Taking your agents to the clinic: phase I study design

Elizabeth Garrett-Mayer, PhD
Associate Professor
Director of Biostatistics, Hollings Cancer Center
Dose finding

- Traditional goal: Find the highest dose with acceptable toxicity
- New goals:
  - Find dose with sufficient effect on biomarker
  - Find dose with acceptable toxicity and high efficacy
  - Find dose with acceptable toxicity in the presence of another agent that may also be escalated.
Acceptable toxicity

- What is acceptable rate of toxicity?
  - 20%?
  - 30%?
  - 50%?

- What is toxicity????
  - Standard in cancer: Grade 4 hematologic or grade 3/4 non-hematologic toxicity
  - Always?
  - Does it depend on reversibility of toxicity?
  - Does it depend on intensity of treatment?
    - Tamoxifen?
    - Chemotherapy?
Phase I study design

○ “Standard” Phase I trials (in oncology) use what is often called the ‘3+3’ design (aka ‘modified Fibonacci’):

<table>
<thead>
<tr>
<th>Treat 3 patients at dose K</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. If 0 patients experience dose-limiting toxicity (DLT), escalate to dose K+1</td>
</tr>
<tr>
<td>2. If 2 or more patients experience DLT, de-escalate to level K-1</td>
</tr>
<tr>
<td>3. If 1 patient experiences DLT, treat 3 more patients at dose level K</td>
</tr>
<tr>
<td>A. If 1 of 6 experiences DLT, escalate to dose level K+1</td>
</tr>
<tr>
<td>B. If 2 or more of 6 experiences DLT, de-escalate to level K-1</td>
</tr>
</tbody>
</table>

○ Maximum tolerated dose (MTD) is considered highest dose at which 1 or 0 out of six patients experiences DLT.

○ Doses need to be pre-specified

○ Confidence in MTD is usually poor.
Should we use the “3+3”?

- **It is terribly imprecise and inaccurate in its estimate of the MTD**
- **Why?**
  - MTD is not based on all of the data
  - Algorithm-based method
  - Ignores rate of toxicity!!
- Likely outcomes:
  - Choose a dose that is too high
    - Find in phase II that agent is too toxic.
    - Abandon further investigation or go back to phase I
  - Choose a dose that is too low
    - Find in phase II that agent is ineffective
    - Abandon agent
Two examples:

Example 1: total N=21

<table>
<thead>
<tr>
<th></th>
<th>Cohort 1</th>
<th>Cohort 2</th>
<th>Cohort 3</th>
<th>Cohort 4</th>
<th>Cohort 5</th>
<th>Cohort 6</th>
<th>Cohort 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dose</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>DLTs</td>
<td>0/3</td>
<td>1/3</td>
<td>0/3</td>
<td>1/3</td>
<td>0/3</td>
<td>1/3</td>
<td>1/3</td>
</tr>
</tbody>
</table>
Observed Data

Dose vs. DLT Rate

- Dose values: 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0
- DLT Rate values: 0.0, 0.2, 0.4, 0.6, 0.8, 1.0
Observed Data: with 90% CIs
Example 2:  total N=12

<table>
<thead>
<tr>
<th></th>
<th>Cohort 1</th>
<th>Cohort 2</th>
<th>Cohort 3</th>
<th>Cohort 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dose</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>DLTs</td>
<td>0/3</td>
<td>0/3</td>
<td>0/3</td>
<td>2/3</td>
</tr>
</tbody>
</table>
Observed Data
Observed Data: with 90% CIs
Why is the 3+3 so popular?

- People know how to implement it
- “we just want a quick phase I”
- It has historic presence
- FDA (et al.) accept and promote it
- There is a level of comfort from the approach
- The “better” approaches are too statistical(!)
USE A SMARTER DESIGN!

- Phase I is the most critical phase of drug development!
- What makes a good design?
  - Accurate selection of MTD
    - dose close to true MTD
    - dose has DLT rate close to the one specified
  - Relatively few patients in trial are exposed to toxic doses
- Why not impose a statistical model?
- What do we “know” that would help?
  - Monotonicity
  - Desired level of DLT
Continual Reassessment Method (CRM)

- Allows statistical modeling of optimal dose: dose-response relationship is assumed to behave in a certain way.
- Can be based on “safety” or “efficacy” outcome (or both).
- Design searches for best dose given a desired toxicity or efficacy level and does so in an efficient way.
- This design REALLY requires a statistician throughout the trial.
- ADAPTIVE
CRM history in brief

- Originally devised by O’Quigley, Pepe and Fisher (1990) where dose for next patient was determined based on responses of patients previously treated in the trial.

- 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 reasons why to use CRM

Table 4  Characteristics of five Phase 1 studies. Designs 1–3 are CRM designs with different sample sizes and cohort sizes. Designs 4–6 are “3 + 3” designs with different prespecified dose levels. All six designs have the same true model of dose-toxicity. Dose levels for the “3 + 3” designs are shown in Figure 5. Results are based on 1000 simulated trials for each design.

<table>
<thead>
<tr>
<th>Design 1 (CRM) (Example 2)</th>
<th>Design 2 (CRM)</th>
<th>Design 3 (CRM)</th>
<th>Design 4 (“3 + 3”)</th>
<th>Design 5 (“3 + 3”)</th>
<th>Design 6 (“3 + 3”)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Total sample size: fixed for CRMs, median for “3 + 3”</td>
<td>30</td>
<td>50</td>
<td>60</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>2  Patients per cohort</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>3  Number of cohorts fixed for CRMs, median for “3 + 3”</td>
<td>10</td>
<td>10</td>
<td>20</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>4  % of trials with recommended dose within 250mg of true dose (1656 mg)</td>
<td>57%</td>
<td>72%</td>
<td>71%</td>
<td>41%</td>
<td>35%</td>
</tr>
<tr>
<td>5  % of trials with recommended dose within 400mg of true dose (1656 mg)</td>
<td>80%</td>
<td>91%</td>
<td>89%</td>
<td>41%</td>
<td>54%</td>
</tr>
<tr>
<td>6  % of trials with recommended dose dose-limiting toxicity rate of &gt;40%</td>
<td>9.5%</td>
<td>5.8%</td>
<td>5.9%</td>
<td>7.1%</td>
<td>21%</td>
</tr>
<tr>
<td>7  % of trials with recommended dose dose-limiting toxicity rate of &gt;50%</td>
<td>0.9%</td>
<td>0.2%</td>
<td>0.6%</td>
<td>7.1%</td>
<td>2.0%</td>
</tr>
<tr>
<td>8  % of trials with recommended dose dose-limiting toxicity rate of &lt;20%</td>
<td>13%</td>
<td>5.7%</td>
<td>6.2%</td>
<td>52%</td>
<td>44%</td>
</tr>
<tr>
<td>9  % of trials with recommended dose dose-limiting toxicity rate of &lt;10%</td>
<td>0.0%</td>
<td>0.1%</td>
<td>0.0%</td>
<td>11%</td>
<td>16%</td>
</tr>
<tr>
<td>10 Average % of patients treated at doses with 40% or greater dose-limiting toxicity rate</td>
<td>7.6%</td>
<td>7.8%</td>
<td>5.7%</td>
<td>17%</td>
<td>23%</td>
</tr>
<tr>
<td>11 Average % of patients treated at doses with 20% or less dose-limiting toxicity rate</td>
<td>32%</td>
<td>19%</td>
<td>24%</td>
<td>62%</td>
<td>53%</td>
</tr>
<tr>
<td>12 Average % of patients with dose-limiting toxicities</td>
<td>26%</td>
<td>28%</td>
<td>26%</td>
<td>21%</td>
<td>22%</td>
</tr>
</tbody>
</table>
Basic Idea of CRM

\[ p(\text{toxicity} | \text{dose} = d_i) = \frac{\exp(3 + \alpha d_i)}{1 + \exp(3 + \alpha d_i)} \]
Modified CRM
(Goodman, Zahurak, and 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 (e.g., p = 0.30)**
- At end of trial, we can estimate dose response curve.
Some other mathematical models we could choose

A. Hyperbolic tangent: \[ p(\text{dose}) = \left( \frac{\tanh(\text{dose}) + 1}{2} \right)^b \]

B. One-parameter logistic: \[ p(\text{dose}) = \frac{e^{4+b \cdot \text{dose}}}{1 + e^{4+b \cdot \text{dose}}} \]

C. One-parameter logistic: \[ p(\text{dose}) = \frac{e^{3+b \cdot \text{dose}}}{1 + e^{3+b \cdot \text{dose}}} \]

D. Two-parameter logistic: \[ P = \frac{e^{a+b \cdot \text{dose}}}{1 + e^{a+b \cdot \text{dose}}} \]
○ **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, in this case, ‘2+2’)
  - **Upon first toxicity, fit the dose-response model using observed data**
    - Estimate $\alpha$
      - 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.
Simulated Example

- Shows how the CRM works in practice
- Assume:
  - Cohorts of size 2
  - Escalate at fixed doses until DLT occurs
  - Then, fit model and use model-based escalation
  - Increments of 50mg are allowed
  - Stop when 10 patients have already been treated at a dose that is the next chosen dose
Result

- 450mg is determined to be the optimal dose to take to phase II
- 30 patients (?!)
- Confidence interval for true DLT rate at 450mg: 15% - 40%
- Used ALL of the data to make our conclusion
**Real Example**  Samarium in pediatric osteosarcoma:

**Desired DLT rate is 30%**.

- 2 patients treated at dose 1 with 0 toxicities
- 2 patients treated at dose 2 with 1 toxicity
- Fit CRM using equation below

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

- Estimated \( \alpha = 0.77 \)
- Estimated dose is 1.4mCi/kg for next cohort.
**Example**  Samarium study with cohorts of size 2:
- 2 patients treated at 1.0 mCi/kg with no toxicities
- 4 patients treated at 1.4 mCi/kg with 2 toxicities

⇒ Fit CRM using equation on earlier slide

- Estimated $\alpha = 0.71$
- Estimated dose for next patient is 1.2 mCi/kg
**Example**  Samarium study with cohorts of size 2:
- 2 patients treated at 1.0 mCi/kg with no toxicities
- 4 patients treated at 1.4 mCi/kg with 2 toxicities
- 2 patients treated at 1.2 mCi/kg with 1 toxicity

⇒ Fit CRM using equation on earlier slide

- Estimated $\alpha = 0.66$
- Estimated dose for next patient is 1.1 mCi/kg
**Example**  Samarium study with cohorts of size 2:

- 2 patients treated at 1.0 mCi/kg with no toxicities
- 4 patients treated at 1.4 mCi/kg with 2 toxicities
- 2 patients treated at 1.2 mCi/kg with 1 toxicity
- 2 patients treated at 1.1 mCi/kg with no toxicities

⇒ Fit CRM using equation on earlier slide

- Estimated $\alpha = 0.72$
- Estimated dose for next patient is 1.2 mCi/kg
When does it end?

- Pre-specified stopping rule
- Can be fixed sample size
- Often when a “large” number have been assigned to one dose.
- This study should enroll at least two more cohorts.
Dose increments

- Can be discrete or continuous
- Infusion?
- Tablet?
- Stopping rule should depend on nature (and size) of allowed increment!
A little more on the statistics:

- Original design was purely Bayesian
- Requires a prior distribution
  - Prior is critically important because it outweighs the data early in the trial
  - Computationally is somewhat challenging
- Some revised designs use ML
  - Simpler to use
  - Once a DLT is observed, model can be fit
  - Some will “inform” the ML approach using “pseudo-data” (Piantadosi)
Simple prediction, but backwards(?)

- Usual prediction:
  - Get some data
  - Fit model
  - Estimate the **outcome** for a **new** patient with a particular **characteristic**

- CRM prediction
  - Get some data
  - Fit model
  - Find the **characteristic** (dose) associated with a particular **outcome** (DLT rate)
Finding the next dose: ML approach

- Use maximum likelihood to estimate the model.
- What likelihood do we use? Binomial.

\[
L(p; y) = \prod_{i=1}^{N} p^{y_i} (1 - p)^{(1-y_i)}
\]

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

- Algorithmic estimation of \( \alpha \)
Finding next dose

- Recall model, now with estimated \( \alpha \):

\[
p_i = \frac{\exp(3 + \hat{\alpha}d_i)}{1 + \exp(3 + \hat{\alpha}d_i)}
\]

- Rewrite in terms of \( d_i \):

\[
d_i = \frac{\log(\frac{p_i}{1-p_i}) - 3}{\hat{\alpha}}
\]
Finding next dose

- Use desired DLT rate as $p_i$

$$d_i = \frac{\log\left(\frac{3}{7}\right) - 3}{\hat{\alpha}} = -3.85$$
Negative dose?

- Doses are often mapped to another scale
- **dose coding:**
  - -6 = level 1 (1.0)
  - -5 = level 2 (1.4)
  - -4 = level 3 (2.0)
  - -3 = level 4 (2.8)
  - -2 = level 5 (4.0)
- **WHY?** Makes the statistics work....
CRM Software:
http://www.cancerbiostats.onc.jhmi.edu/software.cfm
Escalation with Overdose Control

- EWOC (Babb et al.)
- Similar to CRM
- Bayesian
- Advantage: overdose control
  - “loss function”
  - Constrained so that the predicted proportion of patients who receive an overdose cannot exceed a specified value
  - Implies that giving an overdose is greater mistake than an underdose
  - CRM does not make this distinction
  - This control is changed as data accumulates
Figure 3. Proportion of patients given doses for which the probability of a severe toxic reaction is less than or equal to 1/5. Each × and + represents the results from all simulation runs for a particular parameter combination. Each point, (●), is the average of the results obtained for a particular method at the six parameter combinations considered. For each of the UD schemes the results obtained when only 6 dose levels were used are shown to the left of the results obtained when 11 levels were used.
Figure 9. Average bias of \( \hat{\gamma} \), the estimate of the MTD.
Figure 12. Proportion of patients exhibiting dose-limiting toxicity
EWOC Software

- http://www.sph.emory.edu/BRI-WCI/ewoc.html
Other Novel Ideas in Phase I

- Outcome is not always toxicity
- Even in phase I, efficacy can be outcome to guide dose selection
- Two outcomes: safety and efficacy
Efficacy Example: Rapamycin in Pancreatic Cancer

- Outcome: response
- Response = 80% inhibition of pharmacodynamic marker
- Assumption: as dose increases, % of patients with response will increase
- Desired proportion responding: 80%
Efficacy Example: Rapamycin in Pancreatic Cancer
Safety and Efficacy

- Zhang, Sargent, Mandrekar
- Example: high dose can induce “over-stimulation”
- Three categories:
  - 1 = no response, no DLT
  - 2 = response, no DLT
  - 3 = DLT
- Use the continuation ratio model
- Very beautiful(!)
- Not particularly friendly at the current time for implementation
Safety and Efficacy

- I’m working on less beautiful, more practical approach
- \( Y = 1 \) if toxicity
  \( = 0 \) if no toxicity
- \( Z = 1 \) if efficacy
  \( = 0 \) if no efficacy
- Simultaneously search for doses with constraints based on toxicity
Summary: “Novel” Phase I trials

- Offer significant improvements over “traditional” phase I design
  - Safer
  - More accurate
- Slightly larger phase I: worth it!
- Related methods: Bayesian Adaptive
Why isn’t everyone using these?

- Change in paradigm
- Larger N
- “I just want a quick phase I”
- Large investment of time from statistician
- Need time to “think” and plan it.
- IRB and others (e.g. CTEP) worry about safety (unjustified!)
References