A Bayesian decision-theoretic approach to incorporating pre-clinical information into phase I clinical trials

Haiyan Zheng, Lisa V. Hampson

Medical and Pharmaceutical Statistics Research Unit
Department of Mathematics and Statistics, Lancaster University

Industry collaborator: Dr Alun Bedding (AstraZeneca)
Clinical advisor: Dr Malcolm Mecleod (Edinburgh University)

IDEAS Think Tank, Traunkirchen
23rd June 2016

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 633567. (IDEAS - http://www.ideas-itn.eu/)
Motivation
FDA (2005), Sharma and McNeill (2009), Reigner and Blesch (2002)

Q: How can we use pre-clinical toxicology and pharmacology data to improve the design, conduct and analysis of phase I dose-escalation trials?

Current approaches use pre-clinical data to determine a maximum recommended starting dose (MRSD) using allometric scaling:

- Using toxicology data: Human dose (mg/kg) = NOAEL × (WA/WH)^0.33
- Using PK data: Estimate human PK parameters using allometric scaling, e.g., Cl_H = Cl_A(W_A/W_H)^b
- Scale doses by a safety factor of 10 in case of size-independent differences.

Simply allometry can produce inaccurate predictions of human doses (e.g., diazepam, warfarin) leading to conservative or toxic starting doses.

- May be uncertainty about the best choice of allometric exponent
- Likely to be data on several animal species - which species is most relevant?
Research objectives

To develop a **Bayesian model-based procedure** that uses pre-clinical information based on concerns about its **degree of agreement** with dose-toxicity relationship in humans:

*Is the drug predicted more (or less) potent in humans than it actually is?*
Pre-clinical toxicology data

Commensurability issues

Prior 1: animal data predict excess of DLTs

Prior 2: animal data predict insufficient DLTs

Prior 3: animal data indicate a shallow curve

Prior 4: animal data indicate a steep curve
Phase I clinical trials

- **Principal aim**: to estimate the maximum tolerated dose (MTD) using cumulative toxicity data from an ongoing phase I trial
  - MTD
    - subject to an acceptable risk of dose-limiting toxicity (DLT): $\pi_d$
    - target: $\pi_d \in (0.20, 0.35)$
  - Assumption: toxicity increases monotonically with the dose

- **Approaches/Designs**
  - Algorithmic
    - ‘3+3’ design and more sophisticated up-and-down schemes
  - Model-based
    - Continual reassessment method (CRM) *with a one-parameter ‘working’ model*
    - Bayesian logistic regression method (BLRM) *with a two-parameter dose-toxicity model*
    - …
Bayesian logistic regression method (BLRM)

- Cumulative toxicity data
  - At each dose $d_j$
  - Number of patients: $n_j$
  - Number of DLTs observed: $t_j$

- BLRM
  $$t_j | n_j \sim \text{Binomial}(n_j, p(d_j)),$$
  The dose-toxicity model: \( \log \frac{p(d_j)}{1 - p(d_j)} = \theta_1 + \exp(\theta_2) \log d_j \)

- Prior
  - Bivariate normal distribution: \((\theta_1, \theta_2) \sim MVN(, )\)
  - Pseudo-data envisaged for the lowest and highest doses

<table>
<thead>
<tr>
<th>Dose</th>
<th># Patients</th>
<th># DLTs</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{-1}$</td>
<td>$n_{-1}$</td>
<td>$t_{-1}$</td>
</tr>
<tr>
<td>$d_0$</td>
<td>$n_0$</td>
<td>$t_0$</td>
</tr>
</tbody>
</table>
Bayesian methods for dose-escalation studies

Expressing the prior

- Consider two potential doses $d_i$, $i = \{-1, 0\}$
- $p(d_i) \sim Beta(t_i, u_i)$, where $u_i = n_i - t_i$
- Two beta distributions are independent of one another
- $p(d_i)$ is written as $p_i$ for illustration purposes
- The joint prior density is therefore given by

$$g_0(p_{-1}, p_0) = \prod_{i=-1}^{0} \frac{p_i^{t_i-1}(1 - p_i)^{u_i-1}}{B(t_i, u_i)},$$

and that of $\theta_1$ and $\theta_2$ by

$$h_0(\theta_1, \theta_2) = \prod_{i=-1}^{0} \frac{[\exp(-\theta_1 - \exp(\theta_2) \log d_i) + 1]^{-t_i} [\exp(\theta_1 + \exp(\theta_2) \log d_i) + 1]^{-u_i}}{B(t_i, u_i)} \times \exp(\theta_2) \left| \log \left( \frac{d_{-1}}{d_0} \right) \right|$$

using the Jacobian transformation.

(Tsutakawa, 1975; Whitehead and Williamson, 1998)
Bayesian methods for dose-escalation studies

Expressing the prior

- Pseudo-data indicates both the position of the parameters \((\theta_1, \theta_2)\) and the strength of such an opinion: as if
  - \(n_1\) patients were administered the lower dose, \(t_1\) DLTs observed
  - \(n_0\) patients were administered the higher dose, \(t_0\) DLTs observed
- Strength \(n_1 = n_0 = 3\) is deemed proper

Comments:
1. These are operational priors, calibrated to ensure the dose-escalation scheme has favourable operating characteristics
2. Informative priors can be formed using pre-clinical toxicology data
Bayesian methods for dose-escalation studies

Expressing the prior

- Pseudo-data indicates both the position of the parameters \((\theta_1, \theta_2)\) and the strength of such an opinion: as if
  - \(n_{-1}\) patients were administered the lower dose, \(t_{-1}\) DLTs observed
  - \(n_0\) patients were administered the higher dose, \(t_0\) DLTs observed
- Strength \(n_{-1} = n_0 = 3\) is deemed proper

\[ \text{Strength } n_{-1} = n_0 = 3 \text{ is deemed proper} \]
Bayesian methods for dose-escalation studies

Expressing the prior

- Pseudo-data indicates both the position of the parameters \((\theta_1, \theta_2)\) and the strength of such an opinion: as if
  - \(n_{-1}\) patients were administered the lower dose, \(t_{-1}\) DLTs observed
  - \(n_0\) patients were administered the higher dose, \(t_0\) DLTs observed
- Strength \(n_{-1} = n_0 = 3\) is deemed proper

Comments:

1. These are operational priors, calibrated to ensure the dose-escalation scheme has favourable operating characteristics
2. Informative priors can be formed using pre-clinical toxicology data
Addressing the potential prior-data conflict

Marginal distribution function for $p_j$

- The joint prior density can be given as

$$h_0(\theta_1, \theta_2) = \prod_{i=-1}^{0} \frac{\left[\exp(-\theta_1 - \exp(\theta_2) \log d_i) + 1\right]^{-t_i} \left[\exp(\theta_1 + \exp(\theta_2) \log d_i) + 1\right]^{-u_i}}{B(t_i, u_i)} \times \exp(\theta_2) \left|\log \left(\frac{d_{-1}}{d_0}\right)\right|$$

- We further express the prior as the joint pdf. of $p_j$ and $\theta_2$:

$$m_0(p_j, \theta_2) = \prod_{i=-1}^{0} \frac{\left[\exp \left(-\log \frac{p_j}{1-p_j} + \exp(\theta_2) \log \frac{d_j}{d_i}\right) + 1\right]^{-t_i} \left[\exp \left(\log \frac{p_j}{1-p_j} - \exp(\theta_2) \log \frac{d_j}{d_i}\right) + 1\right]^{-u_i}}{B(t_i, u_i)} \times \frac{1}{p_j(1-p_j)} \cdot \exp(\theta_2) \cdot \left|\log \left(\frac{d_{-1}}{d_0}\right)\right|$$

- The marginal distribution for $p_j, j = \{-1, 0, 1, 2, \ldots, J\}$, is

$$g(p_j) = \int m_0(p_j, \theta_2) d\theta_2.$$

Notation: $p(d_j)$ is written as $p_j$ for illustration purposes.
Addressing the potential prior-data conflict

1. The considered future observations of DLT ($\tilde{y} = 1$) or not ($\tilde{y} = 0$) are predicted using the prior predictive distribution of $\tilde{y}$,

$$P\{Y = \tilde{y}\} = \int_{p_j} f(\tilde{y}|p_j)g(p_j)dp_j,$$

where $f(\cdot)$ is the link function with the DLT probability $p_j$, and the prior $g(p_j)$ is formed based on pre-clinical studies.

2. Predictions are optimal in the sense of maximising the prior expected utility

$$\bar{u}(\eta) = \sum_{\tilde{y}} u(\tilde{y},\eta)P\{Y = \tilde{y}\},$$

where $u(\tilde{y},\eta)$ is the utility function that rewards/penalises predictions of $\tilde{y}$ as $\eta$:

$$u(\tilde{y},\eta) = \begin{cases} 0, & \text{if } \eta = 0 \text{ while } \tilde{y} = 1 \\ c, & \text{if } \eta = 1 \text{ while } \tilde{y} = 0 \\ 1, & \eta = \tilde{y} \in \{0, 1\} \end{cases}$$

Note that $0 < c < 1$. 
Addressing the potential prior-data conflict (Cont’d)

3. The optimal prediction $\hat{\eta}$ is therefore chosen out of the whole decision set $\mathcal{H} = \{0, 1\}$ by maximising the prior expected utility $\bar{u}(\eta)$:

$$\hat{\eta} = \arg\max_{\eta \in \mathcal{H}} \sum_{\tilde{y}} u(\tilde{y}, \eta) \mathbb{P}\{Y = \tilde{y}\}.$$

4. A $2 \times 2$ contingency table for the actual versus predicted DLTs and no-DLTs

<table>
<thead>
<tr>
<th></th>
<th>Rewards and Penalties</th>
<th>Human counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual obs ($y$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>No-DLT</td>
<td>DLT</td>
</tr>
<tr>
<td>Prior predictions ($\hat{\eta}$)</td>
<td>$u_{00}$</td>
<td>$u_{10}$</td>
</tr>
<tr>
<td></td>
<td>$u_{01}$</td>
<td>$u_{11}$</td>
</tr>
<tr>
<td>No-DLT</td>
<td>$n_{00}$</td>
<td>$n_{10}$</td>
</tr>
<tr>
<td>DLT</td>
<td>$n_{01}$</td>
<td>$n_{11}$</td>
</tr>
</tbody>
</table>

5. The predictive utility is then calculated at dose $d_j$ as $U_{j}^{\text{pred}} = \sum_{l=0}^{1} \sum_{m=0}^{1} u_{lm} n_{lm}$, and the predictive accuracy as

$$a_j = \frac{U_{j}^{\text{pred}}}{\sum_{l=0}^{1} \sum_{m=0}^{1} n_{lm}}.$$

6. The average $\bar{a} = \sum a_j / K$ measures the commensurability of animal and human data.
Prior tuning

- $\bar{a} \in [0, 1)$, computed at each interim analysis, quantifies the degree of agreement between animal and human toxicology data.

- Mixture prior with a weakly informative component will be considered:
  
  \[ f(\theta) = w \times g(\theta) + (1 - w) \times p(\theta), \]

- the weight $w$ is a function of $\bar{a}$, allowing for a flexible borrowing especially when the human data is sparse.
- $w$ will determine how quickly pre-clinical information is discounted in the case of a prior-data conflict.
- $p(\theta)$ constructed based on the operational priors (Whitehead and Williamson, 1998) works reasonably well.
The pre-clinical data weight

\[ f(\theta) = w \times g(\theta) + (1 - w) \times p(\theta) \]

- The weight \( w \) is defined as a function of predictive accuracy \( \bar{a} \) in the relevance of the information time \( n/N \) (say, \( N = 24 \))
- It governs how influential the pre-clinical data are as the trial proceeds
- One possible function considered in our study is \( w = \bar{a}^{1 - \log(n/N)} \)

With the accumulation of human data, \( \bar{a} \), serving as a measurement of prior-data conflict, becomes more reliable

\( \bar{a} \) is computed dynamically at each interim analysis
Dose recommendation

- Fully Bayesian ‘patient gain’ criterion

\[ G_1 = \int (p(d_i) - \pi_d)^{-2} f(p(d_i)|x) dp(d_i), \quad \pi_d \in (0.20, 0.35) \]

- Fully Bayesian ‘determinant gain’ criterion

\[ G_2 = \int (\det I(\theta))^{-1} f(\theta|x) d\theta, \quad \theta = (\theta_1, \theta_2) \]

where \( x \) denote the human data accumulated from each patient cohort, and \( f(\cdot|x) \) the (mixture) posterior distribution.

Safety constraint:
Controlling the probability of excessive toxicity at level \( \delta \):

\[ \int_\gamma^1 g(p(d_j)) dp(d_j) \leq \delta, \]

where \( \gamma \) is some threshold above which the risk of toxicity is considered excessively high.
Simulations

- Six doses to be assessed: 2.2, 3.1, 3.3, 3.7, 4.0, 4.9 mg/m²
- A total number of 24 patients, cohort size of 1
- Targeted probability $\pi_d = 0.20$; that is, $TD_{20}$ is sought
- For all simulations, the true dose-response curve was assumed to be

$$\log\frac{p(d_j)}{1 - p(d_j)} = \theta_1 + \exp(\theta_2) \log d_j,$$

- Prior distribution for the model parameters was constructed as a mixture prior

$$f(\theta) = w \times g(\theta) + (1 - w) \times p(\theta),$$

with given information on the lowest and highest doses:

- For instance, $p(d_1) \sim Beta(5, 95)$ and $p(d_6) \sim Beta(95, 5)$ are summarised from pre-clinical toxicology studies
- Stochastic optimisation approach technique is used to derive the closed form for the bivariate normal distributions $g(\theta)$ and $p(\theta)$
- Full (mixture) posterior distributions are obtained via MCMC simulations
- Results are presented based on 2000 simulated trials per scenario
Prior calibration

Prior1 approximated as:

\[MVN : (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)\]
\[= (-8.20, 1.92, 0.075, 0.037, 0.98)\]

Prior2 approximated as:

\[MVN : (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)\]
\[= (-5.76, 1.25, 0.094, 0.029, 0.78)\]

Prior3 approximated as:

\[MVN : (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)\]
\[= (-3.28, 0.66, 0.120, 0.032, 0.75)\]
Prior calibration (Cont’d)

Prior4 approximated as:

$$MVN: (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$$
$$= (-29.78, 3.16, 0.082, 0.062, 0.28)$$

Prior5 approximated as:

$$MVN: (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$$
$$= (-10.74, 2.11, 0.101, 0.025, 0.95)$$

Operational priors approximated as:

$$MVN: (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$$
$$= (-3.73, 0.84, 0.645, 0.367, 0.057)$$

H Zheng, LV Hampson (Lancs Uni)
Incorporating pre-clinical info
Simulation results

**Table:** Results for incorporating pre-clinical toxicology data or not are presented based on 2000 simulated trials, each with size of 24, per scenario. TD20 was sought with fully Bayesian determinant gain criterion used for dose-escalation.

<table>
<thead>
<tr>
<th>Dose (mg/m²)</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>true ( p(d_j) )</td>
<td>0.014</td>
<td>0.197</td>
<td>0.291</td>
<td>0.513</td>
<td>0.667</td>
<td>0.914</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario</th>
<th>HLInc</th>
<th>SelMTD (%)</th>
<th>avgNo.Pts</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SC1: animal data predict excess of DLTs</td>
<td>5.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLInc</td>
<td>SelMTD (%)</td>
<td>17.50</td>
<td>56.40</td>
<td>18.50</td>
<td>6.90</td>
<td>7.00</td>
</tr>
<tr>
<td>avgNo.Pts</td>
<td>5.77</td>
<td>11.89</td>
<td>3.62</td>
<td>1.94</td>
<td>0.66</td>
<td>0.12</td>
</tr>
<tr>
<td>SC2: animal data predict insufficient DLTs</td>
<td>5.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLInc</td>
<td>SelMTD (%)</td>
<td>13.70</td>
<td>44.30</td>
<td>36.10</td>
<td>5.70</td>
<td>2.00</td>
</tr>
<tr>
<td>avgNo.Pts</td>
<td>5.66</td>
<td>7.68</td>
<td>6.60</td>
<td>3.84</td>
<td>0.16</td>
<td>0.06</td>
</tr>
<tr>
<td>SC3: animal data indicate a shallow curve</td>
<td>4.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLInc</td>
<td>SelMTD (%)</td>
<td>13.10</td>
<td>72.80</td>
<td>11.70</td>
<td>2.30</td>
<td>1.00</td>
</tr>
<tr>
<td>avgNo.Pts</td>
<td>5.22</td>
<td>14.47</td>
<td>3.02</td>
<td>0.68</td>
<td>0.56</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Simulations results (*Cont’d*)

Table: *Continued.*

<table>
<thead>
<tr>
<th>Dose (mg/m²)</th>
<th>d₁</th>
<th>d₂</th>
<th>d₃</th>
<th>d₄</th>
<th>d₅</th>
<th>d₆</th>
</tr>
</thead>
<tbody>
<tr>
<td>true $p(d_j)$</td>
<td>0.014</td>
<td><strong>0.197</strong></td>
<td>0.291</td>
<td>0.513</td>
<td>0.667</td>
<td>0.914</td>
</tr>
</tbody>
</table>

SC4: animal data indicate a steep curve

<table>
<thead>
<tr>
<th>HLInc</th>
<th>SelMTD (%)</th>
<th>17.20</th>
<th><strong>57.50</strong></th>
<th>17.70</th>
<th>7.10</th>
<th>0.50</th>
<th>0.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>avgNo.Pts</td>
<td>6.78</td>
<td><strong>10.14</strong></td>
<td>4.32</td>
<td>1.99</td>
<td>0.64</td>
<td>0.13</td>
<td></td>
</tr>
</tbody>
</table>

SC5: animal data are perfectly commensurate with human data

<table>
<thead>
<tr>
<th>HLInc</th>
<th>SelMTD (%)</th>
<th>0.00</th>
<th><strong>95.40</strong></th>
<th>4.60</th>
<th>0.00</th>
<th>0.00</th>
<th>0.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>avgNo.Pts</td>
<td>1.00</td>
<td><strong>22.83</strong></td>
<td>0.17</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

SC6: animal data are completely discarded

<table>
<thead>
<tr>
<th>$w = 0$</th>
<th>SelMTD (%)</th>
<th>16.90</th>
<th><strong>56.10</strong></th>
<th>21.50</th>
<th>5.00</th>
<th>0.50</th>
<th>0.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>avgNo.Pts</td>
<td>6.10</td>
<td><strong>10.93</strong></td>
<td>4.27</td>
<td>1.97</td>
<td>0.62</td>
<td>0.10</td>
<td></td>
</tr>
</tbody>
</table>
Simulations results (Cont’d)

Prior1: animal data predict excess of DLTs

Prior2: animal data predict insufficient DLTs

Incorporating pre-clinical info

IDEAS Think Tank
Simulations results (Cont’d)

Prior 3: animal data indicate a shallow curve

Prior 4: animal data indicate a steep curve

H Zheng, LV Hampson (Lancs Uni)

Incorporating pre-clinical info

IDEAS Think Tank
Summary

- Incorporating pre-clinical data when they are commensurate with human data will lead to more efficient dose-escalation decisions and greater estimation precision
  - More patients are allocated to the target dose
  - Dose recommendations are sensible and robust
- The pre-clinical data will essentially be discounted if they are in clear conflict with the dose-toxicity relationship in humans
References

FDA. Guidance for Industry. Estimating the maximum safe starting dose in initial clinical trials for therapeutics in adult healthy volunteers. 2005


Mixture prior derivation

\[ f(\theta) = w \times g(\theta) + (1 - w) \times p(\theta) \]

1. Synthesise information from pre-clinical toxicology studies to describe probabilities of toxicity \( p(d) \) especially for the lowest and highest doses

2. Derive \( g(p(d_j)) \), the distribution for \( p(d_j) \), at each candidate dose \( d_j \)

3. Calibrate \( g(p(d_j)) \) to find a prior \( g(\theta) \) in a way such that its implied prior information on the \( p(d) \) scale is in good agreement with \( g(p(d_j)) \)
   - using stochastic optimisation methods to minimise the discrepancy between the specified quantiles \( Q \) and the quantiles \( Q' \) arising from \( g(\theta) \):
     \[
     C(Q, Q') = \max_{j, k} |q_{jk} - q'_{jk}|, \ j = 1, \ldots, J, \ k = 1, \ldots, K
     \]
Mixture prior derivation

\[ f(\theta) = w \times g(\theta) + (1 - w) \times p(\theta) \]

4. The BVN \( p(\theta) \) can also be approximated using this quantile-based approach.

5. Specifically, prior information at a given dose is summarised with three quantiles \( q_{jk} = \{ q_j(0.025), q_j(0.5), q_j(0.975) \} \), which correspond to the median and 95% percent credible intervals.

6. Introducing a weakly-informative component can potentially guarantee the dose-escalation procedure performs properly, even (in an extreme case) when the animal data are completely incommensurate with the human data.