Discounting phase 2 results when planning phase 3 clinical trials

Simon Kirby

Full reference

Discounting phase 2 results when planning phase 3 clinical trials

S. Kirby, J. Burke, C. Chuang-Stein and C. Sin

Pharmaceutical Statistics 2012; 11: 373-385

Layout

1) Introduction and background
2) Launch criteria
3) Sample sizing phase 3
4) Assessing bias
5) A simulation study
6) Some empirical evidence
7) Overall conclusions

1) Introduction and background

- Observation
  - Not uncommon for a positive phase 2 trial to be followed by a negative phase 3 trial

- Many possible reasons but focus in this paper on two
  - Only phase 2 trials with good results lead to phase 3 planning
  - Patient populations for phase 3 trials are often more diverse than for phase 2 trials. A more heterogeneous population tends to report a smaller treatment effect.

2) Launch criteria

- To study the idea that only good phase 2 results lead to phase 3 planning we can consider launch criteria

- Launch criteria are used in conjunction with a threshold δ_{0L} to decide whether to launch a phase 3 study

[Previously considered by Wang Hung and O'Neill (WH&ON), 2006]

2) Launch criteria

- WH&ON considered launching a phase 3 study if

$$(1)\ \tilde{\delta} > \delta_{0L}\ \text{or}$$

$$(2)\ L_{kse} > \delta_{0L}, k = 1,2$$

2) Launch criteria

- We additionally consider

$$(3)\ M_f > \delta_{0L}$$

$$(4)\ L_{0.10} > MAV \text{ and } U_{0.25} > TV$$

$$(5)\ L_{f,0.10} > MAV \text{ and } U_{f,0.25} > TV$$

$$M_f = \tilde{\delta} * f$$

$$MAV = \text{Minimum Acceptable Value}$$

$$TV = \text{Target Value}$$

3) Sample sizing phase 3

- Consider the simple case of a parallel group phase 3 study with two treatment groups and a continuous response that is normally distributed ($\sigma = 1$)

$$n = \frac{2(Z_{0.025} + Z_{0.20})^2}{\delta^2}$$

3) Sample sizing phase 3

- If a phase 3 trial is launched the following treatment estimates are used to sample size the phase 3 trial

$$(1) \quad \tilde{\delta}$$

$$(2) \quad \tilde{\delta} - kse \ , \ k = 1,2$$

$$(3) \quad \tilde{\delta} * f$$

$$(4) \quad \tilde{\delta}$$

$$(5) \quad \tilde{\delta} * f$$

4) Assessing bias

- WH&ON assessed bias by comparing empirical power and planned power

- We compared estimated Assurance after the phase 2 trial with theoretical Assurance

- Assurance $\int p(R \mid \delta) p(\delta) d\delta$

where R represents achieving statistical significance

4) Assessing bias

- Estimated Assurance calculated as

$$\Phi\left(\frac{-1.96 * \tau + \delta}{\sqrt{\tau^2 + \upsilon}}\right)$$

$$\delta = \tilde{\delta}, L_{kse}, M_f, \tau^2 = \frac{2\hat{\sigma}^2}{m}, \upsilon = \frac{2\hat{\sigma}^2}{n}$$

- For Theoretical Assurance true standard deviation and true phase 3 treatment effect used

5) A simulation study

To study patient populations possibly being more diverse for phase 3 trials than for phase 2 trials

δ_{02} - true phase 2 treatment effect

allowed to differ from

δ_0 - true phase 3 treatment effect

5) A simulation study

- Four scenarios studied (σ = 1)

Scenario 1 $\qquad \delta_{02} = 0.3, \delta_0 = 0.3, m = 50,100,200, \delta_{0L} = 0.2,0.15,0.1$

Scenario 2 $\qquad \delta_{02} = 0.3, \delta_0 = 0.2, m = 50,100,200, \delta_{0L} = 0.2$

Scenario 3 $\qquad \delta_{02} = 0.3, \delta_0 = 0.2, m = 50,100,200, \delta_{0L} = 0.1$

Scenario 4 $\qquad \delta_{02} = 0.4, \delta_0 = 0.2, m = 50,100,200, \delta_{0L} = 0.2,0.15,0.1$

5) A simulation study

- For M_f larger multipliers (e.g., >=0.5) appropriate when phase 2 sample size is large and threshold low

- Dual launch criteria approach with multiplicative adjustment leads to larger discounting to equate estimated and theoretical assurance

5) A simulation study

- Scenario 1
  - Provided a larger phase 2 sample size (n=100,200) is used and the launch threshold is <=50% of the target value a multiplicative factor of 0.9 performs well in equating estimated and theoretical assurance

- Other scenarios
  - Provided the threshold is set at no greater than 50% of the target value then a multiplier of 0.9 times the ratio of true phase 3 to phase 2 effect appears to perform reasonably for the larger phase 2 sample sizes

5) A simulation study

The probability of launching a phase 3 study was also estimated

Found that this probability was lower for the WH&ON criterion $L_{1se} > \delta_{0L}$

than for $M_f > \delta_{0L}$

when estimated and theoretical assurance equated

5) A simulation study

However more trials also launched for the multiplicative adjustment under the null case of no true treatment effect in phase 2 or phase 3 studies

This probability becomes smaller as the phase 2 sample size becomes larger and the launch threshold increases

6) Some empirical evidence

- Looked at Pfizer development projects with completed phase 2 and phase 3 studies between 1998 and 2009 (excluded projects without comparative phase 2 trials and those with a different primary endpoint in phase 2 and phase 3)

- 11 drug projects spanning osteoporosis, obesity, hypertension, pain, over-active bladder, afibrillation, smoking cessation, migraine and arthritis

6) Some empirical evidence

- Aim was to compare average estimated assurance with phase 3 success probability

- Rather than try to follow project history we matched all phase 2 studies with all phase 3 studies

- Excluded all matches that gave an estimated assurance of less than 0.10 as these would not have led to a phase 3 programme

6) Some empirical evidence

- Obtained 104 matches from 11 drug projects

- 32 distinct phase 2 and 54 distinct phase 3 studies

6) Some empirical evidence

- For continuous endpoints estimated assurance calculated using a normal prior with mean equal to the phase 2 estimated treatment effect and standard deviation equal to corresponding standard error

- For binary endpoints beta distributions used with means equal to estimated phase 2 proportions and standard deviations equal to estimated standard errors

6) Some empirical evidence

Averaging estimated assurance from 104 matches gave
an estimated assurance of

0.801

The proportion of statistically significant results was

0.78

Averaging planned power gave an average power of

0.86

6) Some empirical evidence

- Using a multiplicative factor of 0.9 essentially equalised estimated assurance and the proportion of significant results

6) Some empirical evidence

- However, a number of caveats

  - Proportion of statistically significant results amongst distinct phase 3 trials, 0.81, is very high so an atypical set of trials compared with historically reported rate
    - Multiple phase 2 trials
    - Sample sizing using MCID
    - Whether pre-specified decision rules used unclear

7) Overall conclusions

- Some discounting of phase 2 results recommended when used to plan phase 3 trials

- Multiplicative factor 0.9 looks reasonable for scenario 1 for larger phase 2 sample sizes (can multiply by ratio of phase 3 effect to phase 2 effect for other situations) when used in conjunction with threshold no bigger than half the target value for the treatment effect

25

7) Overall conclusions

- Some limited empirical evidence that discounting of phase 2 results needed when used to plan phase 3 trials

Reference

Wang, S., Hung, H.M.J. and O'Neill, R.T. Adapting the sample size planning of a phase III trial based on phase II data. Pharmaceutical Statistics 2006; 5: 85-97.