Similarity and Comparability

Helmut Schütz

Similarity in terms of dissolution testing

Similarity of dissolution important in various areas

- **Product development.**
  - Candidate formulations with different release characteristics.
  - Selection of a candidate matching the reference.
  - Selection of a reference batch for an *in vivo* study.
    - Russia, Egypt: Must pass f_2 before a biostudy can be performed. Bizarre.
- **Quality control (Session 7).**
  - Set specifications which likely not affect *in vivo* performance.
- **Biowaivers (Session 9).**
  - Dose proportionality: Biostudy of different strenghts waived.
  - BCS-based biowaivers: Biostudy waived based on f_2 similarity in three media. Class I (Class III drugs under certain conditions).
- **Life cycle.**
  - Support changes of the formulation (EMA minor variation, FDA SUPAC).

Difference factor f_1, similarity factor f_2

Difference factor f_1 (Russia, Brazil)

- Percent difference between dissolution profiles at each time point.

- Measurement of the relative error between the curves.

$$f_1 = 100 \left\{ \sum_{t=1}^{t=n} \left| \overline{R}_t - \overline{T}_t \right| \Bigg/ \sum_{t=1}^{t=n} \overline{R}_t \right\}$$

Similarity factor f_2 (all jurisdictions)

- Logarithmic reciprocal square root transformation of the sum of squared error.

- Measurement of the similarity in the percent dissolution between the curves.
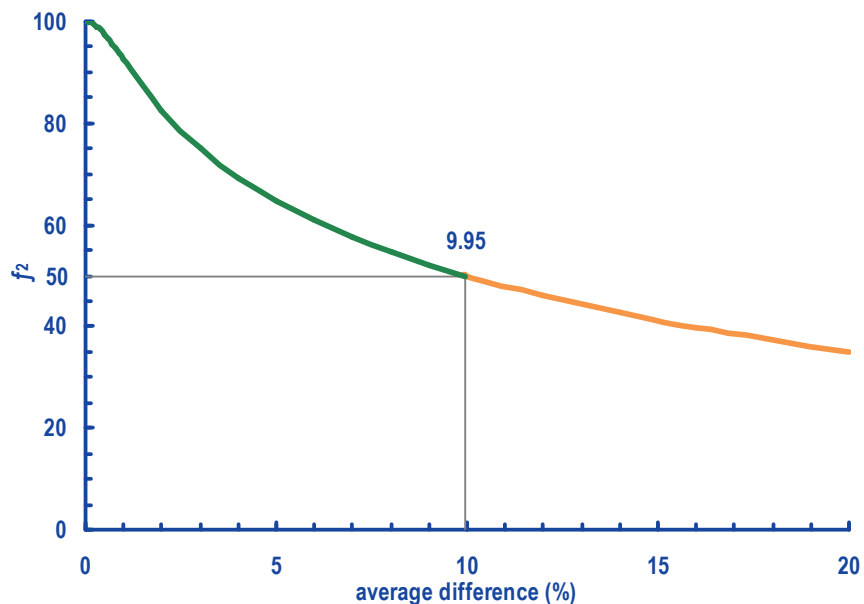
$$f_2 = 50 \cdot \log \left\{ 100 \cdot \left[1 \Bigg/ \sqrt{1 + \frac{1}{n} \sum_{t=1}^{t=n} \left(\overline{R}_t - \overline{T}_t \right)^2} \right] \right\}$$

Moore JW, Flanner HH. *Mathematical Comparison of curves with an emphasis on in vitro dissolution profiles*. Pharm Tech. 1996;20(6):64–74.

Fleming.

Difference factor f_1, similarity factor f_2

Similarity factor f_2

- *Average* difference between two profiles of ~10%
  at *all* sampling data points corresponds to f_2 of 50.

Difference factor f_1, similarity factor f_2

Simple example

n	3
$\Sigma\,(R_t - T_t)$	10
$\Sigma\,|R_t - T_t|$	10
$\Sigma\,(R_t - T_t)^2$	38
$\Sigma\,R_t$	258
f_1	3.9
f_2	71.6

t (min)	R_t (%)	T_t (%)	$\Delta\,(R_t - T_t)$	$\Delta\,|R_t - T_t|$	Δ^2
15	83	78	5	5	25
30	85	83	2	2	4
45	90	87	3	3	9

- **Somewhat strange concept…**
  - In statistics we would compare T with R and hence, use T – R and not R – T.
  - If we reverse the values, we would get f_1 4.0.
  - However, the same f_2 because it is based on the squared differences, where the order is not relevant.

Fleming.

Difference factor f_1, similarity factor f_2

## Certain conditions must be fullfilled for the application of f_2

- Three media: pH 1.2, 4.5, 6.8

- f_2 is *not required* if products release ≥85% in all three media.

- At least three time points, identical for both formulations.

- 12 units of test and reference product.
  R_t and T_t are their arithmetic means.

- Sampling time points after 85% release
  - EMA: Not more than one mean value for any of the formulations.
  - FDA: Only one measurement included for the test formulation.
  - WHO: Only one measurement included for the reference formulation.

- Similarity concluded if f_2 ≥50.

Difference factor f_1, similarity factor f_2

Example 1

- Simulated data, T exactly 90% of R at each time point. EMA-rule: We stop the calculation at 9 h (only one time point with >85% dissolved).

t (h)	R_t (%)	T_t (%)	$\Delta\,(R_t - T_t)$	$\Delta\,\lvert R_t - T_t \rvert$	Δ^2
1.0	20.9	18.8	+2.1	2.1	4.4
2.0	37.5	33.7	+3.7	3.7	14.1
3.0	50.6	45.5	+5.1	5.1	25.6
4.0	60.9	54.8	+6.1	6.1	37.1
5.0	69.1	62.2	+6.9	6.9	47.8
6.0	75.6	68.0	+7.6	7.6	57.1
7.0	80.7	72.6	+8.1	8.1	65.1
8.0	84.7	76.3	+8.5	8.5	71.8
9.0	87.9	79.1	+8.8	8.8	77.3
10	90.5	81.4	+9.0	9.0	–

$$diss(\%) = 100(1 - e^{-0.235 \cdot t})$$

$f1 = 10.0$
$f2 = 58.6$
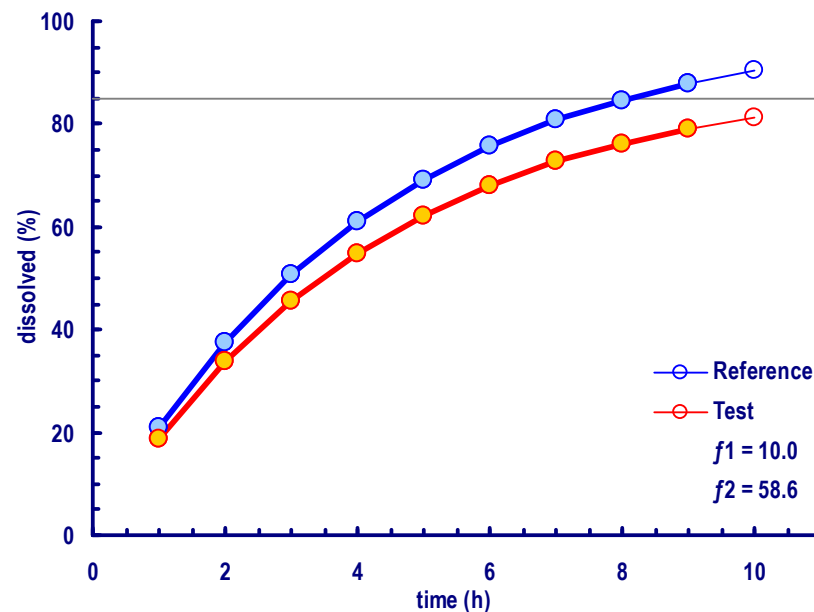
Example 2

- Same function but without the 7 h time point.
  Identical f_1; based on f_2 formulations are 'more similar' (58.6 → 59.2).

t (h)	R_t (%)	T_t (%)	$\Delta (R_t - T_t)$	$\Delta |R_t - T_t|$	Δ^2
1.0	20.9	18.8	+2.1	2.1	4.4
2.0	37.5	33.7	+3.7	3.7	14.1
3.0	50.6	45.5	+5.1	5.1	25.6
4.0	60.9	54.8	+6.1	6.1	37.1
5.0	69.1	54.8	+6.9	6.9	47.8
6.0	75.6	62.2	+7.6	7.6	57.1
8.0	84.7	68.0	+8.5	8.5	71.8
9.0	87.9	79.1	+8.8	8.8	77.3
10	90.5	81.4	+9.0	9.0	–
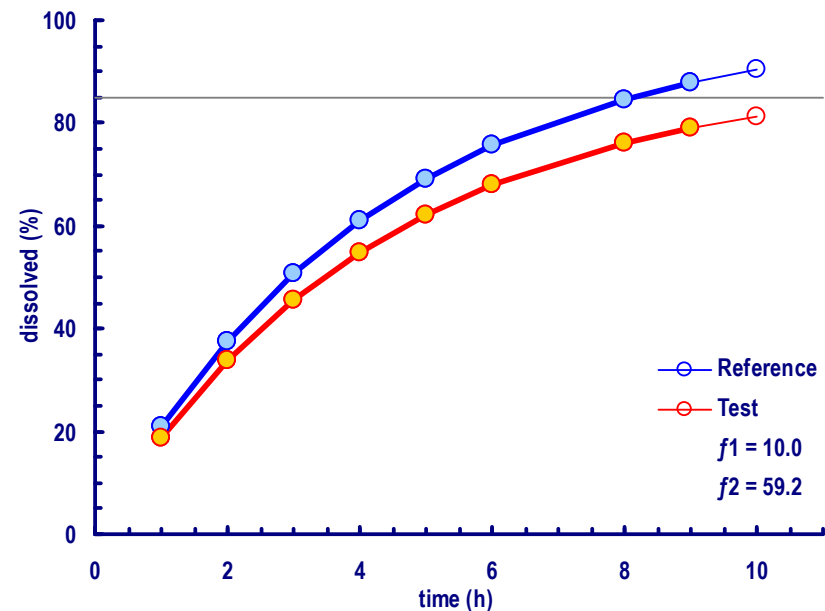
$f1 = 10.0$
$f2 = 59.2$

Example 3

- Same function and same number of time points like in Example 1 but different early time points. Identical f_1 but f_2 gets 'better' (58.6 → 59.2).

t (h)	R_t (%)	T_t (%)	$\Delta (R_t - T_t)$	$\Delta |R_t - T_t|$	Δ^2
0.5	11.1	10.0	+1.1	1.1	1.2
1.5	29.7	26.7	+3.0	3.0	8.8
2.5	44.4	40.0	+4.4	4.4	19.7
3.5	56.1	50.5	+5.6	5.6	31.4
4.5	65.3	58.7	+6.5	6.5	42.6
6.0	75.6	68.0	+7.6	7.6	57.1
7.0	80.7	72.6	+8.1	8.1	65.1
8.0	84.7	76.3	+8.5	8.5	71.8
9.0	87.9	79.1	+8.8	8.8	77.3
10	90.5	81.4	+9.0	9.0	–
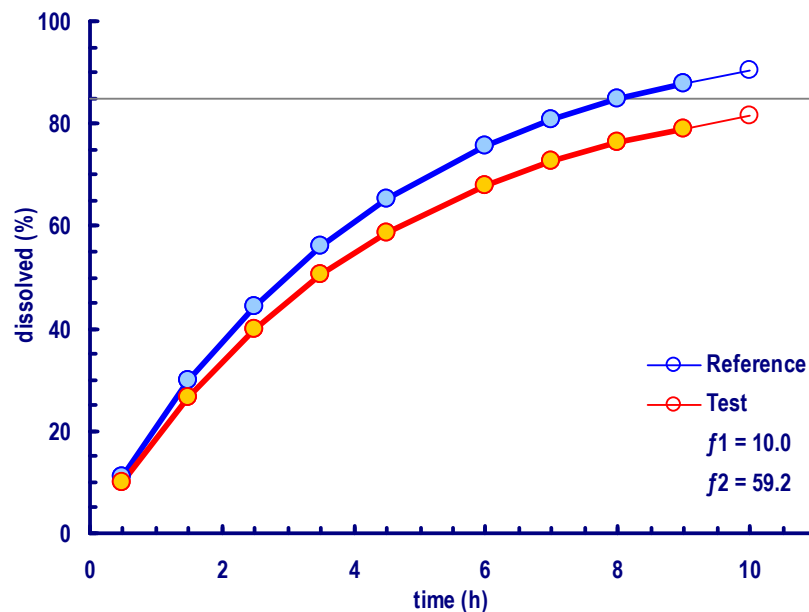
Reference
Test
$f1 = 10.0$
$f2 = 59.2$

Example 4

- We could even shuffle the values and get identical f_1 and f_2. Nonsense, of course but should not be possible for a correct statistical method.

t (h)	R_t (%)	T_t (%)	$\Delta\,(R_t - T_t)$	$\Delta\,|R_t - T_t|$	Δ^2
1.0	65.3	58.7	+6.5	6.5	42.6
1.5	20.9	18.8	+2.1	2.1	4.4
2.5	44.4	40.0	+4.4	4.4	19.7
3.5	56.1	50.5	+5.6	5.6	31.4
4.5	80.7	72.6	+8.1	8.1	65.1
6.0	75.6	68.0	+7.6	7.6	57.1
7.0	84.7	76.3	+8.5	8.5	71.8
8.0	29.7	26.7	+3.0	3.0	8.8
9.0	87.9	79.1	+8.8	8.8	77.3
10	90.5	81.4	+9.0	9.0	–

Reference

Test

$f1 = 10.0$

$f2 = 59.2$
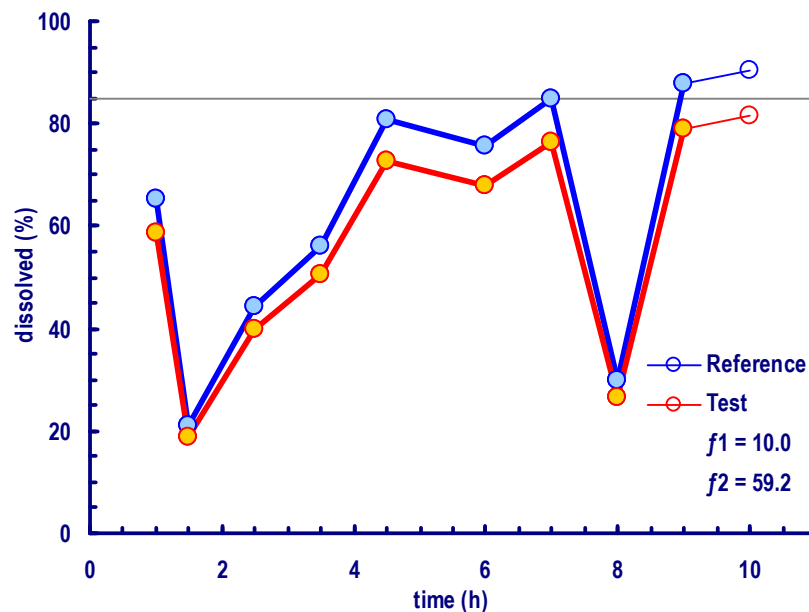
Similarity factor f_2

Problems

- f_2 is *not* a statistic but an *arbitrary* (read: conventient) measure.
  - Different time points give different f_2 values.
  - Different number of time points give different f_2 values.
  - Was criticized from the statistical community.[*]
    - Mean of the underlying distribution is difficult to derive.
    - Variance even more difficult; confidence intervals cannot be derived analytically (requires bootstrapping).
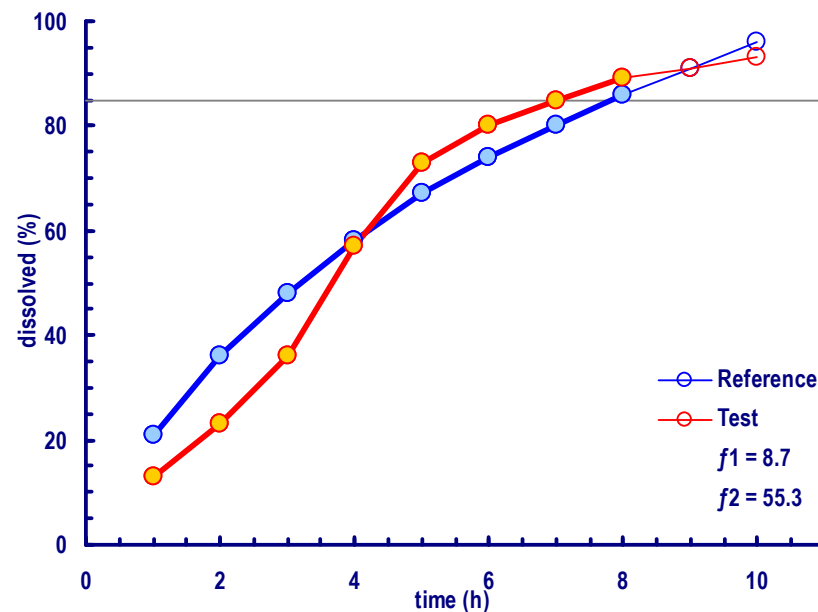   - Shape of profiles and correlation of time points is not taken into account.

[*] Liu J-P, Ma M-C, Chow S-C. *Statistical Evaluation of Similarity Factor f_2 as a Criterion for Assessment of Similarity between Dissolution Profiles*. Drug Inf J. 1997;31:1255–71.

Similarity factor f_2

Different release characteristics

- Although f_1 (8.7) and f_2 (55.3) suggest similarity, the comparison is not suitable because formulations exhibit different release characteristics.

t (h)	R_t (%)	T_t (%)	$\Delta (R_t - T_t)$	$\Delta |R_t - T_t|$	Δ^2
1	21	13	+8	8	64
2	36	23	+13	13	169
3	48	36	+12	12	144
4	58	57	+1	1	1
5	67	73	−6	6	36
6	74	80	−6	6	36
7	80	85	−5	5	25
8	86	89	−3	3	9
9	91	91	±0	0	−
10	96	93	+3	3	−

Reference: Zero order function
Test: Sigmoidal (Weibull)

$f1 = 8.7$
$f2 = 55.3$

Vivian Gray, *Dissolution Workshop*. 10 December 2010.

Fleming.

Similarity factor f_2

Additional criteria (variability)

- **All guidelines:**
    - *CV* should not be >20% at ≤15 minutes.
    - *CV* should not be >10% at other time points.

Similarity factor f_2

Example 5

- Alhough f_1 and f_2 are calculated from the means of units, we have to observe the *CV* as well.

t (h)	R_t (%) mean	CV	T_t (%) mean	CV
0.25	8.2	20.2	7.0	22.9
0.50	16.0	7.5	15.0	9.0
0.75	23.0	8.1	20.9	6.4
1.0	29.9	4.8	25.7	10.1
1.5	39.0	6.8	37.4	7.5
2.0	50.8	8.9	45.0	8.2
3.0	65.6	5.9	56.5	7.6
4.0	73.9	7.0	68.8	8.0
5.0	83.1	4.3	75.6	8.3
6.0	89.0	7.4	80.7	4.7

Reference
Test
$f1 = 9.6$
$f2 = 63.0$

Fleming.
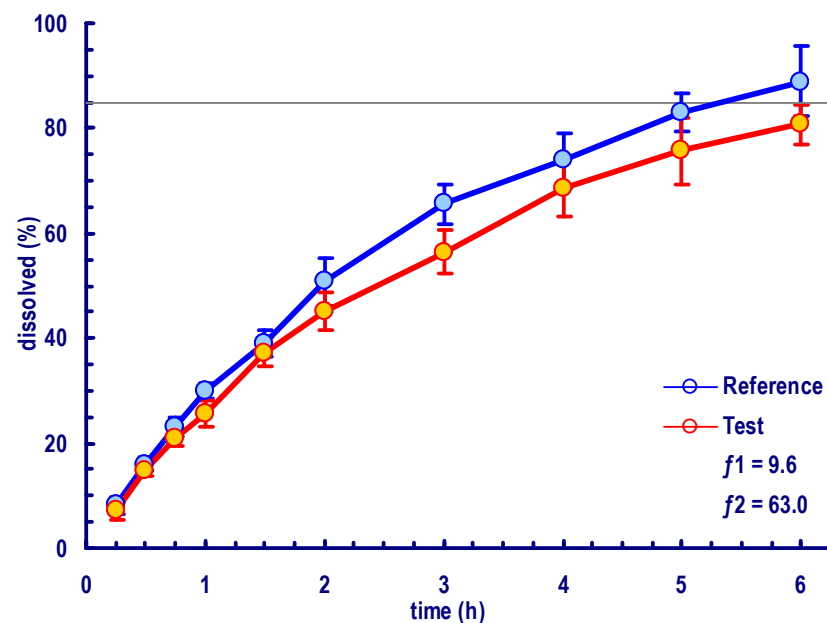
Alternatives to f_2 if conditions not fulfilled?

Suggested if variability (especially in early time points) is high

- **Multivariate Statistical Distance (MSD) [1]**
  - **MSD is estimated.**
    - **Its 90% confidence interval calculated.**
    - **The upper limit compared to the similarity limit.**
  - **A subset of MSD is the Mahalanobis' Distance (MD) [2]**
    - **Not acceptable for the EMA (Q&A July 2018).**
- **Model-dependent approaches**
  - **Select a suitable model (quadratic, logistic, probit, Hill, Weibull, …).**
  - **Similarity region is specified based on the variability.**
  - **Calculate MSD and CI as above.**

1 Cardot J-M, Roudier B, Schütz H. *Dissolution comparisons using a Multivariate Statistical Distance (MSD) test and a comparison of various approaches for calculating the measurements of dissolution profile comparison*. AAPS J. 2017;19(4):1091–101. doi:10.1208/s12248-017-0063-y.
2 Mangas-Sanjuan V, Colon-Useche S, Gonzalez-Alvarez I, Bermejo M, Garcia-Arieta A. *Assessment of the Regulatory Methods for the Comparison of Highly Variable Dissolution Profiles*. AAPS J. 2016;18(6):1550–61. doi:10.1208/s12248-016-9971-5.

Bootstrapping

Suggested if variability (especially in early time points) is high

- EMA/810713/2017 (May 2018).
  - Any approach based upon confidence intervals for f_2 would, however, be considered appropriate whether the validity criteria outlined in CHMP guidance are met or not [CPMP/EWP/QWP/1401/98 Rev. 1/ Corr **].
  - Similarity if the confidence interval for f_2 entirely above 50.
  - f_2 sampling distribution does not allow the derivation of exact confidence intervals to adequately quantify the uncertainty of the f_2 estimate.
  - Bootstrap methodology[1,2,3] could be used to derive confidence intervals for f_2 based on quantiles of resampling distributions, and this approach could actually be considered the preferred method.

1 Shah VP, Tsong Y, Sathe P, Liu J-P. *In Vitro Dissolution Profile Comparison—Statistics and Analysis of the Similarity Factor, f_2*. Pharm Res. 1998;15(6):889–96. doi:10.1023/A:1011976615750.

2 Paixão P, Gouveia LF, Silva N, Morais JAG. *Evaluation of dissolution profile similarity – Comparison between the f_2, the multivariate statistical distance and the f_2 bootstrapping methods*. Eur J Pharm Biopharm. 2017;112:67–74. doi:10.1016/j.ejpb.2016.10.026.

3 Mendyk A, Pacławski A, Szlęk J, Jachowicz R. *PhEq_bootstrap: an Open Source software for simulation of f2 distribution in cases of a large variability in the dissolution profiles*. Diss Technol. 2013;20(1):13–7. doi:10.14227/DT200113P13.
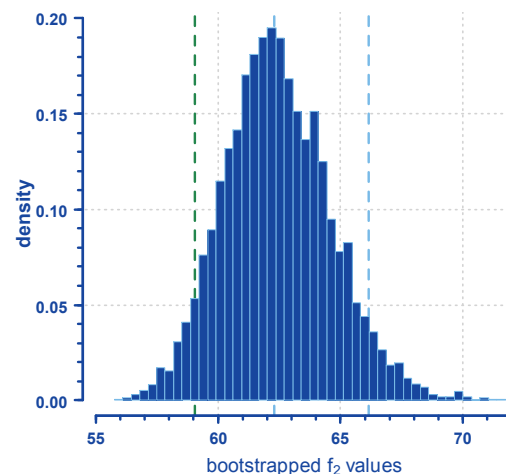
Bootstrapping

Data of Example 5 (f_2 62.97)

- 5,000 bootstrap samples (seed 123456).
  - Four methods implemented in boot2BCA for R (Mendyk 2019).

method	90% confidence interval of f_2	
normal approximation	59.88	67.05
basic bootstrap	59.72	66.85
bootstrap percentile	59.08	66.22
bias corrected and accelerated	60.23	68.93

- Four methods not *enough*?
  - Deficiency (SÚKL, Sep 2019):
    - The applicant provided bootstrapped confidence interval […] based on 1,000 and on 5,000 bootstrap samples. In both cases similarity of dissolution profiles was concluded. However, to see if result is robust, the applicant is asked to provide several types of confidence intervals based on … [SÚKL named five]

Is f_2 history?

Q&A document (rearranged and reworded for clarity)

- **Bootstrap** methodology to derive confidence intervals for f_2 **could** actually **be considered the preferred method over** f_2, **even if the validity criteria** outlined in CHMP guidance **are met**.

- Can we expect 'regulatory creep'?
  - *Preferred* easily turns into *mandatory*.
    - Will bootstrapping be required retrospectively?
    - False positive rate of f_2 can be extremely high – very difficult to meet the lower confidence limit for low but still passing f_2.[*]
  - The only way the decrease the *CV* – and hence, the width of the confidence interval – is to substantially increase the number of units.

| CV	units
–	12
¾	21
⅔	27
½	48
⅓	108

[*] Hofman J. *Simulations – bootstrapping, Q and A*. Prague: BioBridges; 27 September 2019.

Problem?

Example 6

- One $\geq 85\%$; $CV \leq 20\%$ at 15 min, $CV \leq 10\%$ at >15 min: validity criteria met.
- Passes f_2.

t (h)	R_t (%) mean	R_t (%) CV	T_t (%) mean	T_t (%) CV
0.25	8.1	19.8	6.7	17.1
0.50	16.3	5.7	13.5	8.8
0.75	23.6	9.1	18.2	7.3
1.0	29.3	6.3	24.6	7.3
1.5	40.9	8.3	33.2	9.3
2.0	51.1	6.6	41.4	8.6
3.0	64.5	8.0	53.8	9.0
4.0	74.7	8.3	61.5	4.3
5.0	80.9	9.2	69.2	9.0
6.0	85.8	7.0	72.2	6.3

Reference
Test
$f1 = 17.0$
$f2 = 52.0$
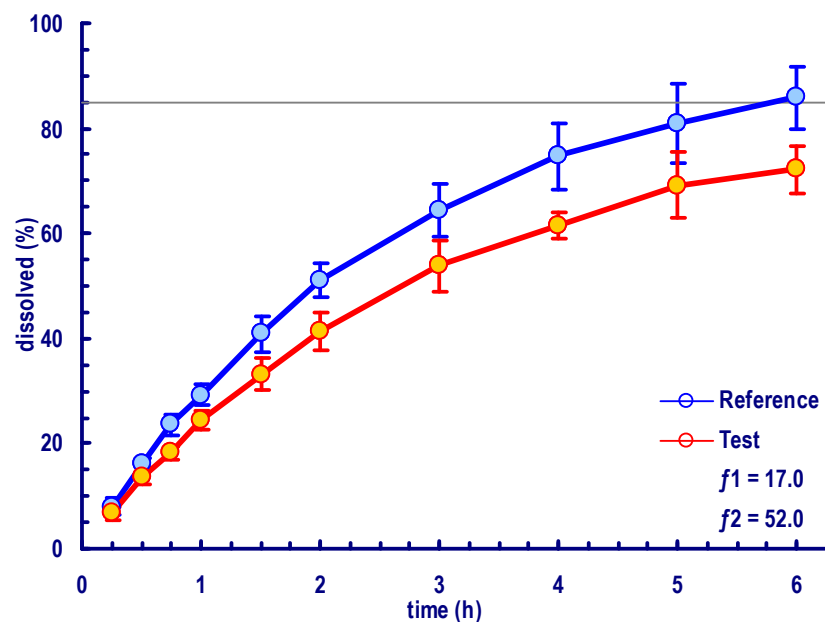
- **What if the bootstrapped confidence interval becomes mandatory?**

Problem!

Data of Example 6 (f_2 51.98)

- 5,000 bootstrap samples.

method	90% confidence interval of f_2	
normal approximation	49.82	54.51
basic bootstrap	49.76	54.43
bootstrap percentile	49.53	54.20
bias corrected and accelerated	49.97	54.74

- **Will such an outcome be accepted (lower confidence limit <50)?**

- **Possibly not…**

  – Bootstrapped confidence interval *preferred over f₂, even if the validity criteria are met*.
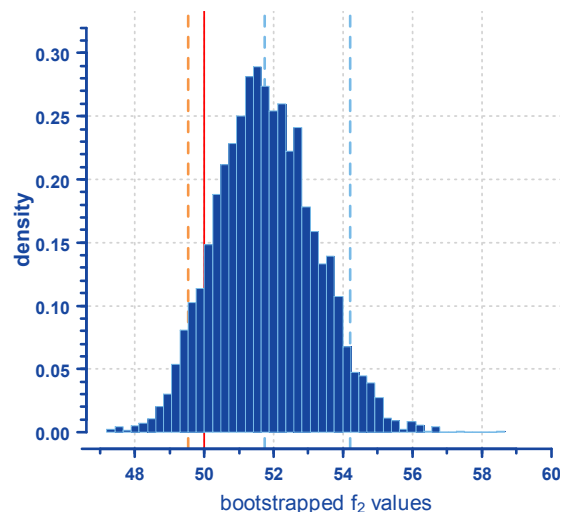
Solution?

We add another 12 units to the 12 we already have → f_2 51.85

- 5,000 bootstrap samples.

| method	90% confidence interval of f_2	
normal approximation	50.58	55.12
basic bootstrap	50.51	55.07
bootstrap percentile	50.50	55.06
bias corrected and accelerated	50.80	55.40

- We are saved but it comes with a price.

- Variability was low – number of units needed to pass the confidence limit might be extreme for high variability and low f_2…
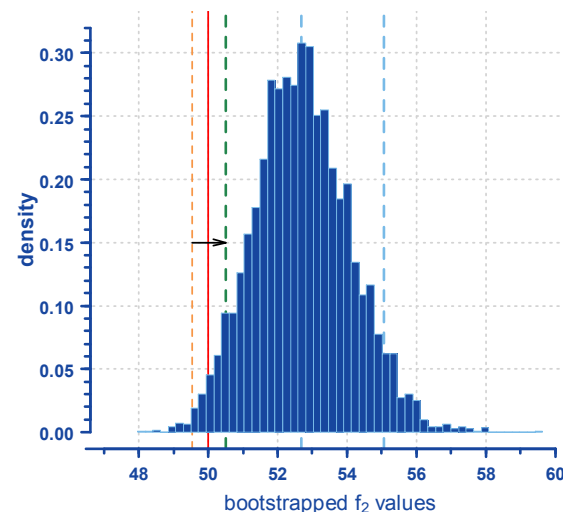
Comparing dissolution to biostudy results

(L)ADME: *In vivo* **profile described by absorption (A) and elimination (metabolization + excretion)**

- **One-compartment model does not have D (distribution).**
  - **Example:** $t_{\frac{1}{2}a}$ **1 h,** $t_{\frac{1}{2}e}$ **8 h**
    - **After** $3 \times t_{\frac{1}{2}a}$ **(3 h) 87.5% are absorbed.**
    - **After** $3 \times t_{\frac{1}{2}e}$ **(24 h) 87.5% are eliminated.**
    - **In the** *in vivo* **profile the inflection point (where the curve changes from concave to convex) is seen at** $2 \times t_{max}$ **(6 h).**
      At this time absorption is essentially complete (98.44%) and the *in vivo* **profile practically represents elimination only.**
- **We can get** *in vivo* **absorption by subtracting the estimated elimination.**
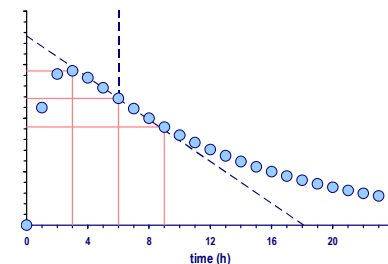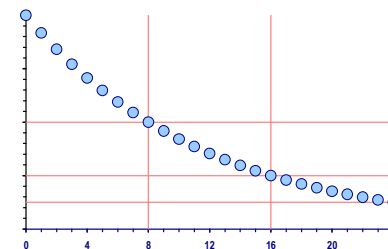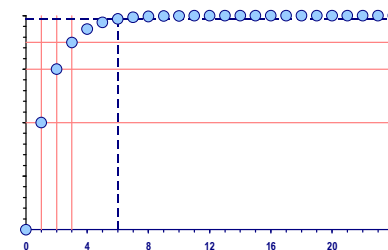
Comparing dissolution to biostudy results

Reconstructing *in vivo* absorption (residual method)

- Fit elimination (λ_z from $2\times t_{max}$ or later to t_z).

- Predict *in vivo* elimination.

- *In vivo* absorption is the *in vivo* curve minus the predicted elimination.

Different other methods exist

- For a one-compartment model.
  - Wagner-Nelson

  $$abs(\%) = 100\,\frac{C_t + k_{el}\cdot AUC_{0-t}}{k_{el}\cdot AUC_{0-\infty}}$$

- For a two-compartment model.
  - Loo-Riegelman (needs true elimination from iv); the distribution phase is reconstructed.
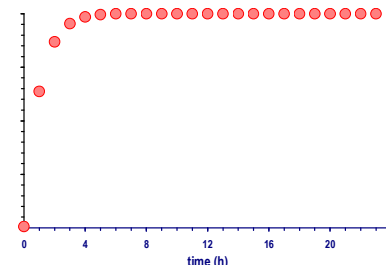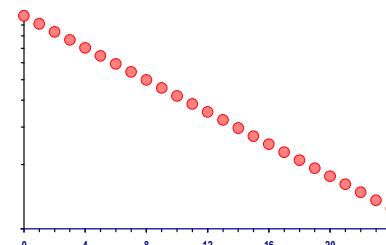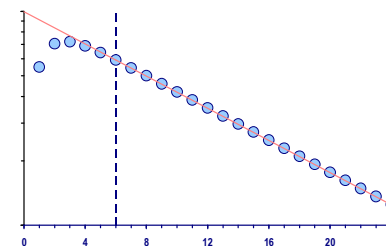
Wagner-Nelson

D 100 mg, V 4 L, F 1, k_a 1 h^{-1} ($t_{1/2}$ 0.69 h), k_{el} 0.25 h^{-1} ($t_{1/2}$ 2.77 h)

- Lin-up/log-down trapezoidal method for AUC_{0-t}.
- λ_z (estimated from 4 to 12 hours) $= 0.2444$.
- $AUC_{0-\infty} = AUC_{0-12} + C_{12} / \lambda_z = 99.68$.
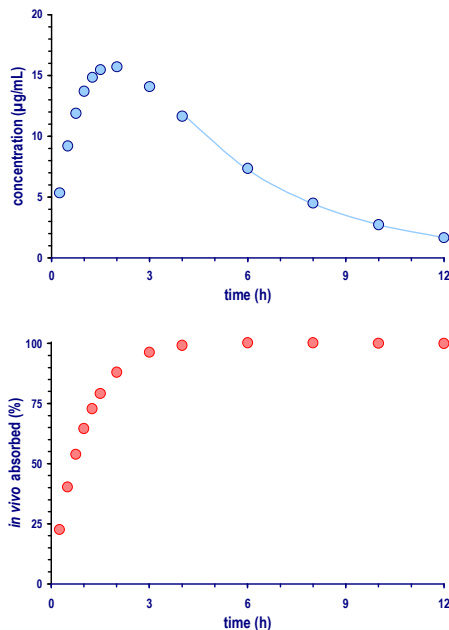
t (h)	C (µg/mL)	AUC_{0-t}	abs (%)
0.00	BQL	–	–
0.25	5.35	0.67	22.63
0.50	9.20	2.49	40.26
0.75	11.89	5.12	53.94
1.00	13.70	8.32	64.58
1.25	14.84	11.89	72.84
1.50	15.47	15.68	79.22
2.00	15.71	23.47	88.03
3.00	14.09	38.36	96.31
4.00	11.65	51.19	99.17
6.00	7.36	69.87	100.31
8.00	4.50	81.50	100.23
10.00	2.73	88.88	100.08
12.00	1.66	92.68	100.00

Outlook: *IVIVC*

Quite often what one thinks to be 'different' (based on a QC dissolution method) turns out to be similar *in vivo*.

- **Develop candidate formulations, perform *in vivo* pilot studies until you see a difference *there*.**
  - **Then (!) develop a discriminatory *in vitro* method (Session 10) which is able to predict *in vivo* absorption**
    - **Try different agitation speeds, use surfactants, change the apparatus, or – as a last resort – explore biorelevant media.**
    - **The final *in vitro* method possibly has nothing in common with the one used in QC.**
      If Earl Grey with a sip of milk is predictive, use it! **(Jean-Michel Cardot)**
- **Once you found a discriminatory method, modify formulations to find one which matches the reference.**
  - **This does not guarantee that the reference will behaves *in vivo* like your best candidate. Another pilot (T *vs.* R) makes sense (to estimate *CV* and *GMR*).**

Fleming.

Thank You!
Open Questions?

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