Mahalanobis Distance Based Approaches, Performance and Limitations

Dr. Thomas Hoffelder

M-CERSI Workshop

In Vitro Dissolution Profiles Similarity Assessment in Support of Drug Product Quality: What, How, and When

May 21-22, 2019

Boehringer Ingelheim

Contents

Introduction

Equivalence analyses of dissolution profiles - statistical description

- **hypotheses:**
  H_0: Non-equivalence of both dissolution profile groups
  versus
  H_1: Equivalence of both dissolution profile groups (goal of study)

- **type I error:**
  wrong decision in favor of equivalence
  regulatory need: control of type I error

- **power:**
  probability of a correct decision in favor of equivalence
  probability of a successful study in case of no relevant differences
  practitioners' need: sufficiently high power

Introduction

- **Distance measure:** Mahalanobis distance (**MD**)

$$MD = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' \, \boldsymbol{\Sigma}^{-1} \, (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$$

$\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$ are the expected values of both reference and test group and $\sum$ is the common covariance matrix

- MD is a multivariate generalization of the **standardized difference** between the expected values: $\frac{\mu_1 - \mu_2}{\sigma}$

- Note: f_2 is based on the Euclidean distance $\sum_{i=1}^{n} (\mu_{1i} - \mu_{2i})^2$ which is the multivariate generalization of the **non-standardized difference** between the expected values: $\mu_1 - \mu_2$

μ_{1i}: expected value of REF at time point nr. *i*

μ_{2i}: expected value of TEST at time point nr. *i*

n: number of dissolution time points

Introduction

- **Distance measure:** Mahalanobis distance (**MD**)

$$MD = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \, \boldsymbol{\Sigma}^{-1} \, (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

$\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are the expected values of both reference and test group and $\sum$ is the common covariance matrix

- **Equivalence hypotheses:**

$$H_0 : MD \geq EM \quad \text{versus} \quad H_1 : MD < EM$$

EM: Equivalence margin (similarity limit / acceptance criterion)

Contents

1) Introduction: Equivalence hypotheses based on the Mahalanobis distance

2) **Equivalence procedures based on the Mahalanobis distance (MD)**
   - **T^2-test for equivalence, ACLMD approach („MSD-test“), bootstrapped MD, T2EQ approach**
   - **properties of MD based approaches**
   - **Discussion on equivalence margin (similarity limit, acceptance criterion)**
   - **Critical comments on MD in the current literature**

3) Experiences with the T2EQ approach
   - dialogue with regulatory agencies, academia and other companies
   - software

4) The Mahalanobis distance: performance and limitations
   - MD versus (bootstrapped) f2 - comments
   - MD versus model dependent approaches – comments
   - The correlation problem

5) Conclusions

MD based equivalence tests

Approximate procedures:

- ACLMD (Approximate Confidence Limit for Mahalanobis Distance) approach according to Tsong et al. (1996), "MSD-test"
- Bootstrapped MD

➔ approximate procedures may be affected by the bias of the MD point estimate

Exact procedure: The T^2-test for equivalence according to Wellek (2010):

- for normally distributed data: **UMPI-test**
- **robust** under deviations from normal distribution assumption
- statistically equivalent to **exact CI procedure for MD** (→ Hoffelder et al., 2015)

- **Problem**: Determination of a **fixed EM** is **not practically feasible**

MD based equivalence tests

Why is a fixed EM for the Mahalanobis distance not practically feasible?

Regulatory perspective:

- manufacturer benefits from an increased variability
- the higher the variability the higher the allowed difference between the profiles
- would not be compliant with EMA (2010) guideline ("similarity acceptance limits should … not be greater than a 10% difference")
- see EMA (2018)

Practitioners' perspective:

- manufacturer punished by a decreased variability
- the lower the variability the lower the allowed difference between the profiles
  ➔ test very sensitive regarding differences at time points with very low variability

MD based equivalence tests

EMA (2018):

Question and answer on the adequacy of the Mahalanobis distance to assess the comparability of drug dissolution profiles

Draft agreed by Biostatistics Working Party	June 2018
Adopted by CHMP	26 July 2018

Keywords	Bioequivalence, dissolution profiles, f2, Mahalanobis distance, biowaiver

MD based equivalence tests

Why is a fixed EM for the Mahalanobis distance not practically feasible?

Regulatory perspective:

- manufacturer benefits from an increased variability
- the higher the variability the higher the allowed difference between the profiles
- would not be compliant with EMA (2010) guideline ("similarity acceptance limits should … not be greater than a 10% difference")
- see EMA (2018)

Practitioners' perspective:

- manufacturer punished by a decreased variability
- the lower the variability the lower the allowed difference between the profiles
  ➔ test very sensitive regarding differences at time points with very low variability

MD based equivalence tests

Solution: Specific choice of the EM

Tsong et al. (1996):

- "**global similarity limit**" for MD in the context of comparing dissolution profiles
  → product independent, always practically feasible EM.
- EM is defined by a max. allowed constant shift between REF and TEST $\boldsymbol{D} = (d,d,\ldots,d)$
- this max. allowed shift is related to the present variability $\mathrm{EM} = \boldsymbol{D}'\boldsymbol{S}^{-1}\boldsymbol{D}$
  $\boldsymbol{S}$: pooled empirical covariance matrix

EMA (2010):

- "similarity acceptance limits should … not be greater than a 10% difference"

➔ $EM = \boldsymbol{D}'\boldsymbol{S}^{-1}\boldsymbol{D}$, where $\boldsymbol{D}$ = (10,10,…,10)

➔ EM defined by a difference of 10% at all time points
same interpretation as for f_2 !
Acceptance criterion: "Average" profile difference < 10%

MD based equivalence tests

T2EQ approach

- T^2-test for equivalence according to Wellek (2010)
- EM according to Tsong et al. (1996)
- EM restrictions according to EMA (2010)

} **T2EQ approach**

- **p-value of the test** can be calculated
  => "p-value < 0.05" is not more difficult to understand as "$f_2 > 50$"
  => no simulations or numerical methods necessary
- test result can be reported via p-value **or** via upper confidence limit for the Mahalanobis distance

MD based equivalence tests

What about control of type I error and power?

In Hoffelder (2018b) several methods, i.e.

- T2EQ approach
- ACLMD ["MSD-Test"] → Tsong et al. (1996)
- bootstrapped MD
- bootstrapped f_2

were compared regarding **control of type I error, robustness and power.**

Note that all methods use **the same acceptance criterion** (limit between equivalence and non-equivalence is defined by a shift in location of 10% at all dissolution time points).

Most advantageous approach in the simulation study: T2EQ

MD based equivalence tests

Critical comments on MD in the current literature (→ §5.1 in Hoffelder, 2018b)

Mahalanobis distance shrinks when the variance of the experiment increases

- ➔ the higher the variability the higher is the power
- ➔ "poorly designed experiments would be rewarded"
- ➔ not compliant with EMA (2010) because a decision in favor of equivalence might be possible in spite of a profile difference > 10% (see EMA, 2018)

T2EQ approach not affected by these concerns because

- the covariance matrix enters into both sides of the equivalence hypotheses, into **MD as well as into EM.**
  Increased variance ➔ **EM decreases in the same way as MD.**
- EM **defined by a shift in location of 10%** at all dissolution time points.
  ➔ **T2EQ compliant with the EMA guideline** independent of the present covariance matrix
- Remark: Simulations in Hoffelder (2018b): increased variability reduces the power

MD based equivalence tests

Discussion on the equivalence margin $EM = \boldsymbol{D}'\boldsymbol{S}^{-1}\boldsymbol{D}$, $\boldsymbol{D}$ = (10,10,...,10)

- EM is a **random variable** via $\boldsymbol{S}$
- critisized as being **"data driven"** (equivalence margins should usually be pre-defined and fixed so that **no cherry picking** is possible)

But:

- **construction of the T2EQ EM is pre-defined and fixed ➔** No cherry picking possible
- perspective of using a non-standardized distance measure.
  For a standardized distance measure (e.g. MD) it holds:
  EM fixed ➔ maximum allowed constant shift in location between the profiles is "data driven" via empirical covariance matrix $\boldsymbol{S}$
  T2EQ EM being a random variable, not fixed
  ➔ **maximum allowed constant shift in location is fixed to 10%!**
  EMA (2010) and EMA (2018) emphasize that the focus on dissolution profile equivalence is the **shift in location.**

Contents

1) Introduction: Equivalence hypotheses based on the Mahalanobis distance

2) Equivalence procedures based on the Mahalanobis distance (MD)
   - T^2-test for equivalence, ACLMD approach („MSD-test“), bootstrapped MD, T2EQ approach
   - properties of MD based approaches
   - Discussion on equivalence margin (similarity limit, acceptance criterion)
   - Critical comments on MD in the current literature

3) **Experiences with the T2EQ approach**
   - **dialogue with regulatory agencies, academia and other companies**
   - **software**

4) The Mahalanobis distance: performance and limitations
   - MD versus (bootstrapped) f2 - comments
   - MD versus model dependent approaches – comments
   - The correlation problem

5) Conclusions

Experiences with T2EQ

Dialogue with regulators, academia and industry:

- DIA/FDA Statistics Forum 2016:
  - Roundtable discussion „Similarity of Dissolution Profiles“
  - Poster „Comparison of Dissolution Profiles: A Statistician’s Perspective“

  ➔ Paper Hoffelder (2018a)

- CEN-ISBS Vienna 2017:
  - Presentation of T2EQ simulation results on size, robustness, power

  ➔ Papers Hoffelder (2018b,2019)

- T2EQ approach and the evaluation strategy addressed in Hoffelder (2018a) submitted to more than 20 countries/regions

  ➔ Sample size calculations, increased sample size ✓

  ➔ no pairwise batch-to-batch comparisons ✓

  ➔ T2EQ approach ✓

Experiences with T2EQ

T2EQ approach summary

1) „Easy“ to understand for non-statisticians:
   Analogue to f_2 => acceptance criterion, p-value

2) Regulatory perspective: sufficient control of type I error

3) Industry perspective: sufficiently high power, sample size calculations can be done

4) R package T2EQ available on CRAN,
   SAS/IML modules published in Hoffelder (2018b)

=> T2EQ ready for use in pharmaceutical practice

Contents

1) Introduction: Equivalence hypotheses based on the Mahalanobis distance

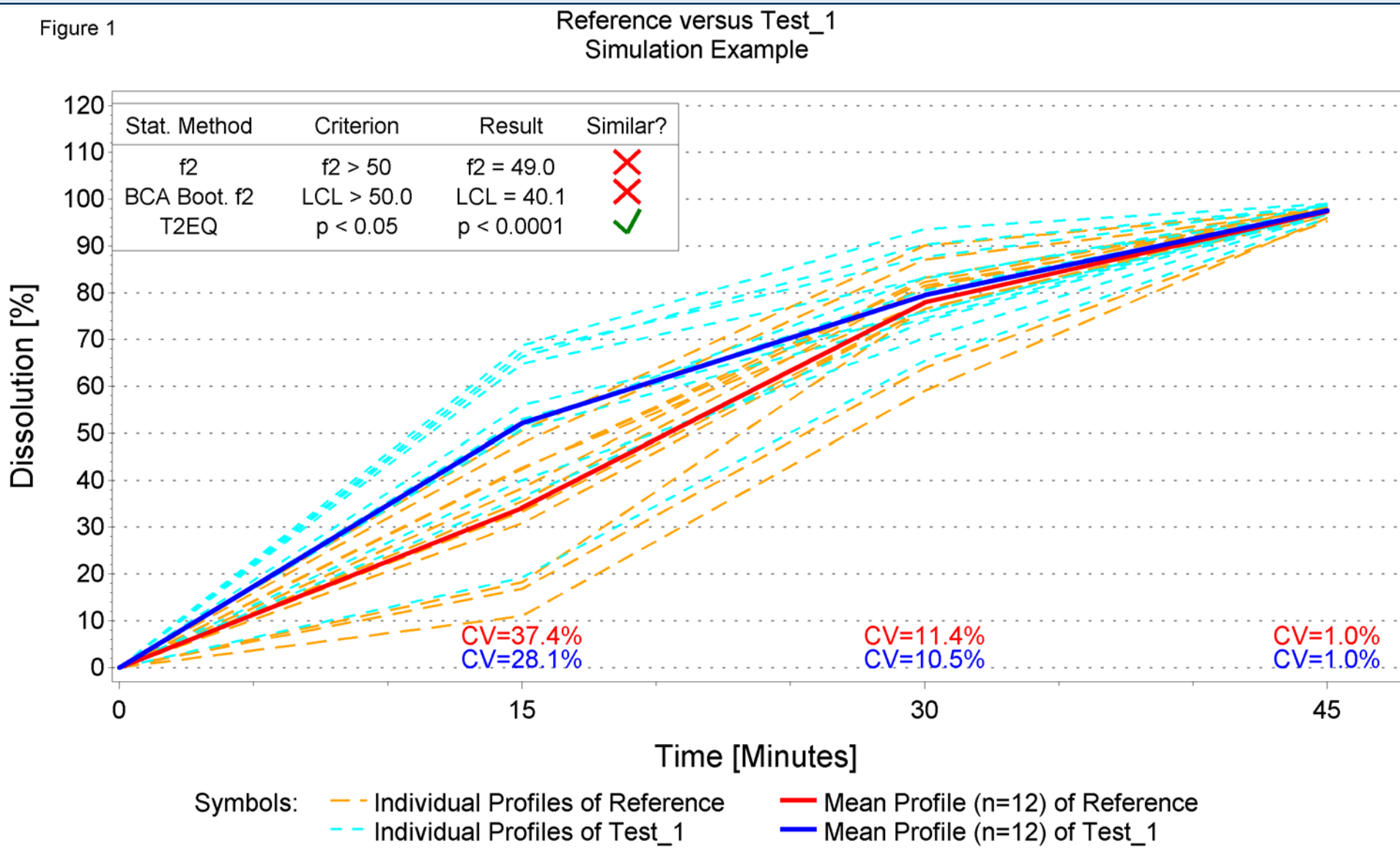
2) Equivalence procedures based on the Mahalanobis distance (MD)
   - T^2-test for equivalence, ACLMD approach („MSD-test“), bootstrapped MD, T2EQ approach
   - properties of MD based approaches
   - Discussion on equivalence margin (similarity limit, acceptance criterion)
   - Critical comments on MD in the current literature

3) Experiences with the T2EQ approach
   - dialogue with regulatory agencies, academia and other companies
   - software

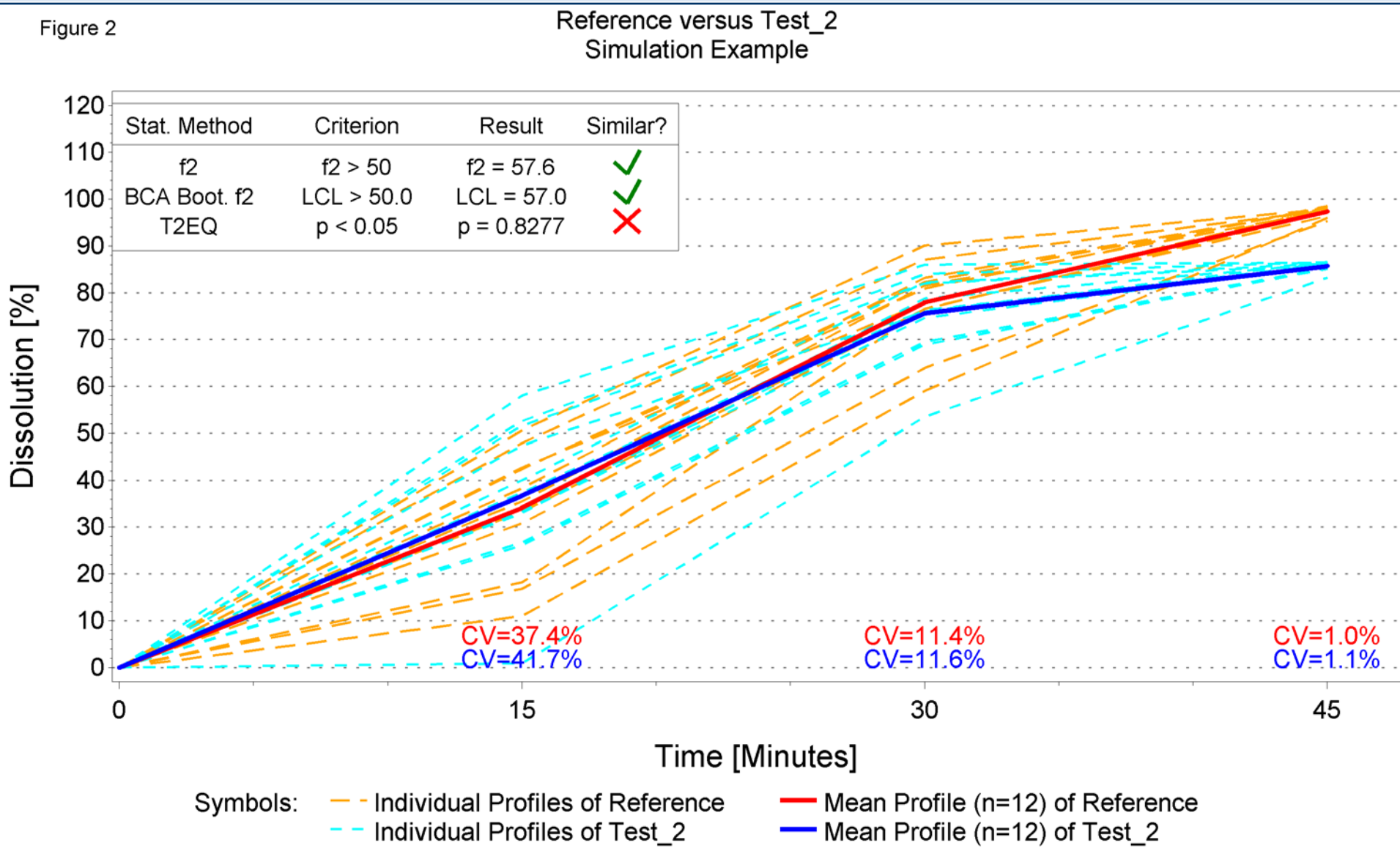
4) The Mahalanobis distance: performance and limitations
   - **MD versus (bootstrapped) f2 - comments**
   - **MD versus model dependent approaches – comments**
   - **The correlation problem**

5) Conclusions

T2EQ versus f_2 (source of figures: Hoffelder, 2018a)



T2EQ versus f_2 (source of figures: Hoffelder, 2018a)



T2EQ versus f_2

Mahalanobis distance versus Euclidean distance (f_2):

- **ED** summarizes **absolute differences** between the mean values.
  - ➔ Differences from time points with maybe considerably **different variabilities** are amalgamated into one distance measure **without any standardization.**
  - ➔ One cannot distinguish whether a certain difference mainly stems from random effects or whether the difference is caused by a systematic effect.
- **MD** summarizes **standardized differences**
  - ➔ allows to distinguish between systematic and random effects as root cause for the profile differences.
- Recall the profile difference in Figure 2
  - ➔ highly significant difference in the last time point ➔ irrelevant for patients?

➔ Compared with ED, a **standardized distance measure as MD** gives a better assessment on the question which differences are ***practically relevant*** and which are not

MD versus Model Dependent Approaches

Mahalanobis distance versus Model Dependent Approaches I:

- for products where only three or four time points are available (e.g. immediate release products completely dissolving within 20 minutes) **sufficient data** for a reliable fit of the profiles might **not be available.**
- a **product-independent** optimal nonlinear model does not exist
- **model selection problems** for Model Dependent Approaches
  → similarity decision might depend on chosen nonlinear regression model
- **Definition of the equivalence criterion** in terms of the parameters of nonlinear functions
  ➔ complex, practically feasible, product-independent equivalence margins / acceptance criteria available?
- In contrast, methods based on MD or ED directly evaluate the dissolution profile data and a **theoretical assumption on the shape of the profiles is not necessary.**

MD versus Model Dependent Approaches

Mahalanobis distance versus Model Dependent Approaches II:

- Maximum deviation based approach (MDBA) from Collignon et al. (2018)
  + Decision in favor of equivalence if **maximum absolute deviation** between both fitted nonlinear regression curves is below a certain acceptance limit, e.g. 10%.
  + The maximum difference between the fitted curves might be at a time point that is not measured (recall example evaluation in Collignon et al., 2018).
  + Possible situation: estimated maximum deviation > 10% but a difference above 10% is not observed at any measured time point.
- In contrast, methods based on MD or ED directly evaluate the dissolution profile data and a theoretical assumption on the shape of the profiles is not necessary.

➔ **Question for BO sessions: Focus of dissolution profile equivalence:**

Average difference or maximum difference?

Limitation of MD based approaches

Limitation: The correlation problem

- One can interpret the T2EQ acceptance criterion as "MD based weighted mean of the profile differences lower than 10%". The **weights** can be interpreted as a **combination of variances and correlations.**

- It is possible that the correlations play an important role in the similarity decision.

- Example: Let $\Sigma = \begin{pmatrix} 120 & 39 & -9 \\ 29 & 146 & 111 \\ -9 & 111 & 113 \end{pmatrix}$ ➔ Difference (3,3,3) ➔ MD = 0.22

  ➔ Difference (3,-3,3) ➔ MD = 2.49

- No practically relevant difference between shifts in location (3,3,3) and (3,-3,3)

➔ strong influence of correlation / hard to understand for non-statisticians.

Limitation of MD based approaches

Limitation: The correlation problem

Possible solution:

- **Standardization with variances only** - not with complete covariance matrix
- Differences (3,3,3) and (3,-3,3) would result in the same (estimated) distance
- **Approach already exists:** SE-test according to Hoffelder et al. (2015)
  EM can be analogically derived as for T2EQ approach
- acceptance criterion can be interpreted as "**weighted mean of the profile differences lower than 10%**". The weights are the variances.
- To-do-list:
  paper with detailed simulation study on type I error, robustness and power.

Contents

1) Introduction: Equivalence hypotheses based on the Mahalanobis distance

2) Equivalence procedures based on the Mahalanobis distance (MD)
   - T^2-test for equivalence, ACLMD approach („MSD-test“), bootstrapped MD, T2EQ approach
   - properties of MD based approaches
   - Discussion on equivalence margin (similarity limit, acceptance criterion)
   - Critical comments on MD in the current literature

3) Experiences with the T2EQ approach
   - dialogue with regulatory agencies, academia and other companies
   - software

4) The Mahalanobis distance: performance and limitations
   - MD versus (bootstrapped) f2 - comments
   - MD versus model dependent approaches – comments
   - The correlation problem

5) **Conclusions**

Conclusions

1) **Standardized distance measures (as e.g. MD) are statistically preferable to decide which differences are practically relevant and which are not.**

2) **T2EQ is ready for use in pharmaceutical practice**

3) **Solution of the „correlation problem“: work in progress ➔ teamwork?**

Literature

Collignon, O., Möllenhoff, K., Dette, H. (2019). Equivalence analyses of dissolution profiles with the Mahalanobis distance: a regulatory perspective and a comparison with a parametric maximum deviation-based approach. *Biometrical Journal* **61**, 779–782.

EMA (2010). *Guideline on the Investigation of Bioequivalence*. London: European Medicines Agency, CHMP.

EMA (2018). *Question and answer on the adequacy of the Mahalanobis distance to assess the comparability of drug dissolution profiles*. London: European Medicines Agency.

FDA (1997). *Guidance for Industry: Dissolution Testing of Immediate Release Solid Oral Dosage Forms*. Rockville: Food and Drug Administration FDA, CDER.

Hoffelder, T., Gössl, R., Wellek, S. (2015). Multivariate equivalence tests for use in pharmaceutical development. *Journal of Biopharmaceutical Statistics* **25**, 417–437.

Hoffelder, T. (2018a). Comparison of dissolution profiles – a statistician's perspective. *Therapeutic Innovation & Regulatory Science* **52:4**, 423–429.

Hoffelder, T. (2018b). Equivalence analyses of dissolution profiles with the Mahalanobis distance. *Biometrical Journal*. Accepted for the CEN-ISBS Vienna 2017 Special Issue. *https://doi.org/10.1002/bimj.201700257*

Hoffelder, T. (2019). Author response to the Letter to the Editor "Equivalence analyses of dissolution profiles with the Mahalanobis distance: A regulatory perspective and a comparison with a parametric maximum deviation-based approach". *Biometrical Journal*. Accepted. *https://doi.org/10.1002/bimj.201900047*

Tsong, Y., Hammerstrom, T., Sathe, P., Shah, V. P. (1996). Statistical assessment of mean differences between two dissolution data sets. *Drug Information Journal* **30**, 1105--1112.

Wellek, S. (2010). *Testing Statistical Hypotheses of Equivalence and Noninferiority. Second Edition*. Boca Raton: Chapman & Hall / CRC Press LLC.

End

Thank you!

End

Back up

MD based equivalence tests

Standard approach: similarity factor f_2

- n: number of dissolution time points, $\boldsymbol{R} = (R_1,\ldots,R_n)$: reference mean profile, $\boldsymbol{T} = (T_1,\ldots,T_n)$: test mean profile
  f_2 is a transformation of the quadratic mean (over time) of the differences between reference and test mean:

$$QMD = \sqrt{(1/n)\sum_{t=1}^{n}(R_t - T_t)^2} \quad \Rightarrow \quad f_2 = 50\log_{10}\left(\frac{100}{\sqrt{1+QMD^2}}\right)$$

- acceptance criterion: $f_2 > 50 \Leftrightarrow QMD < \sqrt{99} = 9.95 \approx 10$

  ➔ profiles similar if **<u>average difference between profile means is below 10%</u>**

- f_2 depends on means only, point estimate, no control of type I error

  ➔ the less reliable the higher the variability of the underlying data

  ➔ guideline restrictions for f_2 if variability exceeds certain thresholds

- no guideline recommendation for highly variable dissolution profiles

  ➔ **WANTED: suitable multivariate equivalence procedure**