Harvard Catalyst Journal Club:
“Performance of toxicity probability interval based designs in contrast to the continual reassessment method “
Horton, Wages and Conaway, Statistics in Medicine, 2017

Edie Weller, Ph.D.
Director, Biostatistics and Research Design Core of the Institutional Centers for Clinical and Translational Research at Boston Children’s Hospital
August 29, 2017
Other papers…

• CRM
  – O’Quigley, Shen (Biometrics 1996, ref 1)
  – O’Quigley, Pepe, Fisher (Biometrics 1990, ref 16)
  – O’Quigley, Iasonos (Statistics in Biopharmaceutical Research 2014, ref 8)
  – Wages, Conaway and O’Quigley (Clinical Trials 2013)
  – Iasonos, O’Quigley (Journal of Clinical Oncology 2014, ref 11) review of model guided phase I clinical trials
  – Many others….
Other papers…

• mTPI
  – Ji, Liu, Li, Bekele (Clinical Trials 2010, ref 12)

• BOIN design
  – Liu and Yuan (Applied Statistics 2015, ref 13)
  – Yuan, Hess, Hilsenbeck, and Gilbert (Clinical Cancer Research 2017)
  – Lin and Yin (Stats methods research 2015, combination trials)
Background

Phase I oncology trials:

• Most common is 3+3 design
  – Easy to implement but less efficient than model based methods
• Continual reassessment method (CRM)
  – Most widely recognized model based method
  – Challenge to implement but more efficient
• Two potential alternatives
  – Modified toxicity probability interval (mTPI)
  – Bayesian optimal interval (BOIN)
• Horton et. al paper evaluates these alternatives relative to CRM by simulation
Background: Pediatric Oncology Phase I Trials

90 studies published between 2009-2014

Notation:

DLT = Dose limiting toxicity
\( \theta \) = Target probability of toxicity
\( K \) = Number of available dose levels
\( d_k \) = Dose at level \( k \)
\( \pi_k \) = Probability of DLT at dose level \( k \)
\( n_k \) = Number of subjects at dose level \( k \)
\( y_k \) = Number of subjects who experienced DLT(s) at dose level \( k \)
Additional subjects often treated at MTD (expansion cohort) to improve precision.
3+3 Design and Variations

• Uses pre-defined set of rules to assign next subject’s dose level
• Advantages:
  – Easy to implement and understand
  – Conservative in terms of safety
• Disadvantages
  – Slow dose escalation
  – Many subjects may be treated at sub-therapeutic doses
  – “Memoryless”: uses information from most recent dose to determine next dose; ignores other observed dose information

CRM Design

• Continual Reassessment Method (CRM):
  – Assume an a priori dose-toxicity curve (e.g. logistic model, power model)
  – Select a target toxicity rate (e.g. 30%)
  – Update dose-toxicity curve after each subject’s outcome is observed
  – Model recommends optimal dose for next subject
  – End trial using stopping rule: (e.g. enrolled pre-specified maximum N)

• Many variations exist: modified CRM, time-to-event (TiTE) CRM, etc.

Example Dose-Toxicity Curves

Garrett-Mayer. Understanding the CRM. 2005

$\Theta$=Target DLT rate
CRM and Simulation Assumptions:

- Used 2-stage likelihood version of CRM (O’Quigley, Shen 1996)
- Power model used for probability of toxicity at dose $d_k$
  
  $$\psi(d_k, a) = p_k \exp(a)$$

  - $a$ is scalar parameter
  - $p_i$ are pre-specified constants (skeleton values), $0 < p_1 < p_2 < \ldots < p_k < 1$
  - Spacing between the skeleton values is important
  - Used Lee and Cheung method to obtain skeleton spacing
CRM and Simulation Assumptions:

- Approach to ensure heterogeneity in DLT responses:
  - 1st stage: single subject cohorts assigned escalating doses until one DLT and one non-DLT observed
  - 2nd stage: ML is used to estimate power model scalar parameter
- Allowed for early termination if the 90% lower confidence limit for the 1st dose level is greater than target toxicity (θ).
- Used following functions in R package dfcrm:
  - CRMOOUT function ("MLE" method, "empiric" model)
  - getprior function for skeleton (half width= 0.06 , MTD prior guess=dose 2)
mTPI and Simulation Assumptions:

- Combines simplicity of 3+3 method and allows for specification of target toxicity probability
- Extension of toxicity probability interval method
- Employs simple beta-binomial model
- Partition interval (0,1) into 3 sub-intervals based on pre-specified constants $\varepsilon_1$ and $\varepsilon_2$. Default values are $\varepsilon_1 = \varepsilon_2 = 0.05$.
  
  $[0, \theta - \varepsilon_1)$ => Dose de-escalation
  $[\theta - \varepsilon_1, \theta - \varepsilon_2)$ => Dose remains the same
  $[\theta - \varepsilon_2, 1)$ => Dose escalation
mTPI and Simulation Assumptions:

• Calculate the unit probability mass (UPM) for each sub-interval
  – Assumes toxicity probabilities ($\pi_k$) have independent beta distributions with shape parameters ($\alpha_k$, $\beta_k$) and ($\alpha_k = \alpha + y_k$ and $\beta_k = \beta + n_k - y_k$).
• Interval with highest UPM dictates the decision for the next patient
• Trial is terminated when pre-specified sample size is reached.
• Safety rules incorporated:
  – Prevent escalation to a dose that has previously deemed too toxic
  – Allow for trial termination if lowest dose too toxic
• Used mTPI webtool found at [http://www.compgenomie.org/NGDF/](http://www.compgenomie.org/NGDF/)
• Used default values of $\varepsilon_1 = \varepsilon_2 = 0.05$. 

YI and Wang, JCO, 2013, 21(14); 1785-91
BOIN Design:

- Optimization is anchored to 3 point (or local) hypotheses.
- Interval boundaries selected to minimize decision error rates.
- Define key parameters:
  - Max N
  - Target toxicity (θ)
  - Pre-specified toxicity tolerance interval (default is [0.6θ, 1.4θ])
  - Number pts per dose level (n_k)
  - Probability the dose is less than, equal to, or higher than MTD (e.g. non-informative prior = 33% for each group)
- With non-informative prior the boundaries are independent of d_k and n_k

BOIN Design:

- Calculate the boundaries $\gamma_{1k}(n_k, \theta)$ and $\gamma_{2k}(n_k, \theta)$
- Approach:
  1. Enter $n_1$ subjects on first dose level
  2. If observed toxicity rate is:
     - $< \gamma_{1k}(n_k, \theta)$ then escalate
     - $> \gamma_{2k}(n_k, \theta)$ then de-escalate
     - Between $\gamma_{1k}(n_k, \theta)$ and $\gamma_{2k}(n_k, \theta)$ then stay at current dose
  3. Repeat steps 1 and 2
  4. Continue trial until max N is reached, or unable to de-escalate further.
- Calculate the MTD with isotonic regression using all toxicity data

BOIN and Simulation Assumptions:

• Not clear how the following parameters were defined
  • Pre-specified tolerance values - use default $(0.6\theta, 1.4\theta)$?
  • Probability that the dose is less than, equal to, or higher than MTD. Use non-informative prior?

Figure 2: A hypothetical phase I clinical trial using the BOIN design. The numbers indicate the patients identification. Three patients in each box from a cohort.

Yuan, Hess, Hilsenbeck, and Gilbert, Clin Cancer Res; 22(17) 4291-301
## Simulation Performance Measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Termination</td>
<td>Proportion of simulations which terminated early</td>
</tr>
<tr>
<td>MTD Selection</td>
<td>Proportion of simulations selecting the true MTD</td>
</tr>
<tr>
<td>Accuracy index for dose selection</td>
<td>Incorporates information at all doses levels into 1 number; Penalizes the method for selecting doses further from the true MTD</td>
</tr>
<tr>
<td>Accuracy index for subject allocation</td>
<td>Same as above but substitute proportion of subjects allocated to dose $k$ for the probability</td>
</tr>
<tr>
<td>Percent correct selection (PCS)</td>
<td>Probability of selecting true MTD (proportion of simulations select the dose that minimizes $</td>
</tr>
<tr>
<td>Percent acceptable selection</td>
<td>Same PCS but considers those doses with $\pi_k$ within 5% of the target probability toxicity.</td>
</tr>
</tbody>
</table>
Simulation Parameters

- $\theta = 0.20$
- Number of simulations = 2000
- Designs: 4, 6, 8 dose levels
- $N=24, 36, 48$ for 4, 6, 8 dose levels
- 16 dose toxicity curves/design
  - 12 with MTD at lower doses
  - 4 with MTD at higher doses
- Scenario numbers do not align across the designs

Figure 1. Dose toxicity curves
• Proportion higher for BOIN and mTPI as compared to CRM for 4 dose levels (S9, S12), 6 dose levels (S1), 8 dose levels (S1) [boxes above]

• In all other scenarios (MTD is not higher than target toxicity) BOIN proportion is a) lower as compared to mTPI, b) within 0.05 of CRM in 45/48 scenarios except 4 dose level (S4,S6) and 6 dose level (S4) [circled above]
- BOIN proportion $\geq$ mTPI proportion in all designs and scenarios
- BOIN proportion within 5% of CRM in 23/48 scenarios (9/16 4 dose scenarios; 10/16 6 dose scenarios; and 4/16 8 dose scenarios)

### Table II. Proportion of simulations recommending the true MTD

<table>
<thead>
<tr>
<th>Scenario</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Four-dose levels</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRM</td>
<td>0.51</td>
<td>0.44</td>
<td>0.66</td>
<td>0.68</td>
<td>0.71</td>
<td>0.64</td>
<td>0.77</td>
<td>0.51</td>
<td>0.50</td>
<td>0.67</td>
<td>0.55</td>
<td>0.34</td>
<td>0.74</td>
<td>0.54</td>
<td>0.72</td>
<td>0.53</td>
</tr>
<tr>
<td>mTPI</td>
<td>0.39</td>
<td>0.33</td>
<td>0.60</td>
<td>0.40</td>
<td>0.87</td>
<td>0.43</td>
<td>0.74</td>
<td>0.51</td>
<td>0.26</td>
<td>0.60</td>
<td>0.42</td>
<td>0.19</td>
<td>0.60</td>
<td>0.46</td>
<td>0.58</td>
<td>0.46</td>
</tr>
<tr>
<td>BOIN</td>
<td>0.44</td>
<td>0.40</td>
<td>0.67</td>
<td>0.53</td>
<td>0.87</td>
<td>0.54</td>
<td>0.82</td>
<td>0.56</td>
<td>0.30</td>
<td>0.66</td>
<td>0.49</td>
<td>0.24</td>
<td>0.65</td>
<td>0.48</td>
<td>0.69</td>
<td>0.48</td>
</tr>
<tr>
<td><strong>Six-dose levels</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRM</td>
<td>0.43</td>
<td>0.49</td>
<td>0.56</td>
<td>0.53</td>
<td>0.38</td>
<td>0.66</td>
<td>0.42</td>
<td>0.54</td>
<td>0.41</td>
<td>0.69</td>
<td>0.47</td>
<td>0.65</td>
<td>0.35</td>
<td>0.34</td>
<td>0.68</td>
<td>0.55</td>
</tr>
<tr>
<td>mTPI</td>
<td>0.22</td>
<td>0.35</td>
<td>0.49</td>
<td>0.39</td>
<td>0.26</td>
<td>0.57</td>
<td>0.30</td>
<td>0.40</td>
<td>0.27</td>
<td>0.58</td>
<td>0.32</td>
<td>0.88</td>
<td>0.29</td>
<td>0.23</td>
<td>0.54</td>
<td>0.43</td>
</tr>
<tr>
<td>BOIN</td>
<td>0.24</td>
<td>0.40</td>
<td>0.54</td>
<td>0.48</td>
<td>0.36</td>
<td>0.63</td>
<td>0.32</td>
<td>0.49</td>
<td>0.33</td>
<td>0.67</td>
<td>0.42</td>
<td>0.90</td>
<td>0.34</td>
<td>0.26</td>
<td>0.63</td>
<td>0.52</td>
</tr>
<tr>
<td><strong>Eight-dose levels</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRM</td>
<td>0.76</td>
<td>0.65</td>
<td>0.38</td>
<td>0.46</td>
<td>0.32</td>
<td>0.51</td>
<td>0.40</td>
<td>0.48</td>
<td>0.47</td>
<td>0.44</td>
<td>0.56</td>
<td>0.50</td>
<td>0.40</td>
<td>0.37</td>
<td>0.92</td>
<td>0.57</td>
</tr>
<tr>
<td>mTPI</td>
<td>0.53</td>
<td>0.47</td>
<td>0.22</td>
<td>0.32</td>
<td>0.22</td>
<td>0.34</td>
<td>0.24</td>
<td>0.34</td>
<td>0.32</td>
<td>0.28</td>
<td>0.37</td>
<td>0.32</td>
<td>0.37</td>
<td>0.22</td>
<td>0.65</td>
<td>0.44</td>
</tr>
<tr>
<td>BOIN</td>
<td>0.58</td>
<td>0.59</td>
<td>0.32</td>
<td>0.38</td>
<td>0.31</td>
<td>0.43</td>
<td>0.33</td>
<td>0.43</td>
<td>0.42</td>
<td>0.36</td>
<td>0.47</td>
<td>0.44</td>
<td>0.37</td>
<td>0.34</td>
<td>0.80</td>
<td>0.47</td>
</tr>
</tbody>
</table>
Figure 2. Accuracy index for dose selection

Note: y-axis ranges differ

Question: how is y-axis calculated?

Authors comments: at 8 dose levels ranking of methods with CRM > BOIN>mTPI
Figure 3. Accuracy index for subject allocation

Note: y-axis ranges differ slightly

Authors comments: Similar performance for methods; as increase dose levels ranking of methods with CRM > BOIN>mTPI
Figure 4. Percent correct selection

Note: y-axis ranges differ slightly

Authors comments: results similar to accuracy; PCS very similar for BOIN and mTPI but with 8 doses BOIN outperforms mTPI.
Figure 5. Percent acceptable within 5%

Note: y-axis ranges differ slightly

Authors comments: results similar to accuracy and PCS.
Questions:

• Accuracy index is calculated with results from simulations that make a dose recommendation (page 294)
  – Does this imply that the studies who terminate early are not included?
  – In table II is the denominator number of simulations or number of simulations that made a dose recommendation?
  – PCS include all simulations in denominator?
Questions:

• For the Scenario 9 (S9) and S12 4 dose designs & S1 6 and 8 dose designs:
  – Terminating early is the correct thing to do and have correctly not identified the MTD
  – So, should all of the simulations be included when compute the proportion of simulations recommending the true MTD (Table II)?
  – Here termination counts as identifying the true MTD. If so, then the results would be as follows in green.

<table>
<thead>
<tr>
<th></th>
<th>CRM</th>
<th>mTPI</th>
<th>BOIN</th>
<th></th>
<th>CRM</th>
<th>mTPI</th>
<th>BOIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>D4S9</td>
<td>0.5</td>
<td>0.26</td>
<td>0.3</td>
<td>D4S9</td>
<td>0.74</td>
<td>0.76</td>
<td>0.77</td>
</tr>
<tr>
<td>D4S12</td>
<td>0.34</td>
<td>0.19</td>
<td>0.24</td>
<td>D6S1</td>
<td>0.74</td>
<td>0.79</td>
<td>0.8</td>
</tr>
<tr>
<td>D6S1</td>
<td>0.43</td>
<td>0.22</td>
<td>0.24</td>
<td>D6S1</td>
<td>0.8</td>
<td>0.71</td>
<td>0.72</td>
</tr>
</tbody>
</table>
Questions:

- For all other designs except S9 and S12 4 dose designs and S1 6 and 8 dose designs-
  - Terminating early is the incorrect thing to do and therefore, have not found the MTD when should have.
  - Therefore, should all of the simulations be included when compute the proportion of simulations recommending the true MTD for these designs (Table II)?
Questions:

• How do these results compare to the paper by Liu and Yuan?
• This paper compared CRM, mTPI and BOIN approaches
• Simulation parameters:
  – 6 dose levels
  – Maximum sample size of 36 patients in 12 cohorts of size 3
  – Target toxicity $\theta=0.25$
  – Tolerance level $(0.15,0.35)$. Used default tolerance values $(0.6\theta,1.4\theta)$
  – Equal prior probabilities
Questions:

• Liu and Yuan simulation parameters:
  – CRM used power model with
    • $a \sim N(0, 1.24^2)$
    • Skeleton based on Lee and Cheung (2009) method: (0.01, 0.08, 0.25, 0.46, 0.65, 0.79)
    • Skipping dose level was not allowed in CRM
    • Toxicity scenario randomly selected using approach of Paoletti et al (2004)
Questions:

• Liu and Yuan simulation performance measures:
  – MTD selection percentage
  – Average percentage of patients treated at MTD
  – Average toxicity rate
  – Average sample size
  – Risk of poor allocation – defined as % of simulations in which number of patients allocated to MTD is less than standard non-sequential design (equal n’s/dose)
  – Risk of high toxicity - % simulations with total number of toxicities greater than that observed if treat all subjects at MTD.
Questions:

• Liu and Yuan Results:
  – Similar average level of performance for CRM, BOIN, mTPI
    • MTD selection %
    • Average % of subjects who are treated at the MTD
    • Average toxicity rate
  – Risk of poor allocation decisions
    • Local BOIN design outperformed the other designs-14-16% lower than CRM and 11% lower than mTPI
  – Risk of high toxicity – BOIN performed better
Questions:

• Yuan, Hess, Hilsenbeck and Gilbert CCR paper compared 3+3, BOIN, mTPI
  – 5 dose levels
  – Maximum sample size of 30 patients
  – Four target toxicity rates (0.15,0.20,0.25,0.30)
  – 16 various scenarios for each target toxicity rate
  – 10,000 simulations

• Performance measures
  – Percent correct selection of MTD (PCS)
  – Average number of patients allocated to MTD
  – Risk of overdosing
  – Risk of underdosing
Questions:

• Results comparing BOIN to mTPI:
  – Correct selection of MTD- BOIN better performance at the lower DLT rates (0.15,0.20)
  – Average number of patients allocated to MTD- BOIN better performance at lower DLT rates (0.15,0.20)
  – Risk of overdosing patients-
    • mTPI highest risk when target DLT rates are 0.2-0.3-in some cases assigning more than 60-80% of patients to doses above the MTD
    • 3+3 conservative-as expected
    • BOIN in between 3+3 and mTPI
  
• CRM not considered in this paper