MEDICAL UNIVERSITY OF VIENNA

BEBAC

Data Manipulation in Bioequivalence

Helmut Schütz
Center for Medical Data Science of the Medical University of Vienna
BEBAC Vienna

SAAMnow
Scientists Advancing Affordable Medicines

Background

Data manipulation detected in the past

- Ranbaxy (2004 – 2008), GVK Bio (2014), Semler (2016), Panexcell (2019), Synchron Research (2022), Synapse (2023)

Various ‘methods’ used by the CROs

- Only the reference administered
- Fake sequences, e.g. TT | RR
- Unblinded interim analysis and – if BE unlikely due to ‘bad’ T/R-ratio
  - Swap the code of T and R in subsequent subjects
  - If T/R-ratio in the interim is very ‘bad’, additionally dilute T- or R-samples
- Analyze backup samples of yet another study

Risk

Regulatory agencies use an arsenal of tools to detect fraud

- In the ideal situation a whistleblower gives details, supporting inspectors (Ranbaxy and GVK cases)
- Software
  - T/R-ratios of C_{max} vs analytical batch (spreadsheet or any statistical software)
  - FDA's 'DABERS' (Data Anomalies in BioEquivalence R Shiny)
  - 'Buster' and 'SaToWIB' routines* (R)
  - BEBAC's 'FraudDetection' (R)

* Fuglsang A. *Detection of data manipulation in bioequivalence trials*. Europ J Pharm Sci. 2021; 156: 105595. https://doi.org/10.1016/j.ejps.2020.105595.

Risk

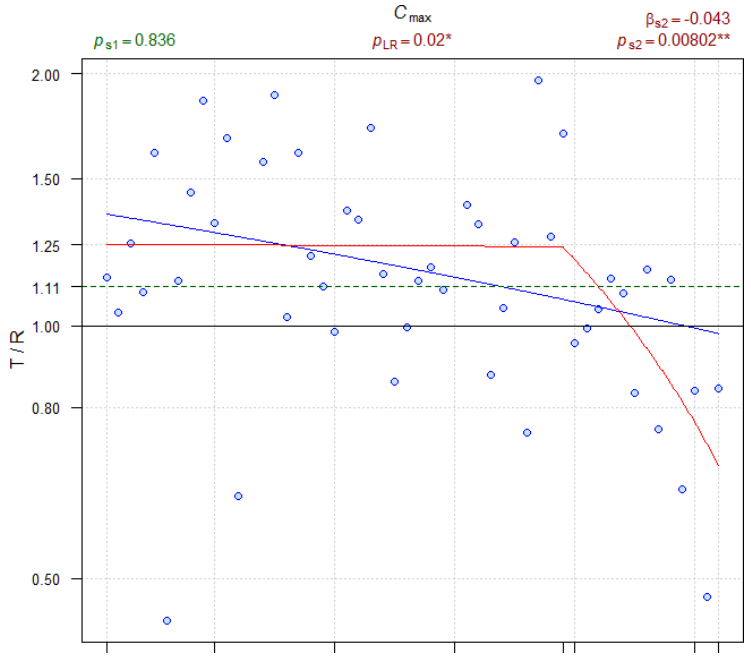
Applicants should not wait for a regulatory action

- Request full data of the CRO *before* submission
- Assess the data by various approaches to detect a signal of potential manipulation
- Consider a thorough audit

2×2×2, n=50, suspected interim analysis

Simple approach:
Assessing the T/R-ratios

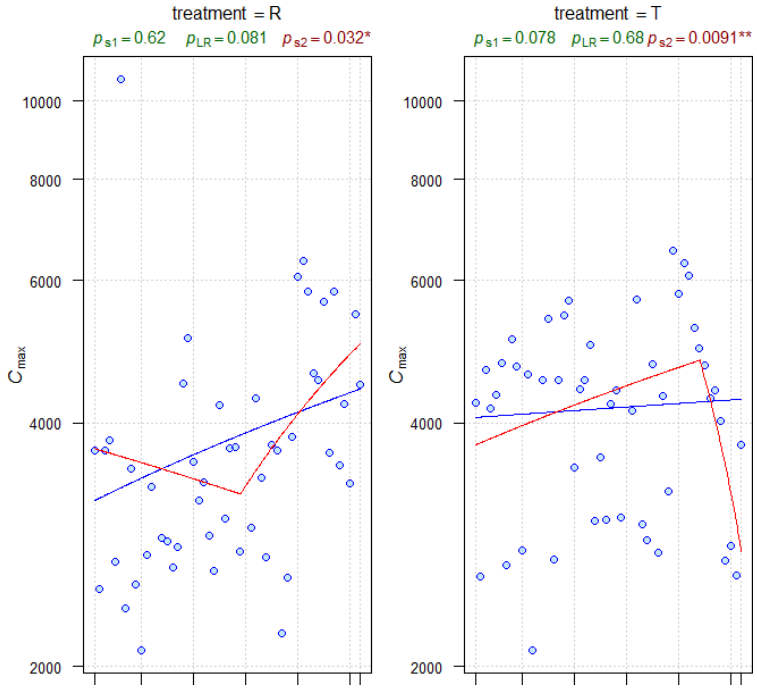
- Linear and segmented regressions
- Changing trends to ‘save’ the otherwise failing study
- ‘Bioequivalent’ in the final analysis



2×2×2, n=50, suspected interim analysis

C_{max}-values by treatment: Any differences, trend?

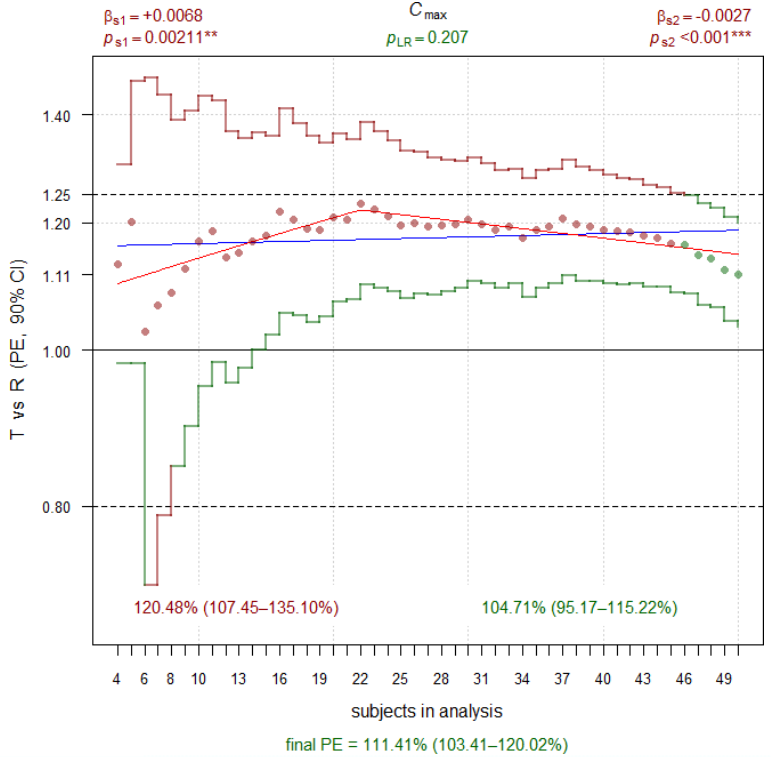
- A similar pattern like before
- Was T swapped with R in the later batches?
- Were the T-samples even diluted to ‘improve’ the T/R-ratio?



2×2×2, n=50, suspected interim analysis

BE assessed with an increasing number of subjects analyzed

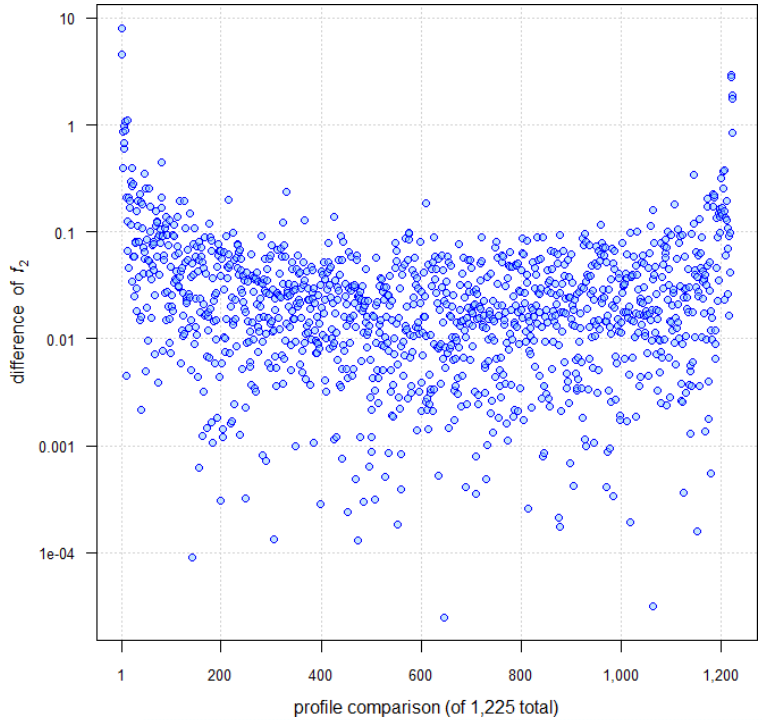
- Circles point estimates
- Stairs 90% confidence intervals (red if outside BE margin, green if passing BE)
- Both segments are significant
- Did the manipulation start already earlier than we assumed?



2×2×2, n=50, suspected interim analysis

Comparison of similarity of plasma profiles by f_2

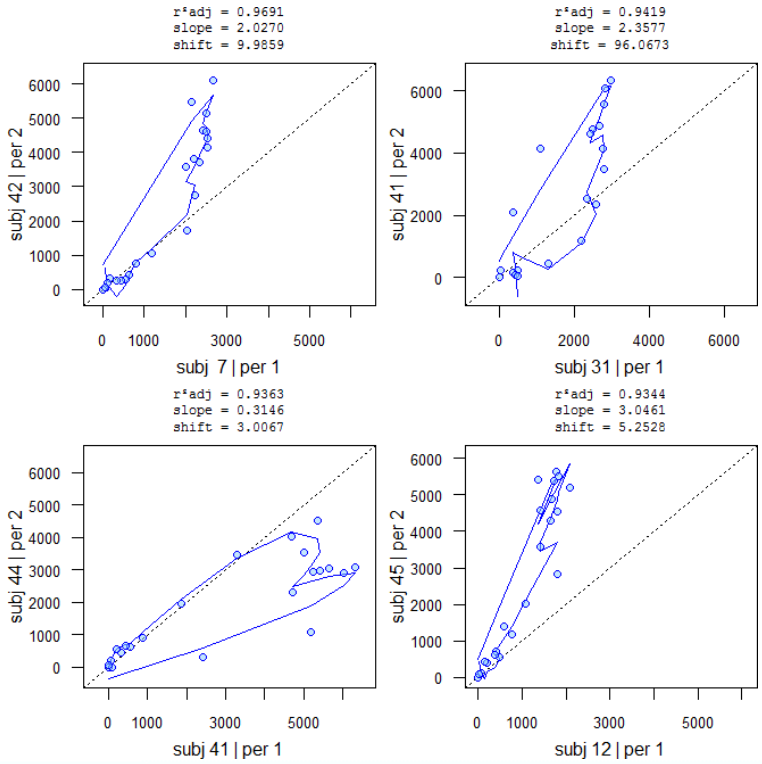
- Each profile with any other (irrespective of the treatment)
- Profiles with very small differences in their f_2-values are suspect



2×2×2, n=50, suspected interim analysis

Correlation of plasma profiles by the measured concentration

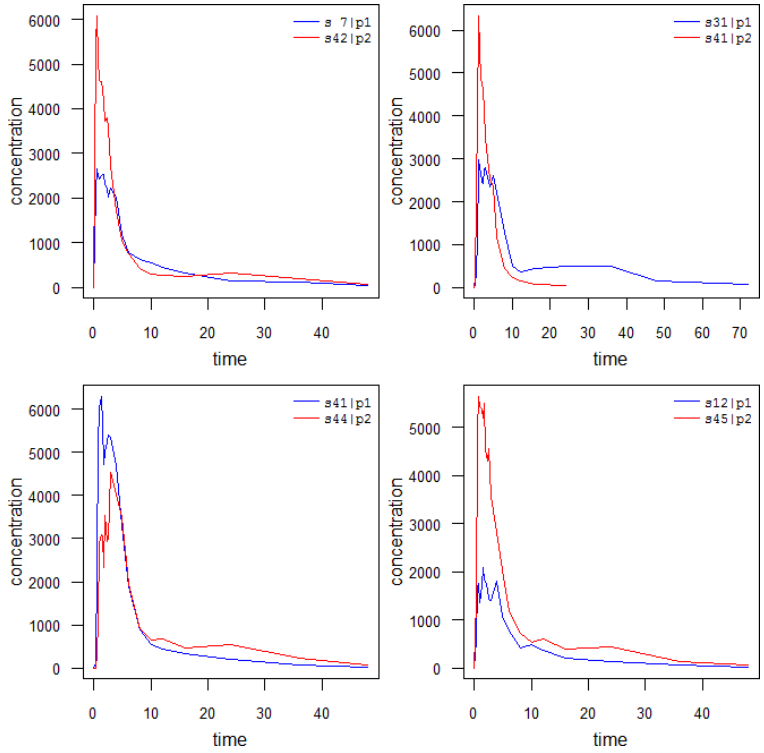
- Each profile with any other (irrespective of the treatment)
- We have to take the time into account – otherwise similar profiles with different lag-times will be falsely appear highly correlated
- Highly correlated concentrations are suspect



2×2×2, n=50, suspected interim analysis

Comparison of plasma profiles by the measured concentration

- We suggest to compare suspect highly correlated concentrations visually



‘FraudDetection’

Required data

- Sampling schedule
- Concentrations (any number of analytes)
- Analytical batches and / or dates of analysis
- Randomization (currently 2×2×2 and higher-order crossover designs)

Optional

- Actual sampling time points
- Method used by the CRO to calculate *AUC*
- PK metrics reported by the CRO

‘FraudDetection’

Supported data formats

- CSV, XLS(x), ODS, SAS XPT, Phoenix Project file
- CDISC (via Phoenix 8.4.0)

Recalculation by NCA

- C_{max} / t_{max}, C_{last} / t_{last}
- AUC_{0-t} (linear trapezoial or linear-up / logarithmic down)
- Optional
  - λ_z (start- and end-time, number of data points)
  - $AUC_{0-\infty}$ (observed or predicted)
  - Extrapolated fraction

‘FraudDetection’

Methods

- Spaghetti (grouped by treatment) and treatment (grouped by subject) plots
- PK metric by treatment vs batch, date of analysis, etc.
- T/R-ratios vs batch, date of analysis, etc.
- $\log_e(PK)$ – mean[$\log_e(PK)$]; runs test
- $\log_e(PK_T/PK_R)$ – mean[$\log_e(PK_T/PK_R)$]; runs test
- BE by subjects analyzed (≥ 4)
  - Plot (PE, 90% CI)
  - Table (MSE, PE, 90% CI, pass|fail)

‘FraudDetection’

Methods cont’d

- MSE of model by subjects analyzed
- Model residuals by subjects analyzed
- Difference factor f_1 by subject
- Similary factor f_2 by subject
- Comparison of f_2 of profiles with any other
  - Plot of differences
  - Table of most and least similar profiles
- If data provided by the CRO, comparison of NCA

Problems

Caveat

- Multiple analytes with the same method
  - Might give contradictory outcomes
  - Judgement required

Unresolved

- No statistical method in the strict sense
  (ideal: null hypothesis = no manipulation, alternative = manipulation)
  - Exploratory
  - Assessment is subjective – open to interpretation

Problems

Unresolved cont'd

- Breakpoint of segmented regression
  - Not unique in the different methods
  - Most reliable possibly the BE plot
    - However, CIs of the segments likely overlap
- Comparison of f_2
  - How similar is similar?
- Correlation of plasma profiles
  - Thresholds of r^2 and slope for detecting dilutions?
- Runs test has low power

Problems

Unresolved cont'd

- If study is performed in groups*
  - Different PK might be detected by pure chance
  - Should *not* be interpreted as a signal of manipulation
- The arms race
  - The CRO manipulated in a way that 'fools' the software
  - The CRO manipulated in an unexpected manner that the software fails to detect

* Schütz H, Burger DA, Cobo E, Dubins DD, Farkás T, Labes D, Lang B, Ocaña J, Ring A, Shitova A, Stus V, Tomashevskiy M. *Group-by-Treatment Interaction Effects in Comparative Bioavailability Studies*. AAPS J. 2024; 26(3): 50. https://doi.org/10.1208/s12248-024-00921-x.

Problems

FDA about 'DABERS'

- **Despite its demonstrated effectiveness,** a major drawback is that the pharmacokinetics and pharmacodynamics may be too complicated to describe with a single statistic. **Indeed,** the current practice offers no practical guidelines regarding how similar PK profiles from different subjects can be in order to be considered valid. **This makes it difficult to assess the adequacy of data to be accepted for an ANDA and requires additional information requests to applicants. This project will address the current gap in identifying the data anomalies and potential data manipulations by use of state-of-the-art statistical methods, specifically focusing on machine learning and data augmentation. […] from a regulatory perspective, our project will provide a data driven method that can model complex patterns of PK data to identify potential data manipulations under an ANDA.**

Data Manipulation in Bioequivalence

Thank You!

Helmut Schütz

Center for Medical Data Science

MEDICAL UNIVERSITY OF VIENNA

1090 Vienna, Austria

helmut.schuetz@muv.ac.at

BEBAC

1070 Vienna, Austria

helmut.schuetz@bebac.at

Acknowledgments: Osama Abd Elrahman, Susana Almeida, Anders Fuglsang, Alfredo García-Arieta, Olivier LeBlaye