Objective calibration of the Bayesian CRM

Ken Cheung
Department of Biostatistics, Columbia University
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):
    \[
    \arg \min_k \left| \pi(d_k) - p \right|
    \]
    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
  \[ \arg \min_k \ | \ F(d_k, b) - p \ | \]
  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 \( 1/3 + 0/3 + 1/1 \ldots ? \)

- **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 $F(x, \beta) = P(\text{DLT at dose } x)$
2. Choose a prior distribution $G(\beta)$ of $\beta$.
3. Evaluate the dose labels $\{d_1, d_2, \ldots, d_K\}$ for the $K$ test doses via back ward 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 $\beta$. 
CRM model

Thus, the model parameters are $(F, G, p_{10}, p_{20}, \ldots, p_{K0})$

- Dose-toxicity function, e.g., empiric $F(x,\beta) = x^\beta$
- Prior distribution, e.g., $\beta \sim \text{Exp}(1)$
- Initial guesses of DLT rates “Skeleton”
CRM model: Literature

- **Chevret (1993):** For $G = \text{Exp}(1)$ 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 $p_{0_k}$’s?

- To get $p_{0_k}$’s we need:
  - The prior MTD, $v_0 = \text{Starting dose level}
  - An acceptable range of DLT rate $\theta \pm \delta$, where $\theta$ is the target DLT rate. E.g., $0.25 \pm 0.05$
  - Dose-toxicity function $F$
  - Number of test doses $K$
  - Target DLT rate $\theta$ …
How to choose $p_{0k}$’s?

- For any given $\delta$, a skeleton can be obtained using the function `getprior` in the R package `dfcrm`.

```r
> 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 $\delta$

- **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 $\delta$?

- **Goal 2:** Choose $\delta$ empirically (if the clinicians don’t call it)
  - Asymptotically, a small $\delta$ has a small bias.
  - With small-moderate sample size, a small $\delta$ has a large variance of selected MTD.
  - Use simulations to obtain a $\delta$ that yields competitive operating characteristics over a wide range of scenarios.
Step 1 – Iterate $\delta$

Specify a CRM model:

- Logistic function (with a fixed intercept):
  \[
  \text{logit} \{ F(x, \beta) \} = 3 + \exp(\beta) x
  \]

- Normal prior $\beta \sim N(0, 1.34)$

- Target rate $\theta = 0.25$

- $K = 5$ dose levels

- Prior MTD $\nu_0 = 3$ (starting dose)

- **Iterate $\delta$ from 0.01 to 0.24**
Step 2 – Simulate

For each $\delta$,
Run CRM under the plateau scenarios (calibration set): Record average probability of correctly selecting (PCS) the MTD

<table>
<thead>
<tr>
<th>Scene</th>
<th>True $p_1$</th>
<th>True $p_2$</th>
<th>True $p_3$</th>
<th>True $p_4$</th>
<th>True $p_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.25</td>
<td>0.40</td>
<td>0.40</td>
<td>0.40</td>
<td>0.40</td>
</tr>
<tr>
<td>2</td>
<td>0.14</td>
<td>0.25</td>
<td>0.40</td>
<td>0.40</td>
<td>0.40</td>
</tr>
<tr>
<td>3</td>
<td>0.14</td>
<td>0.14</td>
<td>0.25</td>
<td>0.40</td>
<td>0.40</td>
</tr>
<tr>
<td>4</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.25</td>
<td>0.40</td>
</tr>
<tr>
<td>5</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.25</td>
</tr>
</tbody>
</table>
Step 3 – Compare PCS (ave.)

*Choose $\delta$ with the highest average PCS*

---

**Scene 1**

- $\text{N = 20}$
- $\text{N = 30}$
- $\text{N = 40}$

**Scene 3**

- $\text{N = 20}$
- $\text{N = 30}$
- $\text{N = 40}$

**Scene 5**

**Average PCS**
Choice of $\delta$: results

- $N \approx 20$–$40$
- For logistic with fixed intercept 3,
  - For $\theta = 0.10$, the optimal $\delta$ ranges 0.02–0.04
  - For $\theta = 0.20$, the optimal $\delta$ ranges 0.04–0.08
  - For $\theta = 0.25$, the optimal $\delta$ ranges 0.04–0.08
  - For $\theta = 0.33$, the optimal $\delta$ ranges 0.04–0.10
- Optimal $\delta$ is tabulated in Cheung (2011, *DFCRM*)
Choice of prior $G(\beta)$
Problem reduction

- Focus on the logistic model with the following parametrization:
  - **Logistic:** \( \text{logit} \{ F(x, \beta) \} = a_0 + \exp(\beta) x \)
  and a normal prior \( \beta \sim N(0, \sigma^2) \)
- \( p_{01}, \ldots, p_{0K} \) are chosen and fixed.
- The CRM model is then completed by specifying the prior standard deviation \( \sigma \).
Simulation to get $\sigma$

- 1st try: Use the same simulation approach as before:
  1. *Iterate $\sigma$*: Fix all CRM parameters but $\sigma$
  2. *Simulate*: Run CRM under the plateau scenarios
  3. *Compare PCS*: Choose $\sigma$ with the highest average PCS
Simulation to get $\sigma$: Results
Simulation to get $\sigma$: Problem 1

- Average PCS is quite flat once $\sigma$ 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**

![Graph showing standard deviation of PCS and average PCS vs. standard deviation of PCS with a 6-fold increase indicated.](image)
Simulation to get $\sigma$: Problem 2

- Range of good $\sigma$ is dependent on the other design parameters, and is not bounded
  - Good range of $\sigma$ for logistic: 0.25—0.45
  - Good range of $\sigma$ for empiric: 0.75—1.50
  - A general exhaustive search is infeasible
Detour: Least informative prior

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

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>Prior probability $\nu = \text{dose level}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>0.09</td>
</tr>
<tr>
<td>0.33</td>
<td>0.21</td>
</tr>
<tr>
<td>1.16</td>
<td>0.41</td>
</tr>
</tbody>
</table>
Detour: Least informative prior

• **Definition:** A least informative $\sigma^{LI}$ for the normal prior $G(\beta)$ is a value of $\sigma$ that gives a prior distribution of $\nu$ “closest” to the uniform distribution.

• **Observation:** For the logistic model, simulations show that the least informative prior performs well.
Detour: Least informative prior

\[ \theta = 0.25 \]
\[ K = 5 \]
\[ \nu = 3 \]
\[ N = 20 \]
\[ \delta = 0.07 \]
Simulation to get $\sigma$: Aided by $\sigma^{LI}$

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

- Trial design: (TITE-)CRM with
  - $\theta = 0.25$, $K = 5$, $\nu = 3$
  - $p_{01} = .05$, $p_{02} = .12$, $p_{03} = .25$, $p_{04} = .40$, $p_{05} = .55$
  - 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:

<table>
<thead>
<tr>
<th>Scene</th>
<th>True $p_1$</th>
<th>True $p_2$</th>
<th>True $p_3$</th>
<th>True $p_4$</th>
<th>True $p_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.25</td>
<td>0.40</td>
<td>0.45</td>
<td>0.55</td>
<td>0.60</td>
</tr>
<tr>
<td>2</td>
<td>0.05</td>
<td>0.25</td>
<td>0.40</td>
<td>0.45</td>
<td>0.55</td>
</tr>
<tr>
<td>3</td>
<td>0.05</td>
<td>0.05</td>
<td>0.25</td>
<td>0.45</td>
<td>0.55</td>
</tr>
<tr>
<td>4</td>
<td>0.05</td>
<td>0.05</td>
<td>0.08</td>
<td>0.25</td>
<td>0.45</td>
</tr>
<tr>
<td>5</td>
<td>0.05</td>
<td>0.05</td>
<td>0.08</td>
<td>0.12</td>
<td>0.25</td>
</tr>
</tbody>
</table>
### Example: A bortezomib trial

<table>
<thead>
<tr>
<th></th>
<th>Study model $\sigma = 1.16$</th>
<th>Logistic $\delta = 0.07$, $\sigma = 1.16$</th>
<th>Logistic $\delta = 0.07$, $\sigma = 0.35$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCS - 1</td>
<td>0.67</td>
<td>0.69</td>
<td>0.62</td>
</tr>
<tr>
<td>PCS - 2</td>
<td>0.58</td>
<td>0.57</td>
<td>0.60</td>
</tr>
<tr>
<td>PCS - 3</td>
<td>0.68</td>
<td>0.64</td>
<td>0.69</td>
</tr>
<tr>
<td>PCS - 4</td>
<td>0.64</td>
<td>0.61</td>
<td>0.66</td>
</tr>
<tr>
<td>PCS - 5</td>
<td>0.66</td>
<td>0.70</td>
<td>0.61</td>
</tr>
<tr>
<td>PCS (ave)</td>
<td>0.65</td>
<td>0.64</td>
<td>0.64</td>
</tr>
<tr>
<td>PCS (std)</td>
<td>0.04</td>
<td>0.05</td>
<td>0.04</td>
</tr>
</tbody>
</table>
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 $\delta$: available from existing tables
  – Get the least informative $\sigma^{LI}$: 5-line code in R
  – (Optional) Iterate in the neighborhood of $\sigma^{LI}$
• 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](http://www.r-project.org)
- Lee and Cheung (2009, *Clinical Trials*)