Adaptive Group Sequential Designs

Werner Brannath

VO ”Sequential and Adaptive Designs”,
University Bremen

Introduction

Classical frequentist's trials – Issues for planning

Details of design and analysis must be fixed in advance:

- **Population**
- **Treatments (doses)**
- **Main outcome variable(s)**
- **Secondary outcome variable(s)**
- **Analysis strategy**
- **Number of stages and sample sizes**
- **Rejection and acceptance boundaries**
- **...**

Classical frequentist's trials

- We often need to deal with new information emerging from inside or outside the trial.
- Lack of flexibility in design features is well known and has been noticed early.

Examples of older approaches for introducing flexibility:

- *Error spending function approach* (Lan & DeMets 1983): accounts for (independent) random group sizes.
- *Repeated confidence intervals* (Jennison & Turnbull 1984, 1989): leads to an exact inference also after deviating from the pre-specified stopping criteria.

Pre-specified Adaptivity versus Flexibility

Pre-specified adaptivity =

adapting design parameters according to a pre-specified adaptation rule

Aims: Increasing efficiency by optimizing specific cost functions.

Examples: Group sequential trials, play-the-winner allocation rules, multi-armed group sequential designs with treatment selection, . . .

Flexibility (unscheduled adaptivity) =

adapting design parameters without a (complete) specification of the adaptation rule

Examples: Combination tests, conditional error function and conditional rejection probability approach, self-designs designs, . . .

Why flexible designs?

Aims of flexibility:

- Dealing with the *unexpected* (protocol amendments).
- Dealing with *expected unpredictability* (sample size reassessments, treatment or sub-group selection).
- Improving the “quality” of the decision process as a whole in an environment where the parameter assumptions and also the *weighting of gains and costs are unclear* a priori and can change in the course of the trial.
- Dealing with complex adaptation rules.

Data-driven sample size adaptations

- Two-stage design **without** early rejection or acceptance,

  but with data-driven sample size reassessment.

- How large can the type I error become?

- Proschan & Hunsberger (1995):

  With a one-sided z-test at level α the maximum type I error is

$$\alpha_{\max} = \alpha + \frac{1}{4} e^{-z_\alpha^2/2}$$

- Numerical examples:

  $\alpha = 0.05 \quad \longrightarrow \quad \alpha_{\max} = 0.1146$

  $\alpha = 0.025 \quad \longrightarrow \quad \alpha_{\max} = 0.0616$

Comments on data-driven sample size adaptations

- With data-driven sample size adaptations the **type I error rate** can be **more than doubled**.
- It is larger than with repeated testing (and Bonferroni), because $n_2 = 0$ is a possible sample size choice.
- One observes type I error inflations also when restricting n_2 to some pre-specified $n_{min} < n_{max}$; with restrictions on the sample size, the inflation is smaller.
- Can we control the type I error rate also with adaptive sample size calculations?

Flexible Two Stage Tests

Step-wise procedure

Stage 1 (e.g. Phase II part) and *Stage 2* (e.g. Phase III part)

Stage 1 and Stage 2 data are from two independent cohorts.

Adaptivity

The design of Stage 2 (sample sizes, statistical test, . . .) is chosen based on the data of Stage 1 as well as any other internal or external information.

Flexibility

For a control of the type I error rate, we do not want to pre-specify how the Stage 1 data determine the design of Stage 2.

Combination Tests

General idea

- Calculate stage-wise p-values

  $p =$ p-value (e.g. from z-test) of first n_1 patients (stage 1)

  $q =$ p-value (e.g. from z-test) of second n_2 patients (stage 2)

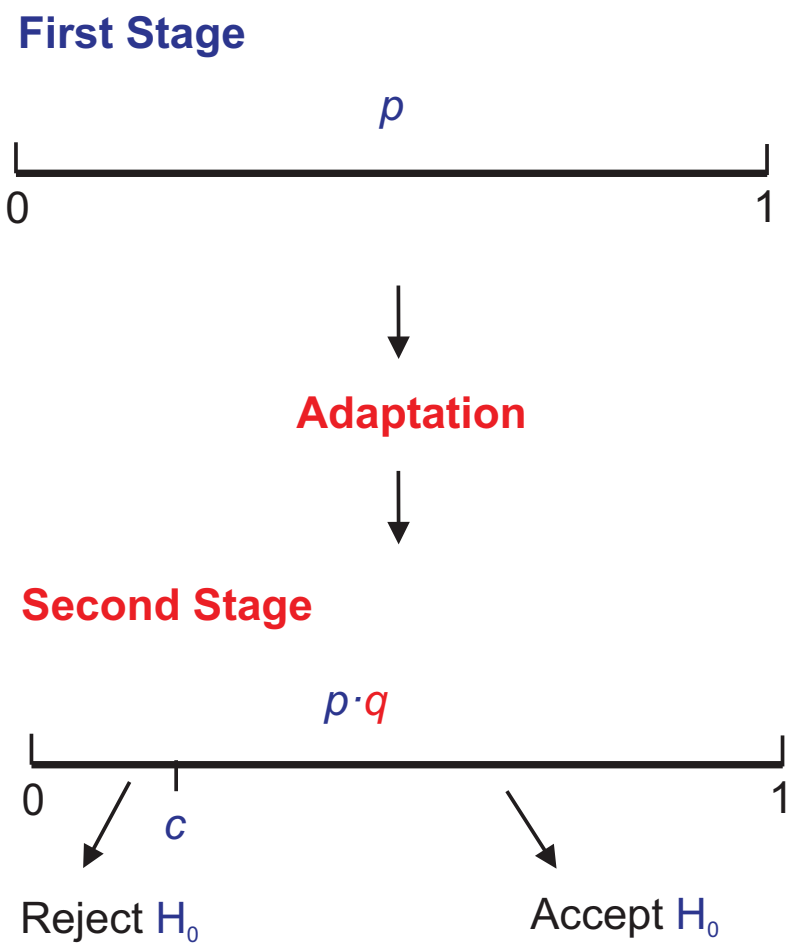
- At stage 2 combine the stage-wise p-values p and q by a pre-specified function ("combination function").
- Compare this with to a pre-specified critical value.

  $\longrightarrow$ Pre-specified critical region in (p, q)-plane

- Control of type I error rate possible, since p and q are independent and on $[0, 1]$ uniformly distributed under H_0.

Fisher's product test

(Bauer 1989, Bauer & Köhne 1994, . . .)



Planning:

- Fix design for stage 1 (sample sizes, test, . . .)

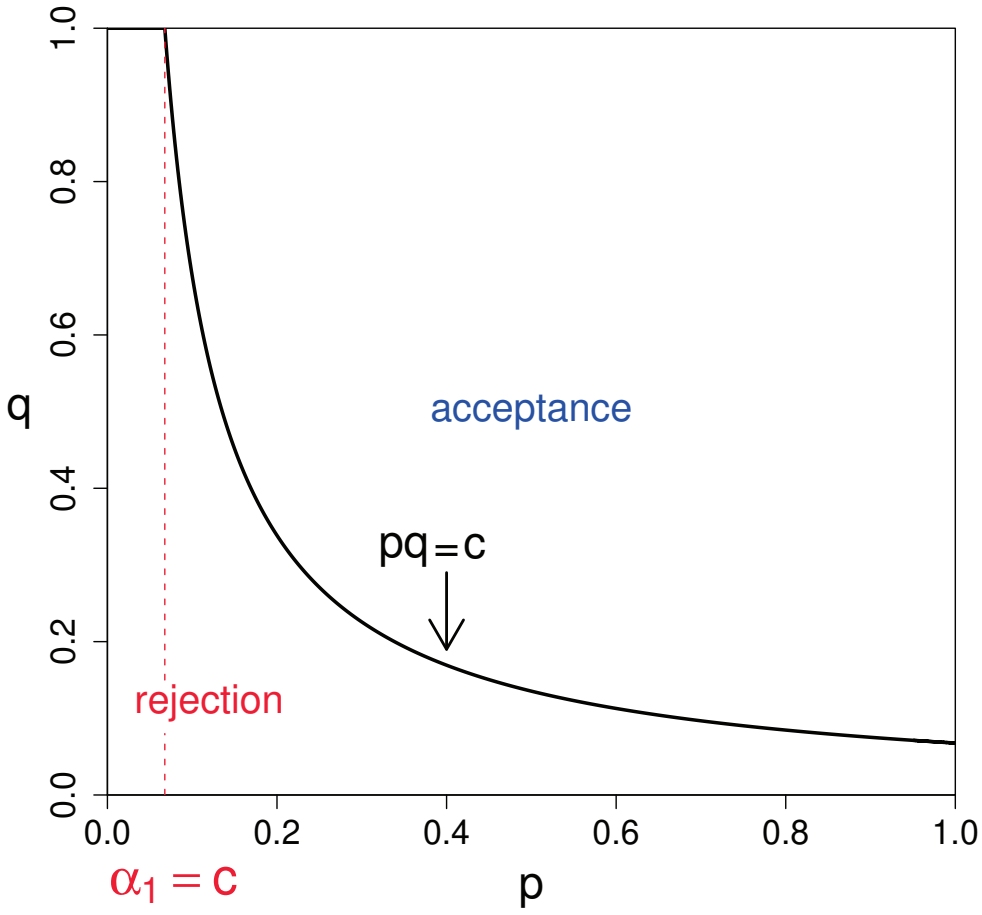
Stage 1:

- Compute p-value p from Stage-1-data
- Fix design for stage 2 based on data from stage 1

Stage 2:

- Compute p-value q from stage-2-data.
- Reject H_0 iff $pq \leq c_\alpha = e^{-\chi^{-2}_{4,1-\alpha}/2}$.
  E.g.
  $c_\alpha = 0.0038$ for $\alpha = 0.025$

Fisher's product test (FISHER 1931, BAUER 1989)



Test $H_0 : \mu_E - \mu_C \leq 0$

p . . . p-value from t-test with stage-1-data

q . . . p-value from t-test with stage-2-data

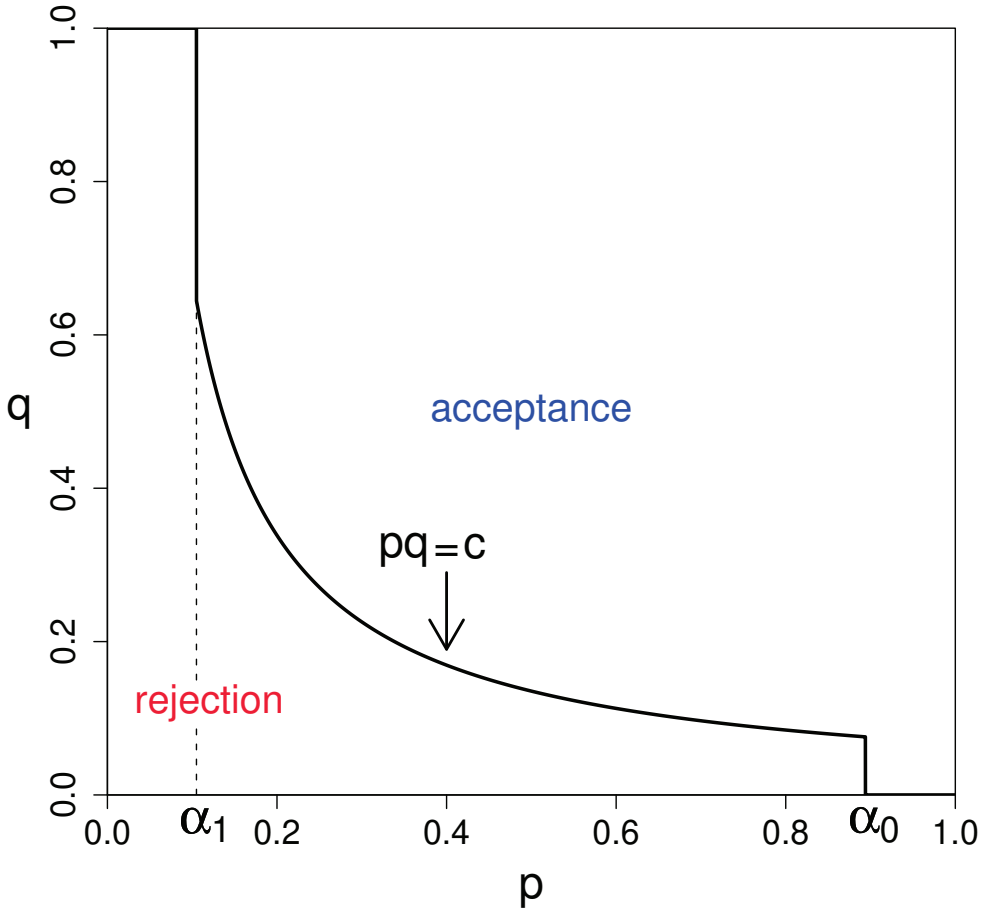
Rejection if $pq \leq e^{-\chi^{-2}_{4,1-\alpha}/2}$

Type I error rate α, since

$-2\log(pq) =$
$-2\log(p) - 2\log(q) \sim \chi^2_4$

Non-stochastic curtailment: $p \leq c \Rightarrow p \cdot q \leq c$ for all $q \leq 1$

Fisher's product test with early rej. and acceptance



Let $H_0 : \Delta = \mu_E - \mu_C \leq 0$
and $\alpha_1, \alpha_2 < \alpha < \alpha_0$

Stage 1:

Calculate p-value p

Reject H_0 if $p \leq \alpha_1$

Accept H_0 if $p > \alpha_0$

Stage 2

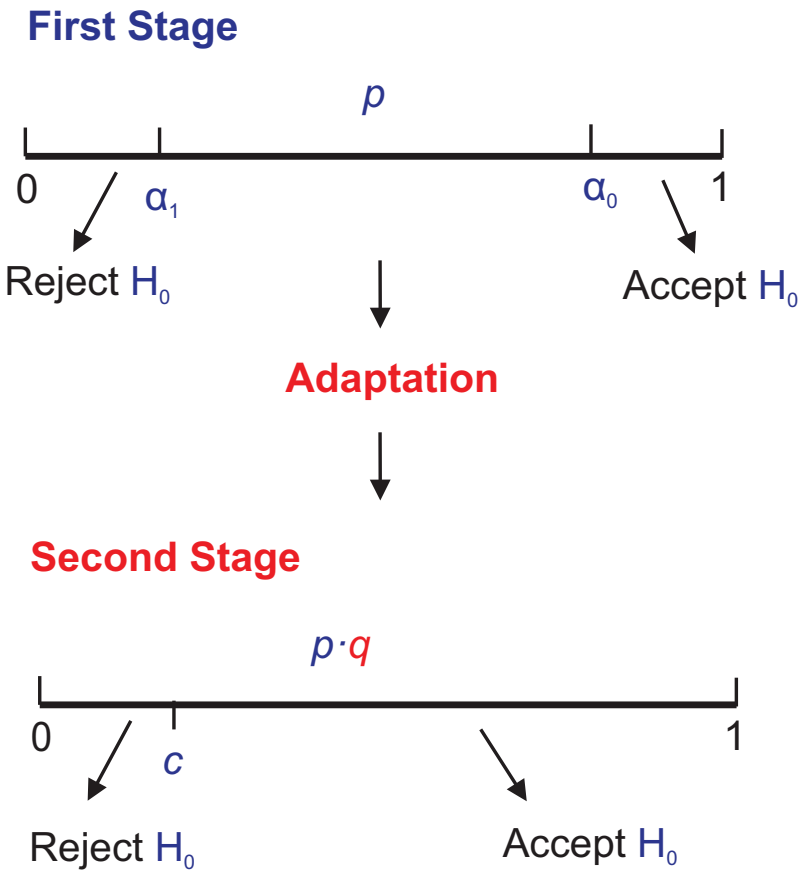
Calculate p-value q

Reject H_0 if

$$pq \leq e^{-\chi^{-2}_{4,1-\alpha_2}/2}$$

Fisher's product test with early rej. and acceptance

(BAUER '89, BAUER & KÖHNE '94)



Planning:

- Fix test, stage 1 sample sizes and $\alpha_1, \alpha_2 < \alpha < \alpha_0$

Stage 1:

- Compute p from Stage 1 data
- Stop and reject, if $p \leq \alpha_1$
- Stop and accept, if $p > \alpha_0$
- Else, fix design for Stage 2 based on data from Stage 1

Stage 2:

- Compute q from Stage 2 data.
- Reject H_0 iff $pq \leq c = c_{\alpha_2}$.

Choice of critical values - 1) full second stage level

BAUER & KÖHNE (1995)

- Choose $\alpha_2 = \alpha$, i.e. critical value $c_\alpha = e^{-\chi^{-2}_{4,1-\alpha_2}/2}$ and $\alpha_0 < 1$.
- Determine α_1 such that

$$\mathbf{P}_{\Delta=0}\left(p_1 \leq \alpha_1\right) + \mathbf{P}_{\Delta=0}\left(\alpha_1 < p_1 \leq \alpha_0,\ pq \leq c_\alpha\right) = \alpha$$

- Type I error rate calculation:

$$\alpha = \mathbf{P}_{\Delta=0}\left(p_1 \leq \alpha_1\right) + \mathbf{P}_{\Delta=0}\left(\alpha_1 < p_1 \leq \alpha_0,\ pq \leq c_\alpha\right)$$

$$= \alpha_1 + \int_{\alpha_1}^{\alpha_0} \int_0^1 \mathbf{1}_{\{pq \leq c_\alpha\}}\, dp\, dq = \alpha_1 + \int_{\alpha_1}^{\alpha_0} \left(\frac{c_\alpha}{p}\right) dp$$

$$= \alpha_1 + c_\alpha \Big[\ln(\alpha_0) - \ln(\alpha_1)\Big]$$

- See Table 6.1 in WaBr2016 for example of critical values α_0 and α_1.

Other choices for critical values

*2) Equal local rejection levels (*BAUER, 1989; WASSMER, 1999*)*:

- Fix $\alpha_0 < 1$ and $\alpha_1 = \alpha_2 = \alpha^* < \alpha$ such that the type I error rate

$$\alpha^* + c_\alpha \Big[\ln(\alpha_0) - \ln(\alpha^*) \Big] = \alpha$$

- See Table 6.2. in WaBr2016

*3) Choice of α, α_1 and α_0 (*BAUER & RÖHMEL, 1995 AND GEN. IN WABR216*):*

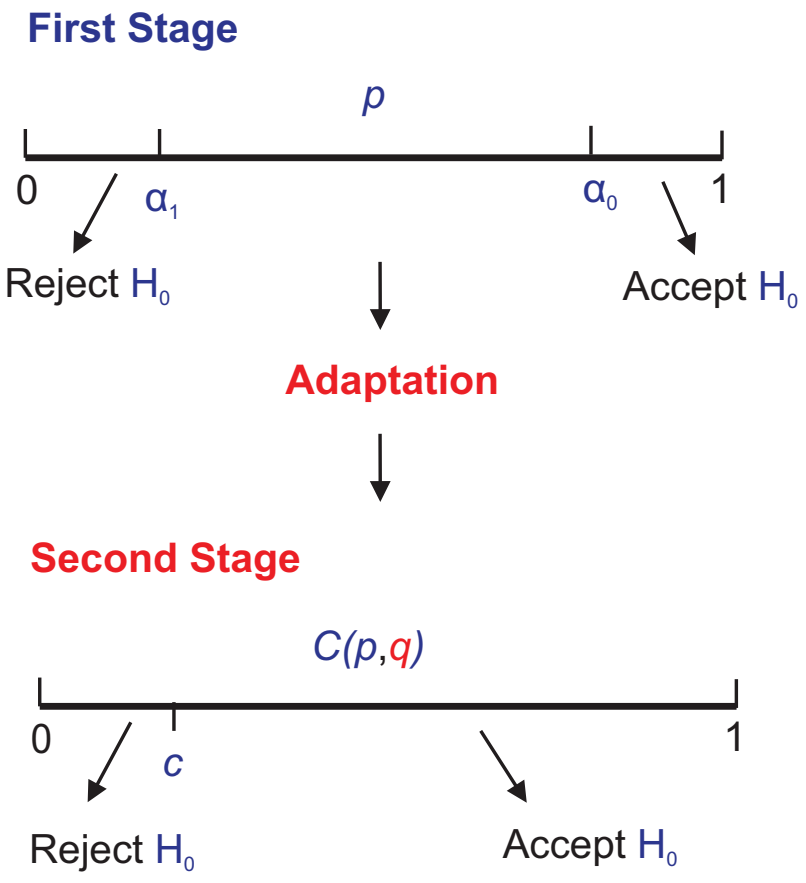
- We fix α, α_1 and α_0 and calculate the critical value c as

$$c = \frac{\alpha - \alpha_1}{\ln(\alpha_0) - \ln(\alpha_1)}$$

- Non-stochastic curtailment:

$$\alpha_1 > c \quad \Longleftrightarrow \quad \alpha_1 + \alpha_1 \big(\ln(\alpha_0) - \ln(\alpha_1) \big) \geq \alpha$$

Combination Tests (Bauer '89, Bauer & Köhne '94)



Planning:

- Fix design for Stage 1 (sample sizes, test, α_1, α_0)
- Fix a monotone combination function $C(p, q)$ and critical value c.

Stage 1:

- Compute p-value p from Stage 1 data
- Either stop or fix design for Stage 2 based on data from Stage 1

Stage 2:

- Compute p-value q from Stage 2 data.
- Reject H_0 iff $C(p, q) \leq c$.

Inverse Normal Combination Test

Inverse normal combination function

(Lehmacher & Wassmer '99; Cui, Hung & Wang, '99)

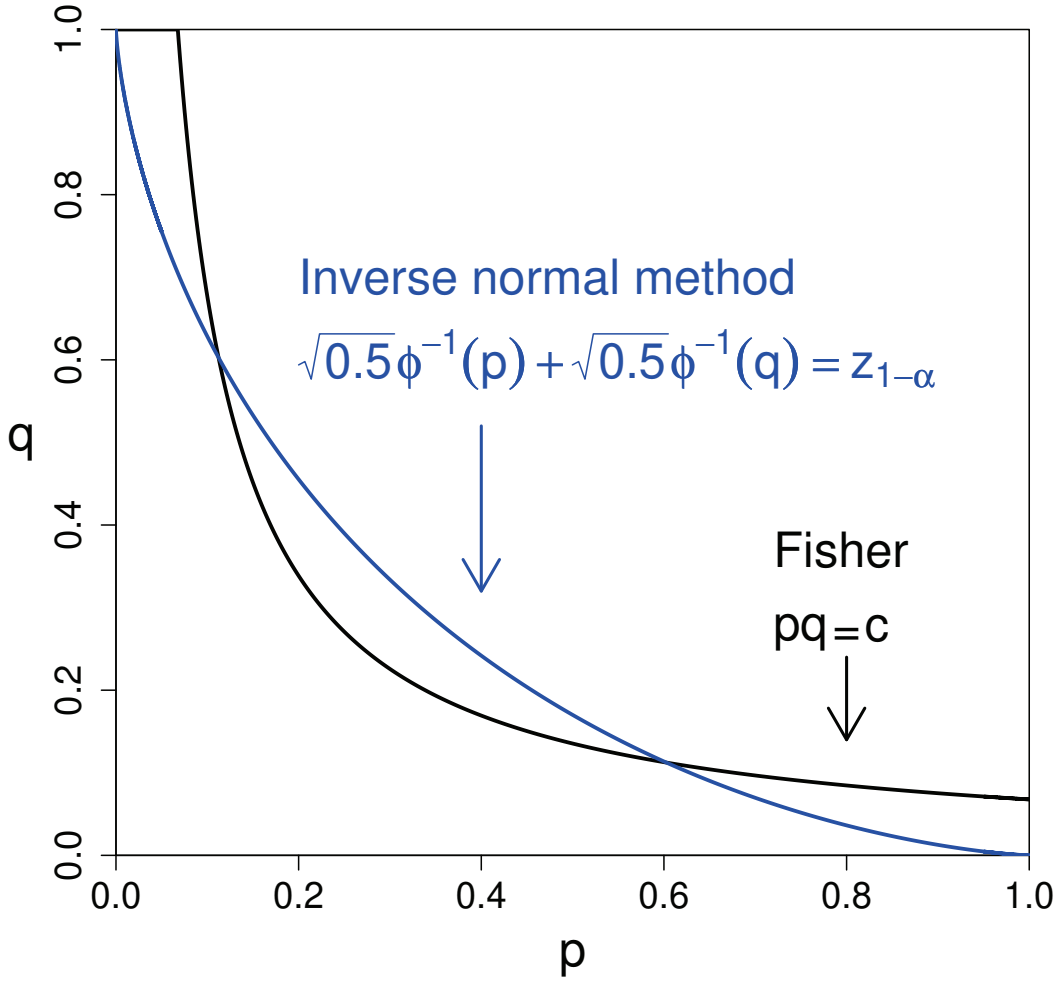
Use of the combination function:

$$C(p,q) = 1 - \Phi\left(\sqrt{0.5}\,\underbrace{\Phi^{-1}(1-p)}_{Z_1} + \sqrt{0.5}\,\underbrace{\Phi^{-1}(1-q)}_{Z_2}\right)$$

We have that

- $Z_1 = \Phi^{-1}(1-p) \sim N(0,1)$ and $Z_2 = \Phi^{-1}(1-q) \sim N(0,1)$
- Z_1 and Z_2 are independent and standard normal.
- Therefore: $Z_2^* = \sqrt{0.5}\,Z_1 + \sqrt{0.5}\,Z_2 \sim N(0,1)$
  ("weighted z-score")
- $C(p,q) = 1 - \Phi(Z)$ is uniformly distributed under H_0.

Comparison to Fisher's product test ($\alpha_0 = 1$)



- For large *p*:
  easier to reject with Fisher test
- For moderate *p*:
  easier to reject with inverse normal method
- **No** non-stochastic curtailment with inverse normal method.
- Possible to incorporate stopping rules also for inverse normal method

Inverse normal combination function

(LEHMACHER & WASSMER '99; CUI, HUNG & WANG, '99)

- When $\alpha_1 = 0$ and $\alpha_0 = 1$ then

$$C(p, q) \leq \alpha \quad \Longleftrightarrow \quad Z \geq \Phi^{-1}(1 - \alpha)$$

  is a combination test at level α.

- Z_1, Z_2^* distributed as in GSD: $Cov(Z_1, Z_2^*) = \sqrt{0.5}$
- We can use rejection and acceptance boundaries, $u_k^0 < u_k^1$, of any equally spaced GSD; gives local levels:

$$\alpha_1 = 1 - \Phi(u_1^1), \quad \alpha_0 = 1 - \Phi(u_1^0), \quad c = 1 - \Phi(u_2^1)$$

Weighted inverse normal method

(LEHMACHER & WASSMER 1999)

Prefix $0 \leq w_1, w_2 \leq 1$ with $w_1^2 + w_2^2 = 1$ and use the combination function:

$$C(p,q) = 1 - \Phi\left(w_1 \underbrace{\Phi^{-1}(1-p)}_{Z_1} + w_2 \underbrace{\Phi^{-1}(1-q)}_{Z_2} \right)$$

This implies

- $Z_2^* = w_1 Z_1 + w_2 Z_2 \sim N(0,1)$ with $Cov(Z_1, Z_2^*) = w_1$
- Distribution as in GSD with interim information time $t = w_1^2$.
- We can use local levels from any GSD with $t_1 = w_1^2$.
- This adaptive GSD is also called “weighted z-score test” (Cui et al., 1999) and can be extended to designs with $K > 2$ stages.

Example

Initial Plan: O'Brien and Fleming GSD with sample sizes $n_1 = 300$ and $n = 470$ ($t_1 = 300/470$) and no futility bound.

$\longrightarrow$ rejection boundaries: $u_1 = 2.5$ and $u_2 = 2.0$

To permit sample size adaptations, we use inverse normal combination test (or equivalently weighted z-score test) with weights

$$w_1 = \sqrt{300/470} = .80 \quad \text{and} \quad w_2 = \sqrt{170/470} = .60$$

and rejection levels

$$\alpha_1 = 1 - \Phi(u_1) = 0.006 \quad \text{and} \quad c = 1 - \Phi(u_2) = 0.023$$

Extending a GSD to an Adaptive Design (Slide 1)

- Start planning a GSD $\quad\rightarrow\quad n_k,\, u_k^0,\, u_1^k, \quad k = 1, \ldots, K.$
- Calculate $\quad \alpha_k^1 = 1 - \Phi(u_k^1)$ and $\alpha_k^0 = 1 - \Phi(u_k^0)$.

- Assume that at stage $k < K$ we want extend the GSD by un-blinded *sample size adaptations:*

  We change the sample size(s) for the next stages, $j > k$, to $\tilde{n}_j$ based on the available un-blinded data.

- At stages $j > k$, calculate: $\quad Z_j^* = \underbrace{\sqrt{\frac{n_1}{N_j}}}_{=w_{k1}} Z_1 + \cdots + \underbrace{\sqrt{\frac{n_j}{N_j}}}_{=w_{jj}} Z_j$

  where w_{ij} is defined by the pre-planned sample sizes and

$$Z_i := \sqrt{\tilde{n}_i}(\bar{X}_i - \mu_0)/\sigma \qquad \sim_{\mu=\mu_0} N(0,1)$$

Extending a GSD to an Adaptive Design (Slide 2)

- Stop and reject (accept) at stage j if $Z_j^* \geq u_j^1$ ($Z_j^* < u_j^0$).
- We can change the sample size $\tilde{n}_j$ at each previous stage.
- We have extended the GSD to an adaptive design:

  - If $\tilde{n}_j = n_j$ for all j, then we perform the pre-planned GSD,

  - otherwise, we follow corresp. weighted z-score (inverse normal combination) test.

Level control with two-stage combination tests

Type I error control in two-stage combination tests

Level Condition:
If we choose α_0, α_1 and the critical value c such that

$$\alpha_1 + P[\,\alpha_1 < p \leq \alpha_0,\ C(p,q) \leq c\,] = \alpha$$

for **independent and uniformly distributed p-values** p **and** q then the type I error rate is at most α.

Example (Fisher' product test): If $C(p,q) = p \cdot q$ then the critical values α_0, α_1 and c must satisfy:

$$\alpha_1 \geq c \quad \text{and} \quad \alpha_1 + c \cdot \{\log(\alpha_0) - \log(\alpha_1)\} = \alpha$$

More combination tests

- Weighted Fisher's product test:

$$C(p,q) = p^w q \leq c, \quad \alpha_1 + c\,(\alpha_0^{1-w} - \alpha_1^{1-w})/(1-w) = \alpha$$

- Sum of p-values (Chung '97) with non-stochastic curtailment condition $\alpha_0 < c$:

$$C(p,q) = p + q \leq c, \quad \alpha_1 + c\,(\alpha_0 - \alpha_1) - (\alpha_0^2 - \alpha_1^2)/2 = \alpha$$

  Chung considers also the weighted sum of p-values.

- Modified Simes' Test (Hommel et al., 2005)

$$C(p,q) = q \leq c \qquad \alpha_1 + c\,(\alpha_0 - \alpha_1) = \alpha$$

Independence of the p-values

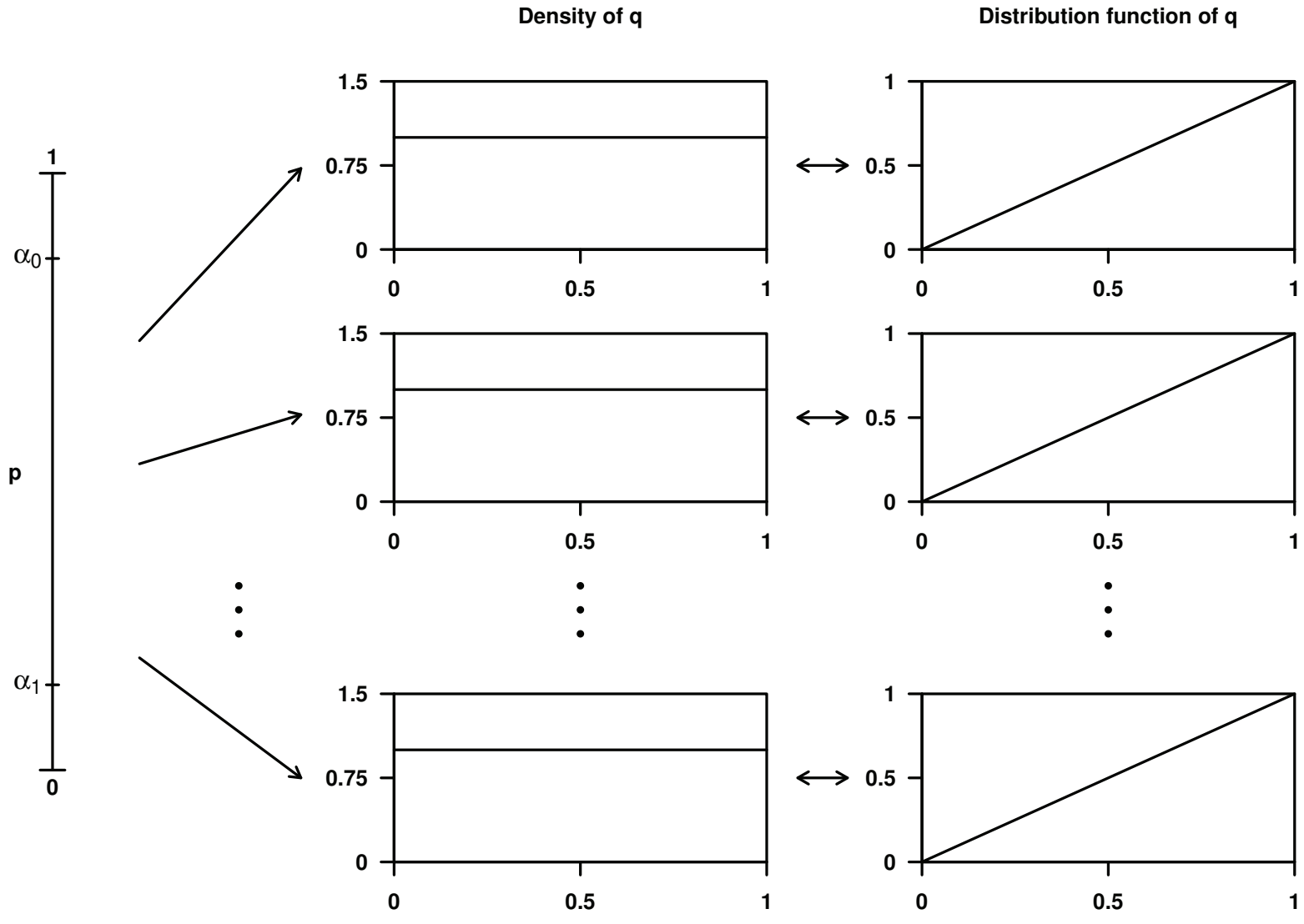
When are p and q independent and uniformly distributed?

This is the case if:

- p is uniformly distributed under H_0,
- q is computed from an *independent* second stage cohort,
- q is uniformly distributed under H_0 for the given second stage design.

Since, the conditional distribution of q given p is the same (namely uniform) for all $p \quad \Rightarrow \quad p$ and q are independent.

Independence of the p-values



The p-clud condition (BRANNATH ET AL., 2002)

Combination tests are conservative (have type I error rate $\leq \alpha$) under the following more general "p-clud" condition:

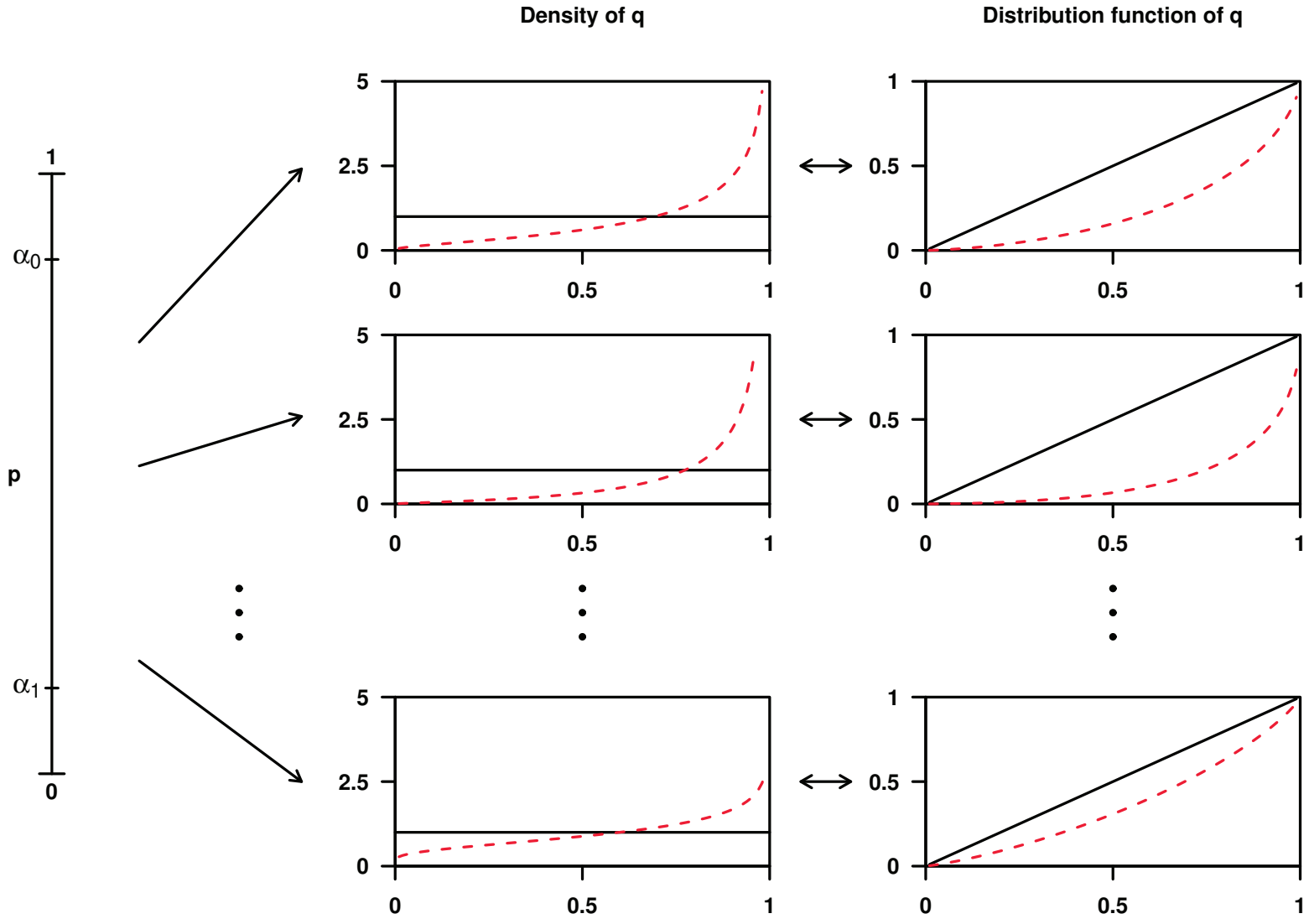
$$\mathbf{P}(\, q \leq u \,|\, p \,) \leq u \quad \text{for all } 0 \leq u \leq 1 \text{ and all } 0 \leq p \leq 1$$

This condition holds when

- q is computed from an *independent* second stage cohort,
- q is from conservative test for given second stage design.

The p-clud condition applies, for instance, to one-sided null hypotheses or discrete tests.

P-clud condition



Invariance Principle

Combination and weighted z-score tests follow a common, general **invariance principle** that guarantees type I error rate control.

- Use for the second stage a test statistic T_2 (e.g. q or Z_2) with a conditional distribution (given the interim data) that is equal to or dominated by a **fixed prototype distribution** (e.g. the uniform or normal dist.) which is **independent** from the adaptations.
- Define the rejection region in terms of the first test statistic (e.g. p or Z_1) and second stage test statistic T_2

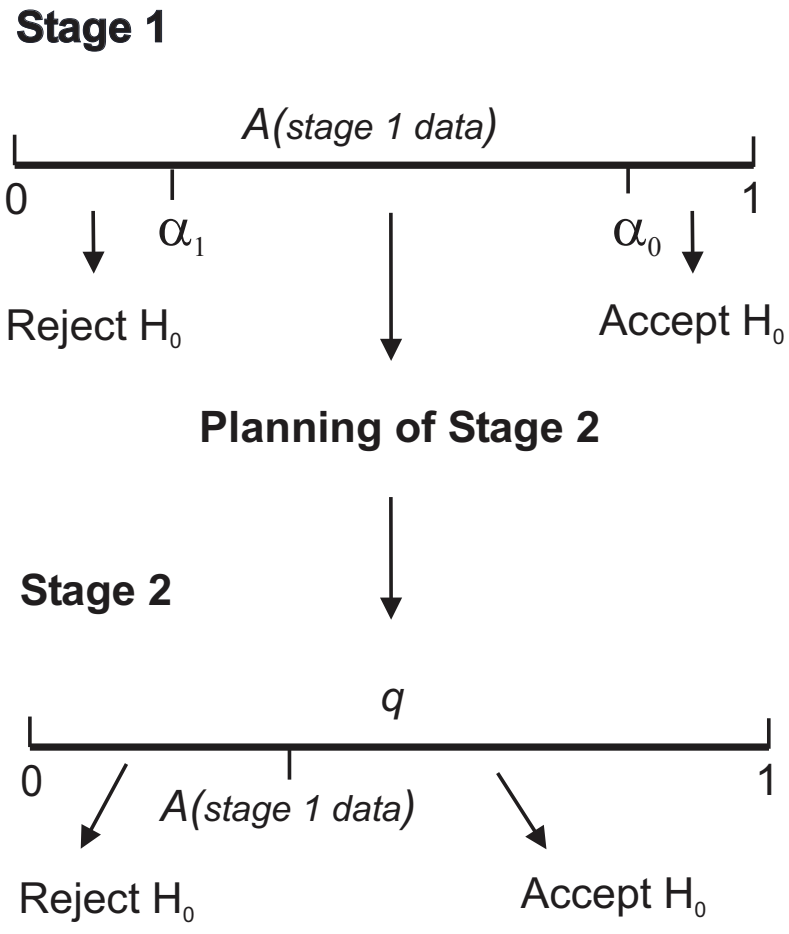
Summary on combination tests

- Combination tests permit data driven sample size adaptations.
- Sample size rule need **not** be (fully) known in advance.
- First stage, number of stages, rejection levels and combination function needs to be preplanned.
- Design of stage k must be specified latest at stage $k-1$.
- GSD can be extended to inverse normal combination tests.
- Fisher's product test can be extended to more than two stages.
- Extension to unblinded adaptations of number of stages possible (CRP, recursive combinations tests, self-designing trials)
- In practice almost always only one adaptive interim analysis.

Conditional Error Function Approach

Conditional error function principle

(PROSCHAN & HUNSBERGER 1995, MÜLLER & SCHÄFER 2001 . . .)



Planning:

- Fix design for Stage 1 (sample sizes, test, α_1, α_0)
- Fix *conditional error function* $0 \leq A(\text{stage 1 data}) \leq 1$ e.g. $A = c/p$

Stage 1:

- Compute A(stage 1 data).
- Either stop or fix design for Stage 2 based on data from Stage 1

Stage 2:

- Compute p-value q from Stage 2 data
- Reject H_0 iff $q \leq A$(stage 1 data).

Type I error control with conditional error function

Level Condition: Type I error rate $\leq \alpha$ if we choose $A(\text{stage 1 data})$ such that

$$\alpha_1 + E_{H_0}\left[A(\text{stage 1 data})\,\mathbf{1}_{\{\alpha_1 < p \leq \alpha_0\}}\right] = \alpha$$

Particular case: If $A = A(p)$ is a function of the first stage p-value, then α_0, α_1 and $A(p)$ must satisfy

$$\alpha_1 + \int_{\alpha_1}^{\alpha_0} A(p)\,dp = \alpha$$

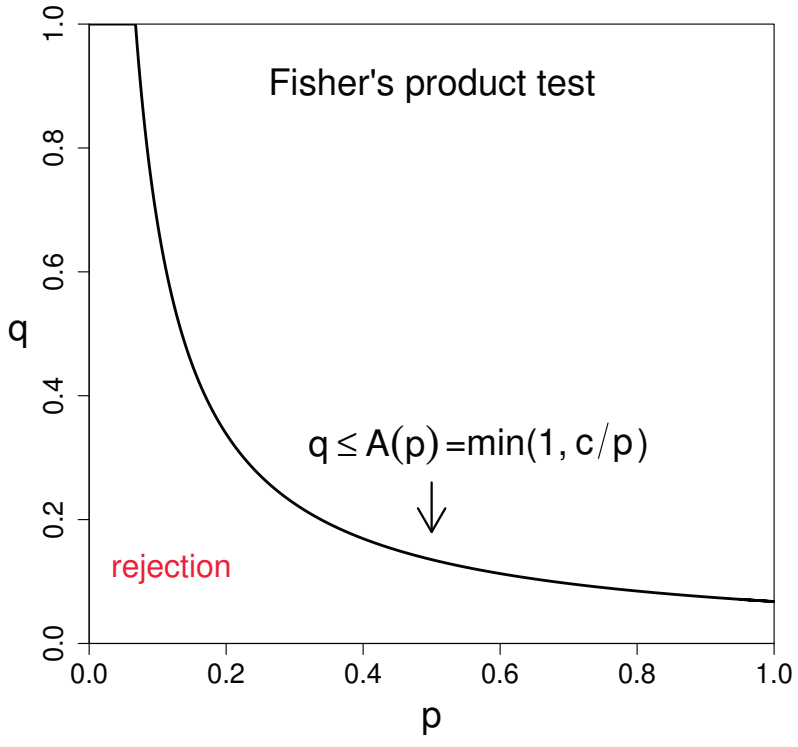
Again, (p, q) need to satisfy the p-clud property.

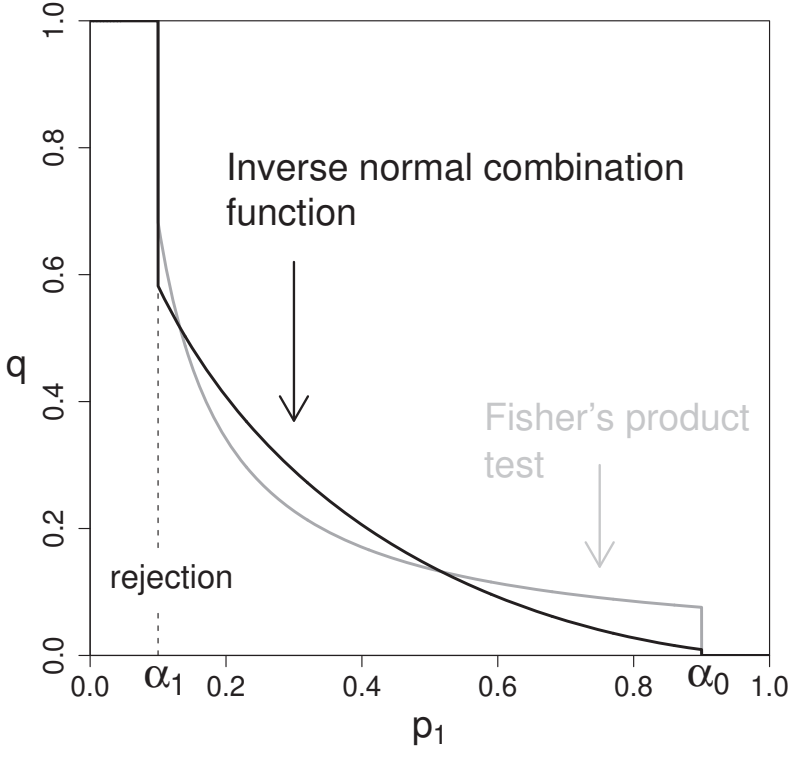
Important: $p \mapsto A(p)$ must be non-decreasing in p, because

- rejection should always be easier for smaller p;
- otherwise no type I error control with (general) p-clud property

Conditional error function of Fisher's test and inverse normal method

(WASSMER 1999, POSCH & BAUER 1999)





Inclusion of early decision boundaries

- We can include the early rejection and acceptance boundary into the conditional error function by the definition

$$\tilde{A}(p) = \begin{cases} 1 & \text{for } p \leq \alpha_1 \\ A(p) & \text{for } \alpha_1 < p \leq \alpha_0 \\ 0 & \text{for } p > \alpha_0. \end{cases}$$

- Since $q \leq 1$ holds always, we can stop and reject if $p \leq \alpha_1$.
- Since $q \leq 0$ never holds (only with prob. 0), we can stop and accept if $p > \alpha_0$.

- The level condition is now simply: $\int_0^1 A(p)dp$.

Combination test vs. conditional error function

(WASSMER 1999, POSCH & BAUER 1999)

Combination tests and conditional error functions $A = A(p)$ give a rejection region in the (p, q)-plane and hence are equivalent:

- Every combination test can be given in terms of the conditional error function
$$A(p) = \sup\{q : C(p, q) \leq c\}$$
- A test based on a conditional error function $A(p)$ can be rewritten as combination test, e.g., with the combination function
$$C(p, q) = q - A(p)$$
and critical values $c = 0$.
- Note that C is not uniquely determined by A.

Conditional error of inverse normal test

- As already mentioned

$$C(p,q) = 1 - \Phi\left(w_1\Phi^{-1}(1-p) + w_2\Phi^{-1}(1-q)\right) \leq c$$

is equivalent to

$$w_1\Phi^{-1}(1-p) + w_2\Phi^{-1}(1-q) \geq u_2$$

where $u_2 = \Phi^{-1}(1-c)$

- Solving $w_1\Phi^{-1}(1-p) + w_2\Phi^{-1}(1-q) = u_2$ for q gives

$$A(p) = \begin{cases} 1 & \text{for } p \leq \alpha_1 \\ 1 - \Phi\left(\frac{u_2 - w_1\Phi^{-1}(1-p)}{w_2}\right) & \text{for } \alpha_1 < p \leq \alpha_0 \\ 0 & \text{for } p > \alpha_0. \end{cases}$$

More on combination tests & conditional error fkts

- There is a one-to-one correspondence between the rejection region in the (p, q)-plane and A:

  The function A describes the boarder of the rejection region.

- By varying c, the combination function $C(p, q)$ defines many (namely, a family of) rejection regions

$$R_c = \{(p, q) : C(p, q) \leq c\}.$$

  By the choice of c we can construct combination test at different significance levels α. (**Exercise:** Try out numerically and plot)

Summary on the conditional error rate principle

- Is a powerful principle that has been suggested in parallel and independently from combination tests.
- Is essentially equivalent to the combination test approach.
- Has the advantage of the one-two-one correspondence with rejection rejections in the (p, q)-plane.
- Has triggered new approaches like e.g. the circular and optimal conditional error functions of Proshan & Hunsberger (1995) and Brannath & Bauer (2004), respectively.
- Combination tests have the advantage to provide a test statistics. This permits to directly define tests on different significance levels and p-values (see later).

Conditional Rejection Probabililty Principle

CRP-Principle of Müller and Schäfer ('01, '04)

- Start with a (classical) test design at level α (e.g. GSD).
- At an interim analysis review the data and possibly also external information.
- No reason to adapt $\rightarrow$ continue as pre-planned.
- Reason to adapt $\rightarrow$ compute the conditional type I error rate of the pre-defined design

$$A(\text{interim data}) = P_{H_0}(\text{reject } H_0 | \text{interim data})$$

- and choose (based on all the interim information) a new design with CRP equal to $A(\text{interim data})$ to finish the trial.

- Can be repeated $\rightarrow$ Design with flexible number of stages

Type I error rate control with CRP-Principle

- X_1 the interim data; X and $\widetilde{X}$ all data of initial design and AD.
- $\varphi(x)$ and $\tilde{\varphi}(\tilde{x})$ test decision functions of initial design and AD.
- Since the initial design has type I error rate α, we have

$$E_{H_0}\Big(A(X_1)\Big) = E_{H_0}\Big(E_{H_0}\big(\varphi(X)\big|X_1\big)\Big) = E_{H_0}\Big(\varphi(X)\Big) = \alpha$$

- Adaptive decision function $\tilde{\varphi}$ satisfies:

$$E_{H_0}\Big(\tilde{\varphi}(\widetilde{X})\Big|X_1 = x_1\Big) \leq A(x_1) \qquad \text{for all interim data } x_1$$

- This implies:

$$E_{H_0}\big(\tilde{\varphi}(X)\big) = E_{H_0}\Big(E_{H_0}(\tilde{\varphi}(\widetilde{X})|X_1)\Big) \leq E_{H_0}\Big(E_{H_0}(A(X_1)\Big) = \alpha$$

and hence type I error rate control.

Example (Slide 1)

- Start with a one-sided z-test at level $\alpha = 0.025$ and sample size $n = 500$.
- We look into the data after $n_1 = 250$ patients and compute first stage z-score $z_1 = 1.75$ ($p_1 = 0.04$).
- Safety profile so far promising and recruitment goes well.
- To increase power, we decide to increase the total sample size to $\tilde{n} = 750$.
- For the final analysis we calculate the CRP of the initial z-test:

$$A(z_1) := P_0\left(Z^* = w_1 Z_1 + w_2 Z_2 \geq z_\alpha \middle| Z_1 = z_1\right) = 1 - \Phi\left(\frac{z_\alpha - w_1 z_1}{w_2}\right)$$

where $w_1 = \sqrt{n_1/n} = \sqrt{0.5}$ and $w_2 = \sqrt{(n - n_1)/n} = \sqrt{0.5}$.

This gives $A(1.75) = 1 - \Phi\left(\left(1.96 - \sqrt{0.5} \cdot 1.75\right)/\sqrt{0.5}\right) = 0.15$.

Example (Slide 2)

- At the final analysis with $\tilde{n} = 750$, we calculate the second stage z-score Z_2 from the new $\tilde{n} - n_1 = 500$ pats and reject H_0 if

$$p_2 := 1 - \Phi(Z_2) \leq A(z_1).$$

  Otherwise, we accept H_0.
- This is the z-test with the second stage data (only) at level $A(z_1)$.
- Due to the p-clud property, this test has CRP $\leq A(z_1)$:

$$\mathbf{P}_0\left(p_2 \leq A(z_1) \,\middle|\, Z_1 = z_1\right) \leq A(z_1)$$

- **Exercise:** Show that this test is equivalent to the inverse normal combination test.

Example (Slide 3) - alternative approach

- At final analysis with $\tilde{n}$ patients we use the z-score of all pats

$$\widetilde{Z}^* = \tilde{w}_1 Z_1 + \tilde{w}_2 Z_2$$

where $\tilde{w}_1 = \sqrt{n_1/\tilde{n}}$ and $\tilde{w}_2 = \sqrt{(\tilde{n} - n_1)/\tilde{n}}$,

and adjust the second stage critical value u_2 such that

$$P_0\left(\widetilde{Z}^* \geq u_2 \,\middle|\, Z_1 = z_1\right) = A(z_1)$$

- This means, to compare $\widetilde{Z}^*$ to the (data dependent) critical value

$$u_2(z_1) = z_\alpha \frac{\tilde{w}_2}{w_2} + \left(\frac{\tilde{w}_1}{w_1} - \frac{\tilde{w}_2}{w_2}\right) w_1 z_1$$

- **Exercise:**
  Show that this is also equivalent to the inverse normal test.

Example (Slide 3)

- Assume now that we do only increase n to $\tilde{n} = 750$ but also want to add another IA after $n_2 = 500$ pats.
- This means that we finally will have two IA and one final analysis.
- We want the possibility for early rejection at the new IA

  with the same critical value $u_2 = u_3$ at the new IA and final analysis.
- This means to apply a GSD with (one-sided) Pocock boundaries.

Example (Slide 4)

There are two possibilities to apply the CRP principle:

1) Pan a GSD with at level $A(z_1)$ with the new data (only).

   This is a (classical) two-stage Pocock GSD with sample size $\tilde{n}_1 = 250$ and $\tilde{n}_2 = \tilde{n} = 500$ and level $A(z_1)$.

2) Plan a GSD with all data and conditional type I error rate (CRP) equal to $A(z_1) = 0.15$.

   This is an adaptive two-stage GSD with sample sizes $\tilde{n}_1 = 500$ and $\tilde{n}_2 = \tilde{n} = 750$.

   Choose identical, data dependent critical value $\tilde{u}_1 = \tilde{u}_2 = \tilde{u}$ s.th.

$$\mathbf{P}_0\left(\left\{Z_1^* \geq \tilde{u}\right\} \cup \left\{Z_2^* \geq \tilde{u}\right\} \middle| Z_1 = z_1\right) = A(z_1) = 0.15$$

   where Z_1^*, Z_2^* are the classical cumulative z-scores.

Example - exercise

- Implement the two approaches from the previous slide and compare their power in a simulation study.
- To this end consider different relative effect sizes δ between 0.10 and 0.15.

- *Hint for the calculation of the CRP of the new GSD:*

  Write each Z_j^* as weighted sum of the stage wise z-scores Z_i

  and bring in the inequality $Z_j^* \geq \tilde{u}$ the first stage z-score $Z_1 = z_1$ on the right side of the inequality.

  The CRP is the type I error rate of a GSD with rejection boundaries given by the right sides of the resulting inequalities.

Summary on the CRP principle

- Is very general approach for adaptations in a initially planned fixed size tests or GSD.
- Is very flexible and permits adaptations of the number of stages.
- Generalizes the weighted inverse normal method and coincides in simple cases (e.g. two-stage GSD with only sample size adaptation).
- Becomes complex in the presence of nuisance parameter (e.g. variance), when accounting form them exactly (e.g. t-test); see Posch et al., 2004; Timmesfeld et al., 2007; Gutjahr et al., 2011.
- The handling of nuisance parameter is much more simple with combination tests by the use of appropriate p-values.
- Recursive combination tests (Brannath et al., 2002) also permit adaptations of the number of stages.