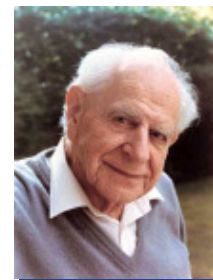


# Basic Statistics for BE

Helmut Schütz

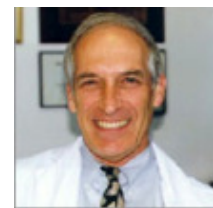
# Keep in memory...

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

# Why logarithmic transformation of the data?

Like most biologic variables PK metrics (e.g.,  $AUC$ ,  $C_{max}$ ) follow a *log-normal* distribution

- If they would follow a *normal distribution* ('bell curve') the range of possible values *by definition* would be  $[-\infty, +\infty]$ 
  - However, negative concentrations are not possible
- The *log-normal* distribution covers a range of  $[>0, +\infty]$ 
  - In statistical methods we apply in bioequivalence (e.g., the ANOVA) we need *normal* distributed data
  - If we log-transform the original data we get exactly what we need
    - Always use the natural logarithm (base e) – not the decadic logarithm (base 10)
  - At the end of the analysis we back-transform the result (e.g., from the 90% confidence interval of  $[-0.1832, +0.0432]$  we get  $[e^{-0.1832}, e^{+0.0432}]$  or [83.26%, 104.41%])

# Why logarithmic transformation of the data?

## Justification

- The basic equation of PK (after an extravascular dose) is
 
$$AUC = f \times D / CL$$
- In BE we are interested in the fraction absorbed ( $f$ ), which leads to
 
$$f = AUC \times CL / D$$
  - which is a *multiplicative* model
  - We get an *additive* model (needed in ANOVA) by taking logs
 
$$\log(f) = \log(AUC) + \log(CL) - \log(D)$$
- Actually we are interested in comparing  $f_{Test}$  with  $f_{Reference}$ 
  - In the study we obtain  $AUC_{Test}$  and  $AUC_{Reference}$
  - We assume (!) that  $D_{Test} = D_{Reference}$  and  $CL_{Test} = CL_{Reference}$
  - Given that, we get
    - $\log(f_{Test}) - \log(f_{Reference}) = \log(AUC_{Test}) - \log(AUC_{Reference})$  or
    - $f_{Test} / f_{Reference} = AUC_{Test} / AUC_{Reference}$

# Why logarithmic transformation of the data?

## Example

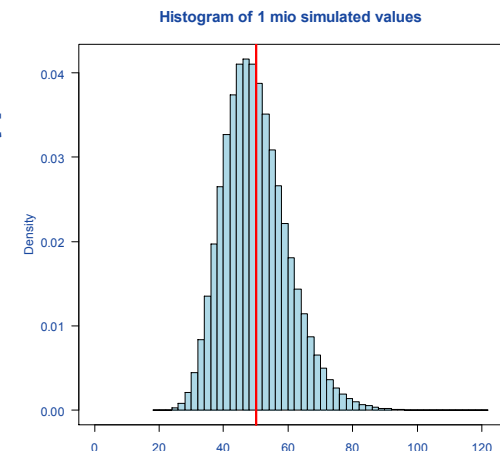
	Reference	log(R)	Test	log(T)	$\Delta \log$	Ratio T/R
<i>AUC</i>	200	5.2983	190	5.2470	-0.0513	95.00%
<i>CL</i>	0.2	-1.6094	0.2	-1.6094		
<i>D</i>	50	3.9120	50	3.9120		
<i>f</i>	80%	-0.2231	76%	-0.2744	-0.0513	95.00%

- The Test has a lower absorption (76%) than the Reference (80%)
  - We assume that the administered doses are equal, as are the clearances (property of the drug, not the formulation)
  - Then we can estimate  $f_{Test}/f_{Reference}$  from the ratio of *AUCs* or the difference of log-transformed *AUCs* ( $\Delta \log$ )
- Practically the analysis is done on log-transformed data
  - We get  $f_{Test}/f_{Reference}$  by the back-transformation of  $\Delta \log$ :  $e^{-0.0513} = 95\%$

# Why geometric means instead of arithmetic means?

In statistics we need an accurate ('unbiased') estimate of the location

- The best unbiased estimate of the location of the *normal distribution* is the *arithmetic mean*
- The best unbiased estimate of the location of the *log-normal distribution* is the *geometric mean*
  - Since we know that concentrations and most derived PK metrics (exception:  $t_{max}$ ) follow a *log-normal distribution* we have to use their geometric means
  - The log-normal distribution is skewed to the right
    - The arithmetic mean is always larger than the geometric mean
    - If we would use the arithmetic mean, the estimate would be positively biased



# Descriptive statistics (transformed and untransformed)

In order to describe the data accurately we have to use suitable descriptive statistics

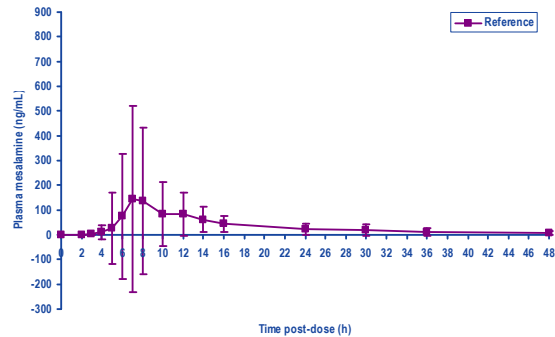
- If we report a certain location (mean, median, ...) and a dispersion (standard deviation, CV, percentiles, ...) we *implicitly* assume a specific distribution
- Arithmetic mean, standard deviation
  - normal distribution (*wrong* in PK...)
- Geometric mean, CV
  - log-normal distribution (concentrations,  $C_{max}$ , AUC, ...)
  - back-transformed arithmetic mean of log-transformed data = geometric mean of raw data
- Median, percentiles, range
  - discrete distribution ( $t_{max}$ ,  $t_{lag}$ )

	raw	log
	1.0000	0.0000
	2.0000	0.6931
	3.0000	1.0986
arithm. mean	2.0000	0.5973
geom. mean	1.8171	
$e^{\text{arithm. mean(log)}}$		1.8171

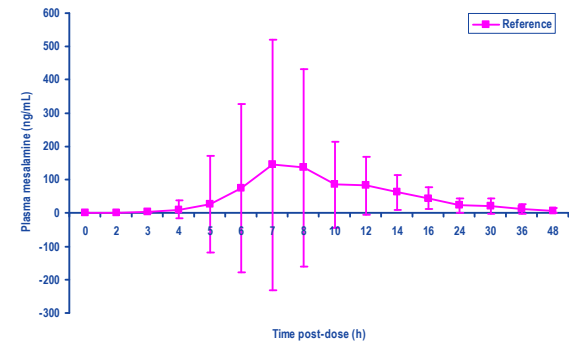
# Descriptive statistics (transformed and untransformed)

## Bad example from the FDA's files (mesalamine, n = 238)

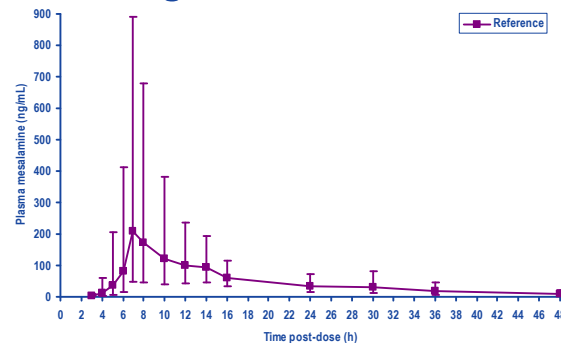
- Wrong: arithmetic means  $\pm$  SD



line plot instead of XY-plot



- Correct: geometric means  $\pm$  SD





# What does the 90% confidence interval mean?

From the study (in statistical terms a ‘sample’) we

- estimate a mean treatment effect  
(in BE the point estimate of the Test/Reference ratio)
- The PE is the best unbiased estimate of the treatment effect in the population of patients

However, we don't know the ‘true’ value

- A confidence interval around the PE tells us where the ‘true’ value might be
- If we use a 90% confidence interval, a wrong decision (*i.e.*, falsely declaring BE of a product which is not) is possible with  $\alpha$ 
  - $\alpha$  is the probability of the Type I Error (the patient's risk) and commonly fixed at 5%
  - The 90% CI is based on  $100(1 - 2\alpha)$

# Excursion: Error(s)

All *formal* decisions are subjected to two ‘Types’ of Error.

- $\alpha$ : Probability of Type I Error (aka Risk Type I)
- $\beta$ : Probability of Type II Error (aka Risk Type II)

Example from the justice system – which presumes that the defendant is *not guilty*:

Verdict	Defendant <i>innocent</i>	Defendant <i>guilty</i>
Presumption of innocence <i>rejected</i> ( <i>guilty</i> )	wrong	correct
Presumption of innocence <i>accepted</i> ( <i>not guilty</i> )	correct	wrong

# Excursion: Hypotheses

## In statistical terminology

- Null hypothesis ( $H_0$ ): **innocent**
- Alternative hypothesis ( $H_a$  aka  $H_1$ ): **guilty**

Decision	Null hypothesis <i>true</i>	Null hypothesis <i>false</i>
$H_0$ rejected	Type I Error	Correct (accept $H_a$ )
Failed to reject $H_0$	Correct (accept $H_0$ )	Type II Error

In BE the Null hypothesis is bio**in**equivalence ( $\mu_T \neq \mu_R$ )!

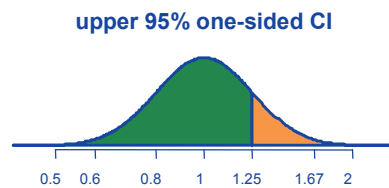
Decision	Null hypothesis <i>true</i>	Null hypothesis <i>false</i>
$H_0$ rejected	Patient's risk ( $\alpha$ )	Correct (BE)
Failed to reject $H_0$	Correct (not BE)	Producer's risk ( $\beta$ )

# Excursion: Type I Error

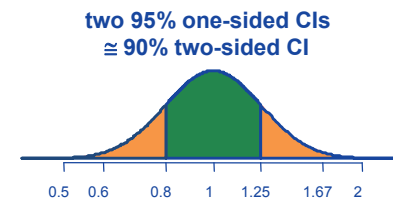
- $\alpha$ : Patient's risk to be treated with an **inequivalent** formulation ( $H_0$  falsely rejected)
- BA of the test compared to reference in a *particular* patient is considered to be risky *either* below 0.80 *or* above 1.25.
    - If we keep the risk of *particular* patients at  $\alpha$  0.05 (5%), the risk of the entire *population* of patients (where  $BA < 0.80$  and  $> 1.25$ ) is  $2\alpha$  (10%) – expressed as a confidence interval:  $100(1 - 2\alpha) = 90\%$ .
    - However, since in a particular patient BA cannot be  $< 0.80$  and  $> 1.25$  *at the same time*, the patient's risk from a 90% CI is still 5%!



5% patients  $< 0.80$



5% patients  $> 1.25$

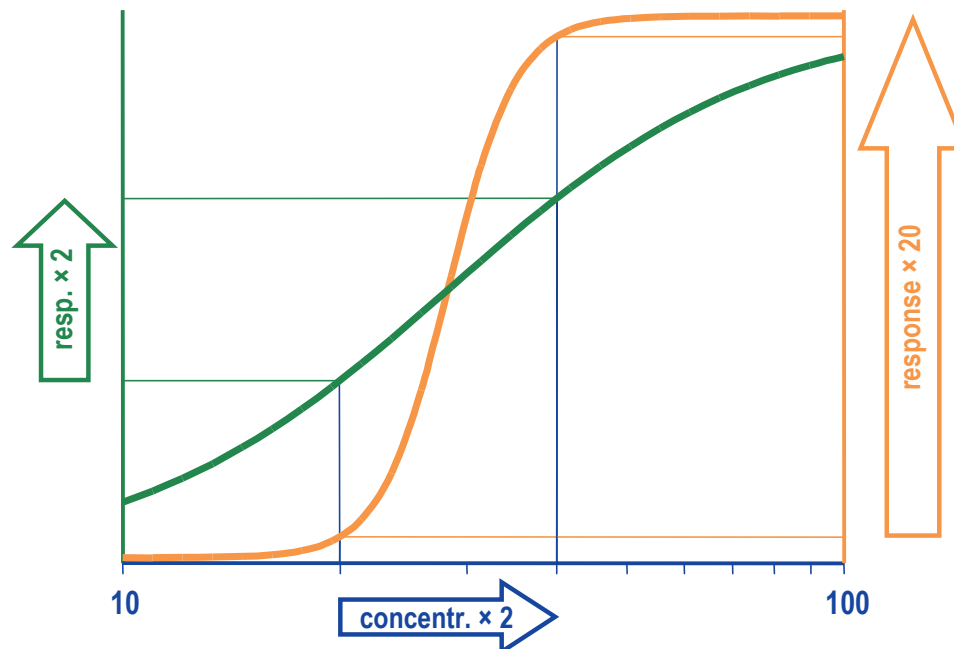


patient population [0.80, 1.25]

# What does $\pm 20\%$ mean and where does it come from?

## Clinically not relevant difference

- Based on PK/PD but extrapolated to similarity of safety and efficacy in the patient population
  - Depends on the dose-response curve! NTID (**steep curve**), HVD (**flat curve**):



# What does $\pm 20\%$ mean and where does it come from?

## Clinically not relevant difference

- Predefined by the authority
  - A difference  $\Delta$  of  $\leq 20\%$  is considered to be clinically not relevant for ‘uncomplicated drugs’
    - The limits  $[L, U]$  of the acceptance range for BE are fixed to  $\log(1 - \Delta) = \log((1 - \Delta)^{-1})$  or  $L \sim -0.2231$  and  $U \sim +0.2231$ , which are back-transformed 80 – 125%
  - Smaller  $\Delta$  for Narrow Therapeutic Index Drugs (NTIDs)
    - EMA  $\Delta 10\%$  leads to BE-limits of 90.00 – 111.11%
    - FDA Scaled (narrowed) based on the variability of the reference
  - Larger  $\Delta$  for Highly Variable Drugs / Drug Products (HVD(P)s)
    - EMA  $\Delta > 20\%$  scaled based on the variability of the reference ( $CV_{wR}$ ), which leads to BE-limits expanded to up to 69.84 – 143.19%
    - HC like EMA, but BE-limits of up to 66.7 – 150.0%
    - FDA Scaled based on the variability of the reference (no upper limit)

# What does $\pm 20\%$ mean and where does it come from?

## Clinically not relevant difference

- Bioequivalence *is not* a scientific concept
  - state a hypothesis
  - perform experiments in order to challenge the hypothesis
  - accept the hypothesis as long as it is not falsified
- Assuming  $\pm 20\%$  to be not clinically relevant was an *ad hoc* concept
- However, empiric evidence of more almost 40 years showed that it ‘works’ (“*No dead people lie in the streets...*”)
- It is a common misconception that BE-limits of 80–125% can lead to approval of products which differ by 45%
  - A survey of 1,636 BE studies submitted to the FDA within 1996–2005 showed  $\Delta$  of 3.19% ( $\pm 2.72$ ) for  $AUC_t$  and 4.50% ( $\pm 3.57$ ) for  $C_{max}$
  - In a strict sense switching *between* generics is *not* supported by (A)BE; nevertheless, it seems to work in practice

# Calculation of point estimate and its 90% CI

Example (2×2 crossover, 8 subjects, 1 dropout,  $CV_{intra} \sim 10\%$ )

subject	sequence	period				LSM (1)	LSM (2)	
		1	2	1 (log)	2 (log)			
1	TR	92.4	97.1	4.526	4.576	T	4.575	4.448
2	TR	86.4	98.0	4.459	4.585	LSM (T)	4.511	mean (T) 4.520
4	TR	114.0	97.9	4.736	4.584	GLSM (T)	91.0	g. mean (T) 91.9
7	TR	97.4	94.6	4.579	4.550			
3	RT	100.9	94.9	4.614	4.553	R	4.589	4.574
5	RT	101.1	71.3	4.616	4.267	LSM (R)	4.581	mean (R) 4.580
6	RT	93.4	92.1	4.537	4.523	GLSM (R)	97.6	g. mean (R) 97.5
8	RT	105.2	-	-	-			
$n_1$ (sequence TR)		4	degr. of freedom ( $n_1+n_2-2$ )					
$n_2$ (sequence RT)		3	5					
Mean Squared Error (MSE)		0.0108184 (from ANOVA)						
Standard Error (SE) of $\Delta$		$0.056173 = \sqrt{[0.5 \times \text{MSE} \times (1/n_1 + 1/n_2)]}$						
$t_{\alpha=0.05, df}$		2.0150						
90% CI = $\Delta \pm t_{\alpha=0.05, df} \times \text{SE}$								
$\Delta = \text{LSM (T)} - \text{LSM (R)}$		-0.0700	93.24%	PE (GMR = $e^{\Delta}$ )				
lower 90% CL		-0.1832	83.26%	90% CI				
upper 90% CL		0.0432	104.41%					



# Calculation of point estimate and its 90% CI

Example (2×2 crossover, 8 subjects, 1 dropout,  $CV_{intra} \sim 10\%$ )

- Important

- Always use the Geometric Least Square Means – *not* the geometric means of treatments

- Only if a design is *balanced*, *i.e.*, there are an equal number of subjects in each sequence, GLSM equals the geometric mean

- In the example (unbalanced;  $n_1 = 4$ ,  $n_2 = 3$ ):

LSM (T) 4.511 (GLSM 91.0) → PE 93.24%

LSM (R) 4.581 (GLSM 97.6)

mean (T) 4.520 (geom. mean 91.9) → PE 93.19%

mean (R) 4.580 (geom. mean 97.5)

- Always use the formula which takes subjects / sequence into account

- There is a ‘simple’ formula which is *only* correct if a study is balanced, namely  $SE = \sqrt{(MSE/n_{ps})}$ , where  $n_{ps} = (n_1 + n_2)/2$

- In the example ( $n_{ps} = 3.5!$ ):

The 90% CI will be wrong (83.36–104.29% instead of 83.26–104.41%)

# Calculation of point estimate and its 90% CI

## Where to find the MSE in software's output

- SAS

The GLM Procedure

Dependent Variable: AUC

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	19	10.8915670	0.5732404	1.86	0.1891
<b>Error</b>	16	4.9439802	<b>0.3089988</b>		
Corrected Total	35	15.8355472			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Treatment	1	1.0469949	1.0469949	3.39	0.0843
Period	1	0.1958572	0.1958572	0.63	0.4376
Sequence	1	1.3052864	1.3052864	2.50	0.1332
Subject(Sequence)	16	8.3434285	0.5214643	1.69	0.1528

- Phoenix/WinNonlin

WINNONLIN LINEAR MIXED EFFECTS MODELING / BIOEQUIVALENCE  
8.0.0.3176  
Core Version 30Jan2014

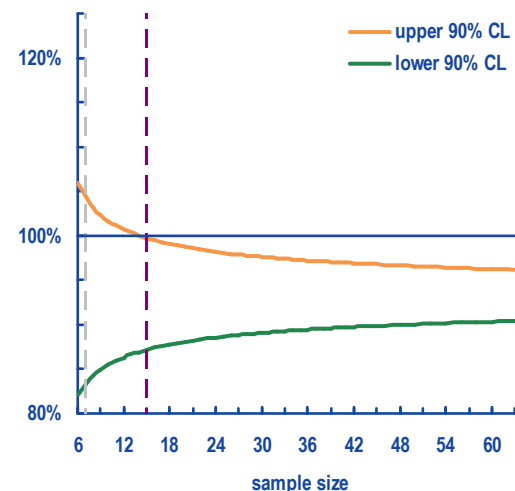
Model Specification and User Settings  
Dependent variable: AUC

Partial Sum of Squares Hypothesis	DF	SS	MS	F_stat	P_value
Sequence	1	1.30529	1.30529	2.50312	0.1332
Sequence*Subject	16	8.34343	0.521464	1.68759	0.1528
Treatment	1	1.04699	1.04699	3.38835	0.0843
Period	1	0.195857	0.195857	0.633844	0.4376
<b>Error</b>	16	4.94398	<b>0.308999</b>		

# Excursion: Treatment effect

## Statistical *significant* $\neq$ clinically *relevant*

- For any given T/R-ratio and variability one will get a significant treatment effect (in the ANOVA  $p < 0.05$ ) if the sample size is only large enough
  - The confidence interval narrows with  $\sqrt{N}$ , *i.e.*, if one uses a four times larger sample size, the CI will be  $\sim$ half as wide
  - If the CI does not include 100% any more, treatments will *significantly* differ
  - However, if the 90% CI is within the acceptance range, this difference is *clinically not relevant*



# Excursion: Period effect

In crossover-studies the period effect is not relevant

- Due to the randomization all treatments will be affected by a true period effect to the same degree
- Period effects mean out, *i.e.*, are handled in the ANOVA
- Previous example, all data in the 2<sup>nd</sup> period multiplied by ten
- Exactly the same PE and 90% CI

subject	sequence	period				LSM (1)	LSM (2)
		1	2	1 (log)	2 (log)		
1	TR	92.4	971	4.526	6.878	T	4.575 6.750
2	TR	86.4	980	4.459	6.888	LSM (T)	5.663
4	TR	114.0	979	4.736	6.887	GLSM (T)	287.9
7	TR	97.4	946	4.579	6.852		
3	RT	100.9	949	4.614	6.855	R	4.589 6.876
5	RT	101.1	713	4.616	6.569	LSM (R)	5.733
6	RT	93.4	921	4.537	6.825	GLSM (R)	308.8
8	RT	105.2	-	-	-		
$\Delta = \text{LSM (T)} - \text{LSM (R)}$				-0.0700	93.24%	PE (GMR = e <sup><math>\Delta</math></sup> )	
lower 90% CL				-0.1832	83.26%	90% CI	
upper 90% CL				0.0432	104.41%		

# Excursion: Sequence effect

In crossover-studies an *equal* sequence effect is not relevant

- However, a *true* sequence effect (better: unequal carry-over) *will* bias the treatment effect
- There is no statistical method to correct for unequal carry-over
- Can only be avoided *by design*, *i.e.*, a sufficiently long enough wash-out between periods
- Previous example, unequal carry-over (TR -5, RT +5)
- Biased PE and CI

subject	sequence	period				LSM (1)	LSM (2)
		1	2	1 (log)	2 (log)		
1	TR	92.4	92.1	4.526	4.523	T	4.575 4.505
2	TR	86.4	93.0	4.459	4.533	LSM (T)	4.540
4	TR	114.0	92.9	4.736	4.532	GLSM (T)	93.7
7	TR	97.4	89.6	4.579	4.495		
3	RT	100.9	99.9	4.614	4.604	R	4.589 4.521
5	RT	101.1	76.3	4.616	4.335	LSM (R)	4.555
6	RT	93.4	97.1	4.537	4.576	GLSM (R)	95.1
8	RT	105.2	-	-	-		
$\Delta = \text{LSM (T)} - \text{LSM (R)}$				-0.0149	98.52%	PE (GMR = $e^{\Delta}$ )	
lower 90% CL				-0.1281	87.98%	90% CI	
upper 90% CL				0.0983	110.33%		

# Basic Statistics for BE

**Thank You!**  
*Open Questions?*



**Helmut Schütz**  
**BEBAC**

Consultancy Services for  
 Bioequivalence and Bioavailability Studies  
 1070 Vienna, Austria  
[helmut.schuetz@bebac.at](mailto:helmut.schuetz@bebac.at)