¡Bienvenidos!

Two-Stage Sequential Designs in Bioequivalence

Helmut Schütz
BEBAC
To bear in Remembrance...

Whenever a theory appears to you as the only possible one, take this as a sign that you have neither understood the theory nor the problem which it was intended to solve.

Even though it’s applied science we’re dealin’ with, it still is – science!

Karl R. Popper

Leslie Z. Benet
BE Study Designs

- long half life and/or patients in unstable conditions?
  - no
    - paired design
    - cross-over design
    - >2 formulations?
      - no
        - reliable information about CV?
          - yes
            - fixed-sample design
            - CV >30?
              - no
                - replicate design (reference scaling)
              - yes
                - 2×2 cross-over design replicate (unscaled)
            - no
              - two-stage sequential design
                - Currently no method if
                  - >2 formulations
                  - replicate design
                  - Futility rules (e.g., maximum sample size) in TSDs problematic.
        - no
          - replicate design or replicate replicate (unscaled)
      - yes
        - multi-arm parallel higher-order cross-over
  - yes
    - parallel design
Add-on / Two-Stage Designs

- Sometimes properly designed studies fail due to
  - ‘true’ bioinequivalence,
  - pure chance (producer’s risk),
  - poor study conduct (increasing variability),
  - false (mainly over-optimistic) assumptions about the CV and/or T/R-ratio – leading to a too small sample size (insufficient power).

- The sample size is planned based on assumptions…
Add-on / Two-Stage Designs

- Dealing with *inconclusive* BE studies (confidence interval not entirely with the acceptance range)
  - Repeat the study in a larger sample size.
  - Optionally perform a meta-analysis of pooled data. Only acceptable if at least one study demonstrates BE.
  - Recruit a second group of subjects and pool data?

- Discussed at Bio-International Conferences (1989, 1992) and guidelines from the 1990s.
  - The patient’s risk must be preserved!
  - Among rivaling methods the one with with the highest power should be selected.
Terminology

- **Add-On Designs**
  - Sample sizes of both groups have a lower limit.

- **Group Sequential Designs**
  - Sample sizes of both groups are pre-specified.

- **Adaptive Two-Stage Sequential Designs**
  - Groups sizes are (generally) not limited.
  - Sample size of the second group is re-estimated from the first group’s data.

H Schütz
*Two-stage designs in bioequivalence trials*
DOI: 10.1007/s00228-015-1806-2
Definition

- For an overview see Schwartz & Denne, Dragalin, Chow & Chang, and Chin

A study design is called *adaptive* if statistical methodology allows modification of a design element (e.g., the sample size) at an interim analysis with full control of the type I error (TIE).

Add-On Designs: Guidelines

- General conditions
  - Intention to perform an AOD has to be stated in the protocol,
  - the same batches of products, and
  - the same clinical and bioanalytical methods have to be employed in both groups.

- Currently only stated in GLs of Japan, Argentina, Mexico, and Korea
  - The patient’s risk might be seriously compromised!
Add-on / Two-Stage Designs

CV 30–90%, α 0.05, n₁ 20–72

\[ n₂ = \frac{1}{2} n₁ \]

\[ \lim_{N \to \infty} \alpha = 0.0746 \]

Japan (2012)
- 1\textsuperscript{st} group \((n₁) ≥ 20\) evaluated with \(α 0.05\) (90\% CI)
- 2\textsuperscript{nd} group \((n₂) ≥ \frac{1}{2} n₁\)
- Pooled data evaluated with \(α 0.05\) (90\% CI)
- Inflation of the patient's risk (up to 7.5\%)

Inflation of the Type I Error: Investigations on Regulatory Recommendations for Bioequivalence of Highly Variable Drugs Pharm Res 32(1), 135–43
DOI: 10.1007/s11095-014-1450-z
Group Sequential Designs

- Long and accepted tradition in clinical research (mainly phase III)
  - Developed for superiority testing, normal distributed data with known variance, fixed and equal sizes of groups.

AL Gould
Group Sequential Extension of a Standard Bioequivalence Testing Procedure
DOI: 10.1007/BF02353786
Group Sequential Designs: GLs

- **Australia (2004), Canada (Draft 2009)**
  - Application of Bonferroni’s correction ($\alpha 0.025$).
  - Theoretical TIE $\leq 0.0494$.
  - For CVs and samples sizes typical in BE $\leq 0.04$.

- **Canada (2012)**
  - Pocock’s $\alpha 0.0294$.
  - $n_1$ based on ‘most likely variance’ + additional subjects to compensate for expected dropout-rate.
  - Total sample size based on ‘worst-case scenario’.
  - If $n_2 \neq n_1$ relevant inflation of the TIE is possible!
Adaptive TS Sequential Designs

- Methods by Potvin et al. (2008) first validated framework in the context of BE
  - Supported by the ‘Product Quality Research Institute’ (members: FDA/CDER, Health Canada, USP, AAPS, PhRMA...).
  - Inspired by conventional BE testing and Pocock’s $\alpha = 0.0294$ for Group Sequential Designs.

Potvin D, DiLiberti CE, Hauck WW, Parr AF, Schuirmann DJ, and RA Smith
Sequential design approaches for bioequivalence studies with crossover designs
Adaptive TS Sequential Designs

- Two ‘types’ of TS Sequential Designs
  1. The *same* adjusted $\alpha$ is applied in both stages (regardless whether a study stops already in the first stage or proceeds to the second stage).
     - Based on Group Sequential Design.
     - In publications called Method B.
  2. An unadjusted $\alpha$ *may* be used in the first stage (dependent on interim power).
     - Based on conventional BE testing + GSD.
     - In publications called Method C, D, C/D.
Review of Guidelines

- **EMA (Jan 2010)**
  Acceptable; Potvin et al. Method B preferred (?)

- **Canada (May 2012)**
  Potvin et al. Method C recommended.

- **FDA (Jun 2012)**
  Potvin et al. Method C/D recommended.
  API specific guidance: Loteprednol

- **Russia (2013)**
  Acceptable; Potvin et al. Method B preferred (?)
Potvin et al. (Method B – type 1)

Evaluate BE at stage 1 ($\alpha = 0.0294$)

- yes
  
  BE met?
  
  - yes
    
    Evaluate power at stage 1 using $\alpha$-level of 0.0294
    
    - yes
      
      $\geq 80\%$?
      
      - yes
        
        Estimate sample size based on $CV_{intra}^T/R$
        
        0.95, $\alpha = 0.0294$; continue to stage 2
      
      - no
        
        Evaluate BE at stage 2 using pooled data from both stages ($\alpha = 0.0294$)
    
    - no
      
      Pass or fail

- no
  
  Pass
Potvin et al. (Method B – type 1)

1150·10^6 Sim’s (Method B)
Potvin et al. (Method B – type 1)

1150·10⁶ Sim’s (Method B)
Potvin et al. (Method B – type 1)

- Technical Aspects
  - Only *one* Interim Analysis (after stage 1).
  - Potvin et al. used a simple power estimation based on the shifted central $t$-distribution. Use software (avoid approximations). Example 2:
  - Should be termed ‘Interim Power Analysis’ – *not* ‘Bioequivalence Assessment’ in the protocol.
  - No *post hoc* Power – only a validated method to guide the decision tree.
  - No adjustment for $T/R$-ratio observed in stage 1!
Potvin et al. (Method B – type 1)

Technical Aspects (cont’d)

- No futility rule preventing to go into stage 2 with a high sample size!
  Must be clearly stated in the protocol (unfamiliar to the IEC because common in Group Sequential Designs).

- Pocock’s $\alpha$ 0.0294 is used in stage 1 and in the pooled analysis (data from stages 1 + 2),
  i.e., the $100(1 - 2\times\alpha) = 94.12\%$ CI is calculated.

- Overall TIE preserved at $\leq 0.05$. 
Technical Aspects (cont’d) + EMA modification

If the study is stopped after stage 1, the statistical model is:

- fixed: sequence + period + treatment
  + subject(sequence)

If the study continues to stage 2, the model for the combined analysis is:

- fixed: stage + sequence + sequence(stage)
  + subject(sequence × stage) + period(stage)
  + treatment

No poolability criterion! Combining is always allowed – even if a significant difference between stages is observed. No need to test this effect.
Potvin et al. (Method B – type 1)

- Technical Aspects (cont’d) + EMA modification
  - Incomprehensible why this modification was introduced by EMA’s Biostatistical Working Party
  - Simulations performed or “gut feeling”?
    - Modification shown to be irrelevant.
  - Furthermore no difference whether subjects were treated as a fixed or random term (unless T/R >1.20).

Karalis V and P Macheras
*On the Statistical Model of the Two-Stage Designs in Bioequivalence Assessment*
Potvin et al. (Method C – type 2)

Evaluate power at stage 1 using $\alpha$-level of 0.050

- **yes**
  - Evaluate BE at stage 1 ($\alpha$ 0.050)
  - **yes**
    - BE met?
      - **yes**
        - Estimate sample size based on $CV_{\text{intra}}$, $T/R$ 0.95, $\alpha$ 0.0294; continue to stage 2
      - **no**
        - Evaluate BE at stage 2 using pooled data from both stages ($\alpha$ 0.0294)
        - **Pass or fail**
  - **no**
    - Evaluate BE at stage 1 ($\alpha$ 0.0294)
    - **yes**
      - BE met?
        - **yes**
          - Estimate sample size based on $CV_{\text{intra}}$, $T/R$ 0.95, $\alpha$ 0.0294; continue to stage 2
        - **no**
          - Evaluate BE at stage 2 using pooled data from both stages ($\alpha$ 0.0294)
          - **Pass or fail**
    - **no**
      - Evaluate BE at stage 2 using pooled data from both stages ($\alpha$ 0.0294)
      - **Pass or fail**

- **no**
  - Evaluate BE at stage 1 ($\alpha$ 0.050)
  - **yes**
    - BE met?
      - **yes**
        - Estimate sample size based on $CV_{\text{intra}}$, $T/R$ 0.95, $\alpha$ 0.0294; continue to stage 2
      - **no**
        - Evaluate BE at stage 2 using pooled data from both stages ($\alpha$ 0.0294)
        - **Pass or fail**
  - **no**
    - Evaluate BE at stage 2 using pooled data from both stages ($\alpha$ 0.0294)
    - **Pass or fail**
Potvin et al. (Method C – type 2)

1150 \cdot 10^6 \text{ Sim's (Method C)}
Potvin et al. (Method B/C – type 1/2)

- Pros & Cons
  - Method C (if power $\geq 80\%$) is a conventional BE study; no penalty in terms of $\alpha$ needs to be applied.
  - Method C proceeds to stage 2 less often and has smaller average total sample sizes than Method B for cases where the initial sample size is reasonable for the CV.
  - If the size of stage 1 is low for the actual CV both methods proceed to stage 2 almost all the time; total sample sizes are similar.
  - Method B slightly more conservative than C.
**Recommendations**

- **Type 2 preferred due to slightly higher power than type 1 (FDA, HPFB). Type 1 for EMA (?)**
- **Plan the study as if the CV is known**
  - If assumptions turn out to be true = no penalty
  - If lower power (CV higher than expected), BE still possible in first stage (penalty; 94.12% CI) or continue to stage 2 as a ‘safety net’.
- **Don’t jeopardize! Small sample sizes in the first stage don’t pay off. Total sample sizes are ~10–20% higher.**
TSDs: Alternatives

- Methods by Potvin et al. (2008) limited to $T/R$ of 0.95 and 80% power

- Follow-up publications ($T/R$ 0.95…0.90, 80…90% power)

<table>
<thead>
<tr>
<th>reference</th>
<th>type</th>
<th>method</th>
<th>$T/R$</th>
<th>target power</th>
<th>CV</th>
<th>$\alpha_{adj}$</th>
<th>max. TIE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potvin et al.</td>
<td>1</td>
<td>B</td>
<td>0.95</td>
<td>80%</td>
<td>10–100%</td>
<td>0.0294</td>
<td>0.0485</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>C</td>
<td>0.95</td>
<td>80%</td>
<td>10–100%</td>
<td>0.0294</td>
<td>0.0510</td>
</tr>
<tr>
<td>Montague et al.</td>
<td>2</td>
<td>D</td>
<td>0.90</td>
<td>10–80%</td>
<td></td>
<td>0.0280</td>
<td>0.0518</td>
</tr>
<tr>
<td>Fuglsang</td>
<td>1</td>
<td>B</td>
<td>0.95</td>
<td>90%</td>
<td>10–80%</td>
<td>0.0284</td>
<td>0.0501</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>C/D</td>
<td>0.90</td>
<td>90%</td>
<td>10–80%</td>
<td>0.0274</td>
<td>0.0503</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>C/D</td>
<td>0.90</td>
<td>90%</td>
<td>10–80%</td>
<td>0.0269</td>
<td>0.0501</td>
</tr>
</tbody>
</table>

Montague TH, Potvin D, DiLiberti CE, Hauck WW, Parr AF, and DJ Schuirmann

Additional results for ‘Sequential design approaches for bioequivalence studies with crossover designs’

A Fuglsang

Sequential Bioequivalence Trial Designs with Increased Power and Controlled Type I Error Rates
Slight inflation of the TIE in some ‘type 2’ designs could easily be avoided

Modifications of published adjusted $\alpha$

<table>
<thead>
<tr>
<th>type</th>
<th>method</th>
<th>$T/R$</th>
<th>$\alpha_{adj}$</th>
<th>max. TIE</th>
<th>$\alpha_{adj}^*$</th>
<th>max. TIE$^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B</td>
<td>0.95</td>
<td>0.0294</td>
<td>0.0485</td>
<td>0.0304</td>
<td>0.0501</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>0.90</td>
<td>0.0280</td>
<td>0.0510</td>
<td>0.0282</td>
<td>0.0500</td>
</tr>
</tbody>
</table>

* Schütz H, Labes D, and A Fuglsang

* Modifications of ‘Sequential design approaches for bioequivalence studies with crossover designs’ in preparation (2015)
Montague et al. (Method D – type 2)

1150·10⁶ Sim’s (Method D)
Parallel Groups (Type 1/2)

Fuglsang (2014)
- Based on Potvin’s Methods B/C ($\alpha_{adj} = 0.0294$, 80% power)
- Framework: $n_1 = 48–120$, CV 10–100%
  - equal allocation ($N_{Test} = N_{Reference}$)
  - equal and unequal variances of groups
  - conventional $t$-test and Welch-Satterthwaite approximation

Results
- No significant inflation of the TIE
- Power $\geq 78.4\%$

Fuglsang
Sequential Bioequivalence Approaches for Parallel Designs
AAPS J 16(3), 373–8 (2014), DOI: 10.1208/s12248-014-9571-1
Futility Rules revised

- EMA GL Section 4.1.8 ‘Two-stage design’
  “[…] the stopping criteria should be clearly defined prior to the study.”
  - What does that mean?
    - Failing in stage 1 or the pooled analysis according to the chosen method.
      → Part of the validated frameworks.
    - Early stopping for futility (e.g., ‘bad’ ratio, extreme stage 2 sample size caused by high CV – better to opt for reference-scaling…).
      → Not validated. A misunderstanding by regulators (stopping criterion ≠ futility rule).
Futility Rules revised

- Introduction of a futility rule does not inflate the TIE, but power may drop substantially!
  - State stopping criteria unambiguously in the protocol.
  - If you want to introduce a futility rule, simulations are mandatory in order to maintain sufficient power.

  "Introduction of [...] futility rules may severely impact power in trials with sequential designs and under some circumstances such trials might be unethical."

A Fuglsang
Futility Rules in Bioequivalence Trials with Sequential Designs
APPS J 16(19), 79–82 (2014) DOI: 10.1208/s12248-013-9540-0
Validation of Frameworks

● Jones and Kenward concluded that
  “[…] before using any of the methods […] , their operating characteristics should be evaluated for a range of values of $n_1$, CV and true ratio of means that are of interest, in order to decide if the Type I error rate is controlled, the power is adequate and the potential maximum total sample size is not too great.”

● Uncomplicated with current software
  ■ Automatically finding a suitable $\alpha_{adj}$ and validating for TIE and power takes ~20 minutes.

Jones B and MG Kenward
*Design and Analysis of Cross-Over Trials*
Chapman & Hall/CRC, Boca Raton (3rd ed. 2014)
D Labes
*Package ‘Power2Stage’, Version 0.2-2* (2014-12-08)
Cost Analysis

- Consider certain questions:
  - Is it possible to assume a best/worst-case scenario?
  - How large should the size of the first stage be?
  - How large is the expected sample size in the second stage?
  - Which power can one expect in both stages?
  - Will introduction of a futility criterion substantially decrease power?
  - Is there a sample size penalty compared to a fixed-sample design?
## Cost Analysis

### Example

- Expected CV 20%, desired power is 80% for a $T/R$-ratio of 0.95. Comparison of a type 1 TSD with a conventional fixed-sample design ($n = 20$, 83.5% power).

<table>
<thead>
<tr>
<th>$n_1$</th>
<th>$E[N]$</th>
<th>Studies stopped in stage 1 (%)</th>
<th>Studies failed in stage 1 (%)</th>
<th>Power in stage 1 (%)</th>
<th>Studies in stage 2 (%)</th>
<th>Final power (%)</th>
<th>Costs (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>20.6</td>
<td>43.6</td>
<td>2.3</td>
<td>41.3</td>
<td>56.4</td>
<td>84.2</td>
<td>+2.9</td>
</tr>
<tr>
<td>14</td>
<td>20.0</td>
<td>55.6</td>
<td>3.0</td>
<td>52.4</td>
<td>44.5</td>
<td>85.0</td>
<td>+0.2</td>
</tr>
<tr>
<td>16</td>
<td>20.1</td>
<td>65.9</td>
<td>3.9</td>
<td>61.9</td>
<td>34.1</td>
<td>85.2</td>
<td>+0.3</td>
</tr>
<tr>
<td>18</td>
<td>20.6</td>
<td>74.3</td>
<td>5.0</td>
<td>69.3</td>
<td>25.7</td>
<td>85.5</td>
<td>+3.1</td>
</tr>
<tr>
<td>20</td>
<td>21.7</td>
<td>81.2</td>
<td>6.3</td>
<td>74.9</td>
<td>18.8</td>
<td>86.2</td>
<td>+8.4</td>
</tr>
<tr>
<td>22</td>
<td>23.0</td>
<td>87.2</td>
<td>7.3</td>
<td>79.8</td>
<td>12.8</td>
<td>87.0</td>
<td>+15.0</td>
</tr>
<tr>
<td>24</td>
<td>24.6</td>
<td>91.5</td>
<td>7.9</td>
<td>83.6</td>
<td>8.5</td>
<td>88.0</td>
<td>+22.9</td>
</tr>
</tbody>
</table>
Cost Analysis

- **Example (cont’d)**
  - With 14 or 16 subjects in the first stage similar costs ($E[N] \sim 20$) are expected; with 16 one has a 66% chance to stop the study already in the first stage (62% chance to pass and 4% to fail).
  - With $n_1$ equal to the fixed design’s $n$ costs are expected to be 8% higher but we have a 75% chance to pass in the first stage and 86% power overall.
  - Power of the TSD is always larger than the one of the fixed-sample design – regardless the initial sample size and even if the assumed CV turns out to be correct.
Cost Analysis

● Example (cont’d)
  ■ If in a fixed-sample design the CV turns out to be higher than the assumed one, power will decrease, whereas in a TSD power is maintained.
  ■ Don’t start the first stage always in a small group and hope for a smaller than expected CV – which would be substantially more economic than a fixed-sample design. This is not necessarily a good idea: With 12 subjects power in the first stage is only 41% and 56% of studies will proceed to the second stage.
Advanced Example

- ‘Must pass’ BE in stage 1 (first to file)
  - Fixed $T/R$ 90% (pessimistic; very likely better).
  - Expected CV 20% (pilot study with two references).
  - Expected dropout rate $\sim$30%; start with 88 to have $n_1 \geq 60$.
  - Targets
    - $>90\%$ power for $n_1 \geq 60$ – even for extreme CV of 45%.
    - $90\%$ power for $n_1 \geq 60$ (CV 20%) in stage 1.
    - Not $<80\%$ power for CV $\geq 25\%$ in stage 1.
    - Low probability to proceed to stage 2.
Advanced Example

- ‘Must pass’ BE in stage 1 (first to file)
  - Sponsor preferred Method B (EU submission...).
  - Fuglsang published $\alpha_{adj}$ 0.0269 for $T/R$ 0.90 and 90% power – but only for Method C...
  - Same $\alpha_{adj}$ applicable for Method B?
  - Likely...
    - Potvin et al. showed less inflation of the TIE with Method B.
    - Fuglsang needed less adjustment in Method B.
    - But we have to justify that!
  - $10^6$ simulations for the TIE and $10^5$ for power.
Advanced Example

- ‘Must pass’ BE in stage 1 (first to file)
  - Targets met
    - 93% power for \( n_1 \leq 60 \) (CV 20%) and 90% for extreme CV of 45%.
    - 90% power for \( n_1 \geq 60 \) (CV 20%) in stage 1.
    - Low chances to proceed to stage 2 with CV 20%:
      - \( n_1 \leq 60: 6\%, \ n_1 \geq 72: 1\%
    - \( \geq 80\% \) power for \( CV \geq 20\% \) – even for a more extreme dropout rate.
    - \( \alpha_{adj} = 0.0271 \) would work as well (with 0.0278 < 0.052).
  - Study passed in the first stage (February 2014)
Case Study 1

- **Method C**: Study passed BE in stage 1
  - (49 subjects, CV 30.65%, 90% CI)
    - **UK/Ireland**: Unadjusted $\alpha$ in stage 1 not acceptable.
    - Study passed BE with 94.12% CI as well
      - *(post hoc* switch to Method B).
    - **Austria**: The Applicant should demonstrate that the type I error inflation, which can be expected from the chosen approach, did not impact on the decision of bioequivalence.
      - One million simulations based on the study’s sample size and CV.
        - TIE 0.0494 (95% CI: 0.0490 – 0.0498)
Case Study 2

- **Method C:** Study stopped in stage 1
  - AUC power >80%: passed BE with 90% CI
  - C<sub>max</sub> power <80%: passed BE with 94.12% CI

  - **The Netherlands:** Adapting the confidence intervals based upon
    power is not acceptable and also not in accordance with the EMA
    guideline. Confidence intervals should be selected *a priori*, without
    evaluation of the power. Therefore, the applicant should submit the
    94.12% confidence intervals for AUC.

    - AUC failed BE with 94.12% CI.
    - Sponsor repeated the study with a very (!) large sample size and
      failed on C<sub>max</sub>. Project cancelled.
Case Study 3

- Method C: Two studies passed in stage 1
  (SD n=15, MD n=16; C_{max} CV 17.9%, 8.54%; 90% CIs)
- Would have passed with Method B as well; however, 94.12% CIs were not reported.
  - RMS Germany. Accepted by CMSs Austria, Denmark, Sweden, and The Netherlands.
  - Spain: Statistical analysis should be GLM. Please justify.
    - Evaluated with fixed-effects model.
      Both studies passed.
      Issue resolved (September 2013)
Conclusions

● Do not blindly follow guidelines. Some current recommendations may lead to inflation of the patient’s risk and/or deteriorate power.

● Validated frameworks can be applied without requiring the sponsor to perform own simulations – though they could further improve power based on additional assumptions.

● Two-stage designs are both ethical and economical alternatives to fixed-sample designs.
Outlook

- Feasibility / futility rules.
- Arbitrary expected T/R and/or power.
- Methods without interim power.
- Dropping a candidate formulation from a higher-order cross-over; continue with 2×2.
- Continue a 2×2 in replicate design for scaling.
- Fully adaptive methods.
- Exact method (not depending on simulations).
- Application to replicate designs / scaling.
Don’t panic!

conventional 2×2 cross-over (fixed sample design)
¡Gracias!

Two-Stage Sequential Designs in Bioequivalence

Open Questions?

Helmut Schütz
BEBAC
Consultancy Services for
Bioequivalence and Bioavailability Studies
1070 Vienna, Austria
helmut.schuetz@bebac.at
To bear in Remembrance...

The fundamental cause of trouble in the world today is that the stupid are cocksure while the intelligent are full of doubt. 

*Bertrand Russell*

In bioequivalence we must not forget the only important – the patient! He/she is living person, not just $\alpha 0.05$.

*Dirk Marteen Barends*

It is a good morning exercise for a research scientist to discard a pet hypothesis every day before breakfast. It keeps him young.

*Konrad Lorenz*
References

- **ICH**
- **EMA-CPMP/CHMP/EWP**
  - Points to Consider on Multiplicity Issues in Clinical Trials (2002)
  - Questions & Answers: Positions on specific questions addressed to the EWP therapeutic subgroup on Pharma-cokinetics (2014)
- **US-FDA**
  - Center for Drug Evaluation and Research (CDER)
    - Statistical Approaches Establishing Bioequivalence (2001)
    - [Draft Guidance on Lotepredenol](http://cran.r-project.org/web/packages/Power2Stage/Power2Stage.pdf) (Apr 2013)
- **DB Owen**
  - *A special case of a bivariate non-central t-distribution*
    - Biometrika 52(3/4), 437–46 (1965)
  - Diletti E, Hauschke D, and VW Steinijans
    - *Sample size determination for bioequivalence assessment by means of confidence intervals*
- **SA Julious**
  - *Sample Sizes for Clinical Trials*
- **AL Gould**
  - *Group Sequential Extension of a Standard Bioequivalence Testing Procedure*
    - [DOI: 10.1007/BF02353786](http://cran.r-project.org/web/packages/Power2Stage/Power2Stage.pdf)
- **Hauck WW, Preston PE, and FY Bois**
  - *A Group Sequential Approach to Crossover Trials for Average Bioequivalence*
    - J Biopharm Stat 71(1), 87–96 (1997) [DOI: 10.1080/10543409708835171](http://cran.r-project.org/web/packages/Power2Stage/Power2Stage.pdf)
- **Potvin D et al.**
  - *Sequential design approaches for bioequivalence studies with crossover designs*
- **Montague TH et al.**
  - *Additional results for ‘Sequential design approaches for bioequivalence studies with crossover designs’*
    - Pharm Stat 11(1), 8–13 (2011) [DOI: 10.1002/pst.483](http://cran.r-project.org/web/packages/Power2Stage/Power2Stage.pdf)
- **D Labes**
  - *Package ‘Power2Stage’, Version 0.2-2* (2014-12-08)
    - [http://cran.r-project.org/web/packages/Power2Stage/Power2Stage.pdf](http://cran.r-project.org/web/packages/Power2Stage/Power2Stage.pdf)
References

- **A Fuglsang**
  *Controlling type I errors for two-stage bioequivalence study designs*

- **García-Arieta A and J Gordon**
  *Bioequivalence Requirements in the European Union: Critical Discussion*

- **BM Davit**
  *Sequential Designs and Interim Analyses in Bioequivalence: FDA’s Experience*
  AAPS Annual Meeting, Chicago, IL, October 13–18, 2012

- **Polli JE et al.**
  *Summary Workshop Report: Facilitating Oral Product Development and Reducing Regulatory Burden Through Novel Approaches to Assess Bioavailability/Bioequivalence*

- **A Fuglsang**
  *Sequential Bioequivalence Trial Designs with Increased Power and Controlled Type I Error Rates*

- **V Karalis**
  *The role of the upper sample size limit in two-stage bioequivalence designs*

- **Karalis V and P Macheras**
  *On the Statistical Model of the Two-Stage Designs in Bioequivalence Assessment*

- **A Fuglsang**
  *Futility Rules in Bioequivalence Trials with Sequential Designs*

- **A Fuglsang**
  *Sequential Bioequivalence Approaches for Parallel Designs*
  AAPS J 16(3), 373–8 (2014) [DOI: 10.1208/s12248-014-9571-1]

- **Jones B and MG Kenward**
  *Design and Analysis of Cross-Over Trials*

- **Wonnemann M, Frömke C, and A Koch**
  *Inflation of the Type I Error: Investigations on Regulatory Recommendations for Bioequivalence of Highly Variable Drugs*

- **H Schütz**
  *Two-stage designs in bioequivalence trials*