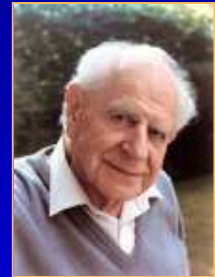# PK–NCA, PK based Design, Biostatistics

Part I

Helmut Schütz
BEBAC

To bear in Remembrance...

Whenever a theory appears to you as the only possible one, take this as a sign that you have neither understood the theory nor the problem which it was intended to solve. *Karl R. Popper*

Even though it's *applied* science we're dealin' with, it still is – *science!*

Leslie Z. Benet

NCA *vs.* PK Modeling

- Pharmacokinetic models
  - Useful for understanding the drug/formulation
    - Study design of BA/BE, *e.g.*, washout, accumulation / saturation to steady state
  - Drawbacks
    - Almost impossible to validate (fine-tuning of side conditions, weighting schemes, software, …)
    - Still a mixture of art and science
    - Impossible to recalculate any given dataset using different software – sometimes even different versions of the same software!
    - Not acceptable for *evaluation* of BA/BE studies!

PK Modeling: AUC

- Based on integration of a PK model;
  e.g., one-compartment open, extravascular
  dose AUC calculation

$$C(t) = \frac{f \cdot D}{V} \frac{k_a}{k_a - k_{el}} \left(e^{k_{el}t} - e^{k_a t} \right)$$

$$AUC_{0-\infty} = \int_0^{\infty} C(t)dt = \frac{f \cdot D}{V} \frac{k_a}{k_a - k_{el}} \left(\frac{1}{k_{el}} - \frac{1}{k_a} \right) = \frac{f \cdot D}{V \cdot k_{el}} = \frac{f \cdot D}{CL}$$

NCA: Single Dose

- Noncompartmental methods do not rely on a pharmacokinetic (=compartmental) model
- Also called SHAM (Shape, Height, Area, Moments)
  - Metrics (plasma, single dose)
    - Extent of absorption (EU…), total exposure (US): AUC (Area Under the Curve)
    - Rate of absorption (EU…), peak exposure (US): C_{max}
    - t_{max} (EU…)
    - Early exposure (US, CAN): AUC_{tmax}; partial AUC truncated at population (CAN: subject's) t_{max} of the reference
    - Others: C_{min}, Fluctuation, MRT, Occupancy time, t_{lag},…

NCA: AUC

- Compartmental models not acceptable in BE, numeric approximation required
  - Linear trapezoidal rule*[)]
  - Lin-log trapezoidal rule*[)]
  - Lin-up log-down trapezoidal rule
  - Cubic splines
 - Lagrange-polynomials
  - Simpson's rule

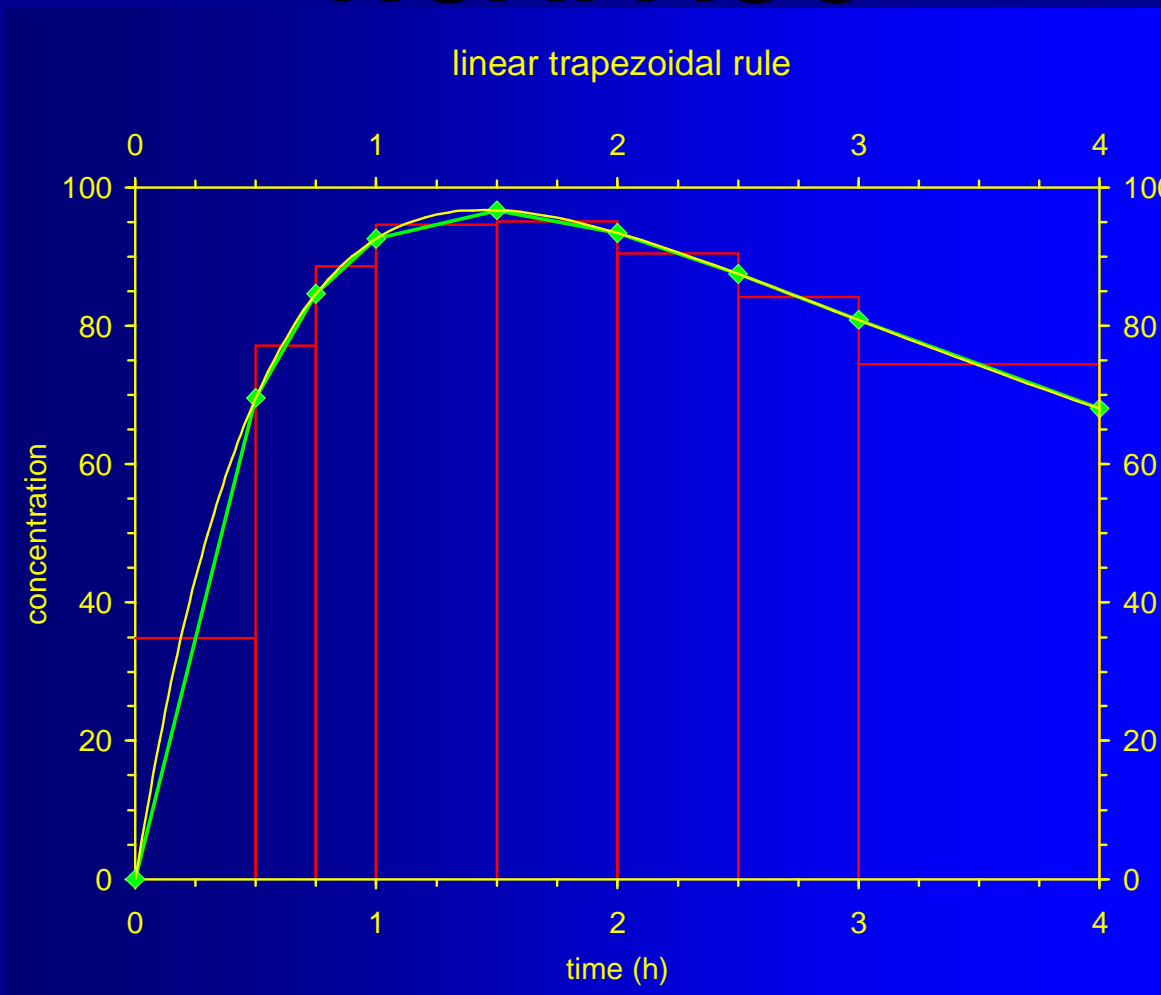 *[)] Stated in Russian GL; only these two acceptable?

NCA: AUC

- Linear trapezoidal rule
  - Linear interpolation between data points
  - Sections represented as trapezoids
  - Sides a, b = neighbouring concentrations
  - h = time interval
  - Area of trapezoid $A = \dfrac{a+b}{2} h$
  - Total

$$AUC_{0-t_n} \approx \sum_{i=1}^{i=n-1} \frac{C_{i+1} + C_i}{2} \left(t_{i+1} - t_i \right) \approx \frac{1}{2} \sum_{i=1}^{i=n-1} \left(t_{i+1} - t_i \right) \cdot \left(C_{i+1} + C_i \right)$$

NCA: AUC

linear trapezoidal rule

# NCA: AUC
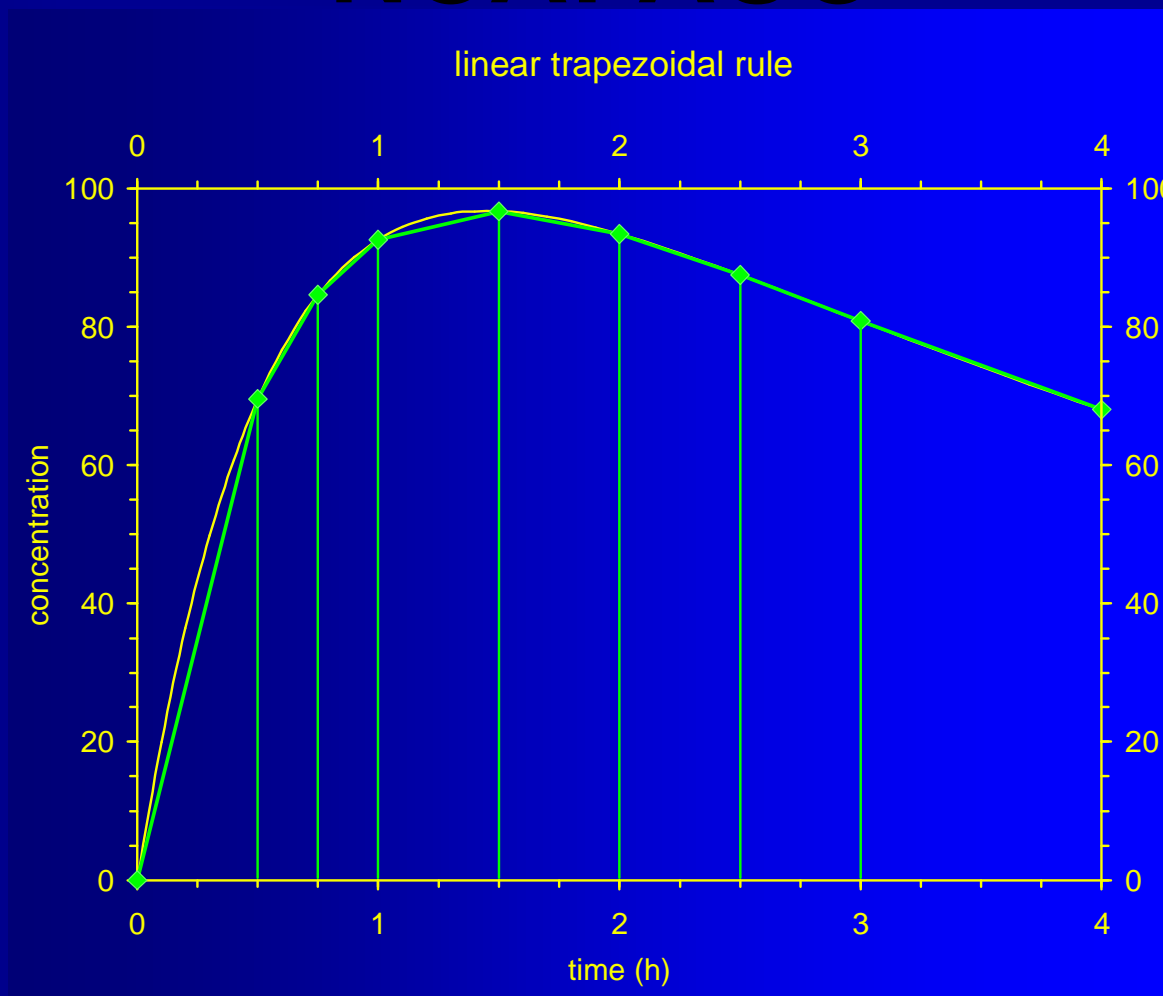
linear trapezoidal rule

NCA: AUC

● Log-linear trapezoidal rule

- ■ Assumes exponential elimination
- ■ Log-linear interpolation between data points
- ■ Only valid for iv administration; sections in absorption phase underestimated if applied to ev
- ■ If $C = 0$ or subsequent concentrations are equal, section calculated by linear trapezoidal
- ■ Total

$$AUC_{0-t_n} \approx \sum_{i=1}^{i=n-1} \left(t_{i+1} - t_i \right) \frac{C_{i+1} - C_i}{\ln \dfrac{C_{i+1}}{C_i}}$$

NCA: AUC

● Lin-up log-down trapezoidal rule

■ Hybrid of linear and log-linear

■ Sections with *increasing or equal* concentrations ($C_{i+1} \geq C_i$) calculated by linear trapezoidal rule

■ Sections with *decreasing* concentrations ($C_{i+1} < C_i$) calculated by log-linear trapezoidal rule

■ Avoids bias in both absorption and elimination phases

■ Suitable for iv and ev

■ Suitable for multiphasic profiles

Lin. *vs.* log-lin. interpolation

PK and approximation methods (absorption phase)

Lin. *vs.* log-lin. interpolation

PK and approximation methods (elimination phase)
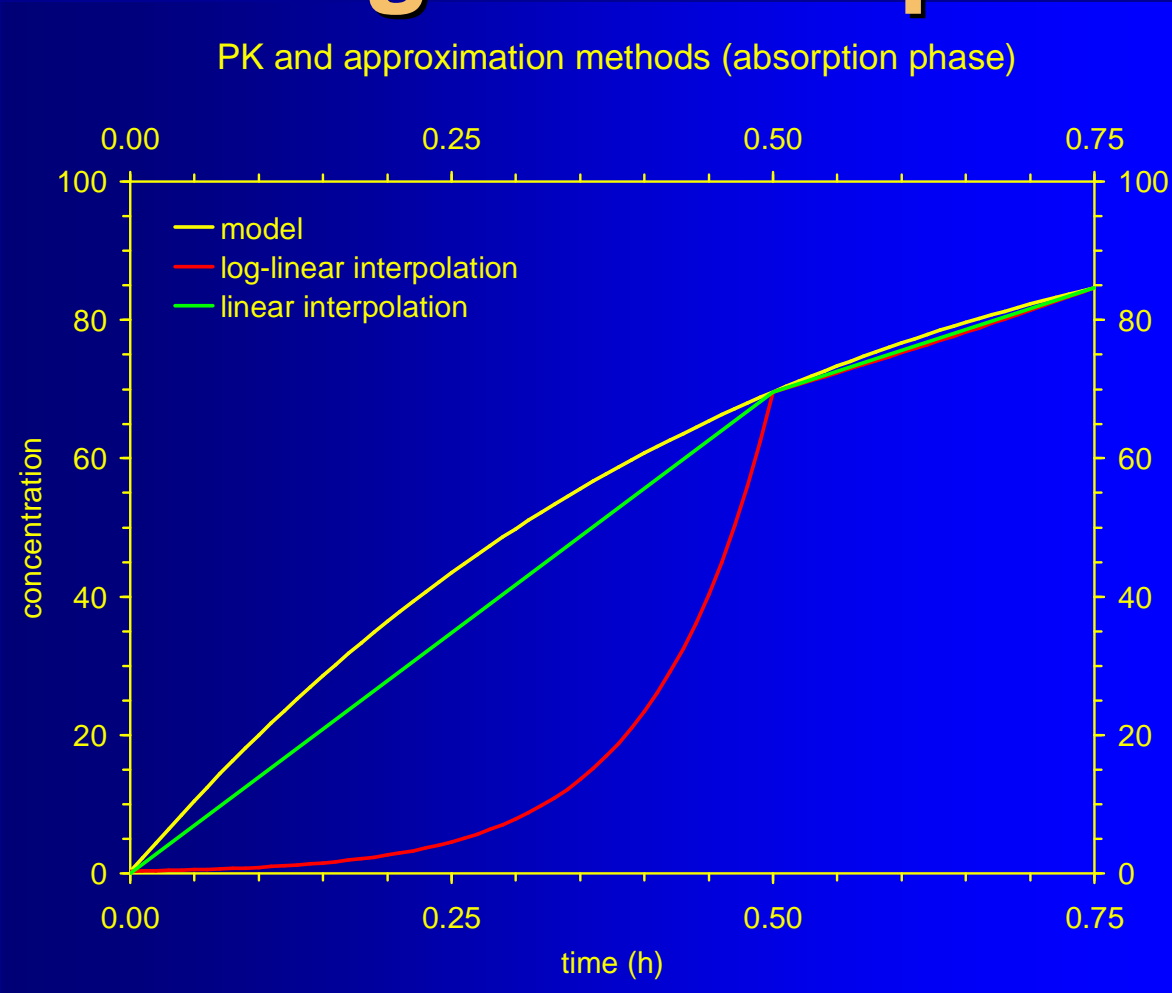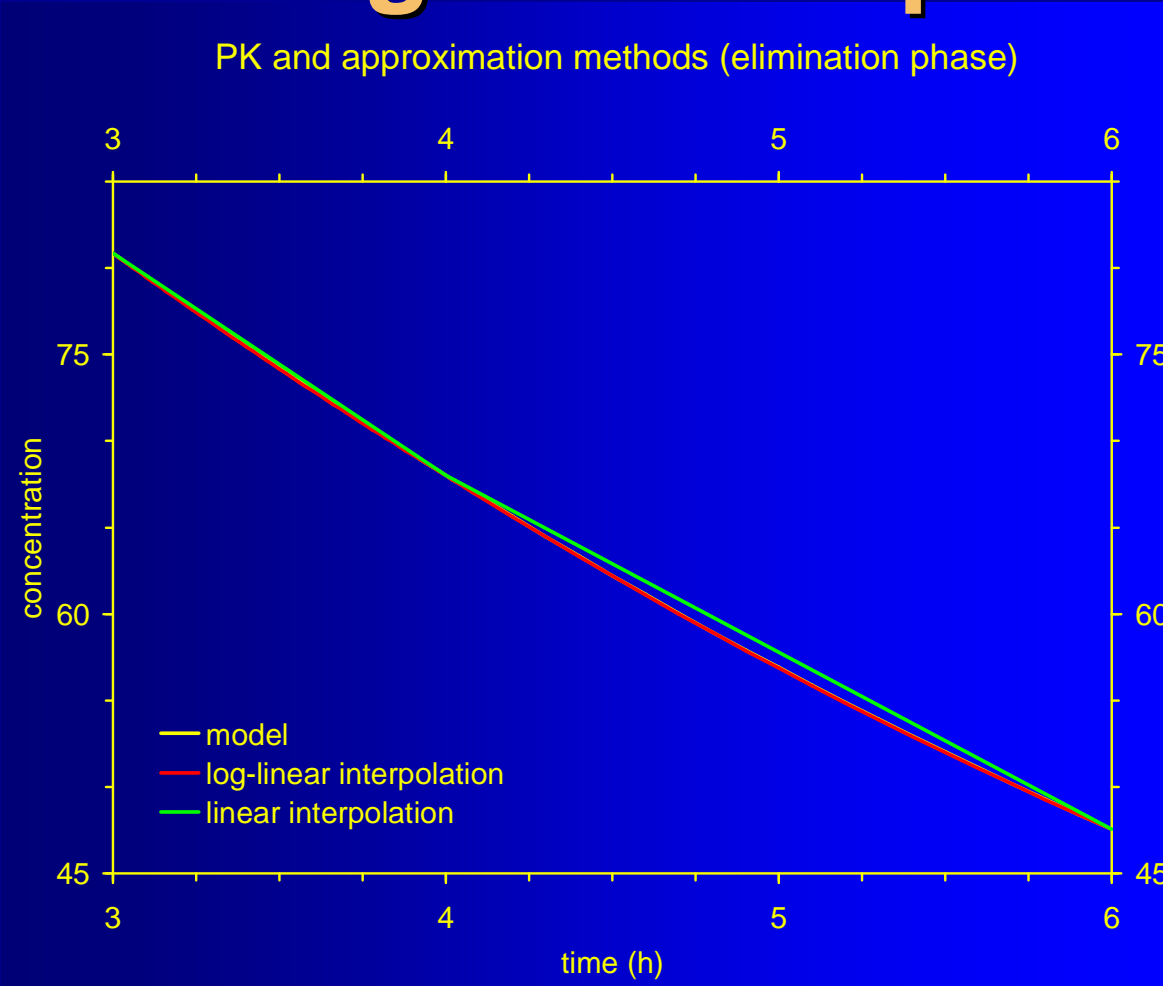
Example 1

AUCi (R) 707.6, AUCi (T) 670.9, T/R 94.8%, bias -0.20%

Model

AUC_R 697.8

AUC_T 662.9

T/R   95.00%

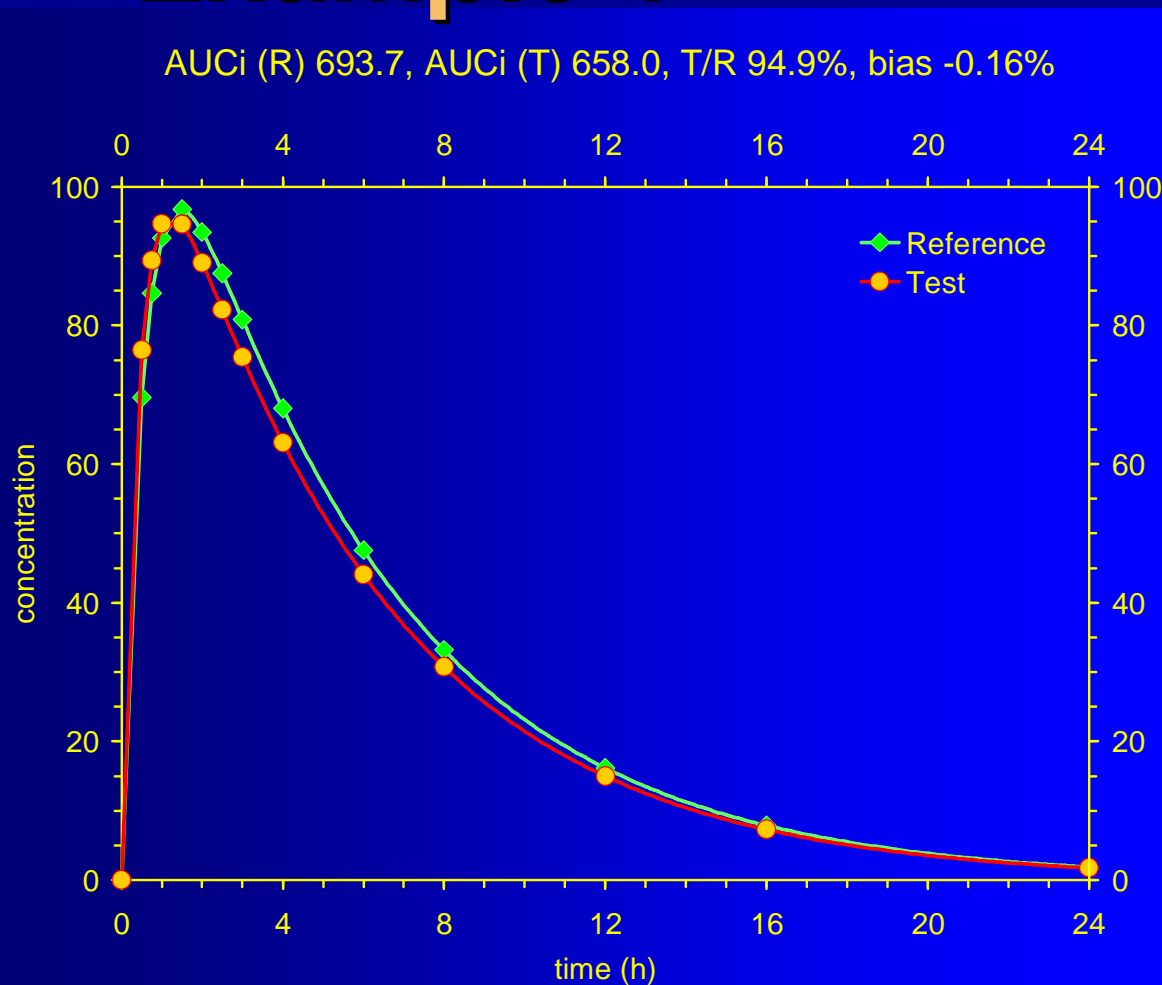linear trapezoidal

T/R   94.85%

Example 1

AUCi (R) 693.7, AUCi (T) 658.0, T/R 94.9%, bias -0.16%
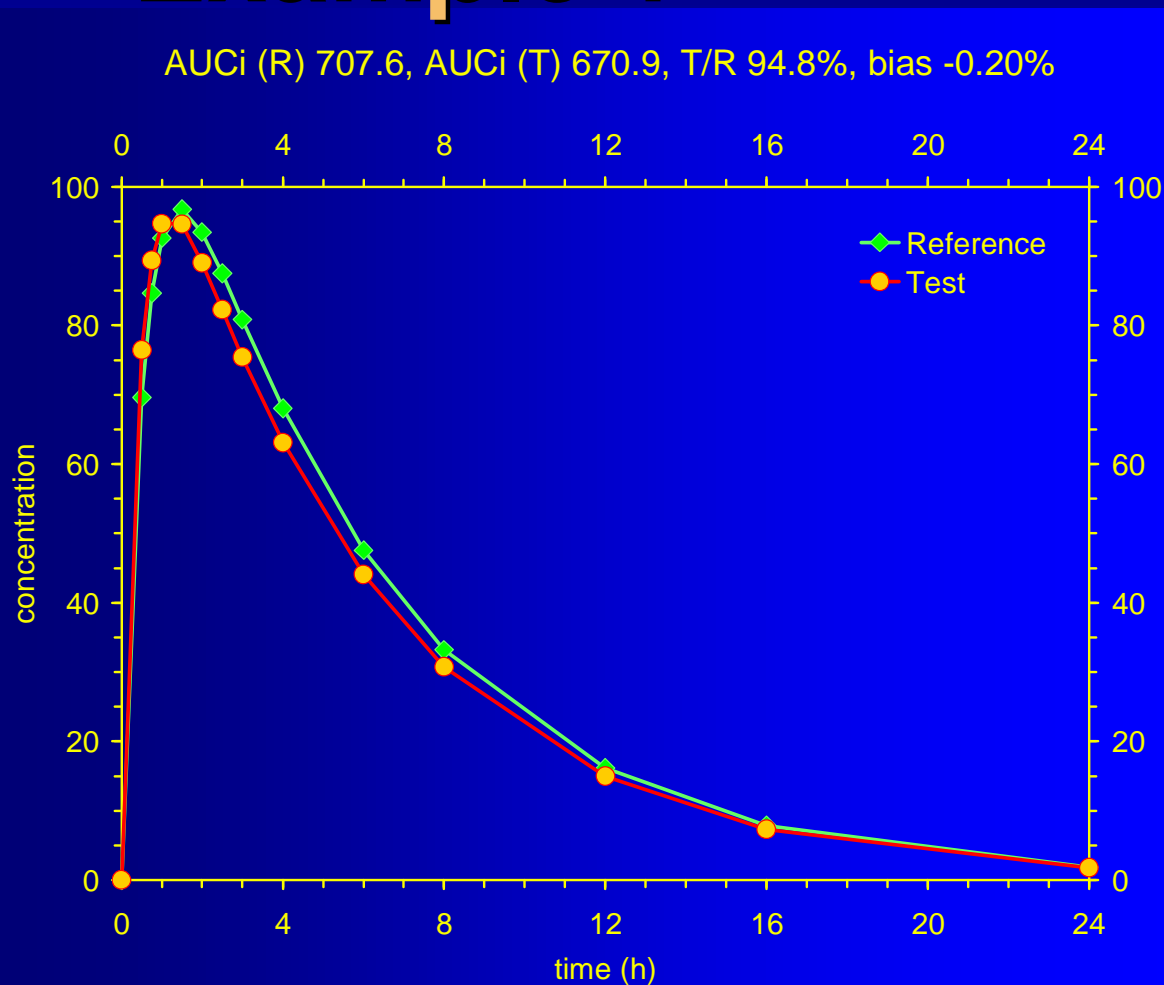
Model
AUC_R 697.8
AUC_T 662.9
T/R   95.00%

lin-up log-down
T/R   94.89%

Example 2

AUCi (R) 725.1, AUCi (T) 670.9, T/R 92.5%, bias -2.60%

Model

AUC_R 697.8
AUC_T 662.9
T/R   95.00%

linear trapezoidal

12 h (R) missing
T/R   92.53%

Example 2

AUCi (R) 693.7, AUCi (T) 658.0, T/R 94.9%, bias -0.15%
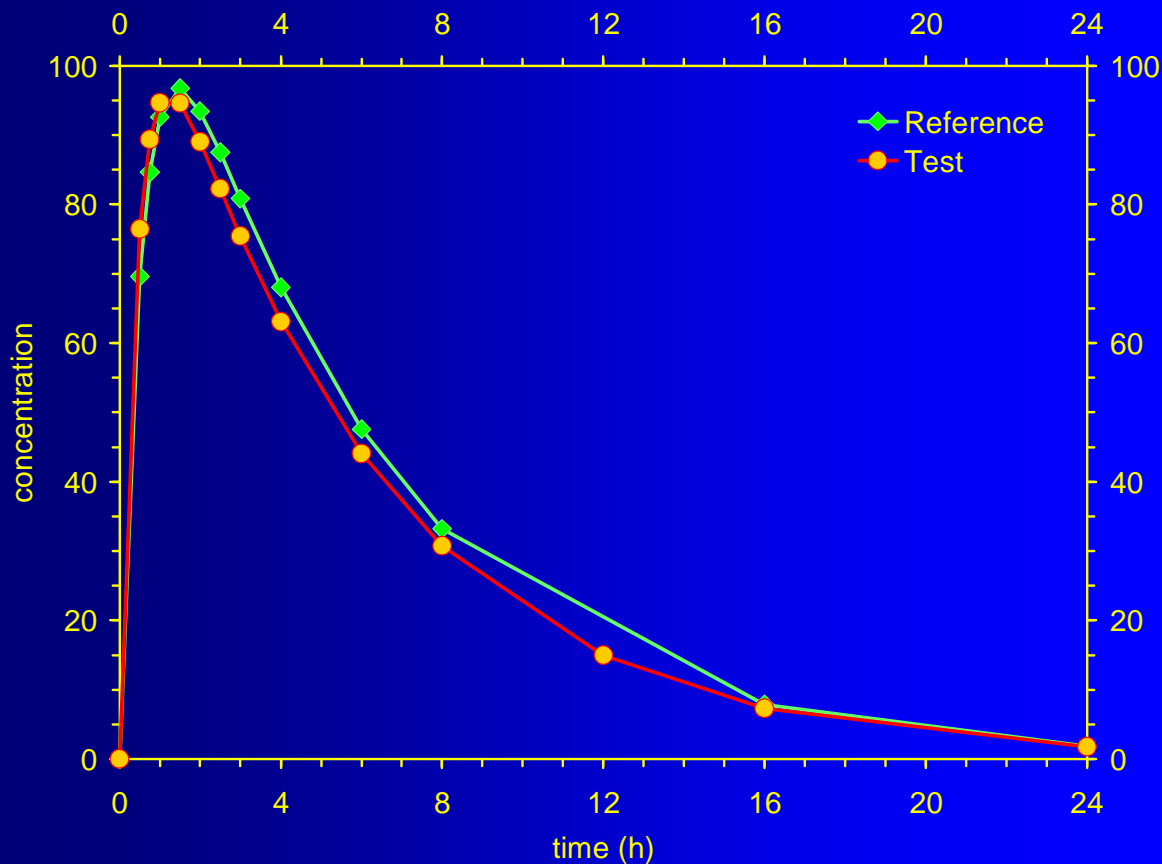
Model

AUC_R 697.8
AUC_T 662.9
T/R 95.00%

lin-up log-down

12 h (R) missing
T/R 94.89%
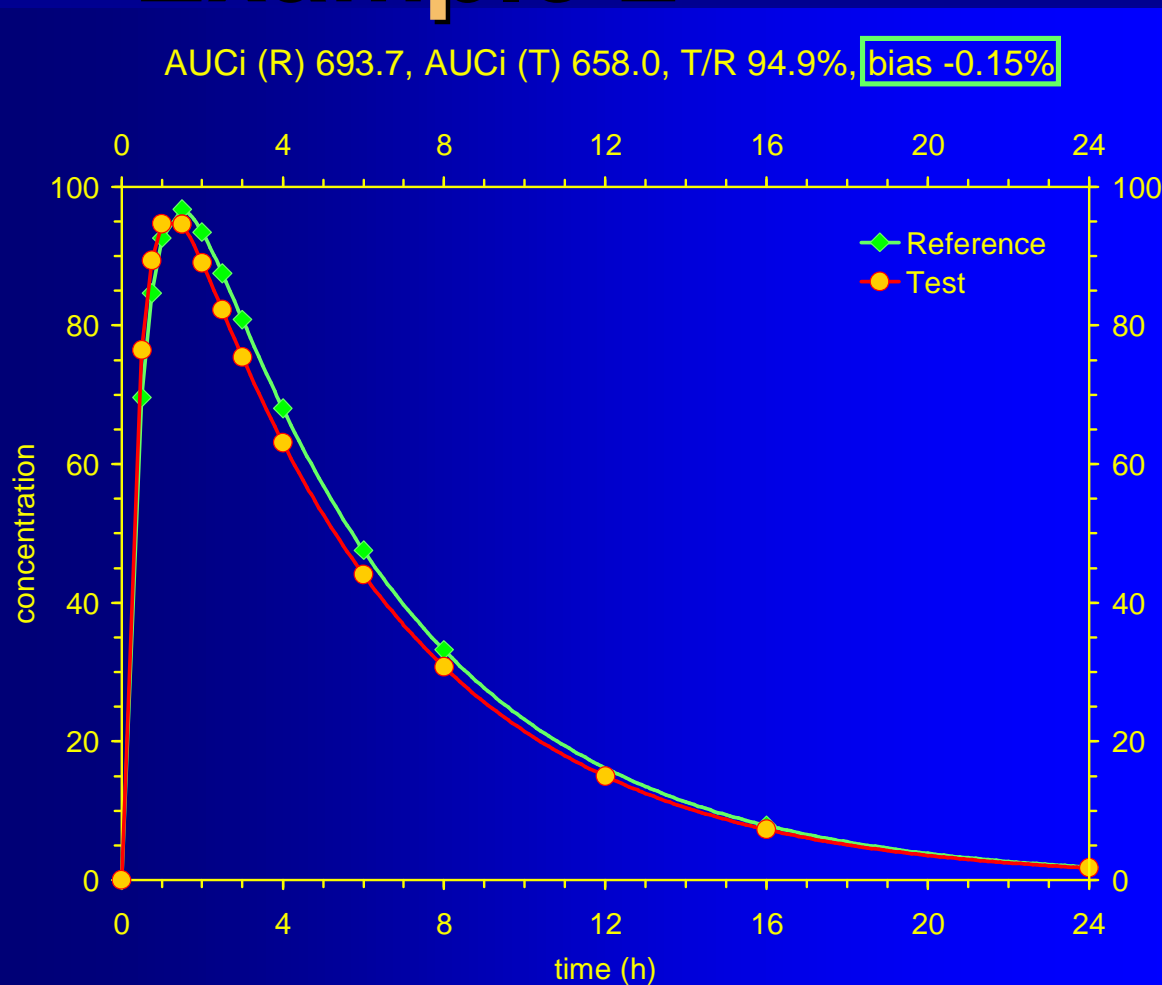
Recommendations

- Don't exclude a subject if only a few data points are missing (loss of power)
  - *Only* if linear rule is required for any reason: data imputation
    - Linear within increasing/equal values ($C_{i+1} \geq C_{i-1}$)

$$\hat{C}_i = C_{i-1} + \frac{C_{i+1} - C_{i-1}}{t_{i+1} - t_{i-1}}\left(t_i - t_{i-1}\right)$$

    - Log-linear within decreasing values ($C_{i+1} < C_{i-1}$)

$$\hat{C}_i = e^{\ln C_{i-1} - \frac{t_i - t_{i-1}}{t_{i+1} - t_{i-1}}\left(\ln C_{i-1} - \ln C_{i+1}\right)}$$

Recommendations

- Don't exclude a subject … (cont'd)
  - Although I had never problems with this procedure in 500+ BE studies (stated in the protocol, according to SOP, and by validated software) data imputation may be unfamiliar to assessors
  - Lin-up log-down trapezoidal 'automatically' corrects for missing values and unbiased estimates are obtained

NCA: AUC Extrapolation

- $AUC_{0-\infty}$
  - EMA (and all countires except US and Russia): No primary PK metric; but demonstrate that AUC_{0-t} is a reliable estimate of extent of absorption (*i.e.*, extrapolated area $\leq 20\%$ of $AUC_{0-\infty}$)
    - FDA: Primary PK metric (*additionally* to AUC_{0-t})
    - What if extrapolated $AUC_{0-t} > 20\%$ of $AUC_{0-\infty}$ in some subjects?
      - Russia: Use $AUC_{0-\infty}$ *instead* of AUC_{0-t} as primary metric of the study
      - Others: State a procedure in the protocol! *Either* exclude the subject *or* switch to $AUC_{0-\infty}$

NCA: AUC Extrapolation

● $AUC_{0-\infty}$

- Unweighted log-linear regression of at least three data points in the elimination phase
- Extrapolation from AUC_{0-t} (regardless the method)

$$AUC_{\infty} = AUC_t + \frac{C_t}{\hat{\lambda}_z} \quad \text{or better} \quad AUC_{\infty} = AUC_t + \frac{\hat{C}_t}{\hat{\lambda}_z}$$

Russia: Only first method stated in GL; mandatory?

NCA: AUC Extrapolation

- Single dose only!
  - Method of estimation of λ_z stated in protocol!
    - One-compartment model: TTT-method [*)]
      (Two times t_{max} to t_z)
    - Maximum adjusted R^2 (Phoenix/WinNonlin, Kinetica)

      $$R^2_{adj} = 1 - \frac{(1 - R^2) \cdot (n - 1)}{n - 2}$$

 | WinNonlin $\leq$5.3: | C_{max} **in**cluded |
      | Phoenix/WNL $\geq$6.0: | C_{max} **ex**cluded |

    - Multi-compartment models: starting point = last inflection
    - Minimum AIC: $AIC = n \cdot \left[\ln(2 \cdot \pi) + 1\right] + n \cdot \ln(RSS/n) + 2 \cdot p$
    - Visual inspection of fit mandatory!

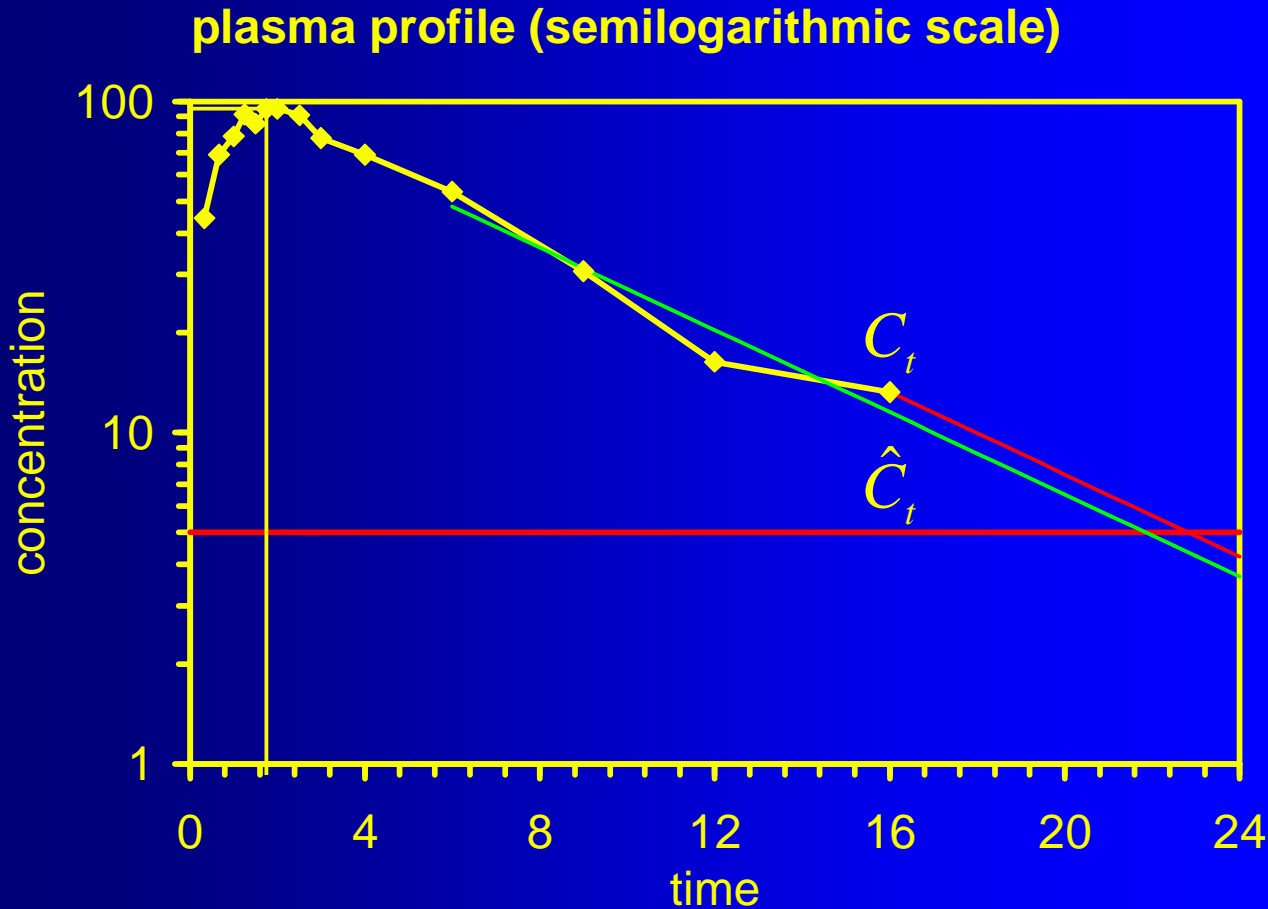   [*)] **Scheerans C, Derendorf H and C Kloft**
    Proposal for a Standardised Identification of the Mono-Exponential Terminal Phase
    for Orally Administered Drugs
    Biopharm Drug Dispos 29, 145–57 (2008)

NCA: AUC Extrapolation

plasma profile (linear scale)

NCA: AUC Extrapolation

plasma profile (semilogarithmic scale)
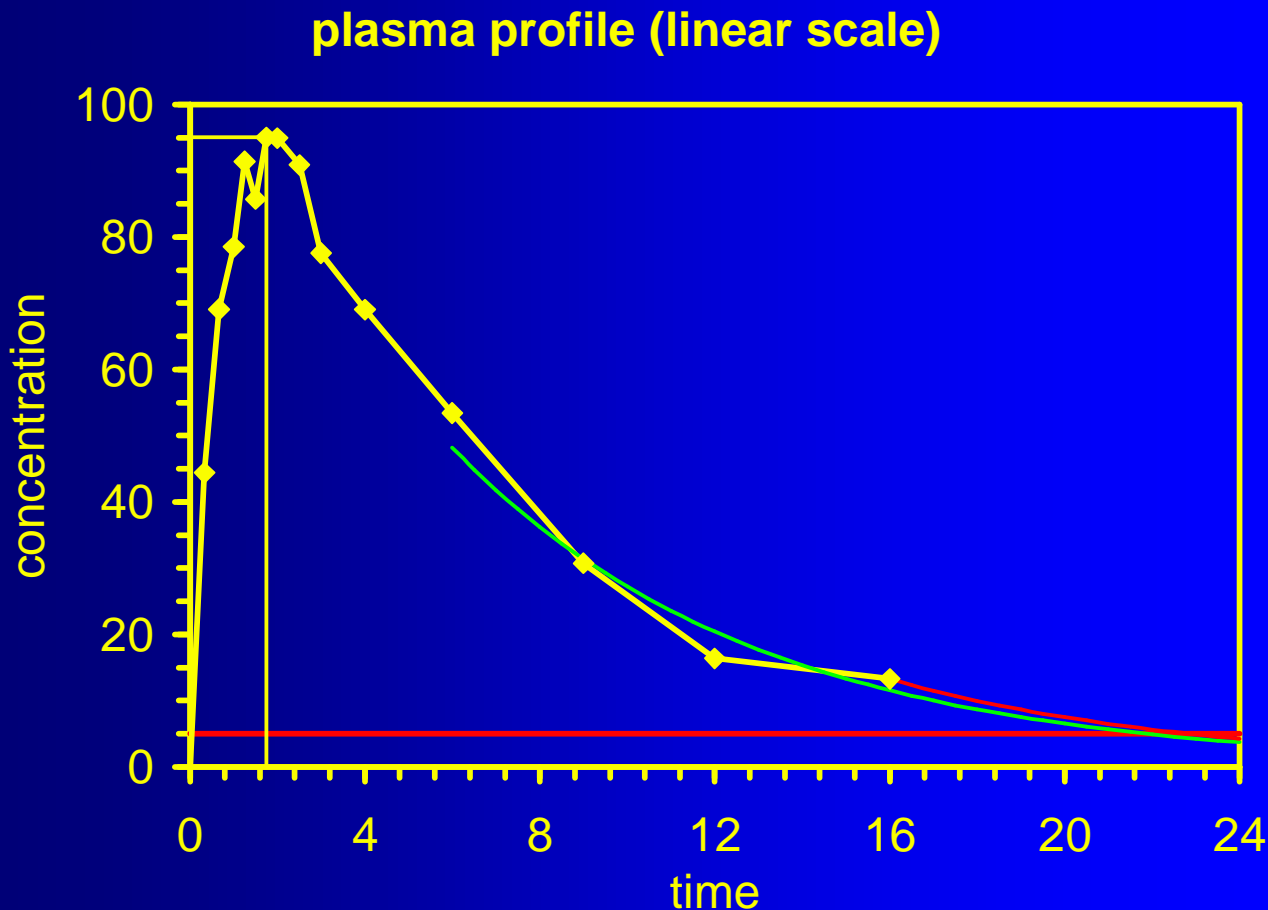
NCA: other PK Metrics

- Single dose
  - C_{max} and t_{max} directly from profile
  - Metrics describing the shape of the profile
    - Early exposure (US, CAN): AUC_{tmax}; partial AUC truncated at population (CAN: subject's) t_{max} of the reference
    - Biphasic MR formulations: Partial $AUCs$ truncated at prespecific cut-off time point
      - FDA: Product specific guidances (methylphenidate, zolpidem)
      - EMA: All products
        Questions & Answers: positions on specific questions addressed to the pharmacokinetics working party
        EMA/618604/2008 Rev. 4 (16 February 2012)
        http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002963.pdf

NCA: other PK Metrics

- Single dose
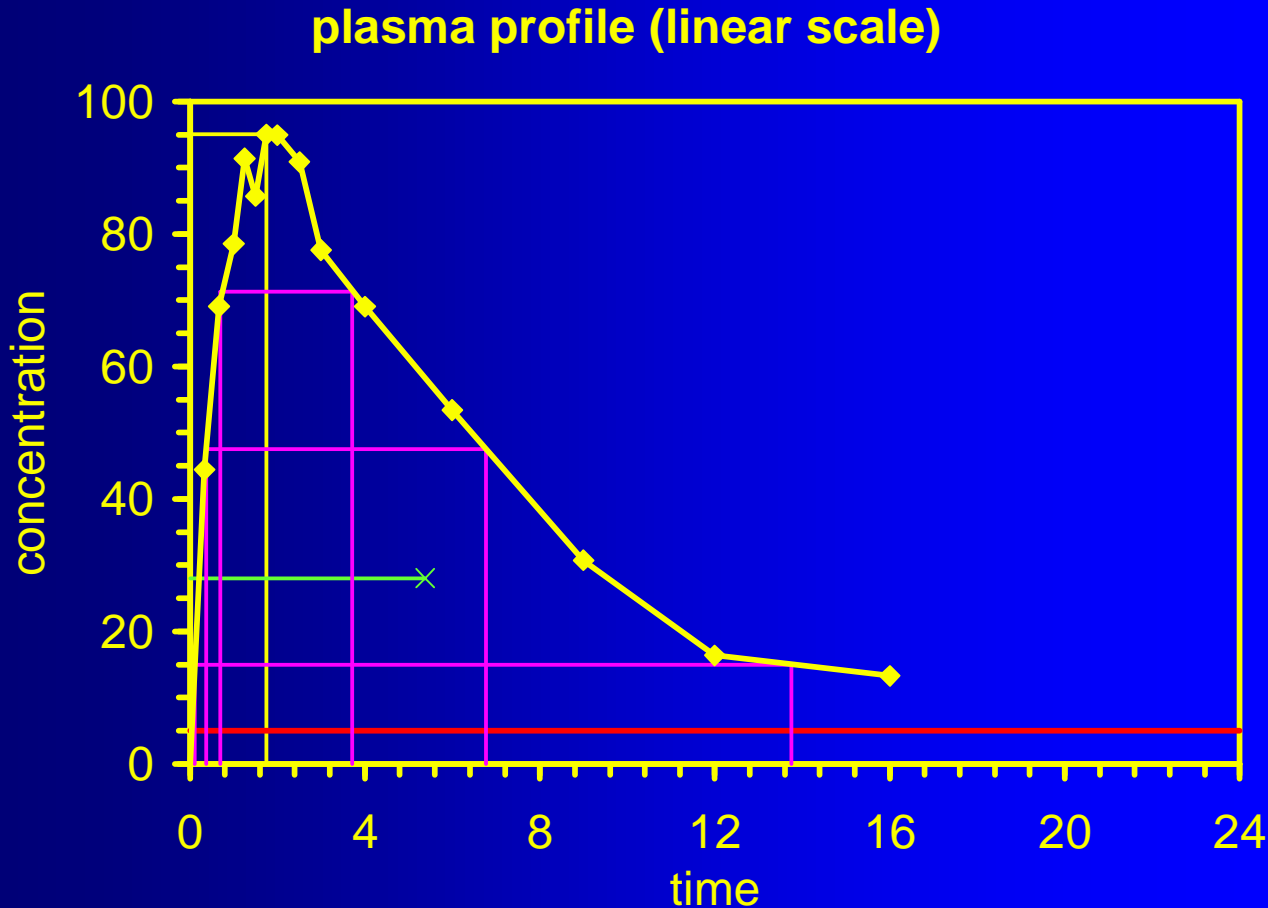 - Metrics describing the shape of the profile
    - C_{max}/AUC
    - $t_{75\%}$ (Plateau time: interval where $C(t) \geq 75\%$ of C_{max})*)
    - HVD (Half value duration: time interval where $C(t) \geq 50\%$ of C_{max})
    - Occupancy time, $t \geq MIC$ (time interval where $C(t)$ is above some limiting concentration)

    *) Russia: mandatory for sustained release formulations

NCA: Overview Single Dose

plasma profile (linear scale)

NCA: Urine

- Noncompartmental methods (cont'd)
  - Extent of absorption (EU…), total exposure (US): Ae_t (cumulative amount excreted); rarely extrapolated to $t = \infty$
  - Rate of absorption, peak exposure (US): ΔAe_{max}, $t\Delta Ae_{max}$
  - EU: C_{max}, t_{max} from plasma!
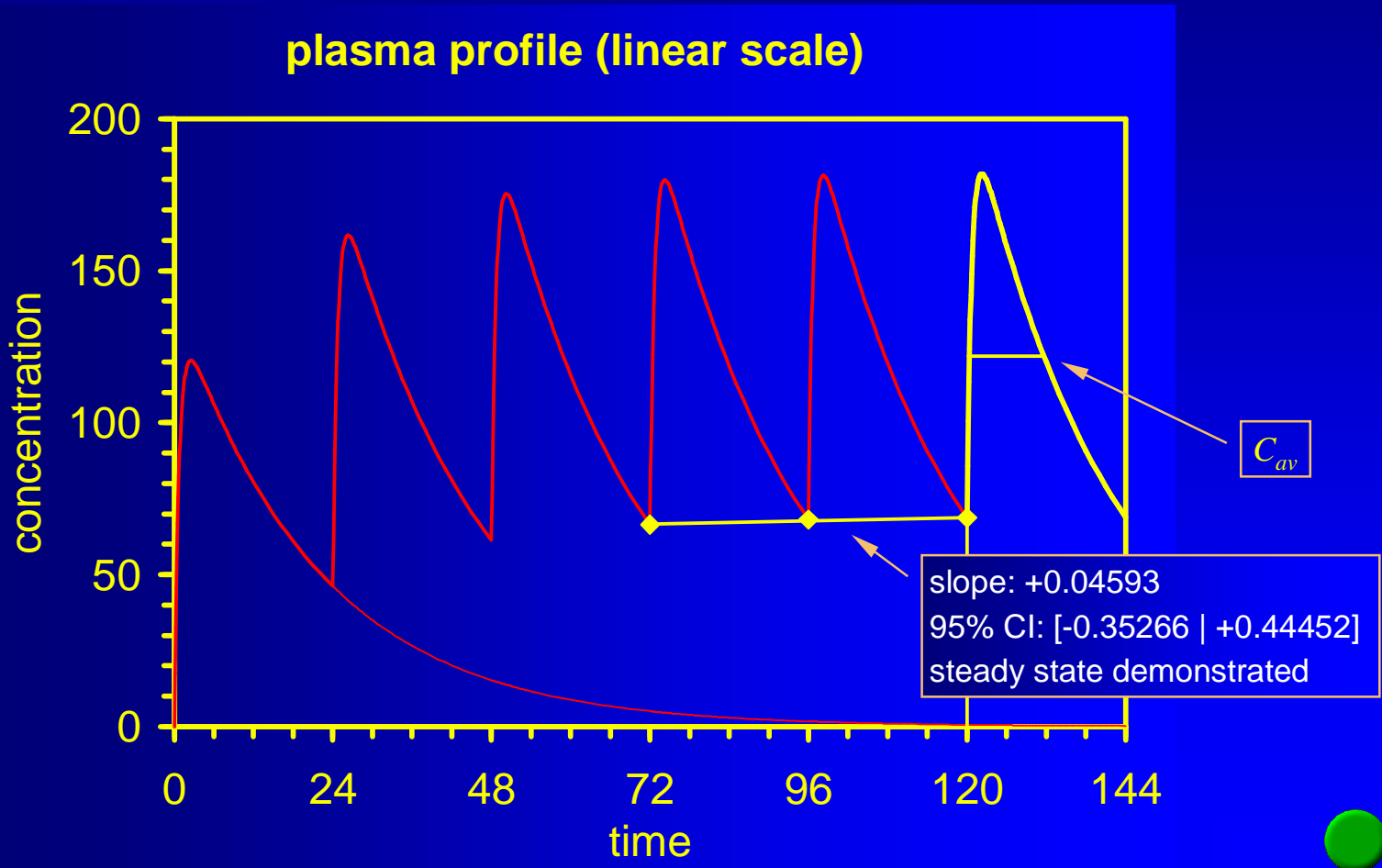
NCA (Methods)

- Multiple dose
  - Calculation of AUC_τ (dosage interval τ); $AUC_{ss,24h}$ if more than *o.a.d.* and chronopharmaco-logical variation)
  - No extrapolation!
  - $C_{ss,max}$ / $C_{ss,min}$ directly from profile
  - Peak-Trough-Fluctuation: $(C_{ss,max} - C_{ss,min}) / C_{ss,av}$, where $C_{ss,av} = AUC_\tau / \tau$
  - Swing: $(C_{ss,max} - C_{ss,min}) / C_{ss,min}$

NCA (Methods)

- ●Multiple dose
  - ■Assessment whether steady state is reached (in a linear PK system: $AUC_\tau = AUC_\infty$)
    - ■ No recommendations in GLs (except EU/US Veterinary)
    - ■ Not required according to comments to EMA BE-GL
    - ■ MANOVA-model (sometimes mentioned in Canada, rarely used)
    - ■ t-test of last two pre-dose concentrations
    - ■ Hotelling's T^2
    - ■ Linear regression of last three pre-dose concentrations, individually for each subject/treatment
  - ■ Only the last method allows the exclusion of subjects being not in stead state. Other methods give only a yes|no result!

NCA (Methods)

plasma profile (linear scale)

C_{av}

slope: +0.04593
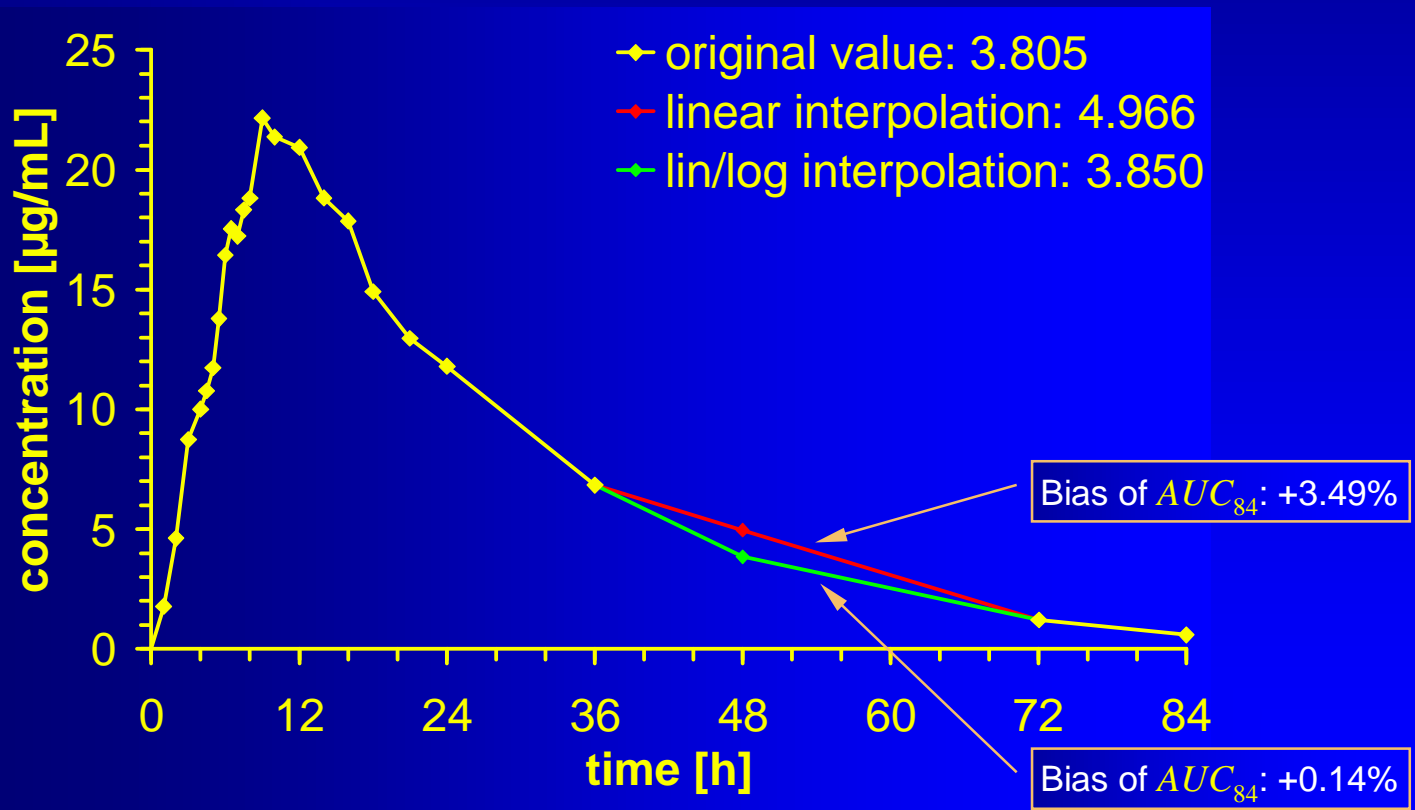95% CI: [-0.35266 | +0.44452]
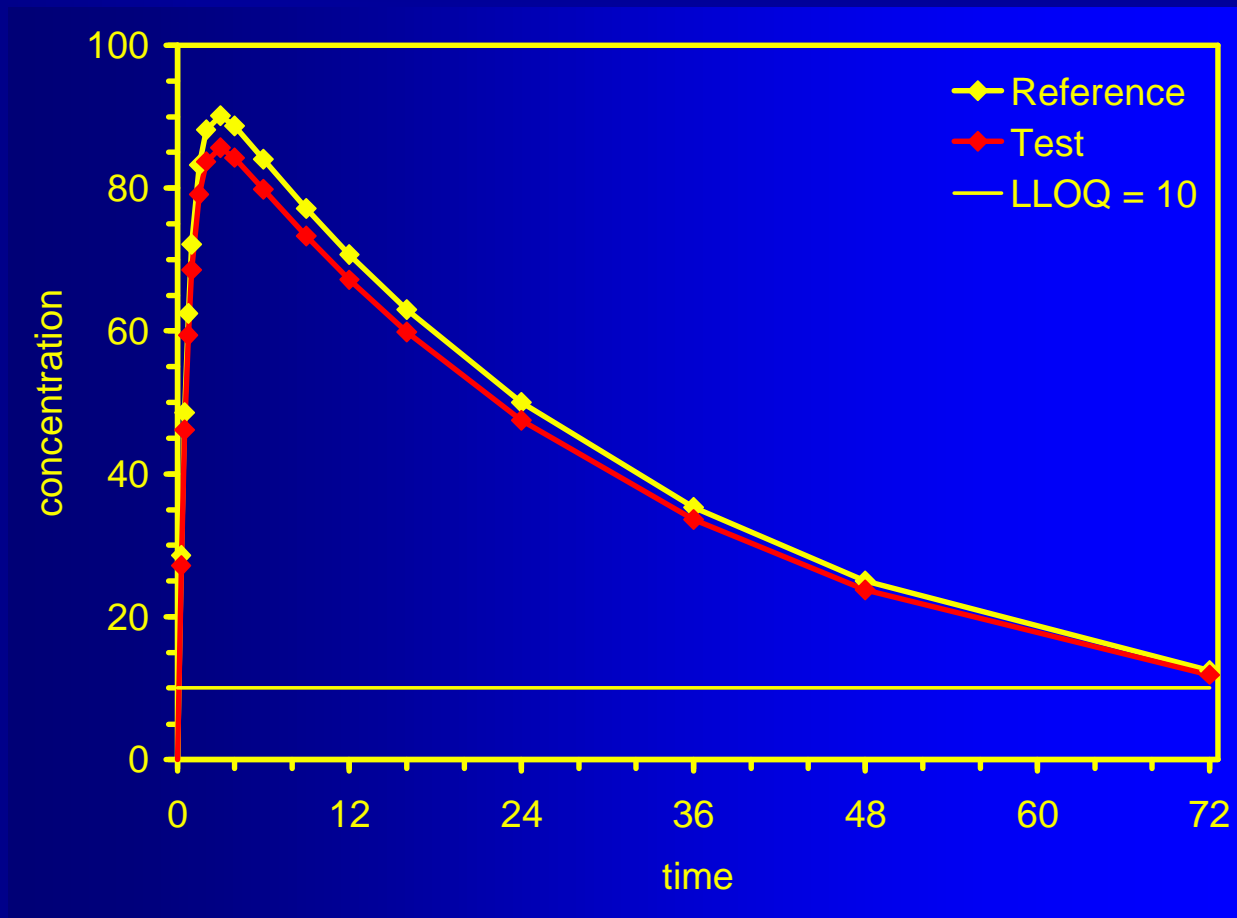steady state demonstrated

Some Problems…

- Missing values I
  - Procedure for Imputation must be stated in the Protocol; recommended:
    - in the Absorption Phase ($t < t_{max}$) by linear Interpolation of two adjacent values
    - in the Elimination Phase ($t \geq t_{max}$) by log/linear Interpolation of two adjacent values
    - estimated value must not be used in calculation of the apparent half life!
  - Don't rely on softwares' defaults!
    - Phoenix/WinNonlin interpolates linear – unless lin-up/log-down trapezoidal method is used
    - Kinetica interpolates log/lin within descending values

Some Problems…
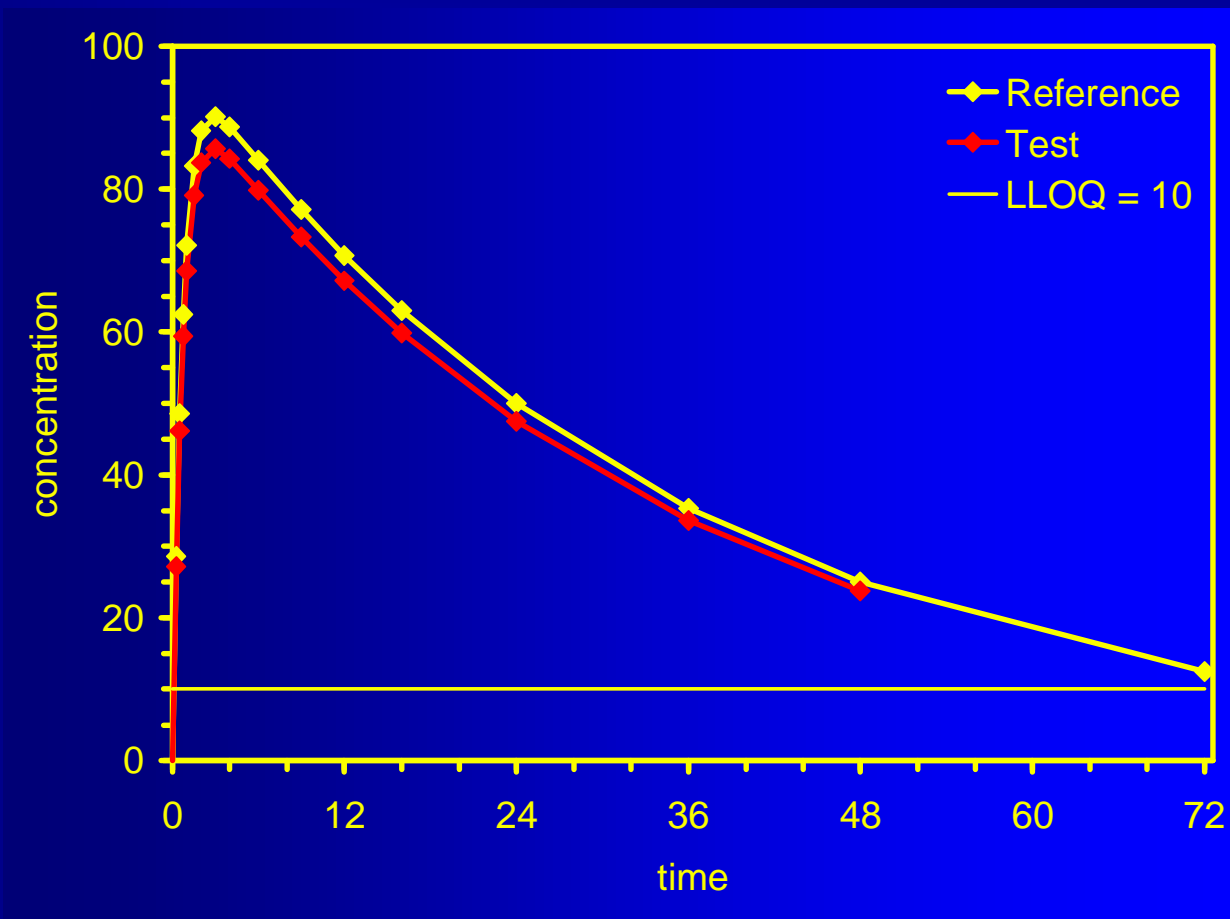
● Missing values I

original value: 3.805
linear interpolation: 4.966
lin/log interpolation: 3.850

Bias of AUC_{84}: +3.49%

Bias of AUC_{84}: +0.14%

concentration [µg/mL]
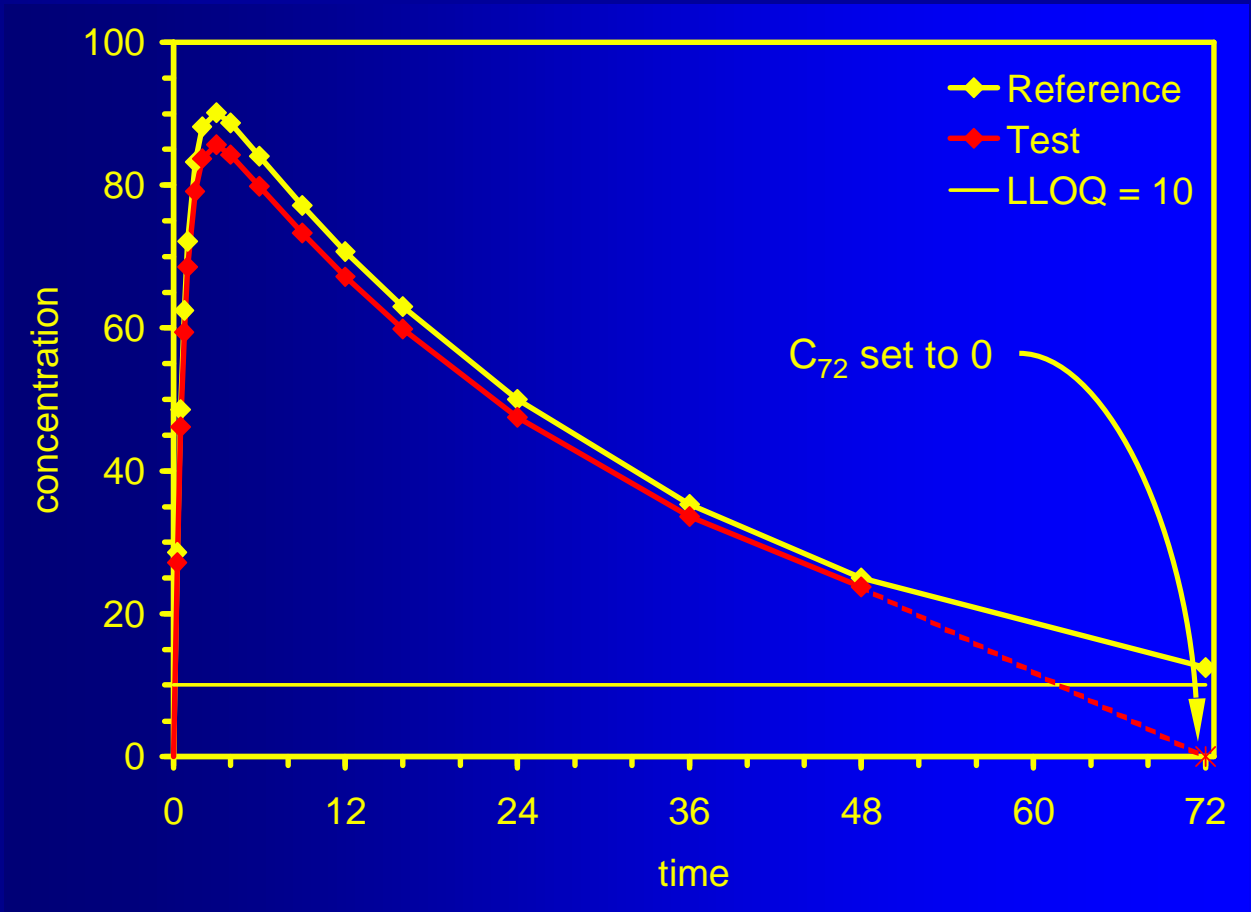
time [h]

Some Problems…

Some Problems…

Some Problems…

● Missing values II

■ Last value of T missing (*e.g.*, vial broken)

■ AUC_{tlast} (48) T = 2407
AUC_{tlast} (72) R = 2984
  T/R = 80.67% biased!

■ Using AUC to t where $C \geq$ LLOQ for both formulations (48)
AUC_{48} T = 2534
AUC_{48} R = 2407
  T/R = 95% ✓

➢ Not available in software

➢ Regulatory acceptance?

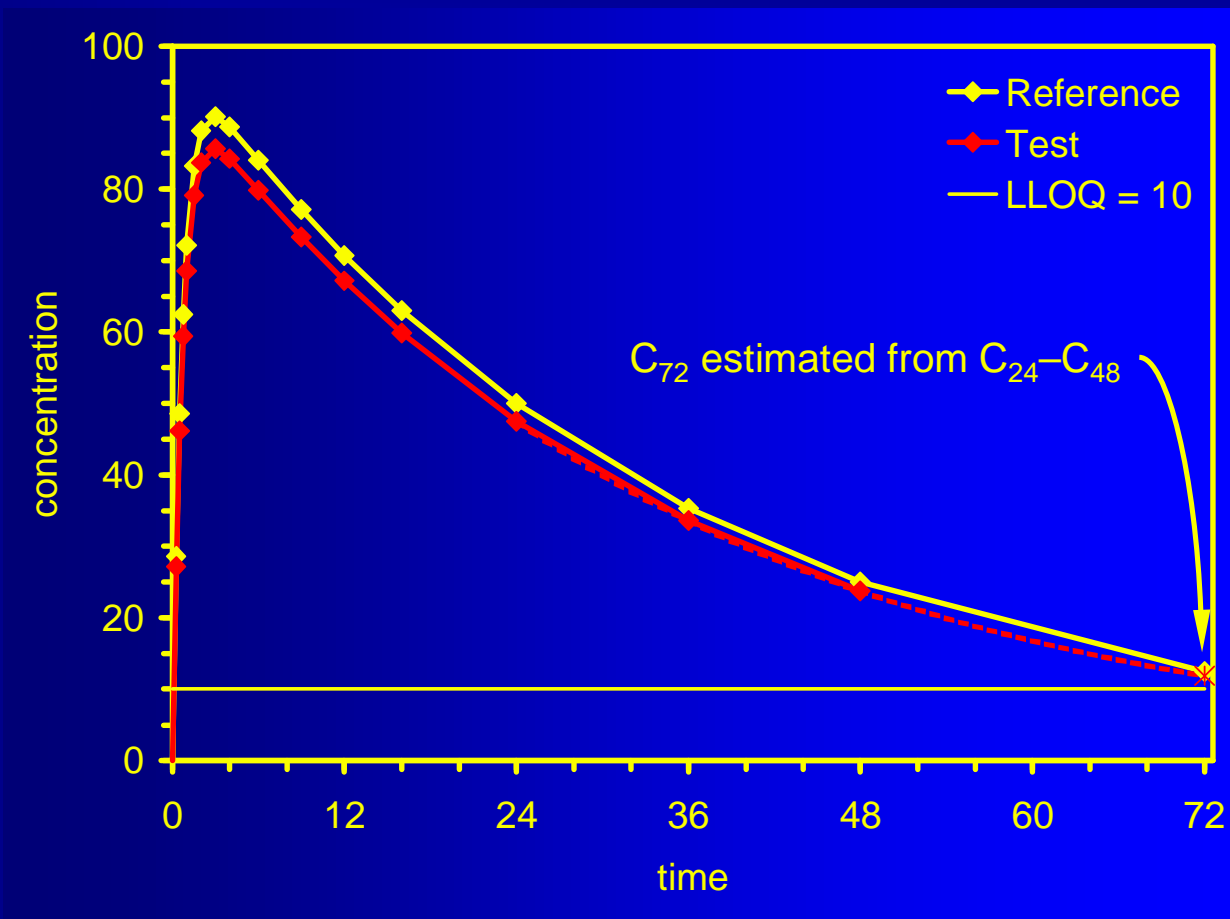| time	Reference conc	Reference AUC_{0-t}	Test conc	Test AUC_{0-t}
0	BLQ	0	BLQ	0
0.25	28.57	4	27.14	3
0.50	48.57	13	46.14	13
0.75	62.50	27	59.38	26
1.00	72.15	44	68.55	42
1.5	83.26	83	79.10	79
2	88.14	126	83.73	119
3	90.14	215	85.63	204
4	88.70	304	84.26	289
6	84.07	477	79.86	453
9	77.11	719	73.25	683
12	70.71	940	67.18	893
16	63.00	1208	59.85	1147
24	50.00	1660	47.50	1577
36	35.36	2172	33.59	2063
48	25.00	**2534**	23.75	**2407**
72	12.50	**2984**	Missing	NA

Some Problems…

Some Problems…

- Missing values II
  - Last value of T missing (*e.g.*, vial broken)
    - Setting the first concentration in the profile where C<LLOQ to zero. AUC_{all}, 'invented' by Pharsight
      AUC_{all} (72) T = 2692
      AUC_{all} (72) R = 2984
      T/R = 90.22% biased!
      - Available in Phoenix / WinNonlin, Kinetica
      - Regulatory acceptance?

time	Reference conc	Reference AUC_{0-t}	Test conc	Test AUC_{0-t}
0	BLQ	0	BLQ	0
0.25	28.57	4	27.14	3
0.50	48.57	13	46.14	13
0.75	62.50	27	59.38	26
1.00	72.15	44	68.55	42
1.5	83.26	83	79.10	79
2	88.14	126	83.73	119
3	90.14	215	85.63	204
4	88.70	304	84.26	289
6	84.07	477	79.86	453
9	77.11	719	73.25	683
12	70.71	940	67.18	893
16	63.00	1208	59.85	1147
24	50.00	1660	47.50	1577
36	35.36	2172	33.59	2063
48	25.00	2534	23.75	2407
72	12.50	**2984**	= *0	**2692**

Some Problems…

Figure: Concentration vs. time plot with Reference (yellow), Test (red), and LLOQ = 10. Annotation: C_{72} estimated from C_{24}–C_{48}.

Some Problems…

● Missing values II

■ Last value of T missing (*e.g.*, vial broken)

■ Estimating the missing value from elimination phase.

AUC_{72*} T = 2835

AUC_{72} R = 2984

T/R = 95% ✔

➤ Not available in software

➤ Regulatory acceptance ±

time	Reference		Test	
	conc	AUC_{0-t}	conc	AUC_{0-t}
0	BLQ	0	BLQ	0
0.25	28.57	4	27.14	3
0.50	48.57	13	46.14	13
0.75	62.50	27	59.38	26
1.00	72.15	44	68.55	42
1.5	83.26	83	79.10	79
2	88.14	126	83.73	119
3	90.14	215	85.63	204
4	88.70	304	84.26	289
6	84.07	477	79.86	453
9	77.11	719	73.25	683
12	70.71	940	67.18	893
16	63.00	1208	59.85	1147
24	50.00	1660	47.50	1577
36	35.36	2172	33.59	2063
48	25.00	2534	23.75	2407
72	12.50	**2984**	*11.88	***2835**

Some Problems…

- ● Missing values II
  - ■ Values below the lower limit of quantitation (LLOQ)
    - ■ Example as before, but LLOQ = 12.5 (instead 10)
    - AUC_{72}: T = ?, R = 2984
      T/R = ?
    - AUC_{48}: T = 2407, R = 2534
      T/R = 95% ✓
    - AUC_{all}: T = 2692, R = 2984
      T/R = 90.22% biased!
   - AUC_{72*}: T = ?, R = 2984
      T/R = ?

	Reference		Test	
time	conc	AUC_{0-t}	conc	AUC_{0-t}
24	50.00	1660	47.50	1577
36	35.36	2172	33.59	2063
48	25.00	**2534**	23.75	2407
72	12.50	**2984**	BLQ	NA

	Reference		Test	
time	conc	AUC_{0-t}	conc	AUC_{0-t}
24	50.00	1660	47.50	1577
36	35.36	2172	33.59	2063
48	25.00	2534	23.75	2407
72	12.50	**2984**	= *0	**2692**

	Reference		Test	
time	conc	AUC_{0-t}	conc	AUC_{0-t}
24	50.00	1660	47.50	1577
36	35.36	2172	33.59	2063
48	25.00	2534	23.75	2407
72	12.50	**2984**	*11.88	NA

Some Problems…

What would you do?
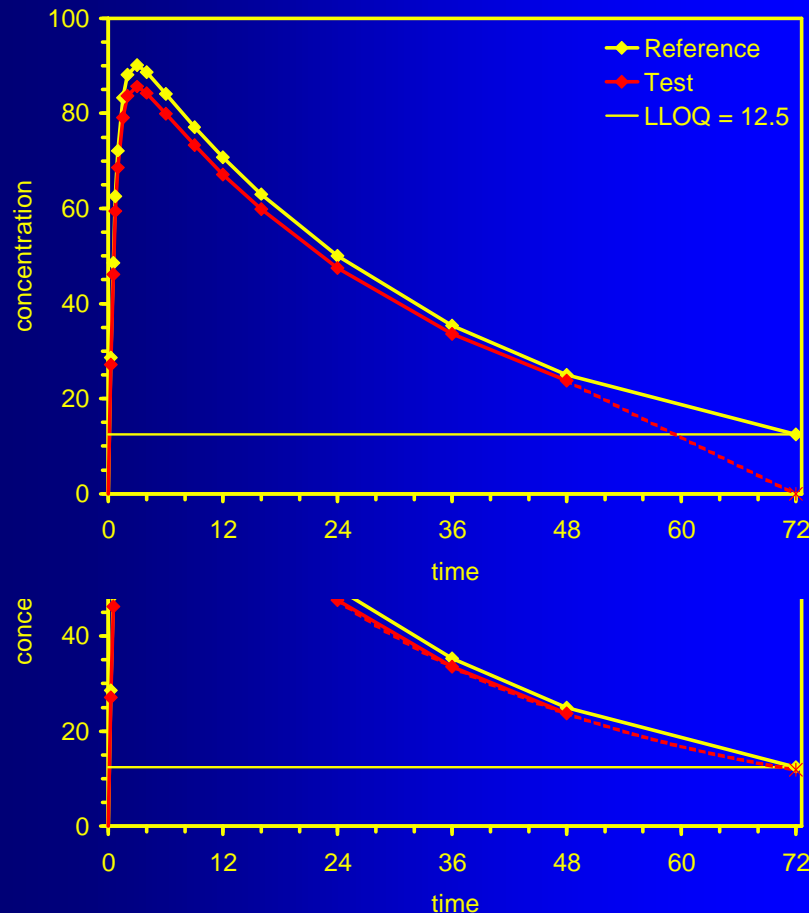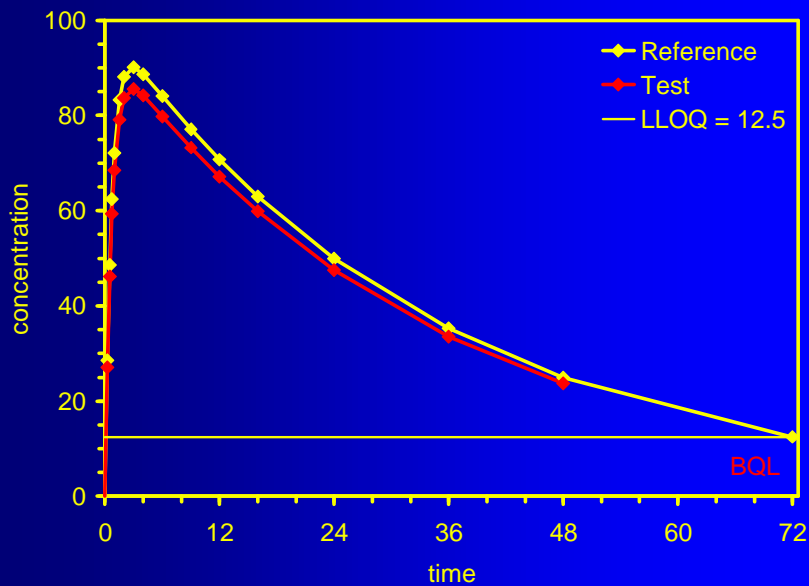
Sampling at C_{max}

- With *any (!)* given sampling scheme the *'true'* C_{max} is missed
  - It is unlikely that we sample *exactly* at the true C_{max} for any given subject
  - High inter- and/or intra-subject variability (single point metric)
  - Variability higher than $AUC's$
  - In many studies the win/loose metric!
  - Try to decrease variability
    - Increase sample size (more subjects)
    - Increase sampling *within* each subject (*maybe* better)

Sampling at C_{max}

- Theoretical values (from PK simulation)
  C_{max}: 41.9/53.5 (81.2%), t_{max}: 6.11/4.02 (Δ 2.09)
  - # samples [2–12h]
    - n = 4
      - C_{max} 78.3%
      - t_{max} Δ 4
    - n = 5
      - C_{max} 78.3%
      - t_{max} Δ 4
    - n = 6
      - C_{max} 79.8%
      - t_{max} Δ 1
    - n = 7
      - C_{max} 81.2%
      - t_{max} Δ 2

Legend: R theoretical, T theoretical, R sampled, T sampled
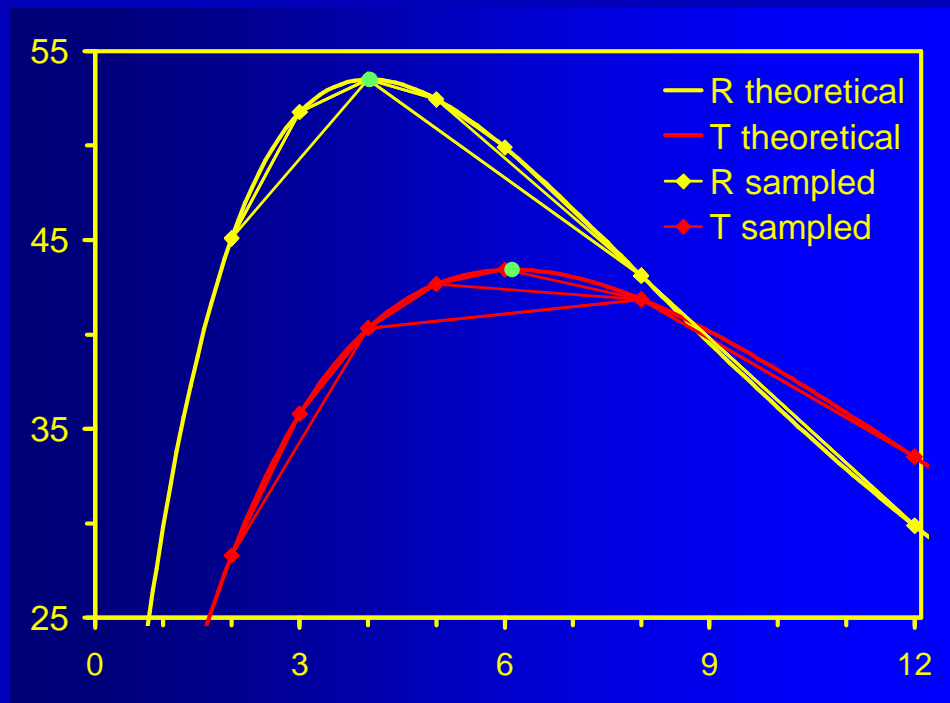
Sampling at C_{max}

- Quote from the literature:
  C_{max} was observed within two to five hours after oral administration…
  - Elimination is drug specific,
  - but what about absorption?
    - Formulation specific!
    - Dependent on the sampling schedule (in a strict sense study-specific)

Sampling at C_{max}

Sampling at C_{max}

2–5 hrs

arithmetic mean
geometric mean
median

$k_a = 0.182$

$t_{lag} = [0 \mid 2.5]$
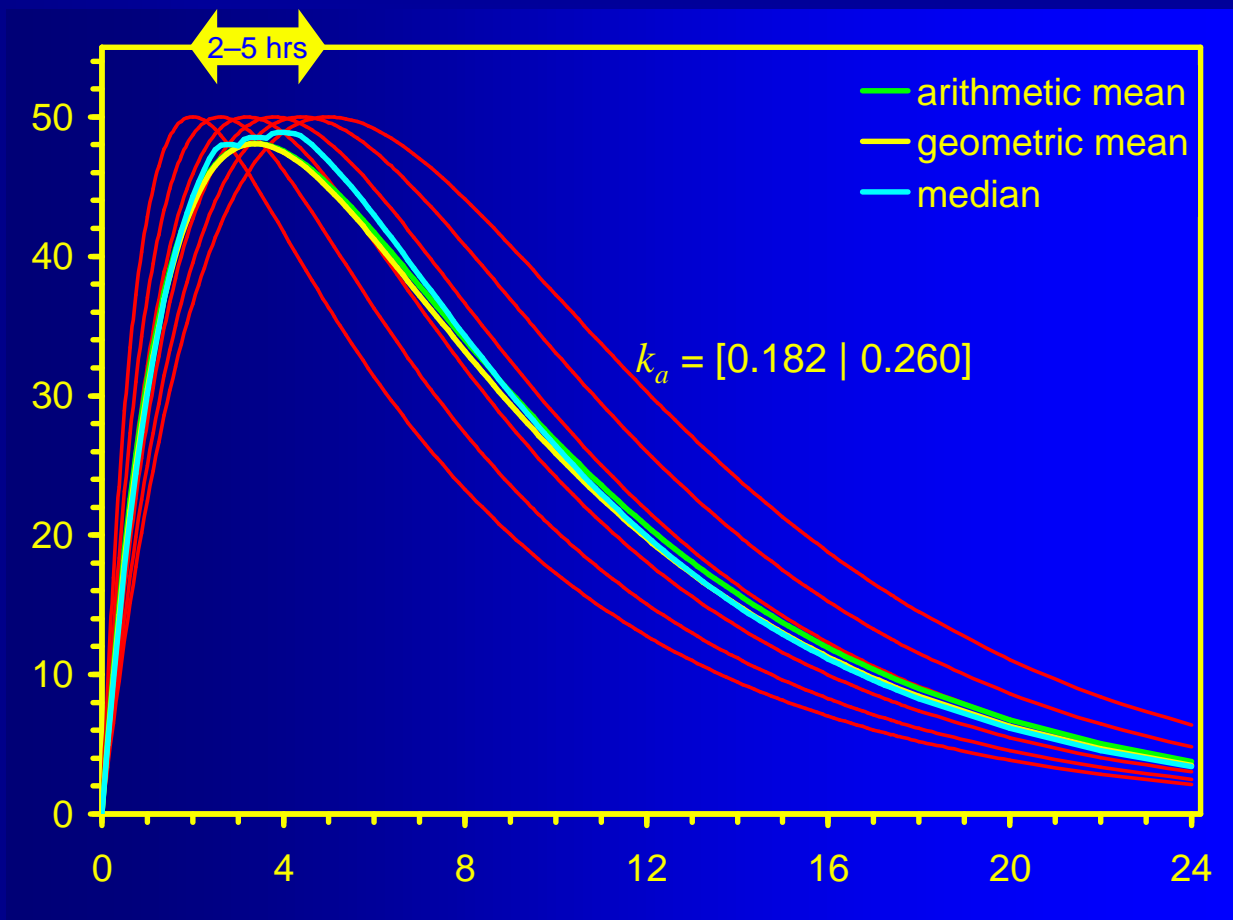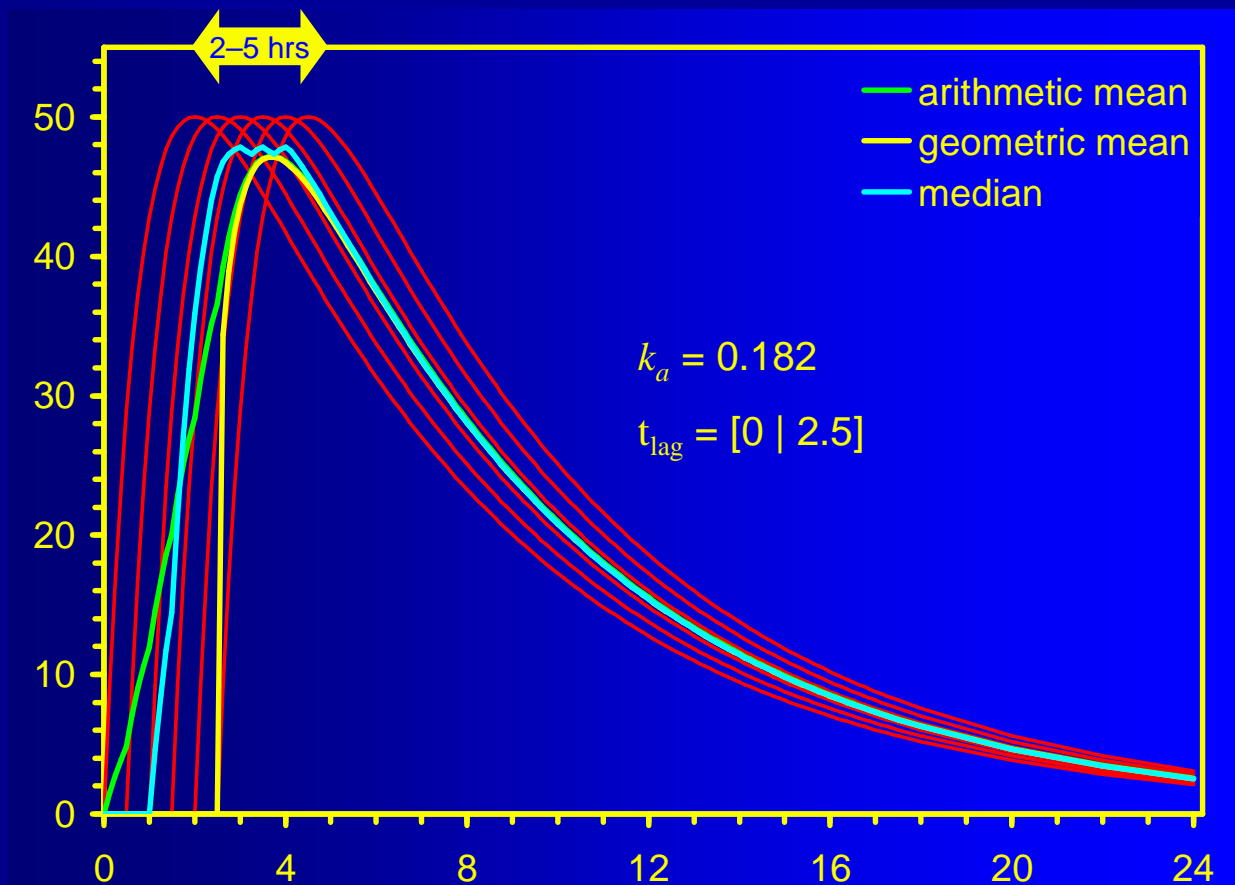
Another Problem

- EMA GL on BE (2010)
  - Section 4.1.8 Reasons for exclusion 1)
    - A subject with lack of any measurable concentrations or only very low plasma concentrations for reference medicinal product. A subject is considered to have very low plasma concentrations if its AUC is less than 5% of reference medicinal product geometric mean AUC (which should be calculated without inclusion of data from the outlying subject). The exclusion of data […] will only be accepted in exceptional cases and may question the validity of the trial.

      Remark: Only possible after unblinding!

Another Problem

- ## EMA GL on BE (2010)
  - ### Section 4.1.8 Resons for exclusion 1) cont'd
    - The above can, for immediate release formulations, be the result of subject non-compliance […] and should as far as possible be avoided by mouth check of subjects after intake of study medication to ensure the subjects have swallowed the study medication […]. The samples from subjects excluded from the statistical analysis should still be assayed and the results listed.

Another Problem

- Gastro-resistant (enteric coated) preparations
  - Gastric emptying of single unit dosage forms non-disintegrating in the stomach is prolonged and highly erratic. The consequences of this effect on the enteric coating of delayed release formulations are largely unpredictable.
    - Sampling period should be designed such that measurable concentrations are obtained, taking into consideration not only the half-life of the drug but the possible occurrence of this effect as well. This should reduce the risk of obtaining incomplete concentration-time profiles due to delay to the most possible extent. These effects are highly dependent on individual behaviour.

Another Problem

● Gastro-resistant (enteric coated) preparations

■ Therefore, but only under the conditions that sampling times are designed to identify very delayed absorption and that the incidence of this outlier behaviour is observed with a comparable frequency in both, test and reference pro-ducts, these incomplete profiles can be excluded from statistical analysis provided that it has been considered in the study protocol.

EMEA, CHMP (EWP-PK)
Questions & Answers: positions on specific questions addressed to the pharmacokinetics working party
EMA/618604/2008 Rev. 4 (16 February 2012)
http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002963.pdf

What is *'comparable'*? For a study in 24 subjects, we get a significant difference for 5/0 (Fisher's exact test: p 0.0496).

t_{lag} – a 'nasty' PK Metric

- Only relevant for gastric resistant (delayed release) formulations
- Highly variable – mainly not due to the formulation but the intrinsic variability in gastric emptying
- Less variability for multiparticulate formulations than for monolithic ones, but still problematic
- Sampling schedule difficult to design
- Assessment (descriptive *vs.* nonparametric)?

t_{lag} – a 'nasty' PK Metric

- Little is published about calculation; five methods assessed [*]

- Commercial software (Phoenix/WinNonlin, Kinetica) treat t_{lag} as the time point prior to the first measurable (non-zero) concentration

- Other methods require programming skills; some of them might be judged by assessors already borderline PK models (?!)

[*] **Csizmadia F and L Endrenyi**
Model-Independent Estimation of Lag Times with First-Order Absorption and Disposition
J Pharmaceut Sci 87/5, 608–12 (1998)

t_{lag} – a 'nasty' PK Metric

- Is t_{lag} *really* clinically relevant – even for formulations where rapid onset of effects is of importance?
- If two formulations follow identical pharmaco-kinetics *except t_{lag}*, this difference is reflected in t_{max} as well (both in SD and MD)
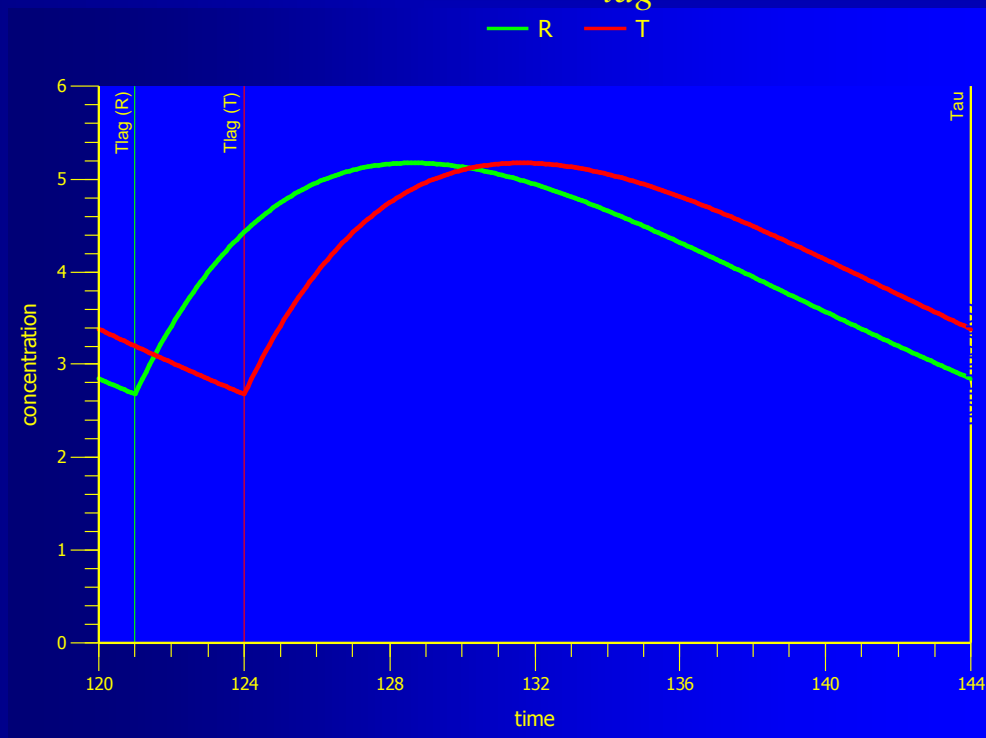
t_{lag} *vs.* t_{max}

- Single dose
  - DR, flip flop PK; V 10, D 100, F 100%, k 0.09902 h^{-1} ($t_{\frac{1}{2}}$ 7 h), $t_{lag,R}$ 1 h, $t_{lag,T}$ 4 h, $t_{max,R}$ 11.1 h, $t_{max,T}$ 14.1 h, C_{max} 3.68, $AUC_{0-\infty}$ 101.0
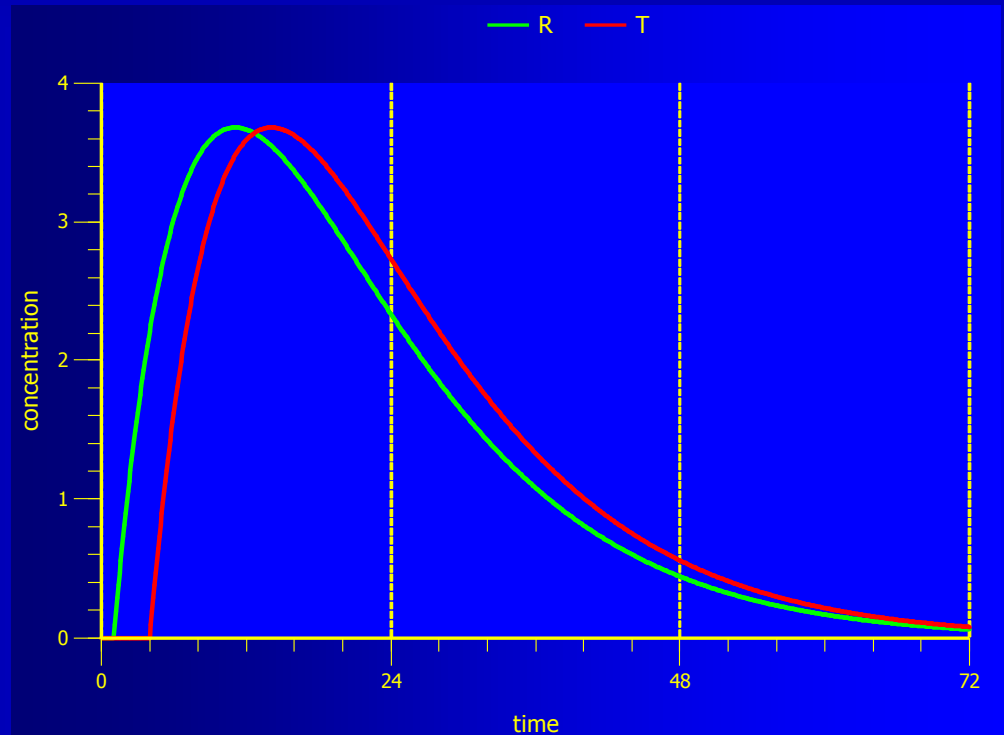
t_{lag} *vs.* t_{max}

- Simulation of steady state (τ 24 h; 6 d $\approx$ 20×$t_{\frac{1}{2}}$)
  - Formulations differ in t_{lag} only!

- t_{lag} is discrimi-natory:
  T   4
  R   1
  T – R   +3
- Might be difficult to measure; frequent sampling required
- Nonparametric statistics (EMA!)

t_{lag} *vs.* t_{max}

- Simulation of steady state (τ 24 h; 6 d $\approx$ 20×$t_{1/2}$)
  - Formulations differ in t_{lag} only! Surrogate possible?
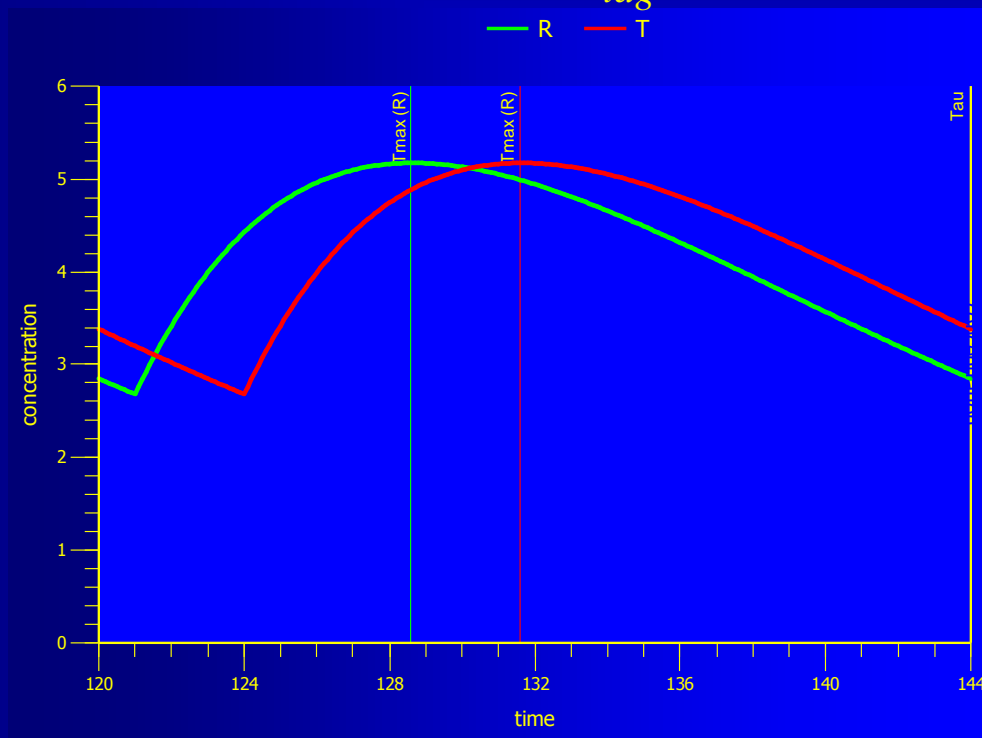
- t_{max} is discrimi-natory as well:
  T   14.1
  R   11.1
  T – R +3
- Maybe better; frequent sam-pling in the area of C_{max} common
- Nonparametric statistics (EMA!)

Case Study (PPI 1)

- Attempt to deal with high variability

Powered to 90% according to CV from previous studies; 140 (!) subjects and to 80% for expect-ed dropout rate. Sampling every 30 min up to 14 hours (7,785 total).
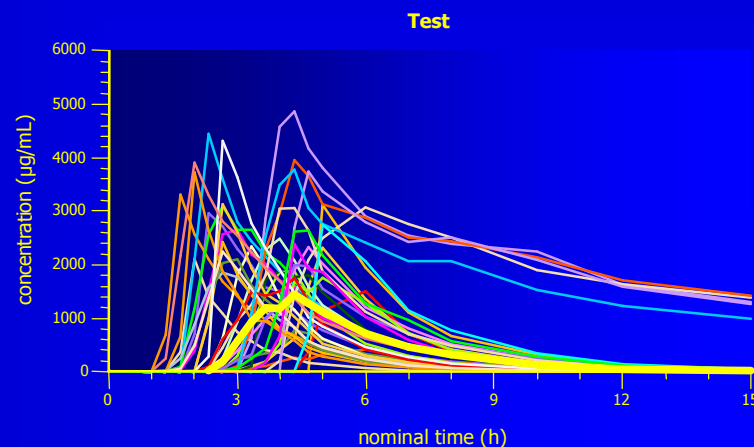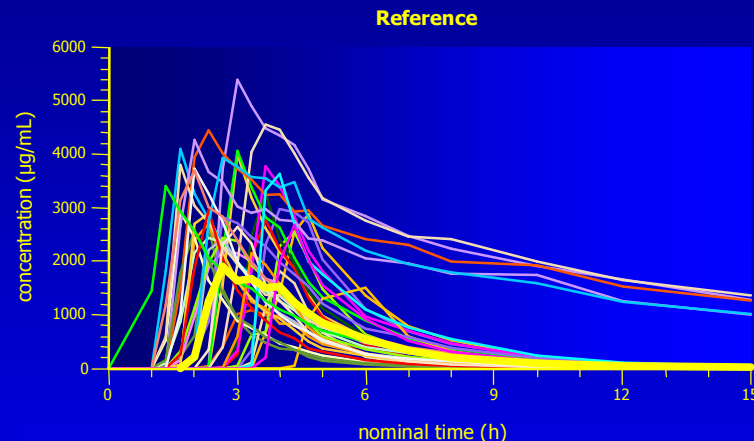
First time C_{max}

$t_{½}$ 12 h

t_{max} 15 h, C_{max} 3.5×LLOQ

t_{lag} 6 h

time (h)
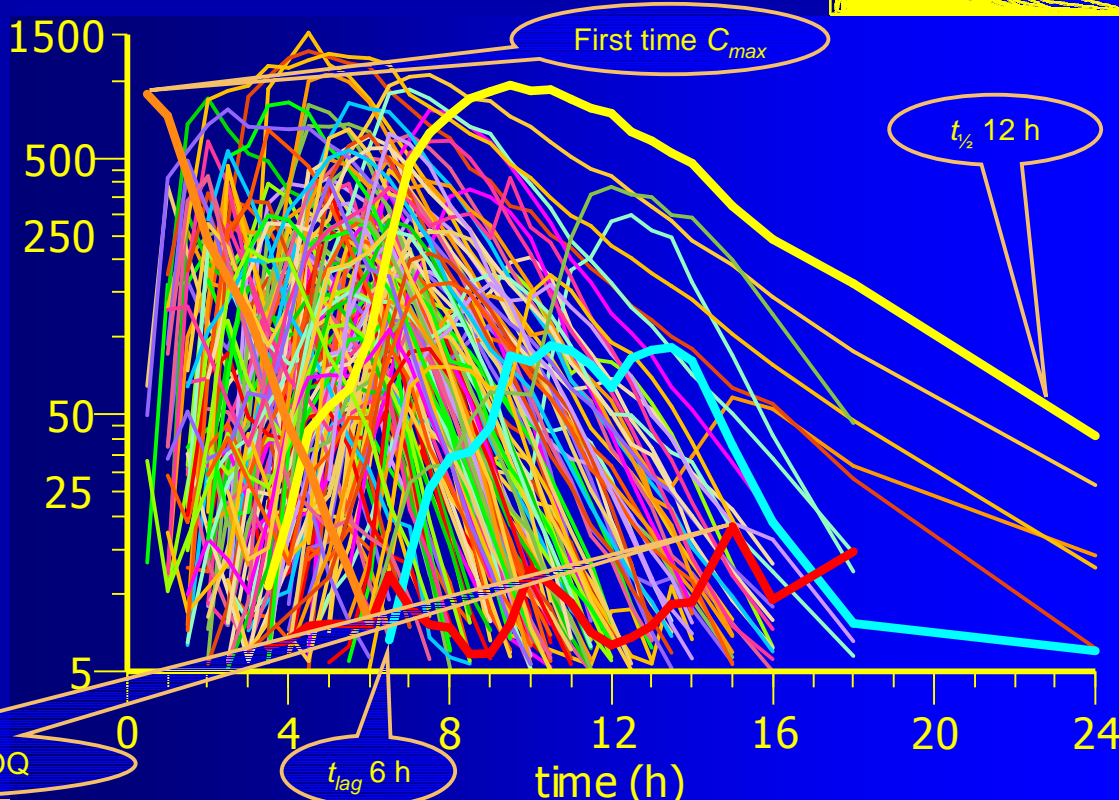
Case Study (PPI 2)

- Submission in China

  AUC_t 87.60, 95.53%
  C_{max} 75.39, 91.84%
  t_{max} +0.500, +1.333
  significantly delayed
  (0 not within CI)

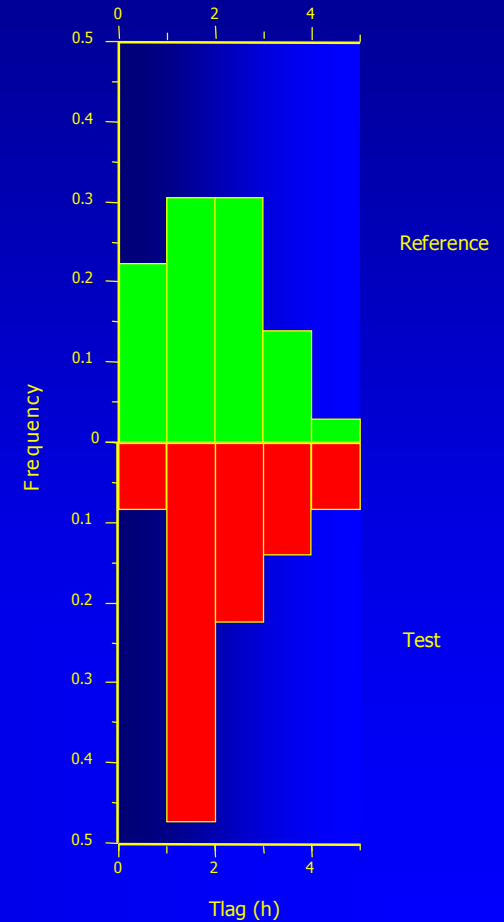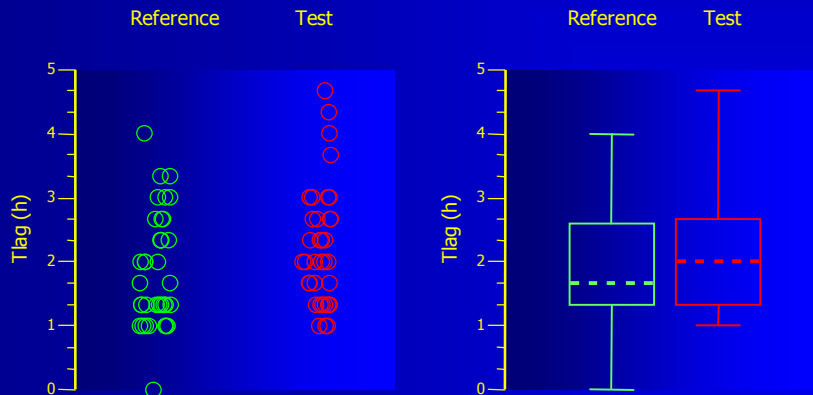- Company's defending argument: caused by highly variable GI-transit manifested in t_{lag}.

- Let's see…
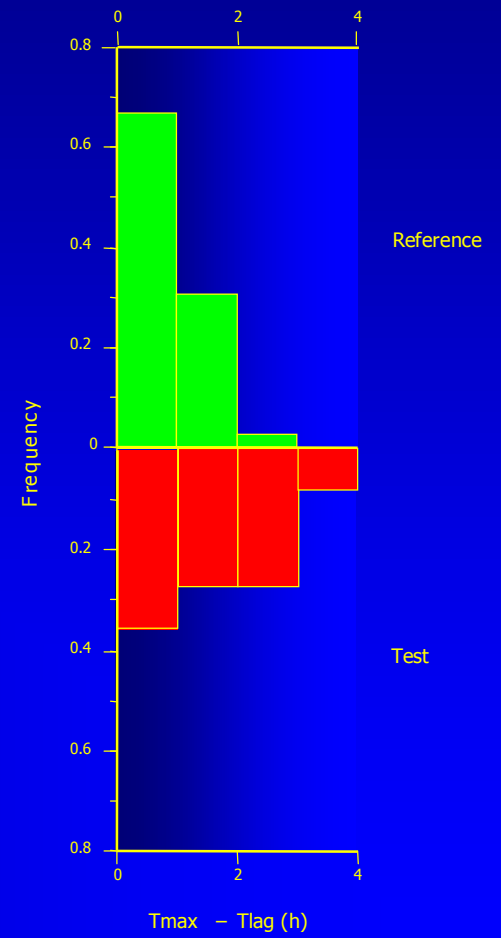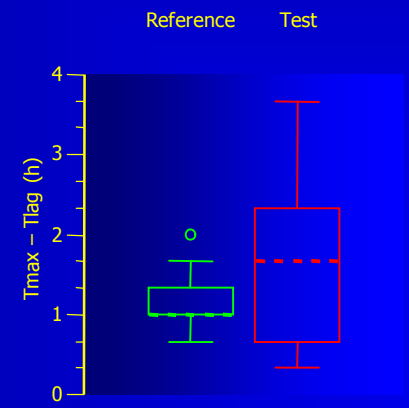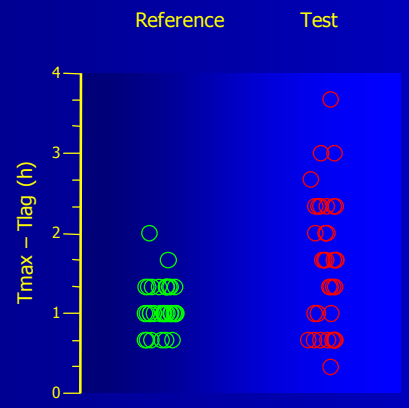
Reference

Test

Case Study (PPI 2)

- Analysis

t_{lag} ±0.000, +0.667
not different (but borderline)

Case Study (PPI 2)

- Analysis

$t_{max}-t_{lag}$ +0.167, +0.667
significantly delayed
(0 not within CI)

Case Study (PPI 2)

Reference

- Assessment
  - Although there was no sig-nificant difference in t_{lag}, the 'corrected' $t_{max}-t_{lag}$ was significantly delayed.
  - Variability of the test formulation was higher.
  - It seems that the com-pany's assumption does not hold – formulations differ.
  - Clinical relevance?

Test
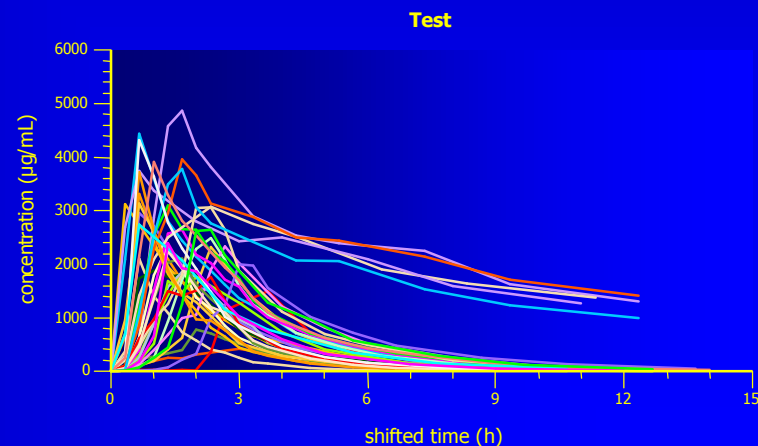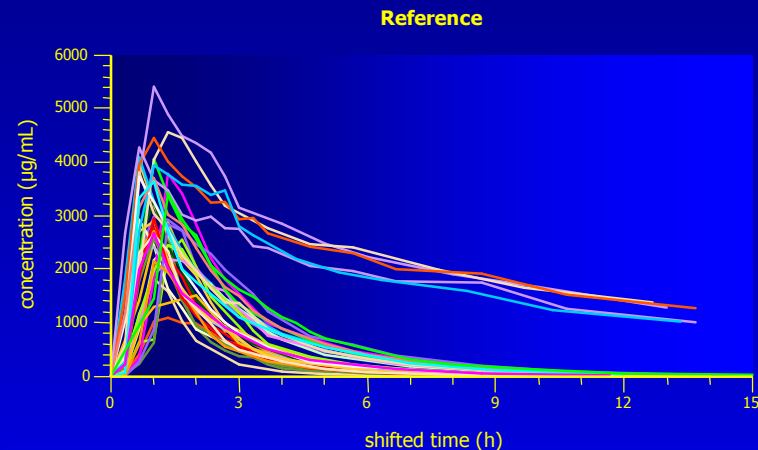
Half lives

- Drug specific, *but*…
  - The *apparent* elimination represents the *slowest* rate constant (controlled release, topicals, transdermals) – not necessarily elimination!
  - Avoid the term 'terminal elimination' – might not be true
  - Important in designing studies
    - To meet $AUC_t \geq 80\%$ AUC_∞ criterion
    - To plan sufficiently long wash-out (avoid carry-over)
    - To plan saturation phase for steady state

Half lives

- Dealing with literature data
  - What if only mean ±SD is given?
    - Assuming normal distribution:
      $\mu \pm \sigma$ covers 68.27% of values (15.87% of values are expected to lie outside of $\mu \pm \sigma$)
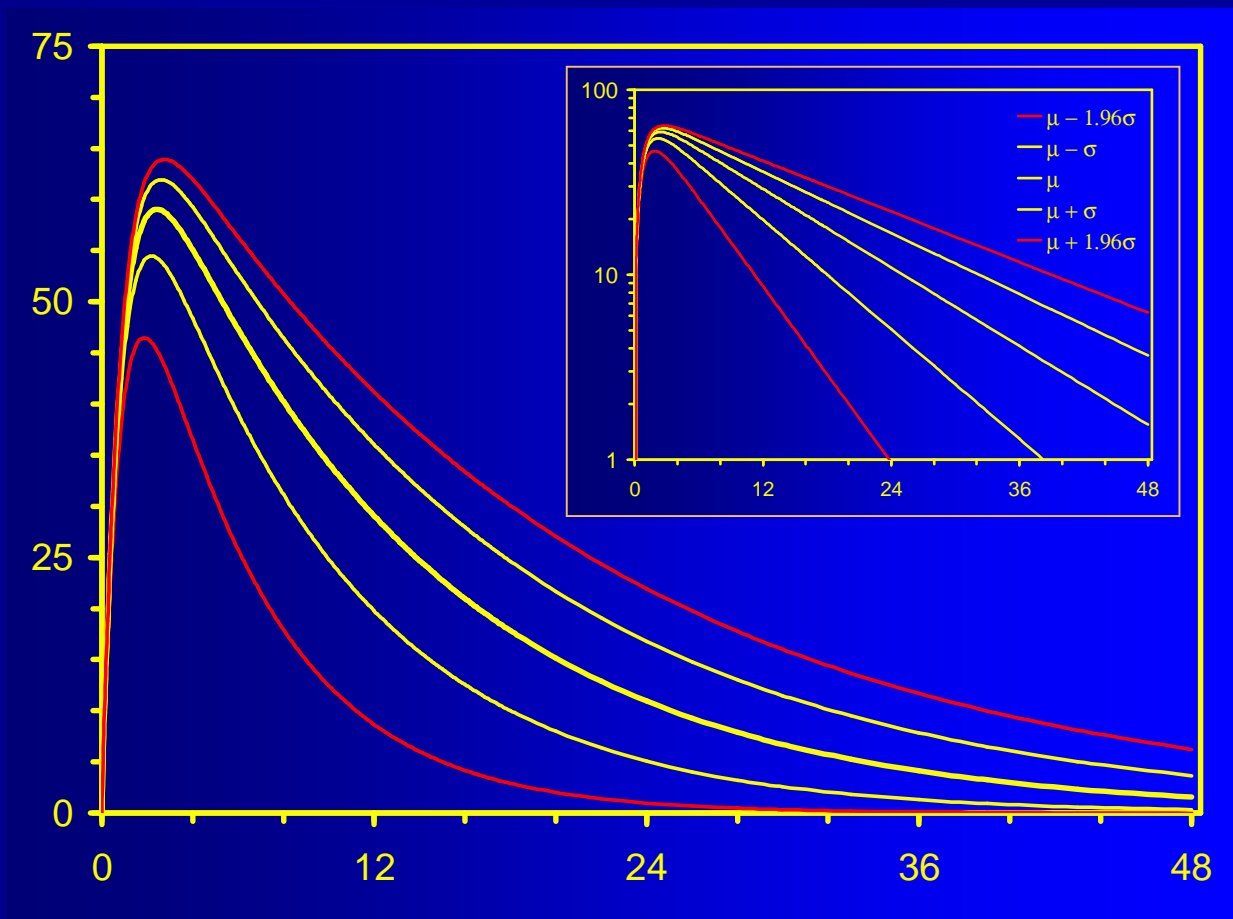   - Example: 8.5 ± 2.4 hours, 36 subjects.
      0.1587 × 36 = 5.71 or in at least five subjects we may expect a half life of > 10.9 hours.
    - Plan for 95% coverage ($z_{0.95}$ = 1.96): $p_{0.95} = \mu \pm z_{0.95} \times \sigma$
      8.5 ± 1.96 × 2.4 = [3.80, 13.2] hours.
      We may expect a half life of >13.2 hours in ~one subject
      (0.05/2 × 36 = 0.90).

Half lives

Washout in MD Studies

- ## EMA GL on BE (2010)

  The treatment periods should be separated by a wash out period sufficient to ensure that drug concentrations are below the lower limit of bioanalytical quantification in all subjects at the beginning of the second period. Normally at least 5 elimination half-lives are necessary to achieve this. In steady-state studies, the wash out period of the previous treatment last dose can overlap with the build-up of the second treatment, provided the build-up period is sufficiently long (at least 5 times the terminal half-life).

  - Justified by PK Superposition Principle
  - 'Switch-over Design'

  2001 NfG: ≥3 half-lives
  Russia: ≥4 half-lives

Washout in MD Studies

washout *vs.* switch-over

$t_{1/2} = 12$

$\tau = 24$
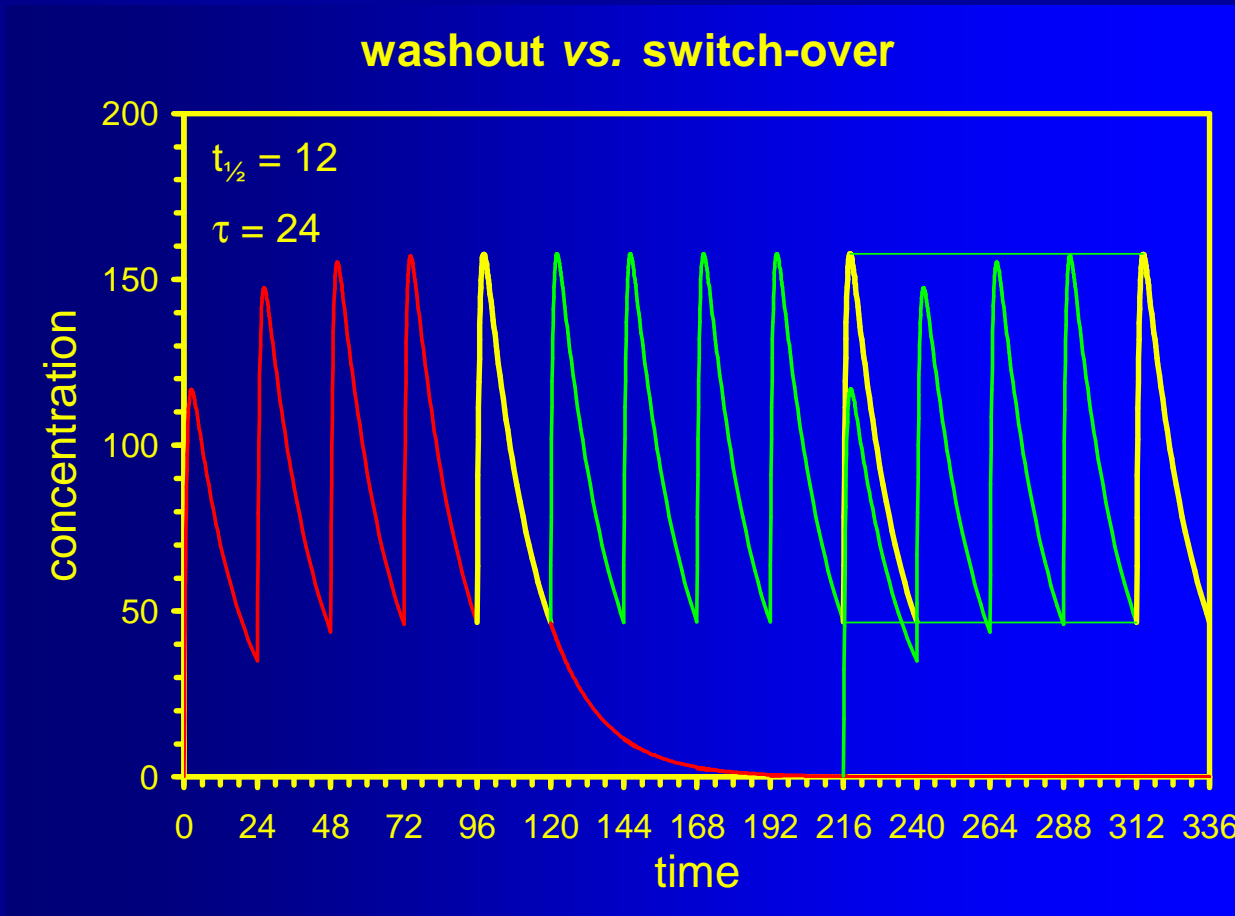
(Bio)statistics

Statistics. A subject which most statisticians find difficult but in which nearly all physicians are expert.

Biostatistician. One who has neither the intellect for mathematics nor the commitment for medicine but likes to dabble in both.

Medical statistician. One who will not accept that Columbus discovered America… because he said he was looking for India in the trial plan.

Stephen Senn

Bioequivalence

- Background / definition (EMA 2010)
  - Two medicinal products containing the same active substance are considered bioequivalent if they are pharmaceutically equivalent or pharmaceutical alternatives and their bioavailabilities (rate and extent) after administration in the same molar dose lie within acceptable predefined limits. These limits are set to ensure comparable *in vivo* performance, i.e. similarity in terms of safety and efficacy.

Bioequivalence

- Background (EMA 2010)
  - In bioequivalence studies, the plasma concentration time curve is generally used to assess the rate and extent of absorption. Selected pharmacokinetic parameters and preset acceptance limits allow the final decision on bioequivalence of the tested pro-ducts. AUC, the area under the concentration time curve, reflects the extent of exposure. C_{max}, the maximum plasma concentration or peak exposure, and the time to maximum plasma concentration, t_{max}, are parameters that are influenced by absorp-tion rate.

Bioequivalence

AUC (T/R) = 96.5%, Cmax (T/R) = 98.6%, Tmax (T–R) = -0.5
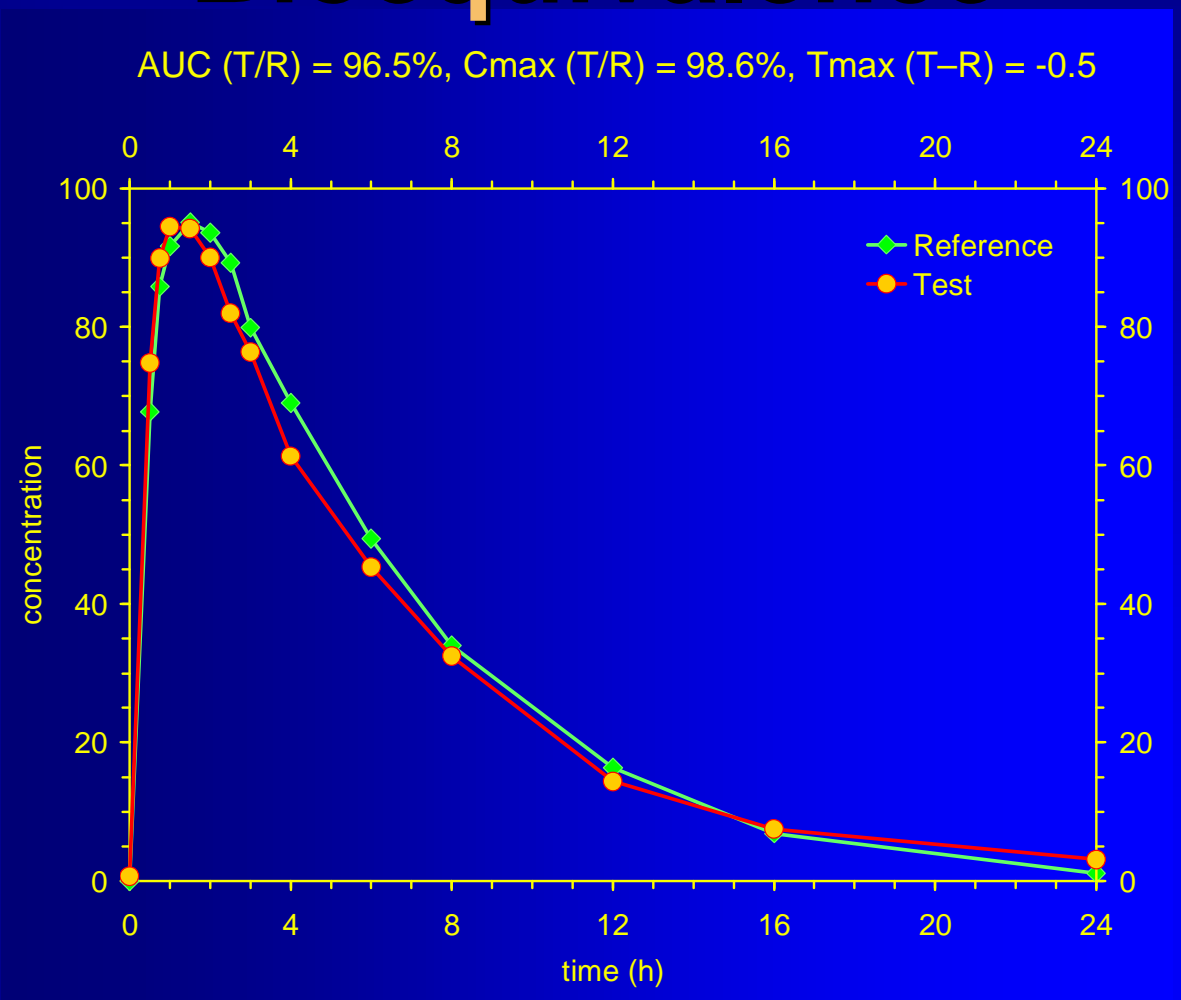
Bioequivalence

- Regulatory background
  - Generic applications
    - EMA: Directive 2001/83/EC, Article 10(1)
    - FDA: Abbreviated New Drug Applications (21CFR320.21)
  - Bridging studies
    - Scale-up from pilot batches used in Phase III to full production batches
    - Major variations of approved formulations EMA: Type II(d)–(f), FDA: SUPAC Level 3
    - Line extensions (*e.g.*, new dosage forms, new strengths if waiving not possible)

Concept of BE…

- Statistical concept of BE also applicable to
  - Food effect studies
  - PK interaction studies
  - Studies of fixed-dose combination products

    '[…] are similar to such degree that their effects, with respect to both efficacy and safety, will be essentially the same.'

EMEA Human Medicines Evaluation Unit / CPMP
Modified Release Oral and Transdermal Dosage Forms: Section II (Quality)
CPMP/EWP/280/96 (1999)
EMEA Human Medicines Evaluation Unit / CPMP
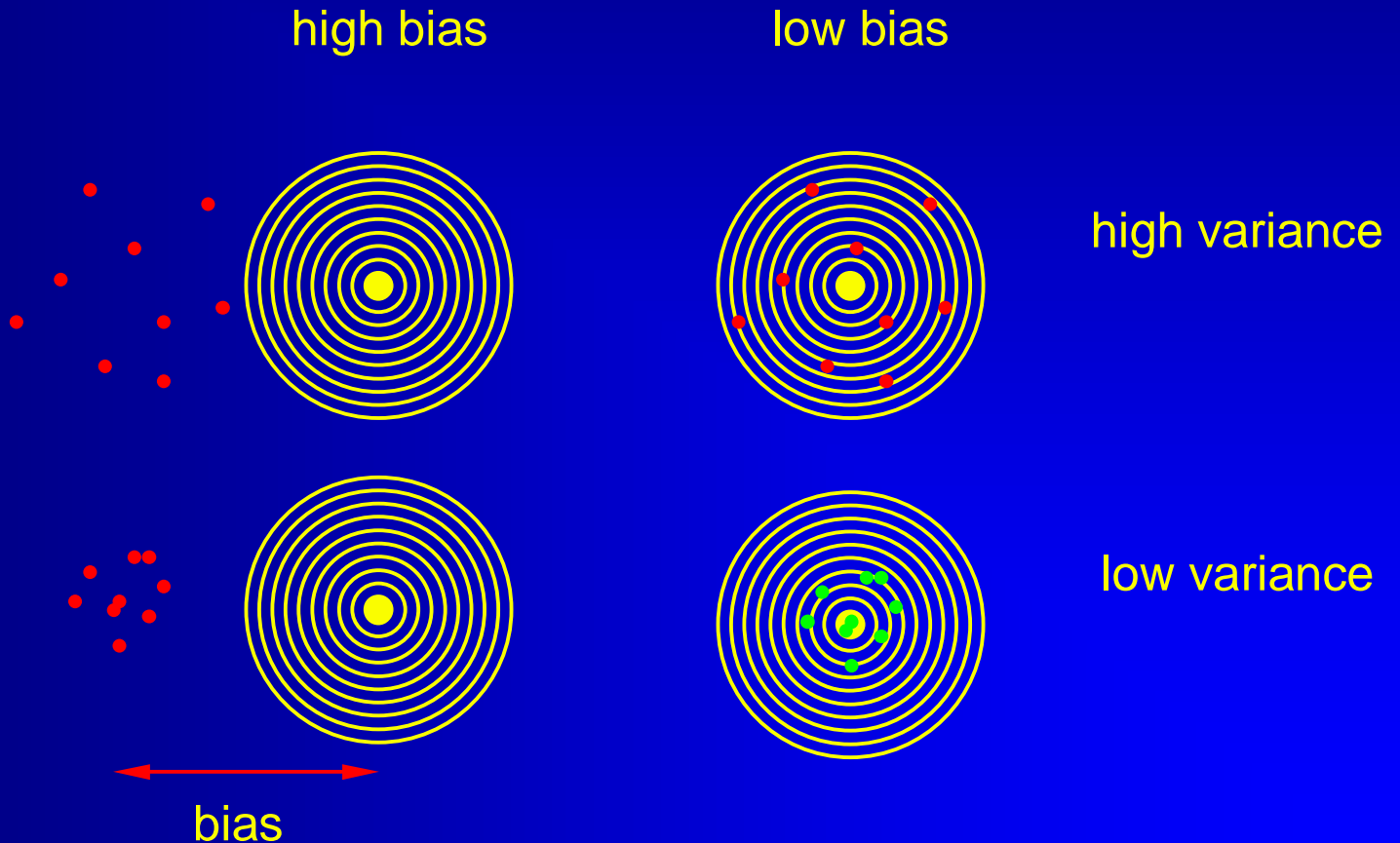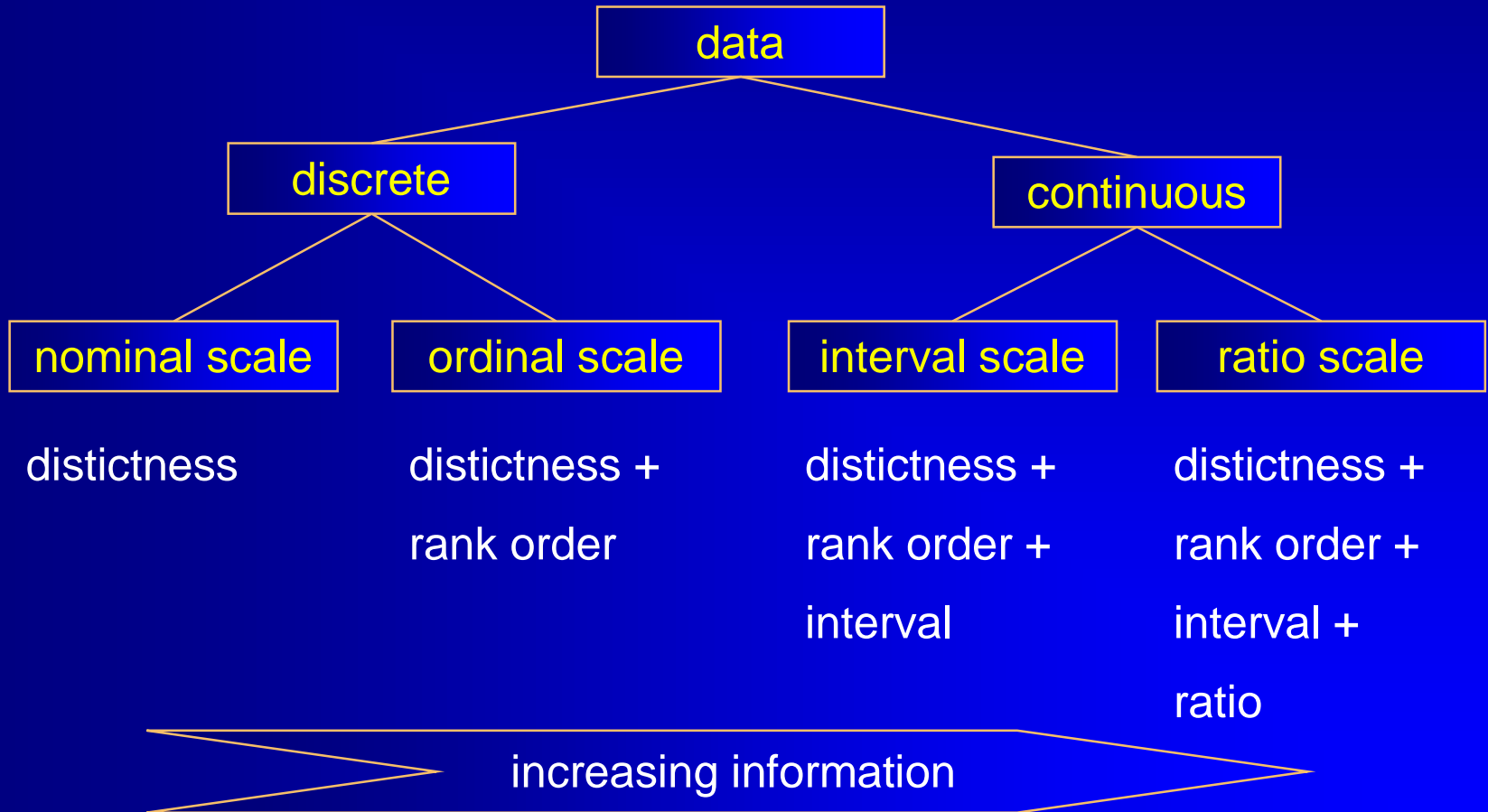*The Investigation of Drug Interactions*
CPMP/EWP/560/95 (1997)
EMEA
Fixed Combination Medicinal Products
CPMP/EWP/240/95 Rev. 1 (2008)

Terminology I

high bias   low bias

high variance

low variance

bias

Terminology II

```
                            ┌──────────────┐
                            │     data     │
                            └──────────────┘
                   ┌───────────────┴───────────────┐
          ┌────────────────┐              ┌──────────────────┐
          │    discrete    │              │    continuous    │
          └────────────────┘              └──────────────────┘
        ┌──────────┴──────────┐         ┌──────────┴──────────┐
┌────────────────┐ ┌────────────────┐ ┌────────────────┐ ┌────────────────┐
│  nominal scale │ │  ordinal scale │ │ interval scale │ │   ratio scale  │
└────────────────┘ └────────────────┘ └────────────────┘ └────────────────┘
```

distictness

distictness +
rank order

distictness +
rank order +
interval

distictness +
rank order +
interval +
ratio

increasing information

Data I

- Nominal scale (aka categorial)
  - Sex, ethnicity,…
    - Statistics: mode, χ^2 test
    - Transformations: equality
- Ordinal scale
  - School grades, disease states,…
    - Statistics: median, percentile, sign test, Wilcoxon test
    - Transformations: monotonic increasing order

Data II

- Interval scale
  - Calendar dates, temperature in °C, IQ,…
    - Statistics: mean, variance (standard devi-ation), correlation, regression, ANOVA
    - Transformations: linear
- Ratio scale
  - Measures with true zero point, temperature in K,…
    - Statistics: all of the above, geometric and harmonic mean, coefficient of variation
    - Transformations: multiplicative, logarithm

Examples from PK

- Ordinal scale

  - t_{max}, t_{lag}
    - Statistics: median, percentile, sign test, Wilcoxon test
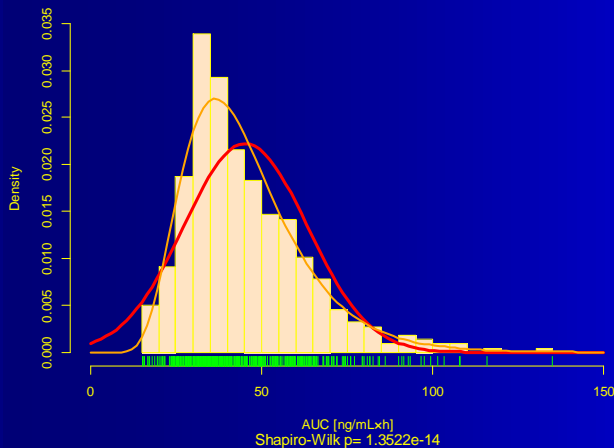   - Transformations: monotonic increasing order

- Ratio scale

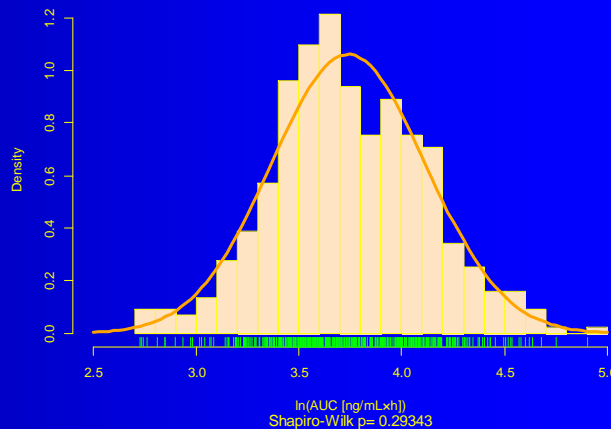 - AUC, C_{max}, λ_z,…
    - Statistics: mean, variance (standard deviation), correlation, regression, ANOVA, geometric and harmonic mean, coefficient of variation
    - Transformations: multiplicative, logarithm
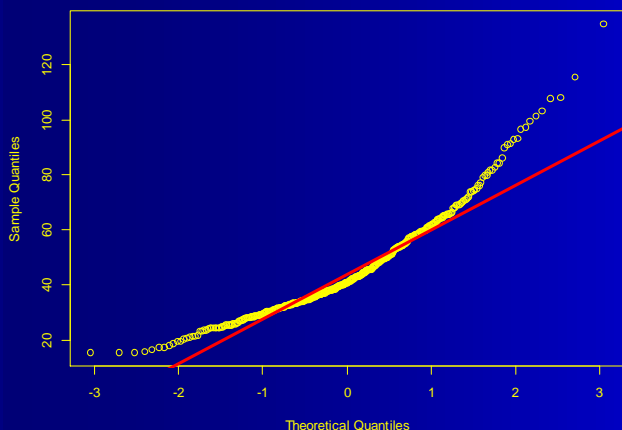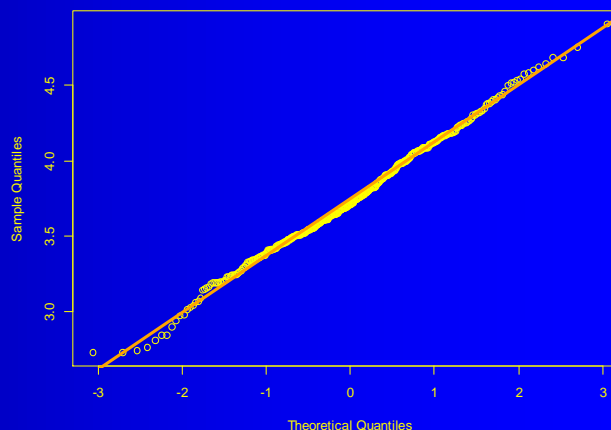
Remark on Transformation

MPH, 437 subjects

AUC [ng/mL×h]
Shapiro-Wilk p= 1.3522e-14

Normal Q-Q Plot

MPH, 437 subjects

ln(AUC [ng/mL×h])
Shapiro-Wilk p= 0.29343

Normal Q-Q Plot

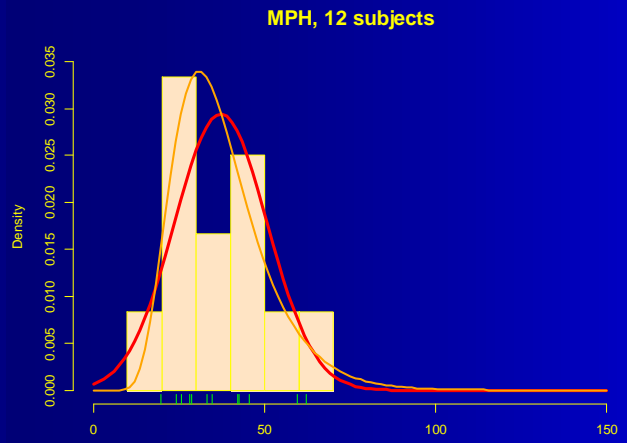Pooled data from studies of MR methyl-phenidate. Clearly in favor of a lognormal distribution.

Shapiro-Wilk test for normal distribution highly signifi-cant (distributi-onal assump-tions rejected).

Remark on Transformation

MPH, 12 subjects

AUC [ng/ml×h]
Shapiro-Wilk p= 0.29667

MPH, 12 subjects
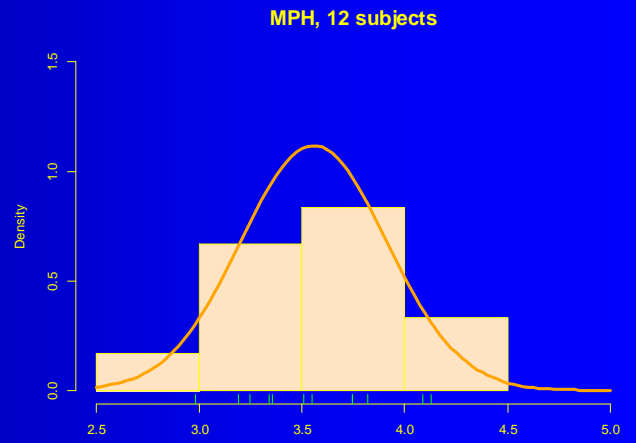
ln(AUC [ng/mL×h])
Shapiro-Wilk p= 0.85764
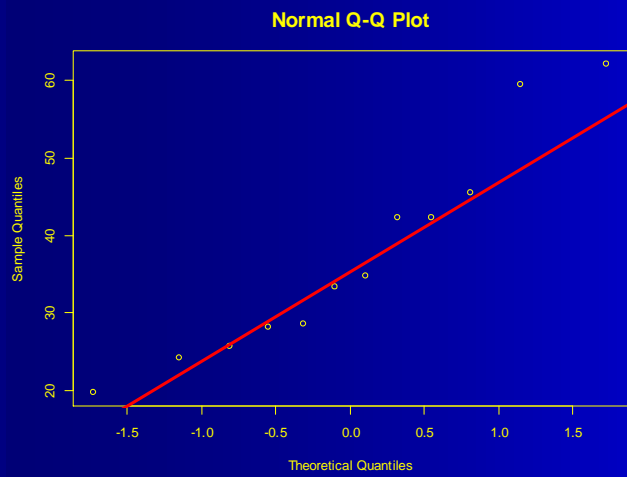
Normal Q-Q Plot
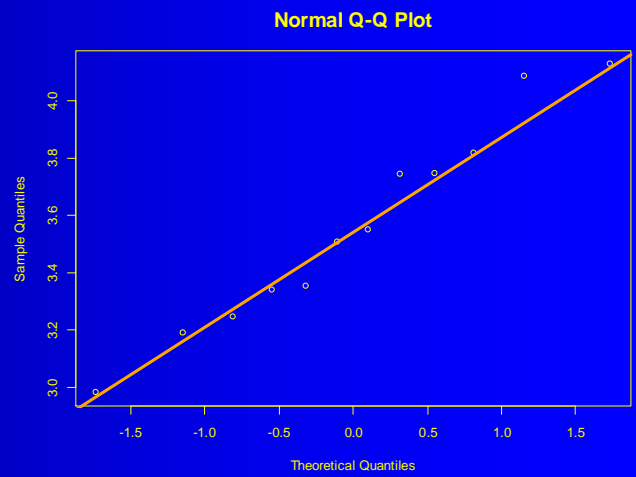
Normal Q-Q Plot

Data set from one of the studies. Both tests *not* signi-ficant (distribu-tional assump-tions *not* rejected).

Tests not acceptable according to GLs; transfor-mation based on prior know-ledge (PK)!

Global Harmonization?

Transformations (e.g. […], logarithm) should be specified in the protocol and a rationale provided […]. The general principles guiding the use of transformations to ensure that the assumptions underlying the statistical methods are met are to be found in standard texts […]. In the choice of statistical methods due attention should be paid to the statistical distribution […]. When making this choice (for example between parametric and non-parametric methods) it is important to bear in mind the need to provide statistical estimates of the size of treatment effects together with confidence intervals […].

ICH Topic E 9
Statistical Principles for Clinical Trials (1998)

Global Harmonization?

No analysis is complete until the assumptions that have been made in the modeling have been checked. Among the assumptions are that the repeated measurements on each subject are independent, normally distributed random variables with equal variances. Perhaps the most important advantage of formally fitting a linear model is that diagnostic information on the validity of the assumed model can be obtained. These assumptions can be most easily checked by analyzing the residuals.

Jones B and MG Kenward
Design and Analysis of Cross-Over Trials
Chapman & Hall, Boca Raton (2nd ed 2003)

Nonparametrics

The limited sample size in a typical BE study precludes a reliable determination of the distribution of the data set. Sponsors and/or applicants are not encouraged to test for normality of error distribution after log-transform-ation […].

FDA, Center for Drug Evaluation and Research (CDER)
Guidance for Industry: Statistical Approaches to Establishing Bioequivalence (2001)

But: acceptable in
Turkey (MOH, November 2005)
Saudia Arabia (SFDA, May 2005)
Japan (NIHS, November 2006)

Nonparametrics

5. In which cases may a non-parametric statistical model be used?

The NfG states under 3.6.1–Statistical analysis: "*AUC and C_{max} should be analysed using ANOVA after log transformation.*"

The reasons for this request are the following:

a) the AUC and C_{max} values as biological parameters are usually not normally distributed;

b) a multiplicative model may be plausible;

c) after log transformation the distribution may allow a parametric analysis.

<u>Comments:</u>

a) – true   b) – true   c) – maybe, but may also terribly fail

EMEA/CHMP/EWP/40326/2006
Questions & Answers on the BA and BE Guideline (2006)

Nonparametrics

5. In which cases may a non-parametric statistical model be used?

However, the true distribution in a pharmacokinetic data set usually cannot be characterised due to the small sample size, so it is <u>not recommended</u> to have the analysis strategy depend on a pre-test for normality. Parametric testing using ANOVA on log-transformed data should be the rule. Results from non-parametric statistical methods or other statistical approaches are nevertheless welcome as sensitivity analyses. Such analyses can provide reassurance that conclusions from the experiment are robust against violations of the assumptions underlying the analysis strategy.

<u>Comment:</u> It is well known that the efficiency of *e.g.,* the Wilcoxon-Mann-Whitney test for normal distributed data is $3/\pi \approx 95.5$ %; for *not normal distributed data* the efficiency is >100 %!

Nonparametrics

4.1.8 Evaluation / Statistical analysis

The pharmacokinetic parameters under consideration should be analysed using ANOVA (or equivalent parametric method). The data should be transformed prior to analysis using a logarithmic transformation. A confidence interval for the difference between formulations on the log-transformed scale is obtained from the ANOVA model. This confidence interval is then back-transformed to obtain the desired confidence interval for the ratio on the original scale. <u>A non-parametric analysis is not acceptable.</u>
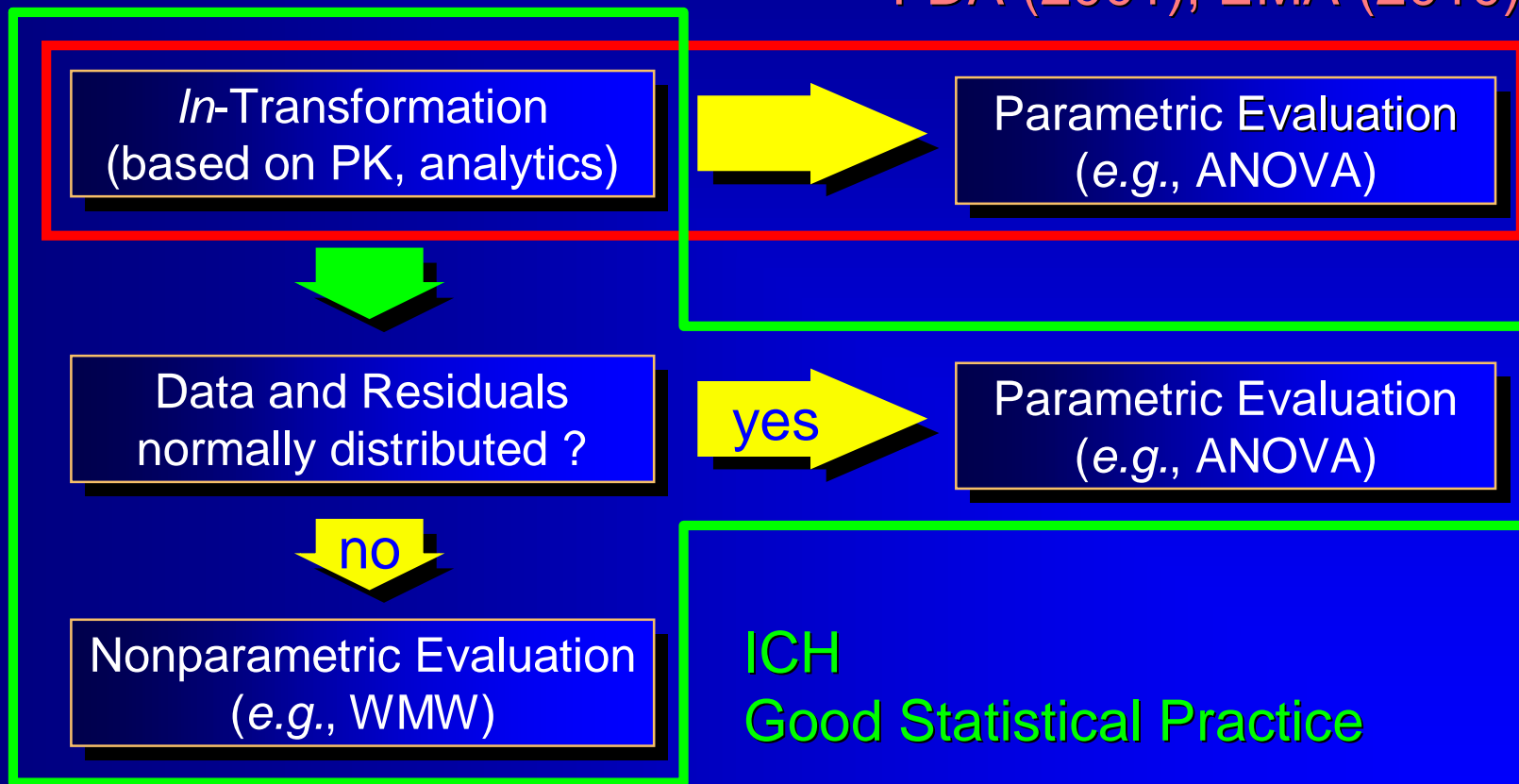
EMEA/CPMP/EWP/QWP/1401/98 Rev. 1
Draft Guideline on the Investigation of Bioequivalence (2008)

Walter Hauck: *'Also interesting that they now say they will not accept non-parametric analyses. That seems a step backwards.'* (personal communication Oct 2008)

Global Harmonization?

FDA (2001), EMA (2010)

| *ln*-Transformation (based on PK, analytics) | → | Parametric Evaluation (*e.g.*, ANOVA) |

| Data and Residuals normally distributed ? | yes → | Parametric Evaluation (*e.g.*, ANOVA) |

| no ↓ | | |

| Nonparametric Evaluation (*e.g.*, WMW) | | ICH Good Statistical Practice |

Hierarchy of Designs

- The more 'sophisticated' a design is, the more information can be extracted

  - Hierarchy of designs:
    Full replicate (TRTR | RTRT) ⇲
       Partial replicate (TRR | RTR | RRT) ⇲
          Standard 2×2 cross-over (RT | RT) ⇲
             Parallel (R | T)

  - Variances which can be estimated:
                Parallel: total variance (between + within)
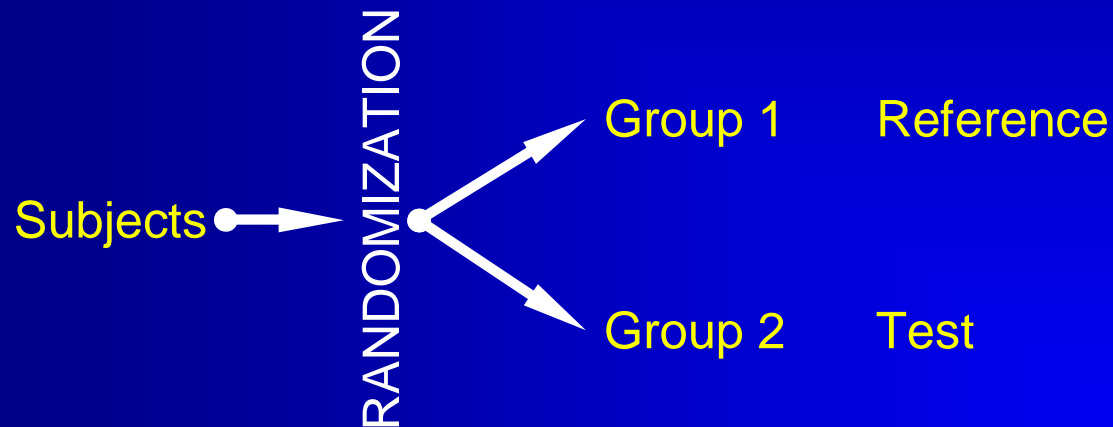   2×2 Xover: + between, within subjects ☞
          Partial replicate: + within subjects (reference) ☞
    Full replicate: + within subjects (reference, test) ☞

Information

Parallel Designs

- One group is treated with the test formulation and another group with reference
- Common that the dataset is imbalanced, *i.e.*, $n_1 \neq n_2$
- Guidelines against the assumption of equal variance
- Not implemented in PK software (Phoenix/WinNonlin, Kinetica)!
- Welch's t-test (available in SAS, SPlus, or R)

Parallel Designs

- Two-Group Parallel Design
  - Advantages
    - Clinical part – *sometimes* – faster than Xover.
    - Straigthforward statistical analysis.
    - Drugs with long half life.
    - Potentially toxic drugs or effect and/or AEs unacceptable in healthy subjects.
    - Studies in patients, where the condition of the disease irreversibly changes.
  - Disadvantages
    - Lower statistical power than Xover (assuming same sample size).
    - Phenotyping mandatory for drugs showing polymorphism.

Cross-over Designs

● Standard 2×2×2 Design
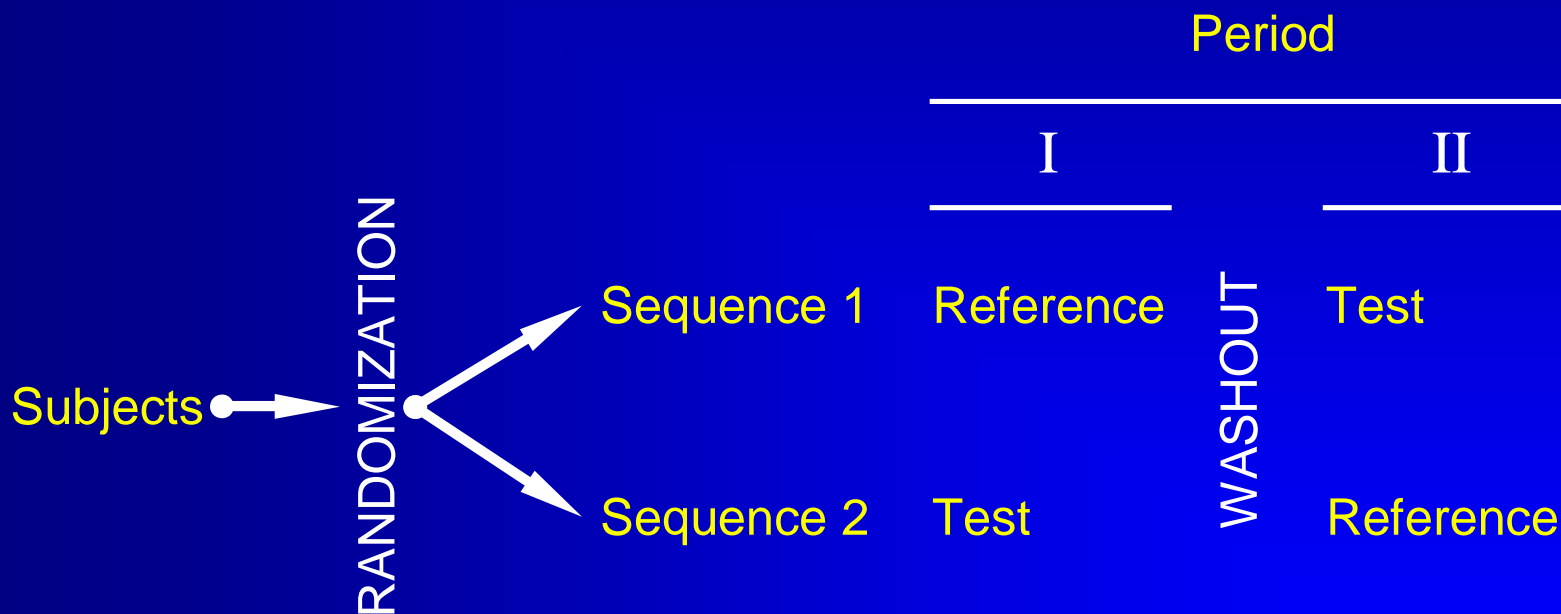
Xover Designs (cont'd)

- Every subject is treated once with both test and reference.

- Subjects are randomized into two groups; one is receiving formulations in the order RT and the other one in the order TR.
  These two orders are called sequences.

- Whilst in a paired design we must rely on the assumption that no external influences affect the periods, a cross-over design will account for that.

Xover Design: Model

Multiplicative Model (Xover without carryover)

$$X_{ijk} = \mu \cdot \pi_k \cdot \Phi_l \cdot s_{ik} \cdot e_{ijk}$$

X_{ijk}: *ln*-transformed response of j-th subject ($j=1,…,n_i$) in i-th sequence ($i=1,2$) and k-th period ($k=1,2$), μ: global mean, μ_l: expected formulation means ($l=1,2$: $\mu_1=\mu_{test}$, $\mu_2=\mu_{ref.}$), π_k: fixed period effects, Φ_l: fixed formulation effects ($l=1,2$: $\Phi_1=\Phi_{test}$, $\Phi_2=\Phi_{ref.}$)

Xover Design: Assumptions

Multiplicative Model (Xover without carryover)

$$X_{ijk} = \mu \cdot \pi_k \cdot \Phi_l \cdot s_{ik} \cdot e_{ijk}$$

- All $ln\{s_{ik}\}$ and $ln\{e_{ijk}\}$ are independently and normally distributed about unity with variances σ^2_s and σ^2_e.
  - ➔ This assumption may not hold true for all formulations; if the reference formulation shows higher variability than the test formulation, a *'good'* test will be penalized for the *'bad'* reference.
- All observations made on different subjects are independent.
  - ➔ This assumption should not be a problem, unless you plan to include twins or triplets in your study…

Xover Designs (cont'd)
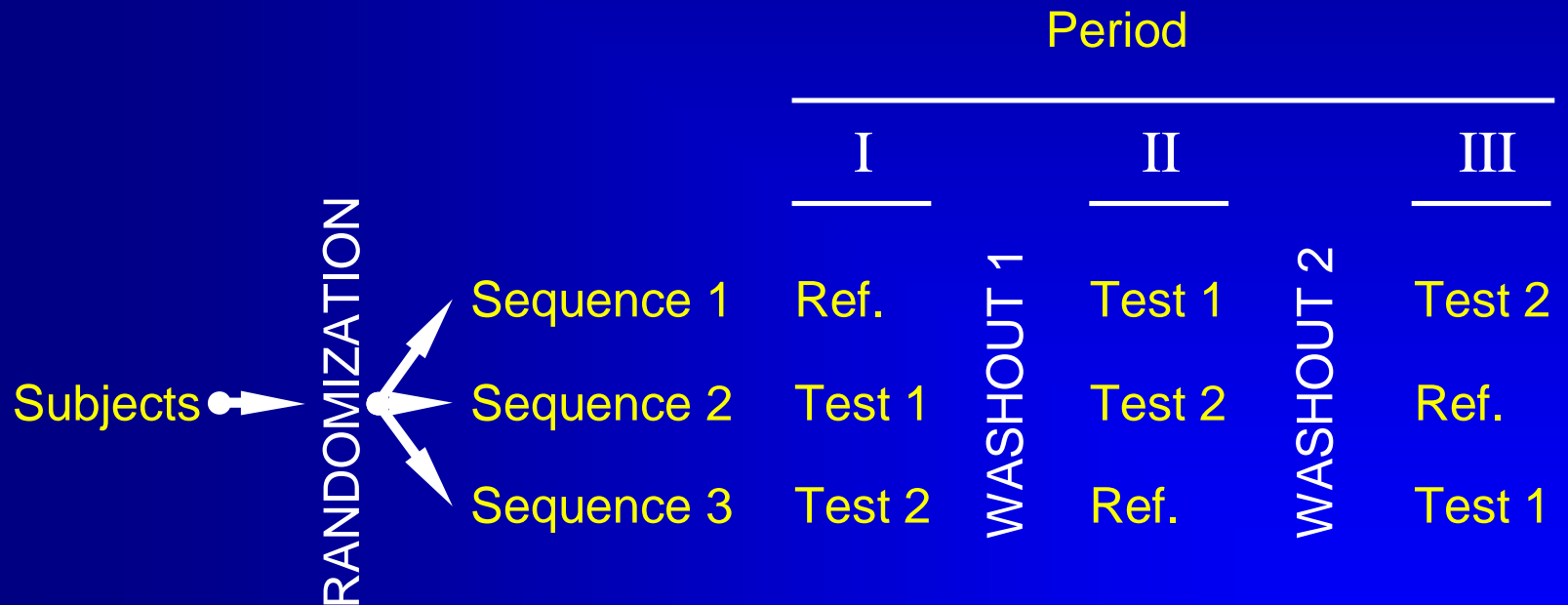
- Standard 2×2×2 design
  - Advantages
    - Globally applied standard protocol for bioequivalence, PK interaction- and food-effect studies.
    - Straigthforward statistical analysis.
  - Disadvantages
    - Not suitable for drugs with long half life ($\rightarrow$ parallel designs).
    - Not optimal for studies in patients with instable diseases ($\rightarrow$ parallel designs).
    - Not optimal if CV uncertain ($\rightarrow$ two-stage sequential design).
    - Not optimal for HVDs/HVDPs ($\rightarrow$ replicate designs).

Xover Designs (cont'd)

- 3×3×3 Latin Square Design

	Period		
	I	II	III
Sequence 1	Ref.	Test 1	Test 2
Sequence 2	Test 1	Test 2	Ref.
Sequence 3	Test 2	Ref.	Test 1

Subjects → RANDOMIZATION

WASHOUT 1 — WASHOUT 2

Xover Designs (cont'd)

- Williams' Design for three treatments

Sequence	Period		
	I	II	III
1	R	T_2	T_1
2	T_1	R	T_2
3	T_2	T_1	R
4	T_1	T_2	R
5	T_2	R	T_1
6	R	T_1	T_2

Xover Designs (cont'd)

- Williams' Design for four treatments

Sequence	Period			
	I	II	III	IV
1	R	T_3	T_1	T_2
2	T_1	R	T_2	T_3
3	T_2	T_1	T_3	R
4	T_3	T_2	R	T_1

HVDs / HVDPs

- For Highly Variable Drugs / Drug Products (HVDs/HVDPs) it may be almost impossible to show BE with a reasonable sample size.

- The common 2×2 Xover assumes Independent Identically Distributions (IID), which may not hold.
  If *e.g.*, the variability of the reference is higher than the one of the test, one obtains a high common (pooled) variance and the test will be penalized for the 'bad' reference.

HVDs / HVDPs (2×2)

Power to show BE with 40 subjects if CV_{intra} 30–50%
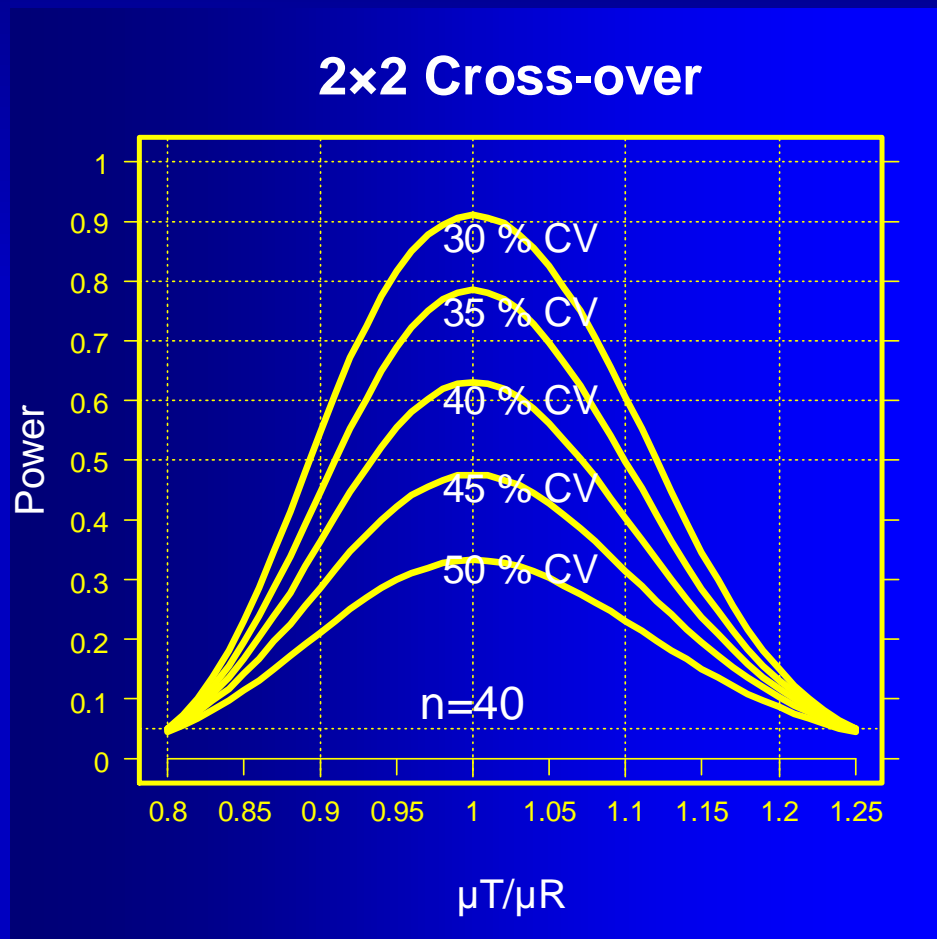
μ_T/μ_R 0.95, CV_{intra} 30%
 → power 0.816
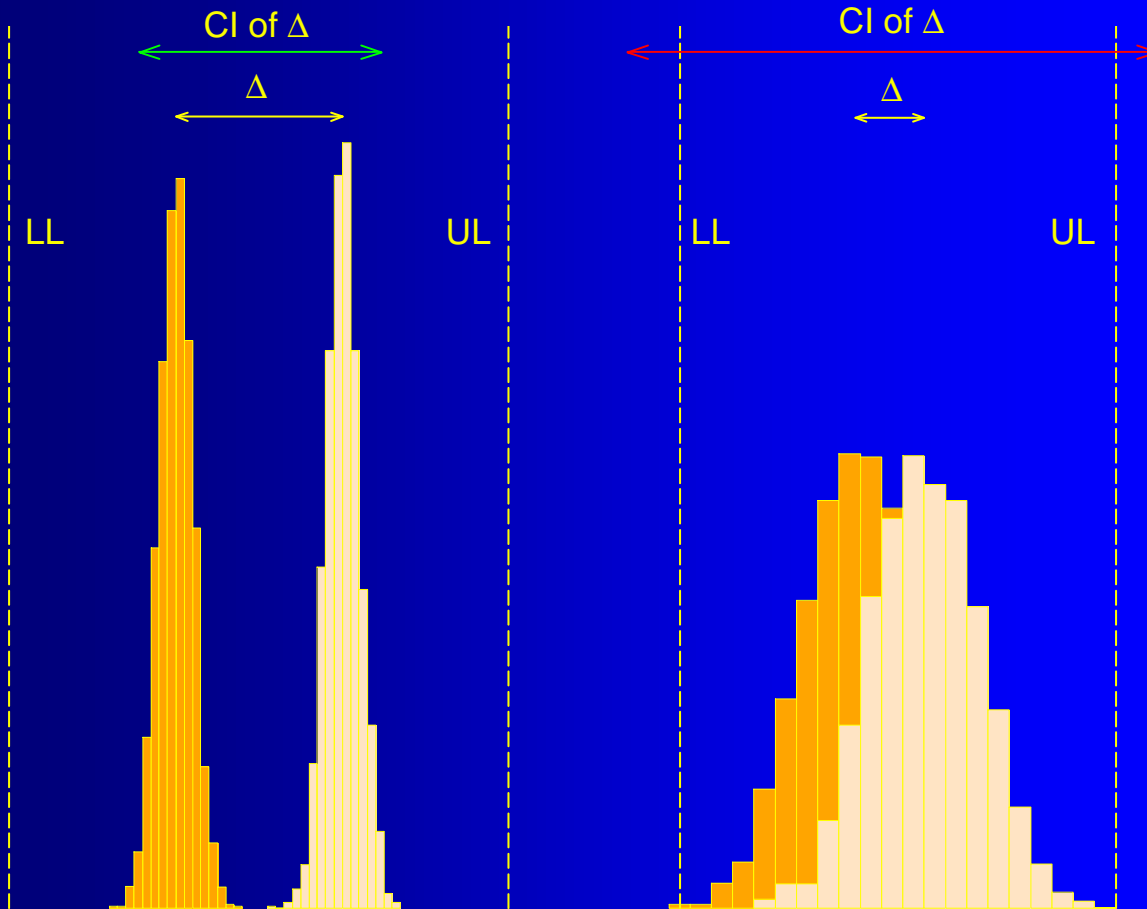μ_T/μ_R 1.00, CV_{intra} 45%
 → power 0.476
 < *Roulette* (0.486!)

μ_T/μ_R 0.95, CV_{intra} 50%
 → n=98 (power 0.803)

2×2 Cross-over

30 % CV
35 % CV
40 % CV
45 % CV
50 % CV
n=40
Power
µT/µR

HVDs / HVDPs (2×2)

Modified from Fig. 1
L Tóthfalusi, L Endrenyi and
A García Arieta
*Evaluation of Bioequivalence
for Highly Variable Drugs
with Scaled Average Bio-
equivalence*
Clin Pharmacokinet 48,
725–43 (2009)

Counterintuitive
concept of BE:
Two formulations with
a large difference in
means are declared
bioequivalent if vari-
ances are low, but not
bioequivalent – even
if the difference is
quite small – due to
high variability.

Replicate Xover Designs

- Each subject is randomly assigned to sequences, where *at least one* of the treatments is administered *at least twice*.
  - Not only the *global within-subject variability*, but also the *within-subject variability per treatment* may be estimated.
  - Smaller subject numbers compared to a standard 2×2×2 design – but outweighed by an increased number of periods.
    Note: Similar *overall* number of administered treatments!

Replicate Xover Designs

- Two-sequence three-period*⁾
  TRT
  RTR
- Two-sequence four-period*⁾
  TRTR
  RTRT

*⁾ Recommended designs:
László Tóthfalusi
Scaled Average Bioequivalence to Evaluate Bioequivalence of Highly Variable Drugs
Dissolution Testing, Bioavailability & Bioequivalence Conference
Budapest, May 24th, 2007

Replicate Xover Designs

- … and many others (examples)
  - Two-period
    - TT | RR | RT | TR
      Balaam's design: not recommended by the FDA – but stated in ANVISA's GL
  - Three-period
    - TRR | RTR
    - TRR | RTR | RRT
      FDA's partial replicate design
  - Four-period
    - TTRR | RRTT
    - TRTR | RTRT | TTRR | RRTT
      completely randomized

Replicate Xover Designs

- Required for
  - Reference-scaled average bioequivalence for AUC and C_{max} (FDA: RSABE)
  - Average BE with expanding limits for C_{max} (EMA 2010: ABEL)
  - Widening of the AR to 75–133% for C_{max} (EMEA's 2001 NfG, Q&A document 2006)

- Advantages
  - Some experience from FDA's initiative on Population Bioequi-valence (PBE) and Individual Bioequivalence (IBE)
  - Mentioned in RSA's GL; FDA's API-specific GLs and EMA
  - Scaling of different metrics acceptable in some countries (FDA and RSA: AUC and C_{max}, EMA: C_{max} only)
  - Handling of outliers (Subject-by-Formulation Interaction may be ruled out).

Replicate Xover Designs

- Disadvantages
  - Statistical analysis quite complicated (especially in the case of drop-outs and if RSABE is the target) – not available in standard software.
  - Many publications, but still no agreement on methodology (!)
  - SAS-code published by the FDA in April 2010:
    http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM209294.pdf
  - For the EMA it has to be shown that CV_{WR} >30% *is not* caused by outliers!
  - SAS-code and two example datasets published by the EMA in March 2011:
    http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002963.pdf
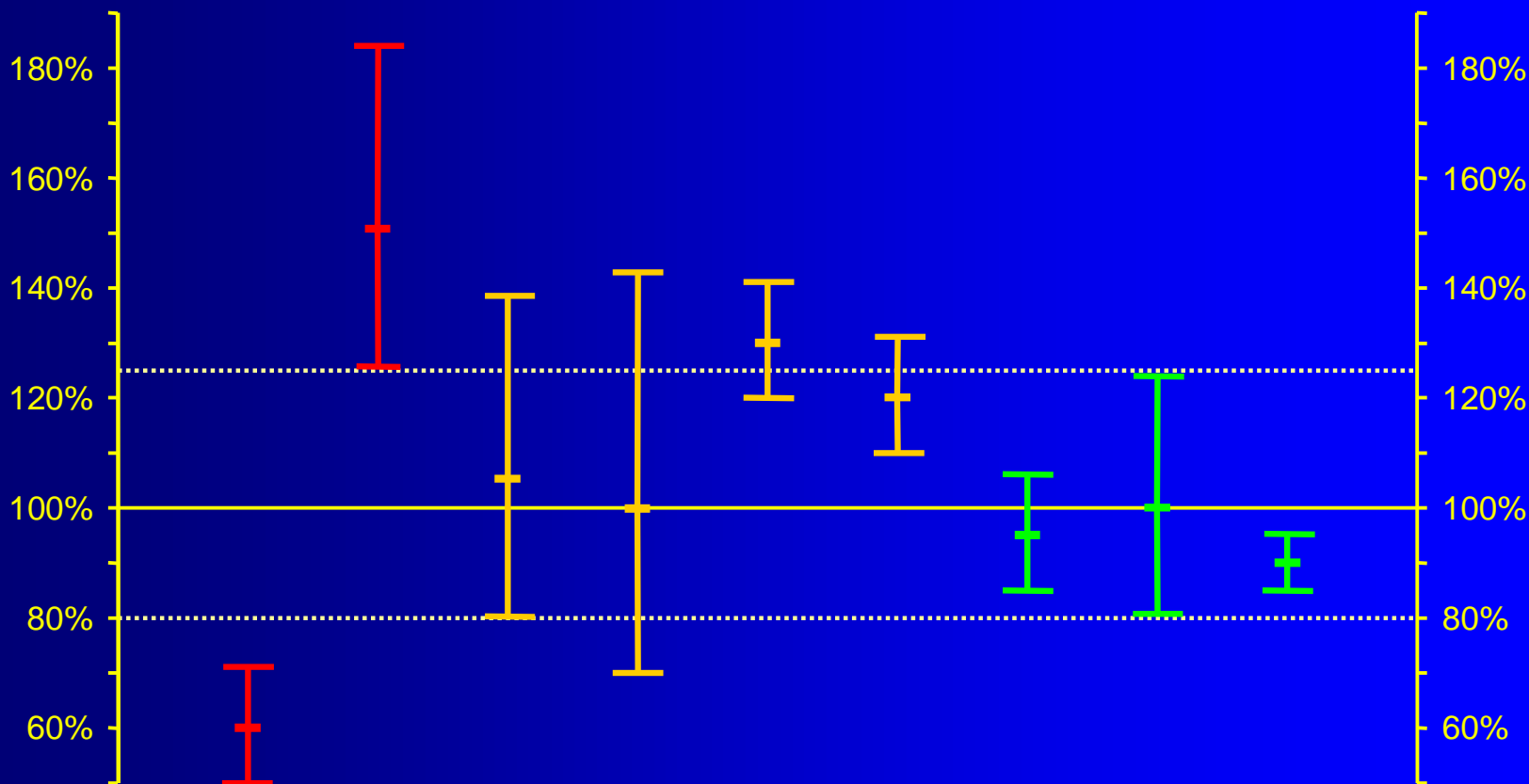
BE Evaluation

● Based on the design set up a statistical model.

● Calculate the test/reference ratio.

● Calculate the 90% confidence interval (CI) around the ratio.

● The *width* of the CI depends on the variability observed in the study.

● The *location* of the CI depends on the observed test/reference-ratio.

BE Assessment

- Decision rules based on the CI and the Acceptance Range (AR)
  - CI *entirely outside* the AR:
    Bioinequivalence proven
  - CI *overlaps* the AR (lies *not entirely within* the AR):
    Bioequivalence not proven
  - CI lies *entirely within* the AR:
    Bioequivalence proven

BE Assessment

EMA *vs.* Rest of the World

- EMA BE GL (2010), 4.1.8 Evaluation / Statis-
tical analysis:

  The terms to be used in the ANOVA model are usually sequence, subject within sequence, period and formulation. *Fixed effects, rather than random effects, should be used for all terms.*

- Adapt your standard setup:

  - SAS: `Proc GLM` instead of `Proc MIXED` (*i.e.*, incomplete data are dropped).
  - Phoenix/WinNonlin: Don't use the default settings!

EMA *vs.* Rest of the World

Sample Size Estimation

- Introduction
  - 'Classical' sample size estimation in BE
    - Patient's & producer's risk
    - Power in study planning
- Details ($\rightarrow$ day 2)
  - Uncertainties in assumptions
    - Variability, Test/Reference-ratio
    - Sensitivity analysis
  - Recent developments
    - Review of guidelines (Two-Stage Design, Replicates)

α- vs. β-Errors

- All formal decisions are subjected to two types of error:
  - Error Type I (α-Error, Risk Type I)
  - Error Type II (β-Error, Risk Type II)
    Example from the justice system:

Verdict	Defendant innocent	Defendant guilty
Presumption of innocence not accepted (guilty)	Error type I	Correct
Presumption of innocence accepted (not guilty)	Correct	Error type II

α- vs. β-Errors

● Or in more statistical terms:

Decision	Null hypothesis true	Null hypothesis false
Null hypothesis rejected	Error type I	Correct (H_a)
Failed to reject null hypothesis	Correct (H_0)	Error type II

● In BE-testing the null hypothesis is bioinequivalence ($\mu_1 \neq \mu_2$)!

Decision	Null hypothesis true	Null hypothesis false
Null hypothesis rejected	Patients' risk	Correct (BE)
Failed to reject null hypothesis	Correct (not BE)	Producer's risk

α- vs. β-Errors

- α-Error: Patient's Risk to be treated with a bioinequivalent formulation (H_0 falsely rejected)
  - BA of the test compared to reference in a *particular* patient is risky *either* below 80% *or* above 125%.
  - If we keep the risk of particular patients at α 0.05 (5%), the risk of the entire population of patients (<80% *and* >125%) is 2×α (10%) – expressed as: 90% CI = 1 − 2×α = 0.90

95% one-sided CI	**95% one-sided CI**	**90% two-sided CI = two 95% one-sided**
0.6 0.8 1 1.25 1.67	0.6 0.8 1 1.25 1.67	0.6 0.8 1 1.25 1.67
particular patient	particular patient	population of patients

α- vs. β-Errors
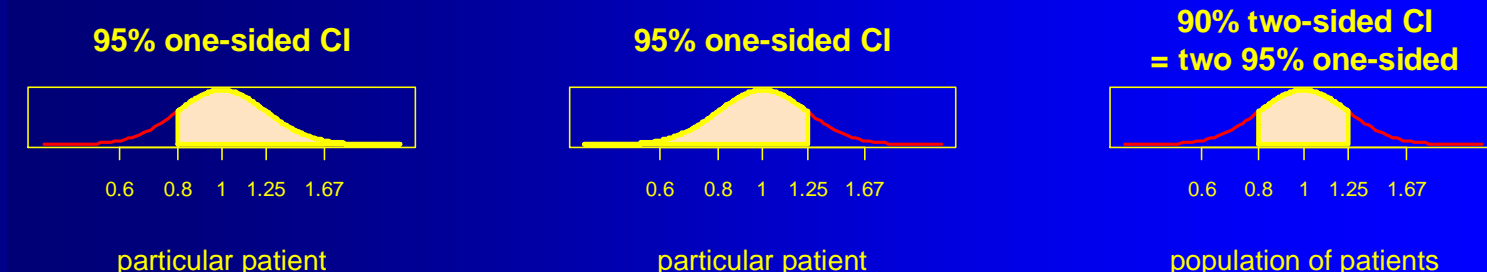
- β-Error: Producer's Risk to get no approval for a bioequivalent formulation (H_0 falsely not rejected)
  - *Set* in study planning to ≤0.2, where power = $1 - \beta$ = ≥80%
  - If power is set to 80 %

    One out of five studies will fail just by chance!

α 0.05	BE
not BE	β 0.20

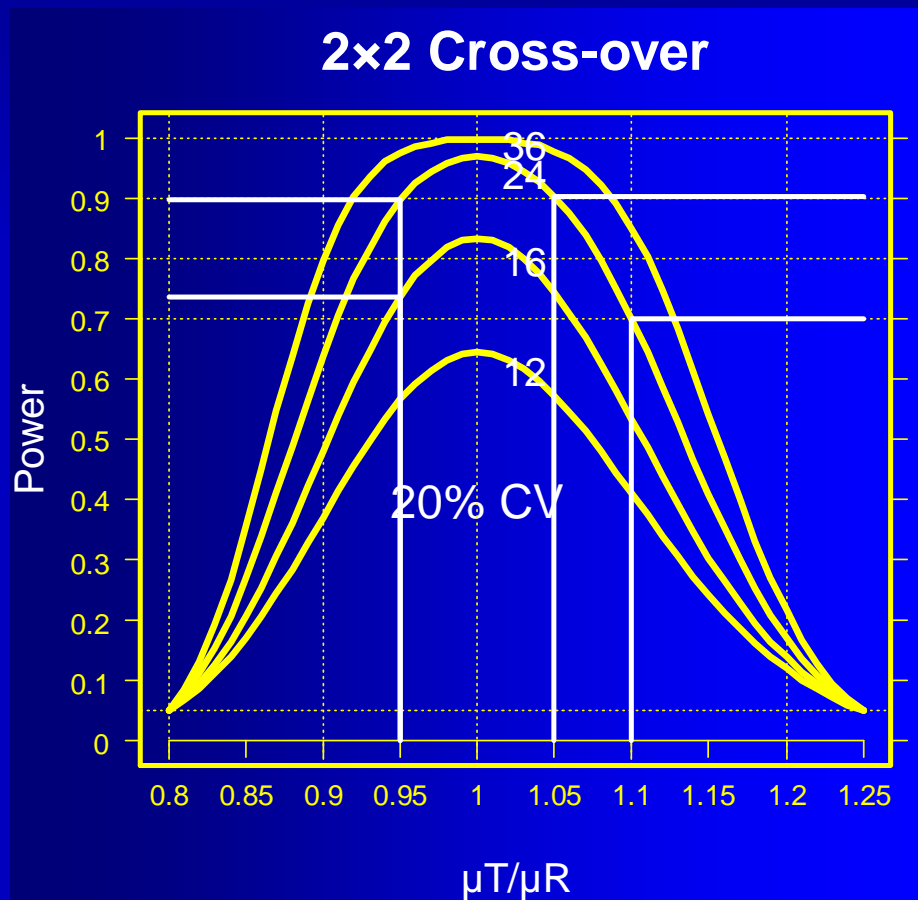 - *A posteriori* (*post hoc*) power does not make sense!
    Either a study has demonstrated BE or not.

Power $(1 - \beta)$

Power to show BE with 12 – 36 subjects for CV_{intra} 20%

n	24	↓	16:
power	0.896	→	0.735

μ_T/μ_R	1.05	↓	1.10:
power	0.903	→	0.700

**2×2 Cross-over**

Power *vs.* Sample Size

- It is not possible to calculate the required sample size *directly*!

- Power is calculated instead; the smallest sample size which fulfills the minimum target power is used

  - Example: α 0.05, target power 80% (β 0.2), T/R 0.95, CV_{intra} 20% $\rightarrow$ minimum sample size 19 (power 81%), rounded *up* to the next *even* number in a 2×2 study to get balanced sequences (power 83%)

n	power
16	73.54%
17	76.51%
18	79.12%
19	81.43%
20	83.47%

Power *vs.* Sample Size

2×2 cross-over, T/R 0.95, 80%–125%, target power 80%

◆ sample size ── power ── power for n=12
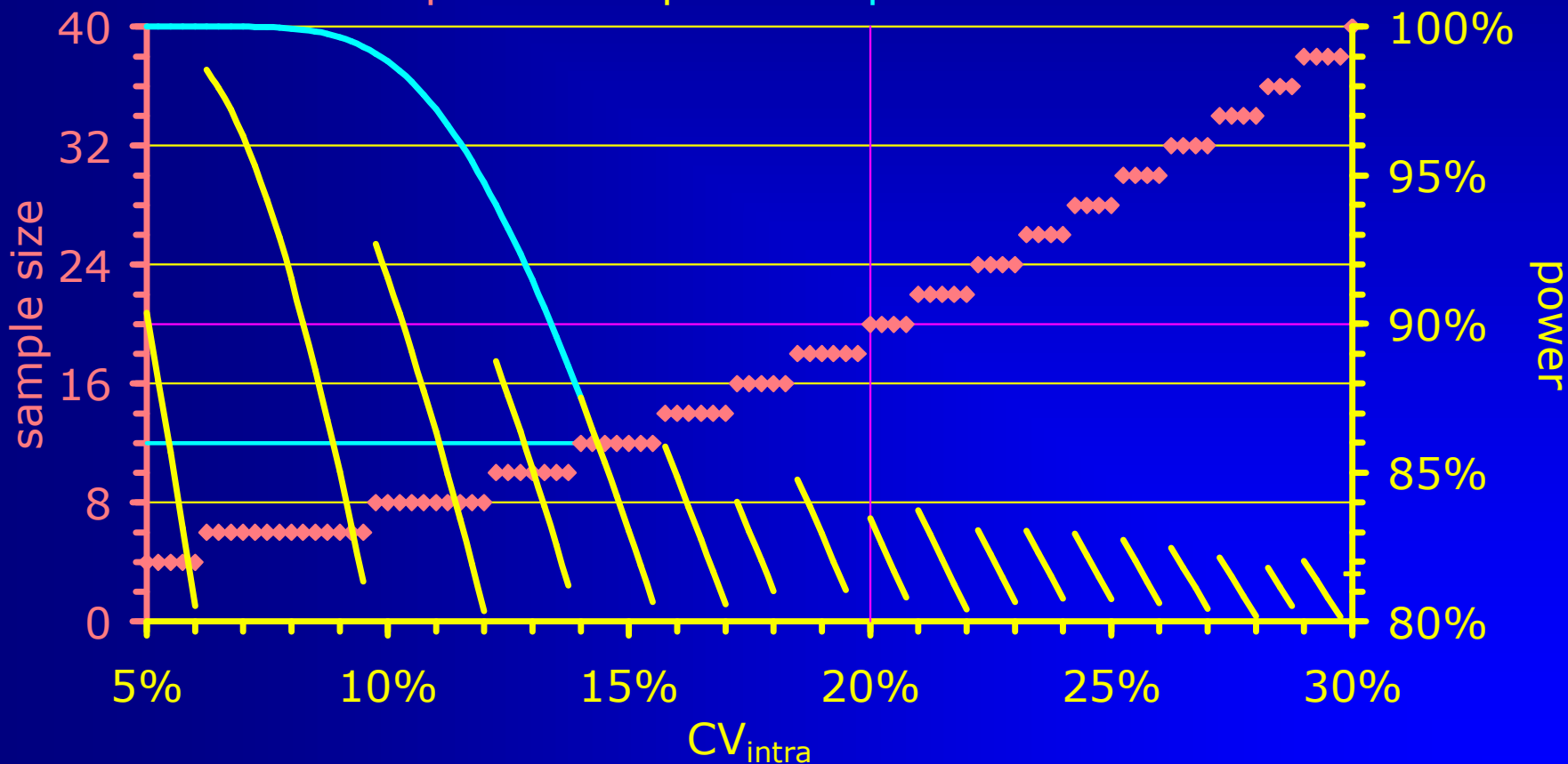
Add-on / Two-Stage Designs

- History / early approaches
  - Add-on studies
  - Problems with α-inflation
- Uncertain Uncertain CV_{intra} …
- Recent developments
  - Review of guidelines
  - Two-stage sequential designs

Add-on / Two-Stage Designs

- Sometimes properly planned and executed studies fail due to
  - Pure chance (producer's risk hit)
  - False assumptions about variability and/or T/R-ratio
  - Poor study conduct (increasing variability)
  - 'True' bioinequivalence
- The patient's risk must be preserved
  - Already noticed at Bio-International Conferences (1989, 1992) and guidelines from the 1990s

Sequential Designs

- Methods by Potvin *et al.* (2008) promising
  - Supported by 'The Product Quality Research Institute' (members: FDA/CDER, Health Canada, USP, AAPS, PhRMA, …)
    - Acceptable by US-FDA
    - Canada? Or Gould (1995) mandatory?
    - Acceptable as a Two-Stage Design in the EU
    - Three of BEBAC's protocols approved by German BfArM, one study accepted

Potvin D, Diliberti CE, Hauck WW, Parr AF, Schuirmann DJ, and RA Smith
Sequential design approaches for bioequivalence studies with crossover designs
Pharmaceut Statist 7/4, 245–62 (2008), DOI: 10.1002/pst.294
http://www3.interscience.wiley.com/cgi-bin/abstract/115805765/ABSTRACT

Sequential Designs

- Open issues
  - Feasibility / futility rules
  - Arbitrary PE and/or power; adaption for stage 1 PE
  - Dropping a candidate formulation from a higher-order cross-over
  - Application to replicated designs (for HVDs/HVDPs)

Open Issues

● Replicated designs (HVDs/HVDPs)

- ■ Nothing published!

- ■ Statistical model?

- ■ Although EMA assumes equal variances of formulations (Q&A document Jan 2010) that does not reflect the 'real world' (quite often $\sigma^2_{WR} > \sigma^2_{WT}$)

- ■ If you set up simulations, allow for different variances of test and reference

Outliers

● Problems

■ Parametric methods (ANOVA, GLM) are very sensitive to outliers

■ A single outlier may underpower a properly sized study

■ Exclusion of outliers only possible if procedure stated in the protocol, *and* reason is justified, *e.g.*,

➢ Lacking compliance (subject did not take the medication),

➢ vomiting (up to 2 × t_{max} for IR, at all times for MR),

➢ analytical problems (*e.g.*, interferences in chromatography);

➢ not acceptable if based on statistical grounds only.

Nuisances in BE Studies

- Period effect
- Sequence (aka unequal carry-over) effect
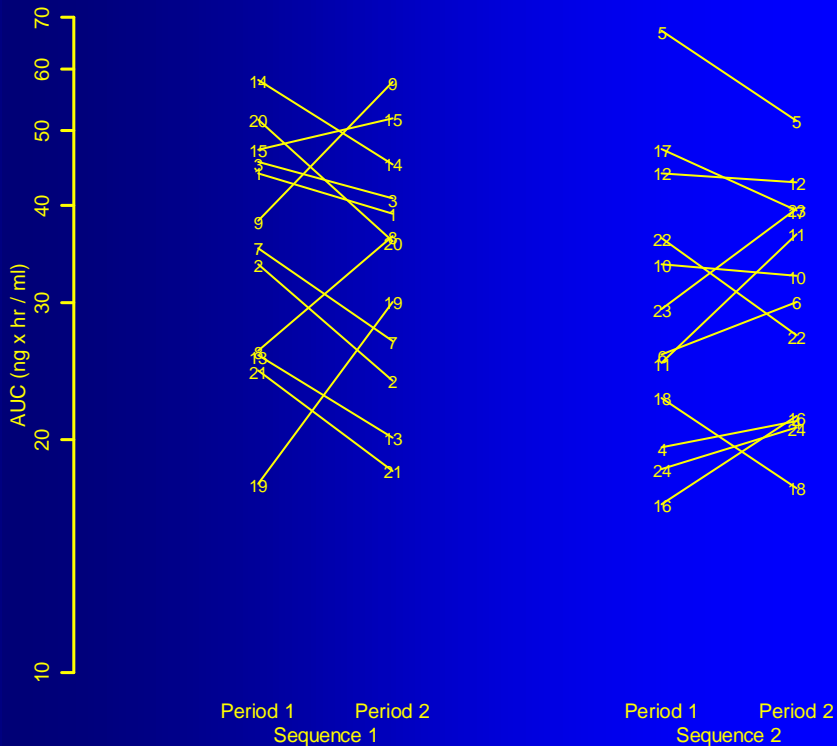- Group effect

Period effect

- Original data
  - AUC(p_2/p_1): 98.4%
  - Period: *p* 0.7856   (95% CI: 87.4% −110.8%)
  - Sequence: *p* 0.3239   (95% CI: 86.0% −154.8%)
  - GMR: 96.5%   (90% CI: 87.5% −106.5%)
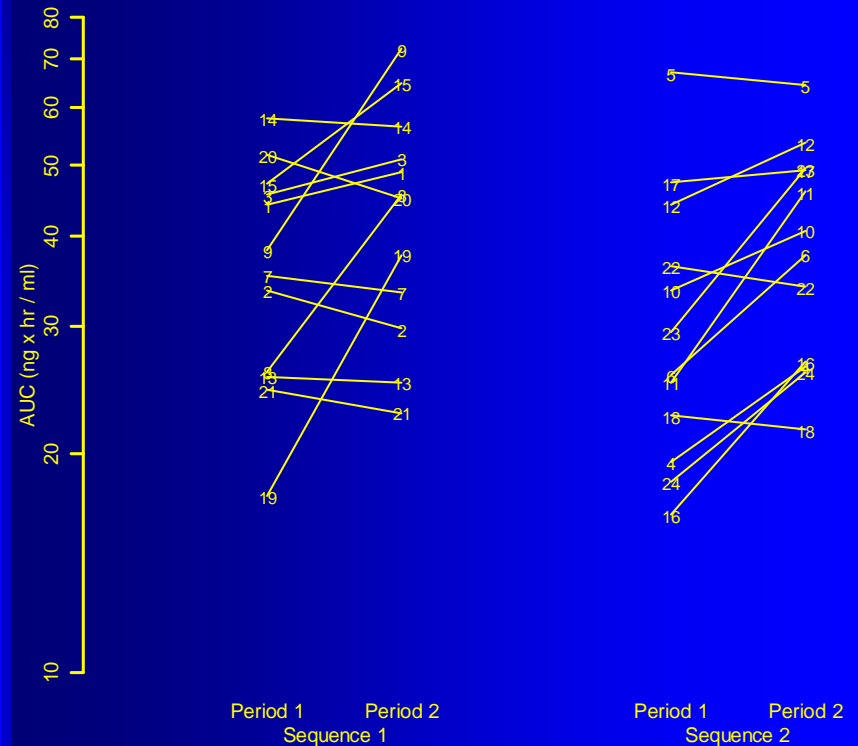- Modified data (p_2 125% of original values)
  - AUC(p_2/p_1): 123.0%
  - Period: *p* 0.0015   (95% CI: 109.3% −138.5%)
  - Sequence: *p* 0.3239   (95% CI: 86.0% −154.8%)
  - GMR: 96.5%   (90% CI: 87.5% −106.5%)

Period effect
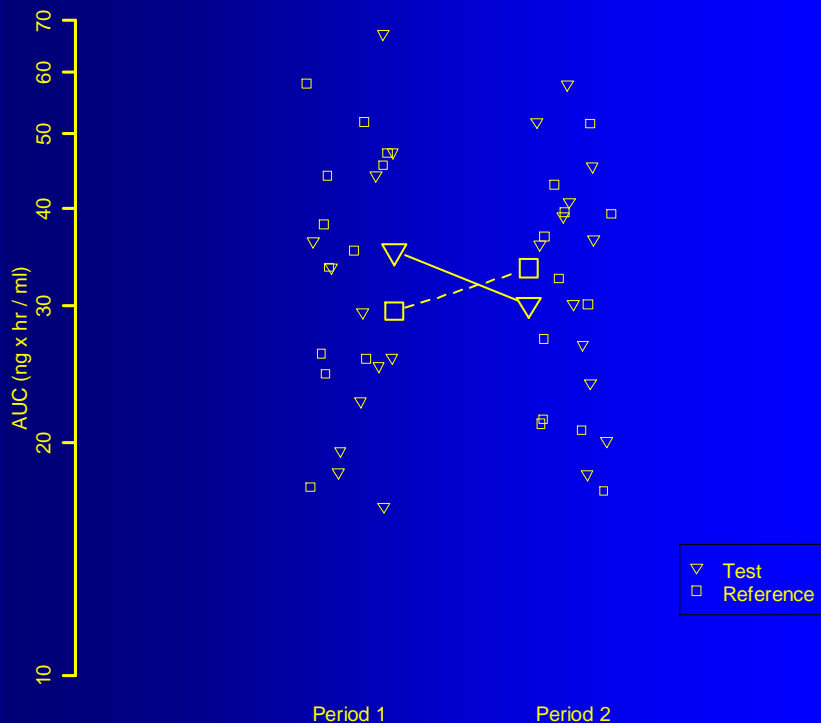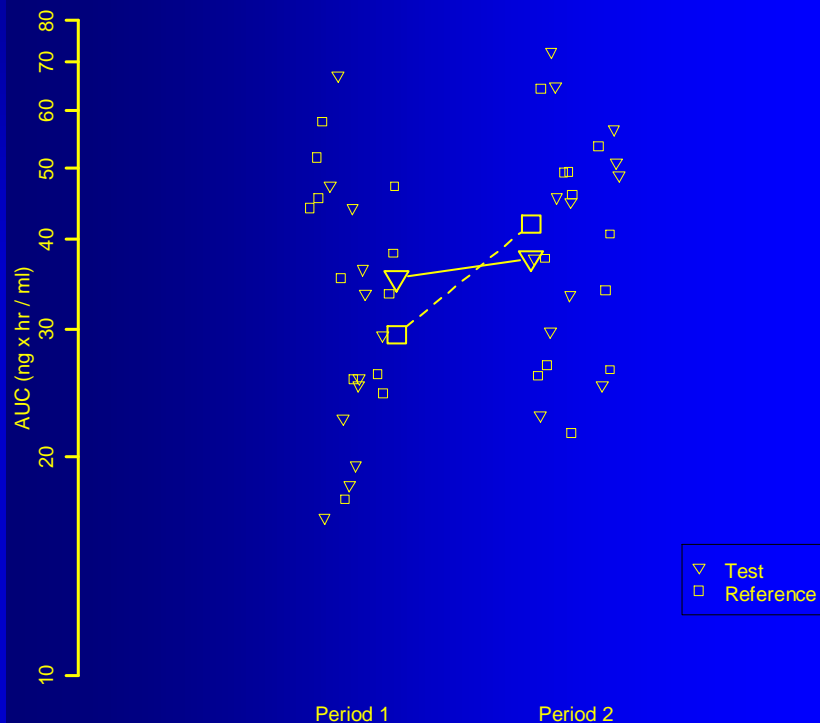
Period effect
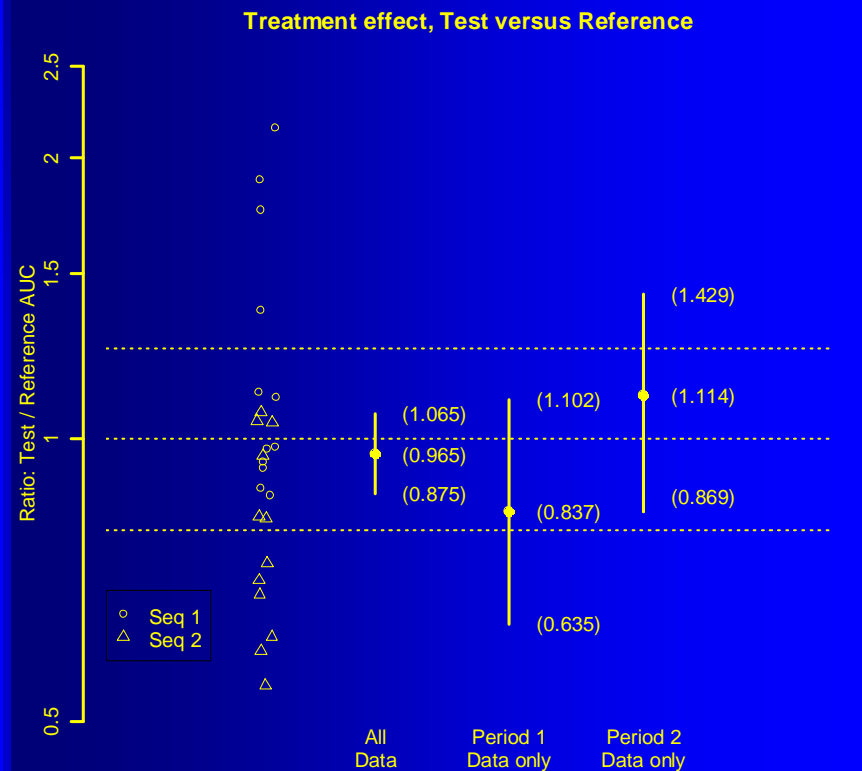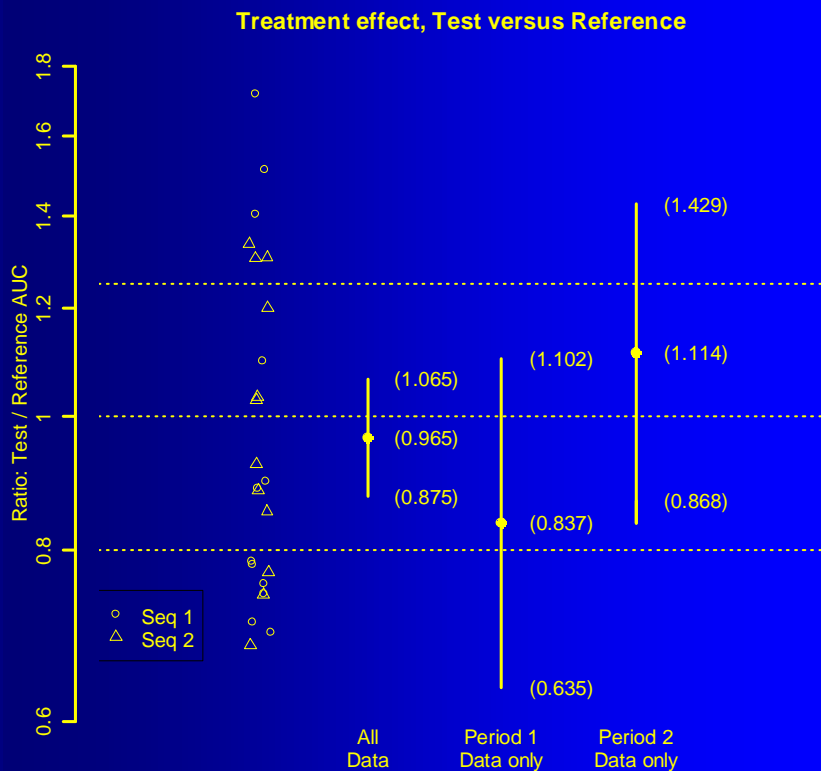
Period effect

Sequence effect

- In a 'standard' 2×2 cross-over design
  - the sequence effect is confounded with
    - the carry-over effect, and
    - the formulation-by-period interaction
  - Therefore, a statistically significant sequence effect could indicate that there is
    - a true sequence effect,
    - a true carry-over effect,
    - a true formulation by period interaction, or
    - a failure of randomization

Sequence effect

- 'Two-stage analysis'[1] was – and regrettably still is – often applied
  - Test for a significant sequence effect at α 0.10
  - If a significant sequence effect is found, evaluation of the first period as a parallel design
- This procedure was shown to be statistically flawed[2]

[1] **JE Grizzle**
The two-period change over design and ist use in clinical trials
Biometrics 21, 467–80 (1965)

[2] **P Freeman**
The performance of the two-stage analysis of two-treatment, two-period cross-over trials
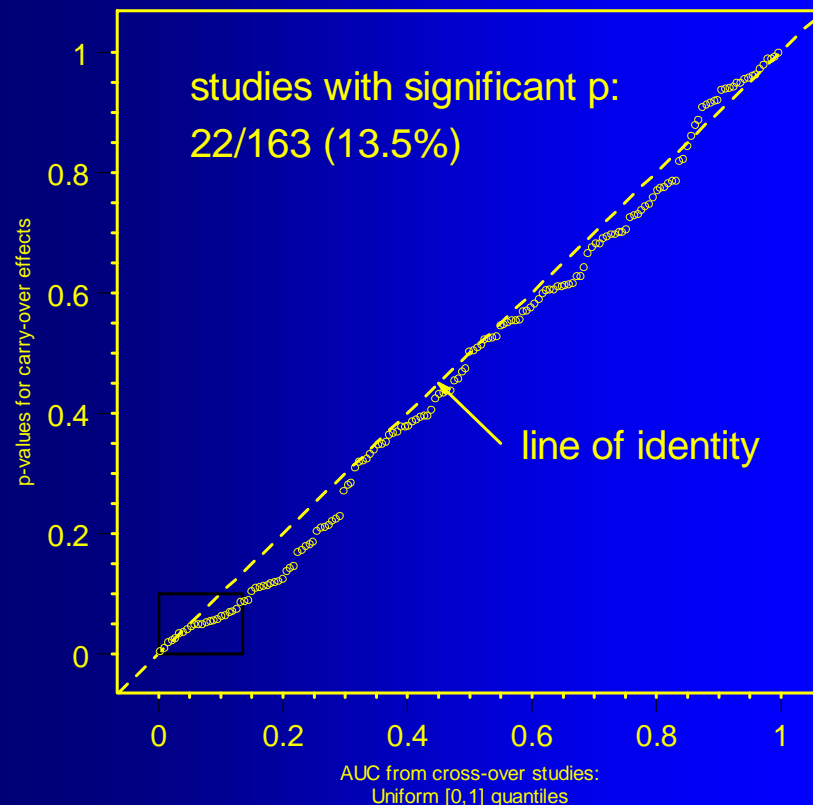Statistics in Medicine 8, 1421–32 (1989)

Sequence effect

- In a large metastudy (n=420) significant sequence effects were found at $\approx \alpha$, both for AUC and C_{max}[*]
  - 2×2 studies (n=324)
    - AUC: 34/324 (10.5%) C_{max}: 37/324 (11.4%)
  - 6×3 studies (n=96)
    - AUC: 4/96   (4.2%) C_{max}:   4/96   (4.2%)
  - For both metrics the distribution of *p* values followed closely Uniform [0,1]

[*] **D'Angelo G, Potvin D and J Turgeon**
    Carry-over effects in bioequivalence studies
    J Biopharm Stat 11, 35–43 (2001)

Sequence effect

- These results could be con-firmed (20 published studies, 143 studies from BEBAC's database; AUC):
  - Significant sequence effects in 22/163 studies (13.5%)
- Significant sequence effects in properly planned studies should be considered a statis-tical artefact (significant results are obtained in α of studies)

studies with significant p:
22/163 (13.5%)

line of identity

p-values for carry-over effects

AUC from cross-over studies:
Uniform [0,1] quantiles

Sequence effect

- Conclusions
  - No valid statistical procedure exists to *correct* for a true sequence/carry-over effect
  - A true sequence/carry-over is *highly unlikely* in a BE study if
    - the study is performed in healthy subjects,
    - the drug is not an endogenous entity, and
    - an adequate washout period was maintained (no predose concentrations >5% of C_{max} observed).
  - Testing for a sequence effect is futile!

Sequence effect

- Conclusions (cont'd)
  - EMA GL on BE (2010)
    - A test for carry-over should not be performed and no decisions regarding the analysis (e.g. analysis of the first period, only) should be made on the basis of such a test. The potential for carry-over can be directly addressed by examination of the pre-treatment plasma concentrations in period 2 (and beyond if applicable).

Group effect

- More than one group of subjects
  - 'If a crossover study is carried out in two or more groups of subjects (*e.g.*, if for logistical reasons only a limited number of subjects can be studied at one time), the statistical model should be modified to reflect the multigroup nature of the study. In particular, the model should reflect the fact that the periods for the first group are different from the periods for the second group.'

FDA, Center for Drug Evaluation and Research (CDER)
Guidance for Industry: Statistical Approaches to Establishing Bioequivalence (2001)

Group effect

- More than one group of subjects
  - Cases where '… the study is carried out in two or more groups and those groups are studied at diffe-rent clinical sites, or at the same site but greatly separated in time (months apart, for example) […] should be discussed with the appropriate CDER review division.'
  - EMEA BA/BE (2001), BE GL (2010)
    - The study should be designed in such a way that the formulation effect can be distinguished from other effects.

Group effect

- Increasing number of referrals (deficiency letters) from
  - Canada
  - Gulf States (Saudia Arabia, Emirates, Oman)
- Extended Statistical model (fixed effects in ANOVA)
  - Group
  - Group × Treatment Interaction
  - If both terms are not significant (p>0.05), pooling of groups is justified.

Group effect

- Recommendations
  - If ever possible, avoid multiple groups
  - Keep the time interval between groups as short as possible
  - Do not split the study into equally sized groups
    - Perform at least one group in the maximum capacity of the clinical site
      (*e.g.*, 24 & 8 instead of 16 & 16 for a total of 32)
    - If a significant group and/or group × treatment interaction is found (preventing a pooled analysis), it may still be possible to demonstrate BE in the largest group

Group effect

- Example
  - T/R 0.95, CV 22.5%, sample size to obtain at least 90% power estimated with 32
  - Two groups due to logistic reasons
  - Assumptions on T/R and CV exactly hold in the actual study, *but*
  - pooling not allowed (significant effect)
    - If group sizes 16 & 16
      Power to show BE is 62.10%
    - If group sizes 24 & 8
      Power to show BE in the larger group is 82.27%

Thank You!
PK–NCA, PK based Design, Biostatistics
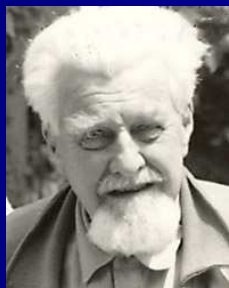## *Open Questions?*

Helmut Schütz

BEBAC

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

To bear in Remembrance...

The fundamental cause of trouble in the world today is
that the stupid are cocksure
while the intelligent are full of doubt.   *Bertrand Russell*

It is a good morning exercise for a research scientist
to discard a pet hypothesis every day before
breakfast.
It keeps him young.   *Konrad Lorenz*

If you shut your door to all errors
truth will be shut out.

Rabindranath Tagore