Statistical Considerations in Assessing BE of IVPT Data

Elena Rantou, PhD
Lead Mathematical Statistician
DBVIII, Office of Biostatistics / OTS
Center for Drug Evaluation and Research, FDA

Disclaimer

1. This presentation reflects the views of the presenter and should not be construed to represent the United States Food and Drug Administration's views or policies.
2. Data sets shown in this presentation have been de-identified

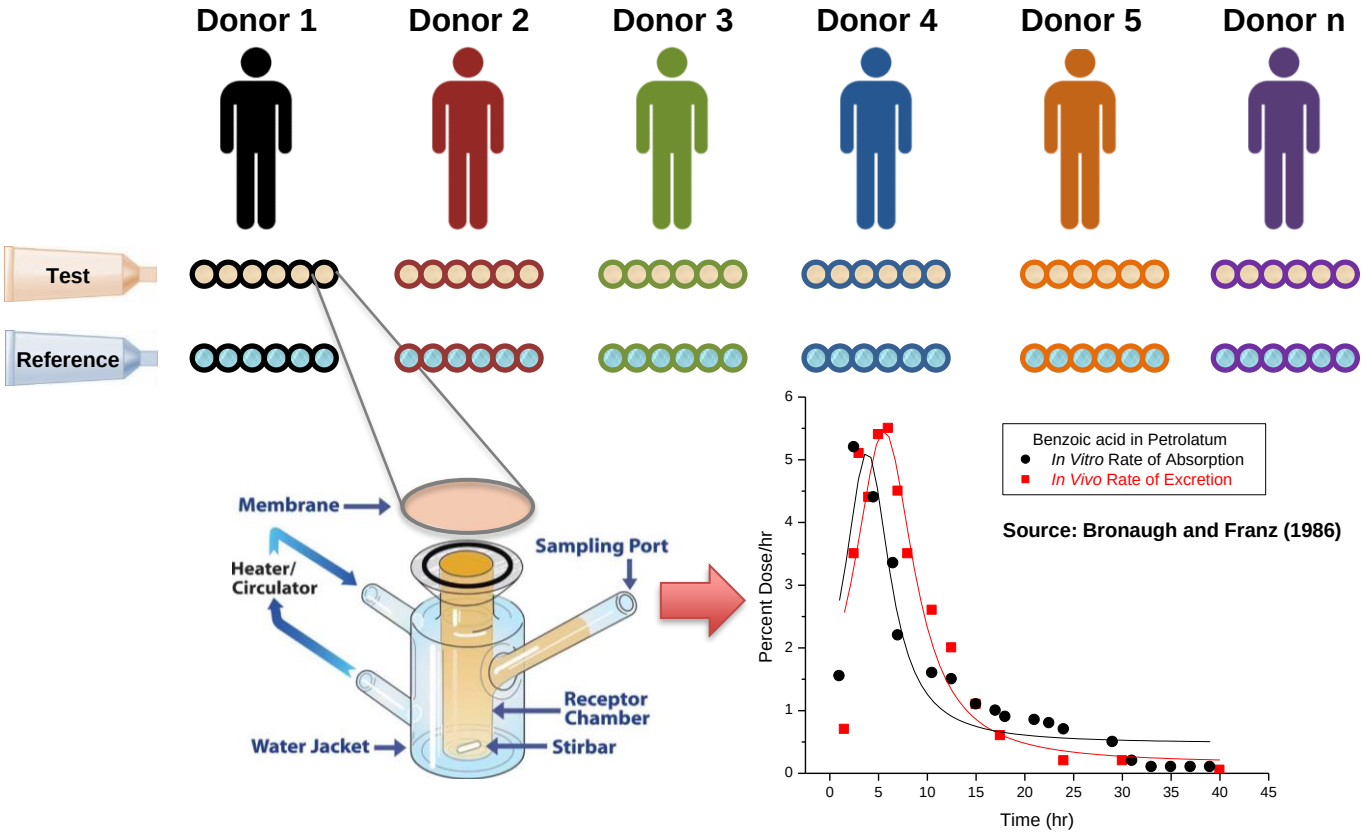
Outline

1. Overview of the IVPT data and study design
2. Assessing bioequivalence (BE)
3. Issues with IVPT

In Vitro Permeation Test (IVPT)

- Uses excised human skin
- Measures drug concentration
- The rate of drug delivery (flux) is measured by sampling at specific, pre-selected time-points in a way analogous to that used in blood (or plasma) concentration sampling in PK studies

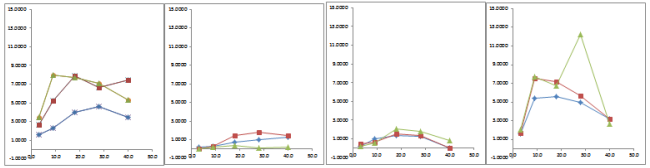
IVPT Study Design (Balanced Data)



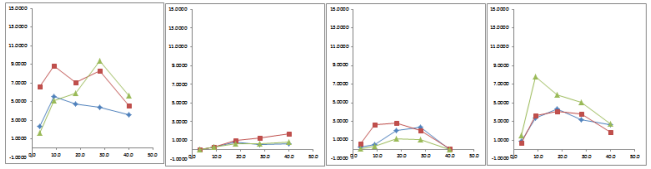
Developing In Vitro BE Standards

- **IVPT Statistical Analysis of Bioequivalence**

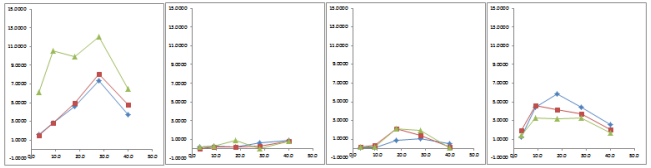
RLD Product
Lot# "**1**" (redacted)



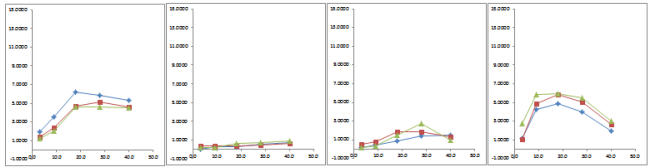
RLD Product
Lot# "**2**" (redacted)



RLD Product
Lot# "**3**" (redacted)



Test Product



Developing In Vitro BE Standards

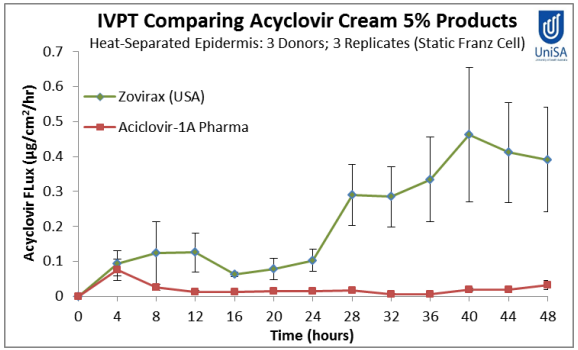
- *Standard procedures for IVPT study statistical analysis of BE had not been established*
- **IVPT Statistical Analysis of Bioequivalence**
  - The approach for Scaled Average Bio-Equivalence (SABE) analysis of highly variable drugs was modified for the IVPT study design
  - The mixed criterion uses the within-reference variability (σ_{WR}) as a cutoff point for bioequivalence analysis
  - When $\sigma_{WR} \leq 0.294$, Average Bio-Equivalence (ABE) is used
  - When $\sigma_{WR} > 0.294$, Scaled ABE (SABE) is used

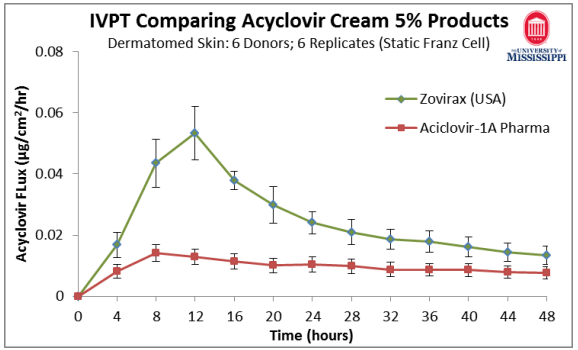
Performance/Results

- The results obtained with IVPT and the suggested statistical analysis, agreed with the original results that led to regulatory approval of these products. This supports the ***validity of this model for assessing BE***
- The test has been used for comparing two batches of the same reference product and successfully captured the similarity of these products in terms of BE. The outcomes support the ***model's sensitivity to meaningful differences and its resistance to the hazard of rejecting good products***

IVPT Statistical Analysis

- **Negative Controls** for BE: Aciclovir-1A® vs. Zovirax® US





Aciclovir-1A® (T) vs. **Zovirax® US (R)**

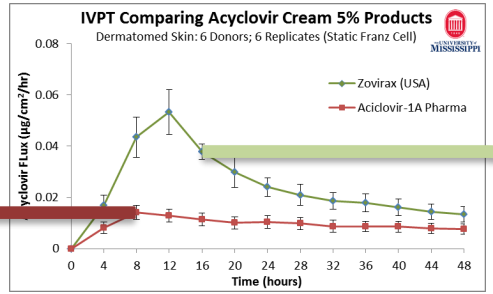
IVPT PK Endpoint	Maximum Flux **(Jmax)**	Total Bioavailability **(AUC)**
Point Estimate	0.172	0.104
$S_{Within\ Reference}$	0.521	0.551
SABE [0.80, 1.25]	4.433 (Non-BE)	7.236 (Non-BE)
N for [0.80, 1.25] with 3 Replicates	6	8

Aciclovir-1A® (T) vs. **Zovirax® US (R)**

IVPT PK Endpoint	Maximum Flux **(Jmax)**	Total Bioavailability **(AUC)**
Point Estimate	0.290	0.366
$S_{Within\ Reference}$	0.575	0.419
SABE [0.80, 1.25]	2.383 (Non-BE)	1.884 (Non-BE)
N for [0.80, 1.25] with 6 Replicates	8	20

IVPT Statistical Analysis

- **Positive Controls** for BE: Aciclovir-1A® and Zovirax® US



Comparison to Self by dividing up 6 replicates

Comparison to Self by dividing up 6 replicates

Aciclovir-1A® (T) vs. **Aciclovir-1A® (R)**

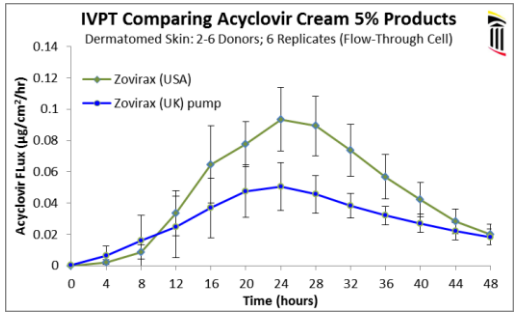
IVPT PK Endpoint	Maximum Flux **(Jmax)**	Total Bioavailability **(AUC)**
Point Estimate	0.983	0.958
$S_{Within\ Reference}$	0.303	0.318
SABE [0.80, 1.25]	-0.026 (**BE**)	-0.041 (**BE**)
N for [0.80, 1.25] with 4 Replicates	26+	15
N for [0.80, 1.25] with 3 Replicates	26+	15

Zovirax® US (T) vs. **Zovirax® US (R)**

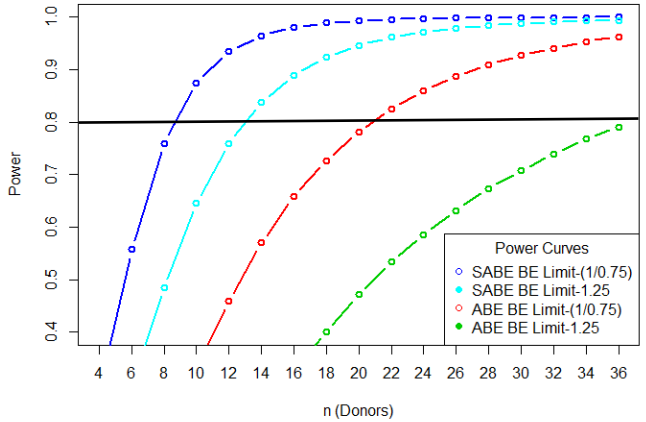
IVPT PK Endpoint	Maximum Flux **(Jmax)**	Total Bioavailability **(AUC)**
Point Estimate	0.962	1.101
$S_{Within\ Reference}$	0.697	0.469
SABE [0.80, 1.25]	-0.214 (**BE**)	-0.020 (**BE**)
N for [0.80, 1.25] with 4 Replicates	12+	14
N for [0.80, 1.25] with 3 Replicates	14	15+

IVPT Bioequivalence Limits

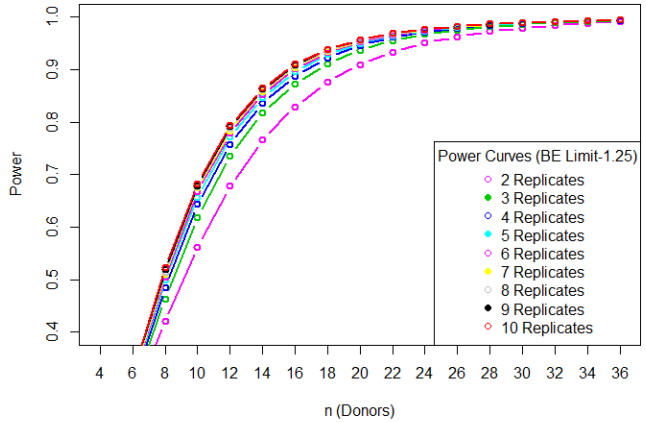
- **Bioequivalence Limits, Study Power and Study Size**



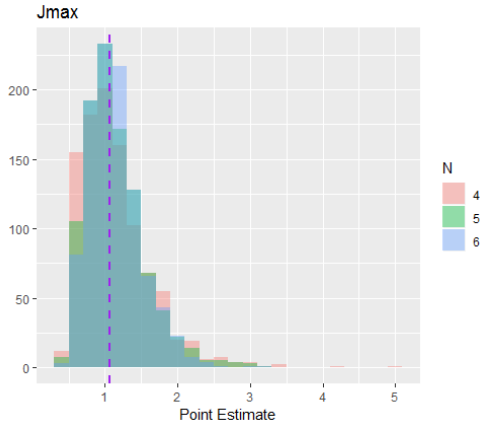
UK-US Jmax

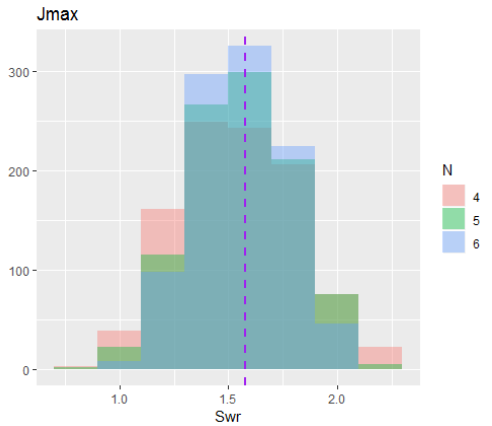


UK-US Jmax



Selecting the Number of Donors for a Pilot Study





- Prior work on pilot study sample size selection indicates a constant improvement in precision, when the sample size increases
- Additionally, the choice of the sample size depends on the characteristics and variability of each data set

Unbalanced Data

Replicate skin sections are withdrawn when

- Samples from the diffusion cell are destroyed (anticipated experimental event)
- There is documented evidence that skin is damaged during the course of the experiment

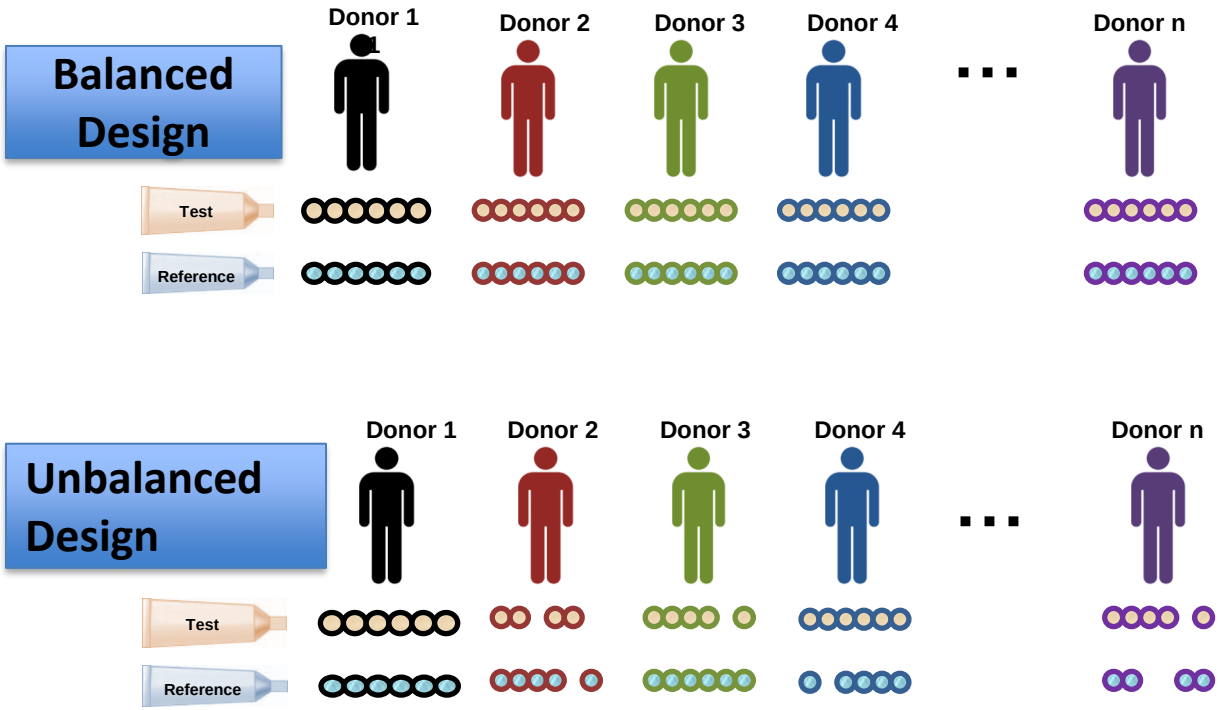
In such cases, replicate values can be replaced so that there is no informational loss

Unbalanced Data

In situations where we are unable to replace the diffusion cell, replicate values are dropped but not uniformly for all donors.

This needed to be addressed by adjusting the statistical test to account for the unbalanced data.

Balanced and Unbalanced Design



Study Design (balanced case)

We consider a sample of

n: donors (per treatment),

r: replicate skin sections from each one of the n donors are collected for each formulation (replicates from each donor are randomly assigned to each product)

2 treatment formulations: test (generic: T) and reference (R)

A confidence interval (CI) should be calculated for each pharmacokinetic (PK) endpoint:

a. the natural log-transformed maximum flux (J_{max})
b. the natural log-transformed total (cumulative) penetration (AMT)

Study Design (unbalanced case)

T_{ij} = the natural log-transformed (J_{max} or AMT) dosed with the test product for the i^{th} replicate from the j^{th} donor $(i = 1, 2, \cdots, r_j^T,\ j = 1, 2, \cdots, n)$

R_{ij} = the natural log-transformed (J_{max} or AMT) dosed with the RLD product for the i^{th} replicate from the j^{th} donor $(i = 1, 2, \cdots, r_j^R,\ j = 1, 2, \cdots, n)$

r_j^T = the number of skin replicates from the j^{th} donor dosed with the test product ($j = 1, 2, \cdots, n$)

r_j^R = the number of skin replicates from the j^{th} donor dosed with the RLD product ($j = 1, 2, \cdots, n$)

$r^* = r_1^R + r_2^R + \cdots + r_n^R$ = the total number of skin replicates in the RLD group

n = the number of donors

Statistical Analysis - S_{WR}

$$S_{WR}^2 = \frac{\sum_{j=1}^{n} \sum_{i=1}^{r_j^R} \left(R_{ij} - \bar{R}_{.j}\right)^2}{r^* - n}$$

where $\bar{R}_{.j} = \frac{1}{r_j^R} \sum_{i=1}^{r_j^R} R_{ij}$ is the average of log-transformed observations across all r_j^R replicates from donor *j* dosed with the RLD product.

Statistical Analysis – Point Estimate ($\hat{I}$) and its Standard Error

Estimate the point estimate for the treatment mean difference, its standard error and the corresponding degrees of freedom by using a linear model with the donor and treatment factors. For example:

```
proc mixed data = IVPT.data;
class DONOR TRT;
model log(AMT) = DONOR TRT;
estimate 'log(AMT) Test-Ref' TRT -1 1/cl alpha=0.1;
```

The output of this model estimates:

$\hat{\boldsymbol{I}}$: the point estimate of the mean difference $\mu_T - \mu_R$

$\boldsymbol{se}(\hat{\boldsymbol{I}})$: the standard error of the estimate

$\boldsymbol{df}^*$: the corresponding degrees of freedom

Statistical Analysis - Regular Average BE (ABE)

Determine the (1-2α)*100% confidence interval for $\mu_T - \mu_R$:

$$\hat{I} \pm t_{(1-\alpha),df^*} * se(\hat{I})$$

where:

$\mu_T - \mu_R$ = mean difference of T and R products

$t_{(1-\alpha),df^*}$ = $(1-\alpha) * 100^{\text{th}}$ percentile of the Student's t-distribution with df^* degrees of freedom

α = 0.05

Statistical Analysis - Scaled Average BE (SABE)

The hypotheses to be tested are:

$$H_0: \frac{(\mu_T - \mu_R)^2}{\sigma_{WR}^2} > \theta$$

$$H_a: \frac{(\mu_T - \mu_R)^2}{\sigma_{WR}^2} \leq \theta$$

Where $\theta = \frac{(\ln(m))^2}{(0.25)^2}$

Assessing Bioequivalence

After constructing the CI for the quantity $(\mu_T - \mu_R)^2 - \theta\ \sigma_{WR}^2$ we observe its $(1-\alpha) * 100\%$ upper bound. If this is less than or equal to zero, H_0 will be rejected. Rejection of the null hypothesis, H_0, supports BE.

This criterion is accompanied by a ***point estimate constraint*** according to which the geometric mean ratio (point estimate of the log-transformed response falls within the pre-specified limits: $[\frac{1}{m}, m]$

Outliers

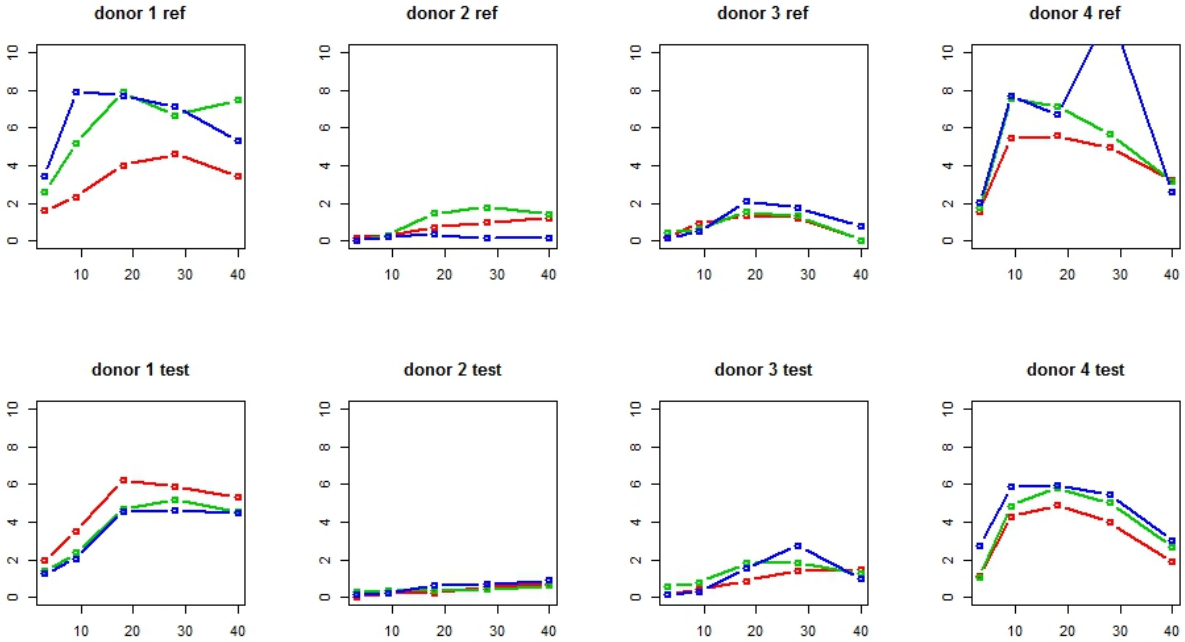
- The nature of an outlying observation in this setup:
  Within-donor, extreme replicate values

- Is it meaningful to consider ‘outlying donors’?

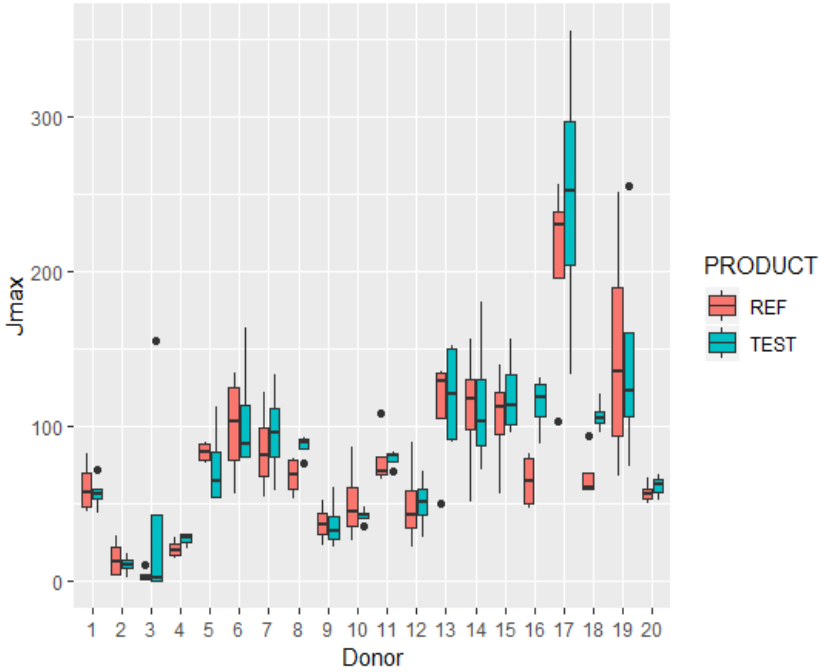
Outlier Detection

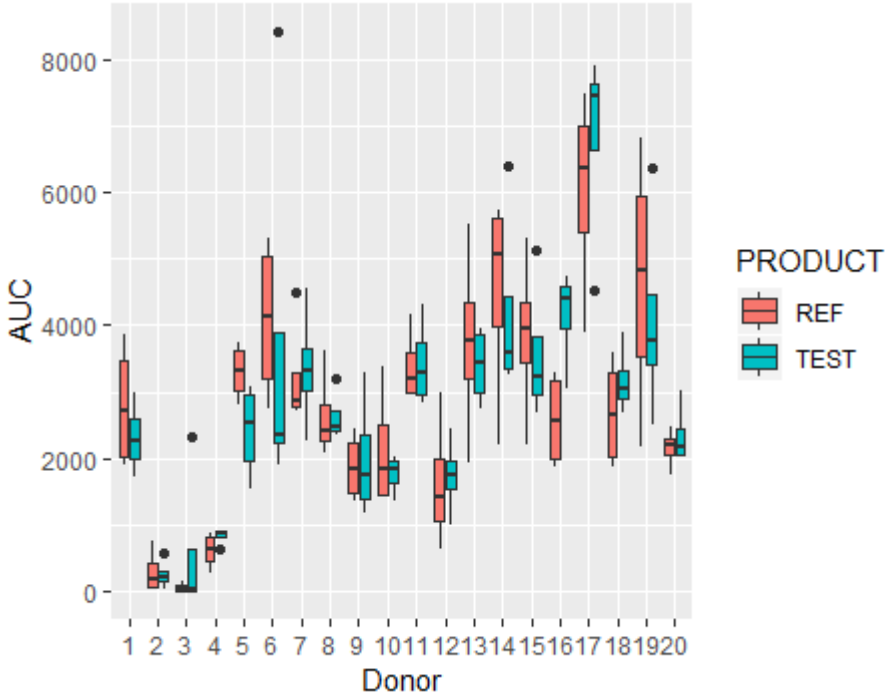
- Standard practices used in PK-studies (standardized residuals) do not apply here because of the small sample size of replicate values within one donor
- How do the results from Dean-Dixon test compared to other tests (such as, Grubbs)?
- Such tests are appropriate for small n in cases of experimental conduct anomalies that are detected once the sample analysis is completed.

Outliers

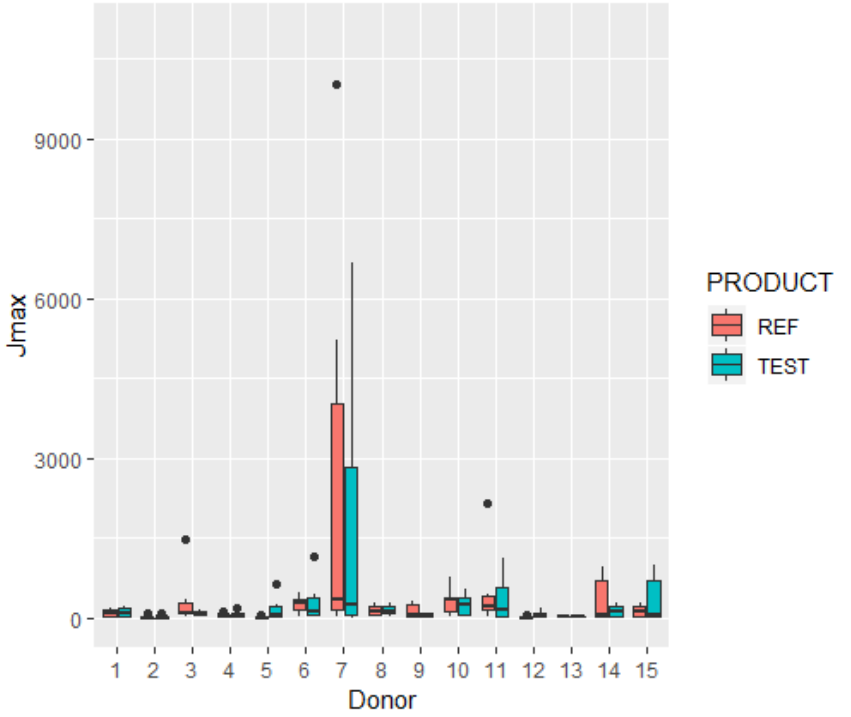

donor 1 ref
donor 2 ref
donor 3 ref
donor 4 ref
donor 1 test
donor 2 test
donor 3 test
donor 4 test
0
2
4
6
8
10
10
20
30
40

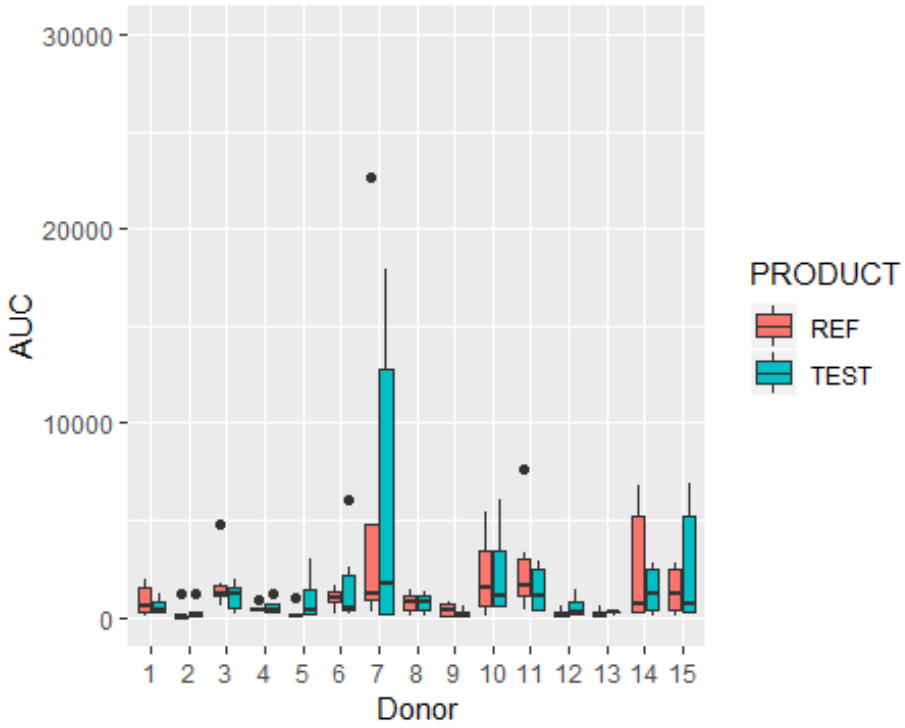
Example (20 donors)



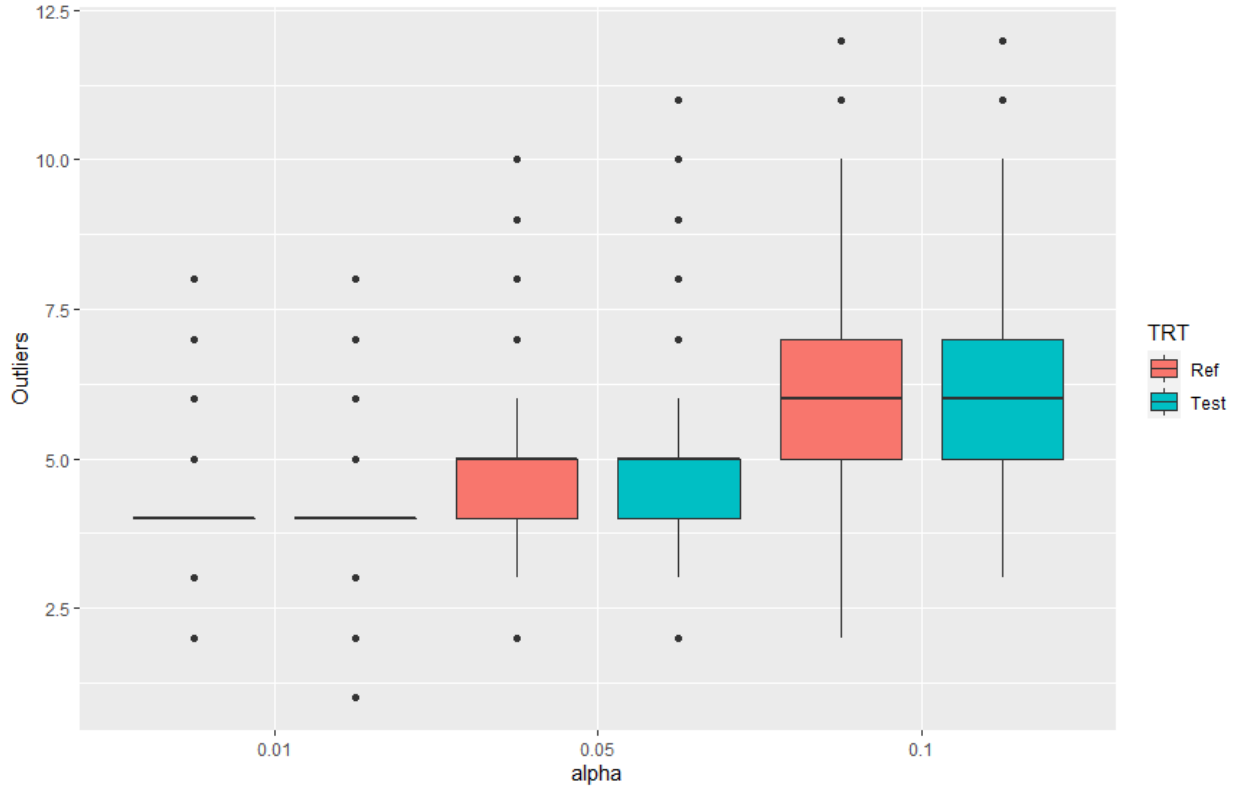


Example (15 donors)

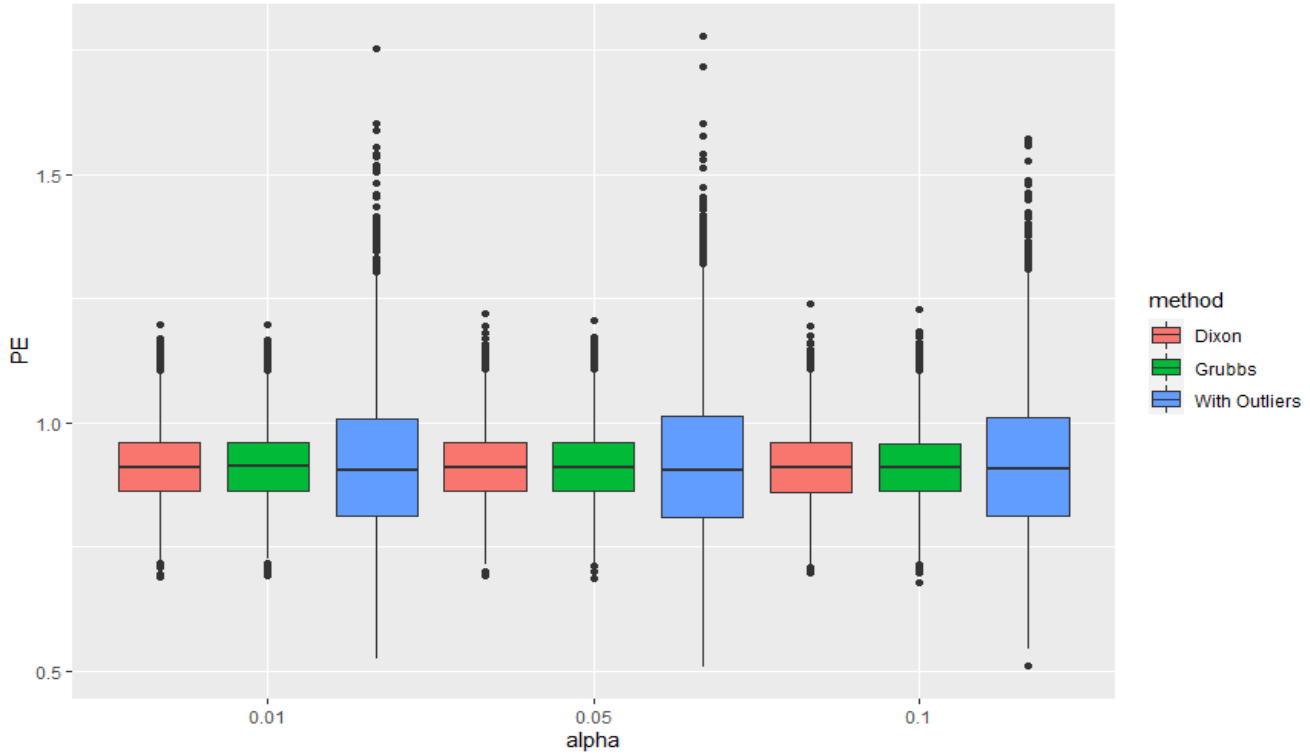




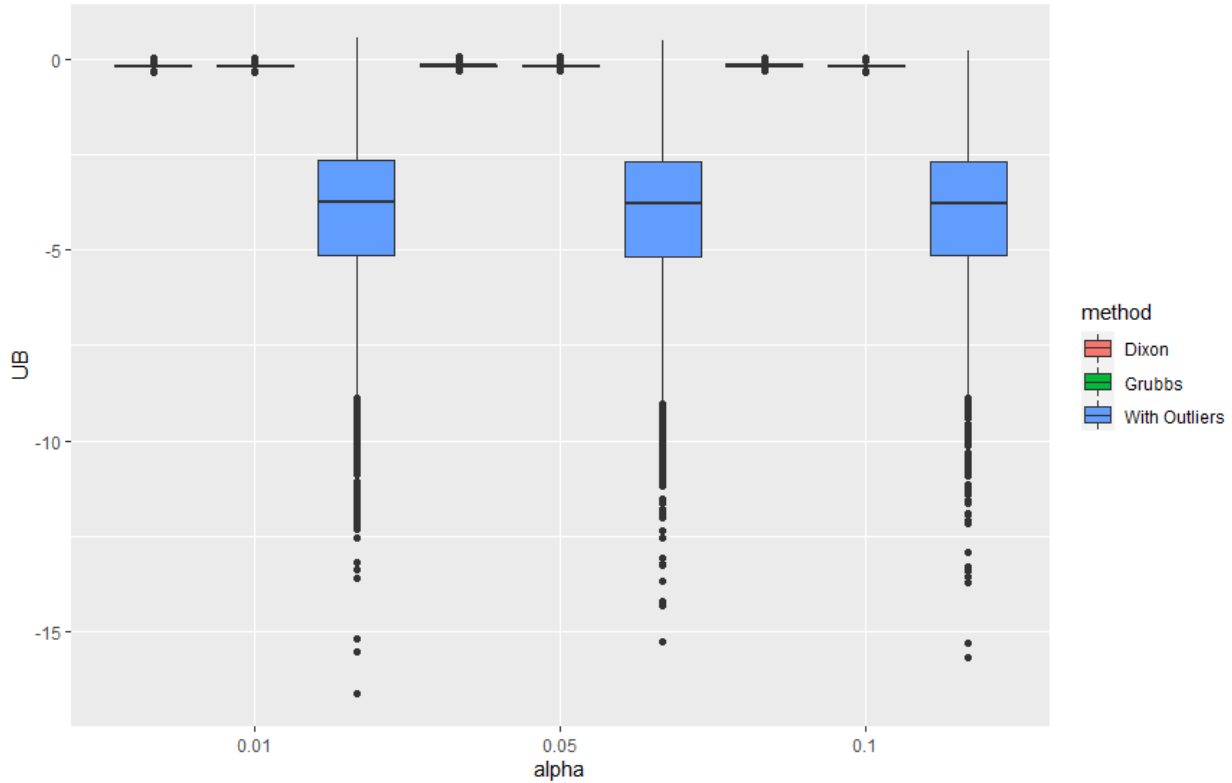
Impact of varying the α-value on the number of identified outliers (Dean-Dixon test)



Impact of changing α value and selecting a specific outlier test on Estimated GMR



Impact of changing α value and selecting a specific outlier test on 95% Upper Bound



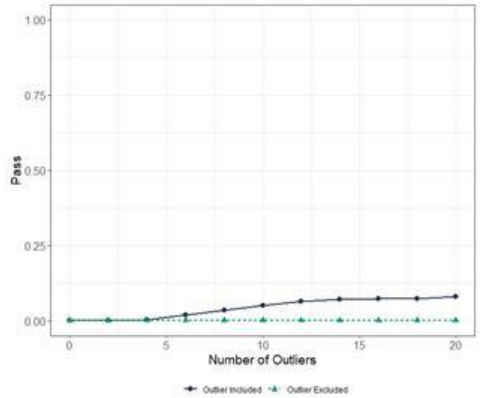
Conclusion

Concluding we can say that:

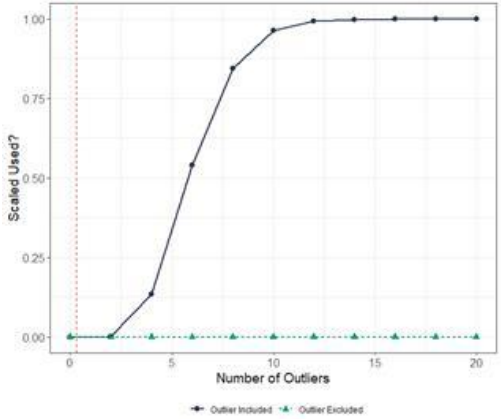
Varying the α value within a specific type of outlier test seems to affect the number of outliers identified. However:

- the impact of a combination of the choice of a specific type of outlier test and the changes of the α value does not seem to be significant *based on the data from the different IVPT studies examined here.*
- *We note that the results are only based on the limited availability of data from IVPT studies.*
- *To further investigate any potential impact of outliers, we may need more data from such applications and/or simulation studies under various configurations.*

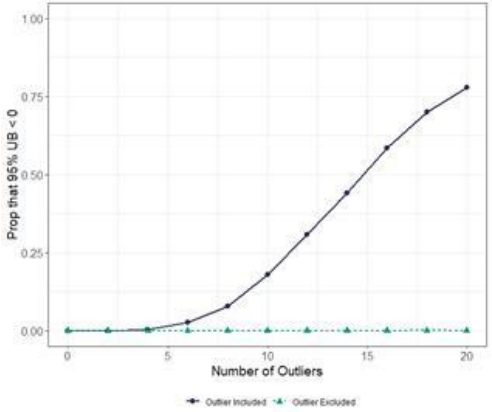
Passing rate



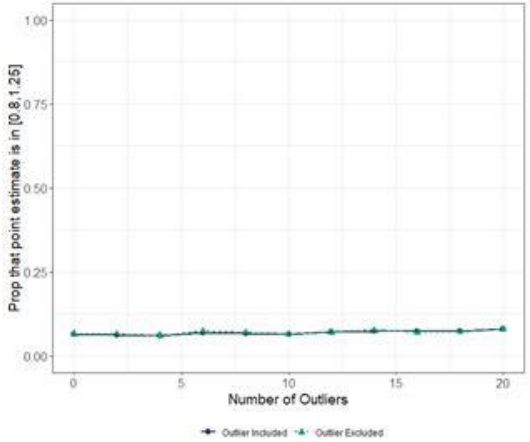
Prop. of time SABE is used



Prop. SABE upper bound ≤0



Prop. point est. is in [0.80, 1.25]

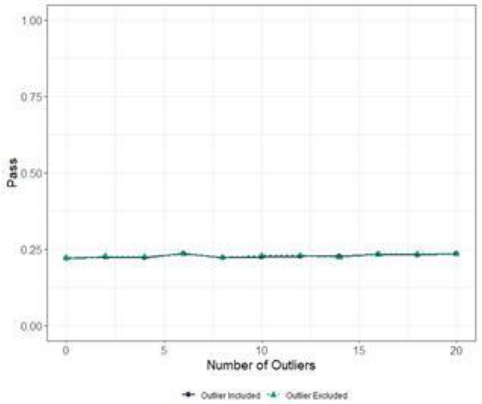


Scenario 1

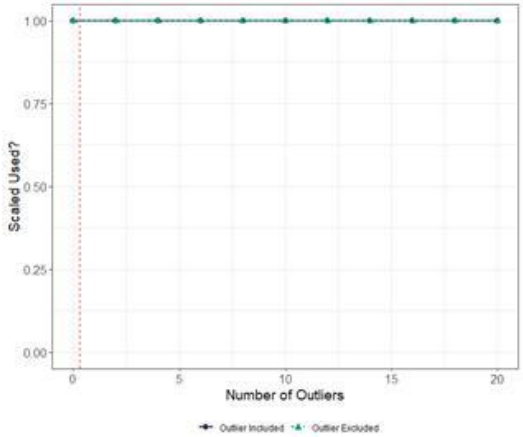
True S_{WR} = 0.2 ,

GMR = 1.3

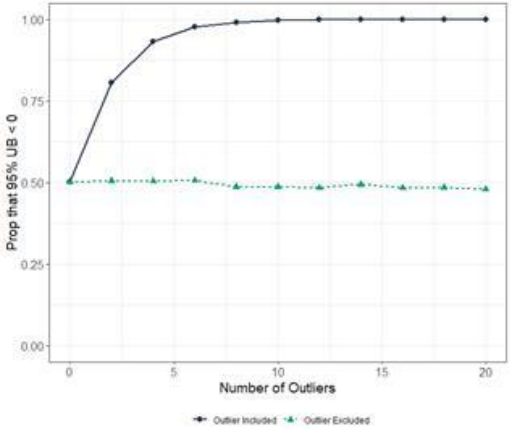
Passing rate



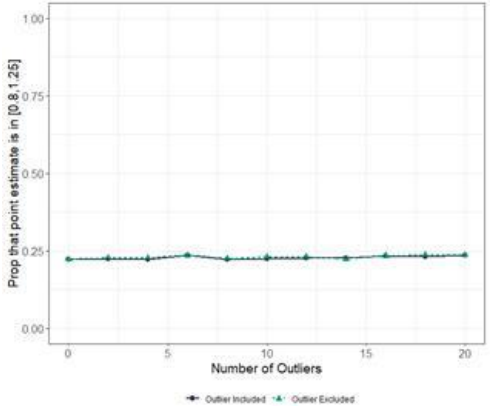
Prop. of time SABE is used



Prop. SABE upper bound ≤0



Prop. point est. is in [0.80, 1.25]



Scenario 2

True S_{WR} = 0.4 ,

GMR = 1.3

Conclusions

Our simulation studies showed, that including outliers does *not necessarily* make a study easier to pass BE, as long as,

- S_{WR} is greater than 0.294 in both cases of including/excluding outliers and
- the point estimate is not affected by including/excluding outliers.

References

Grosser, S., Park, M., Raney, S. G., & Rantou, E. (2015). Determining equivalence for generic locally acting drug products. *Statistics in Biopharmaceutical Research*, *7*(4), 337-345.

Grosser, S., Park, M., Raney, S. G., & Rantou, E. (2013). Clinical Endpoint Studies: Assessing Equivalence of Generic Locally Acting Drug Products. In *Encyclopedia of Biopharmaceutical Statistics* (pp. 1-7). CRC Press.

Wang, W. and Chow, S. (2003). Examining Outlying Subjects and Outlying Records in Bioequivalence Trials. *Journal of Biopharmaceutical Statistics*, 13:1, 43-56.

Schall, R., Endrenyi, L. and Ring, A. (2010). Residuals and outliers in replicate design crossover studies. *Journal of biopharmaceutical statistics*, *20*(4), 835-849.

Hu, D., Rantou, E. and Choi, S. (2021). Detecting Outliers in Replicated Crossover Bioequivalence Studies. *Manuscript in progress*.

Acknowledgments

Stella Grosser

Nam Hee Choi

Jessica Kim

Sungwoo Choi

Neha Agarwala

Sam Raney

Priyanka Ghosh

Paul Lehman

Thank you!


FDA
U.S. FOOD & DRUG
ADMINISTRATION