# Pilot Studies

Purpose
Applications
Limitations

Pilot Studies = Good Scientific Practice

- ## In order to properly design a confirmatory/pivotal study
  - Define targets of the pilot study, *i.e.*, in BE
    - Assess whether the – validated – bioanalytical method is suitable in 'real' samples (in the presence of metabolites and endogenous compounds, stability; co-medications in patients, …)
    - Suitability of chosen sampling schedule and wash-out phase
    - Suitability of chosen PK metrics
    - Obtain information on variability and T/R ratio of PK metrics required for sample size estimation
  - Assumptions
    - Be aware of their limitations and potential impact of deviations from them on the expected outcome
    - Keep their number as small as possible
      Everything should be made as simple as possible, but not simpler. (Albert Einstein)

Murphy's Law

- ## Case Study
  - Solution (reference to a new multiphasic product), 24 subjects
    - Validated LC/MS-MS method (SPE, HILC, structural analogue IS, APCI/SIM, LLOQ 500 pg/mL)
    - Study performed before the EMA's BMV GL was in force (blinded review of data acceptable, assessing matrix effect not mandatory)
    - Bioanalytics terminated after 12 subjects due to suspected matrix effect

      - Irregular profiles
      - In some subjects C_{last} 1.65 ng/mL
      - At 12 hours measurable concentrations in only 3/12 subjects, none at 16 hours
      - Not consistent with $t_{1/2}$ from the literature

Murphy's Law
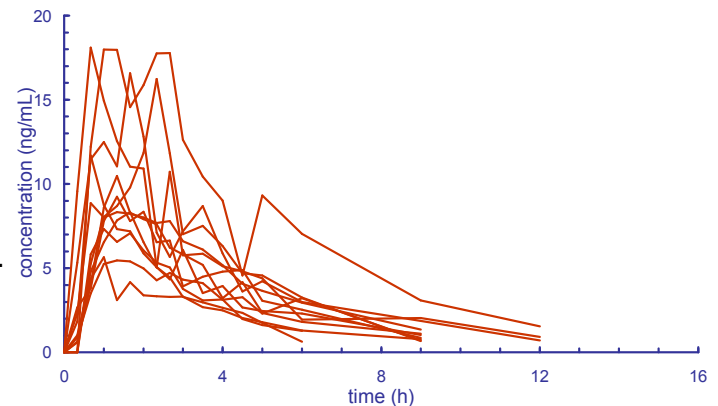
- ## Case Study cont'd

    - GC/MS method (LLE, $^{18}O_2$ labeled IS, derivatization, NICI/SIR, LLOQ 143 pg/mL) developed and validated

        – Expected profiles

        – No matrix effect due to stable isotope labeled IS

        – Concentration at 16 hours measurable in 22/24 subjects (both T and R)

        – $t_{1/2}$ agreed with the literature
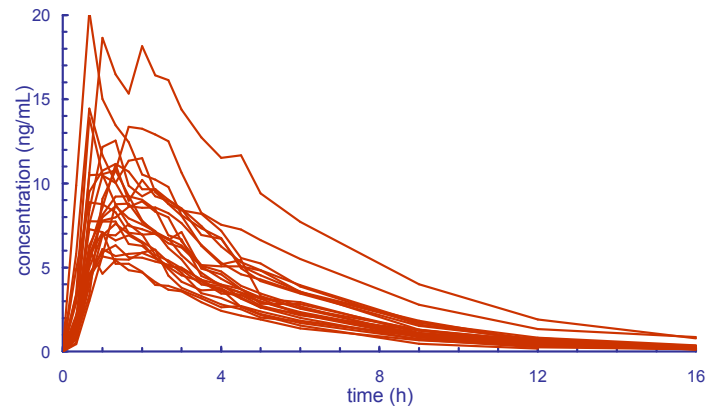
    – Leasons learned

        - A validated method is not necessarily suitable for 'real' samples
        - Sampling schedule was not ideal for the biphasic Test product
        - Study accepted by the authority (supportive in a hybrid application)
        - A pilot study would have prevented the issues

Selection of Candidates (Part I)

- Candidates are developed to match the *in vivo* perfor-mance of the reference product as close as possible
  - The entire arsenal
    - reverse engineering
    - same/similar excipients (Q1/Q2)
    - *in vitro* dissolution (f_2 similarity)

    is applied
  - Patent issues
    - Different salt or polymorph of the API
    - Different release mechanisms of MR products
  - However, without any *in vivo* data we are fishing in the dark (esp. for BCS class II/IV – where f_2 is not informative)
  - Small studies are required to establish an *IVIVC*

Selection of Candidates (Part I)

- ## Candidates should be
  - manufactured with varying process parameters (*e.g.*, compression force, drying time, coating, …)
  - For *IVIVC* at least three formulations are required

- ## ICH E9 [*] states

  The number of subjects in a clinical trial should always be large enough to provide a reliable answer to the questions addressed.

  - If applicable to pilot studies, <u>how</u> large is large <u>enough</u>?
    - For *IVIVC* small sample sizes (6 to 12) are sufficient, since only mean values are used
    - When the purpose of the pilot is sample size estimation for the pivotal study, sample sizes should be – generally substantially – larger

[*] International Council on Harmonisation. *ICH Harmonised Tripartite Guideline. <u>Statistical Principles for Clinical Trials</u>.* 5 February 1998.

Excursion into Terminology

- Noncompartmental Analysis (NCA) give observed (measured) values (*e.g.*, C_{max}/t_{max}) or ones obtained by simple numeric methods (*AUC*)
  - $\rightarrow$ PK <u>metrics</u>

- In modeling we obtain estimates
  - $\rightarrow$ PK <u>parameters</u>

- When comparing PK metrics of treatments, we apply a statistical model[*] (*e.g.*, an ANOVA) and obtain
  - estimates of effects
    (*e.g.* T/R ratios of C_{max}, *AUC*; T–R of t_{max}) and
  - their variabilities (generally given as *CV*)

[*]   Statistics cannot provide *true* values – only estimates, how large their error (uncertainty) is, and a means to deal with it.

Uncertainties

- Results of a pilot study (T/R ratios of PK metrics and their variabilities) are not 'carved in stone' but
  - *estimates* and therefore,
  - not the true values but *uncertain*
- The amount of uncertainty depends
  - on the sample size and
  - (to a lesser degree) on the design
- When using the results *as they are* (*i.e.*, following the 'carved in stone' approach),
  - we leave the area of *assumptions* behind and enter the obscure grounds of *believes*, namely that the T/R ratios and their *CV*s in the pivotal study <u>cannot</u> be 'worse' than in the pilot

Uncertainties

- ## To quote my late father

    If you want to believe, go to church!

- ## Example

    - Results [*] of a pilot study (2×2×2 design, 16 subjects)

        - T/R ratio   0.95
        - CV_w 25%

    - In the 'carved in stone' approach we plug these values into our preferred software, enter the desired power (*i.e.*, 80%) and obtain

        - *n*   28
        - achieved power   80.74%

    - However, is this realistic?
      Let us explore how uncertain the results of the pilot study are

---

[*]   Passes 'BE' with the 90% CI of 81.50–110.74% though by chance (power 50.4%).

Uncertainties

- Example cont'd
  - We can calculate an *e.g.*, 80% * confidence interval of the T/R ratio and the *CV*
    - T/R ratio 0.95   (CI 0.8944 – 1.0090)
    - *CV*   25%   (CI 20.28% – 33.93%)
  - When we based or sample size estimation on the T/R ratio of *exactly* 0.95 and the *CV* of *exactly* 25%, with *any*
    - T/R ratio <0.95 and/or *CV* >25% in the pivotal study we will loose power and possibly fail to show BE
    - Let us explore a bad (though not the worst) case
      - The chance is 10% that the T/R ratio is only 0.8944 (its lower confidence limit) as is the chance that the *CV* is 33.93% (its upper CL)
      - Power (chance to show BE) for this combination will be only ≈32%; time for apostasy…

---

*   In the spirit of a producer's risk of 20%. Gould (doi:10.1007/BF02353786) suggested more liberal 25% (75% CI).

Dealing with Uncertainties

- The larger the sample size of the pilot study, the more reliable (*i.e.*, less uncertain) are the estimates we obtain
  - Statistics is a cruel mistress
    - In order to *double* the precision of an estimate one has to *quadruple* the sample size
  - If you work with a confidence interval, use the lower limit of the T/R ratio and – generally * – the upper limit of the *CV*
    - If the T/R ratio turns out to be 'better' (closer to 1) in the pivotal study, you gain power; money spent but study passes BE
    - If the *CV* is lower, you gain power as well
  - After the pivotal study is performed, prepare for a conversation with the *'Guy in the Armani Suit'* (© Anders Fuglsang)

* In reference-scaled ABE sometimes the lower limit. More about that later.

Dealing with Uncertainties

- – If the study failed and he curses you because you did not use the approach which was 'so successful for years'
    - • Make clear that despite you took the uncertainties into acount (which is definitely more conservative than the 'carved in stone' approach), the study was designed for 80% power, *i.e.*, the chance of failing was still 20%
    - • If he demands a higher chance of passing you will be ready to design the next study for higher power
- – If the study passed and he tells you that you wasted the company's money and should have performed the study in fewer subjects
    - • Make clear that you had/have no crystal ball and it could have been the other way 'round as well
    - • He should be happy that the study passed; repeating a failed study – in a large sample size – would be much more costly
    - • If you want to get troubles: Ask him how many studies he repeated

Not the End of the Tunnel

- Example cont'd
  - Which sample size is required when planning with the upper CL (33.93%) of the *CV* (25%)?
    - *n*   50
    - achieved power   81.44%
      > That will increase the study costs by almost 80%
  - Which sample size is required when planning with the lower CL (0.8944) of the T/R ratio (0.95)?
    - *n*   62
    - achieved power   80.20%
   > That will more than double the study costs
  - Belt plus suspenders (assuming the worst)
    - *n*   110
    - achieved power   80.01%
   > When suggesting that, expect to get fired right away

ICH E9

- Sensitivity analysis to explore the impact on power if values deviate from assumptions

  - The function pa.ABE() of PowerTOST [*] comes handy where we can specify a minimum acceptable power (here 70%)

    - The *CV* can increase to 28.4% (relative +13.7%)

    - The T/R ratio can decrease to 0.927 (relative –2.44%)

   - We can have five dropouts (relative –17.9%)
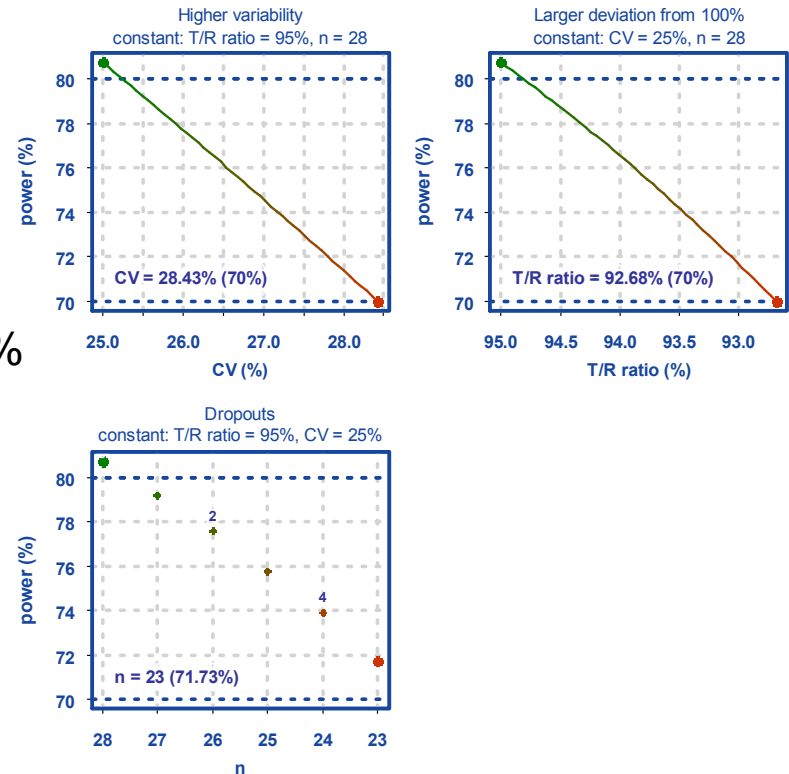
[*] Labes D, Schütz H, Lang B. *PowerTOST: Power and Sample Size for (Bio)Equivalence Studies*. 2019; R package version 1.4-9. https://cran.r-project.org/package=PowerTOST.

ICH E9

- ## Sensitivity analysis cont'd
  - The impact of potential deviations from assumptions is T/R ratio $\gg$ CV $>$ dropouts
  - We have to worry most about the T/R ratio (by far)
    - Power curves are relatively flat close to 1 but get increasingly steep with larger deviations
    - In the study a combination of <u>all</u> deviations (T/R ratio, *CV*, dropouts) occurs simultane-ously – it is up to us to decide on reasonable combinations and analyze their respective impact on power
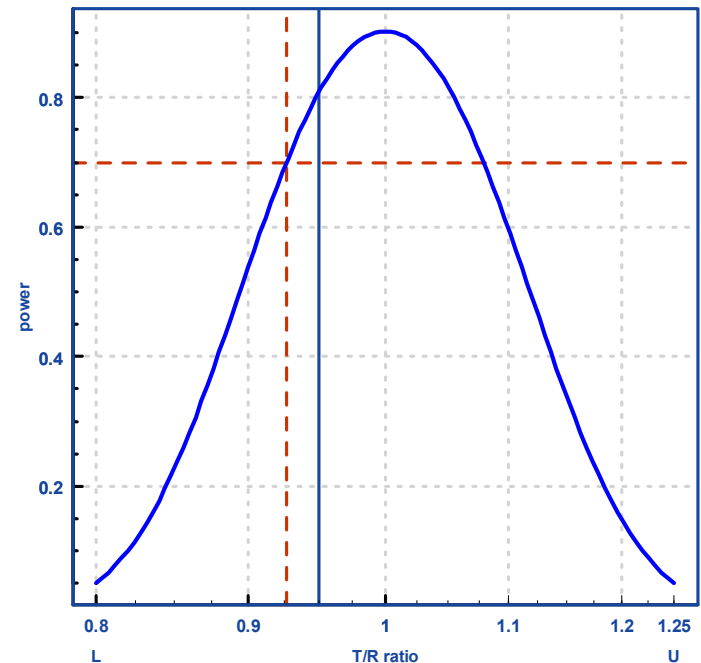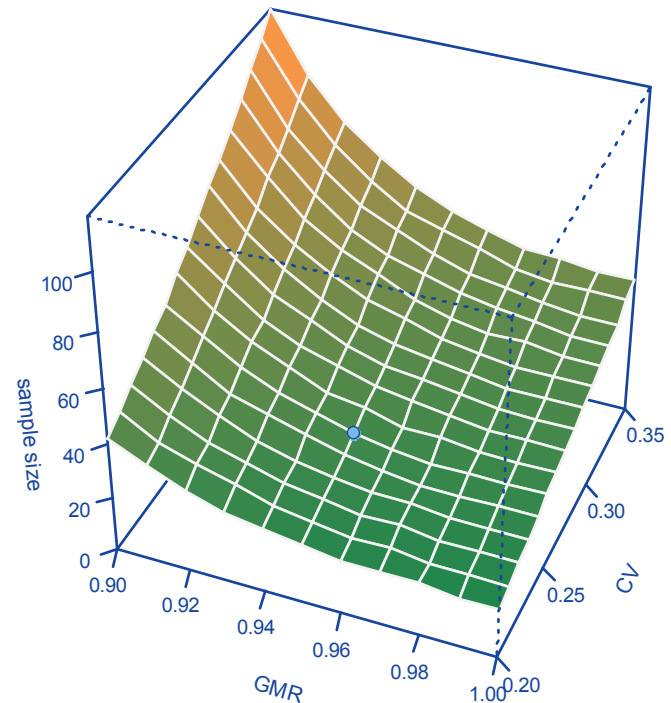
Bayesian Method

- ## Varying T/R ratio and *CV*, required sample size

Pivotal study (80% power) designed on results of a 2x2x2 pilot study with 16 subjects ignoring the uncertainties of estimates (CV, GMR): 'carved in stone'.
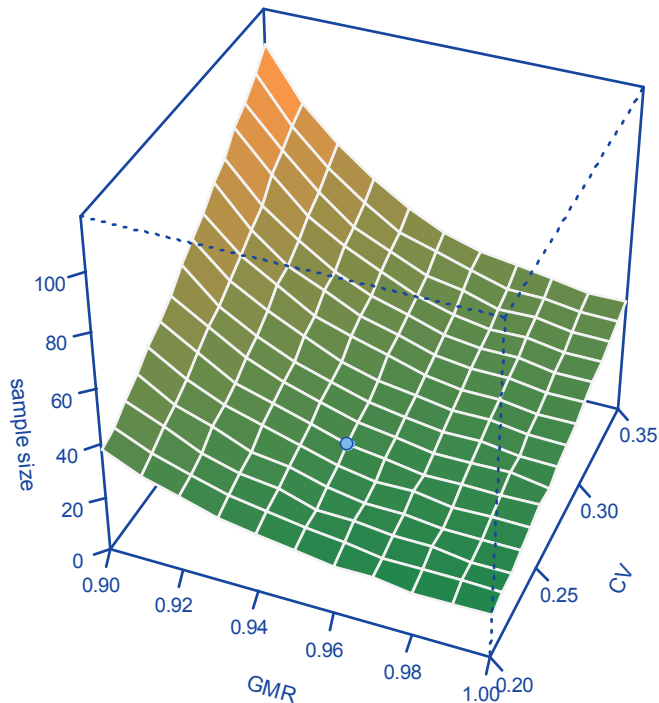
Pivotal study (80% power) designed on results of a 2x2x2 pilot study with 16 subjects taking the uncertainty of estimated CV into account (GMR fixed).
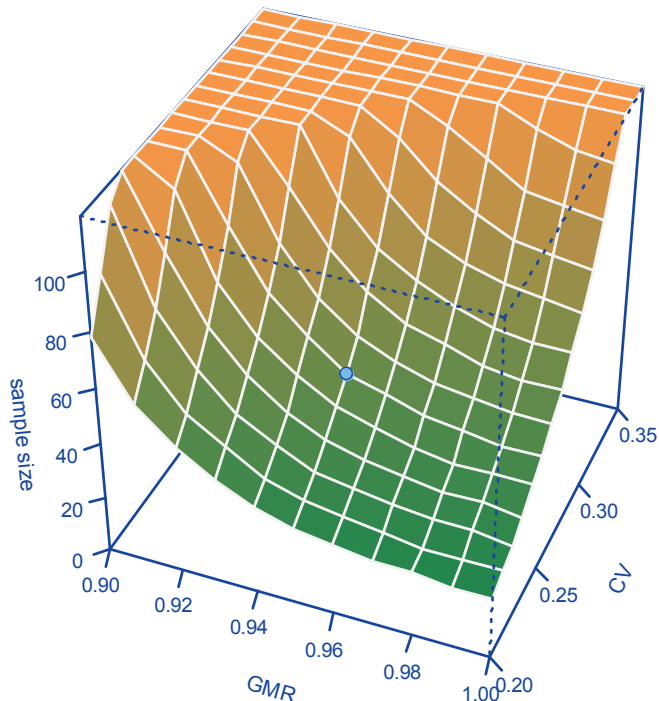
sample size for GMR 0.95 and CV 0.25: 28

sample size for GMR 0.95 and CV 0.25: 32

Bayesian Method

- ## Varying T/R ratio and *CV*, required sample size

Pivotal study (80% power) designed on results of a 2x2x2 pilot study with 16 subjects taking the uncertainty of estimated GMR into account (CV fixed).

Pivotal study (80% power) designed on results of a 2x2x2 pilot study with 16 subjects taking the uncertainties of both estimates (CV, GMR) into account.

sample size for GMR 0.95 and CV 0.25: 54
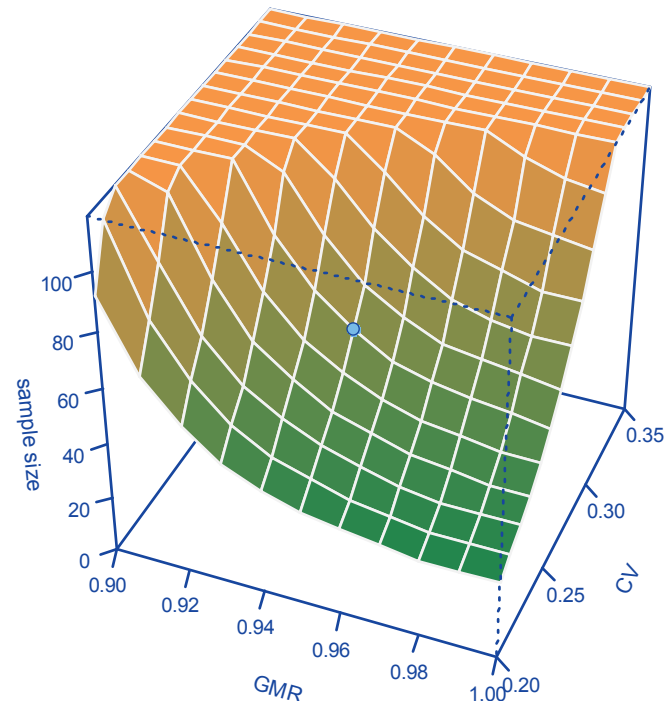
sample size for GMR 0.95 and CV 0.25: 70

Bayesian Method

- ## Feasible in practice?
    - Probably not
    - At least, if the pivotal study fails in a lower sample size, *you* know why and hope to successfully educate the *'Guy in the Armani Suit'* to be more conservative next time …

Reference-scaling (ABEL)

- If the assumed CV_{wR} is 40% and the actual CV_{wR} is larger (up to ~50%), power will *increase* (more expansion of the limits)

  – Different to ABE but this is the basic idea behind ABEL, *i.e.*, preserve power for HVD(P)s

  – Like in ABE the impact of potential deviations from assumptions is
    T/R ratio $\gg$ CV > dropouts

- If the actual CV_{wR} is smaller, power will *decrease* (less expansion of the limits)
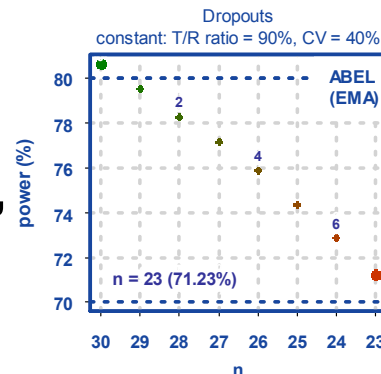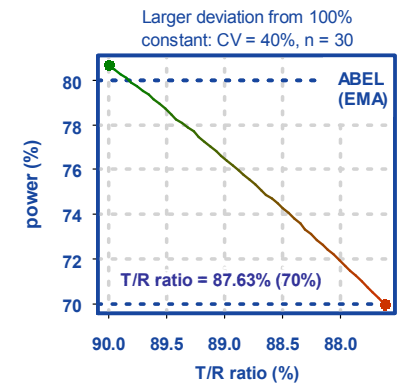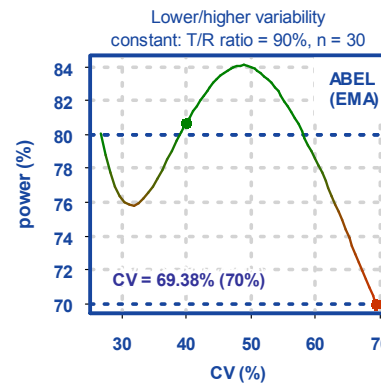
Reference-scaling (ABEL)

- Some large generic companies have a policy for pilot studies of HVD(P)s: Full replicate, 36 subjects
- Even if the pivotal study is planned as a *partial* replicate [*] design (TRR|RTR|RRT), perform the pilot in a *full* replicate to additionally estimate CV_{wT}
  - If $CV_{wT} < CV_{wR}$ there will be an incentive in the sample size Example
    - CV_{wT} 35%, CV_{wR} 50% estimated in the full replicate pilot study → Sample size 33
    - If the pilot was performed in a partial replicate we have no information about CV_{wT} and have to assume that $CV_{wT} = CV_{wR}$ → Sample size 39
  - It is not unusual that $CV_{wT} < CV_{wR}$, since technology improves and the reference might be a lousy product

[*] Not recommended if applying to the FDA. Details in the presentation about replicate designs.

Selection of Candidates (Part II)

- ## General rules
  - Do not assess the pilot with a pooled ANOVA but according to the 'Two-at-a-Time Principle' [1,2]
    - Exclude all candidates but one and perform the analysis as an incomplete block design
    - Repeat for the other candidates
    - A similar procedure is recommended in the EMA's guideline for studies with reference products from two regions
  - We get a set of ratios $\{C_1/R, \ldots, C_n/R\}$ and their *CV*s
    - Since the ratio is most critical select the candidate which is closest to 1
    - If some ratios are similar, select the candidate with the smallest *CV*

1. Schuirmann DE. *Two at a Time? Or All at Once?* IBS – ENAR Spring Meeting. Pittsburgh; March 28–31, 2004.
2. D'Angelo P. *Testing for Bioequivalence in Higher-Order Crossover Designs: Two-at-a-Time Principle Versus Pooled ANOVA*. 2nd GBHI Workshop. Rockville; September 15–16, 2016.

Selection of Candidates (Part II)

- ## HVD(P)s are difficult

  - Two candidates: Design the pilot like a 4-sequence 4-period full replicate and substitute T with C_1 and C_2, *i.e.*, from

    - TRTR | RTRT | TRRT | RTTR

    to

    - C_1RC_2R | RC_1RC_2 | C_2RRC_1 | RC_2C_1R

  - After exclusion we get two partial replicates with missings *

    - $C_1R * R$ | $RC_1R *$ | $* RRC_1$ | $R * C_1R$
    - $* RC_2R$ | $R * RC_2$ | $C_2RR *$ | $RC_2 * R$

  - Select the candidate with the ratio closest to 1

  - Drawback: In sample size estimation we have to assume $CV_{wT} = CV_{wR}$

  - More than two candidates are very difficult; needs many sequences to get balance – consult with a statistican

Selection of Candidates (Part II)

- ## Two candidates in the *pivotal* study
  - Pilot was indecisive (very similar T/R ratios and *CV*s)
    - Some companies are wary to select one based on 'gut feelings' and include *both* in the pivotal study
      - Submit the 'better' one to the authority and stop developing the other
    - Opinion split amongst statisticians
      - Since only one product will be marketed, this approach does not increase the patient's risk (90% CI is sufficient)
      - The company has to two chances to show BE, which will increase the Type I Error
        - » Bonferroni's adjustment (95% CI) to control the patient's risk
        - » ~25% more subjects required to maintain power

- ## If two products should be marketed (*e.g.*, tablet, capsule)
  - Bonferroni's adjustment (95% CI) <u>mandatory</u>

Conclusions

- Design pilot studies as large as the budget allows
  - Increases the precision of estimates
  - Adjusting for the uncertainty of the T/R ratio (even with the Bayesian method) leads to sample sizes of the pivotal study which likely are not feasible
  - Take all available information about the T/R ratio into account (*e.g.*, from f_2 of BCS I(III) or an existing *IVIVC*)

- In designing the pivotal study do not assume perfectly matching products
  - Even if you observe a 'nice' T/R ratio in the pilot study be conservative
  - For ABE do not assume a T/R ratio of 'better' than 0.95 and for ABEL not 'better' than 0.90

Alternatives

- ## Sample size based on statistical assurance
  - Still requires a pilot study
  - Instead of an arbitrary T/R ratio, we assume
    - matching products and
    - how variable the T/R ratio is

- ## Two-Stage Designs
  - Unlike in the combination pilot/pivotal the information is not lost
  - Adjusts the sample size based on the *CV* and/or the T/R ratio observed in the first stage
  - May include futility criteria for early stopping and/or a maximum total sample size

---

* Ring A, Lang B, Kazaroho C, Labes D, Schall R, Schütz H. *Sample size determination in bioequivalence studies using statistical assurance*. Br J Clin Pharmacol. 2019; 85(10): 2369–77. doi:10.1111/bcp.14055.

Thank You!

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