An efficient development program under limited prior information about the association between response and overall survival

Heiko Götte[1], Junyuan Xiong[2], Marietta Kirchner[3], Meinhard Kieser[3]

[1]Merck KGaA, Darmstadt, Germany; [2]EMD Serono, Billerica, MA, USA; [3]Institute of Medical Biometry and Informatics, University of Heidelberg, Germany

INTRODUCTION

- In the highly competitive oncology area, efficiency in drug development is critical. In phase II, **there is a trade-off between reduced investment and high confidence in the potential of the drug before going into phase III.**
  - Adequate decision making after phase II based on long-term endpoints like **overall survival (OS) requires large numbers of subjects and long follow-up times.**
- **Faster decision making is possible based on short-term surrogate endpoints** such as PFS or response (Figure 1A).
- However, this requires a proof of adequate surrogacy which is very difficult. The level of surrogacy may be treatment- and/or indication-dependent. Thus, **when developing a new drug, there is limited reliable information about the relationship between response and OS.**
- The limited information can be transformed into weak/ non-informative priors for Bayesian measures like the predictive probability of success (PoS).
- However, available approaches for PoS calculation based on bivariate normal distribution of the treatment effect measures [1] are very sensitive to assumptions about means, variances and correlation, even with weak/ non-informative priors.
- Therefore, we calculate PoS based on individual patient response and available OS data using a similar model as in [2].
- As **only 40% of oncology drugs entering phase III will get FDA approval** [3], a de-risking strategy is important.
- **To learn about the relationship between response and OS and to de-risk the program**, we follow the concept of a **"phase 2+" design** [4] where after a go-to-phase-III-decision, further follow-up data from phase II is collected and used to make interim decisions on phase III (Figure 1B). **The phase III will be stopped early if the updated phase II and partial phase III data lead to a low PoS.**

OBJECTIVE

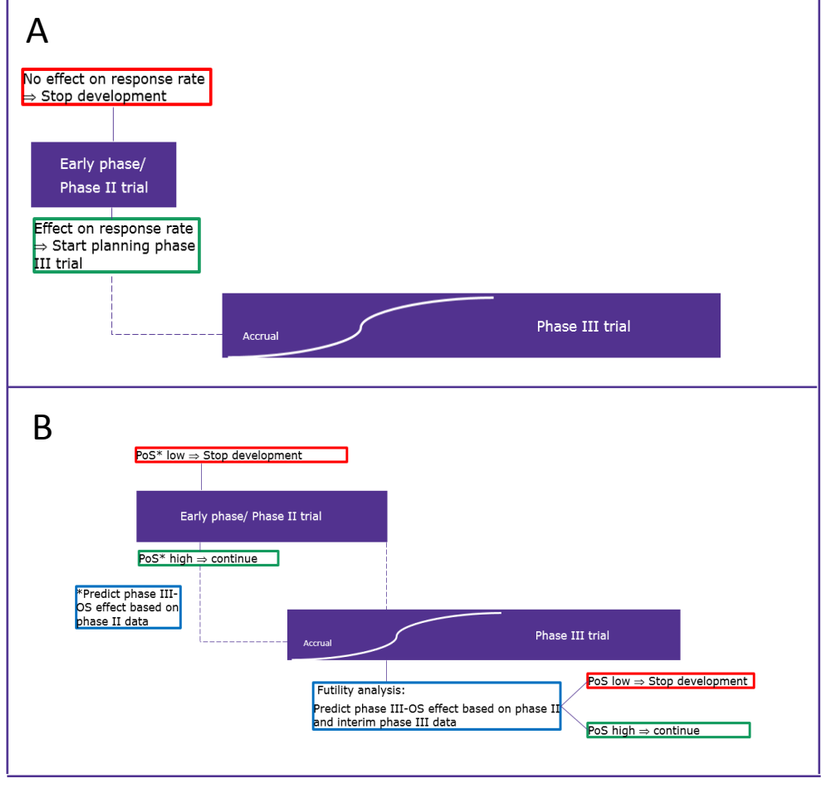
- **Plan an efficient development program with the option to collect phase II further follow-up data to reduce the risk of phase III failure by defining optimal decision boundaries and analysis time points**

METHODS

Basic setting

- Phase II and phase III: one randomized trial each (experimental E and control C)
- Phase III: OS is the primary endpoint and primary analysis is based on Cox model with treatment as the independent variable
- Short-term endpoint with potential relationship to OS:
  - Observed RECIST response status CR/PR, SD & PD
  - Assumption: response status is determined at baseline but only observable after minimum follow-up of 3 months (two 6-week scans)
- There are up to two looks/ decision time points (Figure 1 B)
- Subjects with censored OS time (due to data cut-off) get randomly assigned future OS time based on observed response status and treatment group
  - Uncertainty in individual risk of death is taken into account
- These data are used to predict OS treatment effect (hazard ratio (HR))
- Predicted HR is the basis for PoS estimate which is used for decision making
- Simulation study will determine optimal decision boundaries and time points

Figure 1. Traditonal (A) and proposed (B) development program



Prediction of overall survival

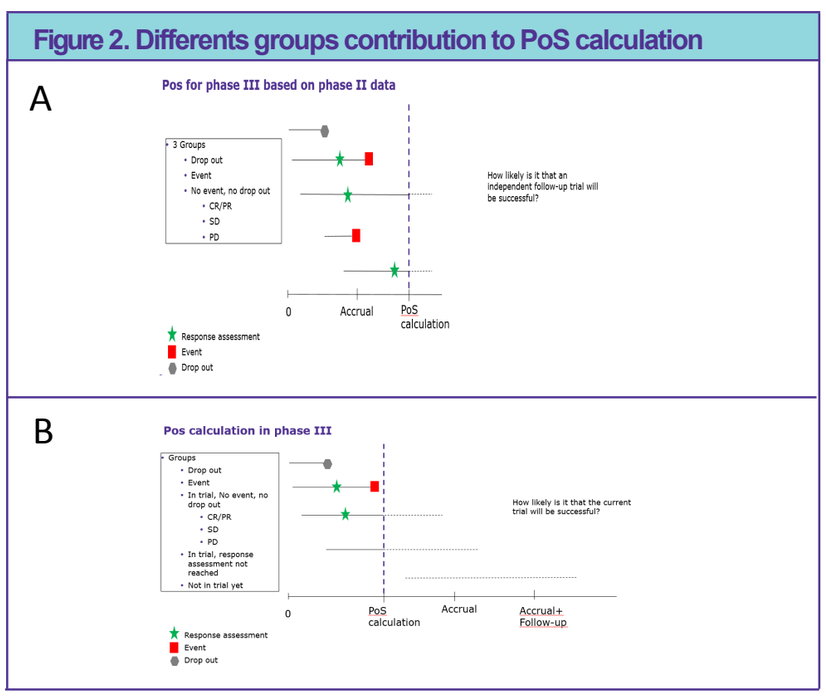
- m subjects with observed OS event or drop-out: survival time t_i
- Predicted survival time based on response status for subject i, $i \in \{m+1,\ldots,n\}$: $t_i^I = t_i + s_i^*$ and $I_i^I = 1$ *if* $s_i^* = t_i^*$ where $s_i^* = \min(t_i^*, d_i^*)$ *with* t_i^* as a random sample from an exponential distribution with rate λ_i which is random sample from
  - posterior distribution λ_{kj}|data ~$\text{Gamma}(\tau_{jk} + e_{jk}, \omega_{jk} + T_{jk})$ with
    - response status k and treatment group j
    - weak optimistic prior: $\text{Gamma}(\tau_{jk}, \omega_{jk})$ with $(\tau_{C1}, \tau_{C2}, \tau_{C3}, \tau_{E1}, \tau_{E2}, \tau_{E3}) = (-\frac{\log(0.5)}{16} \times 5, -\frac{\log(0.5)}{11} \times 5, -\frac{\log(0.5)}{8} \times 5, -\frac{\log(0.5)}{20} \times 5, -\frac{\log(0.5)}{14} \times 5, -\frac{\log(0.5)}{11} \times 5)$ and $(\omega_{C1}, \omega_{C2}, \omega_{C3}, \omega_{E1}, \omega_{E2}, \omega_{E3}) = (5, 5, 5, 5, 5, 5)$
    - e_{jk} , T_{jk} : sum of observed number of events and survival times based on landmark analysis at t_0=3 months
  - d_i^*(drop-out): a random sample from exponential distribution with $d \sim Gamma(-\frac{\log(0.9)}{12} * 100, 100)$

Probability of Success using OS and response data

- Calculation depends on whether prediction is done for a future or the current phase III trial
- Future trial (Figure 2 A)
  - $\text{PoS}_{sim2} = \frac{1}{Q}\sum_{q=1}^{Q} \text{PoS}_q$ *with* $\text{PoS}_q = 1 - \Phi\left(\frac{b - \hat{\theta}_{LIIq}}{\sqrt{se(\tilde{\theta})^2 + se_q(\hat{\theta}_{LII})^2}}\right)$ with $b = \Phi^{-1}(1-\alpha) \cdot se(\tilde{\theta})$ and $\hat{\theta}_{LII_q} = \hat{\theta}_{LII}\left((t_1, I_1, x_1), \ldots, (t_m, I_m, x_m), (t^*_{q,m+1}, I^*_{q,m+1}, x_{m+1}), \ldots, (t^*_{q,n}, I^*_{q,n}, x_n)\right)$ being –log(HR) estimate and $se_q(\hat{\theta}_{LII})$ corresponding standard error from phase II Cox model in the simulation step q, and expected standard error for phase III: $se(\tilde{\theta}) = \sqrt{\frac{4}{d_3}}$
- Same trial (Figure 2 B)

  $\text{PoS}_{sim23} = \frac{1}{Q}\sum_{q=1}^{Q} I\left(\frac{\hat{\theta}_{LIIIq}}{se_q(\hat{\theta}_{LIII})} > \Phi^{-1}(1-\alpha)\right)$ with $\hat{\theta}_{LIII_q} = \hat{\theta}_{LIII}\left((t_1, I_1, x_1), \ldots, (t_m, I_m, x_m), (t^*_{q,m+1}, I^*_{q,m+1}, x_{m+1}), \ldots, (t^*_{q,n}, I^*_{q,n}, x_{n_3})\right)$ being –log(HR) estimate and $se_q(\hat{\theta}_{LIII})$ corresponding standard error from Cox model based on phase II and III data in simulation step q

Figure 2. Differents groups contribution to PoS calculation



Decision making

There are possible two looks/ decision points (Figure 1 B):

$\text{PoS}_{sim2} > b_{go} \Rightarrow$ go to phase III

$\text{PoS}_{sim2} \leq b_{go} \Rightarrow$ stop after phase II

and

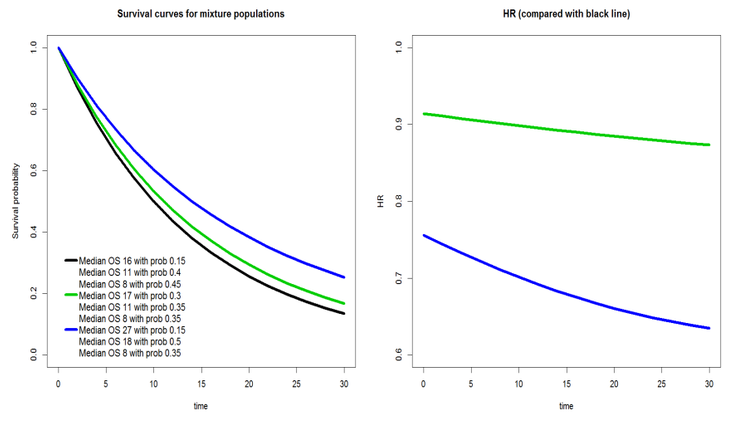
$\text{PoS}_{sim23} > b_{fut} \Rightarrow$ continue with phase III

$\text{PoS}_{sim23} \leq b_{fut} \Rightarrow$ stop phase III for futility

Simulation study for planning purpose

- True model for data generation: categorical baseline variable defining three different risk groups per treatment group. Subjects within a risk group have constant hazards (Figure 3, hazards are transformed to Median OS times)
- **Optimization: Find the combination of decision boundaries b_{go} and b_{fut} and time points of 1st and 2nd look which lead to**
  - **P(Finish phase III | positive scenario)>60%**
  - **P(Stop for futility (2nd look) | negative scenario)>40%**
  - **and minimum mean duration (negative scenario) of the program**
    - Positve/Negative scenario as in Figure 3
    - Scenarios where decisons exclusively based on response would be misleading: OS effect exists but no response effect and vice versa
- Assumed trial parameter
  - Phase II:
    - $n_2 = 80$ recruited in 4 months (20 subjects/month),
    - analysis after follow-up of {3, 6, 9} months
  - "White space"
    - time between phase II analysis and the start of phase III =5 months
  - Phase III:
    - $n_3 = 390$ recruited in 13 months (30 subjects/month),
    - futility analysis {-1, 5, 7} months after start of trial
      - $\Rightarrow$ number of phase III subjects: $n_{3fut} \in \{0,\ 150, 210\}$
      - $\Rightarrow$ time from first phase II analysis is {4, 10, 12} months
    - final analysis: after 256 events or 36 months after start of trial, whatever comes first
      - 256 events based on 90% power, $HR = 2/3$, $\alpha = 0.025$
- Four PoS:
  - Two for decision making: Phase II all data (1st look), Phase II all data (2nd look) plus available phase III all data
  - Two for comparison purpose: based on $1 - \Phi\left(\frac{b - \hat{\theta}_{LII}}{\sqrt{se(\tilde{\theta})^2 + se(\hat{\theta}_{LII})^2}}\right)$ using phase II OS data only -- no cut-off, cut-off at 1st look

Figure 3. Theoretical Survival curve and hazard ratios (HR)



Control: Black line, Positve scenario for experimental: Blue line, Negative scenario for experimental: Green line

RESULTS

Figure 4. PoS distribution per analysis time point: positve scenario (A), negative scenario (B)

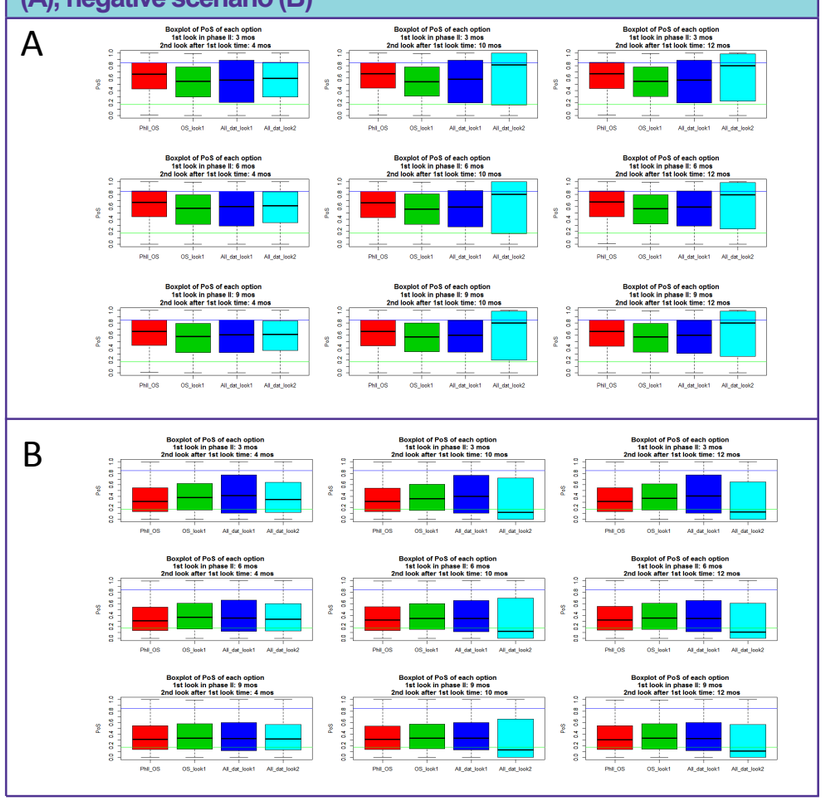


Table 1: Optimal decision boundaries evaluated at different time points

Decision rule: stop if	Phase II analysis*	Phase III futility analysis **	Stop after phase II	Stop after phase III futility	Finish phase III (Conditional finish)	Mean duration (months)
$PoS_{sim2} \leq 0.15$ $PoS_{sim23} \leq 0.45$	3	-1	0.203 0.293	0.237 0.345	0.560 (0.703) 0.362 (0.512)	26.99 19.60
	3	5	0.206 0.299	0.243 0.407	0.551 (0.694) 0.294 (0.419)	30.94 21.93
	3	7	0.210 0.296	0.230 0.438	0.560 (0.709) 0.266 (0.378)	31.63 22.09
	6	-1	0.140 0.284	0.222 0.344	0.638 (0.742) 0.372 (0.520)	32.58 22.91
	6	5	0.148 0.298	0.254 0.394	0.598 (0.702) 0.308 (0.439)	35.73 25.24
	6	**7**	0.141 0.300	0.244 0.428	0.615 (0.716) 0.272 (0.389)	36.86 25.14
	9	-1	0.111 0.297	0.227 0.350	0.662 (0.745) 0.353 (0.502)	36.42 25.34
	9	5	0.114 0.288	0.259 0.400	0.627 (0.708) 0.312 (0.438)	39.88 28.33
	9	7	0.119 0.301	0.251 0.429	0.630 (0.715) 0.270 (0.386)	40.48 28.06

* Months after last subject randomized, ** Months after first subject randomized

- In phase II: impact on PoS distribution by adding response data rather low (Figure 4)
- Using phase III data improves PoS distribution
  - Due to using more data
  - Different formula (assurance formula might lead to a biased PoS estimate [5])
- Requiring high P(finish phase III) in positive scenario and high P(stop at 2nd look) for negative scenario lead to low boundary for first look and higher boundary at 2nd look (Table 1)
- With the same decision boundaries: Mean duration has a range of 13 months (40-27) in the positive scenario; probability of finishing phase III in the negative scenario has a range of 10% (0.37-0.27) (Table 1)

DISCUSSION

- **Optimal choice of boundaries and time points leads to promising compromise between low investment (measured by mean duration) and high confidence (measured by stopping probabilities and correct go probabilities)**
- **Optimal rules strongly depend on side conditions**: ie., probability of finishing phase III etc
- How to handle subjects without response status and no OS event in phase III?
  - Option 1 (used here):
    - Randomly assigned response status based on observed response status rates in each treatment group from phase II data
      - t_0 (3 months) is not baseline (here: differences of up to 3%)
    - Based on response status & treatment, random survival time is assigned
      - Assumes constant hazards
      - Early events before t_0 not considered
  - Option 2:
    - Based on treatment group, random survival time is assigned
      - All events including early events are considered
      - If response distribution differs between phase II and phase III, potential biased results arise
  - Select option1 or 2 depending on the number of events before t0 and observed response distributions between phase II and phase III-futility analysis
  - Alternative model assumption: multi-state model

REFERENCES

1. Hong S, Shi L. (2012).. Statistics in Medicine 31:831–843.
2. Stallard N, Whitehead J, Cleall S (2005). Pharmaceutical Statistics, 4: 119–128
3. BIO Industry Analysis Reports (2016). Clinical Development Success Rates 2006-2015
4. Beckman RA, Clark J, Chen C (2011). Nature Reviews Drug Discovery 10:735-748
5. Götte H, Kirchner M, Sailer M. (2017). Pharmaceutical Statistics, 16:178-191.

Presented at European Statistical Meeting on Decision Making in Drug Development 2018