Novel approaches in adaptive designs and α adjustment, e.g., with futility criteria and for parallel design studies

Helmut Schütz
BEBAC, Vienna, Austria
Institute of Medical Statistics, Vienna, Austria

Simulation-based (2×2×2 crossover)

No futility criteria, unlimited stage 2 sample size

Reference	Type	Method	GMR	Power (%)	*CV* (%)	α_{adj}	TIE_{max}
Potvin *et al.* [3]	1	B	0.95	80	10 – 100	0.0294	0.0489
	2	C					0.0512*
Montague *et al.* [4]	2	D	0.90			0.0280	0.0517*
Fuglsang [5]	1	B	0.95	90	10 – 80	0.0284	0.0497
	2	C/D				0.0274	0.0501
	2	C/D	0.90			0.0269	0.0503

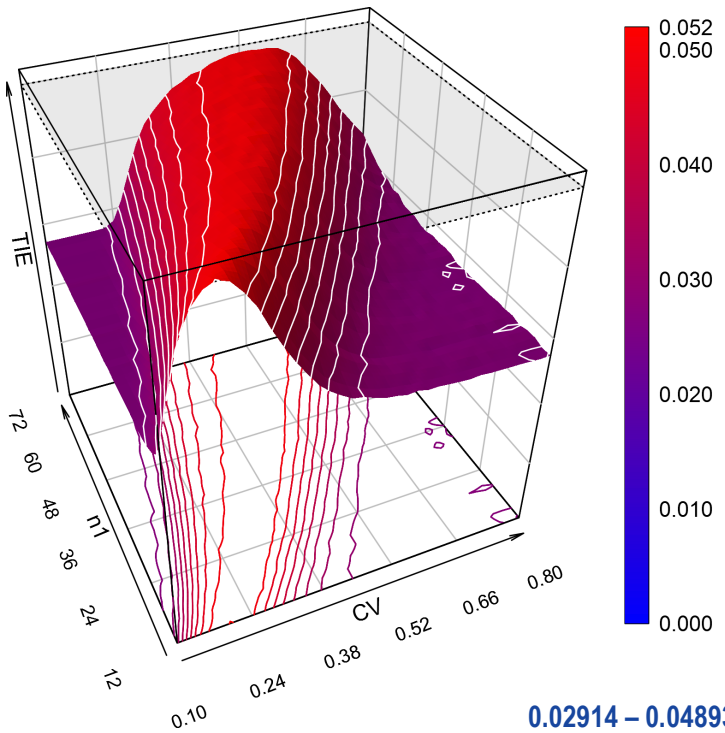
Futility criteria on CI, N_{max} 42 (low *CV*), 180 (high *CV*)

Reference	Type	Method	GMR	Power (%)	*CV* (%)	Futility region	α_{adj1}	α_{adj2}	TIE_{max}
Xu *et al.* [11]	1	E	0.95	80	10 – 30	0.9374 – 1.0667	0.0249	0.0363	0.0490
	2	F				0.9492 – 1.0535	0.0248	0.0364	0.0496
	1	E			30 – 55	0.9305 – 1.0747	0.0254	0.0357	0.0453
	2	F				0.9350 – 1.0695	0.0259	0.0349	0.0455

Operating Characteristics (Type 1 TSD [3])

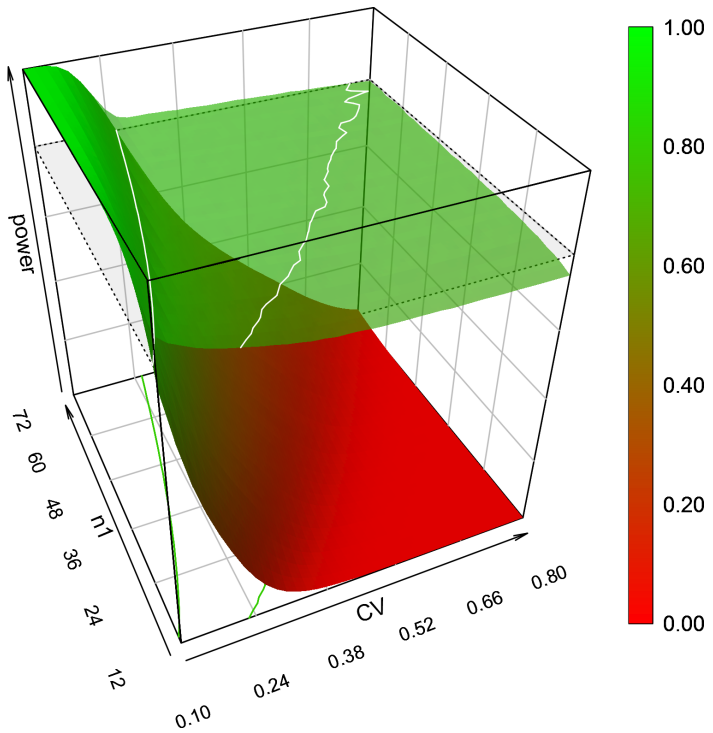
GMR 0.95, power 80%, α_{adj} 0.0294 (Potvin *et al.* 'Method B')

Type I Error



0.02914 – 0.04893

Power



Exact (2×2×2 crossover)

Repeated Confidence Intervals / Inverse Normal Method [15–18] **adapted for BE** [19–21]; **aka »Maurer's method«** [21]

- **Controls the Type I Error in the strict sense**
  - **Analytically proven**
  - **Confirmed in simulations**
- **Two approaches**
  - **Standard Combination Method**
  - **Maximum Combination Test (recommended** [20,21])
- **Weights of the stages have to be pre-specified**
  - **The adjusted α depends on the weights**
    - **The more weights differ, the more adjustment**
    - **Robust against misspecification**
    - **The same adjusted α is used in both stages**
  - **Simulations can be performed to find suitable weights**

Exact (2×2×2 crossover)

»Maurer's method« [21]

- My recommendations
  - Stage 1 sample size
    - 80% of fixed sample design for assumed GMR and *CV*
      - » Reasonably high probability to stop already in stage 1 for BE
      - » Overall power higher than fixed sample design
  - If *CV* expected to be not more than 75% larger than assumed
    - Standard Combination Method more powerful, requires less adjustment [22]
- Sample size re-estimation
  - A fixed GMR is used by default
  - Can be fully adaptive, i.e., based on the GMR of stage 1
- Minimum and maximum stage 2 sample sizes can be pre-specified
  - At least four subjects in two sequences are required in stage 2
  - Too small stage 2 negatively affects power

Exact (2×2×2 crossover)

»Maurer's method« [21]

- Optional futility criteria for stopping in the interim
  - GMR outside specified limits (default 0.80 – 1.25)
  - 90% confidence interval of the GMR entirely outside specified limits (default 0.95 – 0.95^{-1})
  - Maximum total sample size (default $4 \times n_1$)
- Futility criteria
  - Combinations are possible
  - Reduce the Type I Error
  - Negatively affect power
    - Fairly robust on GMR or CI
    - Very sensitive on N_{max}
  - Simulations highly recommended

Operating Characteristics (Exact [21])

Fixed GMR 0.95 (*CV* 0.1–0.8, n_1 12–72)

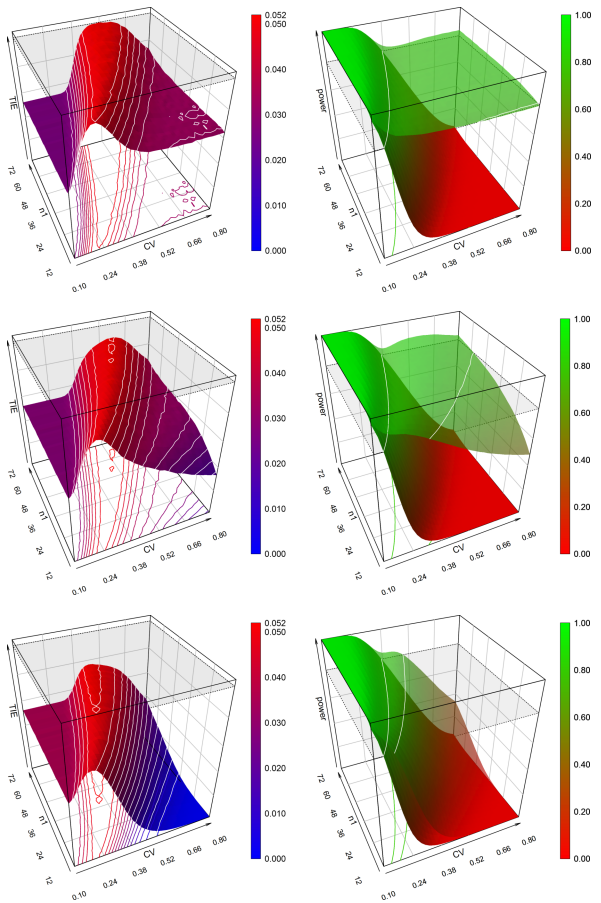
- No futility
- Type I Error 0.02598 – 0.04995

Adaptive GMR (*CV* 0.1–0.8, n_1 12–72)

- Futility on CI (outside 0.95 – 0.95^{-1})
- Type I Error 0.01678 – 0.04523

Adaptive GMR (*CV* 0.1–0.8, n_1 12–72)

- Futility on CI (outside 0.95 – 0.95^{-1})
- Futility on N_{max} ($> 4 \times n_1$)
- Type I Error 0.00006 – 0.03838



Exact (2×2×2 crossover)

»Maurer's method« [21]

- Contrary to simulation-based methods, data are not pooled
  - Stages are evaluated separately and assessed by (repeated) confidence intervals
    - ANOVA (EMA and most other juridictions)
    - Mixed-effects model (FDA, Health Canada, China's CDE)
    - Additional factors, e.g., for multi-group or multi-site studies can be incorporated in the model
  - Required in the interim analysis by Power2Stage [35]
    - GMR_1, CV_1, n_1, assumed GMR for sample size re-estimation, target power
    - If additional factors in the model: df_1, SEM_1
  - Required in the final analysis by Power2Stage [35]
    - GMR_1, CV_1, n_1
    - GMR_2, CV_2, n_2
    - If additional factors in the model: df_1, SEM_1, df_2, SEM_2

Simulation-based (parallel design [14])

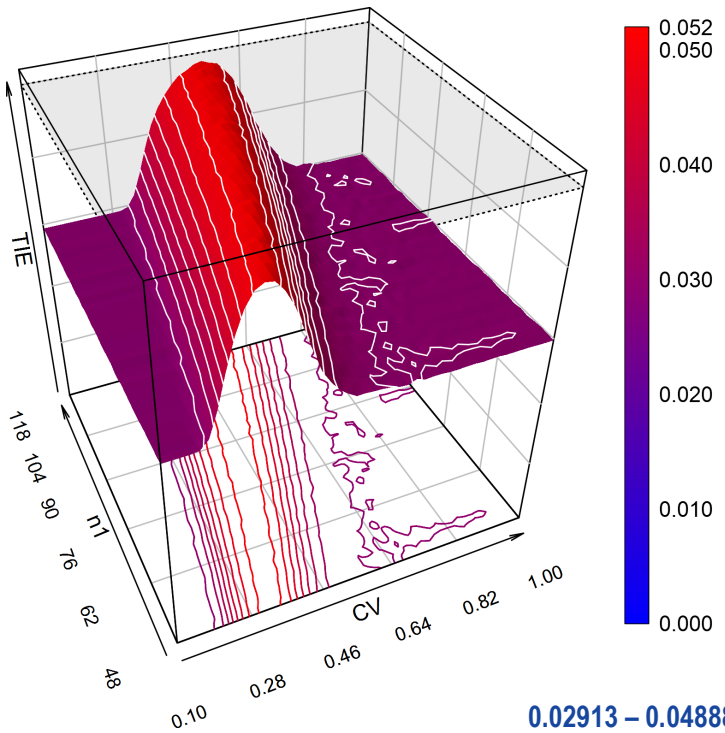
Two methods implemented in Power2Stage [35]

- ANOVA
  - Incorporates a term for stage in the final analysis
    - In line with the EMA's Q&A-document [28]
  - Assumes equal variances
    - Liberal and hence, not recommended
- Welch-Satterthwaite test
  - Approximates degrees of freedom for unequal group sizes and variances
    - No stage term possible; contradicts the EMA's Q&A-document [28]
    - In line with the FDA's guidance [33]
    - Highly recommended
- Modifications (simulations recommended)
  - Fully adaptive, i.e., based on the GMR observed in stage 1
  - Futility criterion on N_{max}

Operating Characteristics (Type 1 TSD [14])

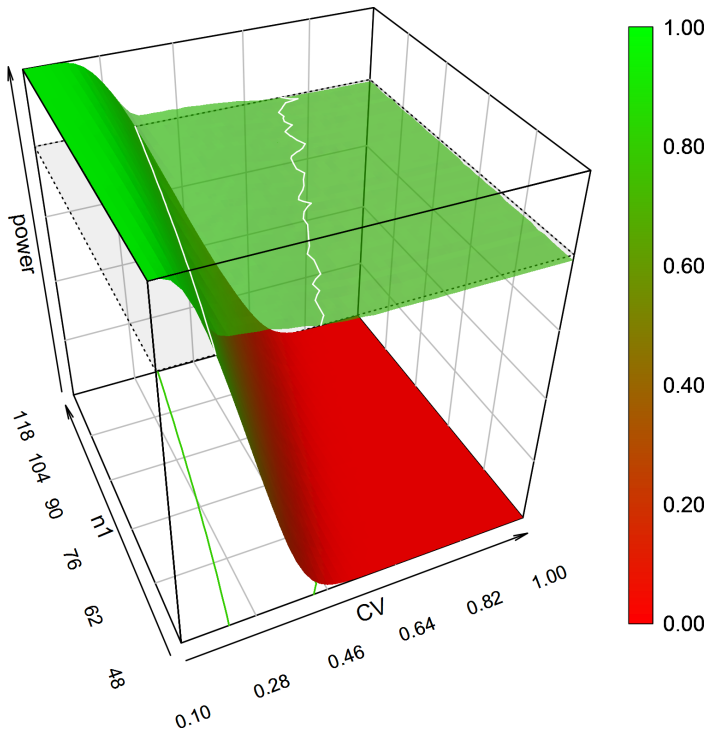
GMR 0.95, power 80%, α_{adj} 0.0294, Welch-Satterthwaite test

Type I Error



0.02913 – 0.04888

Power



State of Affairs

Simulation-based methods for 2×2×2 crossovers [3–13]

- **Ambiguous description in the EMA's guideline, regrettably incurred in other jurisdictions**
  - **Resulted in – unsubstantiated – deficiency letters**
  - **Wariness in the industry about application of adaptive designs**

    There are actually more articles describing the theoretical and statistical base for the application of the two-stage design than there are reported studies. [26]
  - **Type 2 TSD recommended by the FDA and Health Canada** [23,32]

Exact method for 2×2×2 crossovers [21] preferable

- **Strict Type I Error control**
- **Flexible (fully adaptive, futility criteria)**

Simulation-based method for parallel designs [14]

- **At the time being the only available**

Outlook

Expand the exact method for 2×2×2 crossovers

- If a PK metric in the first stage is highly variable, perform the second stage in a replicate design intended for reference-scaling
  - Scaling based on CV_{wR} in the second stage

Development of an exact method for parallel designs

- Not trivial because unequal sample sizes and variances have to be taken into account

Simulation-based methods for RSABE/ABEL

- Practically impossible
  - Stable sample size estimation requires 10^5 simulations taking conditions of the regulatory frameworks into account
  - 10^6 simulations to demonstrate control of the Type I Error
  - With a reasonable narrow grid of n_1 / CV-combinations estimated runtime ≈50 years 24 / 7 on a current workstation

Novel approaches in adaptive designs and α adjustment, e.g., with futility criteria and for parallel design studies

Thank You!

Helmut Schütz
BEBAC
1070 Vienna, Austria
helmut.schuetz@bebac.at
Institute of Medical Statistics, Medical University of Vienna
1090 Vienna, Austria
helmut.schuetz@meduniwien.ac.at

References

Simulation-based sequential Two-Stage Designs

- **2×2×2 crossover design**

1. Gould AL. Group Sequential Extension of a Standard Bioequivalence Testing Procedure. J Pharmacokinet Biopharm. 1995; 23(1): 57–86. https://doi.org/10.1007/BF02353786.
2. Hauck WW, Preston PE, Bois FY. A Group Sequential Approach to Crossover Trials for Average Bioequivalence. J Biopharm Stat. 1997; 7(1): 87–96. https://doi.org/10.1080/10543409708835171.
3. Potvin D, DiLiberti CE, Hauck WW, Parr AF, Schuirmann DJ, Smith RA. Sequential design approaches for bioequivalence studies with crossover designs. Pharm Stat. 2008; 7(4): 245–262. https://doi.org/10.1002/pst.294.
4. 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. https://doi.org/10.1002/pst.483.
5. Fuglsang A. Sequential Bioequivalence Trial Designs with Increased Power and Controlled Type I Error Rates. AAPS J. 2013; 15(3): 659–61. https://doi.org/10.1208/s12248-013-9475-5.
6. Karalis V, Macheras P. An Insight into the Properties of a Two-Stage Design in Bioequivalence Studies. Pharm Res. 2013; 30(7): 1824–1835. https://doi.org/10.1007/s11095-013-1026-3.
7. Karalis V. The role of the upper sample size limit in two-stage bioequivalence designs. Int J Pharm. 2013; 456(1): 87–84. https://doi.org/10.1016/j.ijpharm.2013.08.013.
8. Fuglsang A. Futility Rules in Bioequivalence Trials with Sequential Designs. AAPS J. 2014; 16(1): 79–82. https://doi.org/10.1208/s12248-013-9540-0.
9. Karalis V, Macheras P. On the Statistical Model of the Two-Stage Designs in Bioequivalence Assessment. J Pharm Pharmacol. 2014; 66(1): 48–52. https://doi.org/10.1111/jphp.12164.
10. Zheng Ch, Zhao L, Wang J. Modifications of sequential designs in bioequivalence trials. Pharm Stat. 2015; 14(3): 180–8. https://doi.org/10.1002/pst.1672.

References

11. Xu J, Audet C, DiLiberti CE, Hauck WW, Montague TH, Parr AF, Potvin D, Schuirmann DJ. Optimal adaptive sequential designs for crossover bioequivalence studies. Pharm Stat. 2016; 15(1): 15–27. https://doi.org/10.1002/pst.1721.
12. Molins E, Cobo E, Ocaña J. Two-stage designs versus European scaled average designs in bioequivalence studies for highly variable drugs: Which to choose? Stat Med. 2017; 36(30): 4777–88. https://doi.org/10.1002/sim.7452.
13. Molins E, Labes D, Schütz H, Cobo E, Ocaña J. An iterative method to protect the type I error rate in bioequivalence studies under two-stage adaptive 2×2 crossover designs. Biom J. 2021; 63(1): 122–33. https://doi.org/10.1002/bimj.201900388.

- **Parallel design**

14. Fuglsang A. Sequential Bioequivalence Approaches for Parallel Designs. AAPS J. 2014; 16(3): 373–8. https://doi.org/10.1208/s12248-014-9571-1.

Adaptive designs

- **Conditional power, repeated confidence intervals**

15. Bauer P, Köhne K. Evaluation of Experiments with Adaptive Interim Analyses. Biometrics. 1994; 50(4): 1029–41. https://doi.org/10.2307/2533441.
16. Lehmacher W, Wassmer G. Adaptive Sample Size Calculations in Group Sequential Trials. Biometrics. 1999; 55(4): 1286–90. https://doi.org/10.1111/j.0006-341X.1999.01286.x.
17. Bretz F, König F, Brannath W, Glimm E, Posch M. Tutorial in biostatistics: Adaptive designs for confirmatory clinical trials. Stat Med. 2009; 28(8): 1181–217. https://doi.org/10.1002/sim.3538.
18. Wassmer G, Brannath W. Group Sequential and Confirmatory Adaptive Designs in Clinical Trials. Switzerland: Springer; 2016. ISBN 978-3-319-32560-6. https://doi.org/10.1007/978-3-319-32562-0.

References

- **2×2×2 crossover design**

19. König F, Wolfsegger M, Jaki T, Schütz H, Wassmer G. Adaptive two-stage bioequivalence trials with early stopping and sample size re-estimation. Trials. 2015; 16 (Suppl 2): P218. https://doi.org/10.1186/1745-6215-16-S2-P218.
20. Patterson SD, Jones B. Bioequivalence and Statistics in Clinical Pharmacology. Boca Raton: Chapman & Hall / CRC Press; Second edition 2017. p. 141–87.
21. Maurer W, Jones B, Chen Y. Controlling the type 1 error rate in two-stage sequential designs when testing for average bioequivalence. Stat Med. 2018; 37(10): 1–21. https://doi.org/10.1002/sim.7614. Supplementary Information for "Controlling the Type I error rate in two-stage sequential adaptive designs when testing for Average Bioequivalence". https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002/sim.7614&file=sim7614-sup-0001-supplementary.pdf.
22. Lang B. Selection between Standard Combination Method and Maximum Combination Test based on a 'worst case' CV. Personal communication. 2020.

Reviews

23. García-Arieta A, Gordon J. Bioequivalence Requirements in the European Union: Critical Discussion. AAPS J. 2012; 14(4): 738–48. https://doi.org/10.1208/s12248-012-9382-1.
24. Davit B, Braddy AC, Conner DP, Yu LX. International Guidelines for Bioequivalence of Systemically Available Orally Administered Generic Drug Products: A Survey of Similarities and Differences. AAPS J. 2013; 15(4): 974–90. https://doi.org/10.1208/s12248-013-9499-x.
25. Schütz H. Two-stage designs in bioequivalence trials. Eur J Clin Pharmacol. 2015; 71(3): 271–81. https://doi.org/10.1007/s00228-015-1806-2.
26. Kaza M, Sokolvskyi A, Rudzki PJ. 10th Anniversary of a Two-Stage Design in Bioequivalence. Why Has it Still Not Been Implemented? Pharm Res. 2020; 37(7): Article number 140. https://doi.org/10.1007/s11095-020-02871-3.

References

27. Lee J, Feng K, Xu M, Gong X, Sun W, Kim J, Zhang Z, Wang M, Fang L, Zhao L. Applications of Adaptive Designs in Generic Drug Development. Clin Pharm Ther. 2020; 110(1): 32–5. https://doi.org/10.1002/cpt.2050.

Guidelines

28. EMA, CHMP. Guideline on the Investigation of Bioequivalence. CPMP/EWP/QWP/1401/98 Rev. 1/ Corr. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-investigation-bioequivalence-rev1_en.pdf. London. 20 January 2010.
29. EMA, CHMP. Questions & Answers: positions on specific questions addressed to the Pharmacokinetics Working Party (PKWP). EMA/618604/2008 Rev. 13. https://www.ema.europa.eu/en/documents/scientific-guideline/questions-answers-positions-specific-questions-addressed-pharmacokinetics-working-party_en.pdf. London. 19 November 2015.
30. EMA, CHMP. Guideline on multiplicity issues in clinical trials. Draft. EMA/CHMP/44762/2017. https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-multiplicity-issues-clinical-trials_en.pdf. London. 15 December 2016.
31. WHO Expert Committee on Specifications for Pharmaceutical Preparations. Multisource (generic) pharmaceutical products: Guidelines on registration requirements to establish interchangeability. WHO Technical Report Series No. 992, Annex 6. https://www.who.int/docs/default-source/medicines/norms-and-standards/guidelines/regulatory-standards/trs1003-annex6-who-multisource-pharmaceutical-products-interchangeability.pdf. Geneva. 2017.

Guidance

32. Health Canada. Conduct and Analysis of Comparative Bioavailability Studies. Cat: H13-9/6-2018E-PDF. https://www.canada.ca/content/dam/hc-sc/documents/services/drugs-health-products/drug-products/applications-submissions/guidance-documents/bioavailability-bioequivalence/conduct-analysis-comparative.pdf. Ottawa. Adopted 2012/02/08, revised 2018/06/08.

References

33. FDA, CDER. Statistical Approaches to Establishing Bioequivalence. https://www.fda.gov/media/70958/download. Rockville, MD. January 2001.
34. FDA, CDER/CBER. Adaptive Designs for Clinical Trials of Drugs and Biologics. https://www.fda.gov/media/78495/download. Silver Spring, MD. November 2019.

Packages for the open-source software R https://www.r-project.org/

- **Simulation-based (2×2×2 crossover** [3–12] **and parallel design** [14]**),**
  Exact (2×2×2 crossover design [21]**)**

35. Labes D, Lang B. Schütz H. Power2Stage: Power and Sample-Size Distribution of 2-Stage Bioequivalence Studies. 2021-11-20; Package version 0.5-4. https://cran.r-project.org/package=Power2Stage.

- **Simulation-based (2×2×2 crossover design** [13]**)**

36. Molins E, Labes D, Schütz H, Ocaña J. betsd: Adjusting significance levels in two-stage adaptive 2×2 cross-over designs. 2019; Package version 0.1.5. https://github.com/eduard-molins/betsd.

Maximum empiric Type I Error (10^6 simulations under the Null)

```
library(Power2Stage)
# Simulation-based (2×2×2 crossover, no futility); References 3, 4, 5
# Exact Power estimation by Owen's Q-function
method <- c("B", "C", "C", "B", "C", "C")
CV.loc <- c(0.24, 0.22, 0.20, 0.22, 0.10, 0.18)
n1.loc <- c(12, 12, 12, 12, 16, 12)
a      <- data.frame(Ref    = c(rep("[3]", 2), "[4]", rep("[5]", 3)),
                     Type   = c(1, 2, 2, 1, 2, 2),
                     Method = c("B ", "C ", "D ", "B ", "C/D", "C/D"),
                     GMR    = c(0.95, 0.95, 0.90, 0.95, 0.95, 0.90),
                     Power  = c(rep(80, 3), rep(90, 3)),
                     CV     = c(rep("10 - 100%", 3), rep("10  - 80%", 3)),
                     adj    = c(rep(0.0294, 2), 0.028, 0.0284, 0.0274, 0.0269))
for (i in 1:6) {
  a$TIE.max[i] <- power.tsd(method = method[i], alpha0 = 0.05, alpha = rep(a$adj[i], 2),
                            CV = CV.loc[i], n1 = n1.loc[i], GMR = a$GMR[i],
                            targetpower = a$Power[i] / 100, pmethod = "exact",
                            theta0 = 1.25)$pBE
}
a$Power   <- sprintf("%.0f%%", a$Power)
a$TIE.max <- round(a$TIE.max, 4)
print(a, row.names = FALSE)
```

Maximum empiric Type I Error (10^6 simulations under the Null)

```
library(Power2Stage)
# Simulation-based (2×2×2 crossover, futility criteria); Reference 11
# Exact Power estimation by Owen's Q-function
method  <- rep(c("B", "C"), 2)
CV.loc  <- c(rep(0.3, 2), rep(0.55, 2))
n1.loc  <- c(rep(18, 2), rep(48, 2))
fClower <- c(0.9374, 0.9305, 0.9492, 0.9350) # different futilities on CI
max.n   <- c(rep(42, 2), rep(180, 2))        # different futilities on Nmax
b       <- data.frame(Ref    = rep("[11]", 4), Type = c(1, 2, 1, 2),
                      Method = c("E  ", "F  ", "E  ", "F  "), GMR = rep(0.95, 4),
                      Power  = rep(80, 4), CV = c(rep("10 – 30%", 2), rep("30 – 55%", 2)),
                      adj1   = c(0.0249, 0.0248, 0.0254, 0.0259), # different alphas
                      adj2   = c(0.0363, 0.0364, 0.0357, 0.0349)) # in the stages
for (i in 1:4) {
  b$TIE.max[i] <- power.tsd.fC(method = method[i], alpha0 = 0.05,
                               alpha = c(b$adj1[i], b$adj2[i]), CV = CV.loc[i], n1 = n1.loc[i],
                               GMR = b$GMR[i], targetpower = b$Power[i] / 100,
                               max.n = max.n[i], fCrit = "CI", fClower = fClower[i],
                               pmethod = "exact", theta0 = 1.25)$pBE
}
b$Power   <- sprintf("%.0f%%", b$Power)
b$TIE.max <- round(b$TIE.max, 4)
print(b, row.names = FALSE)
```

Maximum empiric Type I Error (10^6 simulations under the Null)

```
library(Power2Stage)
# Exact (2×2×2 crossover): Maximum Combination Test, no futility criteria; Reference 21
# Exact Power estimation by Owen's Q-function
x       <- power.tsd.in(CV = 0.46, n1 = 62, weight = c(0.5, 0.25),
                        max.comb.test = TRUE, ssr.conditional = "error_power",
                        fCrit = "No", GMR = 0.95, targetpower = 0.8,
                        pmethod = "exact", theta0 = 1.25)
c       <- data.frame(Ref = "[21]", Method = "Maurer's", GMR = 0.95, Power = 80,
                      CV = "10 - 80%", adj = round(x$alpha[1], 5), TIE.max = x$pBE)
c$Power <- sprintf("%.0f%%", c$Power)
print(c, row.names = FALSE)

# Exact (2×2×2 crossover): Standard Combination Method, no futility criteria; Reference 21
# Exact Power estimation by Owen's Q-function
x       <- power.tsd.in(CV = 0.50, n1 = 70, weight = 0.5,
                        max.comb.test = FALSE, ssr.conditional = "error_power",
                        fCrit = "No", GMR = 0.95, targetpower = 0.8,
                        pmethod = "exact", theta0 = 1.25)
d       <- data.frame(Ref = "[21]", Method = "Maurer's", GMR = 0.95, Power = 80,
                      CV = "10 - 80%", adj = round(x$alpha[1], 5), TIE.max = x$pBE)
d$Power <- sprintf("%.0f%%", d$Power)
print(d, row.names = FALSE)
```

Exact (2×2×2 crossover)

Evaluation by Power2Stage [35]: Subjects 1–12 [3]

- **Interim analysis with fixed GMR 0.95**

```
interim.tsd.in(GMR1 = 1.0876, CV1 = 0.18213, n1 = 12, GMR = 0.95)
TSD with 2x2 crossover
Inverse Normal approach
 - Maximum combination test with weights for stage 1 = 0.5 0.25
 - Significance levels (s1/s2) = 0.02635 0.02635
 - Critical values (s1/s2) = 1.93741 1.93741
 - BE acceptance range = 0.8 ... 1.25
 - Observed point estimate from stage 1 is not used for SSR
 - With conditional error rates and conditional estimated target power
Interim analysis after first stage
 - Derived key statistics:
   z1 = 3.10000, z2 = 1.70344
   Repeated CI = (0.92491, 1.27891)
   Median unbiased estimate = NA
 - No futility criterion met
 - Test for BE not positive (not considering any futility rule)
 - Calculated n2 = 6
 - Decision: Continue to stage 2 with 6 subjects
```

Exact (2×2×2 crossover)

Evaluation by Power2Stage [35]: Subjects 1–12 | 13–18 [3]

- **Final analysis**

```
final.tsd.in(GMR1 = 1.0876, CV1 = 0.18213, n1 = 12,
             GMR2 = 1.0893, CV2 = 0.16776, n2 =  6)
TSD with 2x2 crossover
Inverse Normal approach
 - Maximum combination test with weights for stage 1 = 0.5 0.25
 - Significance levels (s1/s2) = 0.02635 0.02635
 - Critical values (s1/s2) = 1.93741 1.93741
 - BE acceptance range = 0.8 ... 1.25
Final analysis after second stage
 - Derived key statistics:
   z1 = 3.70299, z2 = 2.06106
   Repeated CI = (0.95672, 1.23796)
   Median unbiased estimate = 1.0953
 - Decision: BE achieved
```

  - Same conclusion as Potvin *et al.* 'Method B' [3] but slightly more conservative than its 94.12% CI with 0.9664 – 1.2252

Example: Weights and power in stages (minimum $n_2 = 6$, futility $N_{max} = 5 \times n_1$)

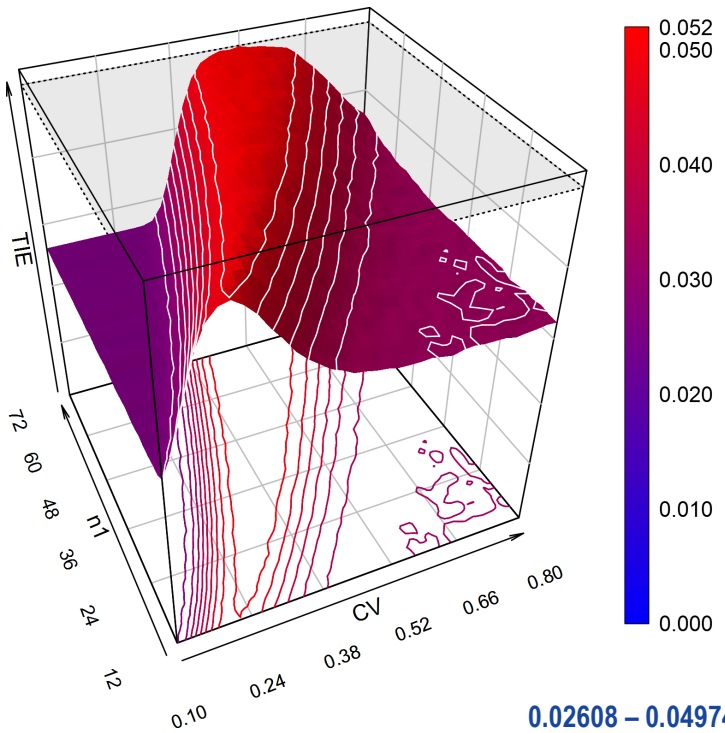
```
library(PowerTOST); library(Power2Stage)
CV  <- 0.3
n1  <- 0.8 * sampleN.TOST(CV = CV, print = FALSE)[["Sample size"]] # my recommendation
w1  <- w <- c(seq(0.9, 0.5, -0.1), 0.50)
w2  <- c(seq(0.1, 0.5, 0.1), 0.25)
MCT <- data.frame(Approach = "Maximum Combination Test",  w1 = w1, w2 = w2, n1 = n1)
SCM <- data.frame(Approach = "Standard Combination Method", w = w, n1 = n1)
for (i in seq_along(w1)) { # nmean is the expected average total sample size E[N]
  x           <- power.tsd.in(CV = CV, n1 = n1, weight = c(w1[i], w2[i]), GMR = 0.95,
                              min.n2 = 6, fCrit = "Nmax", fCNmax = 5 * n1)
  MCT$n2[i]   <- ceiling(x$nmean) - n1; MCT$adj[i] <- round(x$alpha[1], 5)
  MCT$pBE1[i] <- x$pBE_s1; MCT$pBE2[i] <- x$pBE
  x           <- power.tsd.in(CV = CV, n1 = n1, weight = w[i], max.comb.test = FALSE,
                              GMR = 0.95, min.n2 = 6,fCrit = "Nmax", fCNmax = 5 * n1)
  SCM$n2[i]   <- ceiling(x$nmean) - n1; SCM$adj[i] <- round(x$alpha[1], 5)
  SCM$pBE1[i] <- x$pBE_s1; SCM$pBE2[i] <- x$pBE
}
SCM <- SCM[-nrow(SCM), ]; MCT$Approach[2:nrow(MCT)] <- ""; SCM$Approach[2:nrow(SCM)] <- ""
print(MCT, row.names = FALSE, right = FALSE)
print(SCM, row.names = FALSE, right = FALSE)
```

- Similar average total sample sizes like in a fixed sample design
- Always higher power than fixed sample design
- ≈60% chance to stop in the interim for BE

Operating Characteristics (Exact [21])

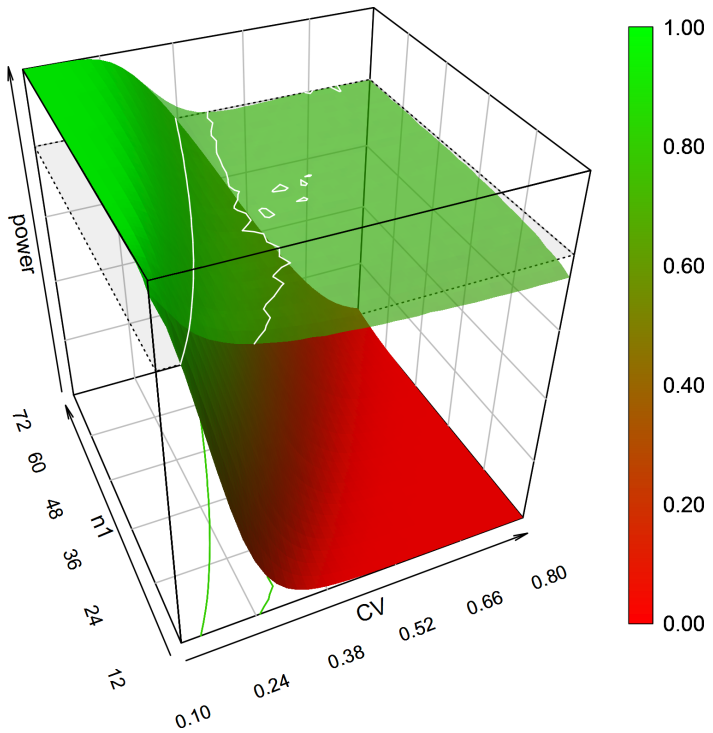
MCT: weights (0.25, 0.5) instead of the default (0.5, 0.25)

Type I Error



0.02608 – 0.04974

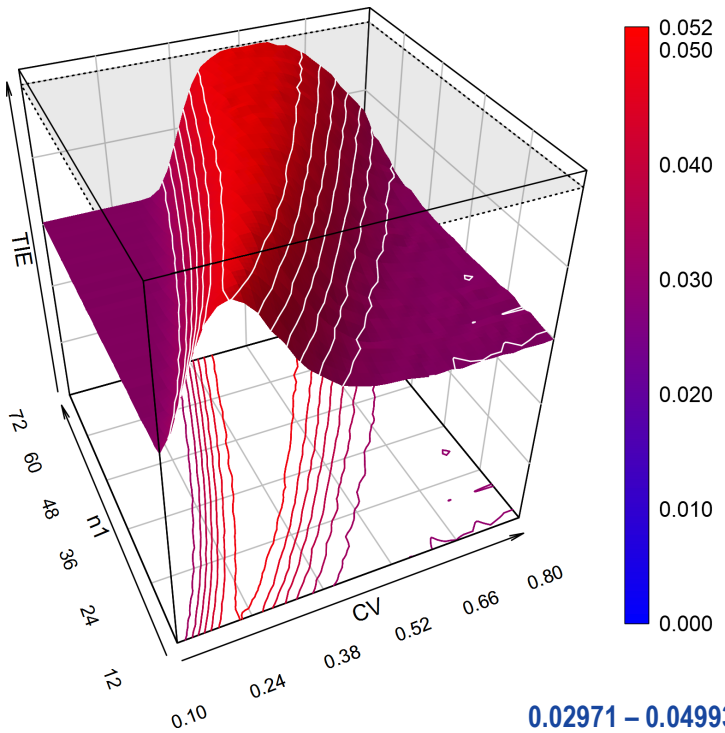
Power



Operating Characteristics (Exact [21])

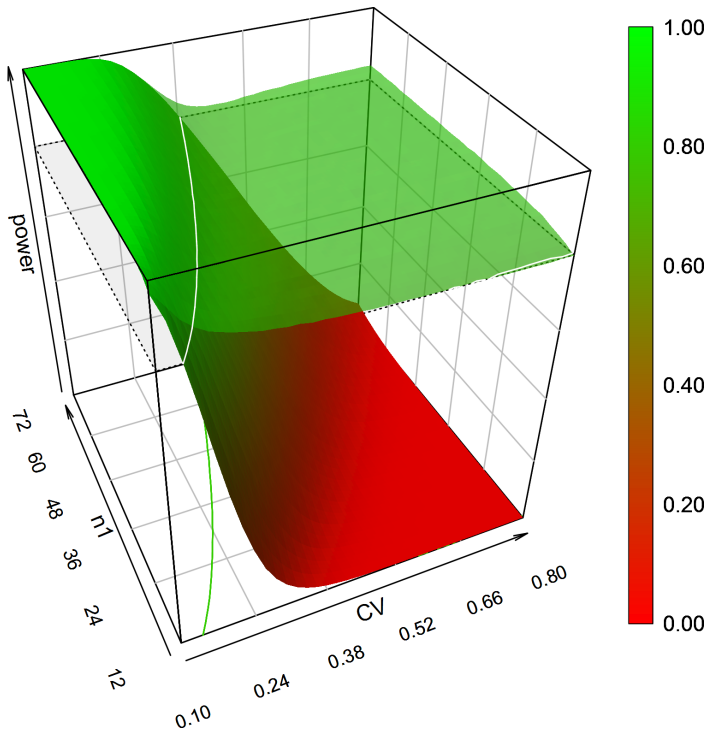
GMR 0.95, power 80%, α_{adj} 0.03037 (Standard Combination)

Type I Error



0.02971 – 0.04993

Power



Standard Combination Method
Minimum six subjects in stage 2

- **Interim analysis with fixed GMR 0.95**

```
interim.tsd.in(max.comb.test = FALSE, min.n2 = 6,
               GMR1 = 1.0876, CV1 = 0.18213, n1 = 12, GMR = 0.95)
TSD with 2x2 crossover
 Inverse Normal approach
  - Standard combination test with weight for stage 1 = 0.5
  - Significance levels (s1/s2) = 0.03037 0.03037
  - Critical values (s1/s2) = 1.87542 1.87542
  - BE acceptance range = 0.8 ... 1.25
  - Observed point estimate from stage 1 is not used for SSR
  - With conditional error rates and conditional estimated target power
Interim analysis after first stage
 - Derived key statistics:
    z1 = 3.10000, z2 = 1.70344
    Repeated CI = (0.92491, 1.27891)
    Median unbiased estimate = NA
 - No futility criterion met
 - Test for BE not positive (not considering any futility rule)
 - Calculated n2 = 6
 - Decision: Continue to stage 2 with 6 subjects
```

Standard Combination Method
Minimum six subjects in stage 2

- **Final analysis**

```
final.tsd.in(max.comb.test = FALSE,
             GMR1 = 1.0876, CV1 = 0.18213, n1 = 12,
             GMR2 = 1.0893, CV2 = 0.16776, n2 =  6)
TSD with 2x2 crossover
 Inverse Normal approach
  - Standard combination test with weight for stage 1 = 0.5
  - Significance levels (s1/s2) = 0.03037 0.03037
  - Critical values (s1/s2) = 1.87542 1.87542
  - BE acceptance range = 0.8 ... 1.25
Final analysis after second stage
 - Derived key statistics:
    z1 = 3.70299, z2 = 2.06106
    Repeated CI = (0.96127, 1.23210)
    Median unbiased estimate = 1.0884
 - Decision: BE achieved
```

  - Same conclusion as Maximum Combination Test; slightly less conservative than its CI 0.9567 – 1.2380 due to α_{adj} 0.03037 instead of 0.02635

Maximum empiric Type I Error (10^6 simulations under the Null)

```
library(Power2Stage)
# Simulation-based (parallel design): Welch-Satterthwaite test; Reference 14
# Exact Power estimation by Owen's Q-function
TIE.max <- power.tsd.p(method = "B", alpha = rep(0.0294, 2),
                       CV = 0.52, n1 = 116, GMR = 0.95,
                       targetpower = 0.8, test = "welch",
                       pmethod = "exact", theta0 = 1.25)$pBE
e       <- data.frame(Ref = "[14]", Type = 1, Method = "B ", GMR = 0.95, Power = 80,
                      CV = "10 - 100%", adj = 0.0294, TIE.max = TIE.max)
e$Power <- sprintf("%.0f%%", e$Power)
print(e, row.names = FALSE)
```

Example: Power (parallel design [14])

```
library(PowerTOST); library(Power2Stage)
CV <- 0.45 # total (pooled from within- and between-subject) CV
n1 <- floor(0.8 * sampleN.TOST(CV = CV, design = "parallel",
                               print = FALSE)[["Sample size"]]) # my recommendation
power.tsd.p(method = "B", alpha = rep(0.0294, 2), CV = CV, n1 = n1,
            pmethod = "exact", test = "welch", npct = c(0.05, 0.25, 0.5, 0.75, 0.95))
TSD with 2 parallel groups
Method B: alpha (s1/s2) = 0.0294 0.0294
CIs based on Welch's t-test
Target power in power monitoring and sample size est. = 0.8
Power calculation via exact method
CV1 and GMR = 0.95 in sample size est. used
No futility criterion
BE acceptance range = 0.8 ... 1.25
CV = 0.45; ntot(stage 1) = 128 (nT, nR = 64, 64); GMR = 0.95

1e+05 sims at theta0 = 0.95 (p(BE) = 'power').
p(BE)    = 0.82637
p(BE) s1 = 0.59414
Studies in stage 2 = 40.55%
Distribution of n(total)
- mean (range) = 155.5 (128 ... 316)
- percentiles
 5% 25% 50% 75% 95%
128 128 128 188 224
```

- 100,000 simulations by default
- Probability to stop in the interim for success ≈59%
- Probability to procees to second stage ≈41%
- Expected average total sample size E[N] 156
- Final power larger than fixed sample design's 80% with 160 subjects

Example: Empiric Type I Error (parallel design [14])

```
library(PowerTOST); library(Power2Stage)
CV <- 0.45 # total (pooled from within- and between-subject) CV
n1 <- floor(0.8 * sampleN.TOST(CV = CV, design = "parallel",
                               print = FALSE)[["Sample size"]]) # my recommendation
power.tsd.p(method = "B", alpha = rep(0.0294, 2), CV = CV, n1 = n1,
            pmethod = "exact", test = "welch", npct = 0.5, theta0 = 1.25)
TSD with 2 parallel groups
Method B: alpha (s1/s2) = 0.0294 0.0294
CIs based on Welch's t-test
Target power in power monitoring and sample size est. = 0.8
Power calculation via exact method
CV1 and GMR = 0.95 in sample size est. used
No futility criterion
BE acceptance range = 0.8 ... 1.25
CV = 0.45; ntot(stage 1) = 128 (nT, nR = 64, 64); GMR = 0.95

1e+06 sims at theta0 = 1.25 (p(BE) = TIE 'alpha').
p(BE)    = 0.044659
p(BE) s1 = 0.029288
Studies in stage 2 = 96.94%
Distribution of n(total)
- mean (range) = 190.5 (128 ... 330)
- percentiles
50%
190
```

- At the limits of the BE range one million simulations by default
- Probability to pass in the interim close to the level of the test
- Type I Error controlled (<0.05, significance limit of the binomial test 0.05036)