Predicting Phase III Success Using a Combination of Phase II Results and Historical Reference Data

David Burt
MBSW
May 21, 2013
It is common for dry-eye programs to fail in late stage development.

Therefore, many experts in this therapeutic area tend to be skeptical about the likelihood of reproducibility.

Our dry-eye project team would like to base milestone decisions on quantitative predictions of phase 3 success.

This presentation will describe a method for doing quantitatively what the experts have done qualitatively and describe why conventional p-value based decisions should be avoided.
Outline

- Historic data and meta-analyses basics
- Importance of inter-study variability
- Simple meta-analysis model
- Case study with more innovative model
- Practical considerations
- A simulation study
- Conclusions
Meta-Analysis Basics

The graph illustrates the distribution of Delta Density across different values of Delta. The density is highest around Delta = 0, indicating a normal distribution with a peak at the center and spreading outwards. The dotted line represents a specific distribution, likely a normal distribution, given the bell shape and symmetry around the center.
Ignoring the Historic Data

P-Value < 0.0001

Delta Density

P-Value < 0.0001
Bayesian Methods (Ignoring Historic Data)

Bayesian Posterior
\[ P(\mu > 0) = 100\% \]
Importance of inter-study variability

True Mean = 0
P(Xbar > 0) = 50%
“How to” Fix the Problem?

- Incorporate historic data via a formal meta-analysis
- This will provide estimates of inter and intra study variability
Historical Data for 15 Studies N=30 per Study
Random Effect Meta-analysis Model

- A simple formulation of this problem is the following Random Effect Model

\[ y_{ij} = \mu + \gamma_i + \varepsilon_{ij} \]

- Where \( \mu \) is the population mean, \( \gamma \) is a random effect for study and \( \varepsilon \) is a random (within-study) error term

- We use a Bayesian construct

\[ \gamma_i \sim N(0, \sigma_B^2) \quad \mu \sim N(\mu_0, \kappa_0 \sigma_W^2 + \sigma_B^2) \]

\[ \varepsilon_{ij} \sim N(0, \sigma_W^2) \quad \tau_X \equiv \frac{1}{\sigma_X^2} \sim G(\text{Shape}, \text{Rate}) \]
Observed Data for Multiple Studies (Normal MCMC)

**Observed Data**

- **YIJ**: for(I IN 1 : K) for(J IN 1 : NI)
- **Study Mean**
- **Intra-Study**
- **Inter-Study**
- **Grand Mean**

**Predictive Density**

- **P(θ)**
- **New Mean**
- **New Data**
- **for(L IN 1 : M)**

*Note: The diagram illustrates the relationships between observed data, study means, inter-study, intra-study, and the predictive density of the parameter θ.*
Incorporating Inter-Study Variability

\[ P(\bar{X} > 0) = 70\% \]

Assurance = 56%
Potential for Innovation

- Estimate the degree of inter-study variability at or before the POC
  - In the absence of full meta-analysis data
  - Without running replicate clinical trials
- Use these estimates to make corrected inferences in the POC
- Build predictive models and guide decision criteria that are more consistently correct
- Inform seamless development programs that are more consistently correct
Hierarchical Assurance Model

- Given single study data for the investigational drug product (no historical data), and historical data for the placebo combine them in the following model:

\[ y_{ij} = \mu + \gamma_i + \varepsilon_{ij} \]

- Historic Meta-data (PBO Only)
- Single Study Data (active & PBO)
Meta Analysis Model

Observed Data

\[ \text{Y}_{IJ} \]

\( \text{for}(I \ \text{IN} \ 1 : K) \)

\( \text{for}(J \ \text{IN} \ 1 : N_I) \)

Predictive Density

\[ P(\theta) \]

\[ \text{New Mean} \]

\[ \text{New Data} \]

\( \text{for}(L \ \text{IN} \ 1 : M) \)

\[ \text{Grand Mean} \]

\[ \text{Inter-Study} \]

\[ \text{Intra-Study} \]
Single Study Model (Broken into two groups)
Hierarchical Assurance Model

for(i IN 1 : K)
TAU.N
M2
M1
ShapeB
RateB
RateW
ShapeW
X beta[i] X [i]
YBAR_meta[i] ... S2N_ACT YNEW_ACT
YBAR_ACT
S2_ACT
S2_meta[i]
MU_ACT
TAU.M
Beta
YNEW_PBO
S2N_PBO
P3TEST_ST
P3Success
Omega
Beta
Tau.Beta
Beta0
Xbeta[i]
X[i]
N[i]
S2_meta[i]
TAU.meta[i]
YBAR_meta[i]
for(i IN 1 : K)
**Single Study Data for Dry-Eye example**

<table>
<thead>
<tr>
<th></th>
<th>Placebo</th>
<th>Active Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Mean</td>
<td>-0.50</td>
<td>-3.0</td>
</tr>
<tr>
<td>S2</td>
<td>16.1</td>
<td>15.8</td>
</tr>
<tr>
<td>Planned P3 Sample Size</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Naïve Assurance Estimate

Effect Size -2.5

Assurance = 92 %
## Historical Placebo Data for Dry-Eye

<table>
<thead>
<tr>
<th>Study</th>
<th>Phase</th>
<th>N per group</th>
<th>Sample Mean</th>
<th>Sample $S^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>103</td>
<td>-1.8</td>
<td>4.12</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>176</td>
<td>-0.48</td>
<td>1.13</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>58</td>
<td>1.5</td>
<td>8.12</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>93</td>
<td>-1.58</td>
<td>4.65</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>103</td>
<td>-3.0</td>
<td>11.33</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>47</td>
<td>-2.0</td>
<td>11.75</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>292</td>
<td>-1.95</td>
<td>26.28</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>142</td>
<td>-3.5</td>
<td>7.10</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>95</td>
<td>-4.8</td>
<td>10.45</td>
</tr>
</tbody>
</table>
Hierarchical Assurance Result

Effect Size -2.5

Hierarchical Assurance = 56 %
**Parameter Estimates**

Parameter Estimates from Hierarchical Assurance Model for the Dry-Eye Example

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>-0.69</td>
<td>0.73</td>
<td>-2.29</td>
<td>0.58</td>
</tr>
<tr>
<td>Intra-Study Variability</td>
<td>11.89</td>
<td>0.51</td>
<td>10.89</td>
<td>12.92</td>
</tr>
<tr>
<td>Inter-Study Variability</td>
<td>3.65</td>
<td>2.56</td>
<td>1.15</td>
<td>10.54</td>
</tr>
<tr>
<td>Treatment Effect</td>
<td>-2.00</td>
<td>2.14</td>
<td>-6.41</td>
<td>2.10</td>
</tr>
</tbody>
</table>

Posterior probability that the active treatment is superior to placebo ($\delta < 0$) is 84.2%
Hierarchical Assurance = 80 %
Effect Size = -4.3
Practical Considerations

- It is helpful to specify the model in terms of sufficient statistics, since subject level historical data are often inaccessible.
- Involve project team members in the selection of studies to be included in the historical data set.
- Try to build a robust set of historical studies that does not include reporting bias.
Hierarchical Model Simulations

- Large Inter-Study Variability & Moderate Bias
- Negligible Inter-Study Variability & Large Bias
- Negligible Inter-Study Variability & No Bias
- Historical Data were simulated in conjunction with one Phase 2 study and one Phase 3 study
  - Phase 2 incorporated in the model
  - Phase 3 used to determine ‘true’ success probability
Simulation Results Scenario 1 (Large Inter-Study)

![Box plots for different methods and assurance levels](image)
Simulation Results Scenario 1 (Large Inter-Study)
Simulation Results Scenario 2 (Large Bias)
Simulation Results Scenario 3
(No Inter-Study Variance & No Bias)
Potential Applications

- Go/No Go Criteria
- Futility and/or Adaptive Allocation
- Study Design
- Business Development Opportunities
- General Portfolio Management
- Much Much More
Conclusions

- Inter-Study Variability can cause our inferences and predictions to be wrong more often than we expect.
- There are solutions that can be used to bring our rate of being wrong back to a controllable level.
- The benefits of applying these solutions early in the development process are quite substantial.
- Clinical statisticians have a unique opportunity to enhance the way drugs are developed if we can build some alignment and a greater understanding of how inter-study variability can be controlled for early in the development process.
Acknowledgements

- Bob Noble
- Mike Fries