Two-Stage Designs

Background and Examples

Helmut Schütz

August 2017

In the EMA's Guideline Two-Stage Designs (TSD) are stated as acceptable when attempting to demonstrate bioequivalence (1): *"If this approach is adopted appropriate steps must be taken to preserve the overall type I error of the experiment and the stopping criteria should be clearly defined prior to the study. The analysis of the first stage data should be treated as an interim analysis…"* Based on group-sequential designs (GSD) with interim analyses (2 – 6) a few methods have been published in the context of bioequivalence (7, 8) – which did not achieve regulatory acceptance. Recently numerous frameworks were developed in order to control the type I error (TIE) without requiring the sponsor to perform own simulations (9 – 19). In a review TSDs were classified into two 'Types' (20):

1. The *same* adjusted α is applied in both stages (regardless whether a study stops in the first stage or proceeds to the second stage).
2. An unadjusted α *may* be used in the first stage, dependent on interim power.

Both types use an *interim* power estimation as a means to guide the decision tree. Clearly, the former (Fig. 1) was inspired by Pocock's GSD (3, 5) with one interim analysis, whereas the latter (Fig. 2) by conventional BE testing. The rationale of conditionally adjusting α in the first stage is the following: If the sample size of the first stage was planned for a given target power P (based on an assumed T/R ratio A and CV), it is reasonable to evaluate power first (for A – *not* the observed T/R ratio). Interim power ≥ P implies that assumptions hold and the framework proceeds with *unadjusted* α like a conventional fixed-sample pivotal BE study (left branch of Fig. 2). Since only the CV is used in interim power, no α has to be 'spent'. On the other hand, interim power < P indicates a higher than expected CV, and the framework proceeds to the sequential part, where adjustment is mandatory (right branch of Fig. 2).

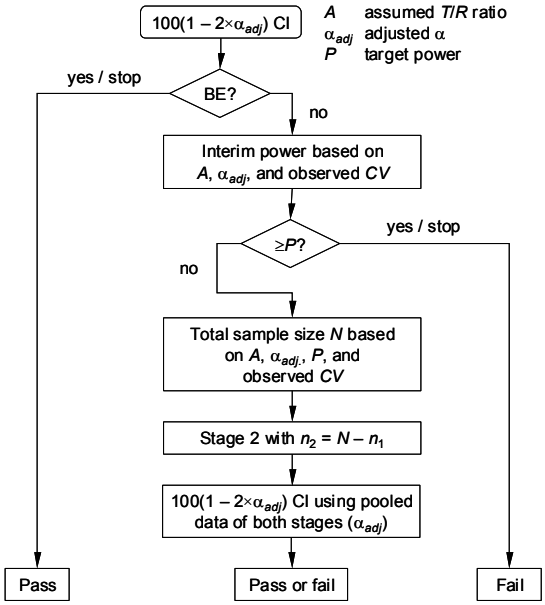


Fig. 1 'Type 1' TSD with adjusted α in both stages (variants of Potvin's 'Method B')

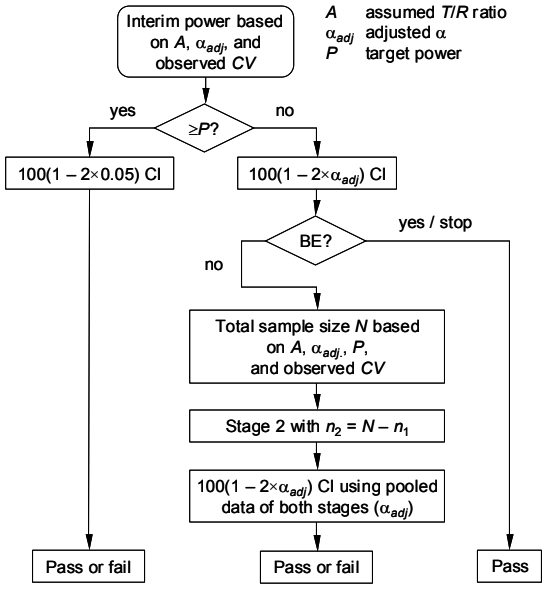


Fig. 2 'Type 2' TSD with conditionally unadjusted α in the first stage (variants of Potvin's 'Method C')

In both types futility rules can be introduced to allow early stopping. It should be noted that introducing a futility criterion which is not described in one of the frameworks will not negatively impact the type I error since the chance to proceed to the second stage will be lower than in the original method. However, especially introducing a maximum total sample size (12 – 14), *i.e.*, stopping in stage 1 if the re-estimated $N > N_{max}$) can render such studies unethical (16, 20). Futility criteria based on the observed T/R-ratio or its confidence interval are a better alternative (6, 19). Full adaptive designs (12 – 14), *i.e.*, re-estimating the sample size based on *both* the observed CV and T/R-ratio) may lead to extreme

sample sizes, which – together with a futility criterion on N_{max} – result in low power since the precision of the estimated T/R-ratio in the first stage is poor (21).

In the recent past the EMA's Biostatics Working Party (BSWP) questioned the validity of TSDs based on simulations in terms of control of the type I error. The package `AdaptiveBE` (22) for `R` (23) supports both applicants and regulatory assessors in *post hoc* exploring the empiric TIE by means of package `Power2Stage` (24).

The workflow is outlined in the following:

1. The study data have to be specified (stage 1 GMR, CV, sample size and – if applicable the same for the second stage). The applied type of the TSD (adjusted α, eventual futility rules) have to be given. If not specified in the protocol and/or report, by default power and sample size re-estimation is done by the noncentral *t*-approximation.
2. The empiric TIE is obtained by simulating one million TSD studies according to the conditions given above (significance limit 0.05036, standard error of the estimate 0.00016) under the true Null $\theta_0 = \theta_2$.
    a. If the empiric TIE is ≤0.05, the results of the study can be accepted as reported (see Example 1, p 4).
    b. If the empiric TIE is >0.05, the equation
    $$power(\alpha_{adj}, \Theta) - \alpha_{nominal} = 0$$
    is numerically solved (25) under the true Null for the study conditions Θ in the interval $\{tol, \alpha_{nominal}\}$, where the defaults are $tol = 10^{-8}$ and $\alpha_{nominal} = 0.05$.
    c. The study is recalculated with the adjusted α (interim, and – if applicable – the sample size re-estimation, and the final analysis).
        i. If results of both evaluations agree, the study can be accepted as reported. Although there might be an inflation of the TIE with the pre-specified α, none of the confidence limits is close to the acceptance range and thus, no relevant impact on the consumer risk is expected (see Example 2, p 5).
        ii. If results do not agree (*i.e.*, the study passes with the pre-specified α and fails with the adjusted α), the potential relative increase of the consumer risk is given (see Example 3, p 7).

Based on the study conditions the code 'guesses' which of the published frameworks might have been used. The frameworks were validated for a range of stage 1 sample sizes and CVs in the interim. Hence, if at least one of the two were outside the validated matrix of n_1/CV-combinations it shows a lack of understanding of the applicant. However, as long as the TIE is controlled, the study should still be acceptable.

Example 1. Data of Potvin *et al.*, Example 2.
'Method B': GMR 0.95, target power 0.80, α_1 0.0294, α_2 0.0294.
Stage 1 MSE 0.032634, ln(T)–ln(R) 0.08396, n_1 12.
Final MSE 0.045896, ln(T)–ln(R) 0.014439, N 20.

```
1       Data for the interim analysis
2       ----------------------------------------
3       CV (MSE)            :  18.21% (0.032634)
4       PE (ln(T)-ln(R))    : 108.76% (0.08396)
5       Sample size         :  12
6
7       Data for the final (pooled) analysis
8       ----------------------------------------
9       CV (MSE)            :  21.67% (0.045896)
10      PE (ln(T)-ln(R))    : 101.45% (0.014439)
11      Total sample size   :  20
12
13      Study conditions and assessment of empiric Type I Error
14      ------------------------------------------------------------
15      Design              : 2×2×2 crossover
16      TSD Type            : 1 (Potvin et al. 2008, Method B)
17      Target power        : 0.80
18      GMR used            : 0.95 (fixed)
19      Interim power check: yes
20      Futility criterion : none
21      Minimum n2          : not specified
22      Maximum N           : not specified
23      Specified α 1, 2    : 0.0294, 0.0294
24      Specified CIs       : 94.12%, 94.12%
25      TIE for specified α: 0.04307 (≤0.05)
26                           Applied adjustment is justified.
27
28      Interim analysis (specified α1 0.0294)
29      ----------------------------------------------------
30      94.12% CI: 92.93–127.28% (failed to demonstrate BE)
31      Power    : 0.5049 (approx. via shifted central t)
32      Second stage with 8 subjects (N=20) is justified.
33
34      Power based on interim data (specified α)
35      ----------------------------------------------------
36      Method              : approx. via shifted central t
37      Stage 1             : 0.5248
38      Both stages         : 0.8560
39      Studies in stage 2 : 44.2%
40      Expected total sample size (N)
41        Average           : 17.5
42        Median            : 12
43        5, 95 percentiles: 12, 34
44
45      Final analysis of pooled data (specified α2 0.0294)
46      ====================================================
47      94.12% CI: 88.45–116.38% (BE concluded)
48      Post hoc power (irrelevant; for validation purposes)
49      Based on GMR        : 0.6627
50      Based on PE         : 0.7685
51
52
53      | Since no inflation of the Type I Error is expected,
54      | can accept the reported analysis.
55
```

In 'Type 1' TSDs BE is assessed with the adjusted α in the interim first and then power. Since the study failed to demonstrate BE (line 30) and power is lower than the target 0.8 (line 31), the second stage can be initiated (line 32). Otherwise, the study should have stopped already in the interim. The code estimates the sample size of the second stage (based on the GMR, target power and α_2). Lines 34–43 give the result of simulating power (argument `pa=TRUE`). The average (total) sample size (called ASN by some authors) is 17.5. With the default setting (`pa=FALSE`) this part is not shown.

In the final analysis BE is demonstrated. *Post hoc* power is only given to compare the result with the reference (with the default setting this part is not shown). The assessment is given in the box.

Example 2. Data from above but 'Method C': α_0 0.05, $\alpha_1 = \alpha_2$ 0.0294.

```
Study conditions and assessment of empiric Type I Error
──────────────────────────────────────────────────────────────────────
Design              : 2×2×2 crossover
TSD Type            : 2 (Potvin et al. 2008, Method C)
Target power        : 0.80
GMR used            : 0.95 (fixed)
Interim power check: yes
Futility criterion : none
Minimum n2          : not specified
Maximum N           : not specified
Specified α 1, 2    : 0.050|0.0294, 0.0294
Specified CIs       : 90.00%|94.12%, 94.12%
TIE for specified α: 0.05062 (>0.05)
                      Applied adjustment is not justified.

Interim analysis (specified α1 0.0294)
──────────────────────────────────────────────────────
94.12% CI: 92.93–127.28% (failed to demonstrate BE)
Power    : 0.6494 (approx. via shifted central t)
Second stage with 8 subjects (N=20) is justified.

Power based on interim data (specified α)
──────────────────────────────────────────────────────
Method              : approx. via shifted central t
Stage 1             : 0.5449
Both stages         : 0.8635
Studies in stage 2 : 40.6%
Expected total sample size (N)
  Average           : 17.4
  Median            : 12
  5, 95 percentiles: 12, 34

Final analysis of pooled data (specified α2 0.0294)
══════════════════════════════════════════════════════
94.12% CI: 88.45–116.38% (BE concluded)
Post hoc power (irrelevant; for validation purposes)
Based on GMR        : 0.6538
Based on PE         : 0.7601

α-optimization (objective function: TIE - 0.05 → 0)
──────────────────────────────────────────────────────────────
Method              : approx. via shifted central t
Convergence         : 18 iterations (run-time 5.15 min)
Estimated precision: 5.07E-09
Adjusted α 1, 2     : 0.050|0.02858, 0.02858
Adjusted CIs        : 90.00%|94.28%, 94.28%
TIE for adjusted α : 0.04992 (n.s. >0.05)

Interim analysis (adjusted α1 0.02858)
──────────────────────────────────────────────────────
94.28% CI: 92.82–127.44% (failed to demonstrate BE)
Power    : 0.6494 (approx. via shifted central t)
Second stage with 8 subjects (N=20) is justified.

Power based on interim data (adjusted α)
──────────────────────────────────────────────────────
Method              : approx. via shifted central t
Stage 1             : 0.5387
Both stages         : 0.8639
Studies in stage 2 : 41.2%
Expected total sample size (N)
  Average           : 17.5
  Median            : 12
  5, 95 percentiles: 12, 34

Final analysis of pooled data (adjusted α2 0.02858)
══════════════════════════════════════════════════════
94.28% CI: 88.36–116.49% (BE concluded)
Post hoc power (irrelevant; for validation purposes)
Based on GMR        : 0.6627
Based on PE         : 0.7685


 Since conclusions of both analyses agree,
 can accept the original analysis.
```

In ‘Type 2’ TSDs power in the interim is assessed first. If power is at least the target (here 0.8), this implies that the assumptions (CV, GMR) which lead to the sample size of the first stage seemingly are correct. According to the framework in this case no adjustment has to be done (BE can be assessed with α_0 0.05) since the study will stop in the interim (pass or fail). In the example power is less than the target (line 19) and therefore, BE must be assessed with α_1 0.0294. The study failed to demonstrate BE in the interim (line 18), and therefore, the second stage can be initiated (line 20).

Since an inflation of the TIE (0.05062) is expected, α is optimized (lines 40–47). With an α_2 of 0.02858 the TIE is controlled (0.04992). The interim with this α justifies a second stage as well (lines 49–53).

In the final analysis with the specified α 0.0294 (lines 33–35) BE is easily demonstrated (CI well within the acceptance range). Repeating the final analysis with the adjusted α_2 0.02858 (lines 66–68) shows BE as well. The assessment is given in the box.

Example 3. Montague *et al.* Method D: GMR 0.90, target power 0.80, α_0 0.05, $\alpha_1 = \alpha_2$ 0.0280.
Stage 1 CV 20%, PE 0.92, n_1 12.
Final CV 23.315%, PE 0.88, N 45 (estimated 46; but one dropout in the second stage). Only part of the output is shown below.

```
1       Design              : 2×2×2 crossover
2       TSD Type            : 2 (Montague et al. 2011, Method D)
3       Target power        : 0.80
4       GMR used            : 0.90 (fixed)
5       Interim power check: yes
6       Futility criterion : none
7       Minimum n2          : not specified
8       Maximum N           : not specified
9       Specified α 1, 2    : 0.050|0.0280, 0.0280
10      Specified CIs       : 90.00%|94.40%, 94.40%
11      TIE for specified α: 0.05153 (>0.05)
12                           Applied adjustment is not justified.
13
14      Interim analysis (specified α1 0.028)
15      ──────────────────────────────────────────────────
16      94.40% CI: 77.25–109.57% (failed to demonstrate BE)
17      Power    : 0.3407 (approx. via shifted central t)
18      Second stage with 34 subjects (N=46) is justified.
19
20      Final analysis of pooled data (specified α2 0.028)
21      ══════════════════════════════════════════════════
22      94.40% CI: 80.00–96.80% (BE concluded)
23
24      α-optimization (objective function: TIE - 0.05 → 0)
25      ──────────────────────────────────────────────────
26      Method              : approx. via shifted central t
27      Convergence         : 19 iterations (run-time 5.56 min)
28      Estimated precision: 5.18E-09
29      Adjusted α 1, 2     : 0.050|0.02709, 0.02709
30      Adjusted CIs        : 90.00%|94.58%, 94.58%
31      TIE for adjusted α : 0.04998 (n.s. >0.05)
32
33      Interim analysis (adjusted α1 0.02709)
34      ──────────────────────────────────────────────────
35      94.58% CI: 77.13–109.74% (failed to demonstrate BE)
36      Power    : 0.3407 (approx. via shifted central t)
37      Second stage with 34 subjects (N=46) is justified.
38
39      Final analysis of pooled data (adjusted α2 0.02709)
40      ══════════════════════════════════════════════════
41      94.58% CI: 79.94–96.88% (failed to demonstrate BE)
42      Post hoc power (irrelevant; for validation purposes)
43      Based on GMR        : 0.6716
44      Based on PE         : 0.4937
45
46
47        Accepting the reported analysis could in-
48        crease the relative consumer risk by ~3.1%.
49
```

This example represents one of the borderline cases; the reported lower confidence limit in the final analysis (lines 20–22) is at the acceptance range. CV 20% and n_1 12 is the location of the maximum inflation of the TIE of this method (both according to the authors' results and obtained by the R-package `Power2Stage`).

Conclusions in the final analyses do *not* agree (the study *passes* with the 94.40% CI but *fails* with the 94.58% CI). See lines 22, 41, and the assessment given in the box.

Note that this represents also a case where already the mandatory rounding of the CI according to the guideline (1) slightly inflates the TIE (even in fixed sample designs). The 94.40% CI is actually 79.99842–96.80191%. The study passes only due to rounding the lower confidence limit up to 80.00%.

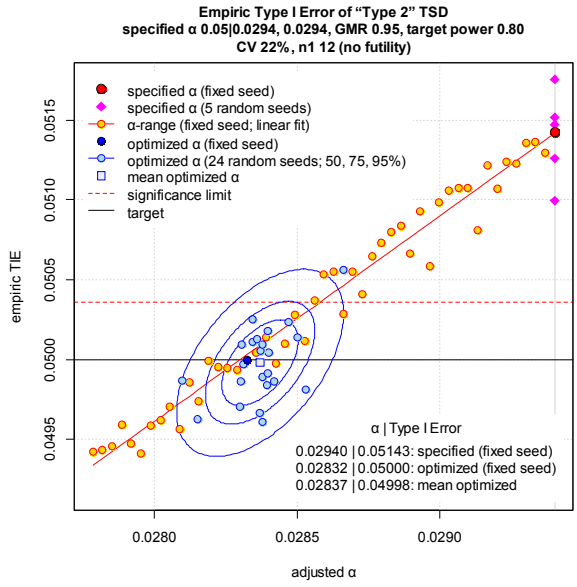
In the references the power approximation by the shifted central *t*-distribution was used for speed reasons (tenths of millions of studies had to be simulated). In actual studies likely the approximation by the noncentral *t*-distribution or even the exact method (Owen's Q) will be used. Both algorithms are available in commercial (`SAS`, `NQuery`) and open-source (`R` package `PowerTOST` (26)) software. The former is also implemented in `PASS`. In the following a comparison of results:

example	TIE by method			adjusted α by			agreement†		
	a	b	c	a	b	c	a	b	c
1	0.04307	0.04269	0.04287	not necessary					
2	0.05062	0.05083	0.05087	0.02858	0.02856	0.02856	yes	yes	yes
3	0.05153	0.05180	0.05180	0.02709	0.02704	0.02704	no	no	no

† agreement in conclusions: analyses by optimized α (if necessary) *vs.* the pre-specified α.
a shifted central *t*-approximation
b noncentral *t*-approximation
c exact
Note: If an inflated TIE is detected with the exact method, adjusting α can take some [*sic*] hours.

It should be noted that in simulations of the references always *exactly* the re-estimated stage 2 sample size n_2 was used. Naturally, if in a study more subjects are dosed in the second stage (based on an assumed dropout-rate) and at the end of the study more than n_2 subjects are eligible, the chance to demonstrate BE increases and thus, potentially the TIE. Therefore, especially in such cases assessing the TIE is recommended.

Can we expect that the Type I Error after optimization will *always* be ≤0.05 in simulations? Only if we use a *fixed* seed of the (pseudo) random generator – which is generally recommended in simulations for reproducibility. Let us explore the location of the maximum TIE for Potvin's 'Method C' (CV 22%, n_1 12) with power and sample size re-estimation by the noncentral *t*-approximation.

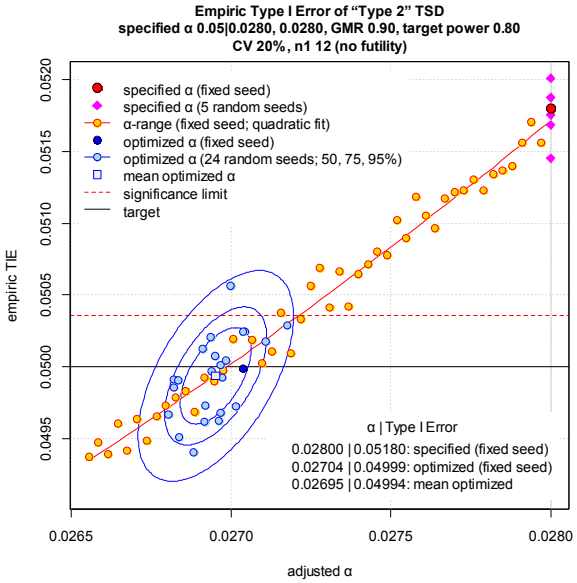


With the specified α 0.0294 we obtain an empiric TIE of 0.05143 (red circle). If we repeat the estimation with *random* seeds we get the magenta rhombi. If we assess lower alphas (*i.e.*, would adjust more), naturally the TIE decreases (yellow circles). With a fixed seed (blue circle) we get an optimized α of 0.02832 (TIE 0.05000) which is far below the significance limit for one million simulations (0.05036, binomial test) and – hopefully – acceptable for the EMA's BSWP.

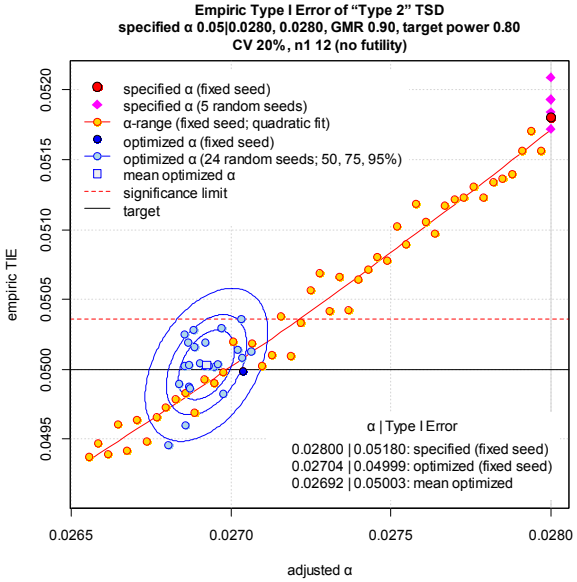
Now we repeat the optimization with random seeds and get the cluster of lightblue circles. The blue lines give the 50, 75, and 95 percentile ellipses (based on the bivariate normal distribution). Of course,

one could use their mean (the square) as a 'best' estimate (here 0.02832), but would that really help? First, it is not reproducible any more (for every run of the code one will get another value) and second there is no guarantee that the TIE will be always ≤0.05. More about it in the next example. The run-time is demanding (almost three hours on my machine) and I do not think that one gets a substantial gain.

This example assesses the maximum TIE for 'Method D' (10; at CV 20%, n_1 12) with the noncentral *t*-approximation.



With the specified α 0.0280 we obtain an empiric TIE of 0.05180. Note that this 'exact' match with the reported 0.0518 is due to chance. Here we are using the noncentral *t*-approximation whereas (10) used the shifted central *t*-approximation. Using the same we would get at TIE of 0.05153 (see Example 3 above). Since the seed is not given in the reference, differences are to be expected. As in the previous example magenta rhombi show results with random seeds. The mean of optimized alphas is slightly lower (0.02695, TIE 0.04994) than the first estimate (0.02704, TIE 0.04999), but the mean TIE can be >0.05 as shown in yet another run.



Validation

		interim						final					
e	T	reported			check.TSD()			reported			check.TSD()		
		CI		P	CI		P	CI		P	CI		P
1	1	104.27	134.17	75.6	104.27	134.17	75.61	102.83	129.71	NR	102.83	129.70	82.17
2	2	106.26	131.66	84.1	106.26	131.66	84.12	NP (failed in the interim)					
3	1	92.93	127.28	50.5	92.93	127.28	50.49	88.45	116.38	66.3	88.45	116.38	66.27
4	2	92.93	127.28	64.9	92.93	127.28	64.94	88.45	116.38	66.3	88.45	116.38	66.27
5	2	NR			77.25	109.57	34.07	NR			80.00	96.80	67.71
6	1	NR			67.27	131.41	55.90	NR			80.37	105.31	64.75
7	1	NR			68.21	132.30	00.00	NR			84.67	106.59	80.06
8	1	78	114	35.8	78.18	113.74	35.66	91	112	NR	91.04	112.05	88.04
9	2	78	114	48.7	78.25	113.65	45.57	91	112	NR	90.98	112.12	87.83
10	1	NR			78.65	109.97	40.92	NR			82.27	102.87	84.51
11	2	NR			73.75	105.00	34.07	NR			82.42	96.11	85.07
12	2	NR			83.07	108.65	76.97	NR			84.47	106.84	80.17
13	2	NR			79.96	112.87	56.50	NR			81.26	104.05	76.97
14	2	99.07	111.85	99.90	99.07	111.85	99.90	NP (passed in the interim)					

e Number of internal validation example: `check.TSD(valid=TRUE, expl=e)`
T Type of design
CI 100(1–2α) confidence interval
P Power (in the interim or *post hoc*)
NR Not reported
NP Not performed

References

1 European Medicines Agency, Committee for Medicinal Products for Human Use. *Guideline on the investigation of bioequivalence.* London, 20 January 2011: CPMP/EWP/QWP/1401/98 Rev. 1.

2 Armitage P, McPherson CK, Rowe BC. *Repeated significance tests on accumulating data.* J R Stat Soc Ser A. 1969; 132: 235–44.

3 Pocock SJ. *Group sequential methods in the design and analysis of clinical trials.* Biometrika. 1977; 64: 191–99.

4 O'Brien PC, Fleming TR. *A multiple testing procedure for clinical trials.* Biometrics. 1979; 35: 549–556.

5 Pocock SJ. *Interim analyses for randomized clinical trials: the group sequential approach.* Biometrics. 1982; 38: 153–62.

6 Armitage P. *Interim Analysis in Clinical Trials.* Stat Med. 1991; 10(6): 925–37. doi:10.1002/sim.4780100613.

7 Gould AL. *Group sequential extension of a standard bioequivalence testing procedure.* J Pharmacokinet Biopharm. 1995; 23: 57–86. doi:10.1007/BF02353786.

8 Hauck WW, Preston PE, Bois FY. *A group sequential approach to crossover trials for average bioequivalence.* J Biopharm Stat. 1997; 71: 87–96. doi:10.1080/10543409708835171.

9 Potvin D, DiLiberti CE, HauckWW, Parr AF, Schuirmann DJ, Smith RA. *Sequential design approaches for bioequivalence studies with crossover designs.* Pharm Stat. 2008; 7(4): 245–62. doi:10.1002/pst.294.

10 Montague TH, Potvin D, DiLiberti CE, Hauck WW, Parr AF, Schuirmann DJ. *Additional results for 'sequential design approaches for bioequivalence studies with crossover designs'.* Pharm Stat. 2011; 11(1): 8–13. doi:10.1002/pst.483.

11 Fuglsang A. *Sequential bioequivalence trial designs with increased power and controlled type I error rates.* AAPS J. 2013; 15(3): 659–61. doi:10.1208/s12248-013-9475-5.

12 Karalis V, Macheras P. *An insight into the properties of a two-stage design in bioequivalence studies.* Pharm Res. 2013; 30(7): 1824–35. doi:10.1007/s11095-013-1026-3.

13 Karalis V. *The role of the upper sample size limit in two-stage bioequivalence designs.* Int J Pharm. 2013; 456(1): 87–94. doi:10.1016/j.ijpharm.2013.08.013.

14 Karalis V, Macheras P. *On the statistical model of the two-stage designs in bioequivalence assessment.* J Pharm Pharmacol. 2014; 66(1): 48–52. doi:10.1111/jphp.12164.

15 Fuglsang A. *Sequential bioequivalence approaches for parallel designs.* AAPS J. 2014; 16(3): 373–8. doi:10.1208/s12248-014-9571-1.

16 Fuglsang A. *Futility rules in bioequivalence trials with sequential designs.* AAPS J. 2014; 16(1): 79–82. doi:10.1208/s12248-013-9540-0.

17 Zheng Ch, Zhao L, Wang J. *Modifications of sequential designs in bioequivalence trials*. Pharm Stat. 2015; 14(3) :180–8. doi:10.1002/pst.1672.

18 Kieser M, Rauch G. *Two-stage designs for cross-over bioequivalence trials.* Stat Med. 2015; 34(16): 2403–16. doi:10.1002/sim.6487.

19 Xu J, Audet C, DiLiberti CE, Hauck WW, Montague TH, Parr TH, Potvin D, Schuirmann DJ. *Optimal adaptive sequential designs for crossover bioequivalence studies.* Pharm Stat. 2015; 15(1): 15–27. doi:10.1002/pst.1721.

20 Schütz H. *Two-stage designs in bioequivalence trials.* Eur J Clin Pharmacol. 2015; 71(3): 271–81. doi:10.1007/s00228-015-1806-2.

21 Tsiatis AA, Mehta C. *On the inefficiency of the adaptive design for monitoring clinical trials.* Biometrika. 2003; 90(2): 367–78. doi:10.1093/biomet/90.2.367.

22 Schütz H. *AdaptiveBE: Check Acceptability of Adaptive Bioequivalence Studies.* 2017; R package version 0.8.3.9000. https://github.com/Helmut01/AdaptiveBE

23 R Core Team. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna, Austria: 2017; https://www.r-project.org/

24 Labes D, Lang B, Schütz H. *Power2Stage: Power and Sample-Size Distribution of 2-Stage Bioequivalence Studies.* 2017; R package version 0.4.5.9000. https://CRAN.R-project.org/package=Power2Stage

25 Brent RP. *Algorithms for minimization without derivatives.* Mineola: Dover Publications; 2003.

26 Labes D, Schütz H, Lang B. *PowerTOST: Power and Sample Size Based on Two One-Sided t-Tests (TOST) for (Bio)Equivalence Studies.* 2017; R package version 1.4.6.9000. https://cran.r-project.org/package=PowerTOST