Sensitivity study of dose-finding methods

Sarah Zohar\textsuperscript{1} John O’Quigley\textsuperscript{2}

1. Inserm, UMRS 717, Biostatistic Department, Hôpital Saint-Louis, Paris, France
2. Inserm, Université Paris VI, Paris, France.
Sensitivity study

Sensitivity in trial allocation and final recommendation

1. To errors in the responses
   - Undetected toxicities
   - Incorrectly recorded toxicities as dose limiting toxicities

2. To the choice of arbitrary parameters of the design
   - Dose-toxicity working model
1. Sensitivity to errors in responses (1)

- Probability of toxicity

\[ R_i = \Pr(Y_j = 1 | X_j = d_i) \]

- To study this we define a new variable,
  - \( V_j \), which corresponds to the true toxicity for patient \( j \).
  - \( Y_j \) our observation which we would hope to be as often as possible the same as \( V_j \).
1. Sensitivity to errors in responses (2)

- $\lambda_1$ corresponds to the probability of incorrectly missing a toxicity and classifying it as a non-toxicity.
  \[
  \Pr(Y_j = 1 | V_j = 1) = 1 - \lambda_1
  \]

- $\lambda_2$ corresponds to the probability of classifying a non-toxicity as a toxicity.
  \[
  \Pr(Y_j = 0 | V_j = 0) = 1 - \lambda_2
  \]

- The probability of observing a toxicity is given by:
  \[
  \Pr(Y_j = 1) = (1 - \lambda_1)\Pr(V_j = 1) + \lambda_2\Pr(V_j = 0)
  \]

- The probability of not observing a toxicity is given by:
  \[
  \Pr(Y_j = 0) = \lambda_1\Pr(V_j = 1) + (1 - \lambda_2)\Pr(V_j = 0)
  \]
Simulations (1)

- Simulations over 200 dose-toxicity relationships
Simulations (2)

- N=36
- Toxicity target=0.3
- 6 dose levels
- 1000 independent replications of each scenario
- Three dose-finding designs:
  - '3+3'
  - group up-and-down design \( UD(s, c_L, c_U) \) with UD \( s = 2, c_L = 0 \) and \( c_U = 1 \)
  - LCRM with working model \( \alpha_1 = 0.1, \alpha_2 = 0.3, \alpha_3 = 0.5, \alpha_4 = 0.6, \alpha_5 = 0.7 \) and \( \alpha_6 = 0.8 \)
$\lambda_1$ : The probability of incorrectly missing a toxicity and classifying it as a non-toxicity

Cumulative distribution of errors where the error itself is defined to be the arithmetic distance between the probability of toxicity at the level recommended and the probability of toxicity at the MTD.
$\lambda_2$: The probability of classifying a non-toxicity as a toxicity
The error associated with incorrectly recorded toxicities has a higher influence on the trial final recommendation than the error associated with undetected toxicities.

The first type of error is where we record an actual toxicity as a non-toxicity and the impact of this is more complex. As far as recommendation of the MTD is concerned then our results show that, as long as the rate of errors of the first type is not too high, the overall design is robust to this.

At the same time, from a clinical viewpoint, those patients treated in the dose-finding study itself have a greater risk of being exposed to doses which are higher than otherwise they would have been.
2. Robustness to dose-toxicity working model

*Would we have recommended the same MTD if we had worked with a CRM design specified differently?*

- Retrospective analysis*

  - Special considerations needed to analyze sequential data retrospectively.

  - Observations need to be correctly weighted. Weights are calculated at each dose based on observed data and response at each dose level.

  - Weights are used to obtain an estimate of model parameters.

  - Estimated model parameters provide an estimate of the MTD

*O'Quigley J., Biometrics 2005; 61: 749–56*
Robustness to dose-toxicity working model

Retrospective analysis

- Weights $w_j(d_i)$ are calculated for all available doses

\[
W_j(a) = \sum_{i=1}^{k} w_j(d_i) U_{ij}(a)
\]

with $U_{ij}(a)$ the log-likelihood function that can be rewritten as:

\[
U_j(a) = \frac{1}{j} \sum_{i=1}^{j} \left\{ y_i \frac{\psi'}{\psi}(x_i, a) + (1 - y_i) \frac{-\psi'}{1 - \psi}(x_i, a) \right\}
\]

- The solution to the equation $W_n(a)$ is obtained when:

\[
\sum_{i=1}^{k} w_n(d_i) \left[ R_i \frac{\psi'}{\psi}(x_i, a) + (1 - R_i) \frac{-\psi'}{1 - \psi}(x_i, a) \right] = 0
\]

- The weighted are used to obtain an estimate of the model parameter $a$
Simulations (1)

### Data Table

<table>
<thead>
<tr>
<th>Working Model</th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$d_3$</th>
<th>$d_4$</th>
<th>$d_5$</th>
<th>$d_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WM 1</td>
<td>0.20</td>
<td>0.30</td>
<td>0.45</td>
<td>0.60</td>
<td>0.70</td>
<td>0.80</td>
</tr>
<tr>
<td>WM 2</td>
<td>0.05</td>
<td>0.10</td>
<td>0.30</td>
<td>0.50</td>
<td>0.60</td>
<td>0.65</td>
</tr>
<tr>
<td>WM 3</td>
<td>0.01</td>
<td>0.02</td>
<td>0.09</td>
<td>0.30</td>
<td>0.40</td>
<td>0.45</td>
</tr>
<tr>
<td>WM 4</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.07</td>
<td>0.30</td>
<td>0.41</td>
</tr>
<tr>
<td>WM 5</td>
<td>0.01</td>
<td>0.02</td>
<td>0.05</td>
<td>0.07</td>
<td>0.10</td>
<td>0.30</td>
</tr>
<tr>
<td>WM 6</td>
<td>0.01</td>
<td>0.10</td>
<td>0.30</td>
<td>0.45</td>
<td>0.65</td>
<td>0.85</td>
</tr>
</tbody>
</table>
Simulations (2)

<table>
<thead>
<tr>
<th>Working Model</th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$d_3$</th>
<th>$d_4$</th>
<th>$d_5$</th>
<th>$d_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WM 7</td>
<td>0.35</td>
<td>0.50</td>
<td>0.55</td>
<td>0.60</td>
<td>0.65</td>
<td>0.70</td>
</tr>
<tr>
<td>WM 8</td>
<td>0.01</td>
<td>0.03</td>
<td>0.07</td>
<td>0.10</td>
<td>0.15</td>
<td>0.22</td>
</tr>
<tr>
<td>WM 9</td>
<td>0.10</td>
<td>0.20</td>
<td>0.30</td>
<td>0.40</td>
<td>0.50</td>
<td>0.60</td>
</tr>
<tr>
<td>WM 10</td>
<td>0.15</td>
<td>0.20</td>
<td>0.25</td>
<td>0.30</td>
<td>0.35</td>
<td>0.40</td>
</tr>
<tr>
<td>WM 11</td>
<td>0.21</td>
<td>0.22</td>
<td>0.23</td>
<td>0.24</td>
<td>0.25</td>
<td>0.26</td>
</tr>
</tbody>
</table>

Toxicity target

NY 2009

13 / 21
Sensitivity study
S. Zohar, J. O’Quigley

Robustness to working model

Simulations (3)

Figure: Cumulative distribution of recommendation errors for 11 working models
For all scenarios and each working model, the percentage of correct selection were as follow 0.97, 0.92, 0.93, 0.97, 0.75, 0.89, 0.92, 0.94, 0.96, 0.95 and 0.61 respectively.

<table>
<thead>
<tr>
<th>Working Model</th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$d_3$</th>
<th>$d_4$</th>
<th>$d_5$</th>
<th>$d_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WM 5</td>
<td>0.01</td>
<td>0.02</td>
<td>0.05</td>
<td>0.07</td>
<td>0.10</td>
<td>0.30</td>
</tr>
<tr>
<td>WM 11</td>
<td>0.21</td>
<td>0.22</td>
<td>0.23</td>
<td>0.24</td>
<td>0.25</td>
<td>0.26</td>
</tr>
</tbody>
</table>
Illustration

Retrospective analysis of the semisynthetic homoharringtonine trial*

- Eighteen patients with advanced acute myeloid leukemia
- CRM
- logistic form with a fixed intercept of 3.0
- Toxicity target 33%
- 5 dose levels: 0.5, 1, 3, 5 and 6 $mg/m^2/d$
- 5 pseudo working doses -5.94, -5.20, -4.73, -3.71 and -3.00
- At the end of the trial the MTD was selected to be the fourth dose levels

Would the estimated MTD have been the same had we used a different model?

*Levy et al., Br J Cancer 2006; 95(3):253-9
# Illustration

## Sensitivity Study

S. Zohar, J. O'Quigley

## Sensitivity to Errors in Responses

### Robustness Analysis

<table>
<thead>
<tr>
<th>Dose</th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$d_3$</th>
<th>$d_4$</th>
<th>$d_5$</th>
<th>Recommended Dose</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_{18}(d_i)$</td>
<td>3</td>
<td>-</td>
<td>3</td>
<td>12</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>DLTs ($t_{18}(d_i)$)</td>
<td>0</td>
<td>-</td>
<td>1</td>
<td>4</td>
<td>-</td>
<td>$d_4$</td>
</tr>
<tr>
<td>Estimated $\hat{R}(d_i)$ by CRM</td>
<td>0.06</td>
<td>0.12</td>
<td>0.17</td>
<td>0.36</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>Empirical $\hat{R}(d_i)$</td>
<td>0</td>
<td>-</td>
<td>0.33</td>
<td>0.33</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

### Robustness Analysis

<table>
<thead>
<tr>
<th>Weights</th>
<th>$w_{18}(d_1)$</th>
<th>$w_{18}(d_2)$</th>
<th>$w_{18}(d_3)$</th>
<th>$w_{18}(d_4)$</th>
<th>$w_{18}(d_5)$</th>
<th>Recommended Dose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working Model 1</td>
<td>0.17</td>
<td>0.18</td>
<td>0.23</td>
<td>0.36</td>
<td>0.07</td>
<td>4</td>
</tr>
<tr>
<td>WM 2</td>
<td>0.17</td>
<td>0.18</td>
<td>0.34</td>
<td>0.27</td>
<td>0.05</td>
<td>3</td>
</tr>
<tr>
<td>WM 3</td>
<td>0.17</td>
<td>0.18</td>
<td>0.34</td>
<td>0.27</td>
<td>0.05</td>
<td>3</td>
</tr>
<tr>
<td>WM 4</td>
<td>0.17</td>
<td>0.18</td>
<td>0.20</td>
<td>0.41</td>
<td>0.04</td>
<td>4</td>
</tr>
<tr>
<td>WM 5</td>
<td>0.17</td>
<td>0.20</td>
<td>0.27</td>
<td>0.19</td>
<td>0.17</td>
<td>3</td>
</tr>
<tr>
<td>WM 6</td>
<td>0.17</td>
<td>0.18</td>
<td>0.34</td>
<td>0.27</td>
<td>0.04</td>
<td>4</td>
</tr>
<tr>
<td>WM 7</td>
<td>0.17</td>
<td>0.21</td>
<td>0.22</td>
<td>0.25</td>
<td>0.16</td>
<td>4</td>
</tr>
<tr>
<td>WM 8</td>
<td>0.17</td>
<td>0.20</td>
<td>0.24</td>
<td>0.24</td>
<td>0.16</td>
<td>3</td>
</tr>
<tr>
<td>WM 9</td>
<td>0.17</td>
<td>0.19</td>
<td>0.22</td>
<td>0.32</td>
<td>0.12</td>
<td>4</td>
</tr>
<tr>
<td>WM 10</td>
<td>0.17</td>
<td>0.20</td>
<td>0.23</td>
<td>0.22</td>
<td>0.18</td>
<td>3</td>
</tr>
<tr>
<td>WM 11</td>
<td>0.31</td>
<td>0.18</td>
<td>0.22</td>
<td>0.04</td>
<td>0.25</td>
<td>3</td>
</tr>
</tbody>
</table>
In this illustration, 5 working models would have recommend the fourth dose level and 6 working models would have recommend the third dose level at the end of the trial.
Conclusions

- Within the class of working models chosen, the continual reassessment method is robust.
- We can make an approximate division of the class of all potential models into those which we consider to be “reasonable” models and those which are not. The study here attempts to tie down in a more rigorous way the concept of being “reasonable.” A “reasonable” model is one that would exhibit good robustness properties.