SANOFI

BE·BAC

Two-Stage Designs

Dealing with Uncertainty

Alternatives to Pilot/Pivotal Studies

- In the conventional approach of pilot/pivotal study
  - only part of the information (T/R ratio, *CV*) of the former is used to design the latter
  - The individual data of the pilot are not used (pilot study only supportive information in the application)
- Alternatives
  - Group-Sequential Designs (GSD)
  - (Adaptive) Sequential Two-Stage Designs (TSD)
  - In both the entire information is used

Comparison

- Data of example 2 of Potvin *et al*. *
  - Pilot/Pivotal 'carved in stone' approach
    - Pilot n 12: T/R ratio 1.0876, *CV* 18.213% →
    - Pivotal n 24: 90% CI passes BE
  - GSD, T/R 0.95 and *CV* 20% assumed → *N* 24
    - Group 1 n 12: outcome like the pilot study
    - Group 2 n 12: T/R ratio 0.9141, *CV* 25.618%
    - Pooled N 24: 94.12% CI passes BE
  - TSD
    - Method B
      - Stage 1 n_1 12: outcome like the pilot study →
      - Stage 2 n_2 8: outcome like GSD, 2nd group
      - Pooled N 20: 94.12% CI passes BE
    - Inverse Normal Method / Maximum Combination Test
      - Stage 1 n_1 12: outcome like the pilot study →
      - Stage 2 n_2 6: repated CI passes BE

* Potvin D, DiLiberti CE, Hauck WW, Parr AF, Schuirmann DJ, Smith RA. *Sequential design approaches for bioequivalence studies with cross-over designs*. Pharmaceut Statist. 2008; 7(4): 245–62. doi:10.1002/pst.294.

Comparison

- Pilot/Pivotal ‘carved in stone’ approach
  - 36 subjects (12 if BE unlikely acc. to pilot)
- GSD
  - 24 subjects (12 with futility criterion)
- TSD
  - Method B
    - 20 subjects (12 with futility criterion)
  - Inverse Normal Method / Maximum Combination Test
    - 18 subjects (12 with futility criterion)
- GSD/TSD *vs*. Pilot/Pivotal
  - Ethically favorable
  - Total cost reduced by 33–50% (same cost if BE unlikely)

Group-Sequential Designs (GSD)

- Standard methodology in clinical research (Phase III) for decades
  - Based on Armitage *et al*. (1969), McPherson (1974), Pocock (1977), O'Brien/Fleming (1979), Lan/DeMets (1983), Jennison/Turnbull (1999), …
- Fixed total sample size (*N*) and – in BE one – interim analysis
- Assumptions are required
  - Total sample size planned on 'worst case'
    - T/R ratio
    - *CV*
- Adjustment of α due to multiplicity required
  - Sample size penalty compared to fixed sample design

GSD

- All published methods were derived for
  - superiority testing
  - parallel groups
  - normal distributed data with known variance, and
  - the interim analysis at *exactly N*/2
- That is not what we have in BE
  - Equivalence (generally crossover)
  - Lognormal data with unknown variance
  - Due to dropouts, the interim analysis might not be exactly at *N*/2, which might inflate the Type I Error (the patient's risk)
- First proposal * in the context of BE did not get regulatory acceptance in Europe

* Gould LA. *Group Sequential Extension of a Standard Bioequivalence Testing Procedure*. J Pharmacokin Biopharm. 1995; 23(1): 57–86. doi:10.1007/BF02353786.

GSD

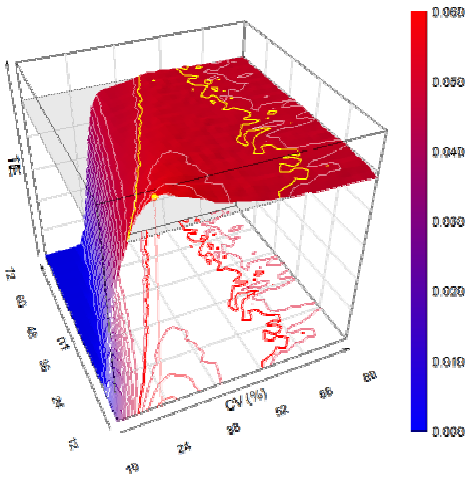
- Asymmetric split of α is possible, *i.e.*,
  - a small α in the interim (*i.e.*, stopping for futility) and
  - a large one in the final analysis (which gives only a small sample size penality compared to a fixed sample design)
  - Examples
    - Haybittle/Peto α_1 0.001, α_2 0.049
    - O'Brien/Fleming α_1 0.005, α_2 0.048
    - Not developed for crossover designs and sample size re-estimation (fixed stage 1 sample size n_1 and variable total sample size N)
      - Lower α_2 or α-spending functions (Lan/DeMets, Jennison/Turnbull) required to control the Type I Error
    - First method for BE in crossovers * with α_1 0.01, α_2 0.04 controls the TIE

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

GSD

- Type I Error
  - 1 mio studies simulated for n_1 12–72, *CV* 10–80%, step size 2

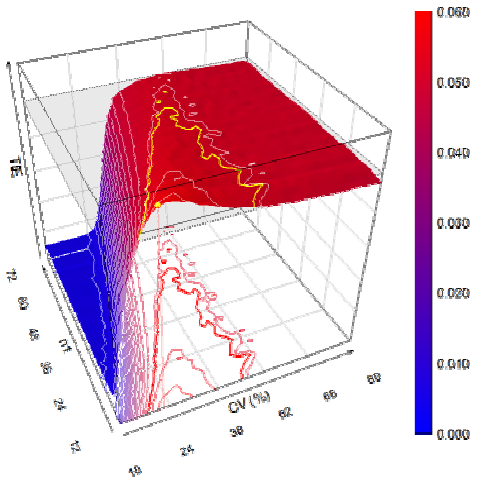
Haybittle/Peto
α_1 0.001, α_2 0.049



Maximum 0.05849

α_2 0.0413 to control the TIE

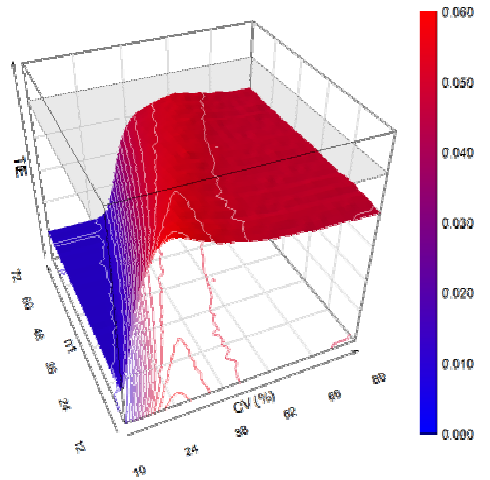
O'Brien/Fleming
α_1 0.005, α_2 0.048



Maximum 0.05700

α_2 0.0415 to control the TIE

Zheng *et al.*
α_1 0.01, α_2 0.04



Maximum 0.04878

GSD (Guidelines)

- Australia (2004), Canada (Draft 2009)
  - Bonferroni's correction (α_{adj} 0.025)
    - Theoretical Type I Error $\leq$0.0494
    - For *CV*s and samples sizes common in BE the TIE is generally $\leq$0.04
- Canada (2012)
  - Pocock's α_{adj} 0.0294
  - n_1 based on 'most likely variance' + additional subjects in order to compensate for expected dropout-rate
  - *N* based on 'worst-case scenario'
  - If $n_1 \neq N/2$ relevant inflation of the Type I Error is possible!
    - α-spending functions can control the TIE
    - Regrettably <u>not</u> mentioned in the guidance…

(Adaptive) Sequential Two-Stage Designs

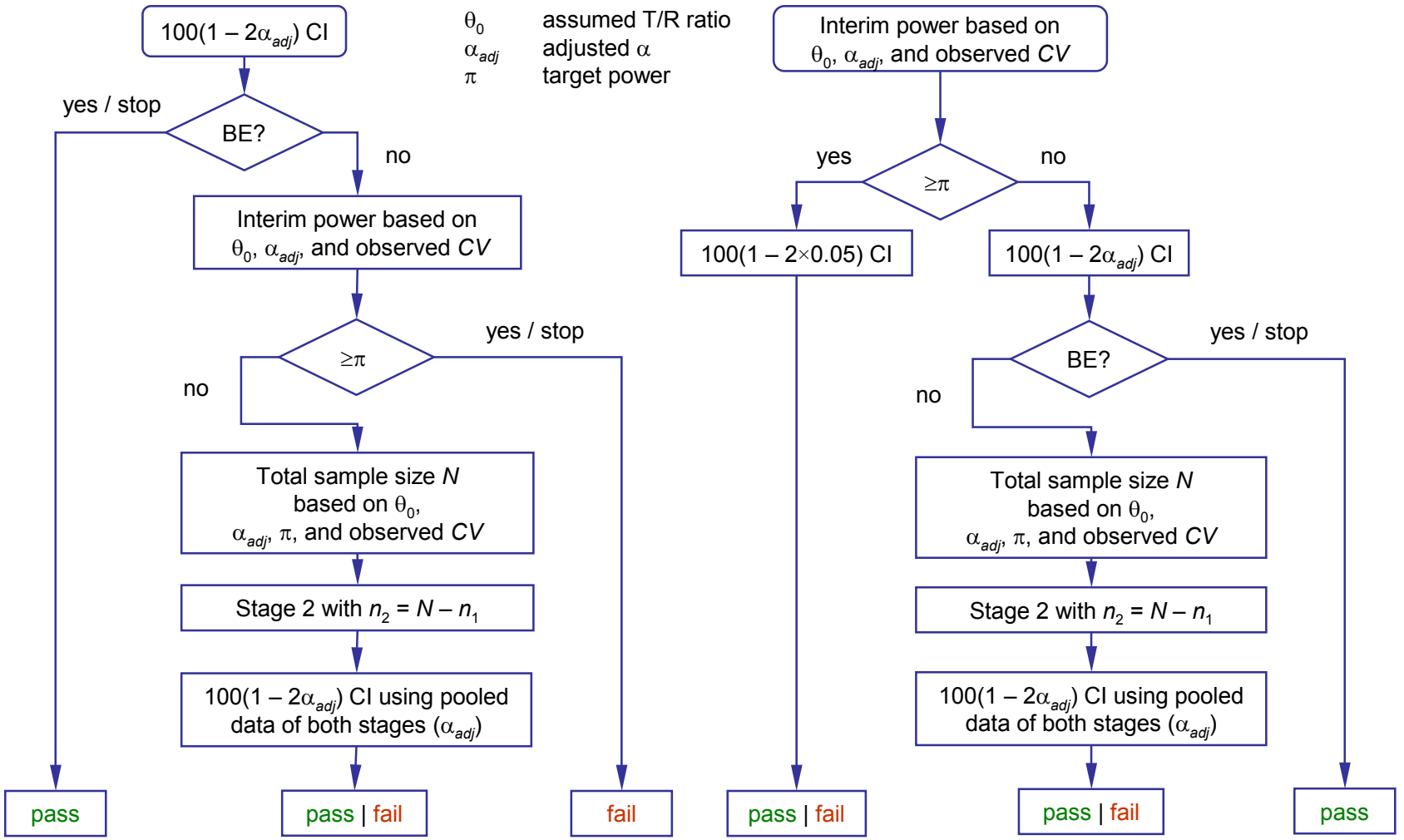
- Fixed stage 1 sample size (n_1) and sample size re-estimation (n_2) in the interim analysis
  - Generally a fixed T/R ratio is assumed
  - All published methods are valid only for a range of combinations of stage 1 sample sizes, T/R ratios, *CV*s, and desired power
  - Fully adaptive methods (*i.e.*, taking also the T/R ratio of stage 1 into account) can be problematic
    - May deteriorate power and require a futility criterion
    - Simulations generally required
  - With one exception (Inverse Normal Method, slide 32) an analytical proof of controlling the TIE does not exists
    - It is the responsibility of the sponsor to demonstrate (*e.g.*, with simulations) that the consumer risk is preserved

(Adaptive) Sequential Two-Stage Designs

- Two 'Types' of TSDs *
  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
     → inspired by GSD
     similar sampe size penalty like GSD
  2. An unadjusted α may be used in the first stage – dependent on interim power and an adjusted α in the second stage
     → a mixture of conventional BE and GSD
     lower sample size penalty compared to Type 1

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

Type 1 and Type 2



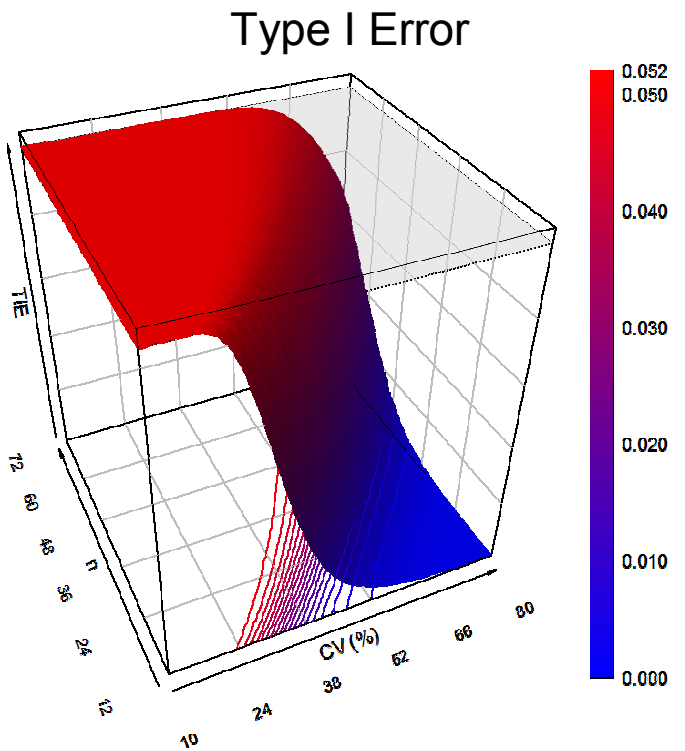
TSD

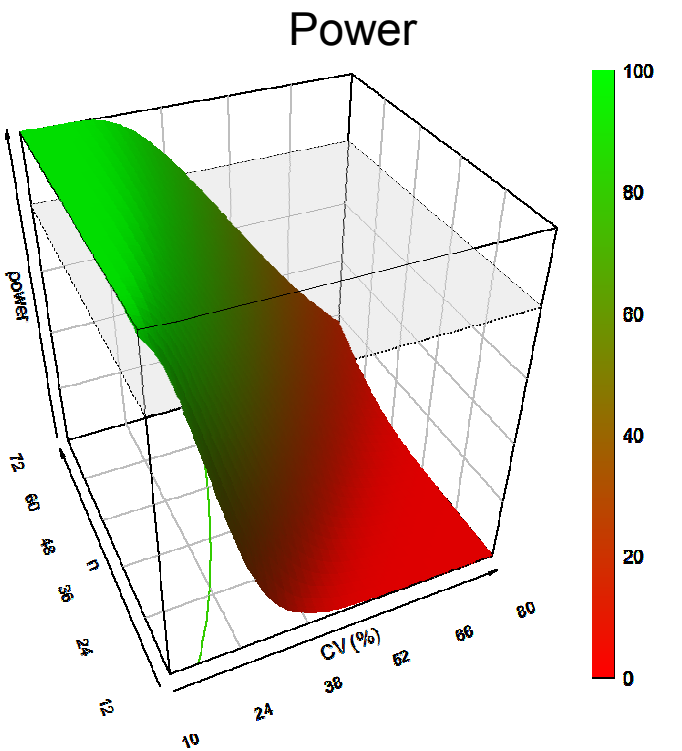
- Methods by Potvin *et al.* first validated frameworks in the context of BE
  - Supported by the 'Product Quality Research Institute' (FDA/CDER, Health Canada, USP, AAPS, PhRMA, …)
  - Inspired by conventional BE testing and Pocock's α_{adj} 0.0294 for Group-Sequential Designs
    - A fixed * T/R ratio is assumed
      - The T/R ratio in the first publication was 0.95; later extended to 0.90 by other authors
    - The *CV* in the interim is taken into account for sample size re-estimation
    - Target power 80% (later extended to 90%)

* A common mistake is to use the T/R ratio observed in the first stage. Whilst possible, not part of the published methods.

Interlude (2×2×2 crossover)

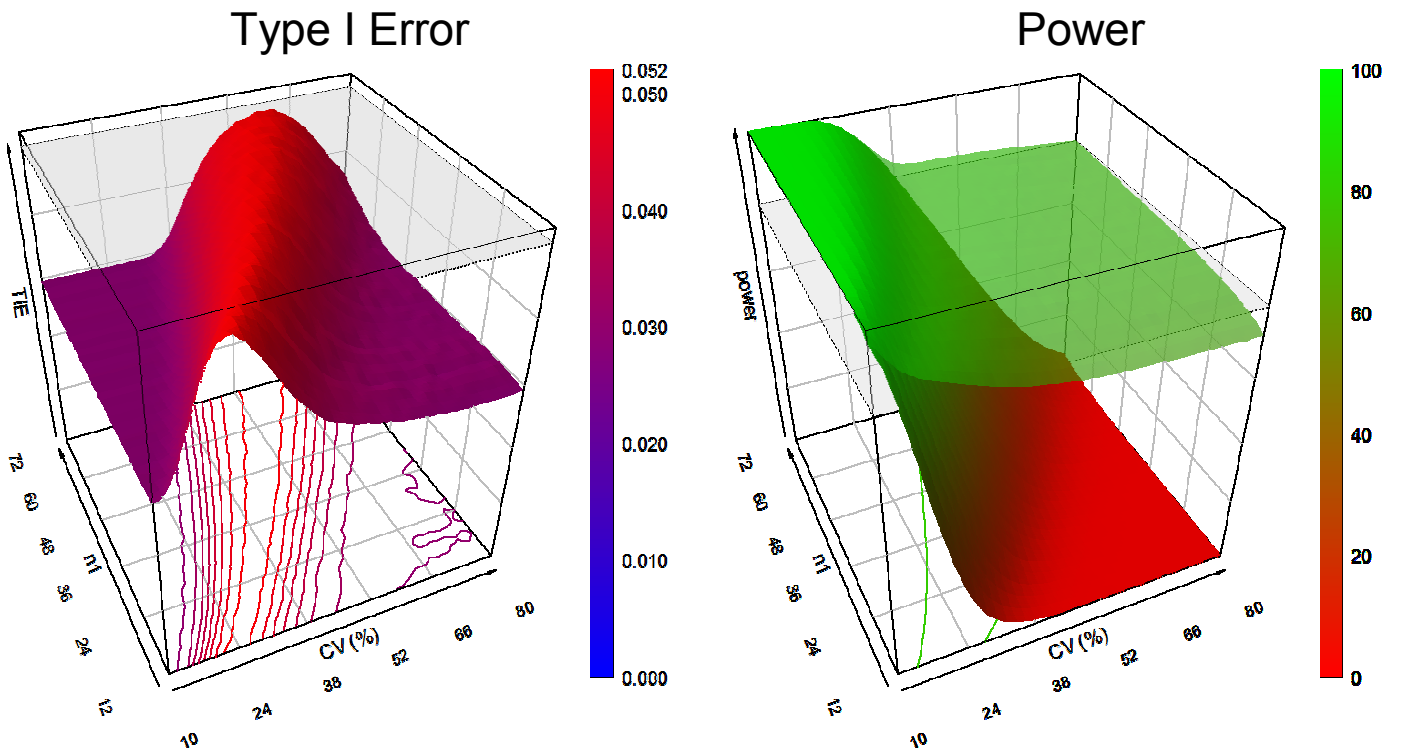
- Operating Characteristics
  - Fixed sample design: T/R 0.95, *CV* 10–80%, *n* 12–72





TSD

- Operating Characteristics
  - ‘Type 1’ TSD (Method B, α_{adj} 0.0294, T/R 0.95, power 80%): *CV* 10–80%, n_1 12–72; 1 mio simulations in all combinations



TSD

- Some published frameworks
  - Stage 1 sample sizes 12 – 60, no futility rules

Reference	Type	Method	T/R	Power	*CV*	α_{adj}	TIE_{max}
Potvin *et al.* (2008)	1	B	0.95	80%	10 – 100%	0.0294	0.0485
	2	C					0.0510
Montague *et al.* (2012)	2	D	0.90			0.0280	0.0518
Fuglsang (2013)	1	B	0.95	90%	10 – 80%	0.0284	0.0501
	2	C/D				0.0274	0.0503
	2	C/D	0.90			0.0269	0.0501

  - Xu *et al*. (2015). Futility for CI, N_{max} 42 (low *CV*), 180 (high *CV*)

Type	Method	T/R	Power	*CV*	Futility region	α_1	α_2	TIE_{max}
1	E	0.95	80%	10 – 30%	0.9374 – 1.0667	0.0249	0.0363	0.050
2	F				0.9492 – 1.0535	0.0248	0.0364	0.050
1	E			30 – 55%	0.9305 – 1.0747	0.0254	0.0357	0.050
2	F				0.9350 – 1.0695	0.0259	0.0349	0.050

TSD

- Futility rules
  - When modifying a published method by adding a futility criterion – or more – for early stopping
    - Own simulations not required since the chance to proceed to the second stage will be lower; the Type I Error will be lower as well
    - However, it may deteriorate power – simulations recommended

> *Introduction of […] futility rules may severely impact power in trials with sequential designs and under some circumstances such trials might be unethical* [1]
>
> *[…] 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* [2]

1. Fuglsang A. *Futility rules in bioequivalence trials with sequential designs*. AAPS J. 2014;16(1): 79–82. doi:10.1208/s12248-013-9540-0.
2. Jones B, Kenward MG. *Design and analysis of crossover trials*. Boca Raton: Chapman & Hall/CRC; 2014. p. 365–80.

TSD

- Own simulations
  - If none of the published methods fits you needs
    - In all methods covering a wide range of n_1 and *CV* the maximum TIE is observed at a combination of small n_1 and low *CV*
    - Since the method has to cover all combinations, the adjusted α is based on this worst case; in the actual study the TIE might be lower (slide 15, left panel)
  - Simulations straightforward with current software
    - Finding a suitable α_{adj} and validating for TIE and power takes ~20 minutes with the R package Power2Stage *
    - Also useful answering a deficiency letter; simulating one mio studies for the empiric TIE with given α_{adj}, n_1, and *CV* for a particular study takes less than 10 seconds

* Labes D, Lang B, Schütz H. *Power2Stage: Power and Sample-Size Distribution of 2-Stage Bioequivalence Studies*. R package version 0.5-2. 2019. https://cran.r-project.org/package=Power2Stage.

TSD

- Cost Analysis
  - Considering 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 average sample size in the second stage?
    - Which power can we expect in the first stage and the final analysis of pooled data?
    - Will there be a substantial sample size penalty compared to a fixed sample design?
    - Will introduction of futility rule(s) substantially decrease power?

TSD

- Example
  - Assumed T/R ratio 0.95, *CV* 20%, target power 80%
    'Type 1' TSD *vs*. fixed sample design (*n* 20, 83.5% power)

n_1	$E[N]$	Studies stopped in stage 1 (%)	Failed in stg. 1 (%)	Passed in stg. 1 (%)	Studies → stg. 2 (%)	Power final (%)	Average increase of cost (%)
12	20.6	43.7	2.3	41.3	56.3	84.3	3.0
14	20.0	55.8	3.0	52.7	44.2	84.9	0.1
16	20.0	66.1	4.1	61.9	33.9	85.2	0.2
18	20.6	74.6	5.2	69.5	25.4	85.7	3.1
20	21.7	81.5	6.3	75.2	18.5	86.3	8.4
22	23.0	87.0	7.2	79.8	13.0	87.1	15.1
24	24.6	91.4	7.9	83.5	8.6	88.0	23.0

  - TSD always more powerful than fixed sample design
  - n_1 ~80% of *n* gives reasonable chance to pass BE already in the first stage

TSD (Guidelines)

- Adopted
  - Partly flawed and/or ambiguous statements
    - 2010 European Economic Area
      (EU + Norway, Iceland, Liechtenstein)
    - 2011 Australia (EMA GL adopted)
    - 2012 Health Canada (Potvin C)
    - 2013 USA (Potvin C, Montague D) *
      Russian Federation
    - 2015 New Zealand (Australian GL adopted)
    - 2016 Eurasian Economic Union
      Gulf Cooperation Council (Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, United Arab Emirates)
    - 2017 Egypt

* 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. doi:10.1208/s12248-013-9499-x.

TSD (Guidelines)

- Public Consultation
  - 2019 Brazil (№ 760.20)
- Ambiguities, misconceptions, flaws
  - EMA IR BE GL

    For example, using 94.12% confidence intervals for both the analysis of stage 1 and the combined data from stage 1 and stage 2 would be acceptable, but there are many acceptable alternatives and the choice of how much alpha to spend at the interim analysis is at the company's discretion. The plan to use a two-stage approach must be pre-specified in the protocol along with the adjusted significance levels to be used for each of the analyses.

    - Ambiguous
      - » In some methods different alphas are used in the stages
      - » In others α_2 is adjusted based on the first stage

TSD (Guidelines)

- Ambiguities, misconceptions, flaws
  - Comments on the EMA's IR BE Draft GL
    - Proposed change

      Suggested to clarify the section and gave the framework of Potvin's Method C as an example
    - Outcome

      Example included – the example chosen was the Pocock approach – very similar to the example in the flow chart included in the comment

      » Bad. Pocock's is a GSD (one interim at *N*/2) and the α 0.0294 (94.12% CI) given in example is for superiority testing.
      Pocock's α for equivalence is 0.0304 (93.92% CI).
      However, a GSD is not a TSD with sample size re-estimation.

TSD (Guidelines)

- Ambiguities, misconceptions, flaws
  - EMA Q&A Document (Rev. 7, Feb 2013)

    […] it is considered that there is no minimal number of subjects to be included in the second stage of a two-stage design, so long as it can be demonstrated that the type I error of the study is controlled. […] minimum of two patients in each stage of the study […] there is at least one subject randomised to each sequence. This does not supersede the requirement for at least 12 subjects overall.

    - Though correct in principle, superfluous
      - » Nobody would start the first stage with less than 12 subjects because – if BE is demonstrated – the study is not acceptable for the authority
      - » If a second stage can be initiated, no software will estimate a sample size of one (always rounds up to get balanced sequences). Hence, two subjects in the second stage is the minimum anyway.

TSD (Guidelines)

- Ambiguities, misconceptions, flaws
  - EMA Q&A Document (Rev. 7, Feb 2013)

    […] the expected ANOVA model for analysis of the combined data from a two-stage design would have the following terms:
    stage, sequence, sequence×stage,
    subject(sequence×stage),
    period(stage), formulation.

    - Strange at least
      - » None of the publications used the term *sequence×stage*
      - » Modification demonstrated to be irrelevant *

* 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.

TSD (Guidelines)

- Ambiguities, misconceptions, flaws
  - ANVISA № 760.20 Public Consultation, Section III (Article 75)
    - IV *This second group must have at least 50% of the previous group*
      - » Wrong. The sample size of the second stage depends only on the estimates of the first stage. Copy/pasted from Japan's Add-On Design or Mexico's GL?
        In few methods $n_1/2$ is indeed the minimum n_2 but generally there is neither a minimum nor a maximum size of n_2.
    - V *[…] to demonstrate bioequivalence the confidence level becomes 94.12*
      - » Wrong. Pocock's α 0.0294 (94.12% CI) is not a 'natural constant' which preserves the Type I Error under all conditions (slide 16).

TSD (Case Studies)

- Case 1 (Potvin 'Method C', 2010/2011)
  - Study stopped in stage 1
    - *AUC* power >80%, passed BE with 90% CI
    - C_{max} power <80%, passed BE with 94.12% CI
  - Deficiency letter (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.
  - The protocol was approved by the German BfArM and performed exactly to the published method …
    - Failed to show BE of *AUC* with 94.12% CI
    - Study repeated in India in a very (!) large fixed sample design
    - Failed on C_{max} – entire project cancelled

TSD (Case Studies)

- Case 2 (Potvin 'Method C', 2011/2012)
  - Study passed in stage 1
    - *CV* in the interim 30.65%, n_1 49
    - 90% CI since power was 87.3%
  - Deficiency letters (UK, Ireland)

    Unadjusted α in stage 1 [is] not acceptable.
    - Study passed with 94.12% CI as well (*post hoc* switch to 'Method B')
  - Deficiency letter (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.
    - 1 mio studies simulated based on the study's *CV* and n_1 $\rightarrow$ empiric Type I Error 0.0494 (95% CI: 0.0490 – 0.0498)

* Unofficial information: Potvin's table contains only the TIE (0.0494) for *CV* 30% and n_1 48…

TSD (Case Studies)

- Case 3 (Potvin ‘Method C’, 2012)
  - Protocol synopsis with statistical details submitted to the Spanish Agency
  - Unofficial feedback (after consultation of AEMPS with the CHMP’s Biostatistics Working Party)

    Potvin’s method is not valid in Europe
- Case 4 (Potvin ‘Method C’, 2013)
  - […] we’d like to ask about the current status of TSD BE study, […] if the BE protocol with Potvin’s Method C is acceptable now […].
  - Answer

    Potvin’s methods are not acceptable for EMA

TSD (Case Studies)

- Case 4 (own simulations 2013, study 2014)
  - Assumed T/R ratio 0.90, no reliable information about *CV* (guesstimate 20%), study in patients (started with 86), expected dropout-rate due to AEs ~30% → n_1 60
  - 'Type 1' TSD; sponsor wanted to pass BE already in stage 1 (first to file) and plan the second stage only as a 'safety net'
    - Simulations covered n_1 43–86 (<43 not needed because with 50% dropouts the study would be stopped for ethical reasons) and *CV* 10–45%
      - With the expected *CV* 20% and n_1 60, power in the first stage 89.9%, chance to proceed to the second stage 6.1%, overall power 92.7%
    - *CV* of C_{max} in the study 18.1%, dropout-rate lower than expected (18.6%)
    - Passed all PK metrics with adjusted CI in the first stage
    - Approach and study accepted by the authority

TSD (Case Studies)

- Case 5 (own simulations 2016)
  - Simulations based on a ‘Type 1’ TSD for parallel groups *
    - Large n_1 (up to 125/group), homo- and heterogenous variances, potentially unequal group sizes due to drop-outs (study in patients of CNS drug); with α_{adj} 0.0274 the maximum TIE was 0.04992
  - Scientific Advice (Sweden)

    According to the guideline, application of a TSD was accepted provided that the patient’s risk is maintained at or below 5%. Confirmed that the statement about Potvin’s methods is not public; these types of TSDs are not proven in a strict sense. However, it was acknowledged that the simulations covered a sufficient range of possible outcomes (unequal variances and drop-out rates).
    […] the empiric type I error rate should be evaluated with the real data (i.e., the actual group sizes and variances of the study).

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

TSD (Exact Method)

- Repeated Confidence Intervals [1] / Inverse Normal Method adapted for BE [2,3,4,5]
  - Implemented since version 0.5-1 of Power2Stage (2018)
  - Possible
    - Futility criteria on the T/R ratio or its CI in the interim
    - Futility criterion on the maximum sample size
    - Use the T/R ratio in the interim for sample size re-estimation (*i.e.*, full adaptive)

1. 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. doi:10.1002/sim.3538.
2. König F, Wolfsegger M, Jaki T, Schütz H, Wassmer G. *Adaptive two-stage bioequivalence trials with early stopping and sample size re-estimation*. Vienna: 2014; 35th Annual Conference of the International Society for Clinical Biostatistics. Poster P1.2.88. doi:10.13140/RG.2.1.5190.0967.
3. 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. doi:10.1186/1745-6215-16-S2-P218.
4. Patterson SD, Jones B. *Bioequivalence and Statistics in Clinical Pharmacology*. Boca Raton: Chapman & Hall/CRC; 2nd edition 2017.
5. 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;1–21. doi:10.1002/sim.7614.

TSD (Exact Method)

- Example
  - Data given by Potvin *et al*.: 12 subjects in stage 1, T/R ratio 1.0876, *CV* 18.213%
    - R code

```
interim.tsd.in(GMR1 = 1.0876, CV1 = 0.18213, n1 = 12)
```

    - Result

```
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)
- 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
```

TSD (Exact Method)

- Example cont'd
  - Although we would need only 6 subjects, we continue with the example of Potvin *et al.*: 8 subjects in stage 2, T/R ratio 0.9141, *CV* 25.618%
    - R code

```
final.tsd.in(GMR1 = 1.0876, CV1 = 0.18213, n1 = 12,
             GMR2 = 0.9141, CV2 = 0.25618, n2 = 8)
```

    - Result

```
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 = 2.87952, z2 = 2.60501,
  Repeated CI = (0.87690, 1.17356)
  Median unbiased estimate = 1.0135
- Decision: BE achieved
```

TSD (Exact Method)

- Example cont'd
  - The data are not pooled like in the others
  - Stages are evaluated by separate ANOVAs
  - The study passed with a (repeated) CI of 87.69–117.36%
  - Although slightly more conservative, same conclusion like based on the 94.12% CI of 88.45–116.38% reported by Potvin *et al*.

Rumors & Chinese Whispers

- TSDs based on simulations
  - One member of the PKWP (2015)

    I made peace with these methods and accept studies – if the confidence interval is not too close to the acceptance limits

    - How close is 'not too' close?
  - Assessors of Spain and Austria (2016)

    *It was shown * that the adjusted α 0.0294 used by Potvin et al. is Pocock's for superiority. The correct value for equivalence is 0.0304. Hence, all studies evaluated with a 94.12% CI in both stages are more conservative than necessary. At least these studies should not be problematic.*

    - It is a misconception that 0.0304 is 'universally valid' for equivalence
    - Other settings (T/R ratio, power) require other adjustements – even for 'Type 1' TSDs (slide 16)

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

Rumors & Chinese Whispers

- TSDs based on simulations
  - Another member of the PKWP asked the BSWP (2015) which inflation of the Type I Error would be acceptable; he gave 0.0501 as an example

    Answer: The TIE must not exceed 0.05

    - Rounding of the CI as required by the GL leads to approval of studies (regardless the design) with confidence limits of 79.995% and/or 125.004% – which inflates the Type I Error up to 0.0508. This is the common approach worldwide and obviously does not worry regulators…

Rumors & Chinese Whispers

- TSDs based on simulations
  - A Portuguese assessor saw a study rejected by one of his colleagues – although BE was shown (2016)
    When asked why, the answer was

    According to the BSWP Potvin's methods are not acceptable
  - He was not aware of such a statement and asked his colleague for an official document

    Answer: Such a document does not exist but all statisticians in the agencies know this statement

– Since 2015 the Type I Error in TSDs was on the work plan of the PKWP/BSWP
  - Discussions suspended *

* Coppola P. *PKWP – BE/PK position on specific questions*. BioBridges. Prague; 26–27 September, 2018.

TSD (the Assessor's Dilemma)

- If a (European) assessor would like to accept a TSD, he/she is facing a dilemma
  - TSDs are stated in the GL and the Q&A document and therefore, studies are submitted
  - The BSWP
    - does not 'like' methods based on simulations and prefers methods which demonstrate by an analytical proof that the patient's risk is preserved
    - Considers even a TIE of 0.0501 not acceptable
    - Methods based on simulations might (!) be acceptable if no alternative which shows analytically control of the TIE is available *
      - That means – for the EMA – simulation-based methods for 2×2×2 crossover designs are essentially 'dead'

* Brandt A. *Personal information*. 4th Annual Biosimilars Forum. Budapest; 17–18 October, 2019.

Conclusions

- Do not blindly follow guidelines; some recommendations are based on misconceptions or even flawed
  - May inflate the patient's risk and/or
  - deteriorate power
- Published frameworks can be applied without requiring the sponsor to perform own simulations – although they could further improve power based on additional assumptions
  - GSDs and TSDs are both ethical and economical alternatives to fixed sample designs
- The Inverse Normal Method
  - Preserves the TIE in the strict sense (no simulations required)
  - Is fully adaptive and provides the most flexibility

Outlook

- Extend the exact method for
  - parallel groups and
  - replicate designs
  - In principle it would be even possible to switch the design in the interim, *i.e.*, if high variability is observed, perform the second stage in a replicate design whilst still using the data of the first for the 90% CI

Thank You!

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