**Arbeitsgemeinschaft
für angewandte
Humanpharmakologie e.V.**

Pitfalls in BA/BE

Attempts in beating Murphy's law: Learnings from failures in study design, bioanalytics and statistics

Helpful (?) quotations

If anything can go wrong, it will. — *Edward A. Murphy Jr.*

He who fails to plan is planning to fail. — *Winston Churchill*

You can't fix by analysis what you bungled by design. — *Richard J. Light, Judith D. Singer, John B. Willett*

100% of all disasters are failures of design, not analysis. — *Ronald G. Marks*

To propose that poor design can be corrected by subtle analysis techniques is contrary to good scientific thinking. — *Stuart J. Pocock*

To call the statistician after the experiment is done may be no more than asking him to perform a *postmortem* examination: He may be able to say what the experiment died of. — *Ronald A. Fisher*

If you think it's simple, then you have misunderstood the problem. — *Bjarne Stroustrup*

Whenever a theory appears to you as the only possible one, take this as a sign that you have neither understood the theory nor the problem which it was intended to solve. — *Karl R. Popper*

Study design

- In a crossover-study the washout between treatments has to be sufficiently long
  - Avoid pre-dose concentrations which are residuals of previous period(s)
  - In order to get an unbiased estimate of treatment differences the physiological state of subjects in higher period(s) has to be the same as in the (drug-naïve) first period
  - Cave
    - Never plan the washout (generally ≥5times the apparent half life) based on an average. Keep the distribution of half lives in mind. Some subjects might show a substantially longer half life – especially if the drug is subjected to polymorphism (poor and extensive metabolizers).
    - Think also about PD. If the drug is an auto-inducer or -inhibitor allow the body to return to its original state.

Case 1 | Study design

- Drug A: $t_{½}$ 60 – 100 h (literature)
  - BA study
    - 10 mg drug A hydrochloride p.o. *vs*. i.v.
    - 12 subjects
    - 2×2×2 crossover, washout 35 days
    - Sampling until 312 hours post dose
    - LC/MS-MS, LLOQ 1 ng/mL (drug A base / plasma)
    - Results considered important for designing other studies
      - $t_{½}$ 49.9 ± 13.0 h (harmonic mean ± jackknife standard deviation)
      - In none of the samples drawn at 312 h a concentration ≥LLOQ was measured
      - Extrapolated *AUC* 10.0% (median)
        3.8% – 13.9% (minimum – maximum)

Case 1 | Study design

- Drug A: $t_{½}$ 60 – 100 h (literature)
  - Comparative BA study aiming to demonstrate BE
    - 10 mg drug A hydrochloride
      (primary target T_2 *vs.* R, descriptive T_2 vs. T_1)
    - 36 subjects
    - 3×6×3 crossover (Williams' design), washout 14 days
      - Washout planned for a worst case $t_{½}$ of 66 h (covering >5 half lives)
    - Sampling until 216 hours post dose
      - No problems with extrapolated *AUC* expected (simulations)
    - GC/MS, LLOQ 0.117 ng/mL (drug A base / plasma)
  - Given that, what happened – and why?

Case 1 | Study design

- Pre-dose concentrations ≥LLOQ: n (% of subjects, geom. means)
  - Period 1: all <LLOQ
  - Period 2: 21 (58%, 0.226 ng/mL)
  - Period 3: 18 (50%, 0.222 ng/mL)

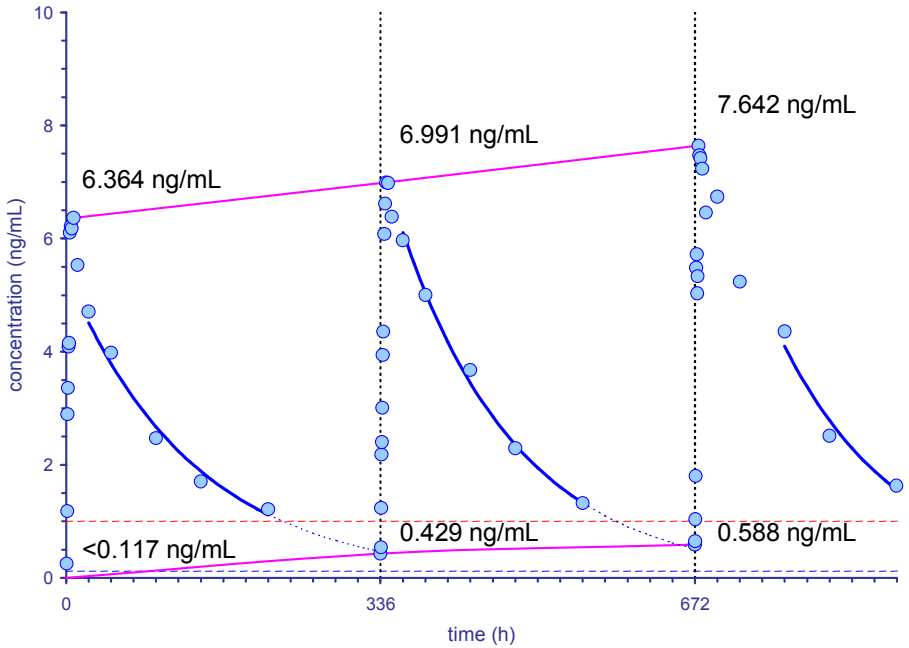
  } carry-over
- Half lives (harmonic means)
  - Period 1: 51.68 h
  - Period 2: 54.20 h
  - Period 3: 63.03 h

  } increasing with time

– Issues
- Improving the bioanalytical method (~9times lower LLOQ) was not a good idea
  - If we would have used the old method we would have seen not a single (!) pre-dose concentration >LLOQ
- The shorter washout (35 days → 14) was not a good idea as well
  - Only if the estimation of λ_z is performed *blinded for* treatment different half lives in the periods (due to accumulation) become evident – even with the less sensitive method

Case 1 | Study design

- Most statisticians unblind studies *before* performing NCA, which will cover potential problems
  - Half lives (harmonic means)
    - » T_1: 54.51 h
    - » T_2: 55.99 h
    - » R: 56.73 h
- Worst case Subject 23



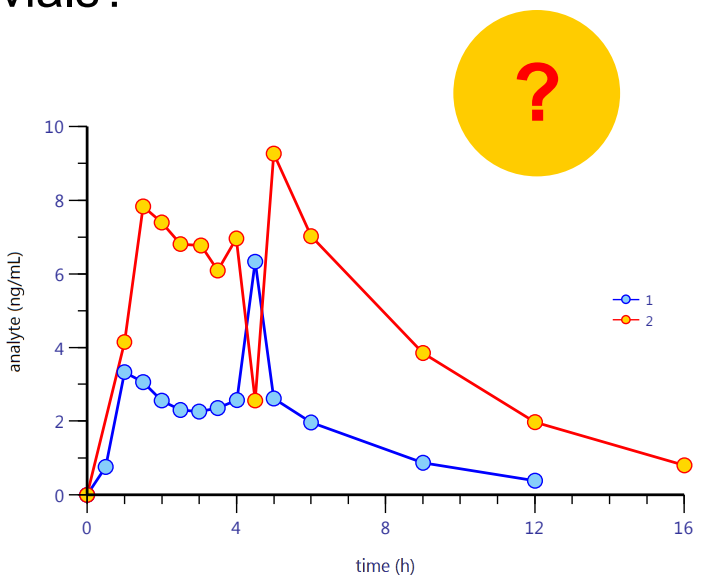
NCA

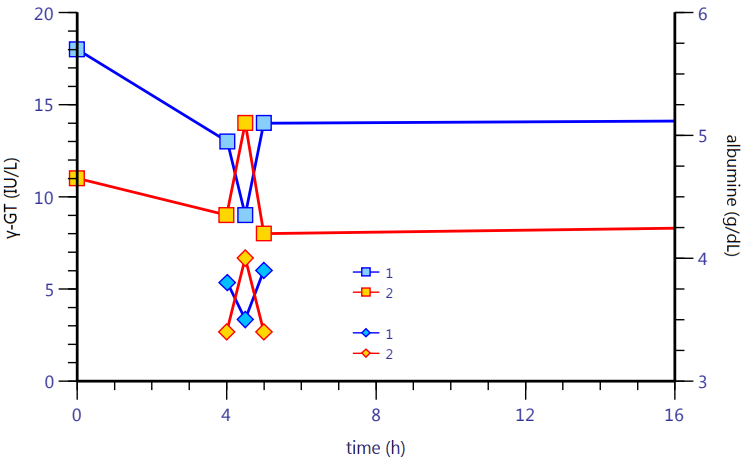
- Requirements for BA/BE studies
  - Bioanalytical method developed and validated *to serve the study's purpose*
    - Calibration range
      - LLOQ ≤5% C_{max} in any of the subjects
      - ULOQ ideally ≥C_{max} in any of the subjects
    - (In)accuracy and (im)precision
      - 15% throughout the range (20% for ligand-binding assays)
      - 20% at LLOQ (30% for ligand-binding assays)
  - Sampling long enough to obtain reliable estimates of
    - λ_z : at least three samples in the log/linear part
    - AUC_{0-t} : covering ≥80% of $AUC_{0-\infty}$ in ≥80% of observations
    - Note that both are *not required* if target metric is AUC_{0-72h} (IR single dose) or $AUC_{0-\tau}$ (steady state)

Case 2 | Sample handling

- Clinical phase
  - Biphasic modified release product of drug B
  - Mix-up in the transfer from sample vials after centrifu-gation to (plasma) sample vials?

Measurable values in clin. chemistry (limited, since anticoagulant citrate)

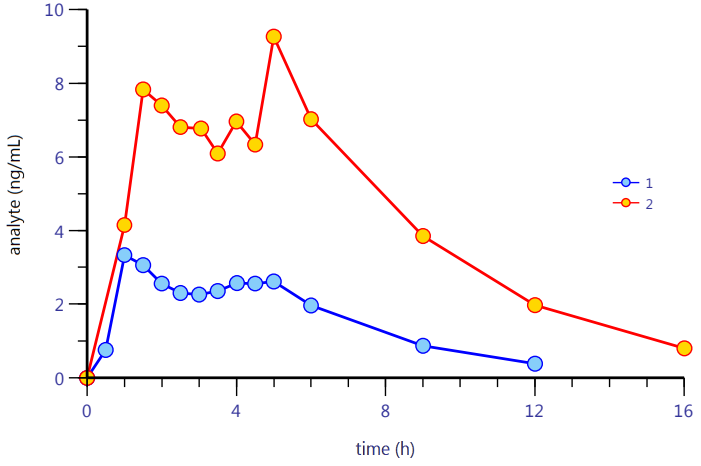


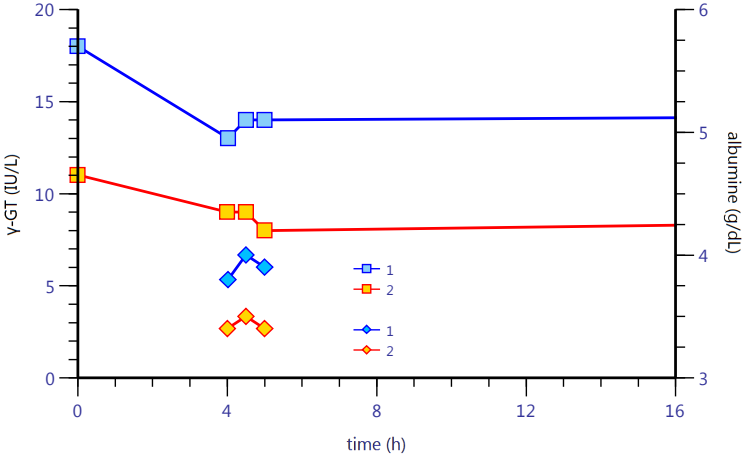


Case 2 | Sample handling

- Clinical phase
  - Biphasic modified release product of drug B
  - Exploratory: values swapped (both analyte and clin.chemistry)
  - Samples of subjects 1 & 2 both taken in the first period

Suspected mix-up confirmed by clin. chemistry values





Case 2 | Sample handling

- Clinical phase
  - Barcode system failed in the first period
  - No bail-out procedure (*e.g.*, four-eye principle)
  - Sponsor monitored plasma separation only up to two hours (when the barcode stystem was still working)
  - Blinded review of data for irregular profiles?
    - EMA
      - According the Bioanalytical Method Validation Guideline measured results are ‘carved from stone’
        - » Exclusion of data only possible if documented error
        - » Not even repeated analysis acceptable
    - FDA
      - Was acceptable until 2022
        - » Exclusion after repeated analysis possible if defined by SOP
      - Currently like EMA

Case 3 | Sample handling

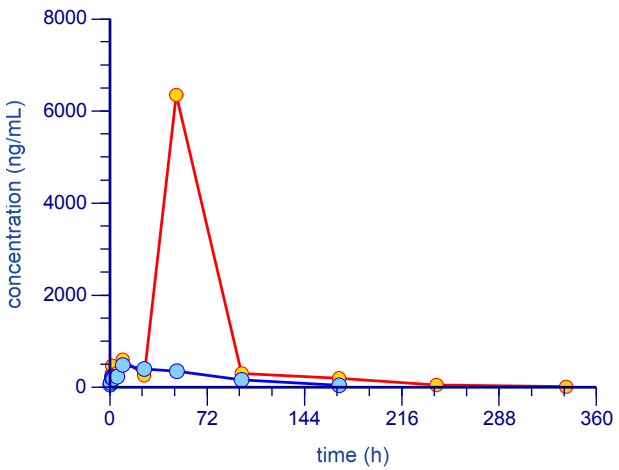
- Clinical phase
  - Liposome encapsulated drug C for infusion
  - Analytes
    - encapsulated drug C
    - unencapsulated drug C (*i.e.*, released from the liposomes)
    - total drug C (encapsulated + unencapsulated)
    - Metabolite (formed from unencapsulated drug C only)
  - Drug may be released from liposomes by
    - Shear forces (infusion pump, infusion needle with narrow diameter)
    - High temperature and long interval until centrifugation
    - High *g* force in centrifugation
    - Only the latter two can be avoided by proper stabilization
      - blood samples on ice
      - addition of DMSO

Case 3 | Sample handling

- Clinical phase
  - Multi-site study in cancer patients
  - Clinical staff educated about critical sample handling, but
    - unfamililar procedure esp. in small clinical sites
    - necessity of following SOPs and documentation of deviations in conformity with GCP not well understood

Case 3 | Sample handling

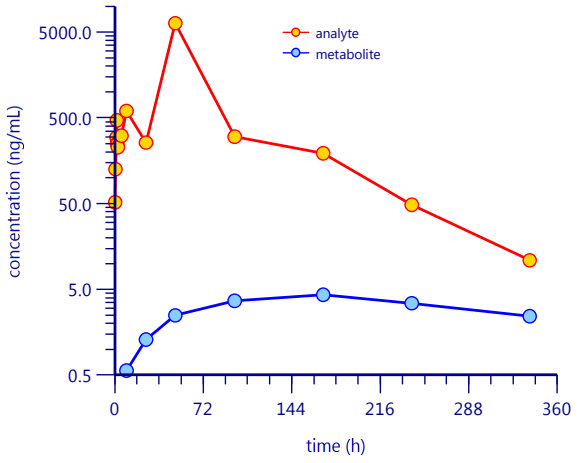
- Clinical phase
  - Surprises in bioanalytics
  - Extremely high concentrations of unencapsulated drug C observed in about 2% of samples
    - All values confirmed in repeated analyses (note: against guidelines)

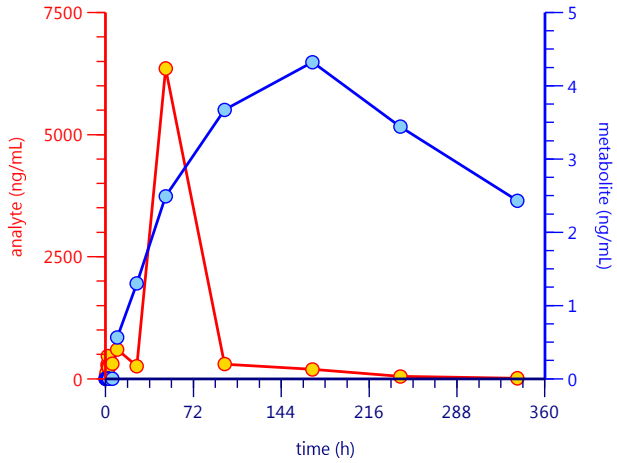


?

Case 3 | Sample handling

- Clinical phase
  - Extremely high concentrations of unencapsulated drug C observed in about 2% of samples
    - However, ‘normal’ concentrations of the metabolite
      - Since the metabolite can only be formed from the unencapsulated drug C, the analyte’s high concentrations were considered an artifact
      - No documentation about improper sample handling (temperature, time, stabilization)



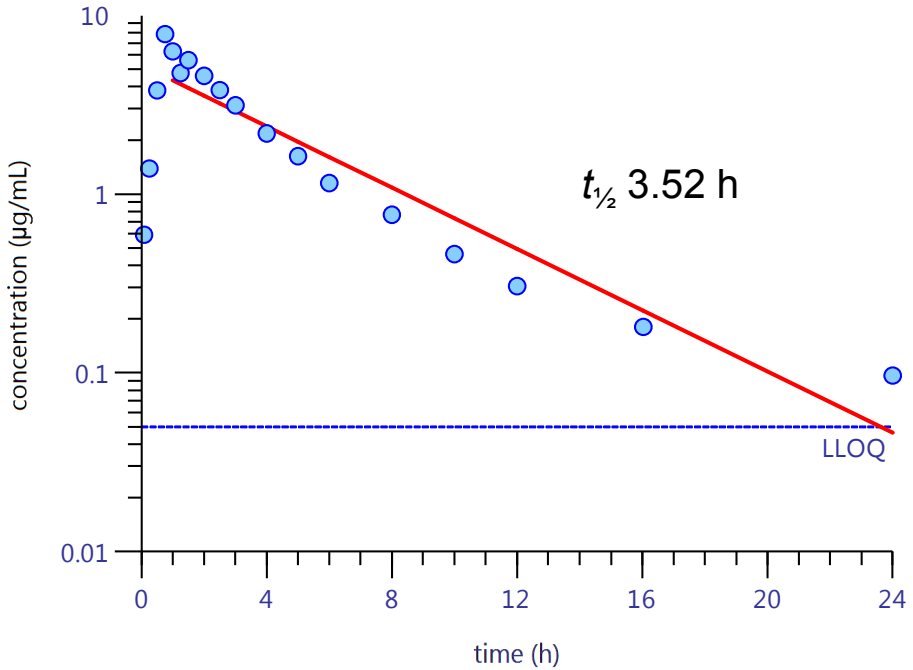


Case 4 | NCA

- Drug D: $t_{½}$ 2 – 3 h (literature)
  - BE study (500 mg D component of a three drug FDC)
    - liquid formulations, T *vs*. R
    - 27 subjects
    - TRR | RTR | RRT replicate design, washout seven days
    - Sampling until 24 hours post dose
    - LC/MS-MS, LLOQ 50 ng/mL
  - Drug D passed ABE with ease
    - $t_{½}$ 3.92 ± 0.88 h (T), 4.98 ± 1.24 h (R)
    - Extrapolated *AUC* (median, minimum – maximum)
      T: 1.76 (0.87% – 3.61%), R: 2.42% (1.14% – 6.19%)
  - Sponsor developed a 4 drug FDC
    - Data of the BE study should be used in a PopPK model to optimize the sampling schedule for a new study

Case 4 | NCA

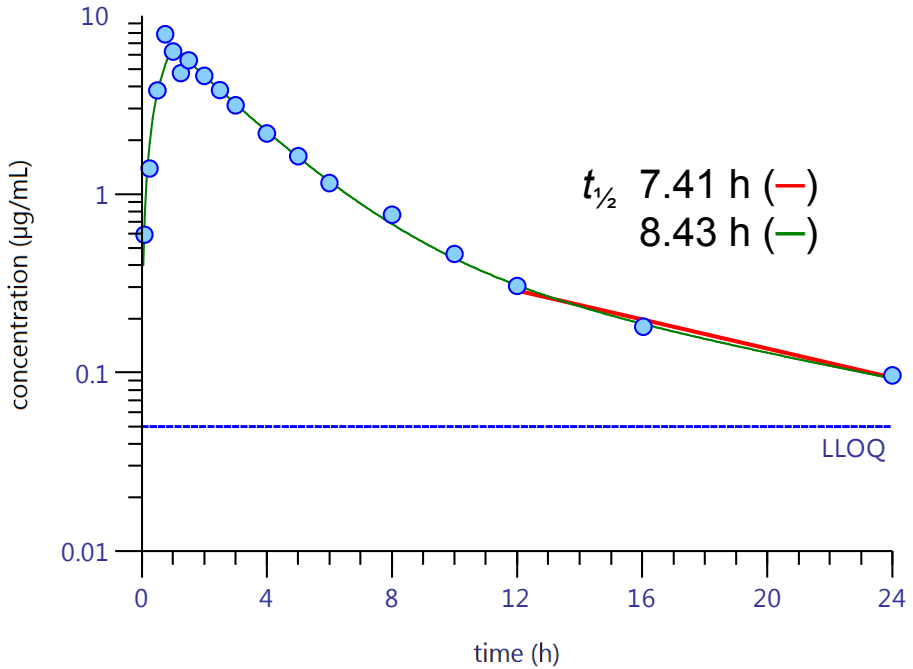
- Drug D: $t_{½}$ 2 – 3 h (literature)
  - No individual λ_z or $t_{½}$ (as well as the time range used in estimation) given in the report, only $AUC_{0–t}$ and $AUC_{0–\infty}$
  - Reproduced the CRO's results by trial and error. Example:



?

Case 4 | NCA

- Drug D: $t_{½}$ 2 – 3 h (literature)
  - Obviously the time range for the estimation of λ_z was wrong
    - Two-compartment model!
  - What I obtained in NCA (—) and by the PK model (—)

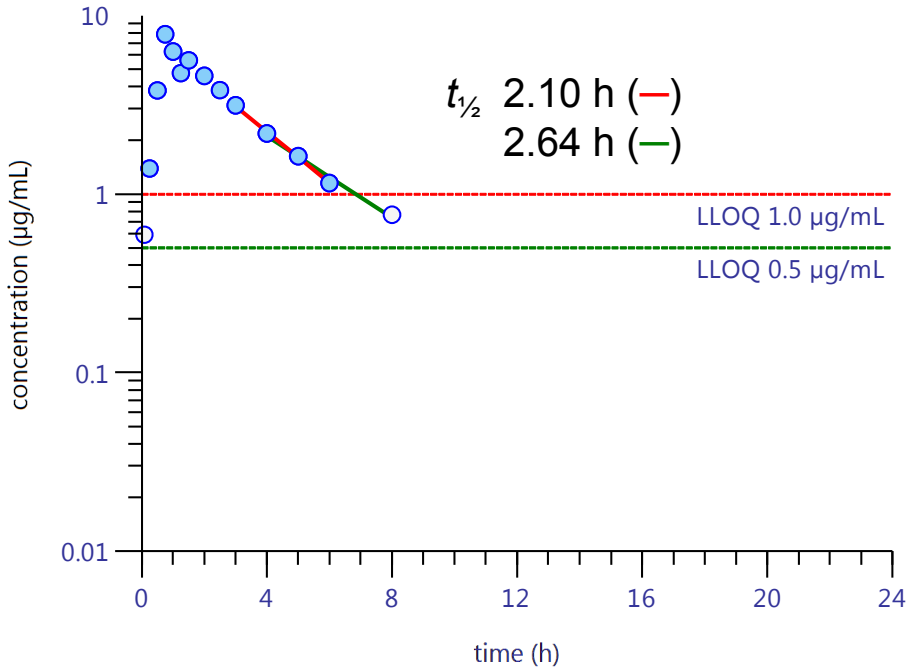


Case 4 | NCA

- Drug D: $t_{½}$ 2 – 3 h (literature)
  - Why? No problems with correct estimation of λ_z
    - $t_{½}$ 4.63 ± 1.07 h (T), 5.59 ± 1.19 h (R)
    - Extrapolated *AUC* (median, minimum – maximum)
      T: 2.08% (1.06% – 4.32%), R: 2.84% (1.47% – 6.19%)
  - Possible explanations
    - 'Push-the-button-pharmacokineticist' at work
      - Relied on an automatic algorithm?
      - No visual inspection of fits?
    - Anticipatory obedience ('vorauseilender Gehorsam')?
      - The bioanalytical method was at least 10times more sensitive than ones used in the past (drug D approved in 1955)
      - Maybe the CRO wanted to avoid a single sentence in the discussion section explaining why a second phase is apparent – explaining a half life longer than the one known from literature

Case 4 | NCA

- Drug D: $t_{½}$ 2 – 3 h (literature)
  - Estimation of λ_z by bioanalytical methods with a LLOQ of 1.0 or 0.5 µg/mL explains short half lives given in ‘old’ literature



Case 4 | NCA

- Drug D: $t_{½}$ 2 – 3 h (literature)
  - Lessons learned
    - Insist that the (draft!) PK report allows independent assessment
    - Mandatory[1,2]
      - All raw data
      - λ_z and/or $t_{½}$ as well as time ranges used in estimation
      - All derived PK metrics
    - Desirable
      - Data in a machine-readable format (CSV, SAS transport, CDISC)
    - Unacceptable
      - A 500+ page PDF generated by SAS
      - As above but a scanned print

1 Schulz H-U, Steinijans, VW. *Striving for standards in bioequivalence assessment: a review*. Int J Clin Pharm Ther Toxicol. 1991;29(8):293–8. PMID 1743802.

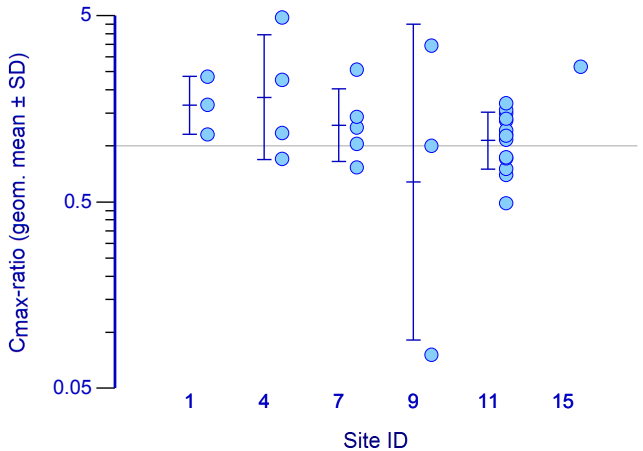
2 Sauter R, Steinijans VW, Diletti E, Böhm E, Schulz H-U. *Presentation of results from bio-equivalence studies*. Int J Clin Pharm Ther Toxicol. 1992;30(Suppl.1):S7–30. PMID 1601535.

Statistics

- Requirements for BA/BE studies
  - Design should allow accurate (unbiased) assessment of the treatment effect
  - EMA (2010), ICH M13A Draft (2023)
    - The study should be designed in such a way that the formulation effect can be distinguished from other effects.
    - The precise model to be used for the analysis should be pre-specified in the protocol. The statistical analysis should take into account sources of variation that can be reasonably assumed to have an effect on the response variable.

Case 5 | Statistics

- Continuing Case 3
  - Extreme C_{max}-ratios of unencapsulated drug C observed only in small clinical sites



  - Pooling data of sites *only* if
    - similar variances
    - no treatment-by-site interaction

?

Case 5 | Statistics

- Planned evaluation of unencapsulated drug C
  - Statistical model *did not* take the multi-site nature of the study into account (*i.e.*, data of all sites were naïvely pooled);
    - Failed: PE 121.01% (90% CI: 96.13 – 152.33%), CV_w 55.6%
- Sensitivity analysis
  - Statistical model suggested by the FDA including the site-by-treatment interaction
    - Highly significant (p 0.00063)
    - Hence, pooling of sites is *not* justified
  - Therefore, analysis of largest site #11 only
    - Passed: PE 103.80% (90% CI: 89.87% – 119.90%), CV_w 21.4%

Statistics

- Adaptive Two-Stage Sequential Design in BE
  - EMA (2010), ICH M13A Draft (2022)
    It is acceptable to use a two-stage approach […]. If this approach is adopted appropriate steps must be taken to preserve the overall type I error of the experiment […]. 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 94.12% CI (α 0.0294) preserves the patient's risk in simulation-based methods *if and only if*
    - GMR 0.95 and
    - Target power 80%

Case 6 | Statistics

- Adaptive Two-Stage Sequential Design in BE
  - GMR 0.90, target power 85%, α 0.0294
  - Stage 1: n_1 24
    - Failed: PE 89.00% (94.12% CI: 77.24 – 102.54%)
    - Since interim power 37.7%, stage 2 with 54 subjects initiated
  - Analysis of pooled data: n_1+n_2 78
    - Passed: PE 91.00% (94.12% CI: 82.16 – 100.79%)

Case 6 | Statistics

- Adaptive Two-Stage Sequential Design in BE
  - GMR 0.90, target power 85%, α 0.0294
  - Not according to the method (GMR 0.95, target power 80%)
  - Inflated patient's risk (5.23%)
  - The study's conditions would require *more* adjustment (at least an α of 0.0279 or a 94.42% CI)
    - *Post hoc* assessment
      - Passed: PE 91.00% (94.42% CI: 82.05 – 100.92%)
      - Type I Error 4.99%
      - Wider CI but conclusion agrees with the original analysis
      - Might (!) be accepted by agencies

Case 6 | Statistics

- Adaptive Two-Stage Sequential Design in BE
  - However, correct would have been to find a suitable α (0.0278) already *before* the study and implement it in the study
  - Stage 1: n_1 24
    - Failed: PE 89.00% (94.44% CI: 77.09 – 102.75%)
    - Since interim power 36.6%, stage 2 with 54 subjects initiated
  - Analysis of pooled data: n_1+n_2 78
    - Passed: PE 91.00% (94.44% CI: 82.05 – 100.93%)
  - Controlled patient's risk (4.99%)