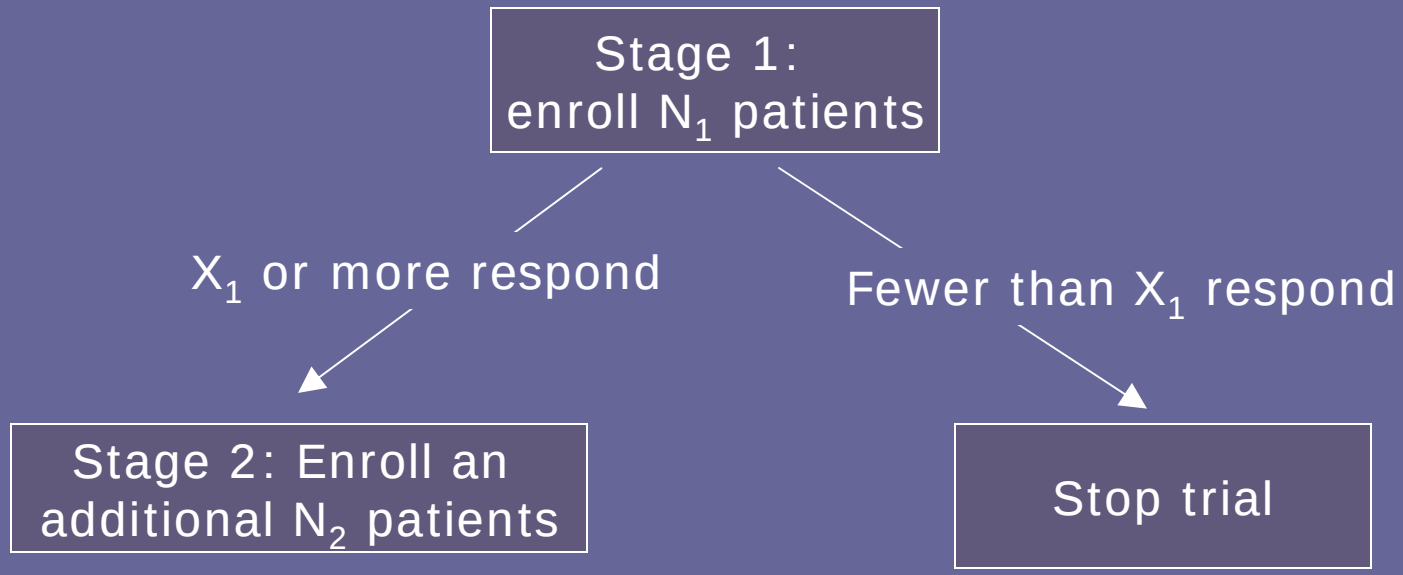
- **Candidate treatments** considered for clinical trials have increased in greater numbers than the number of patients eligible for trials in recent decades creating a demand for greater effiency in trial design.
- Many new agents are **targeted therapies** which may only be tested in patients who have a specific cancer subtype, or particular genetic mutation or protein overexpression.
- **Patient-advocates** are extremely valuable in representing the patient perspective in trial design and have been vocal in their preferences for adaptive designs in oncology.
  (Perlmutter, 2007; Berry, 2004).

Bayesian?

- Sometimes, but not always.
- Bayesian is the trend so it is a misconception that adaptive trials are Bayesian.
- Partly, semantics: what does Bayesian mean?
  - is it the formal use of a prior distribution coupled with a likelihood to create a posterior distribution?
  - or is it the idea of using preliminary information to update your knowledge?

Example: Simon Two Stage design

- Allows early stopping for futility
- If, after a subset of patients have been treated, the response rate is so low that it is unlikely you will reject the null at the end, you stop the study
- Preserves alpha and beta. Usually only minor increase in N compared to a “single stage” approach



What do you think?

- *Wikipedia*: "**Bayesian inference** is statistical inference in which evidence or observations are used to update or to newly infer the probability that a hypothesis may be true."
- Despite
  - Simon is a frequentist
  - Neyman-Pearson hypothesis testing framework
  - type I and type II errors
  - ....it may be Bayesian afterall!
- But, when people distinguish between Bayesian adaptive designs and others, they are *usually* referring to the formal use of priors*likelihoods = posteriors

Are they everywhere?

- Not even close
- A few reasons (simon pointed these out back in 1977, but they are still true today):
  - most clinical trials do not have just one goal
    - e.g. efficacy could be equivalent in two treatments but one would be considered superior if safety outcomes were better
  - seemingly impossible to move away from “rejecting the null” idea. without p-values, people don’t know how to behave!
  - convincing non-statisticians in clinical research is not easy
    - they are harder to understand
    - some concern over letting the computer decide how to treat patients

PhRMA: Adaptive Designs Working Group

- Pharmaceutical Research and Manufacturers of America formed the Adaptive Designs Working Group (ADWG) in 2005.
- The goal of this group are to facilitate and foster:
  - wider usage
  - regulatory acceptance
  - clinical development
- The approach: fact-based evaluation of the benefits and challenges of these types of designs
- Published numerous papers on the topics: Journal of Biopharmaceutical Statistics devoted an entire issue to manuscripts from ADWG.

(Gallo, 2006; J of BioPharm, v 17(6), 2007).

Some ways to adapt trials

- Adaptive Allocation
  - Covariate-Adaptive
  - Response-Adaptive
- Adaptive Sample Size and Stopping
- Adaptive Dose Escalation

Aside: randomization

- What does it mean for a study with *k* treatment arms to be randomized?
- It does NOT mean that patients have an equal chance of being assigned to each arm in a study
- Randomness:
  - there is some probability ($0 < p < 1$) of assignment to each arm.
  - that is, it is not “deterministic” (by patient, physician, or any other means)
  - p can vary across arms!

Adaptive Allocation

- Assign patients to treatment arms based on information collected on patients already in the trial

- Covariate-adaptive: The goal is to achieve balance of covariate factors across the arms so that comparisons of treatment effects will untainted by confounding.

- Response-adaptive:
  - a new patient is assigned to a treatment arm based on the relative success of the patients treated thus far on the trial
  - patients are more likely to be randomized to the arm with greatest success (e.g., arm with the highest response rate).

Response Adaptive Randomization

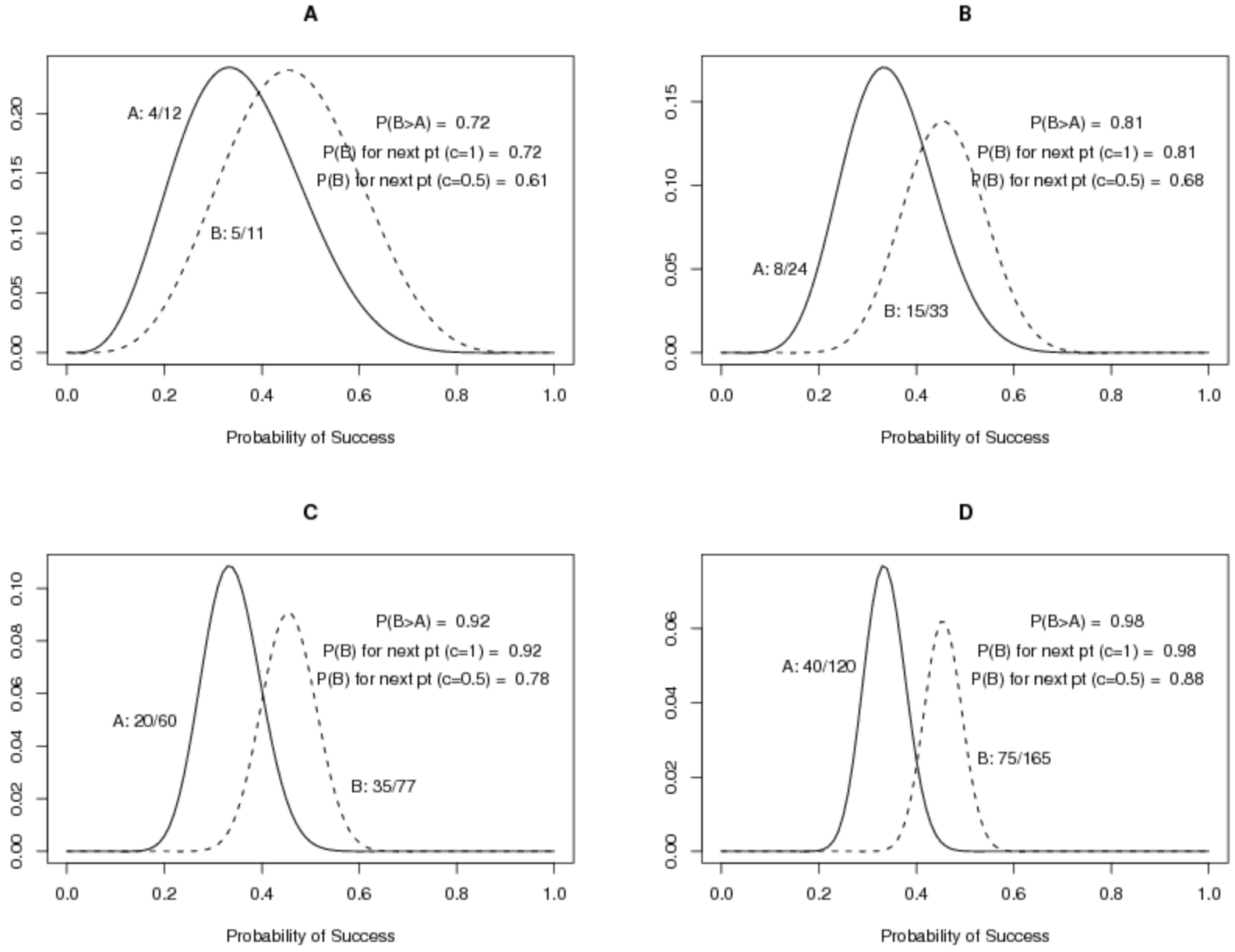
- Frequentist approaches: Karrison et al (2003), Zelen (1978), and Hu and Rosenberger (2006)).
- Thall and Wathen (2007) : Randomized phase II trial of gemcitabine (G) alone versus gemcitabine plus docetaxel (GD) in patients metastatic soft tissue sarcomas by (Maki et al. (2007).
  - success = tumor response (shrinkage by >30%)
  - failure = progressive disease

Simplified example

- Define θ_{G+D} = probability of response on G+D
- Define θ_{G} = probability of response on G

- Randomization probability:

$$p_{G+D} = \frac{P(\theta_{G+D} > \theta_G \mid data)^c}{P(\theta_{G+D} > \theta_G \mid data)^c + P(\theta_G > \theta_{G+D} \mid data)}$$


A
A: 4/12
B: 5/11
P(B>A) = 0.72
P(B) for next pt (c=1) = 0.72
P(B) for next pt (c=0.5) = 0.61
Probability of Success
B
A: 8/24
B: 15/33
P(B>A) = 0.81
P(B) for next pt (c=1) = 0.81
P(B) for next pt (c=0.5) = 0.68
Probability of Success
C
A: 20/60
B: 35/77
P(B>A) = 0.92
P(B) for next pt (c=1) = 0.92
P(B) for next pt (c=0.5) = 0.78
Probability of Success
D
A: 40/120
B: 75/165
P(B>A) = 0.98
P(B) for next pt (c=1) = 0.98
P(B) for next pt (c=0.5) = 0.88
Probability of Success

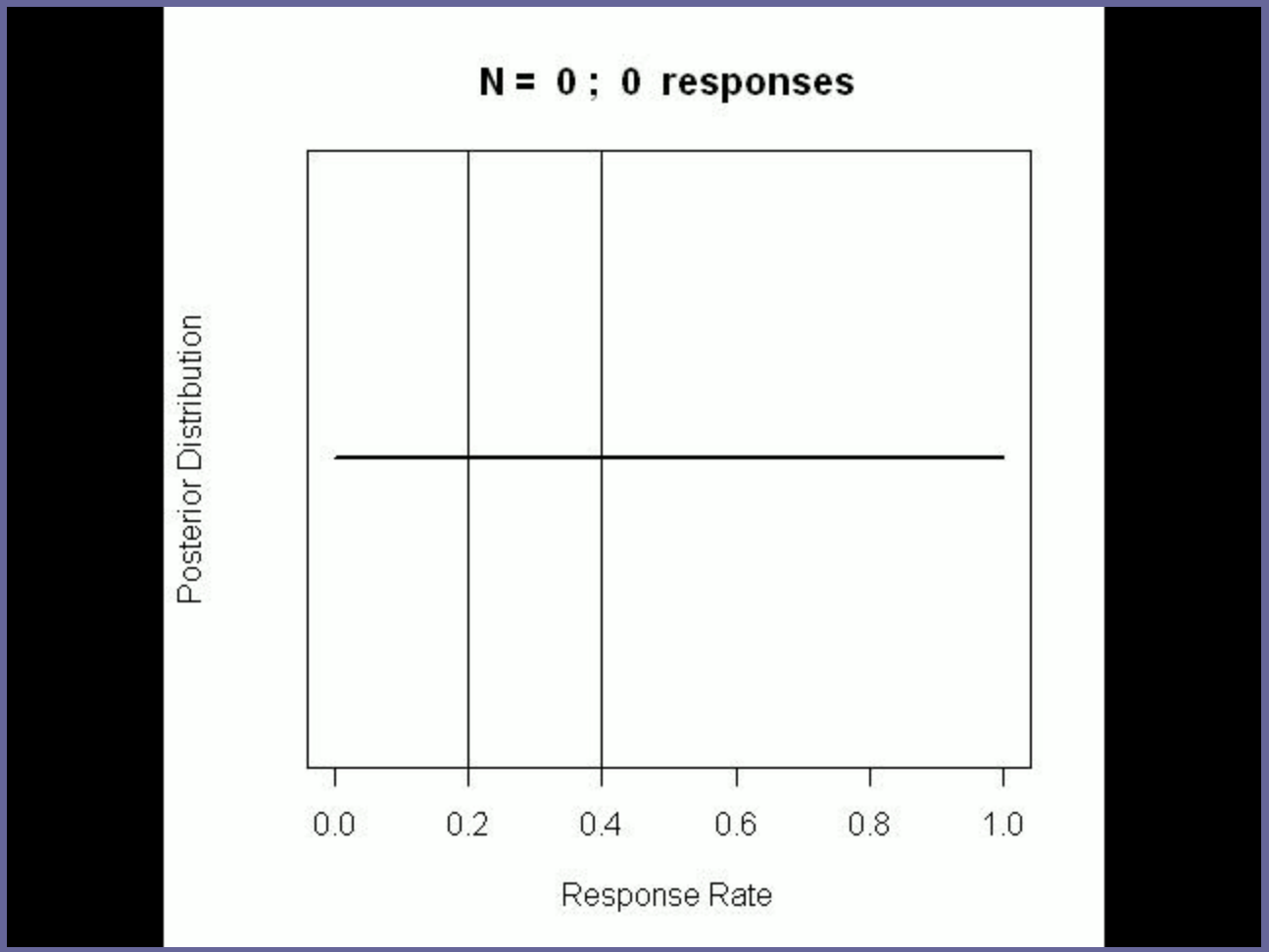
Adaptive Sample Size

- Lots of examples: any design that incorporates early stopping rule for futility, superiority, safety is technically an adaptive sample size design
  - Simon two-stage
  - Group sequential
  - any trial with alpha spending
  - predictive probability
- More complex: sample size calculations are performed as an interim analysis
- Recent NEJM example: Muss et al. Adjuvant Chemotherapy in Older Women with Early-Stage Breast Cancer. v 360(20):2055-2065, May 14, 2009.

Simple example

- Monitor if enough evidence has accumulated to accept/reject a hypothesis
- Example:
  - single arm phase II study of cancer treatment
  - outcome is response
  - Two hypotheses:
    - response rate = 0.20
    - response rate = 0.40
  - Frequentist approach: with alpha = beta = 0.10 requires 39 patients.
  - Do we need to complete the trial to decide?

One simple example: Single arm phase II study with null p = 0.20; alternative p = 0.40

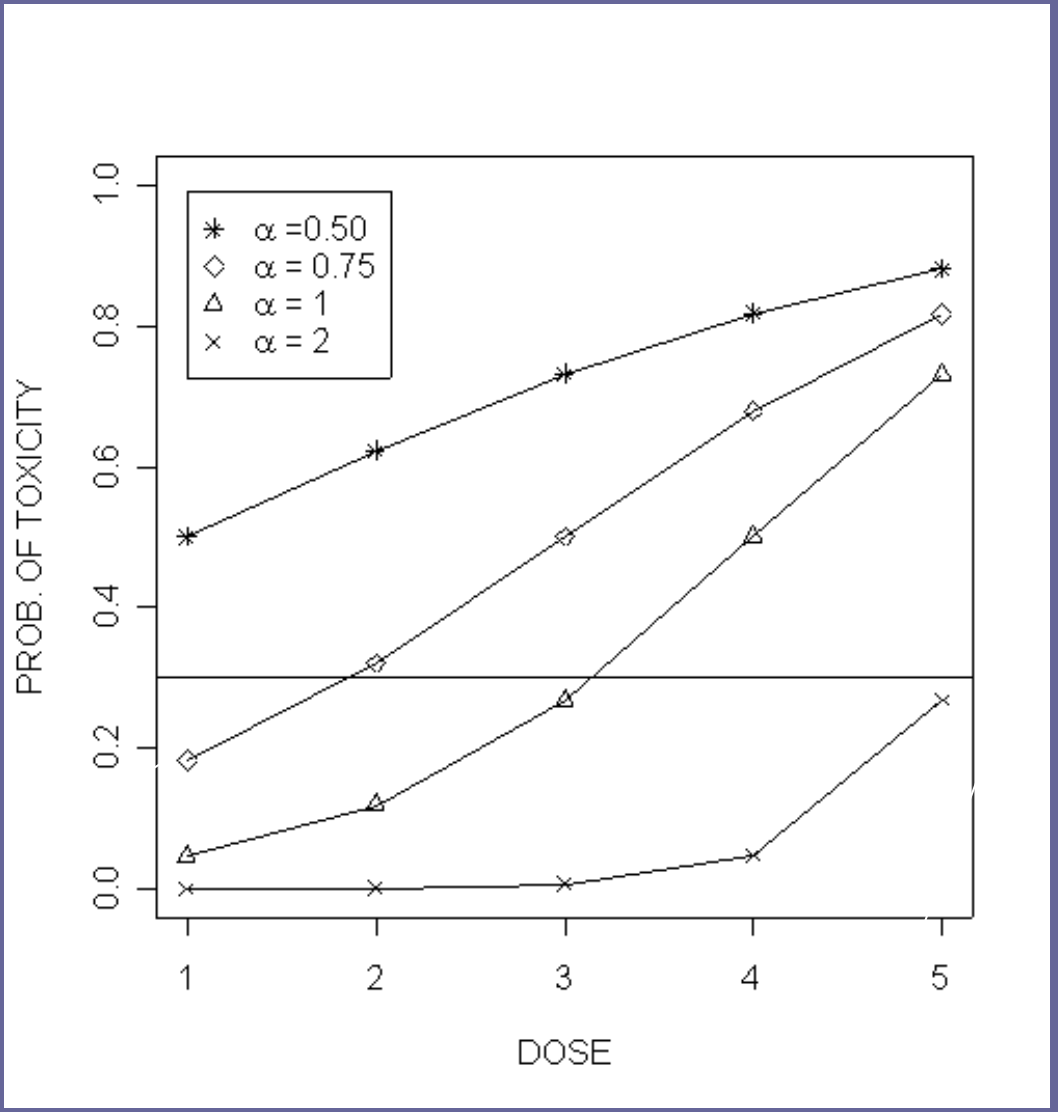


Adaptive Dose Finding

- Most popular adaptive dose finding design is the Continual Reassessment Method (CRM)

- Originally devised by O'Quigley, Pepe and Fisher (1990) **where dose for next patient was determined based on toxicity responses of patients previously treated in the trial**
- Purely Bayesian design
  - Choose a mathematical model (likelihood)
  - Choose a prior distribution
  - Estimate the posterior distribution of parameters of interest
- Find dose that is most **consistent with desired toxicity rate**
  - **need to define what IS a toxicity**
  - **need to choose an acceptable toxicity rate**

Example:



One-parameter logistic model

$$p(toxicity|dose = d_i) = \frac{\exp(3 + \alpha d_i)}{1 + \exp(3 + \alpha d_i)}$$

(where d=dose-7)

What are the goals?

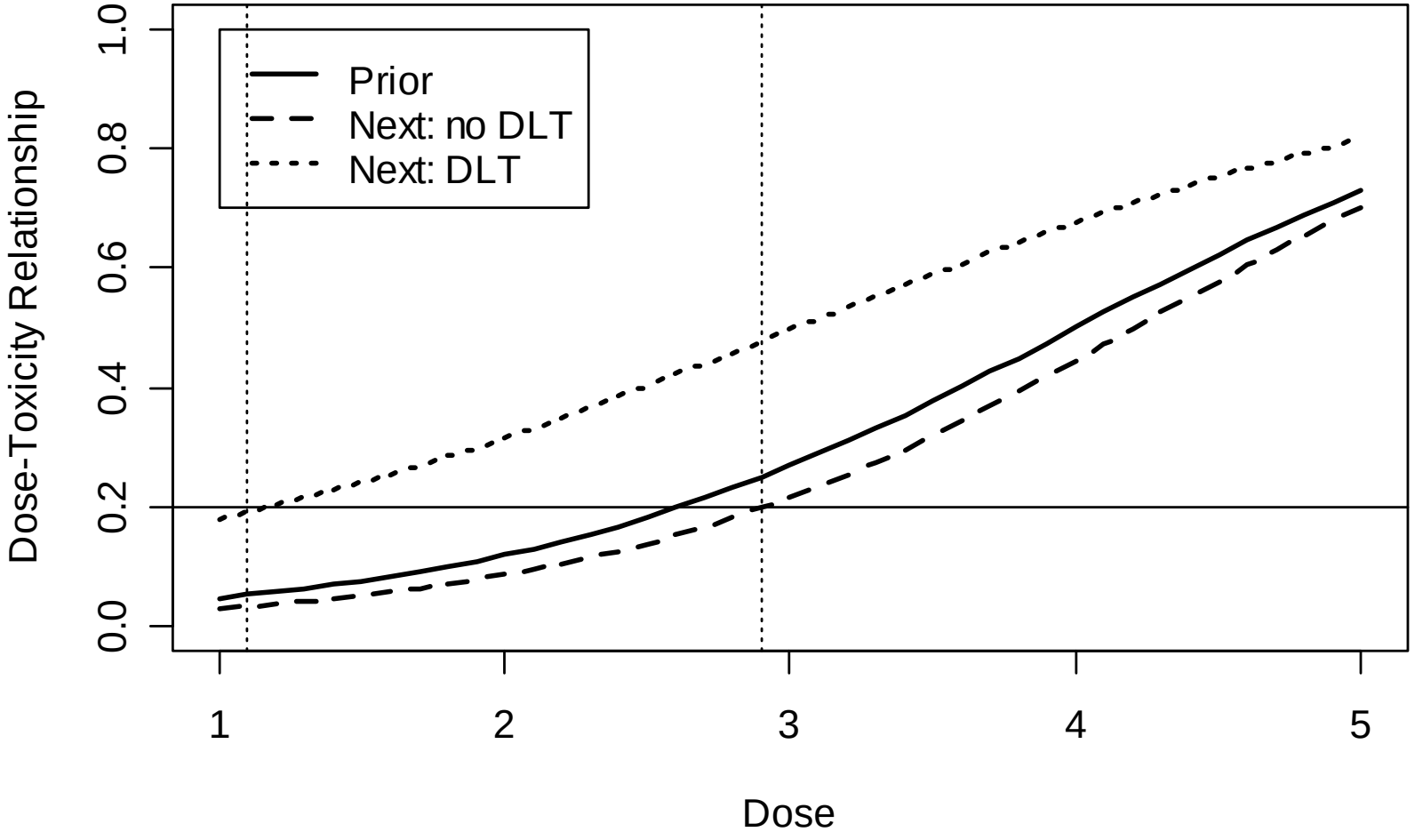
- 1. Find alpha
  - What is the alpha that is most consistent with the model?
  - Recall: Bayesian
    - Prior on alpha
    - Estimate likelihood
    - Find “best” alpha using posterior
- 2. Find the dose for the next patient
  - After alpha is estimated
  - Plug alpha “hat” in model
  - Find dose that is consistent with desired DLT rate

- Notes:
  - doses can be continuous or discrete increments
  - actual doses are “mapped” to another scale

Prior

- VERY IMPORTANT
- **Prior has large impact on behavior early in the trial**
- Requires a lot of simulations in the planning stages of the trial to see how the design will behave under a variety (or all!) of the possible scenarios

Scenarios


Prior
Next: no DLT
Next: DLT
Dose-Toxicity Relationship
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
Dose

Theoretically: a beautiful design!

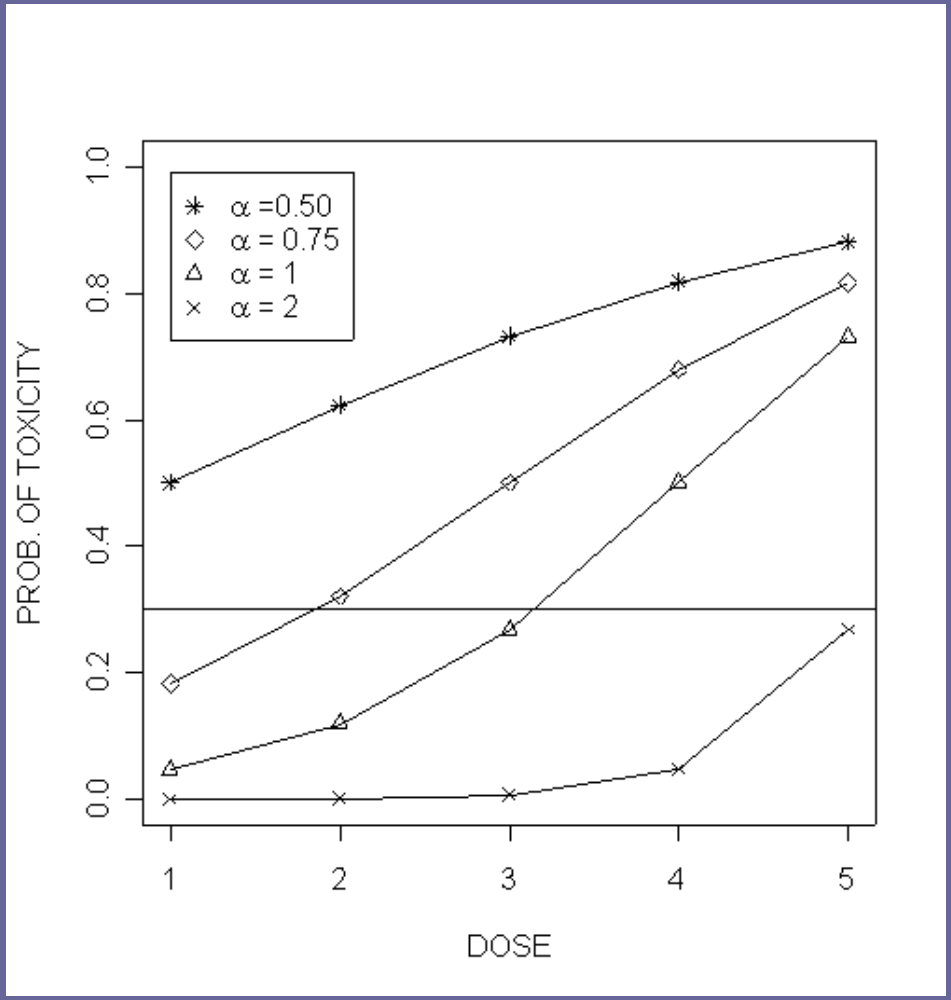
- BUT!
  - Concern over starting in mid-dose range
  - Concern over escalating without enough data
  - Concern over escalating too quickly
- Due to safety concerns, several authors developed variants
  - Modified CRM (Goodman et al. 1995)
  - Extended CRM [2 stage] (Moller, 1995)
  - Restricted CRM (Moller, 1995)
  - and others….
- Some variants are not Bayesian!

Modified CRM

(Goodman, Zahurak, & Piantadosi, Statistics in Medicine, 1995)

Carry-overs from standard CRM

- Mathematical dose-toxicity model must be assumed
- To do this, need to think about the dose-response curve and get preliminary model.
- **We CHOOSE the level of toxicity that we desire for the MTD**
- At end of trial, we estimate dose response curve.



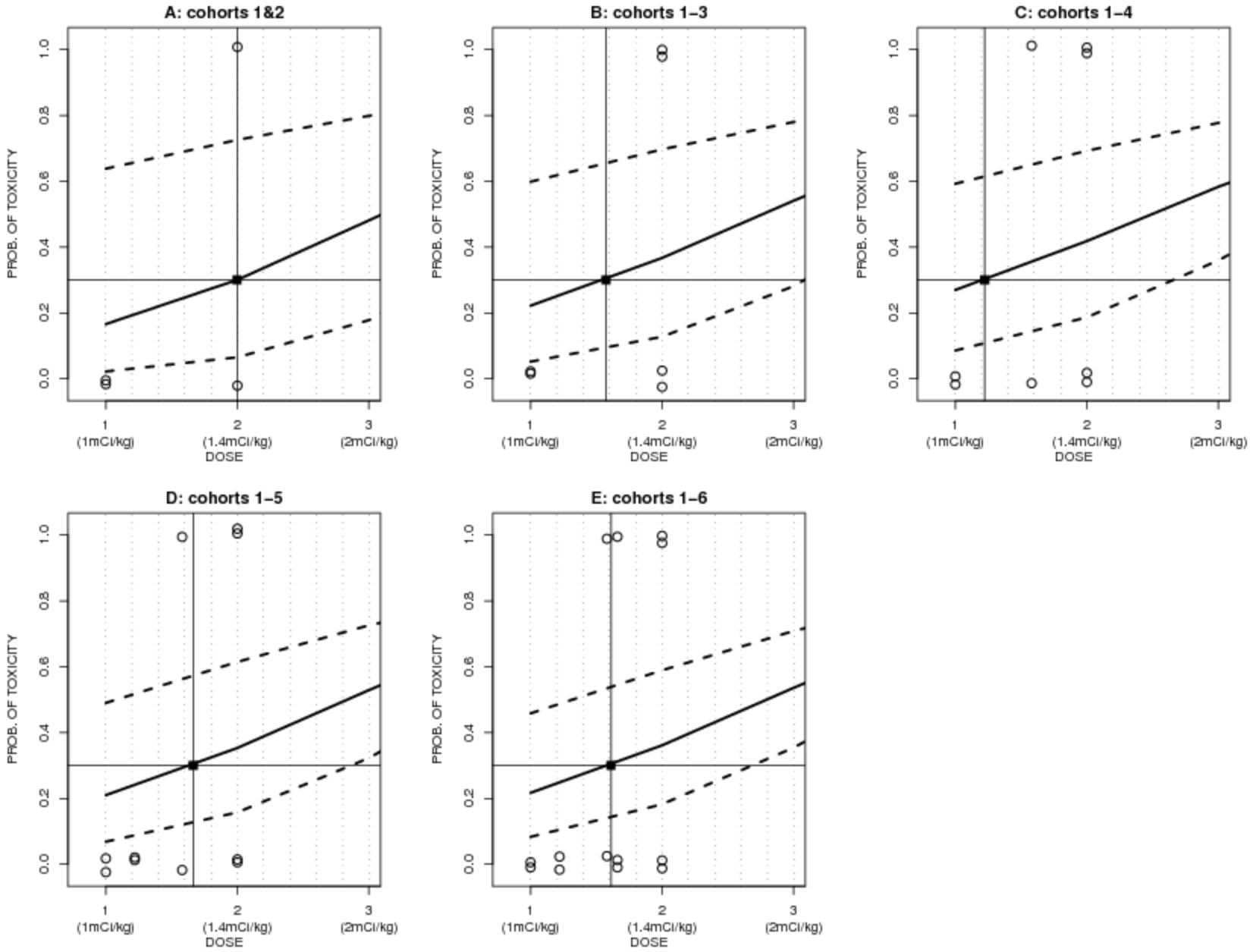
Modified CRM

(Goodman, Zahurak, & Piantadosi, Statistics in Medicine, 1995)

- Modifications by Goodman et al.
  - Use ‘standard’ dose escalation model until first toxicity is observed:
    - Choose cohort sizes of 1, 2, or 3
    - Use standard ‘3+3’ design (or, for example, ‘2+2’)
  - **Upon first toxicity, fit the dose-response model using observed data**
    - Estimate α
    - Find dose that is closest to desired toxicity rate.
  - **Does not allow escalation to increase by more than one dose level.**
  - De-escalation can occur by more than one dose level.

Real Example

- Shows how the CRM works in practice
- dose finding trial in a pediatric population of patients with high risk osteosarcoma to determine the maximum tolerated dose of 153Sm-EDTMP (Samarium).
- Target DLT rate was 30%
- Goodman's modified CRM
  - cohorts of size two
  - one-parameter dose toxicity model.
- The first dose was to be 1.0 mCi/kg
- dose increments increase by 40% up to a maximum dose of 4.0 mCi/kg.


A: cohorts 1&2
B: cohorts 1–3
C: cohorts 1–4
D: cohorts 1–5
E: cohorts 1–6
PROB. OF TOXICITY
0.0
0.2
0.4
0.6
0.8
1.0
1
(1mCi/kg)
2
(1.4mCi/kg)
3
(2mCi/kg)
DOSE