PK-based Design,
Sample Size Considerations

NCA *vs.* PK Modeling in BE

- Pharmacokinetic Models
  - Very useful for understanding the drug and formulation
    - Study design of BA/BE
      - Length of sampling (AUC_{0-t} should cover ≥80% of $AUC_{0-\infty}$) and washout (no residual concentrations from earlier periods)
      - Degree of accumulation / number and of doses / dosing interval to reach steady state
  - Drawbacks
    - Difficult to validate (fine-tuning of side conditions, weighting schemes, software's algorithms, …)
    - Still a mixture of art and science
    - Practically impossible to recalculate any given data set using different software – sometimes even with different versions of the same software
    - Not acceptable for evaluation of BA/BE studies!

NCA *vs.* PK Modeling in BE

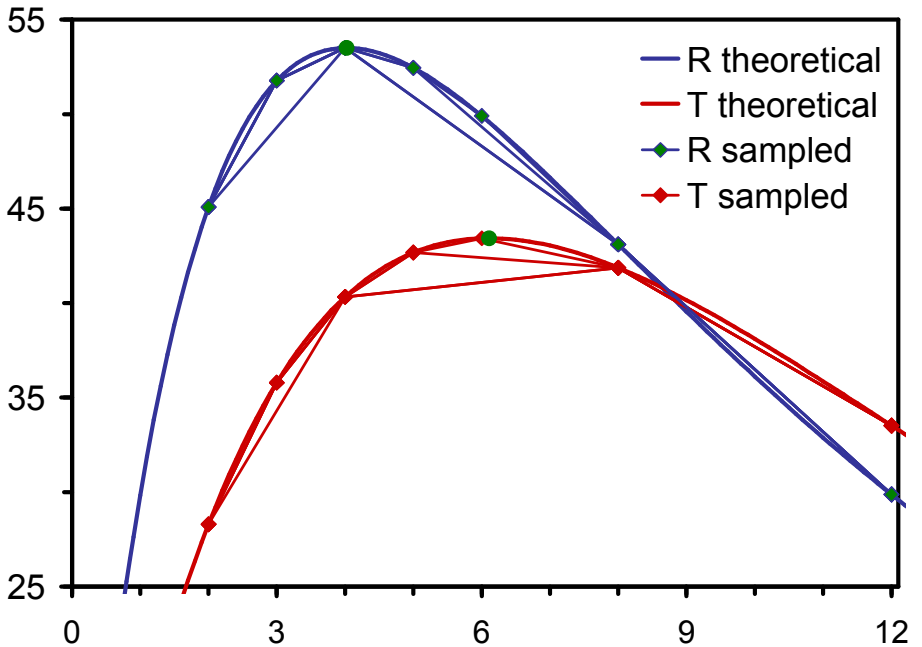
- Nonparametric Superposition is an alternative
  - Designing multiple dose studies based on single dose data
    - Concentrations of a single dose study are stacked according to the desired dosing interval while adding the time course of eliminated concentrations of previous doses (Dost 1953)
    - Limitations
      - Linear PK has to be assumed
      - Requires reliable estimate of λ_z
      - Equal doses
      - Equal dosing intervals
    - Implemented in Phoenix/WinNonlin, Kinetica, ThothPro
    - With experience and patience possible in any spreadsheet and statistical software (SAS, R, MATLAB, …)

PK-based Design

- Sampling at t_{max}
  - With *any* sampling scheme the 'true' C_{max} is missed (one cannot sample exactly at the true C_{max} for any given subject)
  - High inter- and/or intra-subject variability (single point metric)
  - Variability higher than the one of *AUC*
  - In many studies the win/loose metric!
  - Remedies
    - Sample size based on the variability of C_{max}– never of *AUC*
    - Sufficient numer of samples in the area of the expected t_{max}

PK-based Design

- Sampling at t_{max}
  - Theoretical values (from PK simulation)
    C_{max} 41.9 (T) / 53.5 (R), T/R 81.2%
    t_{max} 6.11 (T) / 4.02 (R), Δ 2.09
  - Number of samples within 2 – 12 hours (n), estimated T/R-ratio for C_{max} and for $\Delta\ t_{max}$
    - n = 4
      78.3%, 4
    - n = 5
      78.3%, 4
    - n = 6
      79.8%, 1
    - n = 7
      81.2%, 2

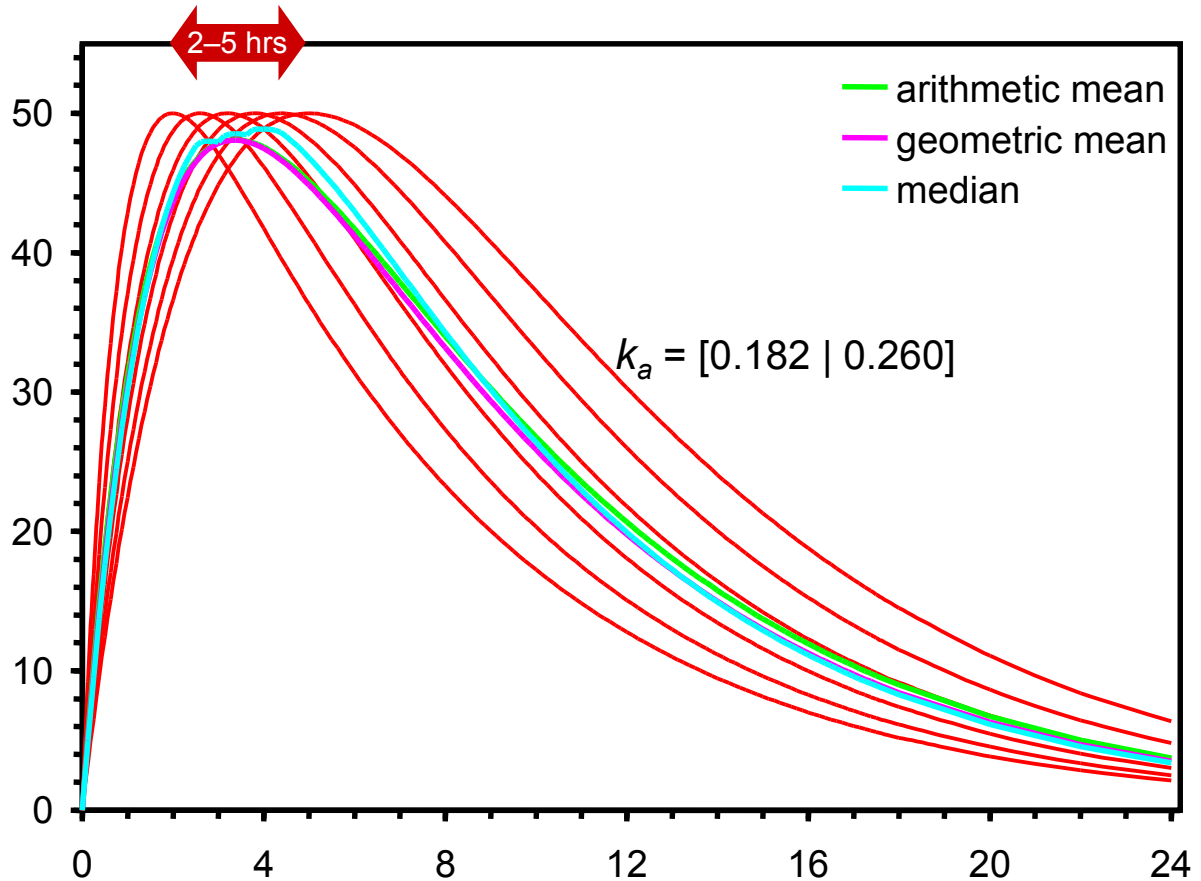


PK-based Design

- Sampling at t_{max}
  - Quote from the literature:
    'Maximum concentrations were observed within two to five hours after oral administration.'
  - Elimination is drug specific,
  - but what about absorption?
    - Formulation specific!
    - Dependent on the sampling schedule (therefore, in a strict sense study-specific)

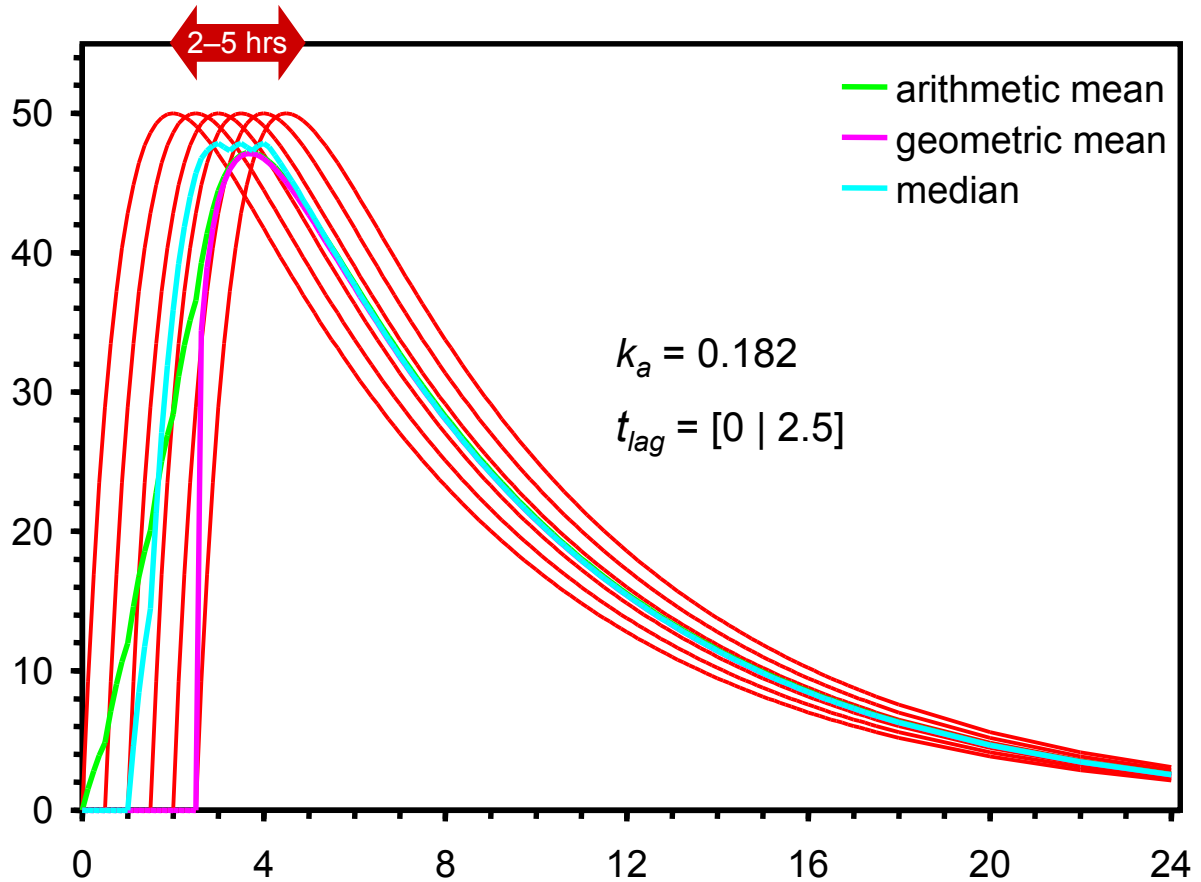
PK-based Design

- Sampling at t_{max} (absorption rate variable, no lag times)



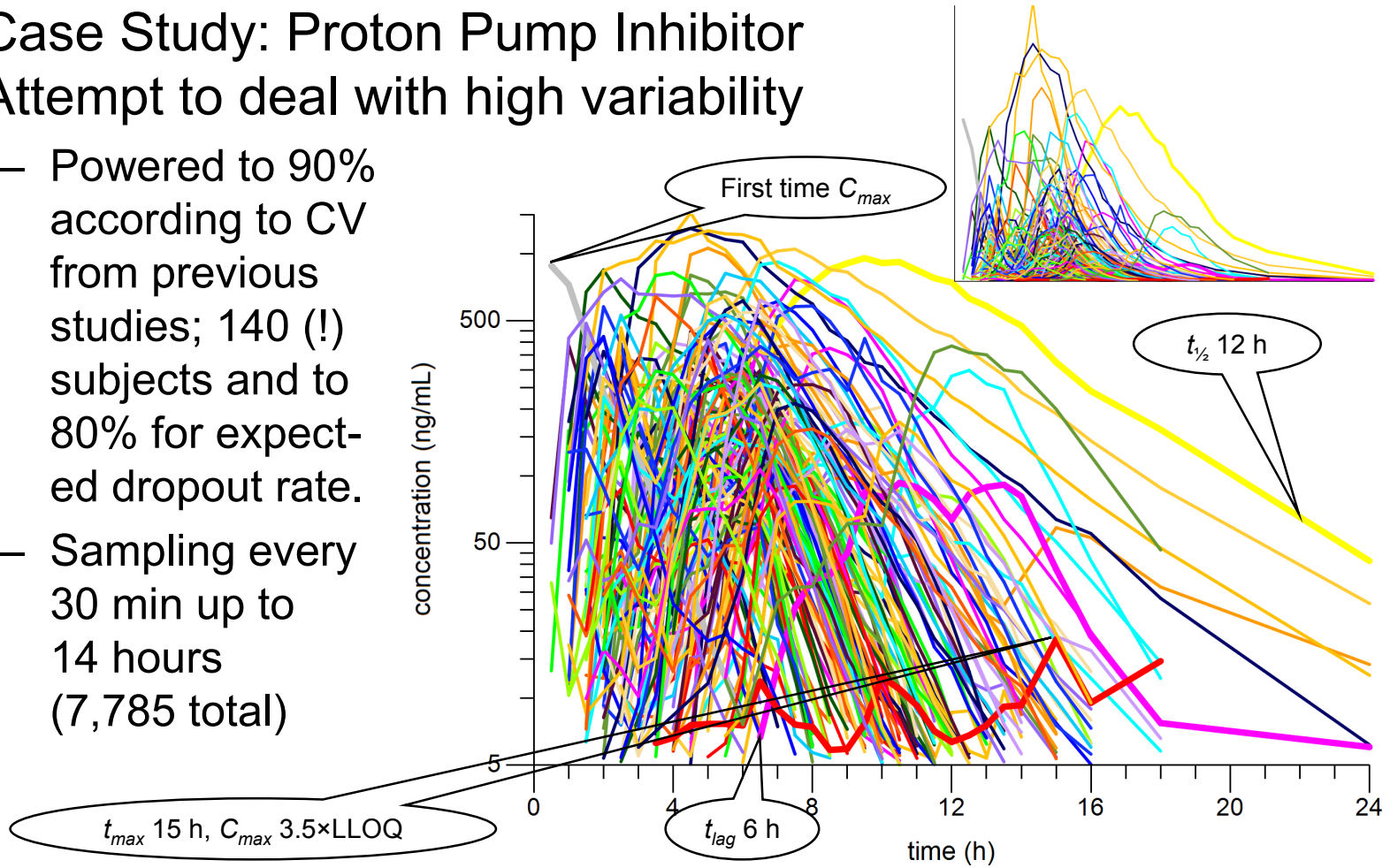
PK-based Design

- Sampling at t_{max} (absorption rate const., lag time variable)



PK-based Design

- Case Study: Proton Pump Inhibitor
  Attempt to deal with high variability
  - Powered to 90% according to CV from previous studies; 140 (!) subjects and to 80% for expect-ed dropout rate.
  - Sampling every 30 min up to 14 hours (7,785 total)



Sample Size Considerations

- Recap
  - Minimum sample size generally 12
  - Maximum not specified in GLs; high ones ethically problematic
  - Recommended power (chance to pass) 80 – 90%
  - ICH E9, Section 3.5

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

Sample Size Considerations

- Power *vs.* Sample Size
  - It is not possible to *directly* obtain the required sample size
    - The required sample size depends on *five* values, namely
      - the acceptance range (AR) for bioequivalence;
      - the error variance (s^2) associated with the PK metrics as estimated from
        - » previous studies, a pilot study, or published data;
      - the fixed significance level (α);
      - the expected deviation (Δ) from the reference product and;
      - the desired power ($1 - \beta$).
  - Three values are known and fixed (AR, α, $1 - \beta$), one is an estimate (s^2), and one an assumption (Δ)
    - Hence, the correct term is ‘sample size estimation’ and not ‘sample size calculation’

Sample Size Considerations

- Power *vs.* Sample Size
  - Only power is accessible
    - The sample size is searched in an iterative procedure until at least the desired power is obtained
      Example: α 0.05, AR 80 – 125%,
      target power 80% (β 0.2), assumed *GMR* 0.95,
      CV_{intra} 20% →
      minimum sample size 19 (power 81.3%),
      rounded up to the next even number in
      a 2×2×2 study (power 83.5%)

n	power (%)
16	73.5
17	76.4
18	79.1
19	81.3
20	83.5

    - Exact methods for ABE in parallel, crossover, and replicate designs are available
    - Simulations recommended for Group-Sequential and Two-Stage Designs
    - Simulations mandatory for reference-scaling methods

Sample Size Considerations

- Power *vs.* Sample Size
  - Can be performed in the open-source package PowerTOST * for R
    - Examples (after `library(PowerTOST)`)
      - *CV* 40%, *GMR* 0.95%, power 80%, parallel design
        ```
        sampleN.TOST(CV=0.40, theta0=0.95, targetpower=0.80,
        design="parallel")[["Sample size"]]
        [1] 130
        ```
      - *CV* 20% *GMR* 0.95%, power 80%, 2×2×2 crossover design
        ```
        sampleN.TOST(CV=0.20, theta0=0.95, targetpower=0.80,
        design="2x2x2")[["Sample size"]]
        [1] 20
        ```
      - *CV* 50% *GMR* 0.90%, power 80%, 2×2×4 full replicate design for the EMA'/WHO' reference-scaling of HVD(P)s
        ```
        sampleN.scABEL(CV=0.50, theta0=0.90, targetpower=0.80,
        design="2x2x4")[["Sample size"]]
        [1] 28
        ```

* Labes D, Schütz H, Lang B. *PowerTOST: Power and Sample Size Based on Two One-Sided t-tests (TOST) for (Bio)Equivalence Studies.* 2018; R package version 1.4.9. https://cran.r-project.org/package=PowerTOST.

Sample Size Considerations

- Power *vs.* Sample Size
  - Examples (cont'd)
    - *CV* 50% *GMR* 0.90%, power 80%, 2×2×4 full replicate design for the FDA's reference-scaling of HVD(P)s

```
sampleN.RSABE(CV=0.50, theta0=0.90, targetpower=0.80,
design="2x2x4")[["Sample size"]]
[1] 28
```

    - *CV* 10% *GMR* 0.975%, power 80%, 2×2×2 crossover design for the EMA/WHO – narrower limits for NITIDs

```
sampleN.TOST(CV=0.10, theta0=0.975, targetpower=0.80,
design="2x2x2", theta1=0.90, theta2=1/0.90)[["Sample size"]]
[1] 22
```

    - *CV* 10% *GMR* 0.975%, power 80%, 2×2×4 full replicate design for the FDA's reference-scaling of NTIDs

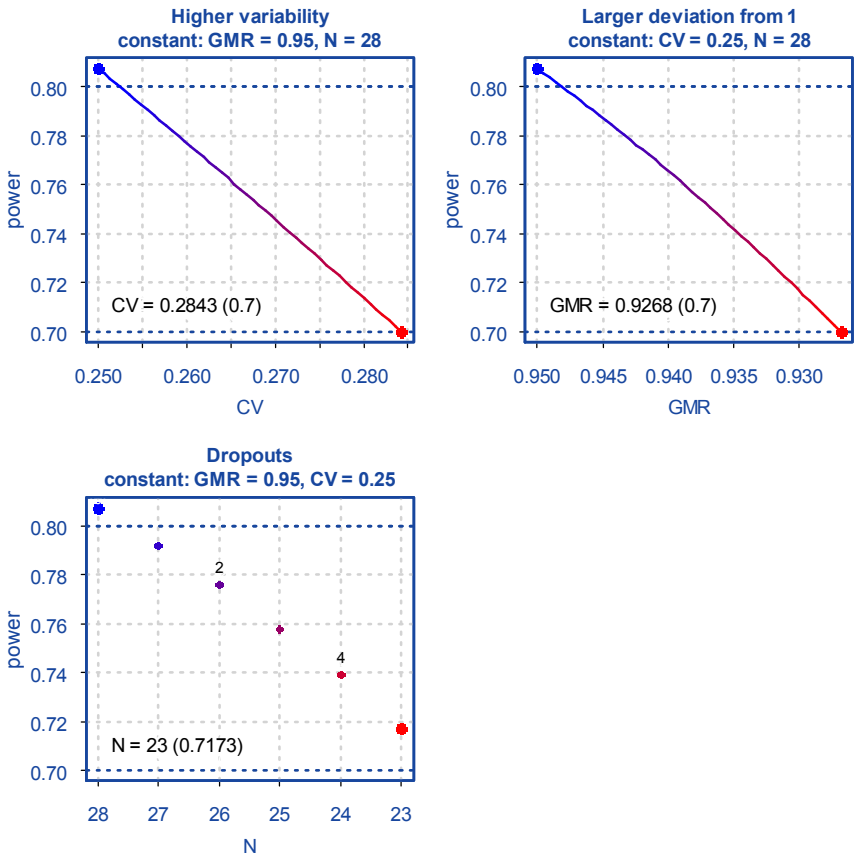
```
sampleN.NTIDFDA(CV=0.10, theta0=0.975, targetpower=0.80,
design="2x2x4")[["Sample size"]]
[1] 18
```

Sample Size Considerations

- Power *vs.* Sample Size
  - However, all results are based on assumptions
  - ICH E9 recommends a sensitivity analysis to explore the impact on power if values deviate from assumptions

Sample Size Considerations

- Power *vs.* Sample Size
  - Example ABE, 2×2×2 Design
    - Assumed *GMR* 0.95, α 0.05, AR 80–125%, CV_{intra} 0.25 (25%) desired power 80%, min. acceptable power 70%
      - Sample size 28 (power 0.807)
      - CV_{intra} ↑ 0.284 (rel. +14%)
      - *GMR* ↓ 0.927 (rel. –2.4%)
      - 5 drop-outs acceptable (rel. –18%)
      - Most critical is the *GMR*



Sample Size Considerations

- Dealing with Uncertainty
  - One should never assume perfectly matching products
  - Recommended ∆
    - Conventional ABE Not better than 5% (*GMR* 0.9500 – 1.0526)
    - HVD(P)s Not better than 10% (*GMR* 0.9000 – 1.1111)
    - NTIDs Not better than 2.5% (*GMR* 0.9750 – 1.0256)
  - The *CV* from previous studies, a pilot study, or the literature is not 'carved in stone'
    - Don't use the value as it is but its (upper) confidence limit
      - As usual, the confidence interval narrows with increasing sample size
      - The larger a previous study was, the more accurate the estimated *CV*
      - Very small pilot studies are practically useless for the estimation of the CV
      - Example: CL of *CV* 25% estimated from a study with *n* subjects
        39.8% (*n* = 6), 32.1% (*n* = 12), 30.6% (*n* = 18)

Sample Size Considerations

- Ethical Issues
  - ‘Demonstrating BE’ in Pilot Study
    - The purpose of a pilot study (amongst others) is to obtain estimates of the *GMR* and *CV* which can be used to design the pivotal study
    - In a strict sense it is not possible to demonstrate bioequivalence in a pilot study which is − by definition − exploratory
    - Acceptable
      - FDA (if at least 12 subjects and properly performed)
      - In the past some agencies (Scandinavian countries, Germany) accepted pilot studies as evidence of BE if stated in the protocol
        - » Repeating a ‘passing’ pilot (even in a larger sample size) may fail by pure chance (producer’s risk = 1 − power)
        - » Hence, this approach was considered unethical
    - Nowadays, European regulatory agencies are seemingly more strict (follow the ‘cook book’)