AptivSolutions
Accelerating the Possibilities

Case Studies for Adaptive Treatment Arm Selection and Population Enrichment Designs

Gernot Wassmer
Innovation Center
Aptiv Solutions

ROeS Dornbirn, Sep 10, 2013

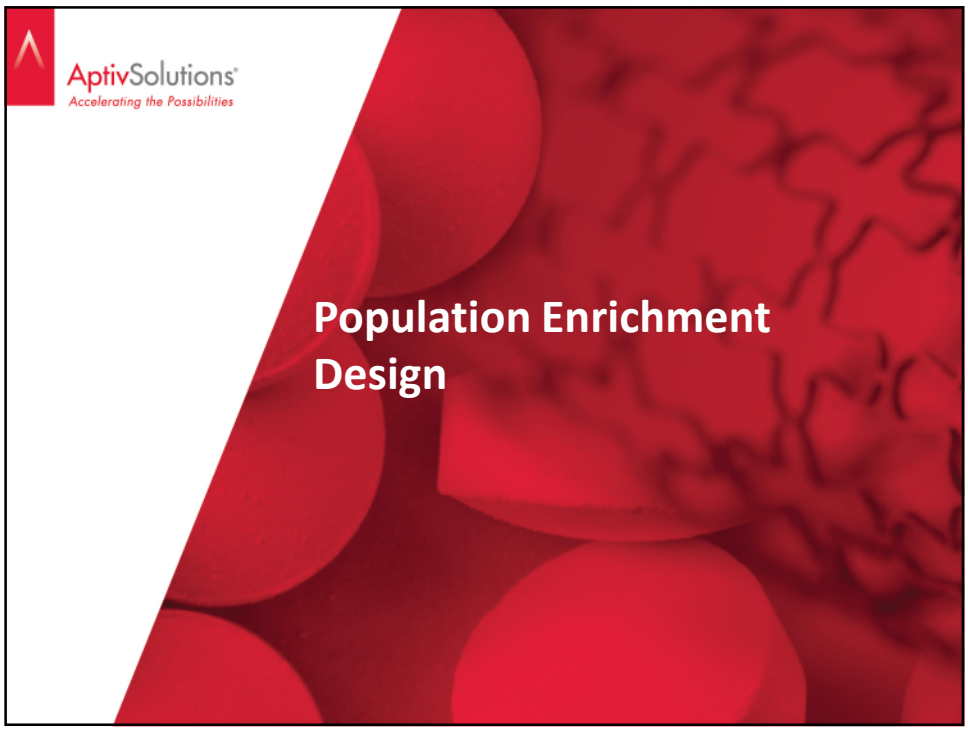
Confirmatory Adaptive Designs

Three particular applications of confirmatory adaptive designs

- Sample size reassessment
- Treatment arm selection in multi-armed designs
- Population enrichment designs

An attractive way to derive such designs is the combination testing principle together with the closed testing principle (Bauer & Köhne, 1994; Bauer & Kieser, 1999; Posch et al. 2005).

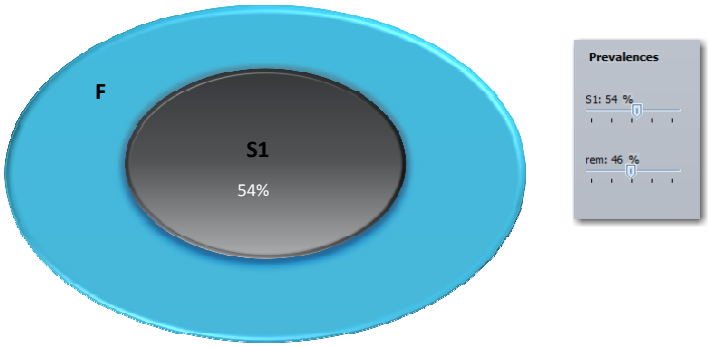
In this talk, two case studies are provided that illustrate the typical way of how to design such trials.



The Enrichment Test Procedure

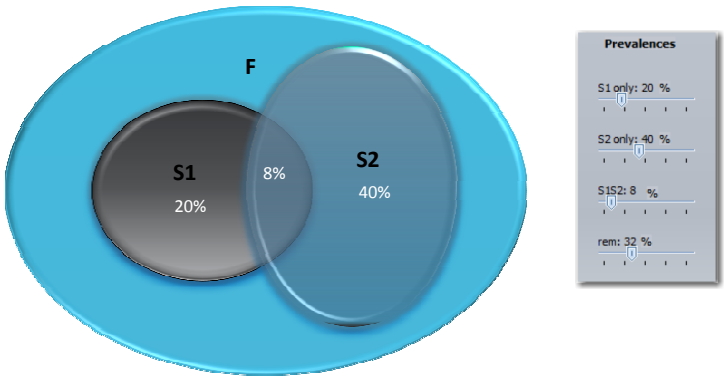
- For simplicity, we consider a two-sample comparison case although an extension to the multi-armed case is straightforward.
- Consider prespecified subpopulation(s) $S_1,\ldots,S_G$, and a full population F.
- At an interim stage it is decided which subpopulation is selected for further inference (including all subpopulations, i.e., full population).
- Not only selection procedures, but also other adaptive strategies (e.g., sample size reassessment) can be performed.
- Use the combination test for p-values of intersection tests within the closed testing procedure.

One sub-population



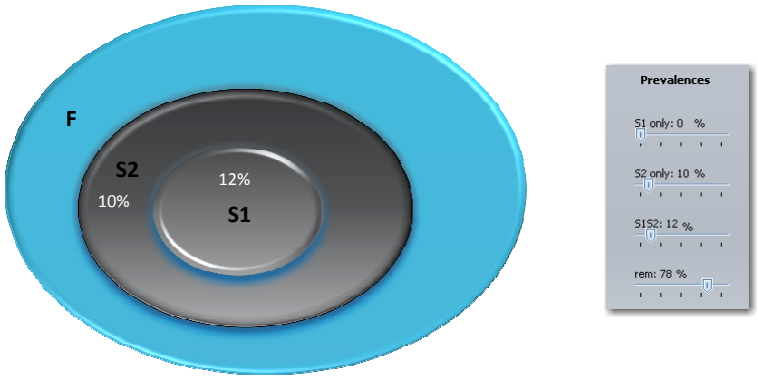
AptivSolutions 5

Two sub-populations of interest

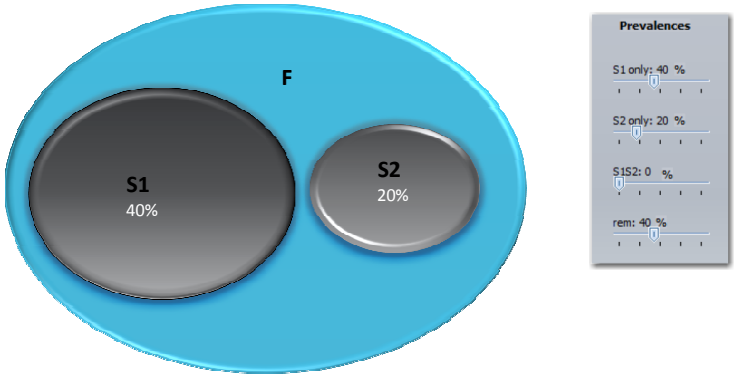


AptivSolutions 6

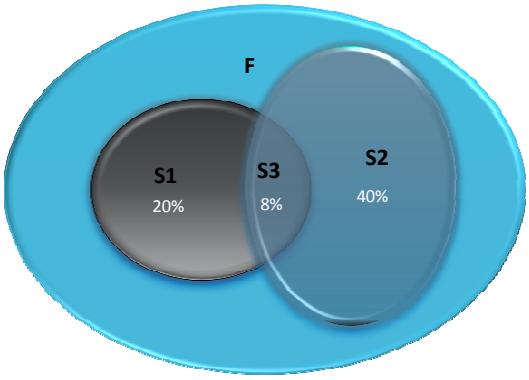
Two nested sub-populations of interest

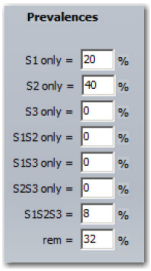


Two non-overlapping sub-populations



Three sub-populations of interest

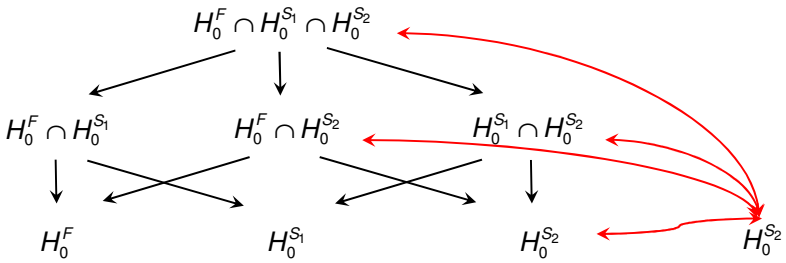




Closed Testing Procedure

Example: 2 stages $S = S_2$

Stage I **Stage II** **…**



$H_0^{S_2}$ can be rejected if all combination tests exceed the critical value u_2.

Closed Testing Procedure

- The choice of combination tests is free. E.g., you might use inverse normal or Fisher‘s combination test.
- The choice of tests for intersection hypotheses is free. E.g., you might use Bonferroni, Simes or Sidak tests.
- For one subgroup also Dunnett‘s test can be applied
- You might also use the CRP principle. i.e., perform conditional Dunnett test (Friede et al., Stat Med, 2012)
- Calculation of RCIs and overall p-values straightforward
- Except conditional Dunnett, all procedures available in ADDPLAN PE, Version 6.0

AptivSolutions
Accelerating the Possibilities

A Case Study Example

Simulation Example
Case Study: Phase 3 Trial in HER2- MBC Patients

- Assume that one of the experimental drugs has been graduated from the I-SPY 2 trial with the biomarker signature of triple negative breast cancer (TNBC) but also with some promising effect in HER2- biomarker signature.
- **Option 1**: a confirmatory Phase 3 trial in TNBC patients only
  - prevalence of TNBC is only about 34%
- **Option 2**: a confirmatory Phase 3 trial in HER2- patients
  - prevalence of HER2- is about 63%
- **Option 3**: Adaptive enrichment design
  - run a confirmatory trial with a two-stage enrichment design
  - starting with the full population (HER2- patients),
  - but with the preplanned option of selecting only the TNBC patients after the 1st stage in case the observed effect is not promising in the HER2- patients with positive hormone-receptor status HR+

Planning the Trial

- Primary Endpoint : pathologic complete response (pCR) at surgery
- Power: 90%
- Sign. Level: 0.025
- Control Rate: pCR=0.3
- TRT Effect: 0.2
- Apply Bonferroni correction

	Plan 1 Rates	Plan 2 Rates	Plan 3 Rates	Plan 4 Rates
alpha	0.0125	0.0125	0.0125	0.0125
Futility stops	-	-	-	-
tails	1	1	1	1
K	1	1	1	1
Design	-	-	-	-
Information rates	-	-	-	-
Hypothesis	diff<=0	diff<=0	diff<=0	diff<=0
Parameters	pi1=0.3 pi2=0.4	pi1=0.3 pi2=0.45	pi1=0.3 pi2=0.5	pi1=0.3 pi2=0.55
Power %	90.0	90.0	90.0	90.0
Total ASN H0	-	-	-	-
Total ASN H01	-	-	-	-
Total ASN H1	-	-	-	-
Total maximum N	1124.9	512.6	293.3	189.5
Allocation	1	1	1	1

Adaptive PE Simulation

Prevalences

	MP Hi-1		MP Hi-2	
	HR +	HR-	HR+	HR-
HER2+	16%	7%	4%	10%
HER2-	23%	6%	6%	28%

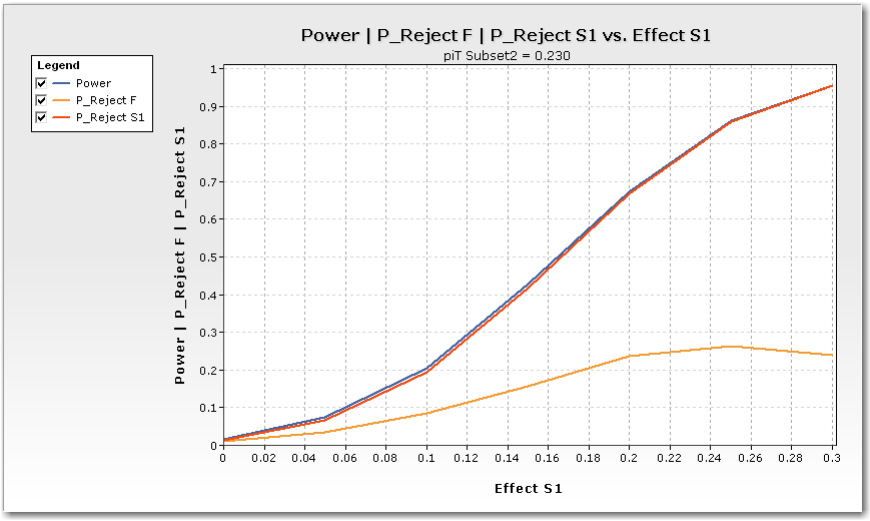
pCR rates

	MP Hi-1		MP Hi-2	
	HR +	HR-	HR+	HR-
HER2+	47%	67%	35%	55%
HER2-	25%	43%	17%	32%

- Prevalence of TNBC in HER2- : 54% (= (6 + 28)/63)
- Control pCR Rate in TNBC: 0.34 (= (6*0.43+28*0.32)/ 34))
- Control pCR Rate in HER2- ∩ HR+: 0.23 (= (23*0.25+6*0.17)/29)
- Total of 21 Simulation Scenarios:
  - TRT effect in TNBC: 0 to 0.3 by 0.05
  - TRT effect in HER2- ∩ HR+: 0, 0.1, 0.2
- Total sample size: 300 patients, stage 1 sample size:150 patients

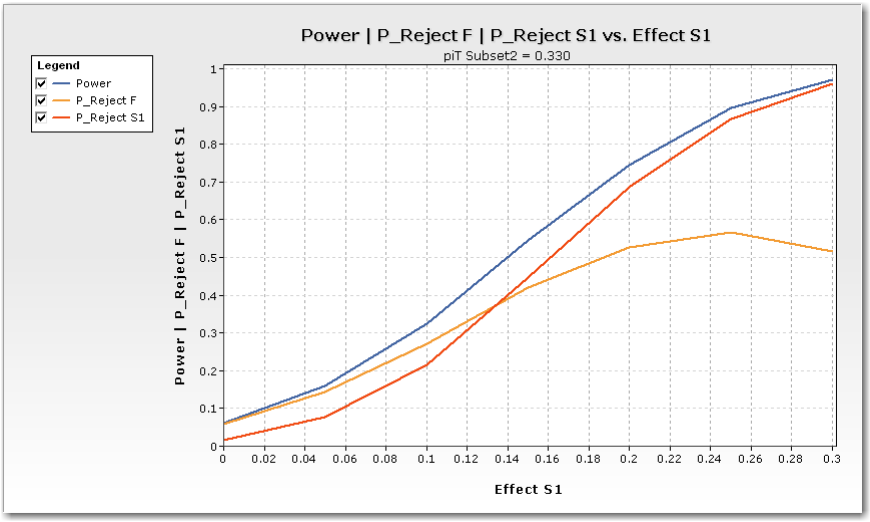
AptivSolutions 15

Operating Characteristics:

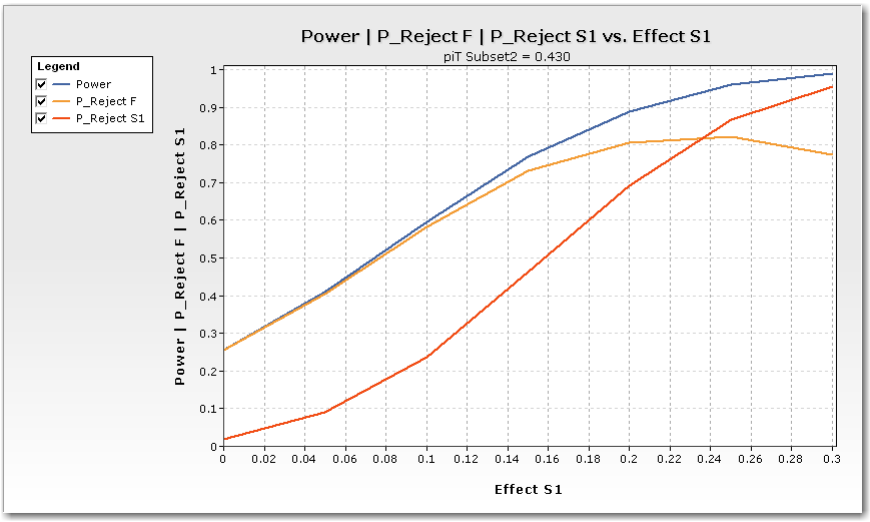


AptivSolutions 16

Operating Characteristics:

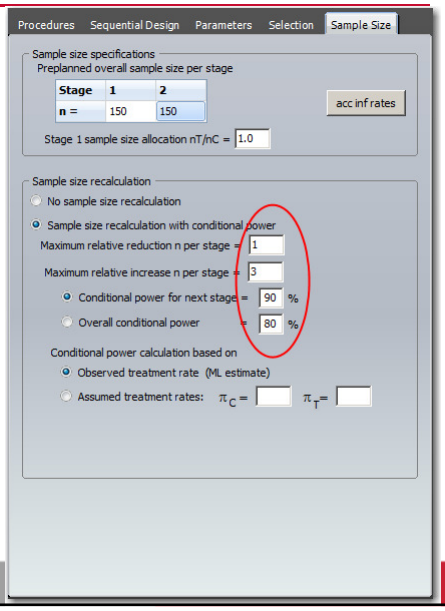


Operating Characteristics:

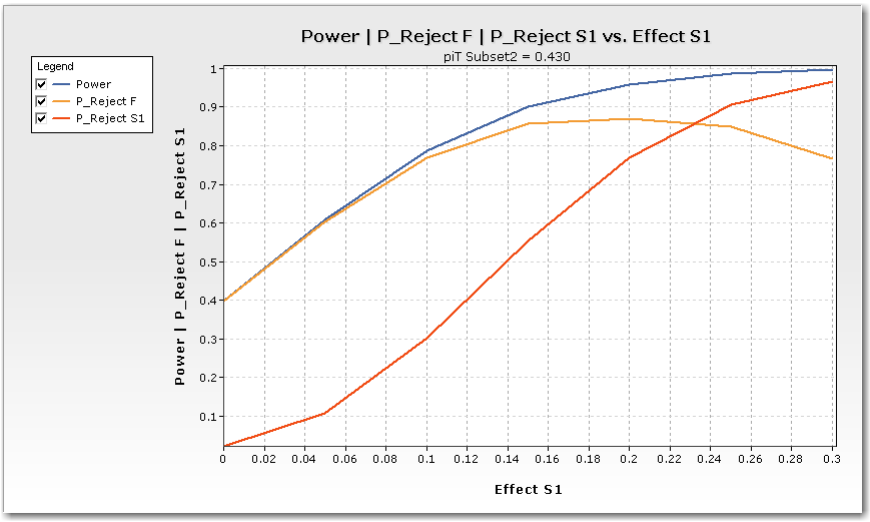


Sample Size Reestimation

- Allow up to a 3-fold sample size increase for Stage 2
- 90% Conditional Power based on observed TRT effect
- Total Sample Size:
  300 - 600



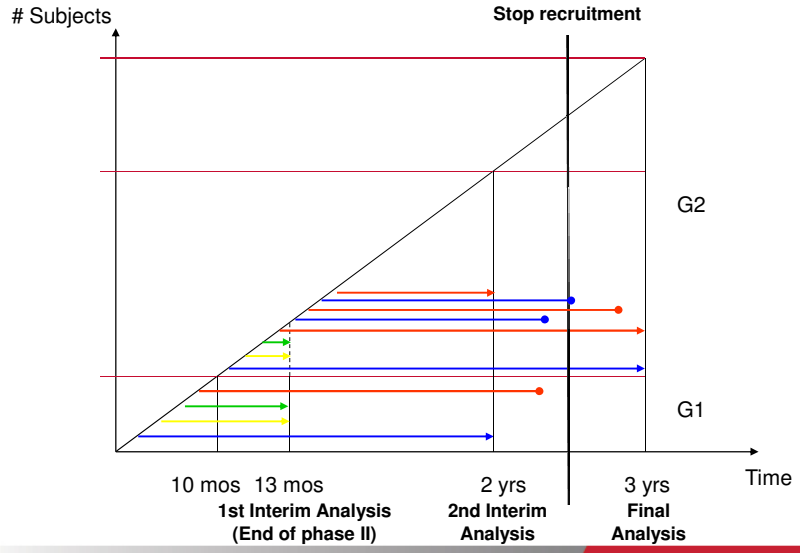
Operating Characteristics



Example Survival design with surrogate parameter for selection

- Indication: Chronic Heart Failure, endpoint time of first event of CV mortality or HF hospitalization
- Three-stage adaptive seamless design using O'Brien & Fleming boundaries and a survival endpoint
- Three doses of a drug against placebo, selection based on efficacy for surrogate parameter, measured after 3 months for each individual patient
- For survival designs, in general, only test statistic for confirmatory phase can be used for subsequent planning, no other information from patients under risk can be used (Bauer & Posch, 2004)
- A solution is to split the two populations into a Phase II and a Phase III part and combine the two populations through the use of a combination test (cf., Jenkins et al., 2011, Friede et al, 2012).

Recruitment and Analysis Pattern



AptivSolutions 23

Assumptions for Power Calculation and Comparison with Separate Phase II/III Design

- ✓ Group sequential boundaries according to O'Brien & Fleming alpha spending with $\alpha_1 = 0$. Information rates 0.40 and 1 acc. to expected events for second interim and final analysis.
- ✓ Inverse normal method (using Dunnett test for first stage) with weights according to sample sizes in Phase II and III (600/3000 = 0.20 and 0.80, resp.)
- ✓ Control hazard = 0.02, effects and correlation (Spearman rank correlation) as specified
- ✓ Selection rule: Select dose with highest effect and ratio >= 2
- ✓ First interim after observation of 600 patients + 3 months observation of short-term endpoint (i.e., after 13 months)
- ✓ Second interim after 24 months, final analysis after 36 months
- ✓ Patient accrual 60 patients per month between month 1 and 10, 90 between month 10 and 15, and 130 between month 15 and 30 (yielding 3000 patients) Drop out rate 16.3% after 24 months (acc. 30% after 4 yrs)

AptivSolutions 24

Gain in Estimated Study Duration

Study Duration and Power for Seamless Design

- Consider fixed hazards (0.02, 0.0166, 0.0156, 0.0149)
- Effect sizes of shortterm (0.1, 0.2, 0.25, 0.3)
- Correlation 0.50
- Simulation yields power 0.893 after month 36

Study Duration and Power for Separate Phase II/III

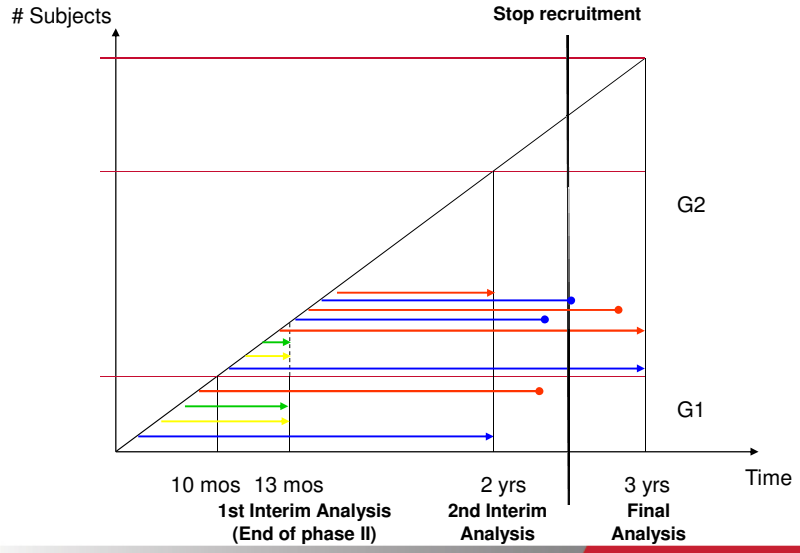
Study duration	Power
36	0.712
37	0.756
38	0.782
39	0.818
40	0.841
41	0.865
42	0.874
43	0.896
44	0.903

Expected gain in study duration is > 6 months with the same number of patients

Caveats

- At end of Phase II, due to the delayed response in the surrogate, a number of patient are randomized to deselected treatment arms and are not used for further analysis.
- Patients (in G1) from deselected treatment arms usually have discontinued follow-up (Friede et al., 2012), *z* statistic is set equal $-\infty$
- G1 population generally small and hence yields larger *p*-values → use of Bonferroni correction might yield adjusted *p*-value = 1.
- Only adaptive selection procedure is possible at first interim, other reassessment procedures are becoming more complicated.
- Procedure relies on asymptotic normality and independent increments structure of test statistic in G1 population. Simulations show that Type I error rate is controlled;
  Or: Calculate independent increments, and use inverse normal method (Wassmer, 2006);
  Or: Do not consider early stops for efficacy at interim.

Recruitment and Analysis Pattern



Caveats

- At end of Phase II, due to the delayed response in the surrogate, a number of patient are randomized to deselected treatment arms and are not used for further analysis.
- Patients (in G1) from deselected treatment arms usually have discontinued follow-up (Friede et al., 2012), *z* statistic is set equal $-\infty$
- G1 population generally small and hence yields larger *p*-values → use of Bonferroni correction might yield adjusted *p*-value = 1.
- Only adaptive selection procedure is possible at first interim, other reassessment procedures are becoming more complicated.
- Procedure relies on asymptotic normality and independent increments structure of test statistic in G1 population. Simulations show that Type I error rate is controlled;
  Or: Calculate independent increments, and use inverse normal method (Wassmer, 2006);
  Or: Do not consider early stops for efficacy at interim.

Discussion

- From a statistical point of view, in many cases there is an increased efficiency of (inferentially) seamless Phase II/III designs as compared to other approaches.
- There are cases where the gain is only small or even reversed, in such cases these designs may not be preferable.
- Our approach assumes neither the rule for the selection nor the number of selected arms to be prespecified.
- These designs require increased resources, e.g., more upfront planning, adequate operational infrastructure, well educated study team, etc.
- For survival designs, selection rules that are based on a surrogate are difficult (though not impossible) to handle.
- Generally, the role of simulations is becoming increasingly important when using these designs. Simulation results help to decide which type of design is reasonable to use. New ADDPLAN 6.1 provides such types of simulations.

References

- Bauer, P., Kieser, M. (1999). Combining different phases in the development of medical treatments within a single trial. Statistics in Medicine 18,1833–1848.
- Bauer, P., Köhne, K. (1994). Evaluation of experiments with adaptive interim analyses. Biometrics 50, 1029–1041.
- Bauer, P., Posch, M. (2004). Letter to the Editor: Modication of the sample size and the schedule of interim analyses in survival trials based on data inspections. Statistics in Medicine 23: 1333-1335.
- Friede, T., Parsons, N., Stallard, N. (2012): A conditional error function approach for subgroup selection in adaptive clinical trials. Statistics in Medicine 31 4309-4320; correction in 32: 2513–2514.
- 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. Pharmaceut. Statist. 10, 347–356.
- Posch, M., König, F., Branson, M., Brannath, W., Dunger-Baldauf, C., Bauer, P. (2005). Testing and estimation in flexible group sequential designs with adaptive treatment selection. Statistics in Medicine 24, 3697–3714.
- Wassmer, G. (2006). Planning and analyzing adaptive group sequential survival trials. Biometrical J. 48 714-729.
- Wassmer, G. (2011). On Sample Size Determination in Multi-Armed Confirmatory Adaptive Designs. J. Biopharm. Stat., 21, 802-817
- Wassmer, G., Dragalin, V. (2013). Designing Issues in Confirmatory Adaptive Population Enrichment Trials. Submitted.