Improving probabilities of correct decision in population enrichment designs

Heiko Götte[1], Margarita Donica[2]*, and Giacomo Mordenti[3]*

[1] Merck KGaA, Darmstadt, Germany

[2] F. Hoffmann – La Roche LTD (Global Medical Affairs Biometrics), Basel, Switzerland

[3] Grünenthal GmbH, Aachen, Germany

* were under employment of Merck Serono S.A. – Geneva, Switzerland, when contributed to the publication work.

Adaptive Designs and Multiple Comparison Procedures
workshop in Köln on June 24-26, 2015

End of Phase II

Biomarker suggests treatment is more effective in a subpopulation

- Biological plausibility
  - Biomarker is related to the mode of action of the experimental treatment
  - External data supporting the assumption about the potential predictive effect
- Subpopulation unambiguously defined
- Biomarker test kit is available and result is reliable

Motivating example – Phase II result

Primary end point of randomized phase II trial: PFS

- HR = 0.71 based on 110 events
  - HR≤ 0.75 is considered as relevant effect

Motivating example – Phase II result

Primary end point of randomized phase II trial: PFS

- HR = 0.71 based on 110 events
  - HR≤ 0.75 is considered as relevant effect
- Biomarker divide population into Subpopulation and Complement
  - HR_S = 0.60 based on 50 events
  - HR_C = 0.89 based on 50 events

Motivating example – Phase II result

Primary end point of randomized phase II trial: PFS

- HR = 0.71 based on 110 events
  - HR≤ 0.75 is considered as relevant effect
- Biomarker divide population into Subpopulation and Complement
  - HR_S = 0.60 based on 50 events
  - HR_C = 0.89 based on 50 events
- Plan phase III trial with one interim analysis for potential subpopulation selection

Phase III Setting

- θ is overall treatment effect, i.e. -log(HR)
  - $\theta > 0 \Leftrightarrow HR < 1$
- Hypothesis tested in Sub and Full population
  - H_S: $\theta_S \leq 0$ against $\theta_S > 0$
  - H_F: $\theta \leq 0$ against $\theta > 0$

Phase III Setting

- θ is overall treatment effect, i.e. -log(HR)
  - $\theta > 0 \Leftrightarrow$ HR<1
- Hypothesis tested in Sub and Full population
  - H_S: $\theta_S \leq 0$ against $\theta_S > 0$
  - H_F: $\theta \leq 0$ against $\theta > 0$
- Relationship between θ and θ_S
  - $\theta = \gamma\, \theta_S + (1-\gamma)\, \theta_C$
  - γ is subpopulation fraction

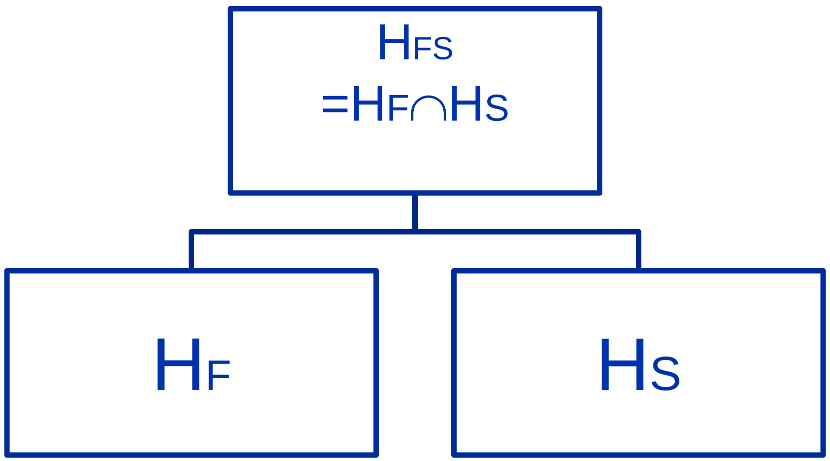
Merck Serono

MERCK Living Innovation

Phase III Setting

- θ is overall treatment effect, i.e. -log(HR)
  - $\theta > 0 \Leftrightarrow HR < 1$
- Hypothesis tested in Sub and Full population
  - H_S: $\theta_S \leq 0$ against $\theta_S > 0$
  - H_F: $\theta \leq 0$ against $\theta > 0$
- Relationship between θ and θ_S
  - $\theta = \gamma\, \theta_S + (1-\gamma)\, \theta_C$
  - γ is subpopulation fraction
- 508 events correspond to 90% Power with one-sided α=0.025 and planned HR=0.75
- One interim analysis is performed after τ% of subjects/events are collected
  - τ is information fraction

Closed testing procedure



Trial Design

Stage 1

Stage 2

- Options after Stage 1
  - Continue with the full population
  - Continue with the sub population
  - Stop for futility

  - Stop for efficacy: no option

Combine data from stage 1 and 2

- Inverse normal method

$$C(p_{1,J}, p_{2,J}) = w_1\Phi^{-1}(1-p_{1,J}) + w_2\Phi^{-1}(1-p_{2,J})$$

  - with $J \subseteq \{F,S\}$
  - Weights: $w_1=\sqrt{\tau}$ $w_2=\sqrt{1-\tau}$ $(w_1^2+w_2^2=1)$

Combine data from stage 1 and 2

- Inverse normal method

$$C(p_{1,J}, p_{2,J}) = w_1\Phi^{-1}(1-p_{1,J}) + w_2\Phi^{-1}(1-p_{2,J})$$

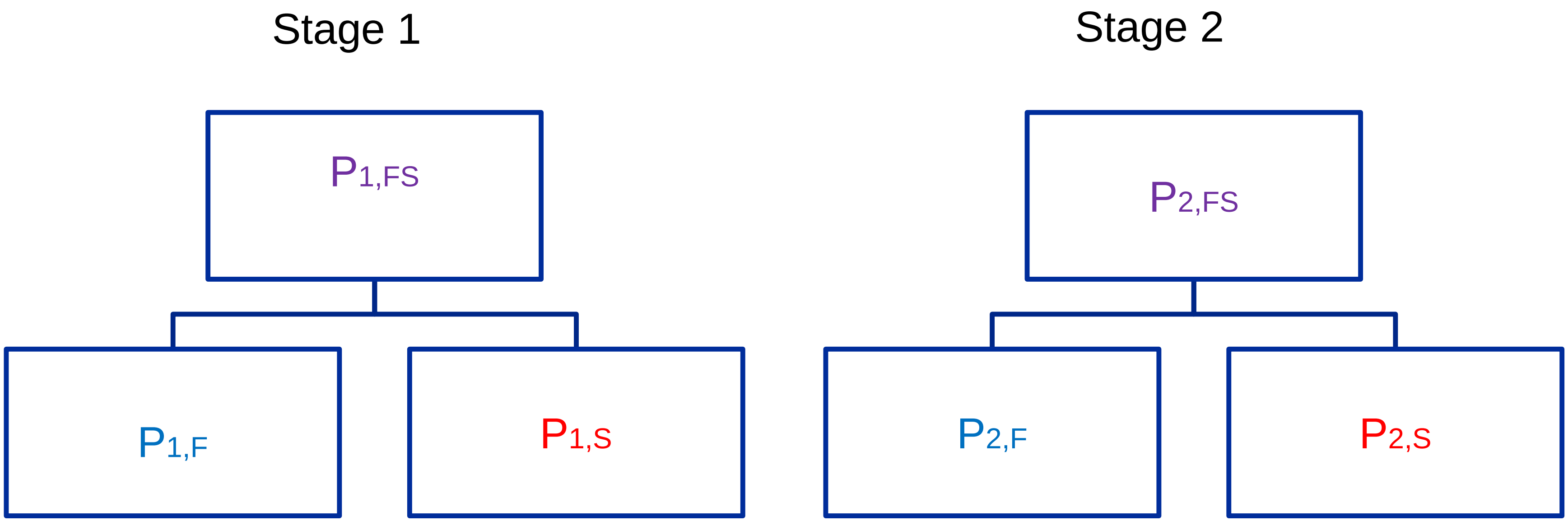
  - with $J \subseteq \{F,S\}$
  - Weights: $w_1=\sqrt{\tau}$ $w_2=\sqrt{1-\tau}$ $(w_1^2+w_2^2=1)$

- Intersection hypothesis: Hochberg procedure
- Second stage p-values based on increments in survival setting

Merck Serono

MERCK Living Innovation

Continue with the full population



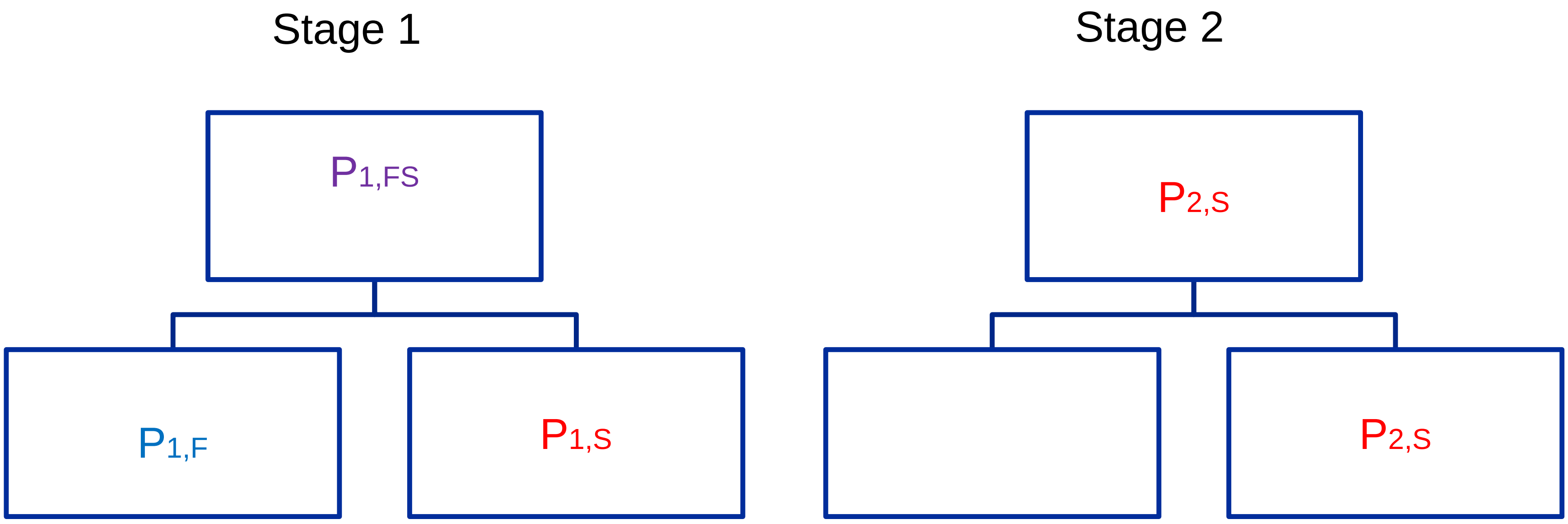
Reject H_F in stage 2, if

$$\min(C(p_{1,FS}\,,\, p_{2,FS}),\, C(p_{1,F},\, p_{2,F})) > \Phi^{-1}(1-\alpha)$$

Reject H_S in stage 2, if

$$\min(C(p_{1,FS}\,,\, p_{2,FS}),\, C(p_{1,S},\, p_{2,S})) > \Phi^{-1}(1-\alpha)$$

Continue with the sub population

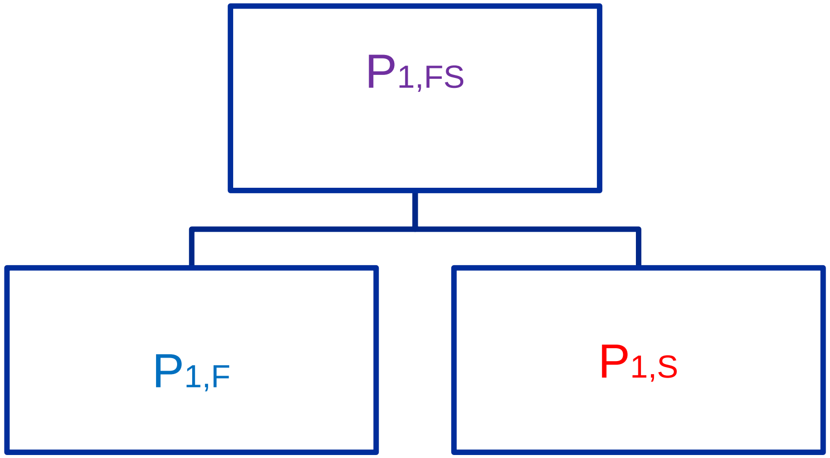


Reject H_S in stage 2, if

$$\min(C(p_{1,FS}, p_{2,S}), C(p_{1,S}, p_{2,S})) > \Phi^{-1}(1-\alpha)$$

Stop for futility

Stage 1



Main focus of this talk

- How to make an interim decision?

Main focus of this talk

- How to make an interim decision?
- How to make the correct interim decision?

Main focus of this talk

- How to make an interim decision?
- How to make the correct interim decision?
  - Let‘s say HR≤0.75 is considered as clinically relevant
  - Let‘s say we know the truth

Main focus of this talk

- How to make an interim decision?
- How to make the correct interim decision?
  - Let's say HR≤0.75 is considered as clinically relevant
  - Let's say we know the truth
  - What would be the decision if (subpopulation fraction γ=0.5)
    - HR_F =0.75, HR_S =0.75, HR_C =0.75 ?

Main focus of this talk

- How to make an interim decision?
- How to make the correct interim decision?
  - Let's say $HR \leq 0.75$ is considered as clinically relevant
  - Let's say we know the truth
  - What would be the decision if (subpopulation fraction $\gamma=0.5$)
    - $HR_F = 0.75$, $HR_S = 0.75$, $HR_C = 0.75$?
    - $HR_F = 0.75$, $HR_S = 0.74$, $HR_C = 0.76$?

Main focus of this talk

- How to make an interim decision?
- How to make the correct interim decision?
  - Let's say $HR \leq 0.75$ is considered as clinically relevant
  - Let's say we know the truth
  - What would be the decision if (subpopulation fraction $\gamma=0.5$)
    - $HR_F = 0.75$, $HR_S = 0.75$, $HR_C = 0.75$?
    - $HR_F = 0.75$, $HR_S = 0.74$, $HR_C = 0.76$?
    - $HR_F = 0.75$, $HR_S = 0.70$, $HR_C = 0.81$?
    - $HR_F = 0.77$, $HR_S = 0.70$, $HR_C = 0.85$?

Main focus of this talk

- How to make an interim decision?
- How to make the correct interim decision?
  - Let‘s say HR≤0.75 is considered as clinically relevant
  - Let‘s say we know the truth
  - What would be the decision if (subpopulation fraction γ=0.5)
    - HR_F =0.75, HR_S =0.75, HR_C =0.75 ?
    - HR_F =0.75, HR_S =0.74, HR_C =0.76 ?
    - HR_F =0.75, HR_S =0.70, HR_C =0.81 ?
    - HR_F =0.77, HR_S =0.70, HR_C =0.85 ?
  - Let‘s focus on the unambiguous scenarios
    - HR_F =0.75, HR_S =0.75, HR_C =0.75 $\Rightarrow$ go with the full population
    - HR_F =0.87, HR_S =0.75, HR_C = 1 $\Rightarrow$ go with the sub population
    - HR_F = ?, HR_S = 1, HR_C = ? $\Rightarrow$ stop for futility

Main focus of this talk

- How to make an interim decision?
- How to make the correct interim decision?
  - Let's focus on the unambiguous scenarios
    - $HR_F = 0.75$, $HR_S = 0.75$, $HR_C = 0.75$ $\Rightarrow$ go with the full population
    - $HR_F = 0.87$, $HR_S = 0.75$, $HR_C = 1$ $\Rightarrow$ go with the sub population
    - $HR_F = ?$, $HR_S = 1$, $HR_C = ?$ $\Rightarrow$ stop for futility
  - Maximally one of these scenarios can be true
    - Make assumption how likely the different scenarios are

Main focus of this talk

- How to make an interim decision?
- How to make the correct interim decision?
  - Let's focus on the unambiguous scenarios
    - $HR_F = 0.75$, $HR_S = 0.75$, $HR_C = 0.75 \Rightarrow$ go with the full population
    - $HR_F = 0.87$, $HR_S = 0.75$, $HR_C = 1 \Rightarrow$ go with the sub population
    - $HR_F = ?$, $HR_S = 1$, $HR_C = ? \Rightarrow$ stop for futility
  - Maximally one of these scenarios can be true
    - Make assumption how likely the different scenarios are
  - Q = P(correct decision in interim analysis)

$$= \sum P(\text{correct decision} \mid \text{true values in sub} \cap \text{in complement}) * P(\text{true values in sub} \cap \text{in complement})$$

Main focus of this talk

- How to make an interim decision?
- How to make the correct interim decision?
  - Let's focus on the unambiguous scenarios
    - $HR_F = 0.75$, $HR_S = 0.75$, $HR_C = 0.75 \Rightarrow$ go with the full population
    - $HR_F = 0.87$, $HR_S = 0.75$, $HR_C = 1 \Rightarrow$ go with the sub population
    - $HR_F = ?$, $HR_S = 1$, $HR_C = ? \Rightarrow$ stop for futility
  - Maximally one of these scenarios can be true
    - Make assumption how likely the different scenarios are
  - Q = P(correct decision in interim analysis)

$$= \sum P(\text{correct decision} \mid \text{true values in sub} \cap \text{in complement}) * P(\text{true values in sub} \cap \text{in complement})$$

$$= \omega_1 P(\text{continue full} \mid \text{effect in sub} \cap \text{effect in complement})$$

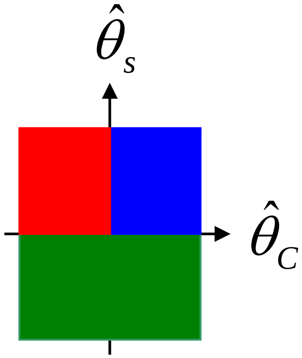
$$+ \omega_2 P(\text{continue sub} \mid \text{effect in sub} \cap \text{no effect in complement})$$

$$+ \omega_3 P(\text{stop for futility} \mid \text{no effect in sub}), \qquad (\omega_1+\omega_2+\omega_3=1)$$

How to make the interim decision?

Sign of the observed treatment effect (“Simple rule”)

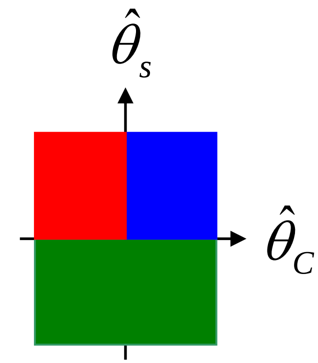
- $\hat{\theta}_s < 0$: Stop for futility
- $\hat{\theta}_s >= 0$ & $\hat{\theta}_c < 0$: Continue sub
- $\hat{\theta}_s >= 0$ & $\hat{\theta}_c >= 0$: Continue full



How to make the interim decision?

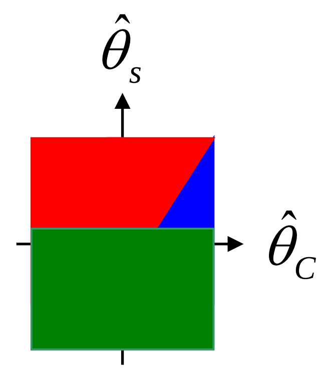
Sign of the observed treatment effect ("Simple rule")

- $\hat{\theta}_s < 0$: Stop for futility
- $\hat{\theta}_s >= 0$ & $\hat{\theta}_c < 0$: Continue sub
- $\hat{\theta}_s >= 0$ & $\hat{\theta}_c >= 0$: Continue full



General "linear rule"

- $\hat{\theta}_s < f_L$: Stop for futility
- $\hat{\theta}_s >= f_L$ & $a_L * \hat{\theta}_s + \hat{\theta}_c < d_L$: Continue sub
- $\hat{\theta}_s >= f_L$ & $a_L * \hat{\theta}_s + \hat{\theta}_c >= d_L$: Continue full



Find optimal decision rule

- $Q_L =$ P(correct decision in interim analysis)

  $= \omega_1 P(X > f_L, Y > d_L \mid E(X) = -\log(0.75), E(Y) = a_L * (-\log(0.75)) + (-\log(0.75)))$

  $+ \omega_2 P(X > f_L, Y < d_L \mid E(X) = -\log(0.75), E(Y) = a_L * (-\log(0.75)) + (-\log(1)))$

  $+ \omega_3 P(X < f_L \mid E(X) = -\log(1))$
- Find optimal values ($\max(Q_L)$) for boundaries
  - a_L, d_L, f_L

Determine "optimal" boundaries

- Assumption about true effects
  - (θs, θc) =(0,) stop for futility
  - (θs,θc)=(log(1/0.75),0) continue sub
  - (θs,θc)= (log(1/0.75), log(1/0.75)) continue full

Determine “optimal” boundaries

- Assumption about true effects
  - $(\theta s, \theta c) = (0,)$ stop for futility
  - $(\theta s, \theta c) = (\log(1/0.75), 0)$ continue sub
  - $(\theta s, \theta c) = (\log(1/0.75), \log(1/0.75))$ continue full
- Subpopulation fraction γ
- Information fraction τ

Determine “optimal” boundaries

- Assumption about true effects
  - (θs, θc) =(0,) stop for futility
  - (θs,θc)=(log(1/0.75),0) continue sub
  - (θs,θc)= (log(1/0.75), log(1/0.75)) continue full
- Subpopulation fraction γ
- Information fraction τ
- Timing of final analysis
  - Continue full population:
    - 508 events in full population, ~ γ*508 events in sub population

Determine "optimal" boundaries

- Assumption about true effects
  - (θs, θc) =(0,) stop for futility
  - (θs,θc)=(log(1/0.75),0) continue sub
  - (θs,θc)= (log(1/0.75), log(1/0.75)) continue full
- Subpopulation fraction γ
- Information fraction τ
- Timing of final analysis
  - Continue full population:
    - 508 events in full population, ~ γ*508 events in sub population
  - Continue sub population
    - 508 events in sub population

Determine “optimal” boundaries

- Assumption about true effects
  - (θs, θc) =(0,) stop for futility
  - (θs,θc)=(log(1/0.75),0) continue sub
  - (θs,θc)= (log(1/0.75), log(1/0.75)) continue full
- Subpopulation fraction γ
- Information fraction τ
- Timing of final analysis
  - Continue full population:
    - 508 events in full population, ~ γ*508 events in sub population
    - Power 90%, …
  - Continue sub population
    - 508 events in sub population
- Weights (ω_1 , ω_2 , ω_3) depending on prior assumption
  - **(full, sub, stop)**
  - (1/3, 1/3, 1/3)
  - (0.4, 0.4, 0.2)

“Optimal” boundaries for linear rule

- $\hat{\theta}_s < f_L$: Stop for futility
- $\hat{\theta}_s >= f_L$ & $a_L * \hat{\theta}_s + \hat{\theta}_c < d_L$: Continue sub
- $\hat{\theta}_s >= f_L$ & $a_L * \hat{\theta}_s + \hat{\theta}_c >= d_L$: Continue full

- a_L often 0 $\Rightarrow$ decision between sub and full based on complement
- Usually $d_L > f_L$

“Optimal” boundaries for linear rule

- $\hat{\theta}_s < f_L$: Stop for futility
- $\hat{\theta}_s >= f_L$ & $a_L * \hat{\theta}_s + \hat{\theta}_c < d_L$: Continue sub
- $\hat{\theta}_s >= f_L$ & $a_L * \hat{\theta}_s + \hat{\theta}_c >= d_L$: Continue full

- a_L often 0 $\Rightarrow$ decision between sub and full based on complement
- Usually $d_L > f_L$

- Example for subpop=0.5, information=0.3, weights **(full, sub, stop)=**(1/3, 1/3, 1/3)
  - $HR_S > 0.95$ Stop for futility
  - $HR_S \leq 0.95$ & $HR_C > 0.86$ Continue sub
  - $HR_S \leq 0.95$ & $HR_C \leq 0.86$ Continue full

"Optimal" boundaries for linear rule

- $\hat{\theta}_s < f_L$: Stop for futility
- $\hat{\theta}_s >= f_L$ & $a_L * \hat{\theta}_s + \hat{\theta}_c < d_L$: Continue sub
- $\hat{\theta}_s >= f_L$ & $a_L * \hat{\theta}_s + \hat{\theta}_c >= d_L$: Continue full

- a_L often 0 $\Rightarrow$ decision between sub and full based on complement
- Usually $d_L > f_L$

- Example for subpop=0.5, information=0.3, weights **(full, sub, stop)=**(1/3, 1/3, 1/3)
  - $HR_S > 0.95$ Stop for futility
  - $HR_S \leq 0.95$ & $HR_C > 0.86$ Continue sub
  - $HR_S \leq 0.95$ & $HR_C \leq 0.86$ Continue full

- Example for subpop=0.5, information=0.3, weights **(full, sub, stop)=**(0.4, 0.4, 0.2)
  - $HR_S > 1.05$ Stop for futility
  - $HR_S \leq 1.05$ & $HR_C > 0.86$ Continue sub
  - $HR_S \leq 1.05$ & $HR_C \leq 0.86$ Continue full

Performance comparison - Simulation

- Simulation of normalized test statistics based on all pairwise combinations of (0.65, 0.75, 0.85, 1) for (1/exp(θs),1/exp(θc))
- Optimal boundaries for
  - (θs,θc) =(0,) stop for futility
  - (θs,θc)=(-log(0.75),0) continue sub
  - (θs,θc)= (-log(0.75), -log(0.75)) continue full
- Results
  - Rate of correct interim decision
  - Power (reject at least one)

Probabilities of Interim Decisions (%)

Optimal Linear rule			(1/3,1/3,1/3)			(0.4,0.4,0.2)		
HR_S	HR_C	HR_F	full	sub	futility	full	sub	futility
0.650	1.000	0.806	24.8	**70.1**	5.1	25.9	**72.3**	1.8
0.750	0.750	0.750	**60.9**	23.9	15.2	**66.5**	26.2	7.2
0.750	1.000	0.866	21.6	**63.4**	15.0	23.3	**70.0**	6.7
1.000	1.000	1.000	11.1	30.2	**58.7**	15.8	42.6	**41.6**

MerckSerono

MERCK Living Innovation

Probabilities of Interim Decisions (%)

Optimal Linear rule			(1/3,1/3,1/3)			(0.4,0.4,0.2)		
HR_S	HR_C	HR_F	full	sub	futility	full	sub	futility
0.650	1.000	0.806	24.8	**70.1**	5.1	25.9	**72.3**	1.8
0.750	0.750	0.750	**60.9**	23.9	15.2	**66.5**	26.2	7.2
0.750	1.000	0.866	21.6	**63.4**	15.0	23.3	**70.0**	6.7
1.000	1.000	1.000	11.1	30.2	**58.7**	15.8	42.6	**41.6**

Simple rule					
HR_S	HR_C	HR_F	full	sub	futility
0.650	1.000	0.806	48.7	**48.3**	3.0
0.750	0.750	0.750	**79.8**	9.6	10.6
0.750	1.000	0.866	45.6	**44.1**	10.3
1.000	1.000	1.000	25.3	24.7	**50.1**

P(Reject at least one) (%)

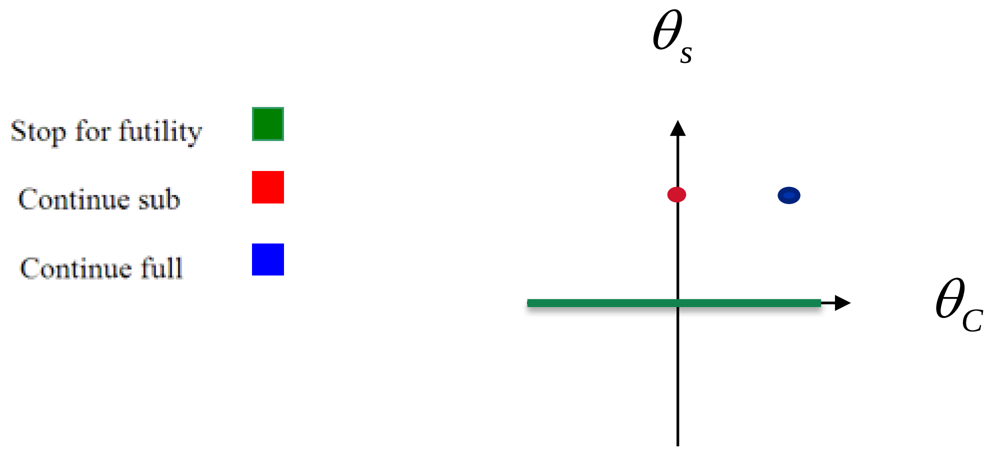
			Optimal Linear rule		Simple rule	Ctp w/o IA
HR_S	HR_C	HR_F	(1/3,1/3,1/3)	(0.4,0.4,0.2)		
0.650	1.000	0.806	93.1	96.2	93.1	90.7
0.750	0.750	0.750	76.7	82.8	78.8	86.2
0.750	1.000	0.866	72.7	78.0	67.6	56.5
1.000	1.000	1.000	1.7	1.8	1.6	2.2

Discussion

- Evaluation of decision rules in planning phase is important
  - Optimizing decision rules can substantially improve probabilities of correct decision and power compared to „intuitive“ decision rules
- Assumption or prior knowledge needed
  - Strong impact on results
  - Recommendation with promising results from phase II: not too much weight on stopping for futility
- Extension to other type of decision rule easy
  - For example: conditional power (CP)
    - $CP_s < f_{CP}$: Stop for futility
    - $CP_s >= f_{CP}$ & $CP_s >= CP_F + d_{CP}$: Continue sub
    - $CP_s >= f_{CP}$ & $CP_s < CP_F + d_{CP}$: Continue full
  - „Optimal“ CP rule lead to similar decisions as „optimal“ linear rule

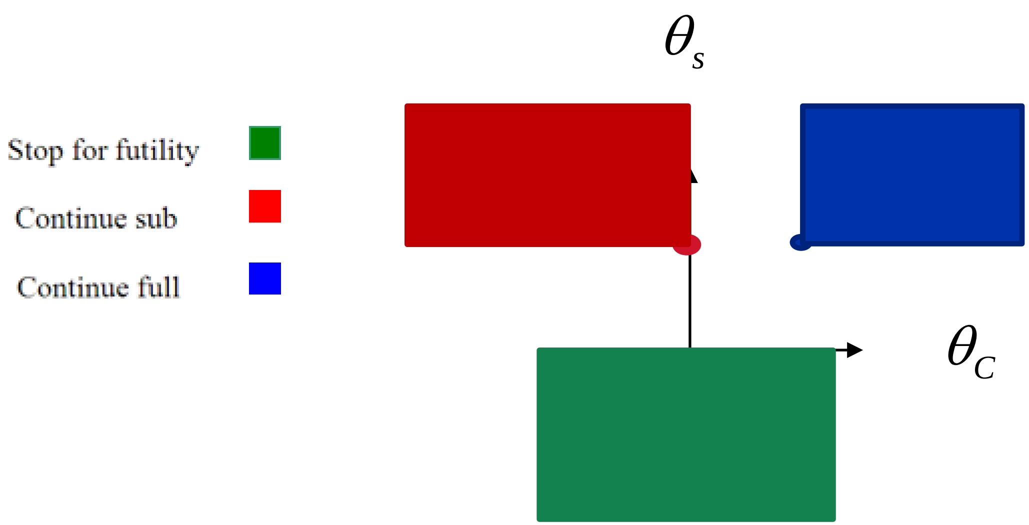
So far…

- “Points/lines” determine correct decisions
- Weights define how likely each case is (e.g. **(full, sub, stop)=**(0.4, 0.4, 0.2))



Extension

- “Areas” determine correct decisions
- Prior distribution based on phase II data define how likely each case is



$$f_{u_s} : N\left(-\log(0.6), 4/50\right), f_{u_c} : N\left(-\log(0.89), 4/50\right)$$

MerckSerono

MERCK Living Innovation

References

Götte H, Donica M, Mordenti G. Improving probabilities of correct interim decision in population enrichment designs. *J Biopharm Stat.* 2014 Jun 10. [Epub ahead of print]

Hommel, G. (2001) Adaptive modifications of hypotheses after an interim analysis. *Biometrical Journal* 43:581–589.

Lehmacher, W., Wassmer, G. (1999). Adaptive sample size calculations in group sequential trials. *Biometrics* 55, 1286–1290.

Marcus, R., Peritz, E., Gabriel, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 63: 655–660.

Wang, S. J., O’Neill, R. T., Hung, H. M. J. (2007). Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. *Pharmaceutical Statistics* 6: 227–244.

Back up

“Optimal” boundaries for linear rule

exp(0.15) =1.16
exp(0.10) =1.11
exp(0.05) =1.05
exp(-0.05)=0.95
exp(-0.10)=0.90
exp(-0.20)=0.82

τ	γ	ω_1	ω_2	ω_3	a_L	d_L	f_L
0.3	0.375	1/3	1/3	1/3	0.00	0.15	0.05
0.3	0.375	0.4	0.4	0.2	0.00	0.15	-0.15
0.3	0.375	0.5	0.35	0.15	0.00	0.10	-0.20
0.3	0.375	0.5	0.5	0	0.00	0.15	-1.00*
0.3	0.5	1/3	1/3	1/3	0.00	0.15	0.05
0.3	0.5	0.4	0.4	0.2	0.00	0.15	-0.05
0.3	0.5	0.5	0.35	0.15	-0.10	0.05	-0.10
0.3	0.5	0.5	0.5	0	0.00	0.15	-1.00*
0.6	0.375	1/3	1/3	1/3	0.00	0.15	0.10
0.6	0.375	0.4	0.4	0.2	0.00	0.15	0.00
0.6	0.375	0.5	0.35	0.15	-0.05	0.10	-0.05
0.6	0.375	0.5	0.5	0	0.00	0.15	-1.00*
0.6	0.5	1/3	1/3	1/3	-0.10	0.10	0.10
0.6	0.5	0.4	0.4	0.2	0.00	0.15	0.05
0.6	0.5	0.5	0.35	0.15	-0.05	0.10	0.00
0.6	0.5	0.5	0.5	0	0.00	0.15	-1.00*

- $\hat{\theta}_s < f_L$: Stop for futility
- $\hat{\theta}_s >= f_L$ & $a_L * \hat{\theta}_s + \hat{\theta}_c < d_L$: Continue sub
- $\hat{\theta}_s >= f_L$ & $a_L * \hat{\theta}_s + \hat{\theta}_c >= d_L$: Continue full

Phase II results often not conclusive