Objective calibration of the Bayesian CRM

Ken Cheung

Department of Biostatistics, Columbia University

King's College Aug 14, 2011

The other King's College

Phase I clinical trials

- Safety endpoint: Dose-limiting toxicity (DLT)
- Dose finding objective:
  - Consider a set of K doses with labels $d_1, d_2, \ldots, d_K$
  - Estimate the maximum tolerated dose (MTD):

    $$arg\ min_k \mid \pi(d_k) - p \mid$$

    where $\pi(x)$ is the probability of DLT at dose x and p is a pre-specified target, i.e., *percentile estimation*
- Sequential dose decisions

CRM

- **Continual Reassessment Method:** treat the next patient at dose level

$$arg\ min_k \mid F(d_k, b) - p \mid$$

  where $F(d_k, b)$ *is an estimate of* $\pi(d_k)$
- *Intuitive and "greedy"*
- Borrowing strength between doses
- *Flexible:* A coherent approach to contingencies via the model. What would the 3+3 rule do if

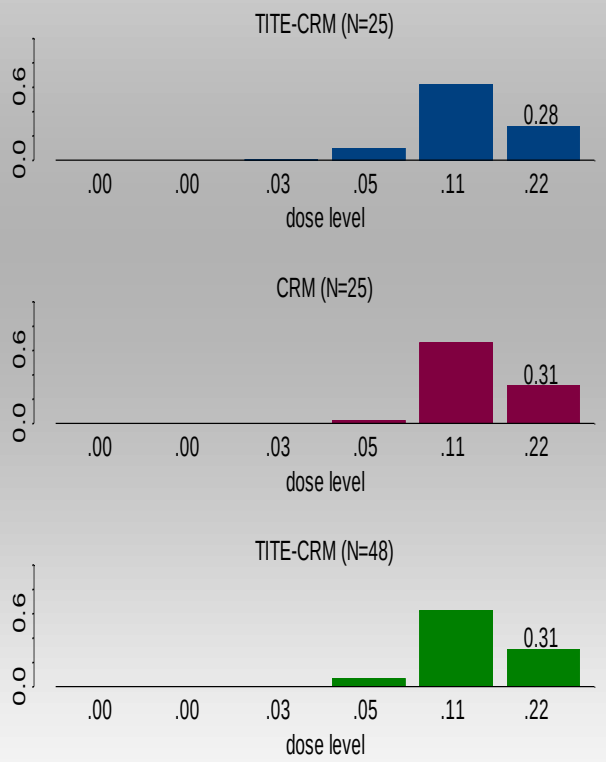
1/3 + 0/3 + **1/1** …?

- ***Assumption:*** The model is properly calibrated.

What may happen when the CRM is poorly calibrated

- Model violates consistency conditions under this true state of nature (Shen & O'Quigley, 1996)
- **Practical problem:** Specifying a (CRM) model can be a complex process … even for statisticians

Target DLT = 20%; MTD ν = 5



Outline

- Components of a **Bayesian CRM model**
  - Dose-toxicity function
  - **Initial guesses of DLT rates ("Skeleton")**
  - **Prior distribution of model parameter**
- **Example:** A bortezomib trial
- **Discussion**

CRM model

Three steps to specify a CRM model:

1. Dose-toxicity function $F(x, \beta)$ = P(DLT at dose x)
2. Choose a prior distribution $G(\beta)$ of β.
3. Evaluate the dose labels $\{d_1, d_2, \ldots, d_K\}$ for the K test doses via *backward substitution*:
   - Let p_{i0} denote initial guess of DLT rate for dose i. The dose labels d_i are obtained such that

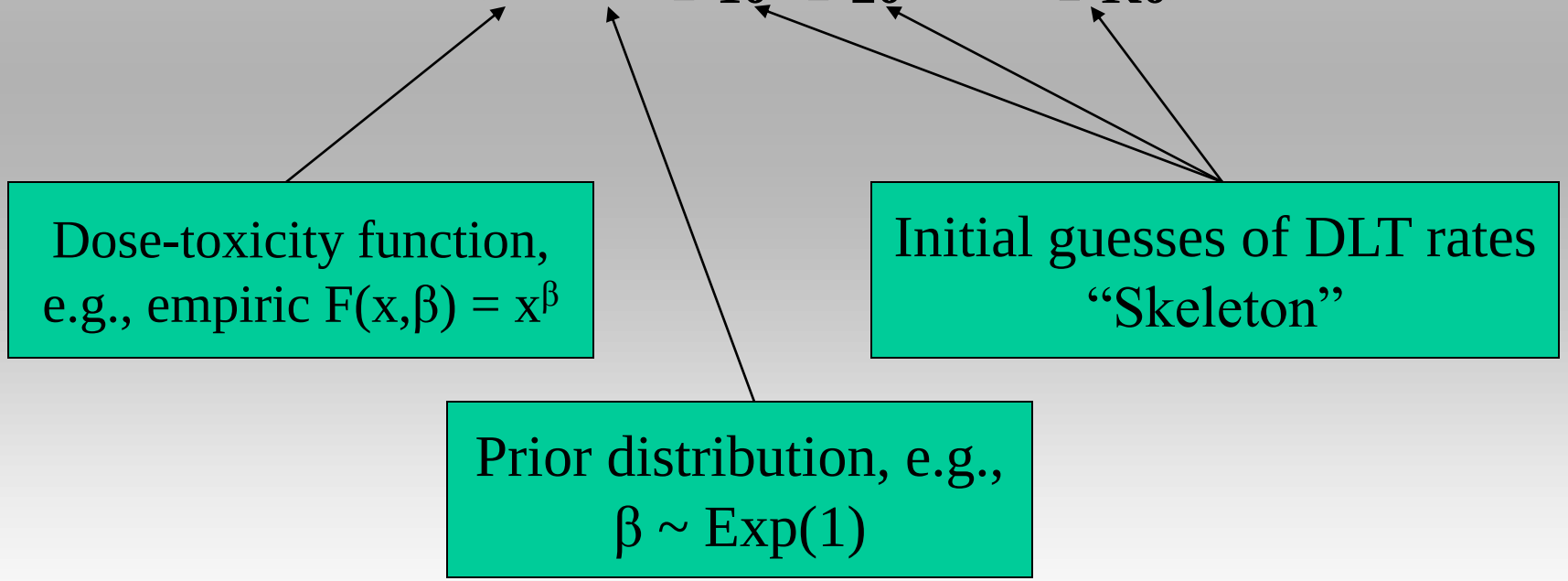
     $$F\{d_i, E_G(\beta)\} = p_{i0}$$

     where $E_G(\beta)$ is the prior mean of β.

CRM model

- Thus, the model parameters are

$$(F, G, p_{10}, p_{20}, \ldots, p_{K0})$$



CRM model: Literature

- Chevret (1993): For **G = Exp(1) and a given set of $p_{10}, p_{20}, \ldots, p_{K0}$**
  - Logistic F with $a_0 = 3$ is superior to empiric
- Lee and Cheung (2009): For **any fixed F and G**
  - we can choose $p_{10}, p_{20}, \ldots, p_{K0}$ to match operating characteristics
- Lee and Cheung (2011): For **any fixed F and $p_{10}, p_{20}, \ldots, p_{K0}$**
  - A *least informative prior* is adequate

Choice of p_{0k}'s

Who should choose p_{0k}'s?

- **Ideal** – clinicians choose the initial guesses for all test doses based on their knowledge/experience
- **Reality** – rarely done; too difficult
- **Goal 1:** Generate the initial guesses p_{0k}'s with minimal inputs from clinicians by reducing the dimensionality of the specification problem:
  - Reduce the initial guesses (K numbers) into two *clinically* interpretable parameters.

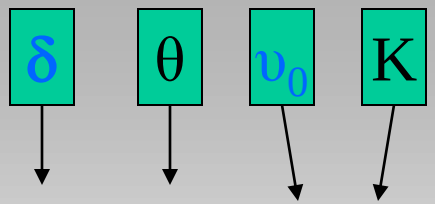
How to choose p_{0k}'s?

- To get p_{0k}'s we need:
  - The prior MTD, υ_0 = **Starting dose level**
  - An acceptable range of DLT rate $\theta \pm \delta$, where θ is the target DLT rate. E.g., 0.25 ± 0.05

  in addition to all other CRM parameters:
  - Dose-toxicity function F
  - Number of test doses K
  - Target DLT rate θ …

How to choose p_{0k}'s?

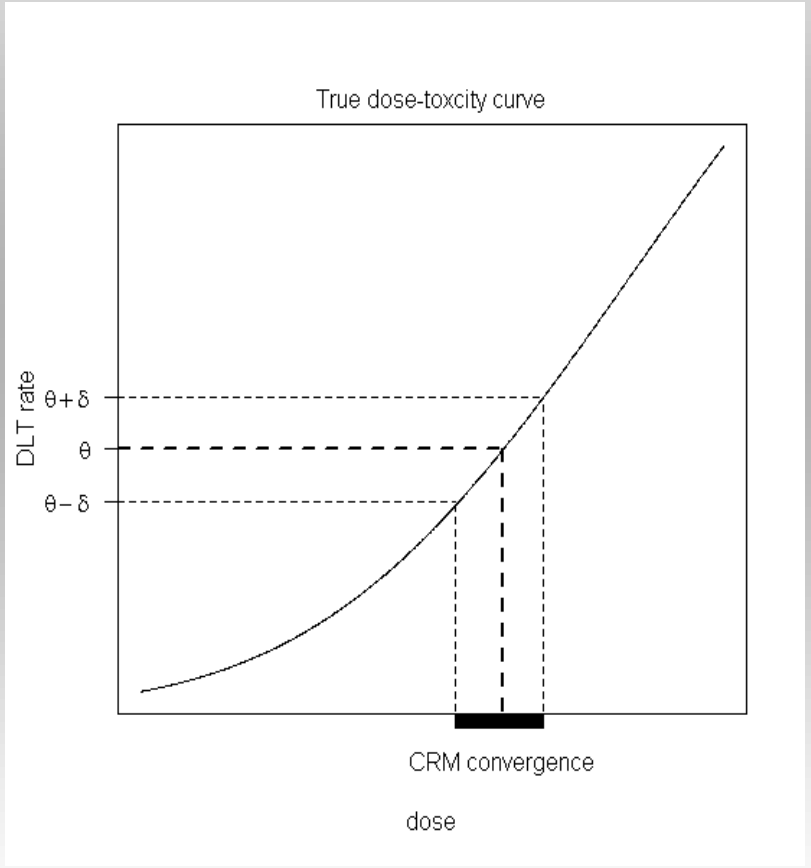
- For any given δ, a skeleton can be obtained using the function **`getprior`** in the R package `dfcrm'




```
> p0 <- getprior(0.05,0.25,3,5,model="logistic")
> round(p0,digits=2)
[1] 0.09 0.16 0.25 0.36 0.46
```

Interpretation of δ

- **Theoretical basis** of p_{0k}'s by the function **`getprior`:** The CRM converges to the acceptable range $\theta \pm \delta$ on the probability scale
- a.k.a. indifference interval (Cheung and Chappell, 2002, *Biometrics*)



How to choose δ?

- **Goal 2:** Choose δ empirically (if the clinicians don't call it)
  - Asymptotically, a small δ has a small bias.
  - With small-moderate sample size, a small δ has a large variance of selected MTD.
  - Use simulations to obtain a δ that yields competitive operating characteristics over a wide range of scenarios

Step 1 – Iterate δ

Specify a CRM model:

- Logistic function (with a fixed intercept):

$$\mathbf{logit\ \{\ F(x,\ \beta)\ \}\ =\ 3\ +\ exp(\beta)\ x}$$

- Normal prior $\boldsymbol{\beta \sim N(0,\ 1.34)}$
- Target rate $\boldsymbol{\theta = 0.25}$
- $\mathbf{K = 5}$ dose levels
- Prior MTD $\boldsymbol{\upsilon_0 = 3}$ (starting dose)
- **Iterate δ from 0.01 to 0.24**

Step 2 – Simulate

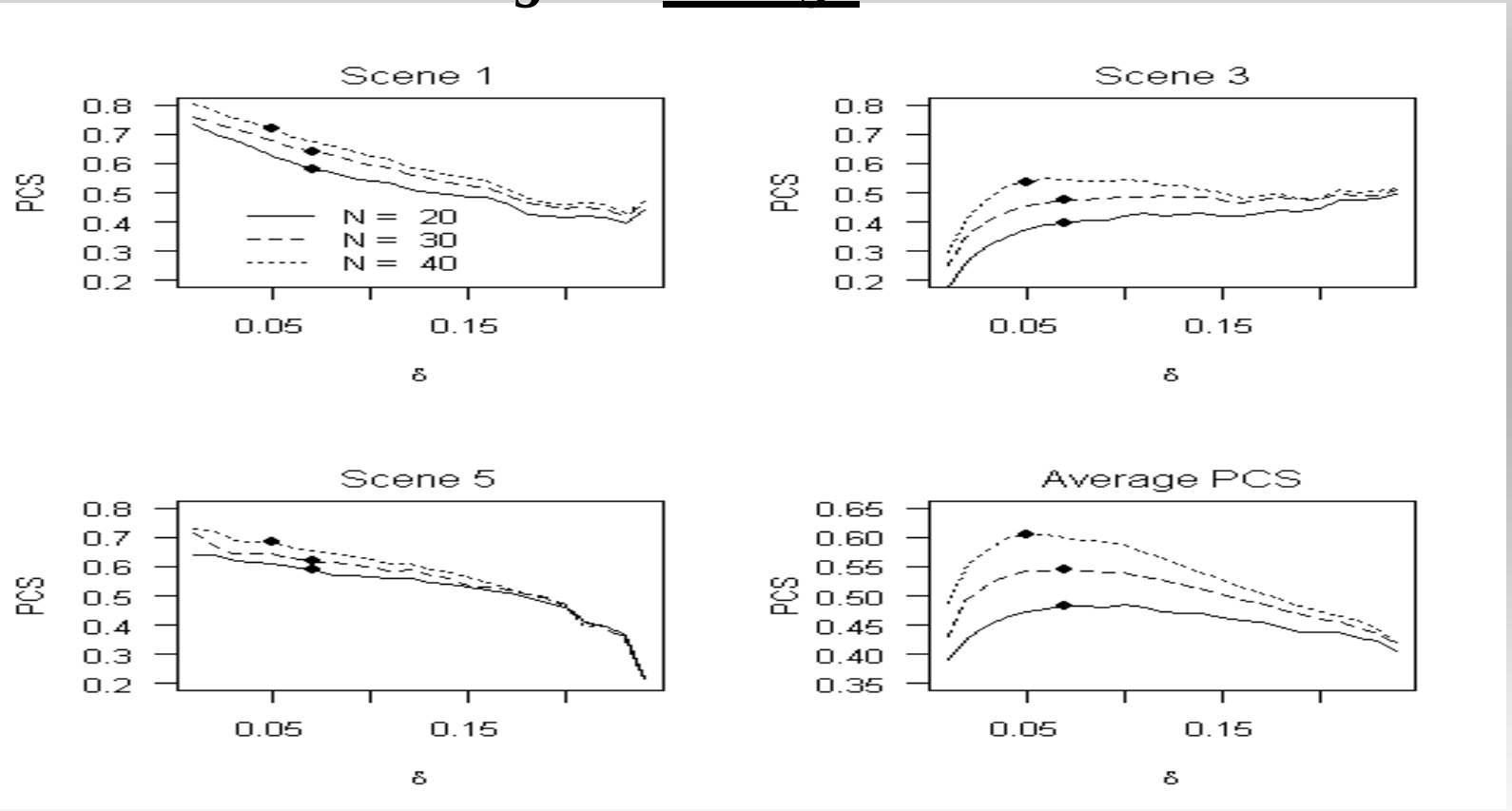
For each δ,

Run CRM under the plateau scenarios (calibration set): Record **average** probability of correctly selecting (**PCS**) the MTD

Scene	True p_1	True p_2	True p_3	True p_4	True p_5
1	**0.25**	0.40	0.40	0.40	0.40
2	0.14	**0.25**	0.40	0.40	0.40
3	0.14	0.14	**0.25**	0.40	0.40
4	0.14	0.14	0.14	**0.25**	0.40
5	0.14	0.14	0.14	0.14	**0.25**

Step 3 – Compare PCS (ave.)

Choose δ with the highest <u>average</u> PCS



Choice of δ: results

- N ≈ 20—40
- For logistic with fixed intercept 3,
  - For θ = 0.10, the optimal δ ranges 0.02—0.04
  - For θ = 0.20, the optimal δ ranges 0.04—0.08
  - For θ = 0.25, the optimal δ ranges 0.04—0.08
  - For θ = 0.33, the optimal δ ranges 0.04—0.10
- Optimal δ is tabulated in Cheung (2011, *DFCRM*)

Choice of prior $G(\beta)$

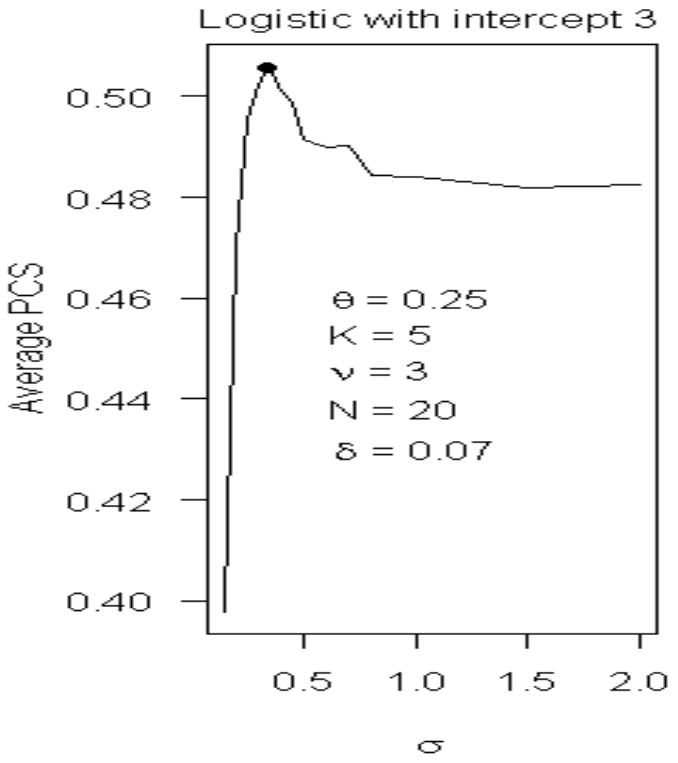
Problem reduction

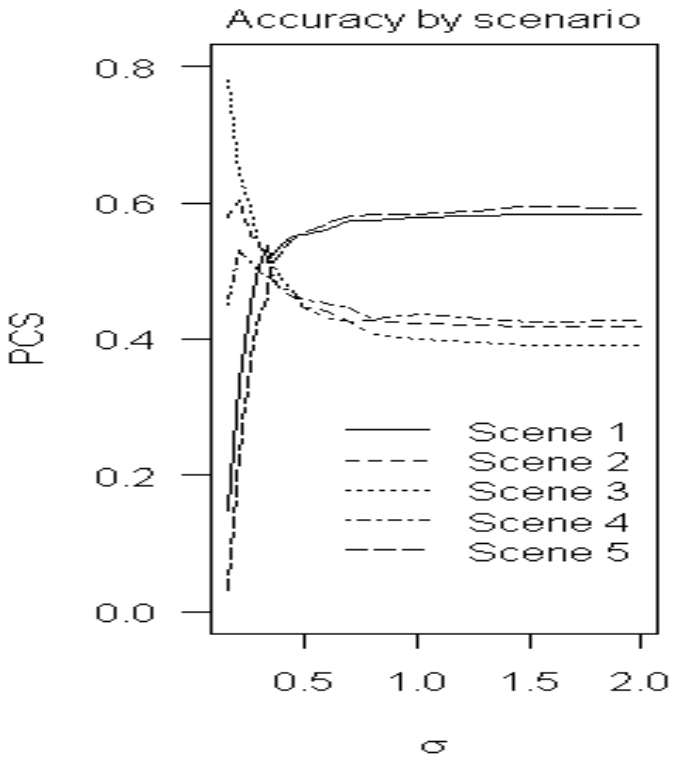
- Focus on the logistic model with the following parametrization:
  - **Logistic:** logit { F(x, β) } = a_0 + **exp(β)** x
    and a **normal prior β ~ N(0, σ^2)**
- **p_{01}, …, p_{0K}** are chosen and fixed.
- The CRM model is then completed by specifying the prior standard deviation **σ.**

Simulation to get σ

- 1st try: Use the same simulation approach as before:
  1. *Iterate **σ**:* Fix all CRM parameters but **σ**
  2. *Simulate:* Run CRM under the plateau scenarios
  3. *Compare PCS:* Choose **σ** with the highest average PCS

Simulation to get σ: Results



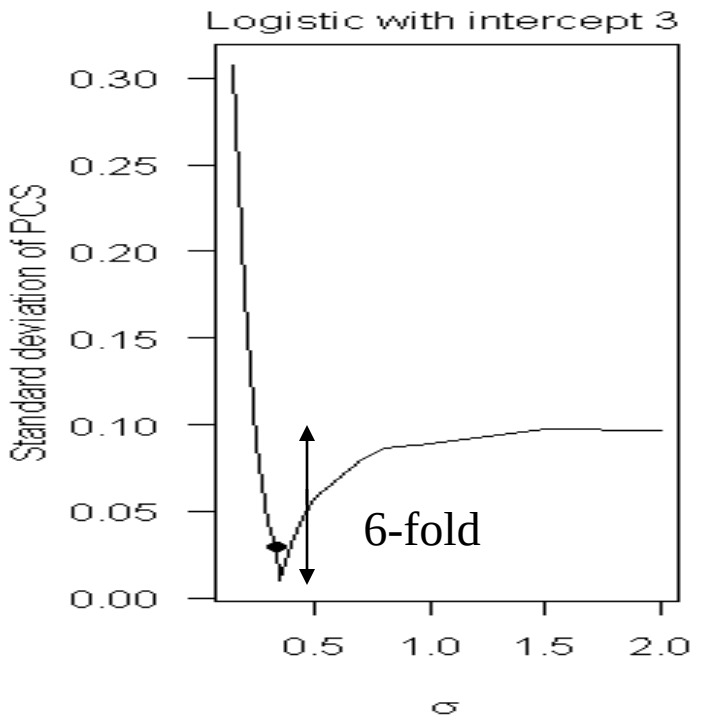


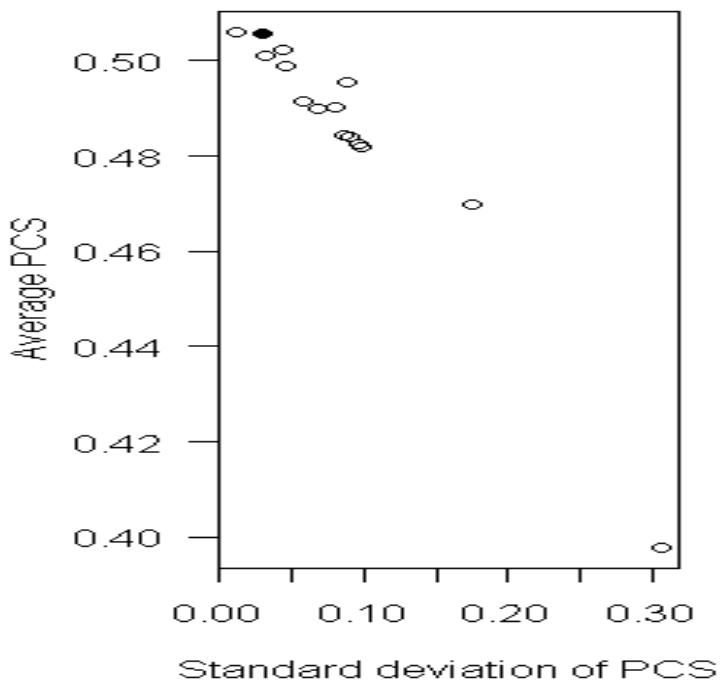
Simulation to get σ: Problem 1

- Average PCS is quite flat once σ is “large” enough
  - difference less than 3 percentage points
  - The average PCS criterion does not seem sensitive and discriminatory

Alternative criterion

Standard deviation of PCS





Simulation to get σ: Problem 2

- Range of good σ is dependent on the other design parameters, and is not bounded
  - Good range of σ for logistic: 0.25—0.45
  - Good range of σ for empiric: 0.75—1.50
  - A general exhaustive search is infeasible

Detour: Least informative prior

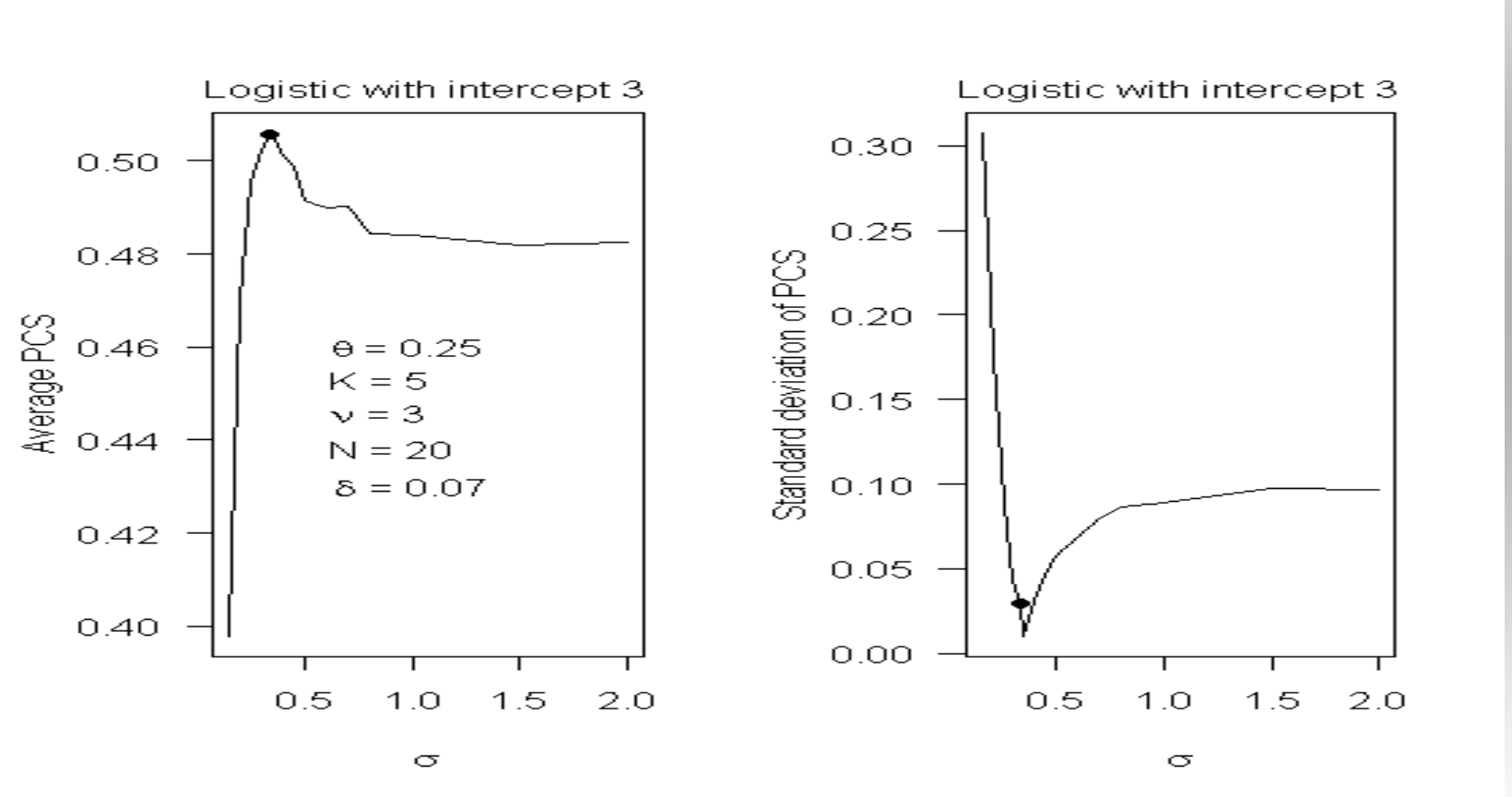
- A large σ is **not** vague – on the MTD scale
- Using the above specified logistic model:

σ	Prior probability υ = dose level				
	1	2	3	4	5
0.20	0.09	0.24	0.36	0.23	0.08
0.33	**0.21**	**0.19**	**0.22**	**0.19**	**0.20**
1.16	0.41	0.06	0.06	0.06	0.40

Detour: Least informative prior

- **Definition:** A least informative $\boldsymbol{\sigma^{LI}}$ for the normal prior $G(\beta)$ is a value of σ that gives a prior distribution of υ "closest" to the uniform distribution.
- **Observation:** For the logistic model, simulations show that the least informative prior performs well.

Detour: Least informative prior



Simulation to get σ: Aided by σ^{LI}

- A general search in the neighborhood of least informative prior
  - Evaluate least informative σ^{LI} (binary search)
  - Iterate σ in the neighborhood of σ^{LI}, e.g., from $0.8\ \sigma^{LI}$ to $1.5\ \sigma^{LI}$.
  - Choose σ that minimizes standard deviation of PCS over the plateau scenarios (calibration set)

Example: A bortezomib trial

- Leonard, Furman, Cheung, et al. (2006): CHOP-R + escalation dose of bortezomib in lymphoma patients
- Trial design: (TITE-)CRM with
  - $\theta = 0.25$, $K = 5$, $\upsilon = 3$
  - **$p_{01} = .05$, $p_{02} = .12$, $p_{03} = .25$, $p_{04} = .40$, $p_{05} = .55$**
  - Empiric $F(d, \beta) = d^{\exp(\beta)}$
  - $\beta \sim N(0, 1.34)$

Example: A bortezomib trial

- These design parameters were chosen by trial-and-error aided by simulations under the validation scenarios:

Scene	True p_1	True p_2	True p_3	True p_4	True p_5
1	**0.25**	0.40	0.45	0.55	0.60
2	0.05	**0.25**	0.40	0.45	0.55
3	0.05	0.05	**0.25**	0.45	0.55
4	0.05	0.05	0.08	**0.25**	0.45
5	0.05	0.05	0.08	0.12	**0.25**

Example: A bortezomib trial

	Study model $\sigma = 1.16$	Logistic $\delta = \mathbf{0.07}$, $\sigma = 1.16$	Logistic $\delta = 0.07$, $\boldsymbol{\sigma = 0.35}$
PCS – 1	0.67	0.69	0.62
PCS – 2	0.58	0.57	0.60
PCS – 3	0.68	0.64	0.69
PCS – 4	0.64	0.61	0.66
PCS – 5	0.66	0.70	0.61
PCS (ave)	0.65	0.64	0.64
PCS (std)	0.04	0.05	0.04

Discussion

- Calibration
  - With respect to objective criteria: indifference interval and least informative prior
  - Aided by objective operating characteristics via simulation
- Simplify the model calibration process
  - Get a reasonable δ: available from existing tables
  - Get the least informative σ^{LI}: 5-line code in R
  - (Optional) Iterate in the neighborhood of σ^{LI}
- NOT to improve upon trial-and-error in terms of accuracy, but to provide competitive operating characteristics with an automated model specification; e.g., bortezomib trial

Resources

- `dfcrm' package in R
  - http://www.r-project.org
- Lee and Cheung (2009, *Clinical Trials*)
- Lee and Cheung (2011, *Stat in Med*)
- Cheung (2011) *DFCRM*. Chapman & Hall