Comparison of clinical development plans for a confirmatory trial with subpopulation selection

Kaspar Rufibach, Methods, Collaboration & Outreach Group, Roche Basel
Joint work with Hoa Nguyen (Genentech) & Meng Chen (Roche Shanghai)
BBS seminar
13th March 2017

Generated 2017-03-07 at 14:45:12.

Question

Setup:

- Phase I done $\Rightarrow$ bring new drug efficiently to registration.
- Primary endpoint: **overall survival** (OS).
- **Binary** biomarker $\Rightarrow$ defines potential subpopulation.

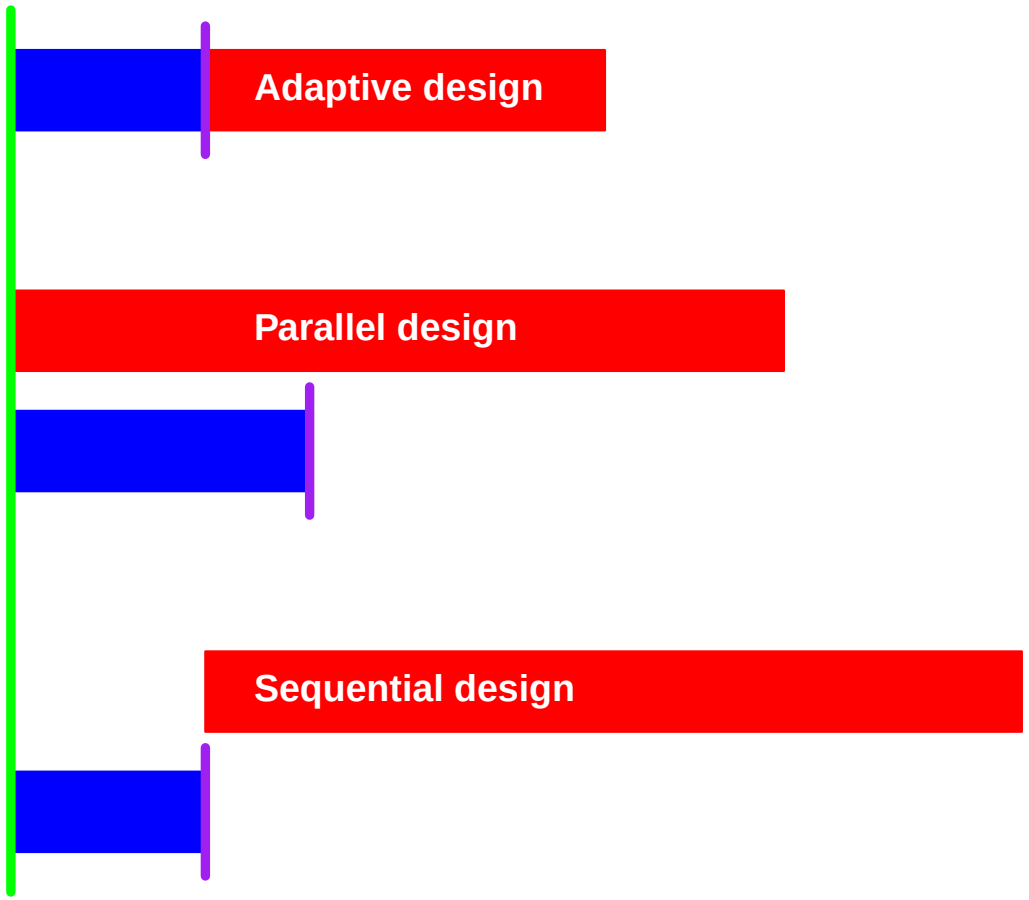
Questions to be asked in clinical trial:

1. Effect in full population?
2. Effect in subpopulation?
3. Effect in full and subpopulation, but enhanced in subpopulation?

Select at interim which scenario to pursue:

- Based on **quick** endpoint, e.g. progression-free survival (PFS).
- Maintain integrity of trial, i.e. protect (overall) type I error, for each way forward after interim.

Potential clinical development plans



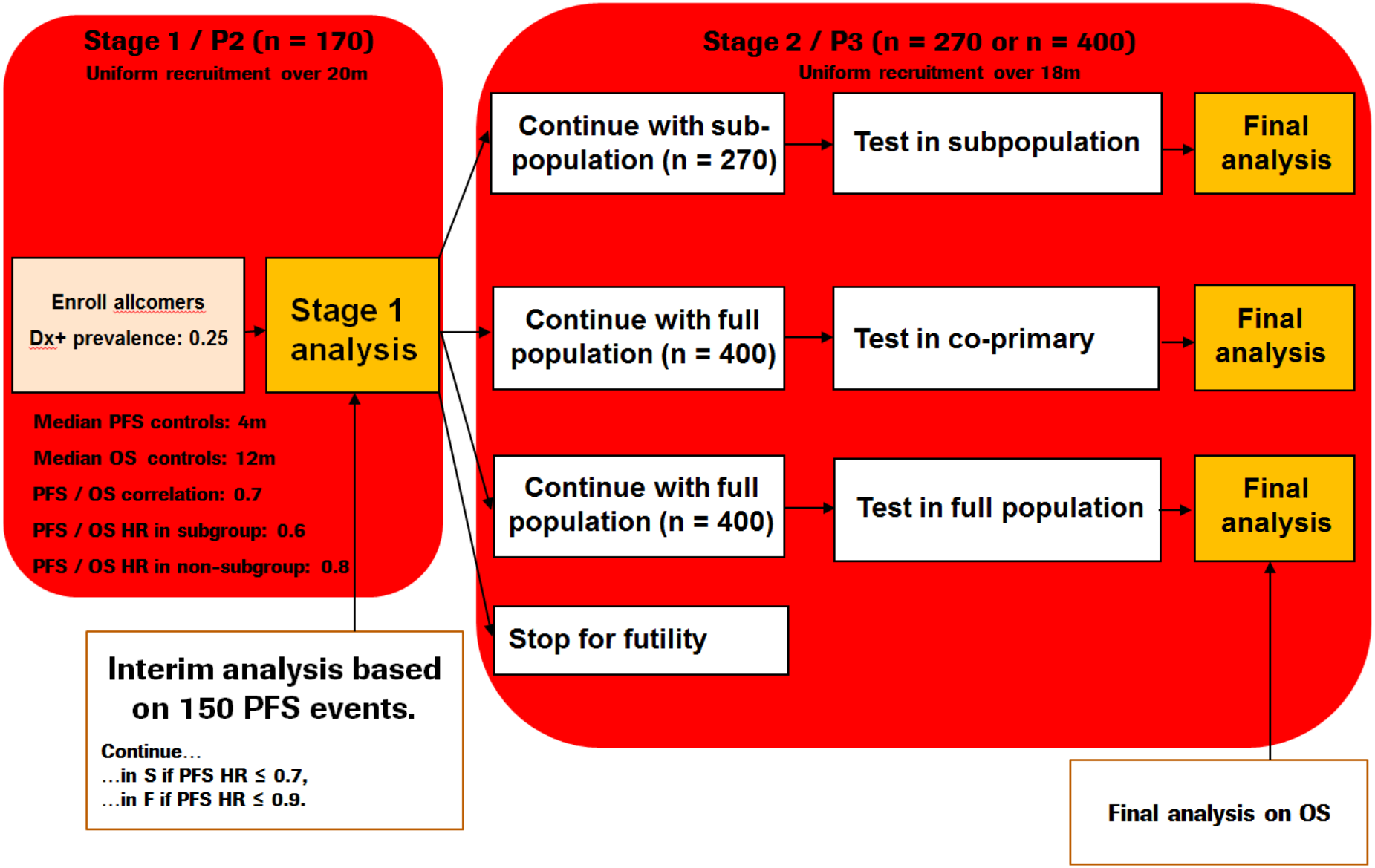
Final analysis combines data from Stage 1 & Stage 2.

Trial final analysis uses data from P3 only. P3 informed by P2 results.

Trial final analysis uses data from P3 only.

- Stage 1 = Phase II
- Stage 2 = Phase III
- Randomization
- Interim analysis

Base case assumptions



Base case assumptions

Base case assumptions:

- **Realistic** scenario in oncology.
- Interim decision instantaneously.
- **No white space** between Phase 2 and Phase 3 in sequential.
- **Accrual rate kept constant** in parallel $\Rightarrow$ longer recruitment time, as we need to fill two trials $\Rightarrow$ delayed interim.
- Exponential PFS and OS times with pre-specified correlation, Michael and Schucany (2002).

Base case assumptions

Interim decision rule: Set targets for PFS hazard ratio at interim:

- Continue in F if PFS hazard ratio ≤ 0.9.
- Continue in S if PFS hazard ratio ≤ 0.7.

Features:

- Easy interpretable and communicable.
- Corresponds to decision based on z-statistic since variance used to normalize based on pre-specified fixed number of events.

Co-primary endpoint: Correct for multiplicity using Hochberg's correction, as for adaptive design. See Jenkins et al. (2011). **Fair** comparison.

Power & timelines

Tune **recruitment** and **cutoffs** such that:

- Adaptive & sequential: Recruitment to Phase II / Stage I has finished prior to PFS interim cutoff.
- Parallel: Recruitment to Phase II has finished prior to PFS interim cutoff.
- Recruitment to Phase III / Stage II has finished prior to OS final cutoff.

Tune **number of events** such that:

- Sum of power to reject either null hypothesis is $\geq 80\%$.
- Stage I and II OS cutoffs are aligned.
- Cutoffs for OS final analysis are aligned in full and subpopulation.

Sequential and parallel: need more events as we do not reuse Phase 2 data.

Metrics

"Traditional" metric in statistical literature: assumptions $\Rightarrow$ power.

Metrics useful when **planning a study**:

1. Define targeted power (80%).
2. Vary patient numbers and/or recruitment rate (# centers) and explore
   - Time to interim PFS cutoff,
   - Time to final OS cutoff,

   among designs that reach targeted power.

"Traditional model": sequential design

Applicable if **not yet well-defined subgroup** at start of Phase 2:

- Still more than one marker to explore after Phase 1.
- Marker still under development.

Biomarker-unrelated: want to make **fully informed** decision at end of Phase 2.

Run randomized Phase 2, analyze data, decide whether to continue with a Phase 3

- in F and analyze in F only,
- in F and analyze in F and S (co-primary),
- in S only,
- without a Phase 3 (stop for futility).

Phase 3:

- Separate trial, do **not re-use data from Phase 2** in final analysis.
- Power: taking into account Phase 2! Power of entire program.
- Get result at final analysis of this trial.
- To get 80% power $\Rightarrow$ need more events $\Rightarrow$ need to wait longer.

Resulting timelines

	Sequential Seamless
PFS interim	26.7m
Final OS for Phase 2 patients	Phase 2 data
	not used
Phase 3 Final OS: F only or co-primary	163.4m
Phase 3 Final OS: S only	157.2m
Power	68%

Parallel Phase II/III design

Applicable if

- **complicated** marker $\Rightarrow$ ongoing assay development and/or cutoff determination.
- **Prevalence** of biomarker subgroup unclear.

Start Phase 2 & Phase 3 at same time. Analyze Phase 2 PFS data and inform Phase 3 whether to

- continue in F and analyze in F only,
- continue in F and analyze in F and S (co-primary),
- continue in S only,
- stop for futility.

Phase 3: still separate trial, Phase 2 data only used to inform Phase 3 $\Rightarrow$ external to Phase 3. **Accepted by regulators**.

Get result at final analysis of Phase 3 by analyzing all Phase 3 data.

Resulting timelines

	Sequential Seamless	Parallel Phase 2 & Phase 3
PFS interim	26.7m	42.3m
Final OS for Phase 2 patients	Phase 2 data not used	separate Phase 2 data used to inform Phase 3
Phase 3 Final OS: F only or co-primary	163.4m	74.3m
Phase 3 Final OS: S only	157.2m	73.5m
Power	68%	80%

Adaptive seamless design

Various approaches in literature: Brannath et al. (2009), Jenkins et al. (2011), Mehta et al. (2014).

We chose Jenkins et al. (2011) for our comparisons:

- Reasonably **simple** methodologically.
- Parameters to be determined are **easy to interpret and communicate**.
- Decision rule **transparent** and simple.
- PFS in Phase 2, OS in Phase 3 $\Rightarrow$ reflects a **realistic** sequential Phase 2 / Phase 3 scenario.

Phase 2 = Stage 1, Phase 3 = Stage 2.

Details in backup.

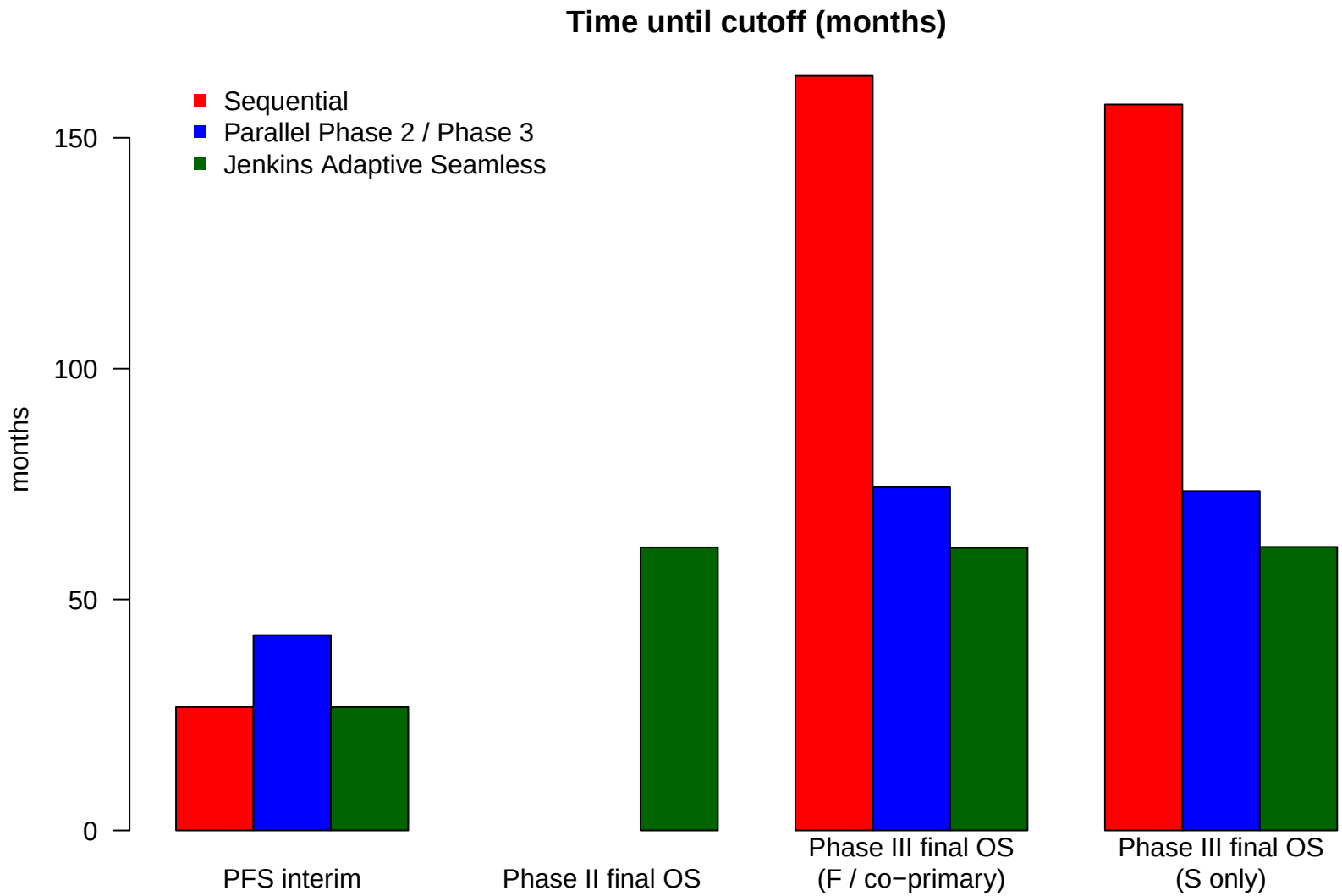
Resulting timelines

	Sequential Seamless	Parallel Phase 2 & Phase 3	Adaptive Seamless
PFS interim	26.7m	42.3m	26.7m
Final OS for Phase 2 patients	Phase 2 data not used	separate Phase 2 data used to inform Phase 3	61.3m
Phase 3 Final OS: F only or co-primary	163.4m	74.3m	61.2m
Phase 3 Final OS: S only	157.2m	73.5m	61.4m
Power	68%	80%	80%

Comments:

- Adaptive design: **tune parameters** so that design has desired features (power, cutoffs).
- Cutoffs for adaptive **aligned**.
- **Massive gain** in time for adaptive: re-use OS data of Stage 1 patients.
- Principle **broadly applicable**, not only for subpopulation selection. GATSBY trial: dose selection at interim. `http://clinicaltrials.gov/show/NCT01641939`

Resulting timelines



Drivers for the trial duration differences?

Sequential **slow**: Start P3 only once P2 is finished. Phase 2 OS data not re-used.

Adaptive faster than Parallel:

- We assume same accrual **rate** in Parallel Phase 2 & 3, but need to fill two trials $\Rightarrow$ interim decision 15.6m later.
- Recruitment rates: Stage 1: 8.5/m; Stage 2: 15/m (S only), 22.2/m (F or co-primary). Parallel **recruits longer at slower rate**. Might compensate through higher Operational effort, e.g. more centers.
- F or co-primary longer for sequential compared to parallel: in parallel, early recruited patients in Phase 3 already started to have events. Sequential: start recruitment from scratch after interim.
- p-value combination for adaptive entails small power loss $\Rightarrow$ duration increase for adaptive (compared to parallel).

Conclusions for timing

Substantial time gains Adaptive $>$ Parallel $>$ Sequential.

Adaptive uses less patients than parallel and is still faster.

Adaptive offers opportunity to substantially accelerate development.

Status of biomarker

Statistical literature: "Assume we have a biomarker that defines S" - typically not realistic **after Phase 1**:

- Biomarker hypothesis not very strong yet.
- Prevalence: Data in limited and highly selected number of patients so far only.
- Most binary biomarkers rely on dichotomizing continuous measurement $\Rightarrow$ estimation of cutoff that determines biomarker positive and negative notoriously difficult.
- Assay performance may vary between Phase I and Phase II/III, or across populations.
- Assay interpretation only consistent for specialized labs $\Rightarrow$ extension to more routine diagnostic?
- Adaptive $\Rightarrow$ filing strategy defined after Phase I $\Rightarrow$ risky.

Sponsor might not feel comfortable to start an adaptive enrichment Phase II/III trial that will potentially run for years, with being blinded to interim decision.

See Rufibach et al. (2015) for detailed discussion.

Overall conclusions

Substantial time gains Adaptive $>$ Parallel $>$ Sequential.

If

- stable binary biomarker and
- accurate idea about subgroup prevalence

after Phase 1 available $\Rightarrow$ opportunity to substantially accelerate development through use of adaptive seamless design.

Outlook: evaluate combination of PFS and OS information for interim decision in Jenkins' design, Brückner et al. (2017).

Acknowledgments

Martin Jenkins for providing initial version of code.

Thank you for your attention.

References

- Bauer, P. and Posch, M. (2004). Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections by H. Schäfer and H.-H. Müller, Stat. Med. 2001; 20: 3741–3751. *Stat. Med.* **23** 1333–1334.
- Brannath, W., Zuber, E., Branson, M., Bretz, F., Gallo, P., Posch, M. and Racine-Poon, A. (2009). Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. *Stat. Med.* **28** 1445–1463.
- Brückner, M., Brannath, W., Rufibach, K. (2017). Interim decisions in adaptive enrichment designs with progression-free and overall survival. *Preprint.*
- Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **75** 800–802.
- Jenkins, M., Stone, A. and Jennison, C. (2011). An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. *Pharm. Stat.* **10** 347–356.
- Mehta, C., Schafer, H., Daniel, H. and Irle, S. (2014). Biomarker driven population enrichment for adaptive oncology trials with time to event endpoints. *Stat. Med.* **33** 4515–4531.
- Michael, J. and Schucany, W. (2002). The mixture approach for simulating bivariate distributions with specified correlations. *The American Statistician* **56** 48–54.
- Rufibach, K., Chen, M. and Nguyen, H. (2015). Comparison of different clinical development plans for confirmatory subpopulation selection. *Contemp Clin Trials* **47** 78–84.

Jenkins' adaptive seamless design

Null hypotheses: $H_0^{\mathrm{F}}, H_0^{\mathrm{S}}$, and H_0^{FS} (no OS difference in neither F nor S). Protect **familywise error rate**.

Put together building blocks of "standard" statistical methods to set up adaptive trial design.

Raw p-values:

- $p_1^{\mathrm{F}}, p_1^{\mathrm{S}}$: From OS data of pts recruited in Stage 1, computed at end of Stage 2.
- $p_2^{\mathrm{F}}, p_2^{\mathrm{S}}$: From OS data of pts recruited in Stage 2, computed at end of Stage 2.

p-value for co-primary hypothesis via **Hochberg procedure**: under positive dependence compute multiplicity corrected p-value for H_0^{FS}:

$$p_i^{\mathrm{FS}} \quad = \quad \min[2\min\{p_i^{\mathrm{F}}, p_i^{\mathrm{S}}\}, \max\{p_i^{\mathrm{F}}, p_i^{\mathrm{S}}\}], \ i = 1, 2.$$

Hochberg (1988).

Jenkins' adaptive seamless design

Phase 2 PFS (used for interim decision) and Phase 2 OS data (used in final analysis) **correlated** $\Rightarrow$ **inverse Normal *p*-value combination** to combine Stage 1 and Stage 2 OS *p*-values:

Co-primary case – when considering both H_0^F and H_0^S

Testing H_0^F: $w_1 \Phi^{-1}(1 - p_1^F) + w_2 \Phi^{-1}(1 - p_2^F)$

Testing H_0^S: $w_1 \Phi^{-1}(1 - p_1^S) + w_2 \Phi^{-1}(1 - p_2^S)$

Testing H_0^{FS}: $w_1 \Phi^{-1}(1 - p_1^{FS}) + w_2 \Phi^{-1}(1 - p_2^{FS})$

F only case – when considering H_0^F only

Testing H_0^F: $w_1 \Phi^{-1}(1 - p_1^F) + w_2 \Phi^{-1}(1 - p_2^F)$

Testing H_0^{FS}: $w_1 \Phi^{-1}(1 - p_1^{FS}) + w_2 \Phi^{-1}(1 - p_2^F)$

S only case – when considering H_0^S only

Testing H_0^S: $w_1 \Phi^{-1}(1 - p_1^S) + w_2 \Phi^{-1}(1 - p_2^S)$

Testing H_0^{FS}: $w_1 \Phi^{-1}(1 - p_1^{FS}) + w_2 \Phi^{-1}(1 - p_2^S)$

Table 1 in Jenkins et al. (2011). Weights appropriately pre-specified. Need to test all hypotheses involved in closed test.

Jenkins' adaptive seamless design

Bauer and Posch (2004):

IF

selection of Stage 2 population is affected by PFS of Stage 1 subjects

AND

OS follow-up of Stage 1 subjects contributes to their Stage 2 logrank statistic

THEN

this statistic might not have desired null distribution.

Solution in Jenkins et al. (2011): Pre-specify length of OS follow-up of Stage 1 patients $\Rightarrow$ Stage 1 follow up not allowed to be affected by interim outcome $\Rightarrow$ combination test p-values **independent** and **uniform** under H_0.

Jenkins' adaptive seamless design

Closed testing for overall assessment for co-primary case: reject H^{F} only if H^{FS} is rejected, same for S.

Interim decision rule: easy interpretable and communicable.

Set targets for PFS hazard ratio at interim. Our base case:

- Continue in F if PFS hazard ratio ≤ 0.9.
- Continue in S if PFS hazard ratio ≤ 0.7.

Corresponds to decision based on z-statistic since variance used to normalize based on pre-specified fixed number of events.

Jenkins' adaptive seamless design

Biased estimates at final analysis. Assess bias via simulation.

Entire set-up pre-specified in protocol.

Building blocks put together such that overall significance level is controlled $\Rightarrow$ design feasible from regulatory perspective, but requires more discussion $\Rightarrow$ timelines.

Operationally more complex: follow-up of Stage 1 patients pre-specified, need quick interim decision (same for parallel), more upfront interactions with HAs, ...

Interim decision taken by iDMC!

Doing now what patients need next

R version and packages used to generate these slides:

R version: R version 3.3.1 (2016-06-21)

Base packages: stats / graphics / grDevices / utils / datasets / methods / base

Other packages:

This document was generated on 2017-03-07 at 14:45:12.