Warwick Medical School

Adaptive clinical trials with subgroup selection

Nigel Stallard[1], Tim Friede[2], Nick Parsons[1]
[1]Warwick Medical School, University of Warwick, Coventry, UK
[2]University Medical Center Göttingen, Göttingen, Germany
n.stallard@warwick.ac.uk

EFSPI Statistical Meeting on Subgroup Analyses
30 November 2012

THE UNIVERSITY OF WARWICK

Setting

Warwick Medical School

Definitive comparison of experimental treatment with control

Single predefined subgroup of interest

Research questions: is treatment effective in whole population

test $H_0^{\{F\}}$

is treatment effective in subgroup

test $H_0^{\{S\}}$

Confirmatory study:

wish to control risk of any false positive result

(Wang et al 2007, Song & Chi 2007, Brannath et al 2009, Spiessens & Debois 2010, Jenkins et al 2011)

Adaptive design – idea

Warwick Medical School

We are testing hypotheses in full population and subgroup

If we knew we could focus on subgroup, could recruit only subgroup patients

Need to adjust for multiple hypotheses to control error rate
This makes it harder to show efficacy
If we knew which hypothesis to test, could focus on that

Adaptive design:

- conduct interim analysis partway through trial
- use interim data to guide recruitment and hypothesis testing in remainder of trial

Adaptive design – details

Warwick Medical School

Two-stage trial

Stage 1: Recruit n_1 patients per group from full population

On basis of interim data decide to continue to

(i) test $H_0^{\{F\}}$ and $H_0^{\{S\}}$ at final analysis

(ii) test $H_0^{\{F\}}$ only final analysis

(iii) test $H_0^{\{S\}}$ only final analysis

(possibly use interim data on short-term endpoint or external information)

Adaptive design – details

Warwick Medical School

Stage 2: Recruit further n_2 patients per group

If testing $H_0^{\{F\}}$ and $H_0^{\{S\}}$ or $H_0^{\{F\}}$ alone:

continue to recruit from full population

If testing $H_0^{\{S\}}$ alone: recruit from subgroup only

(possibly with enrichment)

Final analysis:

Test selected hypothesis/hypotheses

Control familywise error rate allowing for hypothesis selection

(Song & Chi 2007, Brannath et al 2009, Jenkins et al 2011)

Hypothesis testing

Warwick Medical School

Test $H_0^{\{F\}}$ (full population) and $H_0^{\{S\}}$(sub population

Require strong control of family wise error rate (FWER):

$$Pr(\text{erroneously reject either } H_0^{\{F\}} \text{ or } H_0^{\{S\}}) = \alpha$$

under $H_0^{\{F\}}$ or $H_0^{\{S\}}$ or $H_0^{\{F\}} \cap H_0^{\{S\}}$

Adaptive design – ‘ingredients’

Warwick Medical School

We need:

hypothesis testing approach to strongly control FWER allowing for multiple hypotheses

method for combination of evidence from two stages

decision rule for deciding which hypothesis/hypotheses to consider in stage 2

Test statistics

Warwick Medical School

Test $H_0^{\{F\}}$ (full population) and $H_0^{\{S\}}$(sub population

At stage i, let

$p_i^{\{F\}}$ be p-value for testing $H_0^{\{F\}}$ (full population)

$p_i^{\{S\}}$ be p-value for testing $H_0^{\{S\}}$ (sub population)

$Z_i^{\{F\}}$ be test statistic for testing $H_0^{\{F\}}$ (full population)

$$Z_i^{\{F\}} = \Phi^{-1}(1 - p_i^{\{F\}})$$

$Z_i^{\{S\}}$ be test statistic for testing $H_0^{\{S\}}$ (sub population)

$$Z_i^{\{S\}} = \Phi^{-1}(1 - p_i^{\{S\}})$$

Closed testing procedure

Warwick Medical School

Extend family of hypotheses to include

$$H_0^{\{F\}}, H_0^{\{S\}}, H_0^{\{F,S\}} = H_0^{\{F\}} \cap H_0^{\{S\}}$$

Perform 'local' level α test of each hypothesis

Reject $H_0^{\{F\}}$

if and only if $H_0^{\{F\}}$ and $H_0^{\{F\}} \cap H_0^{\{S\}}$ are locally rejected

Reject $H_0^{\{S\}}$

if and only if $H_0^{\{S\}}$ and $H_0^{\{F\}} \cap H_0^{\{S\}}$ are locally rejected

Controls FWER in strong sense

Requires level test of intersection hypothesis $H_0^{\{F,S\}}$

Tests of the intersection hypothesis

Warwick Medical School

At stage i, base test of $H_0^{\{F,S\}}$ on p-value $p_i^{\{F,S\}}$

Bonferroni test

$$p_i^{\{F,S\}} = 2 \min\{p_i^{\{F\}}, p_i^{\{S\}}\}$$

Simes test

$$p_i^{\{F,S\}} = \min\{2 \min\{p_i^{\{F\}}, p_i^{\{S\}}\}, \max\{p_i^{\{F\}}, p_i^{\{S\}}\}\}$$

(Simes 1986)

These tests ignore correlation between $p^{\{F\}}$ and $p^{\{S\}}$

Tests of the intersection hypothesis

Warwick Medical School

Spiessens and Debois test

Under $H_0^{\{S\}} \cap H_0^{\{S\}}$

$$\begin{pmatrix} Z_i^{\{F\}} \\ Z_i^{\{S\}} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sqrt{\tau} \\ \sqrt{\tau} & 1 \end{pmatrix}\right)$$

where τ is proportion in subgroup

Obtain $p_i^{\{F,S\}}$ based on distribution of $\max\{Z_i^{\{S\}}, Z_i^{\{F\}}\}$

(Spiessens and Debois 2010)

Combining stages 1 and 2

Warwick Medical School

Combination test

Obtain overall p-value for each hypothesis using pre-specified combination function

E.g. inverse normal combination test for based on

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

for pre-specified w_1 and w_2 with $w_1^2 + w_2^2 = 1$

Requirement is that p-values are independent (or at least the p-clud condition holds)

(Brannath et al 2002)

Combining stages 1 and 2

Warwick Medical School

Conditional error function approach

Obtain design assuming testing both hypotheses

At interim analysis obtain error rate for second stage conditional on stage 1 data

Can redesign stage 2 so long as conditional error rate is controlled

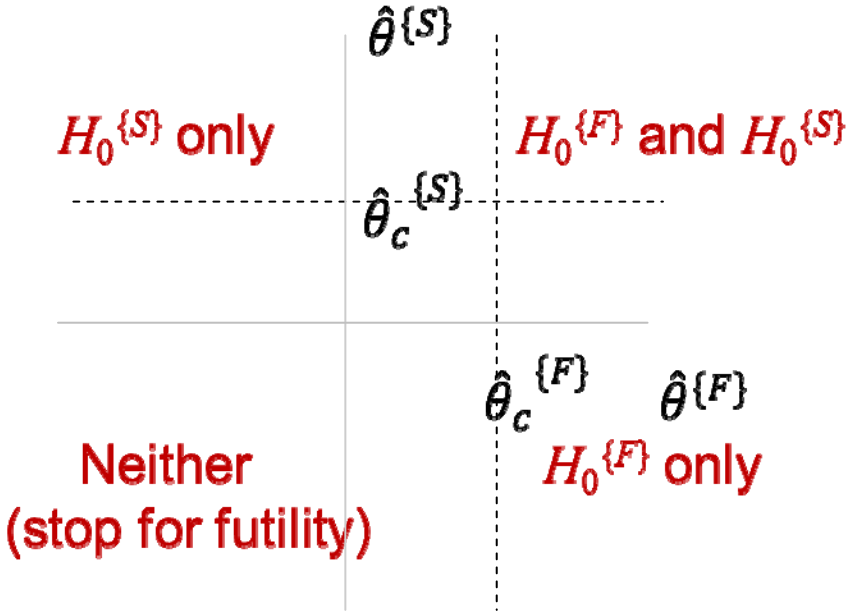
e.g. restrict testing to full population or subgroup alone

(Müller and Schäfer 2001, Friede et al 2011)

Decision rule

Base hypothesis selection on estimated treatment effects:
$\hat{\theta}^{\{F\}}$(full population), $\hat{\theta}^{\{S\}}$ (sub population)

Test $H_0^{\{F\}}$ if and only if
$\hat{\theta}^{\{F\}} \geq \hat{\theta}_c^{\{F\}}$
Test $H_0^{\{S\}}$ if and only if
$\hat{\theta}^{\{S\}} \geq \hat{\theta}_c^{\{S\}}$



(Jenkins et al 2011)

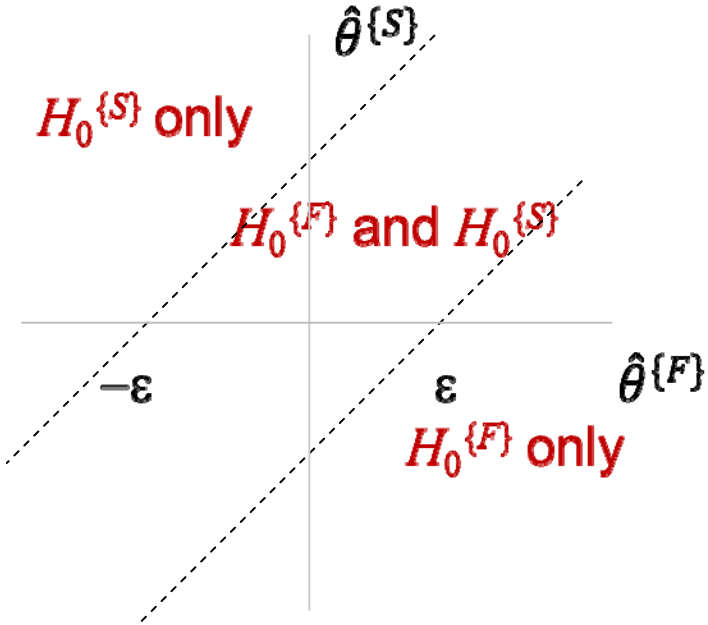
Decision rule

Warwick Medical School

Base hypothesis selection on estimated treatment effects:
$\hat{\theta}^{\{F\}}$(full population), $\hat{\theta}^{\{S\}}$ (sub population)

Test $H_0^{\{F\}}$ or $H_0^{\{S\}}$ according to maximum of $\hat{\theta}^{\{F\}}$ and $\hat{\theta}^{\{F\}}$

Test both if
$\left|\hat{\theta}^{\{F\}} - \hat{\theta}^{\{S\}}\right| \leq \varepsilon$



(Friede et al 2011)

Simulation study

Warwick Medical School

Model:

Normally distributed test statistics

Standardised treatment difference:

0.3 in subgroup

0 outside subgroup

Subgroup prevalence: 1% to 50%

Simulation study

Warwick Medical School

Designs:

Adaptive designs: CEF test with Spiessens and Debois test
combination test with Spiessens test
combination test with Simes test

Fixed designs: single study testing both hypotheses
two separate studies: second to test hypothesis selected by first

Sample size:

$n_1 = 200, n_2 = 200$
fixed single stage: $n = 400$
if enrichment:
all n_2 from subgroup if test $H_0^{\{S\}}$ only

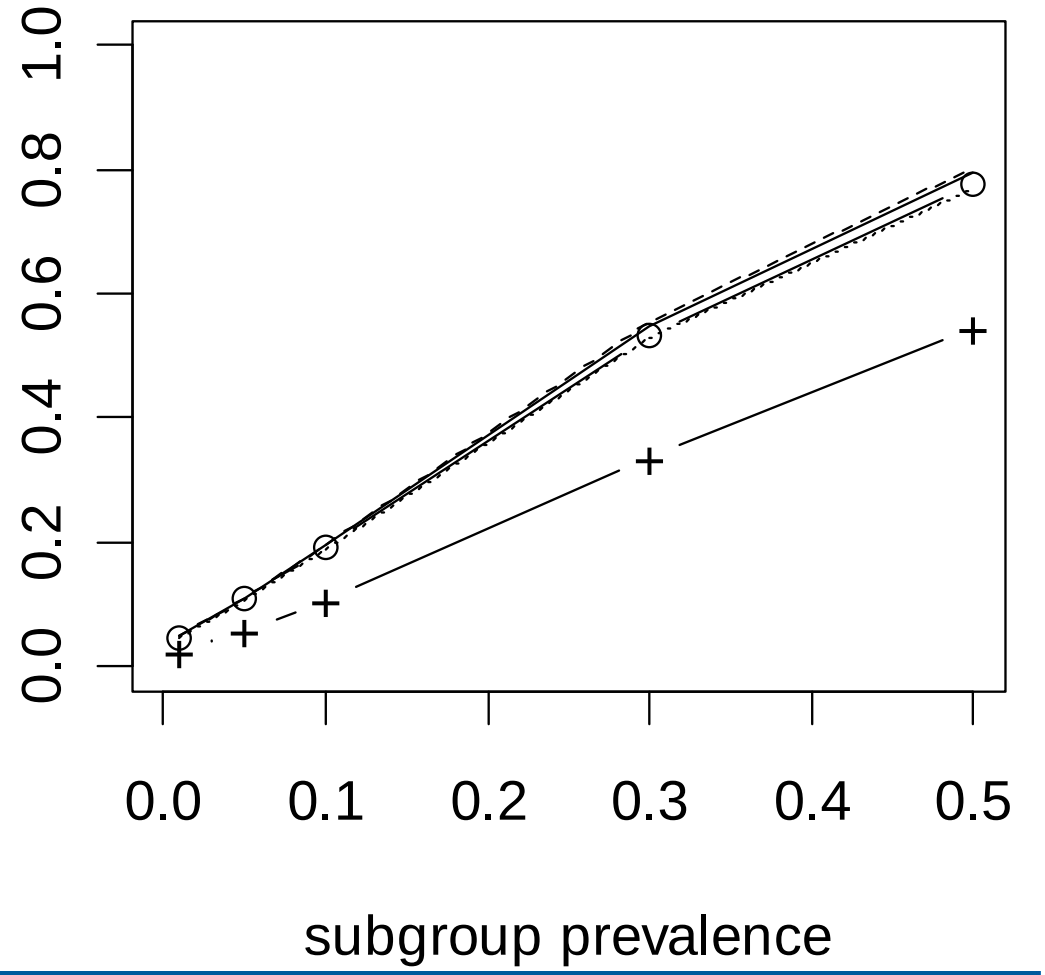
Simulation study results

Without enrichment

Warwick Medical School

Adaptive:
CEF ——
Comb test Simes ------
Comb test Spiessens — — ·

Fixed:
Separate studies + - +
Single study o - o



Simulation study results

Without enrichment

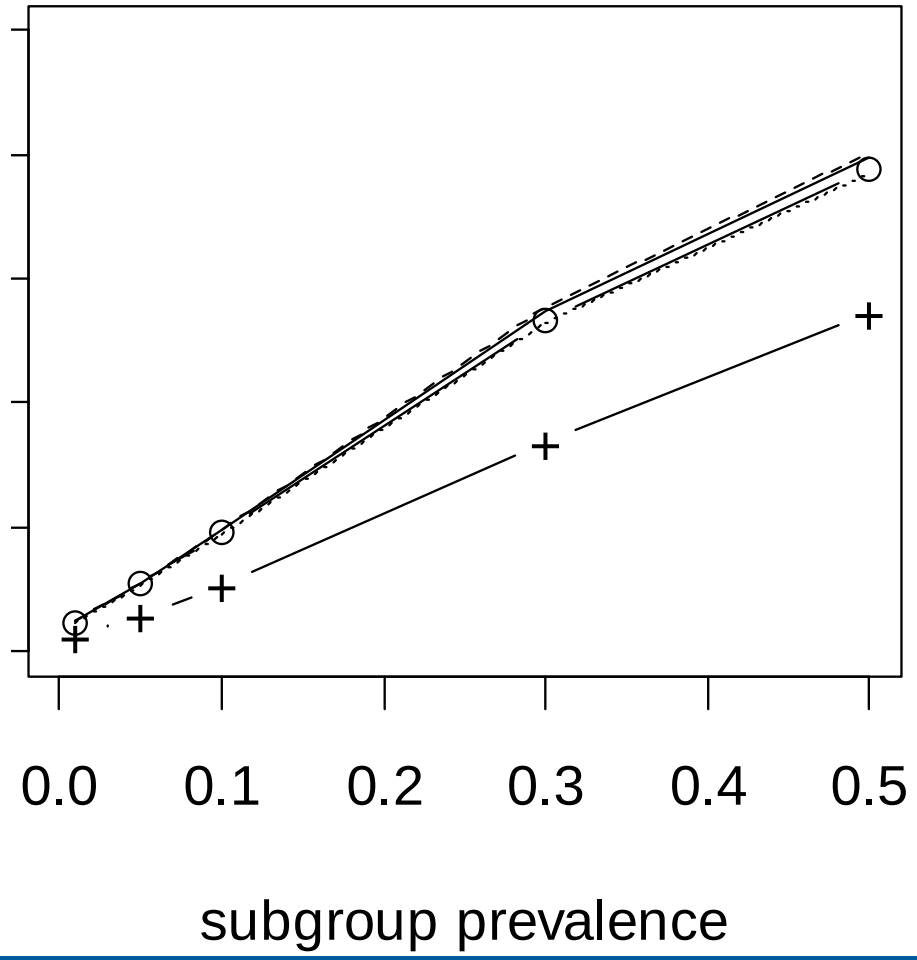
Key messages:

-Power increases with subgroup size

-Including data from both stages increases power

-Gain from allowing testing of only one hypothesis in second stage is relatively small

-Adaptive designs are very similar (Simes method slightly less powerful)



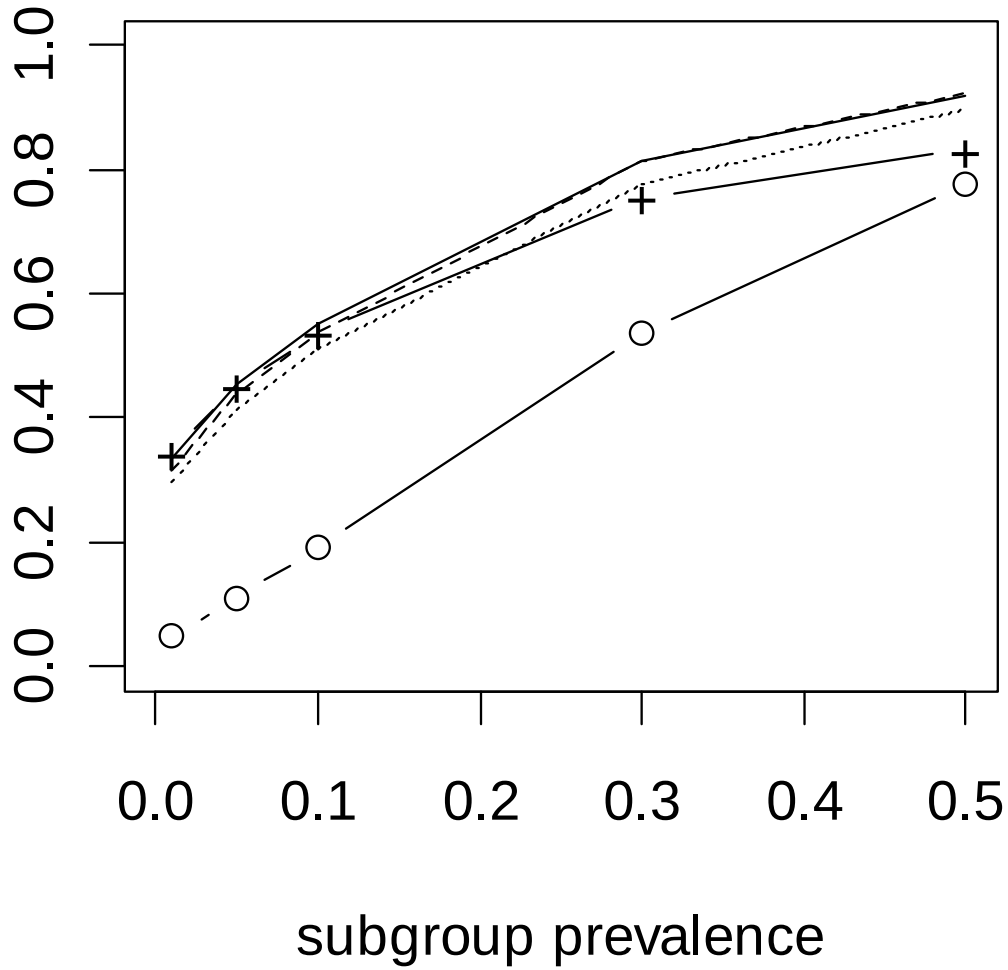
Simulation study results

With enrichment

Warwick Medical School

Adaptive:
CEF ——— power
Comb test Simes ------
Comb test Spiessens — — ·

Fixed:
Separate studies + - +
Single study o - o



Simulation study results

With enrichment

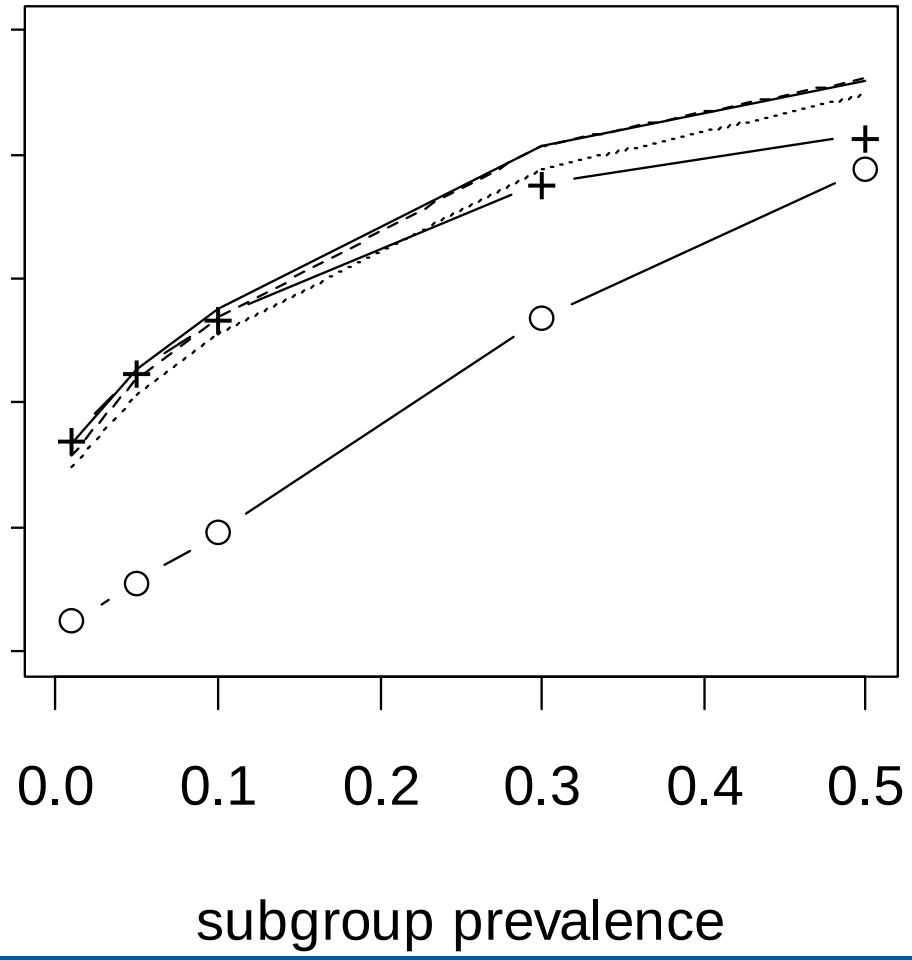
Key messages:

-Enrichment increases power, reduces effect of subgp. size

-Single stage design (without enrichment) is less powerful

-Gain from including first stage data is relatively small, particularly for small subgroup

-Adaptive designs are similar (Simes method slightly less powerful)



Conclusions

Warwick Medical School

We may want to test treatment in a subgroup in addition to full population

In confirmatory test need to control overall error rate

Adaptive design allows use of interim analysis to
(i) select subgroup and enable enrichment
(ii) select hypothesis(hypotheses) for final analysis

A number of methods exists to control overall error rate

Conclusions

Warwick Medical School

When possible, enrichment increases power considerably

Adaptive design can increase power by using data from both stages

Adaptive methods that allow for correlation between full- and subgroup tests are slightly more powerful

We have assumed a pre-defined subgroup. Allowing for use of interim data to select subgroup is harder!

References

Brannath et al. Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. *Stat Med 2009; 28: 1445–1463.*

Brannath, Posch, Bauer. Recursive combination tests. *J. Am. Stat. Ass. 2002, 97, 236-44.*

Jenkins, Stone, Jennison. An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. *Pharm Stat 2011; 10: 347–56.*

Friede, Parsons, Stallard. A conditional error function approach for subgroup selection in adaptive clinical trials. *Stat Med*. In press.

Müller, Schäfer. Adaptive group sequential designs for clinical trials: Combining the advantage of adaptive and of classical group sequential approaches. *Biometrics 2001; 57: 886–91.*

Simes. An improved Bonferroni procedure for multiple test of significance. *Biometrika 1986; 73: 751–4.*

Song, Chi. A method for testing a prespecified subgroup in clinical trials. *Stat Med 2007; 26: 3535-49.*

Spiessens, Debois. Adjusted significance levels for subgroup analysis in clinical trials. *Cont Clin Trials 2010; 31: 647–56.*

Wang, O'Neill, Hung. Approaches to evaluation of treatment effect in randomised clinical trials with genomic subset. *Pharm Stat 2007; 6: 227–4.*