Decision Making in Pharma: Science or Scientology?

Kevin J Carroll

PSI, London, May 2014
Contents

• Context
• Typical Phase III design and the Probability of Success in Phase III given Phase II
• Some considerations vs PoS
• Simple suggestions to enhance effective decision making
• Summary
High failure rates in latter stages of development continues to challenge industry

• Failure rates of 80% in Phase II and 50% in Phase III have recently been reported$^{1,2}$.

• Two thirds of Phase III failures due to not demonstrating a positive treatment effect

• Reflects poorly on the quality of Phase II design and decision making$^2$

• Phase II design and effective use of Phase II data critical to effective business decision making and Phase III PoS

Typical Phase III Situation

• Example: Advanced oncologic setting
• Phase II with N=120 estimates a hazard ratio (HR) of 0.75, 95% CI (0.47, 1.20)
• Result seen as promising; Phase III planned to test the hypothesis drug improves progression-free survival with a HR of 0.75 compared to control
• Statistician calculates the required sample size to provide a desired Power and $\alpha$ level
  • N=950* for 90% power, $\alpha=2.5\%$ 1-sided

• What is the ‘Probability Of Success’ of the planned Phase III?

*9months median on control, 1 yr accrual, 6 mo follow-up
Typical Phase III Set-Up

• HR = 1 with probability 100%
or
HR = 0.75 with probability 100%
• The HR cannot be anything higher or lower or in-between

<table>
<thead>
<tr>
<th>Phase III</th>
<th>Outcome &amp; Decision:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Negative</td>
</tr>
<tr>
<td></td>
<td>(p&gt;0.025 = No Go)</td>
</tr>
<tr>
<td>If Truth H_{null}</td>
<td>0.975</td>
</tr>
<tr>
<td>If Truth H_{alt}</td>
<td>0.10</td>
</tr>
</tbody>
</table>
Typical Phase III Set-Up

- HR = 1 with probability 100%
- or
- HR = 0.75 with probability 100%
- The HR cannot be anything higher or lower or in-between

<table>
<thead>
<tr>
<th>Phase III Outcome &amp; Decision:</th>
<th>Negative (p &gt; 0.025 = No Go)</th>
<th>Positive (p ≤ 0.025 = Go)</th>
</tr>
</thead>
<tbody>
<tr>
<td>If Truth $H_{null}$</td>
<td>0.975</td>
<td>0.025</td>
</tr>
<tr>
<td>If Truth $H_{alt}$</td>
<td>0.10</td>
<td>0.90</td>
</tr>
</tbody>
</table>

But what is the chance that there is no effect of drug (HR=1)?
What is the chance there is an effect of drug (HR=0.75)?
Phase II data can guide us

Pr(HR ≤ 0.85) = 70%

Pr(HR > 0.85) = 30%

Observed Phase II data
HR = 0.75, 95% CI (0.47, 1.20)
# Probability of Success

## Phase III Outcome & Decision:

<table>
<thead>
<tr>
<th>Negative</th>
<th>Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p &gt; 0.025 = No Go)</td>
<td>(p ≤ 0.025 = Go)</td>
</tr>
</tbody>
</table>

| If Truth is HR=1.00 (H\textsubscript{null}) | 0.975 | 0.025 |
| If Truth is HR=0.75 (H\textsubscript{alt}) | 0.10  | 0.90  |

- Pr(HR=1.00) ≈ 0.30
- Pr(HR=0.75) ≈ 0.70
- Pr(Success) = 0.025 × 0.30 + 0.90 × 0.70 = 0.6375
Phase II tells us that the true HR could be a range of values, not only 1.00 and 0.75.

<table>
<thead>
<tr>
<th>HR&lt;sub&gt;TRUE&lt;/sub&gt; Range</th>
<th>Pr(HR&lt;sub&gt;TRUE&lt;/sub&gt; in range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;0.25</td>
<td>0.00%</td>
</tr>
<tr>
<td>0.25-0.35</td>
<td>0.22%</td>
</tr>
<tr>
<td>0.35-0.45</td>
<td>2.61%</td>
</tr>
<tr>
<td>0.45-0.55</td>
<td>9.82%</td>
</tr>
<tr>
<td>0.55-0.65</td>
<td>18.27%</td>
</tr>
<tr>
<td>0.65-0.75</td>
<td>21.49%</td>
</tr>
<tr>
<td>0.75-0.85</td>
<td>18.58%</td>
</tr>
<tr>
<td>0.85-0.95</td>
<td>12.99%</td>
</tr>
<tr>
<td>0.95-1.05</td>
<td>7.82%</td>
</tr>
<tr>
<td>1.05-1.15</td>
<td>4.24%</td>
</tr>
<tr>
<td>1.15-1.25</td>
<td>2.12%</td>
</tr>
<tr>
<td>&gt;1.25</td>
<td>1.84%</td>
</tr>
<tr>
<td></td>
<td>100%</td>
</tr>
</tbody>
</table>
# PIII Probability of Success given PII data

<table>
<thead>
<tr>
<th>Range for $HR_{TRUE}$</th>
<th>Column 1 $\Pr(HR_{TRUE} \text{ in range})$ based on PII data</th>
<th>Column 2 PIII Power if $HR_{TRUE} = \text{midpoint of range}$</th>
<th>PIII Probability of Success $1 \times 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.15-0.25</td>
<td>0.00%</td>
<td>100.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>0.25-0.35</td>
<td>0.22%</td>
<td>100.00%</td>
<td>0.22%</td>
</tr>
<tr>
<td>0.35-0.45</td>
<td>2.61%</td>
<td>100.00%</td>
<td>2.61%</td>
</tr>
<tr>
<td>0.45-0.55</td>
<td>9.82%</td>
<td>100.00%</td>
<td>9.82%</td>
</tr>
<tr>
<td>0.55-0.65</td>
<td>18.27%</td>
<td>99.99%</td>
<td>18.27%</td>
</tr>
<tr>
<td>0.65-0.75</td>
<td>21.49%</td>
<td>98.03%</td>
<td>21.07%</td>
</tr>
<tr>
<td>0.75-0.85</td>
<td>18.58%</td>
<td>71.05%</td>
<td>13.20%</td>
</tr>
<tr>
<td>0.85-0.95</td>
<td>12.99%</td>
<td>21.99%</td>
<td>2.86%</td>
</tr>
<tr>
<td>0.95-1.05</td>
<td>7.82%</td>
<td>2.50%</td>
<td>0.20%</td>
</tr>
<tr>
<td>1.05-1.15</td>
<td>4.24%</td>
<td>0.12%</td>
<td>0.01%</td>
</tr>
<tr>
<td>1.15-1.25</td>
<td>2.12%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

**Total PIII Probability of Success**: 68.3%
Quick Reminder: Power

- Regular Phase III Power = Pr(achieving $p<0.025$)
- Calculations are conditional upon the treatment effect *assumed* under the alternative, $\theta$
- If $x$ is a sufficient statistic for $\theta$ with distribution $f(x|\theta)$
  \[
  \text{PIII Power} = \int_{c}^{\infty} f(x|\theta)dx \quad \text{for some critical value, } c
  \]
- $x$ could be the difference in means, or proportions or the log hazard ratio where often $f(x|\theta) \sim N(\theta,\sigma^2)$
- In this instance $p<0.025$ will be achieved if $x>c$ where $c = z_{0.025} \cdot \sigma$
PoS is given by averaging Phase III Power over all possible values of $\theta$, the treatment effect

- Let $f(\theta)$ represent the probability distribution for the true treatment effect
  - E.g. $f(\theta)$ might be defined by the PII data
- Then Phase III PoS is given by averaging regular power over $f(\theta)$
- Phase III PoS = \[ \int_{x=c}^{\infty} \left[ \int_{-\infty}^{\infty} f(x|\theta)f(\theta)\,d\theta \right] dx = \text{Expected Power} \]
- This is often referred to as ‘assurance’\(^1\)
- Note if $f(\theta) \sim N(m,s^2)$ then
  \[ \int_{-\infty}^{\infty} f(x|\theta)f(\theta)\,d\theta = f(x) \sim N(m, s^2 + \sigma^2) \]
Some Limitations

• PoS = Pr(Phase III yields p < 0.025)
  ▪ Other definitions are possible e.g. Pr(Phase III yields p < 0.025 \( \cap \ x > \) desired difference)

• Phase II population and endpoint reflective of Phase III population and endpoint
  ▪ If Phase II endpoint differs might be possible to rely on established linkage to Phase III endpoint
  ▪ If no established linkage, then any move to Phase III carries high, difficult to quantify risk

• Prior for \( \theta \) defined by Phase II
  ▪ Could use a range of chosen priors, but, in reality, critical business decisions vs Phase III rely on the PII data
Typical Phase III Situation

- Phase II with N=120 estimates a 25% improvement in progression free survival for drug i.e. a hazard ratio (HR) of 0.75, 95% CI (0.47, 1.20) 1-sided p-value = 0.11
- Phase III N= 950 for 90% power, $\alpha$=2.5% 1-sided
- What is the PoS (‘Assurance’) of the planned Phase III?
  - 67.2%
- What if 90% PoS was desired?
  - N=2000 $\rightarrow$ 75.2%
  - N=10,000 $\rightarrow$ 83.5%
  - N=1,000,000 $\rightarrow$ 88.1%

*9months median on control, 1 yr accrual, 6 mo follow-up*
PoS ‘assurance’ cannot exceed the prior probability mass that treatment is effective†

Maximal Assurance = 89%

1-sided Phase II p-value = 0.11

Two Phase III s

- Phase II provides FEV1 treatment effect of 100mL 95% CI (10mL, 190mL), p=0.0297, N=50 per group, SD 230ml
- 2 x identical Phase III s planned with N=120 per group assuming same SD, 90% power $\alpha=2.5\%$ 1-sided
- Phase III PoS by ‘assurance’ = 77.7%
- What is the PoS for the two trials?
  - $\Pr$ (both trials are successful) = $77.7\%^2 = 60.4\%$ Not So

- Outcomes from the two independent, non-overlapping Phase III are correlated by assurance with $\rho = 0.71^+$
  - $\Pr$ (both trials are successful) = 68.5%

‘Discounting’ Phase II data

• Oftentimes seemingly impressive Phase II results are followed by disappointing Phase III outcomes
• To be expected since only positive Phase IIs are taken forward
• Let $x$ represent the treatment effect estimate from Phase II. We take forward to PIII only those with 1-sided p-value \( \leq p_x \)
  - This means $x \geq c$ where $c = \mu + \nu_x \Phi^{-1}(1 - p_x)$ where $E(x) = \mu$ and $\text{var}(x) = \nu_x^2$
  - With $\mu = 0$ and $\nu_x^2 = 1$ then $E(x|x \geq c) = 1.8$
  - With $\mu = 5$ and $\nu_x^2 = 16$ then $E(x|x \geq c) = 12.0$
• By focusing only on positive Phase II trials, the true treatment effect can be over-estimated.
• Discounting has been suggested by some$^\dagger$

‘Assurance’ indirectly suggests a degree of Phase II discounting

- Assurance equivalent to regular power hypothesizing not $\theta$ but rather $\kappa \theta$ where $\kappa = r - (\eta + 1)^{-0.5}(r - m/\theta)$
  - $\eta = s^2/\sigma^2 =$ ratio of PII:PIII estimated treatment effect variance and $r = z_\alpha/(z_\alpha + z_\beta)$
- To restore power in PIII back to $1 - \beta$ requires sample size to be increased by a factor of $\kappa^{-2}$

<table>
<thead>
<tr>
<th>Size of PIII relative to PII</th>
<th>$m/\theta$ Ratio of effect seen in PII (m) to effect hypothesized in PIII ($\theta$)</th>
<th>PoS ‘assurance’</th>
<th>$\kappa$</th>
<th>$\kappa^{-2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4x larger</td>
<td>1.25</td>
<td>82.5%</td>
<td>0.89</td>
<td>1.25</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>71.7%</td>
<td>0.78</td>
<td>1.64</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>61.1%</td>
<td>0.69</td>
<td>2.09</td>
</tr>
<tr>
<td>2x larger</td>
<td>1.25</td>
<td>88.6%</td>
<td>0.98</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>77.0%</td>
<td>0.83</td>
<td>1.44</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>64.3%</td>
<td>0.72</td>
<td>1.94</td>
</tr>
</tbody>
</table>
Decisions Grids to support Phase II to Phase III decision making

<table>
<thead>
<tr>
<th>PoS for FVC decline (1&lt;sup&gt;st&lt;/sup&gt;)</th>
<th>&lt; 30% (&lt;60mL)</th>
<th>30 - 75% (60-100mL)</th>
<th>&gt; 75% (&gt;100mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PoS for PFS (2&lt;sup&gt;nd&lt;/sup&gt;)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 30% (HR&gt;0.85)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 - 75% (0.70≤HR≤0.85)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;75% (HR&gt;0.75)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **GO to PhIII**
- **Review**
- **NO GO to PhIII**
TC-5214: Early Onset of Effect by Two Weeks and Increasing Improvement over Duration of Trial

![Graph showing change in HAM-D score over weeks for different treatment groups.]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Placebo + citalopram Adj. Mean (SE) (N = 132)</th>
<th>TC-5214 + citalopram Adj. Mean (SE) (N = 133)</th>
<th>Difference (95% Confidence Interval)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAMD-17</td>
<td>-7.75 (0.62)</td>
<td>-13.75 (0.62)</td>
<td>-6.0 (-7.72, -4.27)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
Too good to believe?

• Extremely positive Phase II results
  ▪ Results substantially better than that achieved for approved agents for major depressive disorder
  ▪ HAMD well established and accepted rating scale
  ▪ Taking Phase II at face value, PoS very high
  ▪ 50% discount would also result in high PoS

• In licensed late 2009

• In late 2011 & early 2012, 4 randomised, double blind, placebo-controlled Phase III trials all failed to meet their primary endpoint and development terminated

• Not isolated
  ▪ dimebon, inaparib, succinobucol, drisapersen, et al

Big Pharma Execs and the X-Files Syndrome?

I WANT TO BELIEVE
In drug development risk is unavoidable, but blind faith and unflinching belief despite the data is not the answer.
Some suggestions to aid better decision making

• Design:

1. Phase II well-controlled, randomised and double-blind and include an active control arm

2. Consider performing two Phase II trials. Positive outcomes from two Phase II’s of moderate size are more compelling and reliable than a single, larger Phase II
   - 2 Phase IIs with a 10% (or lower) \( \alpha \) level and N/2 patients together carry greater power than a single Phase II with N patients.
   - Prob of an extreme result one PII of size N is greater than observing the same extreme result in each of 2 PIIs of size N/2

3. Choose the most sensitive patient population relevant to the MOA

4. Predefine the Phase III Go/No Go decision rule for ‘success’ and stick to it.

5. Show expected power / regular power on discounted Phase II data
Some suggestions to aid better decision making

• Analysis and Interpretation

1. Analyse PII on an ITT basis to be reflective of what is expected in Phase III.

2. Interims in PII should be properly planned in advance and governed by a fully independent IDMC to protect the integrity of the study.

3. Stick to pre-defined Go/No Go decision rule: If the primary fails
   ▪ Do not be fooled by foraging for ‘signals’ elsewhere in data
   ▪ Be extremely sceptical of extensive (post-hoc) analyses in subgroups
   ▪ Do not retrospectively substitute the primary endpoint with some other secondary endpoint
   ▪ Do not seek to substitute primary with a subset of some secondary endpoint
   ▪ Do not look to retrospectively alter the predefined Go/No Go criteria.
   ▪ All of these things can, and do occur
Some suggestions to aid better decision making

• Analysis and Interpretation

4. Ensure senior leaders and decision makers are talented, experienced drug developers with a proven track record and a sound appreciation of good experimental design, data analysis and interpretation.

4. Enthusiastic scientists, project leaders and product champions are not the best decision makers

5. Include an experienced, technically expert practising statistician in the heart of the decision making process.
Summary

• Well designed Phase IIIs are critical in drug development

• Objective interpretation and careful use concepts like expected power and predefined Go/No Go criteria can help improve Probability of Success and enhance rational, effective and data-driven decision making

• But statistics and expert statistical input is not enough

• Experienced, credible, non-evangelical decision makers who can understand and interpret data objectively, for what they really are and not for what they desire, remain the real key to improving the longstanding high Phase III failure rate in Pharma