Outliers in Bioequivalence

PK Metrics and Irregular Profiles

Outliers

- Definition
  - Depending on the field
    - Some industries $\pm 6\sigma$
    - Particle physics $\pm 5\sigma$
    - BE open issues
- Basics
  - If a normal distribution is assumed $\rightarrow$ statistical tests
  - If no distribution is assumed $\rightarrow$ graphical methods
- Problems in BE
  - Parametric methods (ANOVA, mixed-effect models) are very sensitive to outliers
    - Even a single outlier may underpower a properly sized study
    - Inflates the residual variance $\rightarrow$ wider confidence interval

Guidelines

- EMA, WHO, …

  Unbiased assessment of results from randomised studies requires that all subjects are observed and treated according to the same rules. These rules should be independent from treatment or outcome.
  In consequence, the decision to exclude a subject from the statistical analysis must be made before bioanalysis.

Guidelines

- EMA, WHO, …
  - Exclusion of outliers only possible if procedure stated in the protocol and reason can be justified, e.g.,
    - lacking compliance (subject did not take the medication);
    - vomiting
      - up to 2 × t_{max} for immediate release products and
      - within the intended dosing interval for modified release products;
    - diarrhea for drugs with an absorption window;
    - AEs which may influence gastric motility / liver bloodflow;
    - analytical problems (*e.g.*, interferences in chromatography, equipment malfunction)
  - Other decisions can be made only after bioanalyis
    - Pre-dose concentration >5% C_{max} (potential carry-over) (contrary to other samples reanalysis is aceptable)

Guidelines

- EMA, WHO, …
  - Lacking or very low plasma concentrations, *i.e.*, *AUC* <5% of geometric mean *AUC* of other subjects considered as an outlying subject (only acceptable for the reference product)
    - Great on paper but can be surprising
      - » Examples with 24 subjects; value <10 suspected outlier

        #1 4.9 and 23 cases of 100: $\text{mean}_{geo} = (100^{23})^{1/23} = 100$
        4.9 <5% → outlier

        #2 4.9, 4.9, and 22 cases of 100: $\text{mean}_{geo} = (100^{22})^{1/22} = 100$
        4.9 <5% → two outliers

        #3 4.9, 9.9, and 22 cases of 100: $\text{mean}_{geo} = (100^{22})^{1/22} = 100$
        4.9 <5%, 9.9 >5% → one outlier

        #4 4.9, 10, and 22 cases of 100: $\text{mean}_{geo} = (10 \times 100^{22})^{1/23} = 90.06$
        4.9 = 5.44% → no outlier
      - » In practice it can be *much* more complicated …

Guidelines

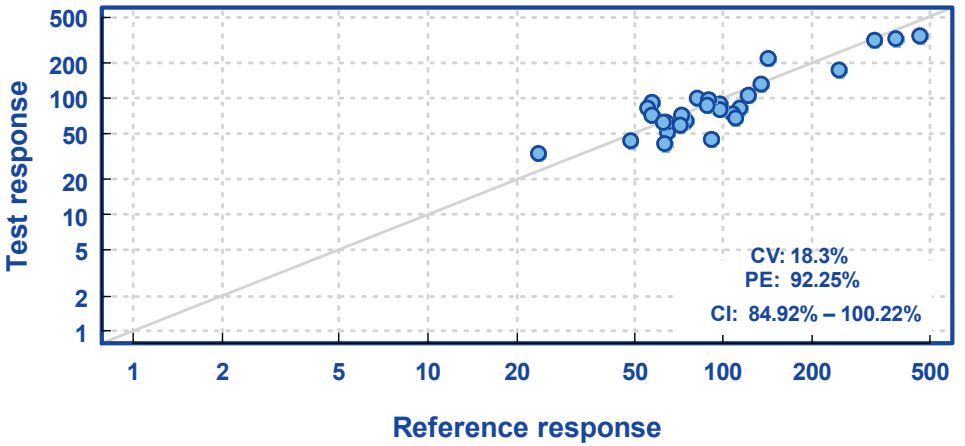
- Most agencies
  - Exclusion is not acceptable if only based on
    - pharmacokinetic reasons, *e.g.*, irregular profiles (good science?)
    - statistical grounds (debatable…)

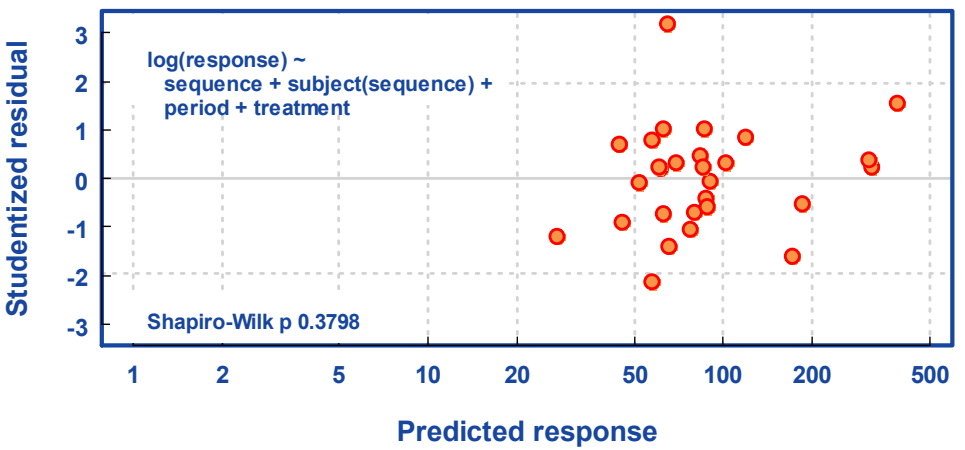
Types of Outliers

- Concordant
  - The PK response in a subject after both test and reference deviates from the majority of the study's subjects
    - Poor metabolizers may lead to high concentrations in 5 – 10% of subjects
    - *Does not* effect the outcome in a crossover study but should be discussed in the report (polymorphism known?)
- Discordant
  - The PK response in a subject after either test or reference deviates from the majority of the study's subjects
    - *May* have a substantial impact on the outcome
    - Is is scientifically sound to treat outliers of the test and reference differently?

Example 2×2×2 crossover

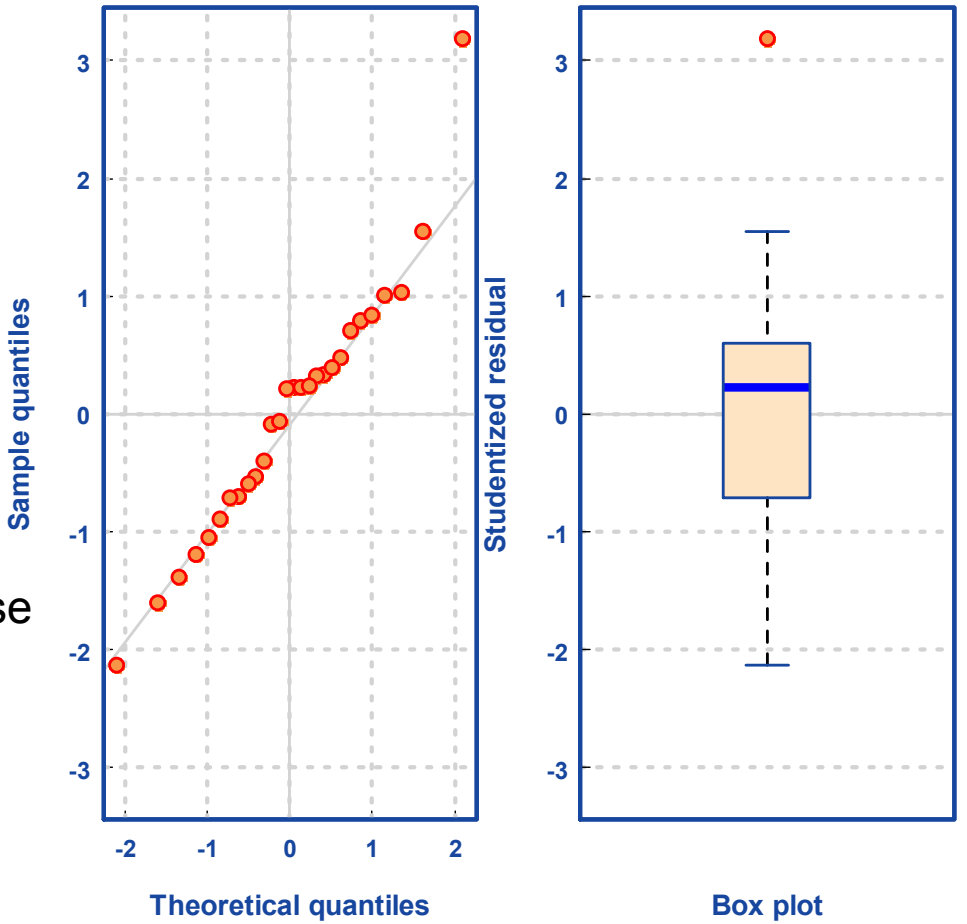
- Simulated data
  - CV_w 25% χ^2 distributed
  - CV_b 75% χ^2 distributed
  - θ_0 0.95 lognormal distributed
  - Power 80% (*n* 28)
  - Shapiro-Wilk test for normality not significant





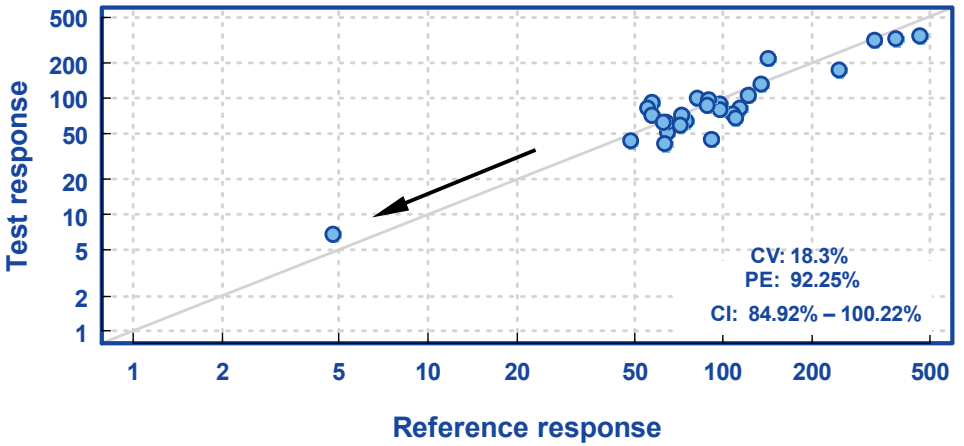
Example cont'd

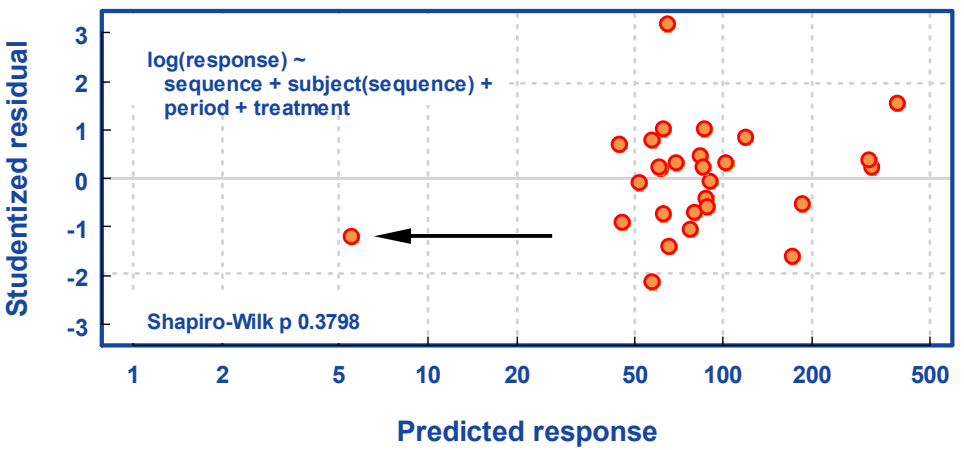
- Simulated data
  - Studentized residuals
    - Parametric
      QQ plot
    - Nonparametric
      Box plot
  - Health Canada
    - Outlier if studentized residual outside ±3
    - False positive because we simulated normal distributed data – pure chance!



Example cont'd

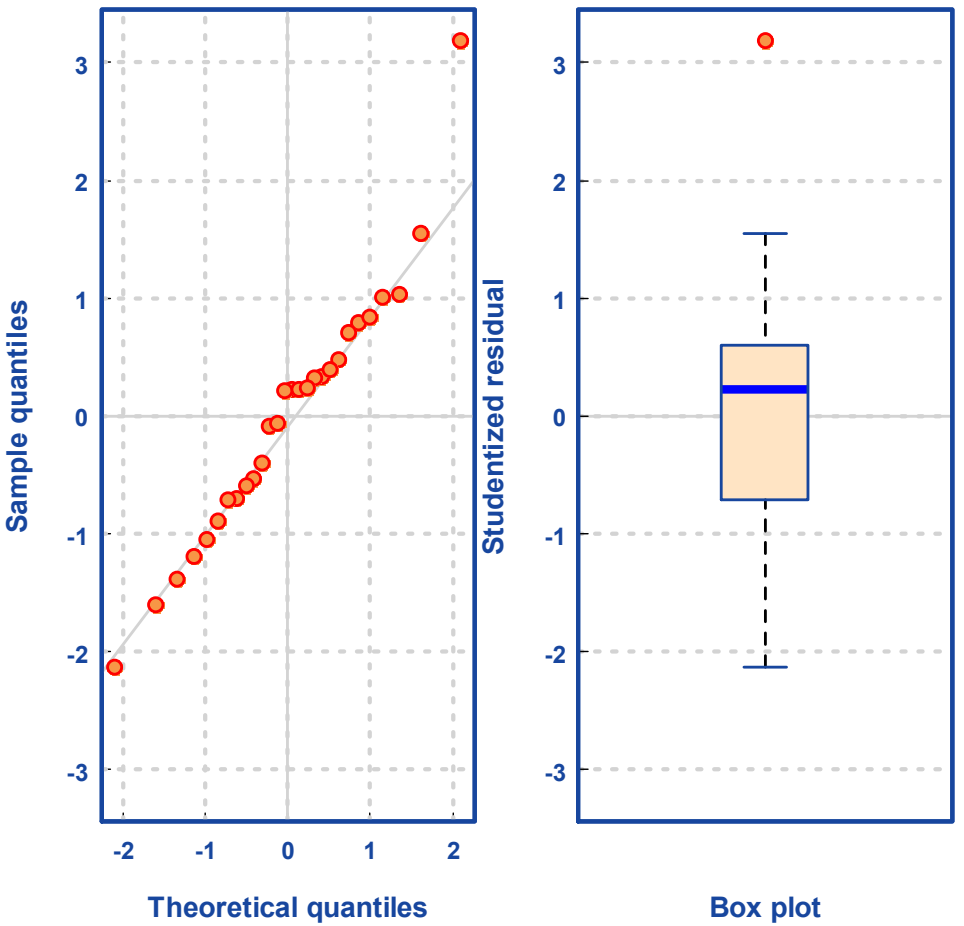
- Concordant outlier
  - Lowest response of T *and* R divided by 5
  - Identical model; same
    - CV_{wR}
    - *PE*
    - 90% CI
  - Residual of outlier
    - Shifted to the left
    - Same value
  - Shapiro-Wilk test for normality not significant (identical to original data)





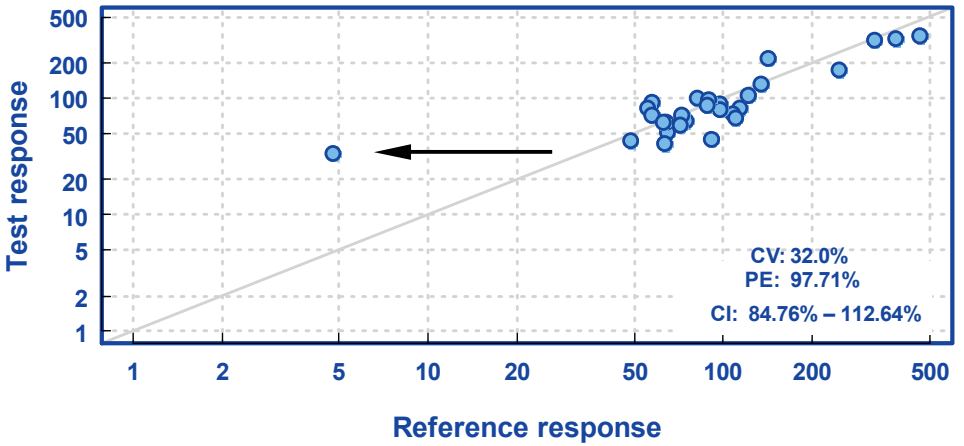
Example cont'd

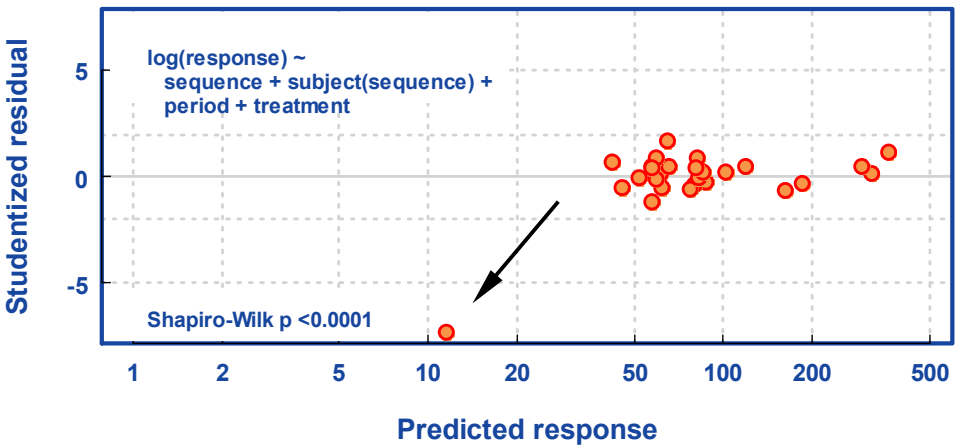
- Concordant outlier
  - Studentized residuals
    - Parametric
      QQ plot
    - Nonparametric
      Box plot
  - Identical to what we observed with the original data



Example cont'd

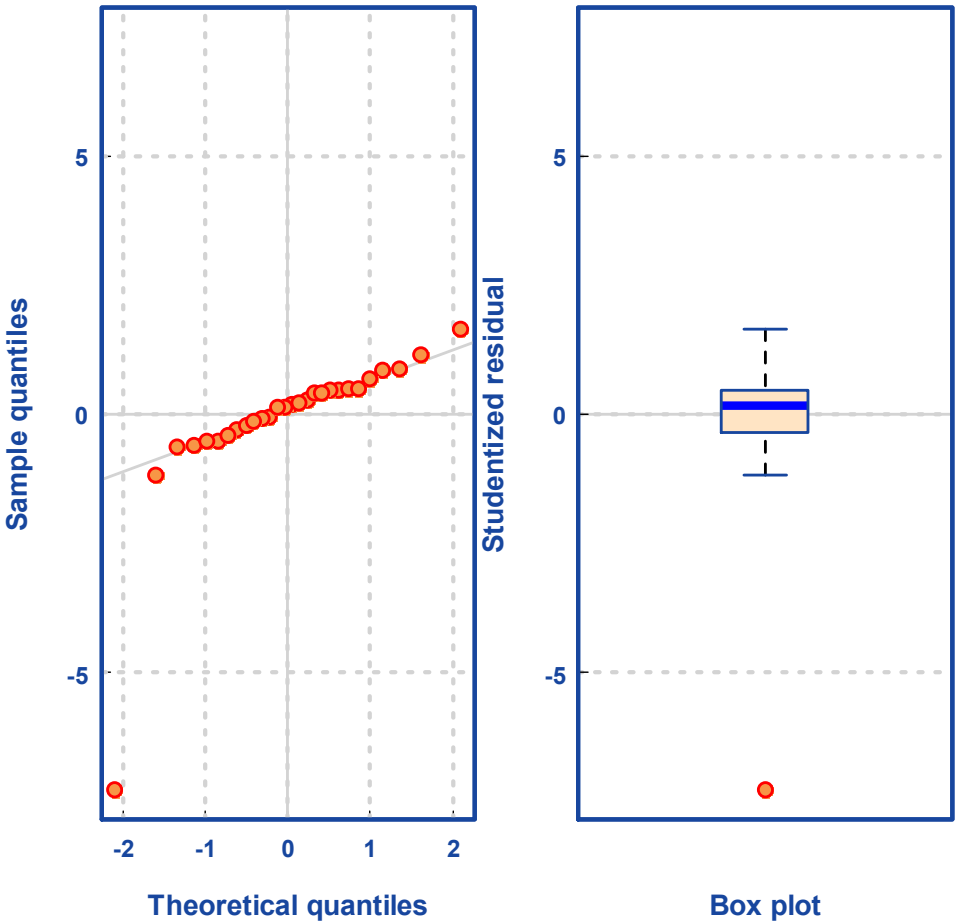
- Discordant outlier
  - Lowest response of *only* R divided by 5
  - Changed model
    - CV_{wR} ↑
    - *PE* ↑
    - 90% CI ↔
  - Residual of outlier
    - Shifted to the left
    - Much lower
  - Shapiro-Wilk test for normality highly significant





Example cont'd

- Discordant outlier
  - Studentized residuals
    - Parametric
      QQ plot
    - Nonparametric
      Box plot
  - Outlier detected
    by both approaches



Example 2×2×4 full replicate

- EMA's Q&A document, dataset 1 (TRTR | RTRT)
  - 77 subjects
  - Unbalanced
    - 39 in sequence TRTR
    - 38 in sequence RTRT
  - Incomplete: Missings / period: 0 | 1 | 7 | 2
    - 7 missings in sequence TRTR
    - 3 missings in sequence RTRT
  - 77 subjects with T and R
    Calculation of 90% CI
  - 73 subjects with 2 administrations of R
    Estimation of CV_{wR} – 'need to know' (calculate expanded limits)
  - 71 subjects with 2 administrations of T
    Estimation of CV_{wT} – 'nice to know' (required for the WHO)

Example 2×2×4 cont'd

- Results
  - CV_{wR} 46.96% → expanded limits 71.23 – 140.40%
  - CV_{wT} 35.16% → lower than CVwR but not significantly
  - 90% CI 107.11 – 124.89% (passes ABEL but ABE as well)
- ASEAN States, Australia, Canada, East African Community, Egypt, Eurasian Economic Union, European Economic Area, New Zealand, Russian Federation, WHO

  The applicant should justify that the calculated intra-subject variability is a reliable estimate and that it is not the result of outliers.
  - It is an open issue how outliers should be handled
- Not required by FDA, CDE, and ANVISA (№ 760.20 of Dec 27, 2019; public consultation until Apr 8, 2020)

Example 2×2×4 cont'd

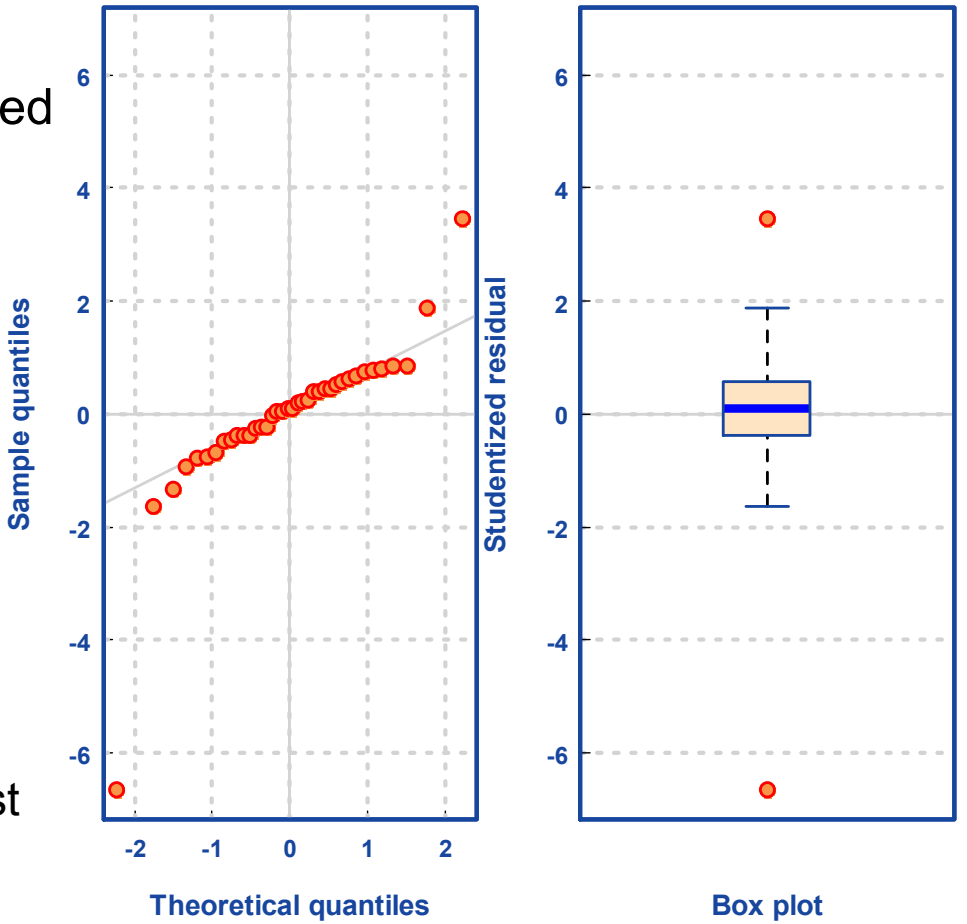
- Regrettably I suggested box plots as a mere *joke* at an EGA/EMA symposium, being aware of their nonparametric nature and the EMA's reluctance towards robust methods; alas, this joke was included in the Q&A document *

  […] a study could be acceptable if the bioequivalence requirements are met both including the outlier subject (using the scaled average bioequivalence approach and the within-subject CV with this subject) and after exclusion of the outlier (using the within-subject CV without this subject).
  An outlier test is not an expectation of the medicines agencies but outliers could be shown by a box plot. This would allow the medicines agencies to compare the data between them.

* European Generic Medicines Association. *Revised EMA Bioequivalence Guideline*. *Questions & Answers*. 3rd EGA Symposium on Bioequivalence. London, 1 June 2010.

Example 2×2×4 cont'd

- Outlier exploration
  - Two extreme studentized residuals in sequence RTRT
    - Subject 45: –6.657
    - Subject 52: +3.453
  - After excluding the subjects / recalc.
    - CV_{wR} decreases from 46.96% to 32.2%
    - New expanded limits 78.79 – 126.93%
    - Passes still but almost the entire expansion vanished



‘Outliers’ of HVD(P)s in Reference-scaling?

- Extreme values are an inherent property of highly variable drugs / drug products
  - HVD(P)s were shown to be safe and efficacious in Phase III trials and in clinical practice – despite extreme values
  - Exclusion not recommended by agencies applying RSABE (US FDA, China CDE) or ABEL (ANVISA № 760.20)
  - Other jurisdictions: Excluding outliers of the reference only
    - Violates a principle stated in guidelines

      Subjects should be observed and treated according to the same rules. These rules should be independent from treatment or outcome.
    - Scientifically questionable
      - Bioequivalence is a surrogate of therapeutic equivalence, *i.e.*, results can be extrapolated to the patient population
      - Excluding outliers of the reference essentially means that one has to demonstrate BE under a condition which does not occur in patients

History

- Cave: Test for normality ≠ outlier test
- Lund test [1] popular in the past
  - Not significant if outliers with opposite signs (low/high values)
  - Modifications exist but not trivial
  - Nowadays regulatory acceptance doubtful
- Robust (nonparametric) methods [2]
  - I evaluated ~300 studies solely by nonparametric methods
    - Provocative question

      Where are the dead people lying in the streets?
  - Quite often the outcome of robust methods is similar to the one of parametric methods after exclusion of outliers

1. Lund RE. *Tables for An Approximate Test for Outliers in Linear Models*. Technometrics 1975; 17(4): 473–6. doi:10.1080/00401706.1975.10489374.
2. Hauschke D, Steinijans VW, Diletti E. *A distribution-free procedure for the statistical analysis of bioequivalence studies*. Int J Clin Pharm Ther Toxicol. 1990; 28(2): 72–8. PMID 2307548.

Back to the first Example

- Parametric *vs.* robust method
  - Full data set (no outlier)
    - ANOVA
      PE 92.25% (90% CI 89.92 –100.22%) width 10.30%
    - Robust
      PE 91.44% (~90% CI 83.80 – 98.83%) width 15.03%
  - Full data set (one discordant outlier)
    - ANOVA
      PE 97.71% (90% CI 84.76 –112.64%) width 27.88%
    - Robust
      PE 91.80% (~90% CI 83.82 – 99.97%) width 16.15%
  - Reduced data set (after exlusion of outlier)
    - ANOVA
      PE 91.22% (90% CI 83.88 – 99.20%) width 15.32%

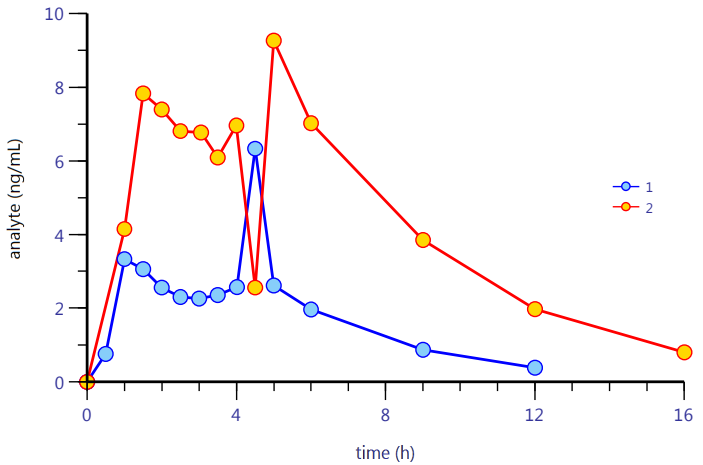
A reasonable Approach?

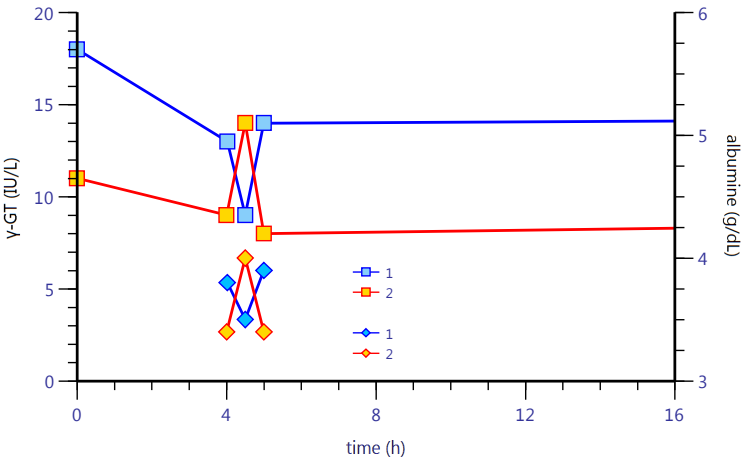
- Health Canada (Jun 2018)
  - Outlier identification performed *before* the study is assessed for BE (*i.e.*, before calculating the PE and CI); procedure stated in the Statistical Analysis Plan
  - Outliers if *e.g.*, the studentized residual outside ±3
    - Identified outliers may be removed from the analysis data set
    - Removal independent from treatment; subject should be indentified as outlier for all PK metrics (conservative!)
  - Rules
    - ≤5% of subjects may be removed, unless 20 or less subjects in the study, in which case only one subject may be removed
    - If >5% of subjects are identified as outliers, the absolute studentized residual is ranked and the highest 5% of subjects are removed
  - Re-testing of subjects identified as outliers is not recommended

Irregular Profiles

- Case Study 1
  - Biphasic release product, pilot study
  - Suspected mix-up in the transfer from sample vials after centrifugation to plasma sample vials

Measurable values in clinical chemistry (limited, since anticoagulant citrate)

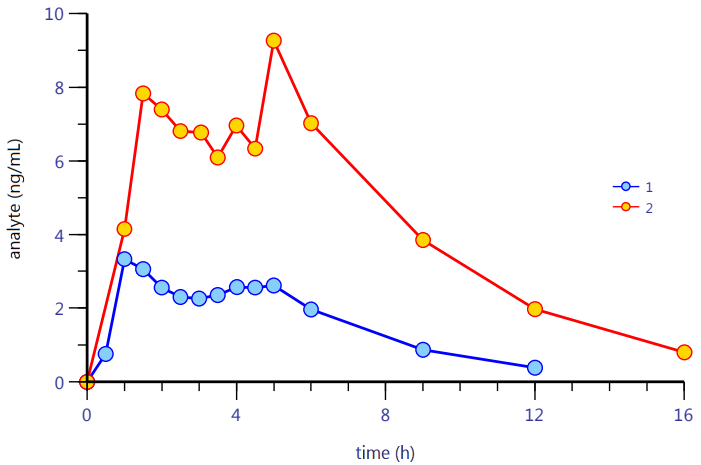


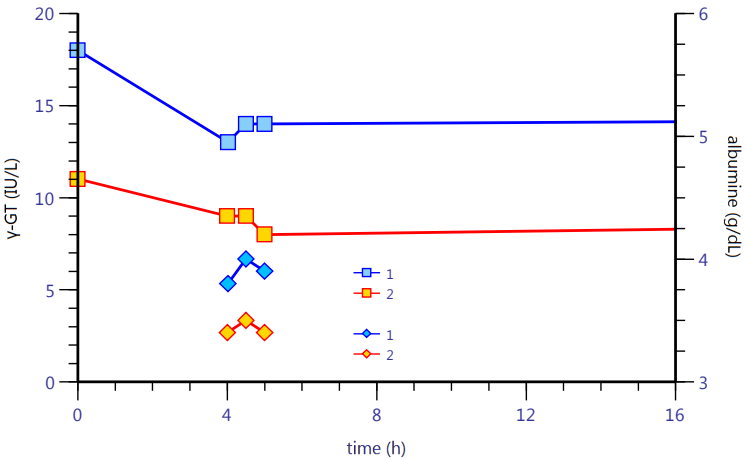


Irregular Profiles

- Case Study 1
  - Biphasic release product, pilot study
  - Exploratory: Values swapped (analyte and clinical chemistry)
  - Samples of subjects 1 & 2 both taken in the first period

Due to clinical chemistry values suspected mix-up very likely





Irregular Profiles

- Case Study 1
  - 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 system was still operable)
  - Blinded review of data for irregular profiles?
    - EMA BMV GL (2011)
      - Exclusion only possible if error documented
      - Measurements are ‘carved from stone’ (not even confirmatory reanalysis is acceptable)
      - Reanalysis of pre-dose sample if >LLOQ acceptable (why?)
    - FDA (Rev.1 Sep 2013)
      - Exclusion after repeated analysis acceptable if defined in SOP
    - FDA (May 2018), ICH M10 (Draft Feb 2019)
      - Like the EMA – generally not acceptable

Irregular Profiles

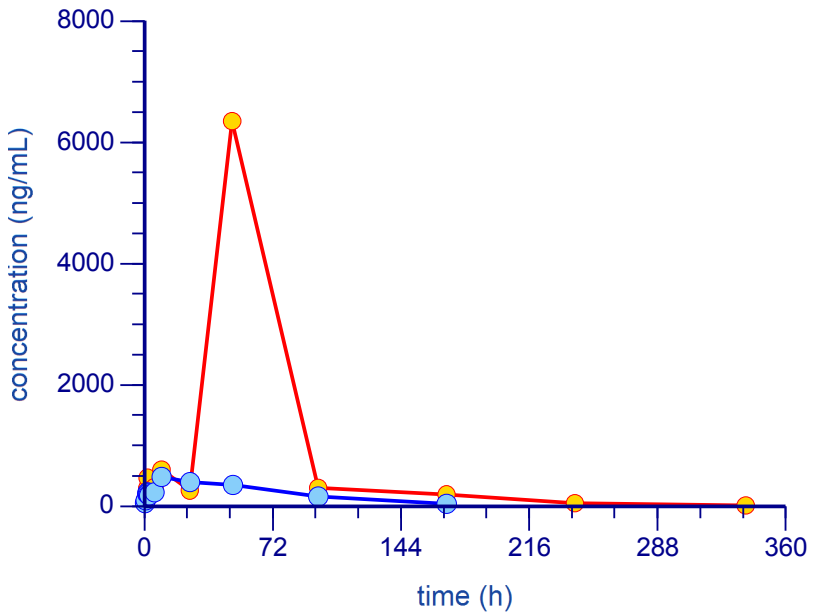
- Case Study 2
  - Liposome encapsulated drug for infusion
  - Analytes
    - Encapsulated drug
    - Unencapsulated drug (*i.e.*, released from liposomes)
    - Total drug (encapsulated + unencapsulated)
    - Metabolite (formed from unencapsulated drug only)
  - Drug may be released from liposomes by
    - shear forces (infusion pump, needle with narrow diameter);
    - high temperature and extended interval until centrifugation;
    - high *g* force in centrifugation
    - Only the latter two can be prevented
      - stabilization by DMSO
      - blood samples on ice, ≤45 minutes until controlled centrifugation

Irregular Profiles

- Case Study 2
  - Multinational study in terminal cancer patients
  - Clinical staff trained about critical sample handling but
    - unfamililar procedure esp. in small sites
    - necessity of following SOPs and documentation of deviations in conformity with GCP not well understood
    - well-being of patients considered by clinical staff of oncology departments of higher priority than “annoying paperwork”

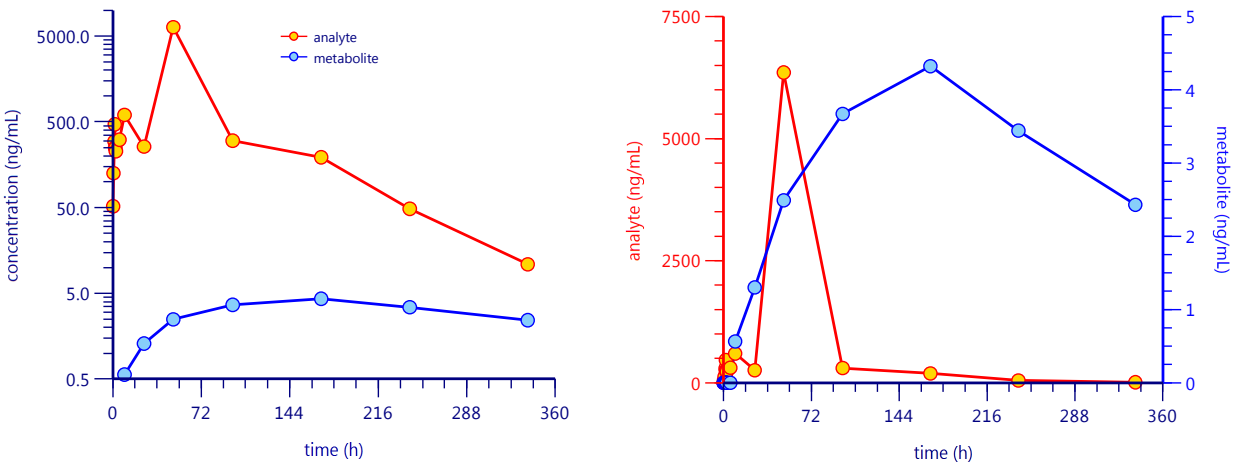
Irregular Profiles

- Case Study 2
  - Unexpected results in bioanalytics
  - Extremely high concentrations of unencapsulated drug observed in about 2% of samples
    - All suspect values confirmed in repeated analyses



Irregular Profiles

- Case Study 2
  - However, “normal” concentrations of the metabolite
    - Since the metabolite can only be formed from the unencapsulated drug, the drug’s high concentrations were considered an artifact
    - No documented improper sample handling (stabilization, temperature, time until centrifugation)
  - EMA (CHMP Referral, May 2019)
    Negative outcome; have to accept results as they are



Current Discussions

- Global Bioequivalence Harmonization Initiative (GBHI) 2nd International Workshop (Rockville, Sep 2016)
  - Session IV: Exclusion of PK Data in BE
    - EMA Not acceptable acc. to GL
      Reanalysis for PK reason not acceptable; only as part of laboratory investigations to prevent the recurrence of similar problems in the future
    - FDA Generally not acceptable but informative
      Workgroup within the FDA to explore potential approaches
    - ANVISA Acc. to RDC № 1170/2006 exclusion should be justified
      Evaluation with and without outlier(s) should be presented
      Max. 5% of subjects excluded?
      No rules for irregular profiles
    - Industry Irregular profiles are a common phenomenon
      Methods / rules should be developed

Current Discussions

- GBHI 3rd WS (Amsterdam, Apr 2018)
  - Session I: Exclusion of PK data in BE assessment
    - Regulatory perspective
      - More research about the possible reasons of outliers and the impact of excluding outliers in PK studies needed
      - Call for a harmonized guidance with a clear procedure and acceptance criteria on outlier exclusion
        - » Develop clear standards for *pre hoc* definition of outlier subjects
        - » Consider re-dosing in handling aberrant observations in a BE study
        - » Consider excluding subjects in some cases (*e.g.*, headache, migraine attack) for BE analysis if the decision criteria are pre-defined in the protocol
        - » Develop scientifically sound statistical approaches for applicants to handle outlier data

Current Discussions

- GBHI 3rd WS cont'd
  - Industry perspective / open issues for discussion
    - Use of re-dosing studies for confirming outliers and how to design and analyze these studies?
      Most re-dosing studies confirm that the original *AUC* and/or C_{max} T/R-ratios are aberrant, questioning the value of a re-dosing study
    - Justification for excluding statistical and/or PK outliers, particularly an 'improbable' PK outlier, when there are no protocol violations or irregularities reported during the clinical or analytical portion of the study and no positive finding from a root-cause investigation to support exclusion of the data point or the subject (*i.e.*, no documentation to support removal)?
    - Utility of evaluating an outlier in replicate designs (*e.g.*, to eliminate subject-by-formulation interaction as a cause for an aberrant T/R-ratio)?
      Regulatory agencies generally do not recommend to exclude outlier data for HVD(P)s because the outlier could be part of the inherent high variability of the product

Current Discussions

- GBHI 3rd WS cont'd
  - Industry perspective / open issues for discussion
    - How to deal with whole PK profiles that show very low or all BLOQ concentrations?
      Could consider harmonizing with guidelines' (EMA, WHO, …) recommendation to allow exclusion of a subject with lack of any measurable concentrations or only very low plasma concentrations for the reference medicinal product (RMP) if its AUC is <5% of RMP geometric mean *AUC* (calculated without inclusion of data from outlying subject).

Current Discussions

- GBHI 3rd WS cont'd
  - Session IV: Liposomal parenteral preparations
    - Irregular profiles not mentioned in guidelines (FDA draft Apr 2017, EMA reflection paper Feb 2013)
    - Blinded review of analytical data by an independent data monitoring committee suggested in the discussion (Charles DiLiberti, Alberto Gabizon, HS)
      - Define acceptance limits of the time-varying drug/metabolite-ratio based on published Population PK models
      - Keep unencapsuled drug as the primary endpoint but use the drug/metabolite-ratio to verify whether the measured concentration is reliable; exclude single concentrations or – if close to C_{max} – the subject

Current Discussions

- GBHI 4th WS (Bethesda, Dec 2019)
  - Session I: Liposomal Parenteral Preparations
    - FDA several product-specific guidances (no partial *AUC*s)
    - EMA reflection paper still in place; product-specific guidance on pegylated liposomal doxorubicin (Dec 2018)
      partial *AUC*s (*e.g.*, $AUC_{0–48}$ and $AUC_{48–tlast}$) required
    - Open issues
      - Proposal to assess BE on encapsulated drug only sufficient because *AUC* of unencapsulated doxorubicin <5% of total?
      - Different views re the reliability of un-encapsulated drug data due to consistent within-lab results but greatly divers between-lab results
      - Different views on the relevance of un-encapsulated drug data since liposomal doxorubicin formulations were developed to substantially reduce the free drug and its toxicity which is evident

Current Discussions

- ANVISA (№ 760.20 of Dec 27, 2019)
  - Article 70
    Exclusion of any research participant who has completed the clinical and bioanalytical stage in accordance with the prepared protocol is not allowed
    - §1° For excluding study participants it is necessary to configure the violation of criteria previously established in the study protocol.
    - §2° The use of a statistical test to identify outliers with aberrant value in order to exclude study participants data from the statistical analysis is not allowed.
  - Article 72
    If a research participant has an interference greater than 5% of his C_{max} at the pre-dose collection (time zero), the statistical calculation shall be presented without the participant in question.
  - Section IV C_{max} range extension
    - Outlier assessment/exclusion in the calculation of CV_{wR} is <u>not</u> mentioned. *Kudos!*

New Developments / Outlook

- Assumption of the normal distribution (after log-trans-formation) might be false
  - Outliers
  - Heavy-tailed distributions (including outliers)
  - Skewness of the distribution
- Replace the normal distribution by the *t* distribution
  - Accommodates heavy tails/outliers
  - Accommodates skewness
- Bayesian approaches

Outliers in Bioequivalence

Thank You!

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