The adaptive design: state of the art and paths for future research

Raphaël Porcher
Département de Biostatistique et Informatique Médicale
Hôpital Saint-Louis, Paris

SISMEC meeting
Milan, 2 February 2007

Outline

1. Introduction

2. Flexible designs

3. Statistical methods and properties

4. Example

5. Discussion and paths for future research

Introduction

- Recent debates about adaptive designs in
  - statistical literature: *Stat Med* (Temple, 2006), *Biom J* (special issue august 2006), *J Biopharm Stat* (white paper from a working group of pharmaceutical industry, 2006)
  - medical literature: *JAMA* editorial, october 25, 2006
- Involving statisticians, trialists and specialists from
  - pharmaceutical companies
  - academic world
  - regulatory agencies: FDA (*Critical paths initiatives, 2004 & 2006*), EMEA

What are we talking about?*

- Adaptive designs: use accumulated data to (possibly) modify some aspects of an ongoing study
- Without compromising the validity and integrity of the trial
- Exemples of adaptive designs
  - Randomized Play-the-Winner$\longrightarrow$ allocation rule
  - Sample size reassessment $\longrightarrow$ sampling rule
  - Group sequential trial $\longrightarrow$ stopping rule
- Includes Bayesian designs, dose finding trials with dose escalation (CRM, Bayesian, . . .), and flexible designs

*Draglin V. Adaptive designs: classification and taxonomy. Adaptive Designs Workshop, 2006.

Pro and Cons

- Pro
  - Allow to better target patients subsets using new technologies (e.g. use of proteomic or genetic markers)
  - Provide a theoretical framework to some common practice protocol amendments
  - Reduce time to decision (in particular, set up next trials)
  - Save resources for the entire drug development process

- Cons
  - Vulnerable to unblinding and bias (Also true for "usual" sequential trials)
  - Over-emphasize statistical over clinical significance
  - Expand the role of DSMB (Are they OK? Are the companies OK?)
  - Complicate the trial's logistic
  - Useful at all? (Shouldn't we simply plan early phase trials more carefully?)

Flexible designs*

- Adaptive group sequential designs
- Where the sequel of a trial can be re-designed after an interim analysis
- "Re-design":
  - sample size reassessment and timing of next IAs (adaptive sampling rule)
  - combination of p-values or test statistics, conditional error, ... (adaptive stopping rule)
  - changing the target treatment difference, the primary endpoint, the test statistic, ... (adaptive decison rule)
  - ...

*Bauer 1989; Bauer & Köhne 1994; Schäfer & Müller 2004

Principles (1)

- Modification of the trial design **while preserving the type I error**

- Concept of conditional error

  – invariance principle: any modification preserving the conditional type I error preserves the overall type I error

  – replace the remainder of a trial by a design preserving the initial type I error, conditional on what has been observed until the IA

- One may adapt the trial design

  – using data collected during the trial (up to the IA)

  – or information from outside the trial (e.g. other trials)

  – no need to prespecify the modifications

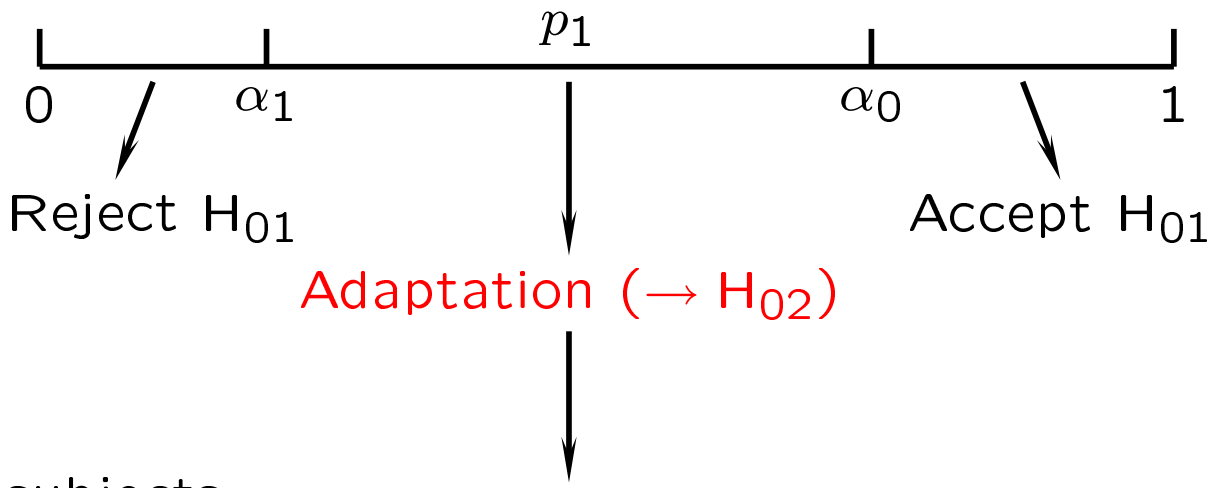
Principles (2)

- Many methods available
- Sequential analysis: integrate information from different stages
- Approaches: variance-spending, conditional error, combination tests
- Notion of combination tests
  - combine normal deviates from independent samples
  - combine independent p-values
  - combination rule has to be specified in advance
- Sequel: focus on recursive combination tests

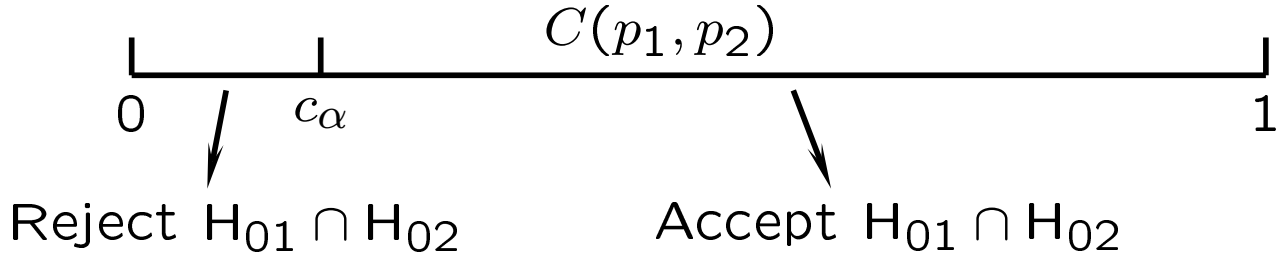
General scheme

Stage 1 Null hypothesis H_{01}

n_1 subjects, p-value p_1



Stage 2 : n_2 subjects, p_2



Stage 1

- Plan a conventional group sequential design
- Determine H_{01} (null hypothesis to be tested), T_1 (test statistic), n_1
- Fix the type I (α) and II (β) error rates
- Choose the combination rule $C(p_1, p_2)$ (has to be fixed, but α_1 and α_0 may be adapted)

Interim analysis

- $T_1 \rightarrow p_1$
- $p_1 \leq \alpha_1 (< \alpha)$ or $p_1 > \alpha_0 (> \alpha) \rightarrow$ stop
- $\alpha_1 < p_1 \leq \alpha_0 \rightarrow$ procede to stage 2
- Possibility to adapt the design
  - using information from inside and outside the trial
  - one may define a new hypothesis H_{02}
- Adaptations to be written down in a protocol

Stage 2

- Include n_2 subjects as planned at the IA
- $T_2 \to p_2$
- p-value $p = C(p_1, p_2)$
- Use p to reject $\mathsf{H}_{01} \cap \mathsf{H}_{02}$ or not, at the level α
- Or plan a third stage . . .

Proposed adaptations

1. Sample size reassessment
2. Conditional power
3. Selecting treatment arms / doses
4. selecting / adding endpoints
5. Inserting / skipping interim analyses
6. Others

Sample size reassessment

- Most published works concerned sample size reassessment
- Few debates on using blinded or unblinded nuisance parameter estimation (e.g. variance)
- To preserve the trial power in case of misspecification of these parameters
- Flexible designs allow the use of information from the first stage to reevaluate the sample size
- More debates on using observed treatment effects (see next)

Conditional power

- Sample size computed for a power level against a given alternative
- Conditional power at IA: power conditional to what has been observed
- Power under which alternative?
  - observed effect size (predictive power): random variable, large apparent effect may have dramatic consequences
  - planned effect size (conditional power), but haow do we account for an unsuspected large effect?
- Flexible trials allow to consider unsuspected effects, in terms of efficacy or safety
- Be careful with parameters estimated at an interim analysis

Selecting treatment arms / doses

- One of the first true applications of flexible designs
- The first stage is used to select a subset of doses and to determine sample size for stage 2
- Possibility to redefine the allocation ratio between doses, or to add new doses
- No need to prespecify the adaptations
- But be careful not to compromise the credibility of the trial

Selecting / adding endpoints

- Select endpoints combined into a single test statistic (e.g. drop those with too much measurement variability)
- Modify the hierarchy of a multiple testing procedure
- Add some new endpoint
- But modifying the main endpoint will cast a reasonable doubt on the trial conclusions (and obviously indicate a poor planning . . .)

Inserting / skipping interim analyses

- Flexible desings = group sequential designs + flexibility
- adapting the upcoming IAs seems natural (number, timing)
- Some authors have tried to incorporate an optimal choice of sample size between two IAs in the sequential decision process, but methods are rather complex
- Add an IA if unblinded data indicate a high chance of obtaining a decision, or drop the next one in the converse case
- Modification of the preplanned sequential boundaries (e.g. Pocock instead of O'Brien & Fleming)

Others

- Choice of a tests statistic better suited to the data (e.g. non-proportional hazards, or adjusting on a variable that was not part of the original protocol)

- Select one or more patients subsets

- Main controversy: Risk to modify a "good" protocol on the basis of a limited number of patients

- But common practice shows that protocol amendments are very frequent anyway

- Flexible designs provide a framework to modify a protocol, with at least control of the type I error

Statistical methods

- Suppose we test H_0 vs H_1

Stage 1
n_1 observations, p_1

  - Reject H_0 if $p_1 \leq \alpha_1$
  - Stop for futility if $p_1 > \alpha_0$

Stage 2

Possibly having adapted H_0

n_2 observations, p_2

Reject H_0 if $C(p_1, p_2) \leq c_{\alpha_2}$

- Particular choices discussed in the literature
  - $\alpha_0 < 1$ and $\alpha_2 = \alpha$ (final test at the nominal level)
  - $\alpha_0 = 1$ (no early stopping for futility)

- Conditions on $C(.,.)$, c_{α_2} ?

Recursive combination test*: combining p-values

- $C(p_1, p_2)$ is increasing in both arguments, strictly increasing in at least one, and left-continuous in p_2, $\forall p_1 \in]\alpha_1, \alpha_0]$ et $p_2 \in [0, 1]$

- (p_1, p_2) have to be p-clud, i.e.
  $\Pr_{H_0}(p_1 \leq \alpha) \leq \alpha$ et $\Pr_{H_0}(p_2 \leq \alpha | p_1) \leq \alpha, \quad \forall \alpha \in [0, 1]$

- **Property:** if p_1 and p_2 are independent and uniformly distributed, (p_1, p_2) are p-clud

- c_{α_2} depends on α, α_0, α_1, α_2 and $C(.,.)$

- Frequent (and simple) combination function: Fisher's product criterion

*Brannath et al. 2002

Fisher's product criterion

- $C(p_1, p_2) = p_1 \times p_2$

- Final test at level $\alpha_2 \Leftrightarrow$ reject H_0 iff

$$p_1 p_2 \leq c_{\alpha_2} = \exp(-0.5\chi^2_{4,1-\alpha_2})$$

- Clearly, $c_{\alpha_2} \leq \alpha_1$ as $p_1 p_2 \leq p_1$

- $c_{\alpha_2} = \dfrac{\alpha - \alpha_1}{\ln \alpha_0 - \ln \alpha_1}$

Recursive combination tests: level

- $\mathsf{Pr}_{\mathsf{H}_0}(p_1 \leq \alpha_1) + \mathsf{Pr}_{\mathsf{H}_0}(C(p_1, p_2) \leq c_{\alpha_2}, \alpha_1 < p_1 \leq \alpha_0) = \alpha$

- If p_1 and p_2 are independent and $\sim \mathsf{U}_{[0,1]}$

$$\alpha_1 + \int_{\alpha_1}^{\alpha_0} \int_0^1 I_{[C(x,y) \leq c_{\alpha_2}]} dx dy = \alpha$$

Recursive combination tests: global p-value

- Let us define

$$q(p_1, p_2) = \begin{cases} p_1 & \text{if } p_1 \leq \alpha_1 \text{ or } p_1 > \alpha_0 \\ \alpha_1 + \int_{\alpha_1}^{\alpha_0} \int_0^1 I_{[C(x,y) \leq C(p_1,p_2)]} dx dy & \text{otherwise} \end{cases}$$

- $q(p_1, p_2)$ is the p-value of the combination test*
  - if p_1 and p_2 anre independent and $\sim \mathsf{U}_{[0,1]}$, $q(p_1, p_2) \sim \mathsf{U}_{[0,1]}$
  - Reject H_0 iff $q(p_1, p_2) \leq \alpha$

*Brannath et al. 2002; although challenged by Liu & Pledger 2005

Recursive combination test: weighted sum of normal deviates

- Other method, equivalent to combine p-values
- Stage 1: n_1, z_1. Reject H_0 if $z_1 \geq z_{\alpha_1}$, accept if $z_1 < z_{\alpha_0}$
- Interim analysis: adapt pre-planned $n_2 \rightarrow \tilde{n}_2$
- Stage 2: $\tilde{n}_2$, $\tilde{z}_2$. Reject if

$$\tilde{z} = w_1 z_1 + w_2 \tilde{z}_2 \geq z_{\alpha_2},$$

  where $w_i = \sqrt{\frac{n_1}{n_1+n_2}}$

- Global level of test procedure α if weights w_i are defined a priori

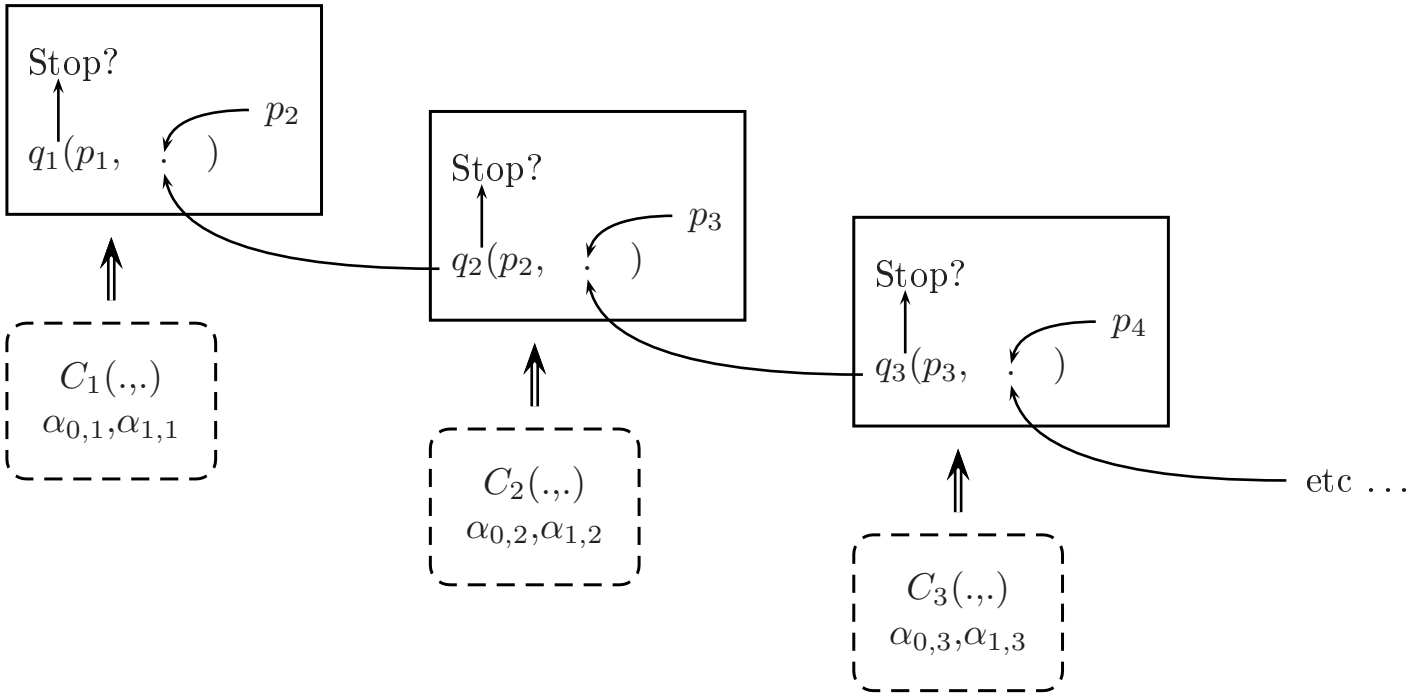
Conditional error function*

- Other approach to construct adaptive designs, equivalent to combination tests
- We observe z_1 at the first IA
- Control of the type I error
- CEF: probability of a type I error at the final analysis, given we have observed z_1
- Any increasing function $A(z)$ with range $[0, 1]$ satisfying $\int A(z)\phi(z)dz = \alpha$

*Proschan & Hunsberger 1995

Generalization to more than two stages

- The stage 2 of previous design is also an IA, and so on . . .



- Global p-value: $p = q_1(p_1, q_2(p_2, q_3(\ldots)))$

- Reject H_0 iff $p \leq \alpha$

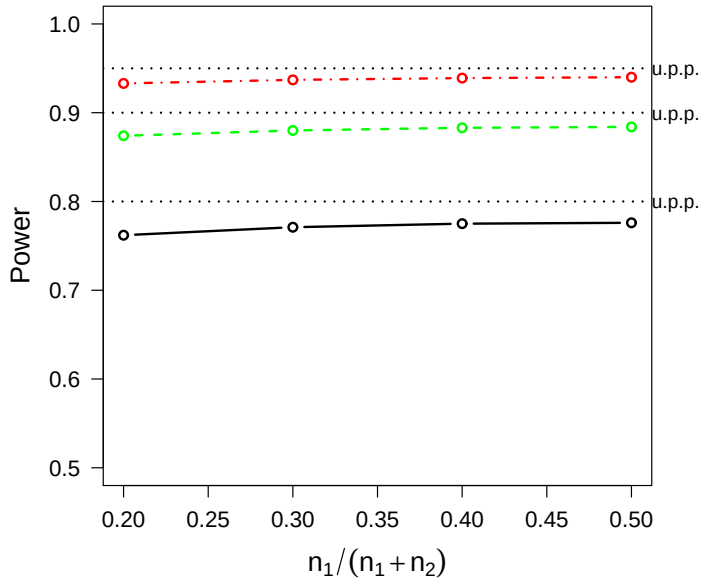
Estimation and confidence intervals

- Estimation following and adaptive design
- Possibility to construct confidence intervals
- Several methods: repeated CI (Lehmacher & Wassmer 1999), alternative approach (Brannath et al. 2002)
- One-sided confidence bounds can be used for conclusion, as an equivalent to combination test p-value

Some properties of adaptive designs (1)

Loss of power of Fisher's combination test

- Power of the test (fixed sample size, no adaptation at the IA)



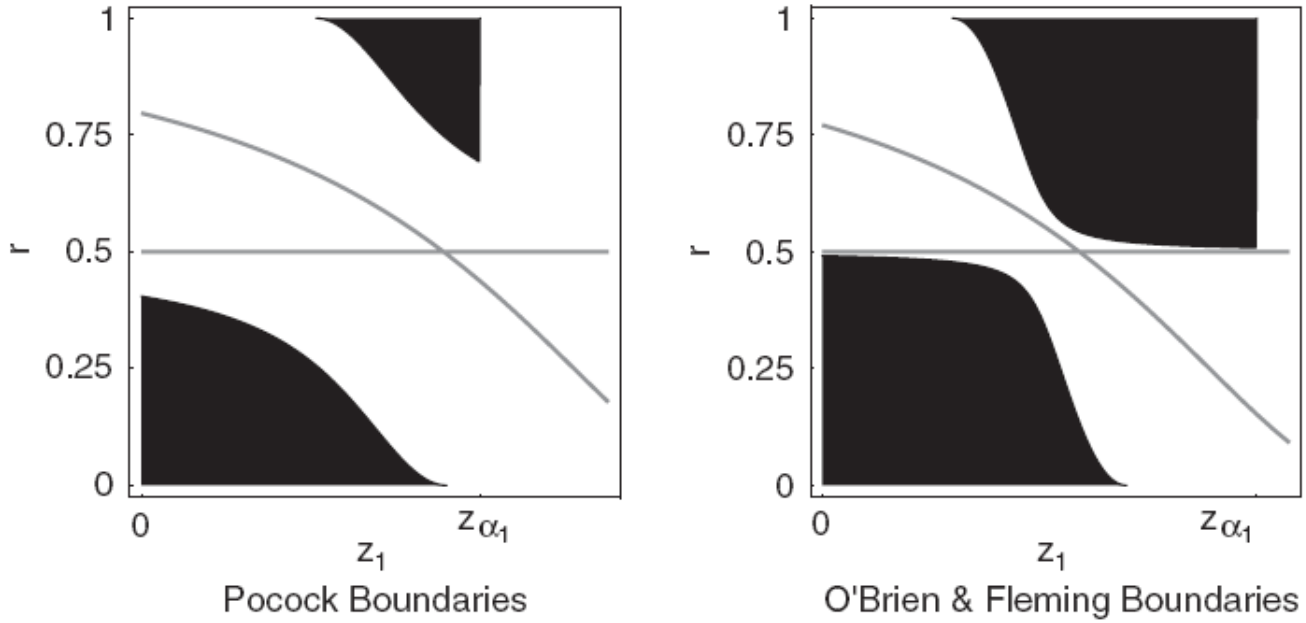
- Aditional loss of power if early stopping is incorporated

Some properties of adaptive designs (2)

Inconsistencies of rejection regions

- When the adaptive test rejects but not the classical pooled sample test

- Two examples (with $n_1 = n_2$, no adaptation)

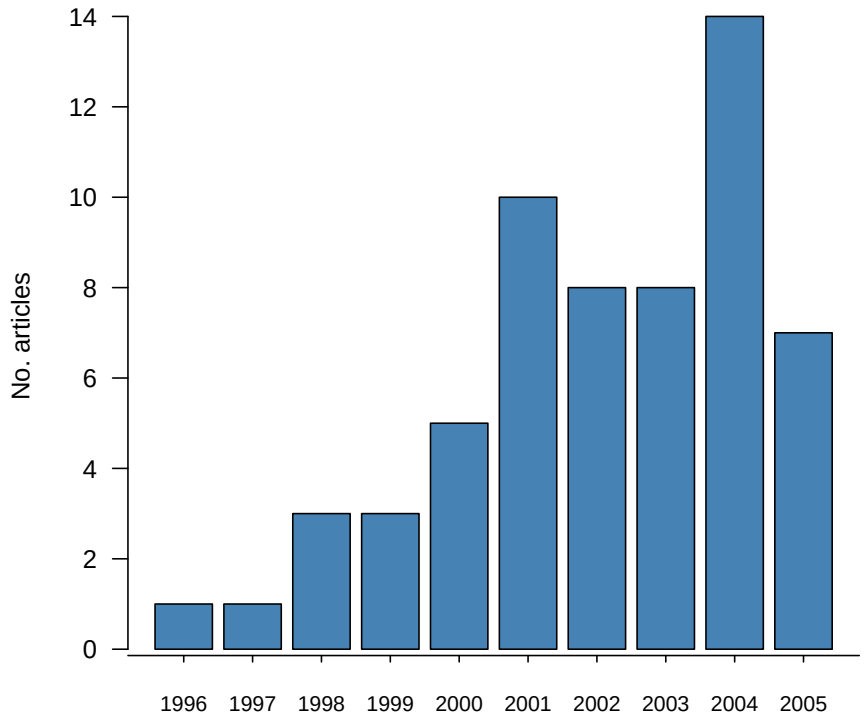


Flexible designs: summary

- Extension of the concept of group sequential designs
- Framework in complience with ICH requirements
- Flexibility = design adaptation using all available information sources
- At planned or unplanned IAs
- Allows to compute a global p-value and confidence intervals, while preserving the type I error
- Not very common yet, but already used in practice

Application of flexible designs*

Number of published applications



From: Bauer P, Einfalt J. Biom J 2006; 48:493–506

*Bauer & Einfalt 2006

Exemple: hypericum extract WS 5570 trial*

- Randomized, double-blind, multicenter, non-inferiority trial
- Comparing hypericum to paroxetine in severe depression
- Primary outcome: absolute decrease of the Hamilton score
- Primary analysis: intent to treat
- Option to test for superiority if non-inferiority is established

*Szegedi et al. *BMJ* 2005;330(7490):503

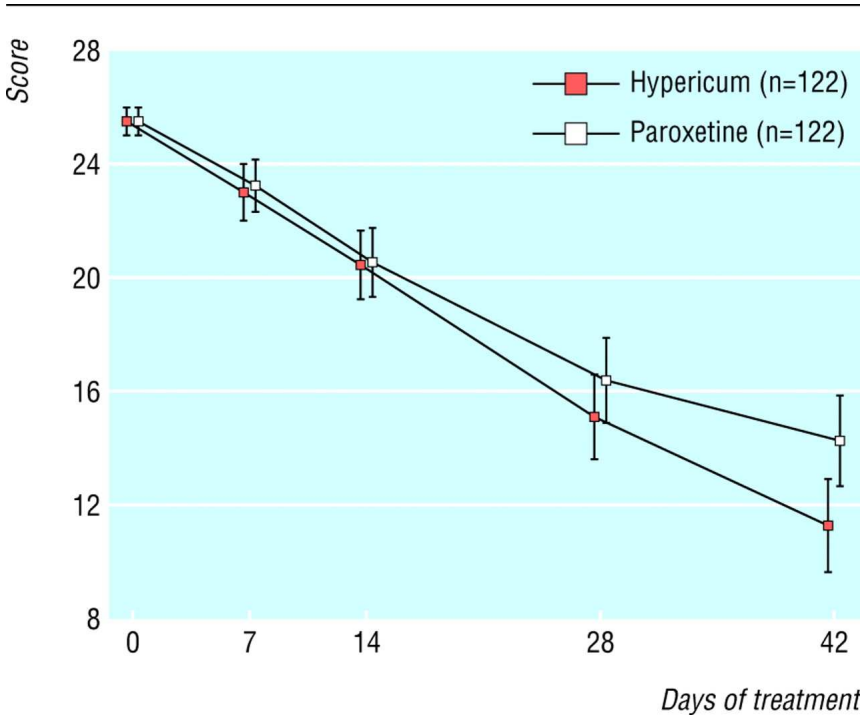
Design

- Non-inferiority margin: $\delta = -2.5$ pts
- Shifted t-test
- Global type I error $\alpha = 0.025$ (one-sided)
- Adaptive interim analysis: $\alpha_1 = 0.01$, $\alpha_0 = 0.5$
- Fisher's product test for final analysis $c_\alpha = 0.0038$
- Analysis: one-sided 97.5% confidence limit adjusted for IA (equivalent to the combination test)

Sample size

- Stage 1: $n_1 = 50$/gp
  - assuming equal changes and common SD of 6 pts
  - power 90% for $p_1 \leq 0.20$ (referred as trend toward non-inferiority)
- Interim analysis: $p_1 = 0.084$
- Stage 2: $n_2 = 75$/gp
  - local type I error level $c_\alpha / p_1 = 0.045$
  - power 80%
  - no adaptation

Results*



- Mean decrease by 14.4 (8.8) with hypericum *vs* 11.4 (8.6) with paroxetine
- Lower one-sided 97.5% confidence limit of difference at 1.5 pt
- $1.5 > -2.5 \rightarrow$ non-inferiority
- $1.5 > 0 \rightarrow$ superiority at the 2.5% level
- Results confired in a per protocol analysis

*©BMJ

Comments

- Flexible designs may be accepted also in high rank journals
- One of the very good examples of how to present results of adaptive designs
- No discussion of the methods in the paper, however
- In this case, no clear gain of a flexible design: fixed sample size would have been $n = 91$/gp instead of 125
- But price to pay for allowing flexibility

Discussion

- In general, inefficient as compared to a group sequential trial
- Price to pay for the possibility of replanning the trial
- The adaptive feature has to be crucial for the trial
  - internal pilot study
  - nuisance parameters difficult to anticipate (suvival, compliance, . . .)
  - phase II/III combined trials
- Importance to preserve the trial credibility (which null hypothesis is tested at the end?)
- **Not an alternative to poorly planned trials**

Future research

- Advantages of adaptive/flexible trials recognized in early phase trials
- More debated in phase III trials
- One of the most promising applications are seamless phase II/III trials
  - considered as confirmatory trials
  - combine a phase II and a phase III trial into a single uninterrupted study
  - stage 1 (phase II): e.g. dose-finding or early biological response (cancer trials)
  - stage 2 (phase III): selected dose(s), terminal endpoint
  - analysis uses data from all subjects

Applications needed

- Combined phase II/III trials bring a valuable answer to physicians and companies expectations
- As do adaptive designs in general
- More practical applications (and worked examples) are crucial
  - to disseminate the methodology
  - to show that adaptive designs are useful
  - to set some practical guidelines in their use