# Confirmatory subgroup analysis: Multiple testing approaches

Alex Dmitrienko

Center for Statistics in Drug Development, Quintiles

JSM 2013

# Outline

**Clinical trials with tailoring objectives**

Clinical trials with pre-specified subpopulations

**Key statistical considerations**

Multiplicity adjustment to control overall Type I error rate

# Clinical trials with tailoring objectives

**Clinical trials with multiple patient populations**

Overall population and one or more subpopulations based on a pre-specified classifier

Enhanced efficacy is expected in the subpopulations

**Confirmatory subgroup analysis**

Overall population and subpopulations are equally important

Efficacy in at least one population provides foundation for registration

# Pre-specified subpopulations

**Temozolomide trial**

Trial in patients with glioblastoma (Hegi et al., 2005)

Classifier: MGMT (Methylguanine DNA-methyltransferase)

**Erlotinib trial**

Trial in patients with advanced NSCLC (Brugger et al., 2011)

Classifier: EGFR (Epidermal growth factor receptor)

# Two-population setting

**Populations**

Population $O$: Overall population

Population $M_+$: Marker-positive population

Population $M_-$: Marker-negative population

**Hypothesis testing problem**

$H_0$ and $H_+$, null hypotheses of no effect in Populations $O$ and $M_+$

Successful outcome if at least one null hypothesis is rejected

# Multiplicity adjustment

## Error rate control

Control familywise error rate for $\{H_0, H_+\}$ at one-sided $\alpha = 0.025$ to enable regulatory claims in both populations

## Clinical information

Account for logical relationships: $H_0$ and $H_+$ are interchangeable

## Statistical information

Utilize all available distributional information: Test statistics for $H_0$ and $H_+$ are strongly positively correlated

# Multiplicity adjustment procedures

**Fixed-sequence procedure**

**Chain procedures**

Bonferroni-based (nonparametric) chain procedures (Bretz et al., 2009)

Parametric chain procedures (Huque and Alosh, 2008; Alosh and Huque, 2009; Millen and Dmitrienko, 2011)

**Feedback procedures**

Family of feedback procedures (Zhao, Dmitrienko and Tamura, 2010)

# Fixed-sequence procedure

## Decision rules


$\alpha = 0.05$, Two-sided familywise error rate

1. Test $H_0$ at 0.05
2. Test $H_+$ at 0.05 only if $H_0$ is rejected

Logical relationships are not taken into account (effect in marker-positive population can be tested only after effect is established in overall population)

# Nonparametric chain procedures

$\alpha$ **allocation rule**

$\alpha w_0$ and $\alpha w_+$ are assigned to $H_0$ and $H_+$

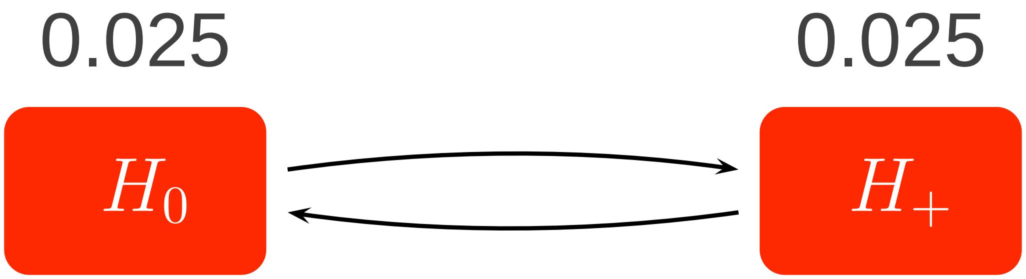
$w_0$ and $w_+$, non-negative weights with $w_0 + w_+ = 1$

Weights reflect relative importance of overall and marker-positive populations

$\alpha$ **propagation rule**

If $H_0$ is rejected, its significance level is transferred to $H_+$ and vice versa

## Decision rules

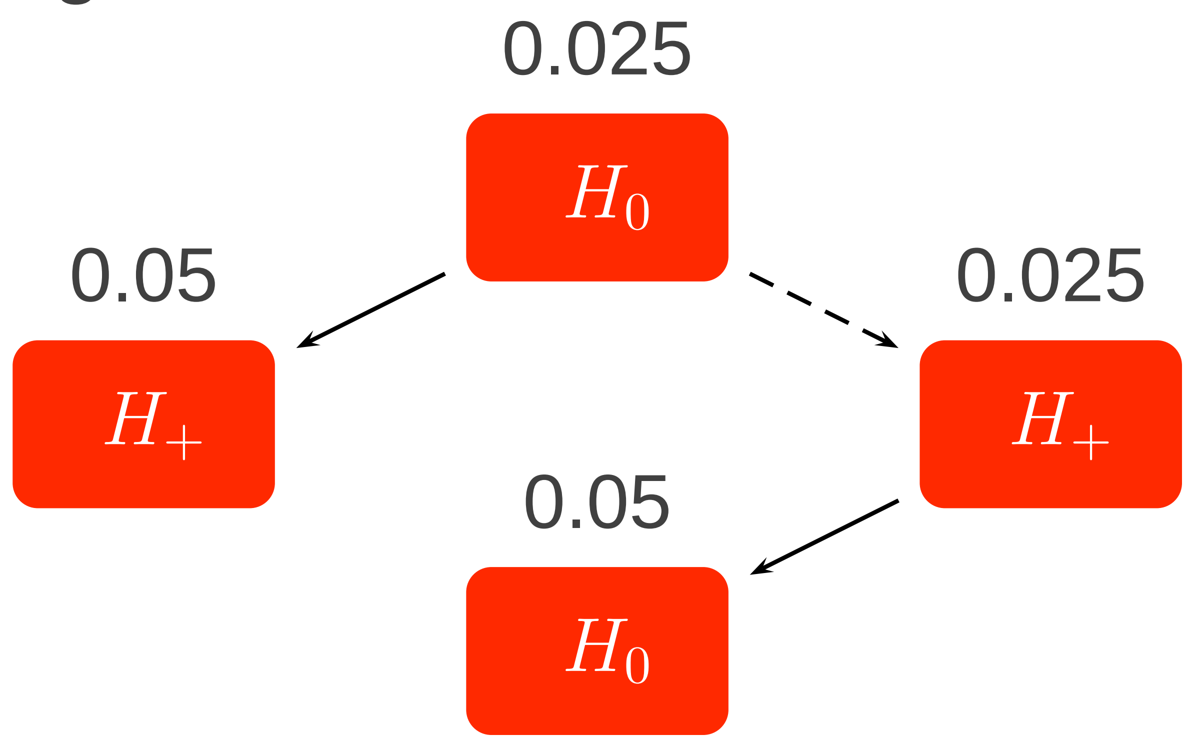



$w_0 = w_+ = 0.5$, Equally weighted analyses

Formal presentation of decision rules (not very useful)

# Nonparametric chain procedures

## Testing algorithm




—, Significant outcome

- -, Non-significant outcome

# Distributional information

## Correlation

Test statistics for $H_0$ and $H_+$ are generally strongly positively correlated

Correlation depends on the relative size of the marker-positive population

## Example

Correlation = 0.7 if 50% of patients are marker-positive ($n_+ = n_0/2$)

# Parametric chain procedures

$\alpha$ **allocation rule**

$\alpha w_0$ and $\alpha w_+$ are assigned to $H_0$ and $H_+$

$w_0$ and $w_+$, non-negative weights with $w_0 + w_+ = 1$

$\alpha$ **propagation rule**

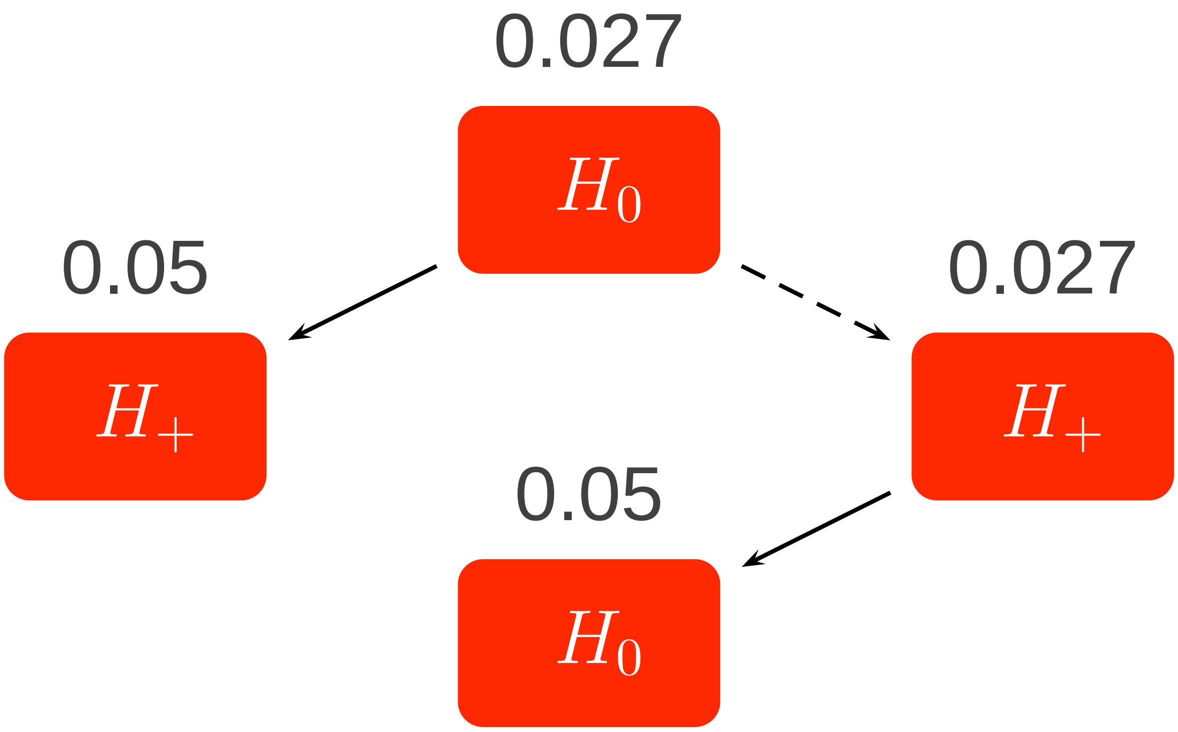
If $H_0$ is rejected, its significance level is transferred to $H_+$ and vice versa

**Distributional information**

Hypothesis test statistics follows a bivariate normal distribution

# Parametric chain procedures

## Testing algorithm




—, Significant outcome

- -, Non-significant outcome

# Selection of hypothesis weights

## Hypothesis weights

Hypothesis weights in confirmatory subgroup analysis must be pre-specified

## Hypothesis weight versus importance

Hypothesis weights in chain and other procedures are selected to help quantify hypothesis importance

# Selection of hypothesis weights

**Clinical trial example**

$n_+/n_0 = 0.2$, Population prevalence of marker-positive patients

Overall and target population tests are both powered at 80%

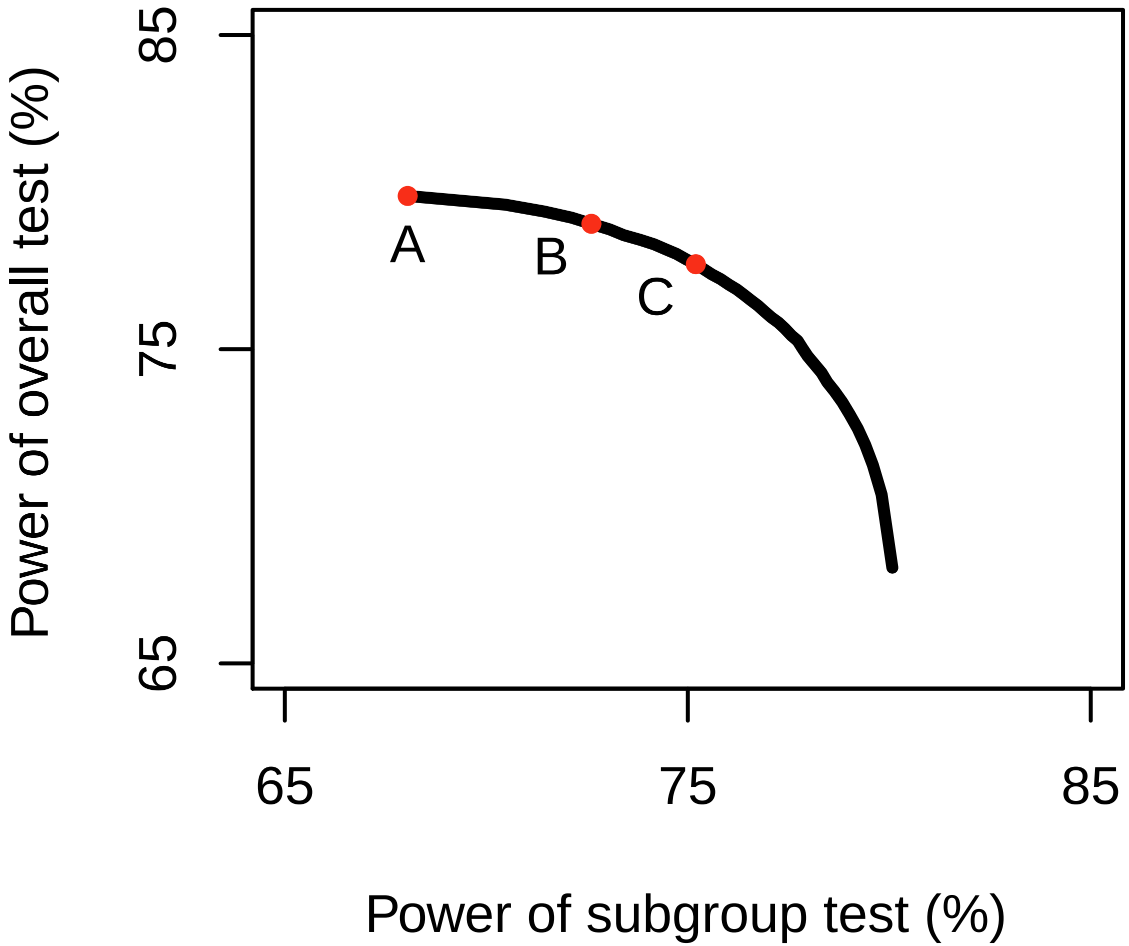
**Nonparametric chain procedure**

Procedure A: $w_0 = 1$, $w_+ = 0$ (fixed-sequence procedure)

Procedure B: $w_0 = 0.9$, $w_+ = 0.1$

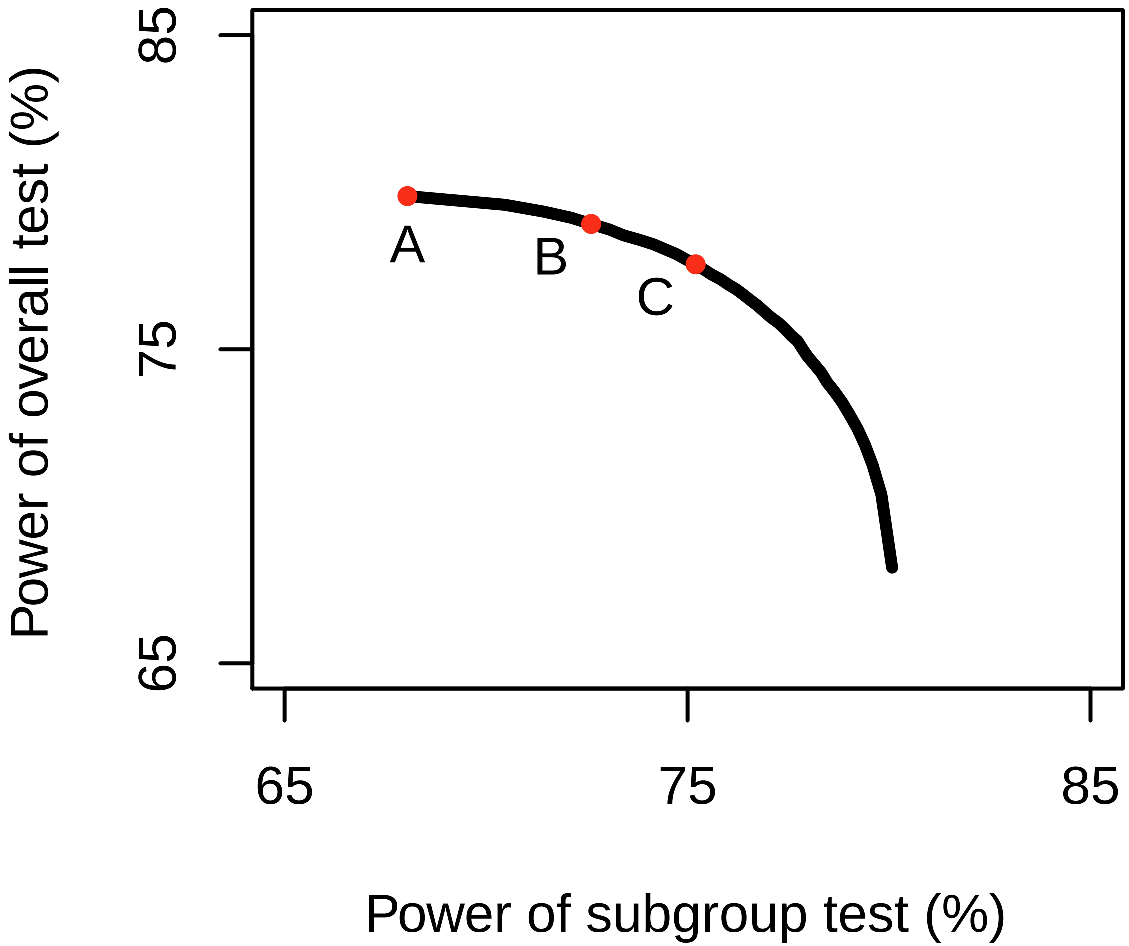
Procedure C: $w_0 = 0.7$, $w_+ = 0.3$

# Power comparison



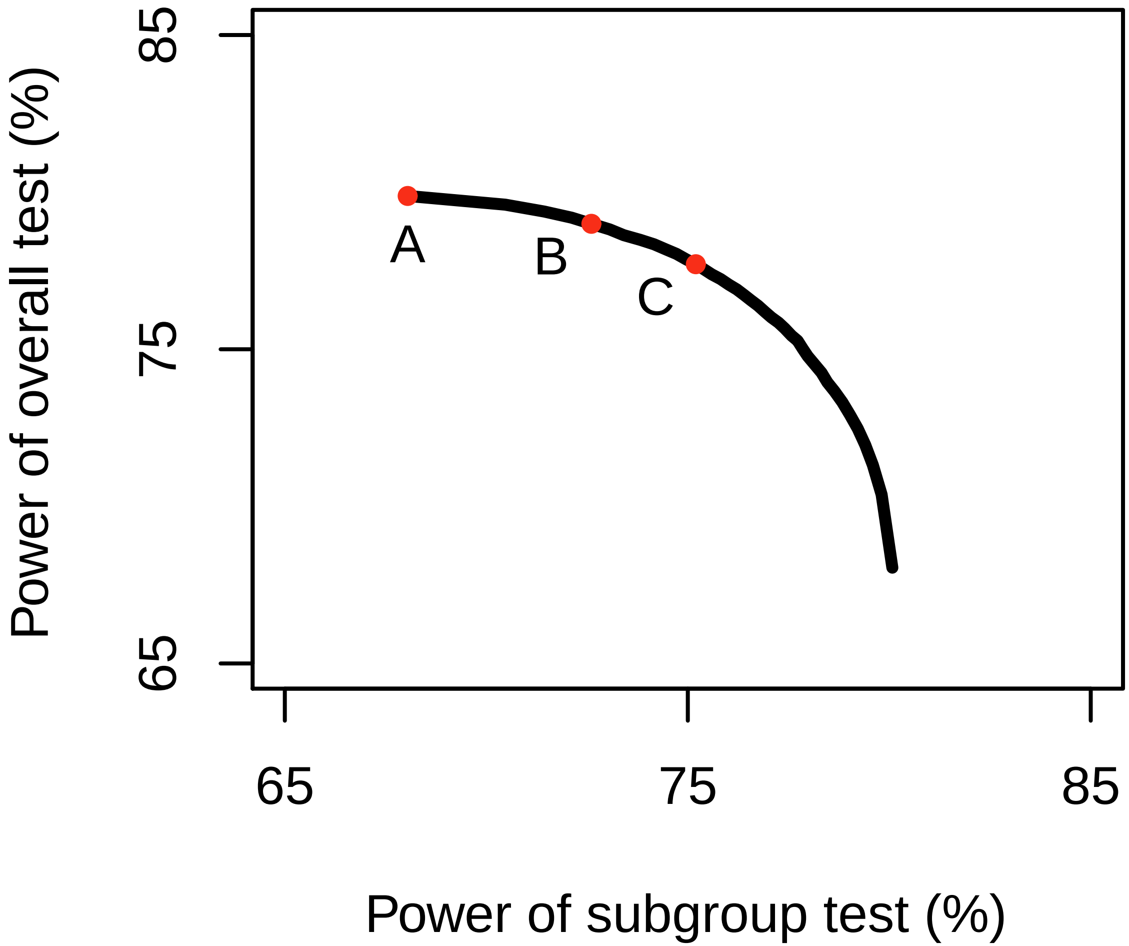

Procedure A: $w_0 = 1$, $w_+ = 0$; Procedure B: $w_0 = 0.9$, $w_+ = 0.1$; Procedure C: $w_0 = 0.7$, $w_+ = 0.3$.

# Power comparison




Procedure A: Overall test 80%, Subgroup test 68%.
Procedure B: Overall test 79%, Subgroup test 73%.

# Power comparison




Procedure B: Overall test 79%, Subgroup test 73%.
Procedure C: Overall test 77%, Subgroup test 75%.

# Optimality criteria

**Disjunctive power**

Probability of rejecting at least one of the hypotheses

$\psi = P\{\text{Reject } H_0 \text{ or } H_+\}$

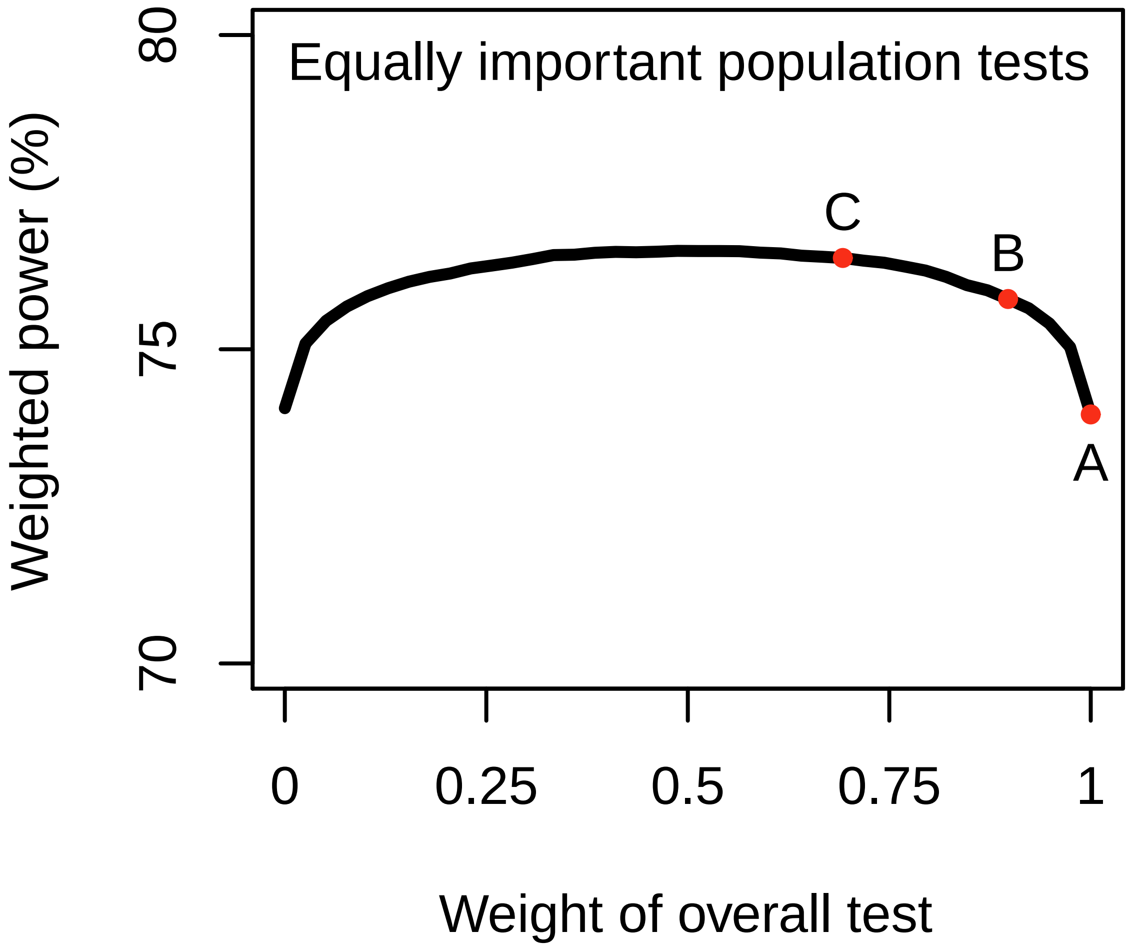
**Weighted power**

Weighted sum of marginal power functions

$\psi = v_0 P\{\text{Reject } H_0\} + v_+ P\{\text{Reject } H_+\}$

$v_0$ and $v_+$, Hypothesis importance ($v_0 > 0$, $v_+ > 0$ and $v_0 + v_+ = 1$)

# Weighted power criterion




Procedure A: $w_0 = 1$, $w_+ = 0$; Procedure B: $w_0 = 0.9$, $w_+ = 0.1$; Procedure C: $w_0 = 0.7$, $w_+ = 0.3$.

# Summary

**Account for available information**

Clinical information: Account for relevant logical relationships between the population tests

Statistical information: Account for positive correlation between the population tests

**Quantitative evaluation and comparison**

Comprehensive quantitative evaluation of candidate multiplicity adjustment procedures to maximize power and select an optimal set of procedure parameters

# References

Alosh, M., Huque, M. (2009). A flexible strategy for testing subgroups and overall population. *Statistics in Medicine*. 28, 3-23.

Brugger et al. (2011). Prospective molecular marker analyses of EGFR and KRAS from a randomized, placebo-controlled study of erlotinib maintenance therapy in advanced nonsmall-cell lung cancer. *Journal of Clinical Oncology*. 29, 4113-4120.

Bretz, F., Maurer, W., Brannath, W., Posch, M. (2009). A graphical approach to sequentially rejective multiple test procedures. *Statistics in Medicine*. 28, 586-604.

# References

Dmitrienko, A., Millen, B.A., Brechenmacher, T., Paux, G. (2011). Development of gatekeeping strategies in confirmatory clinical trials. *Biometrical Journal*. 53, 875-893.

Hegi et al. (2005). MGMT Gene Silencing and Benefit from Temozolomide in Glioblastoma. *New England Journal of Medicine*. 352, 997-1003.

Huque, M., Alosh, M. (2008). A flexible fixed-sequence testing method for hierarchically ordered correlated multiple endpoints in clinical trials. *Journal of Statistical Planning and Inference*. 138, 321-335.

# References

Millen, B.A., Dmitrienko, A. (2011). Chain procedures: A class of flexible closed testing procedures with clinical trial applications. *Statistics in Biopharmaceutical Research*. 3, 14-30.

Millen, B.A., Dmitrienko, A., Ruberg, S., Shen, L. (2012). Statistical considerations in clinical trials with tailoring objectives. *Drug Information Journal*. In press.

Zhao, Y.D., Dmitrienko, A., Tamura, R. (2010). Design and analysis considerations in clinical trials with a sensitive subpopulation. *Statistics in Biopharmaceutical Research*. 2, 72-83.