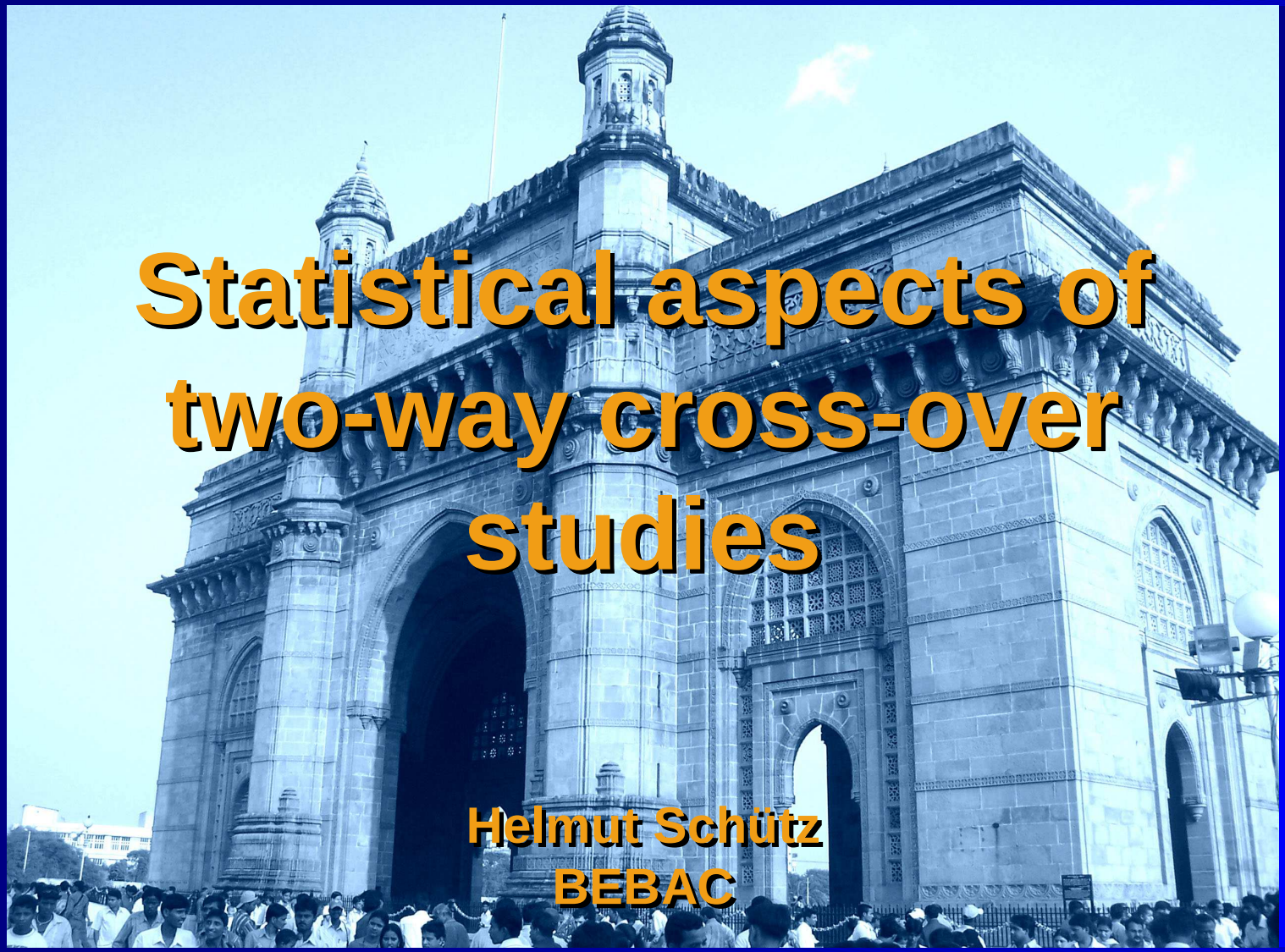
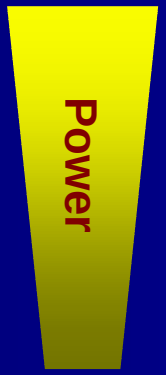


Statistical aspects of two-way cross-over studies

Helmut Schütz
BEBAC

Designs

- The more ‘sophisticated’ a design is, the more information (in terms of variances) we may obtain.
  - Hierarchy of designs:



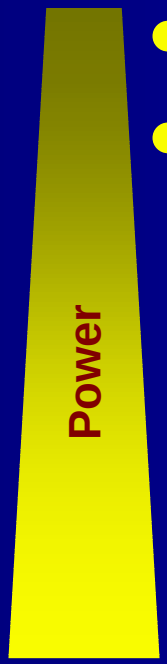
Full replicate (TRTR | RTRT)

Partial replicate (TRR | RTR | RRT)

Standard 2×2 cross-over (RT | TR)

Parallel (R | T)

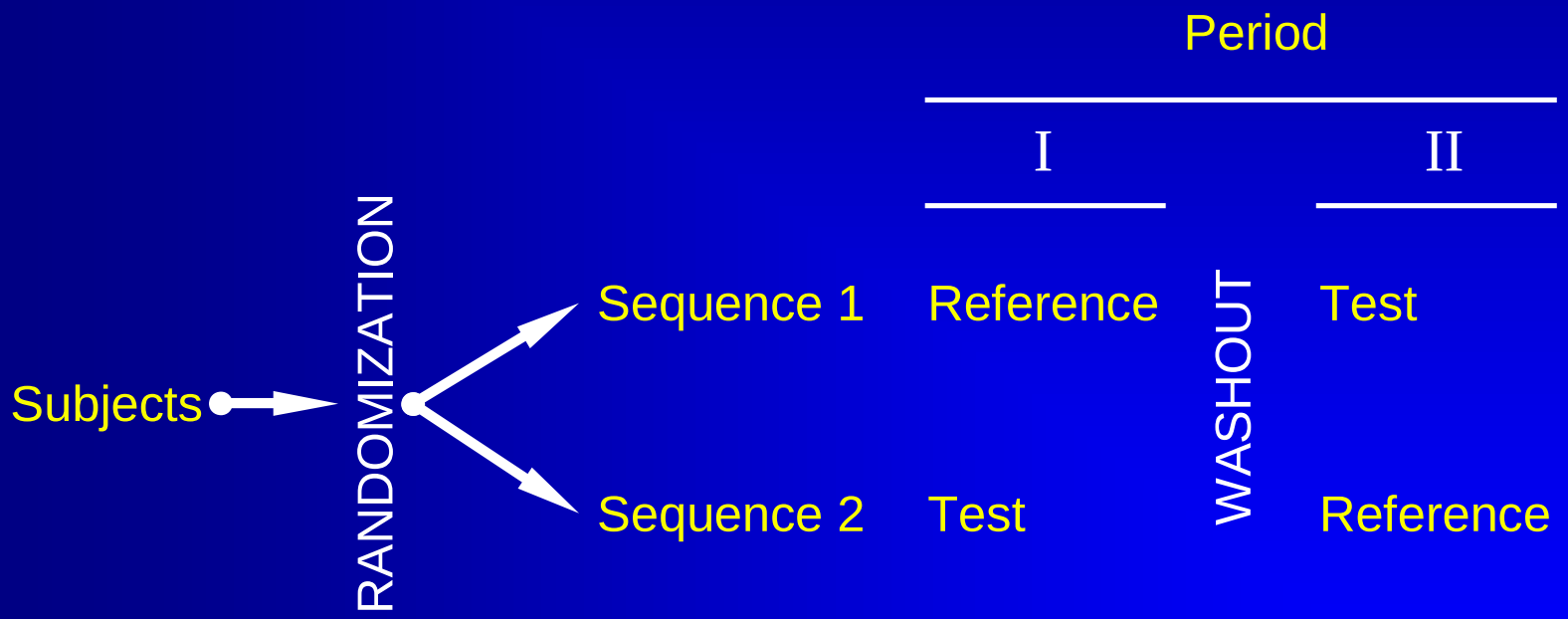
Designs



- Parallel Groups (patients, long half-life drugs)
- Cross-over (generally healthy subjects)
  - Standard 2×2×2
  - Higher Order Designs (more than two formulations)
    - Incomplete Block Designs
    - Latin Squares
    - Variance Balanced Designs (Williams' Designs)
  - Replicate designs

Cross-over designs

- Standard 2×2×2 Design



Cross-over designs (cont'd)

- Every subject is treated once with both test and reference.
- Subjects are randomized into two groups; one is receiving the 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.

Cross-over design: Model

Multiplicative Model (X-over 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 ($j=1,\ldots,n_i$) 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 ($l=1,2$: $\Phi_1=\Phi_{test}$, $\Phi_2=\Phi_{ref.}$)

Cross-over design:
Assumptions

Multiplicative Model (X-over 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…

Cross-over 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 (→ parallel groups)
    - Not optimal for studies in patients with instable diseases (→ parallel groups)
    - Not optimal if CV uncertain (→ two-stage sequential designs)
    - Not optimal for HVDs/HVDPs (→ replicate designs)

Cross-over design: Evaluation

- Mainly by ANOVA and LMEM (linear mixed effects modeling). Results are identical for balanced datasets, and differ only slightly for imbalanced ones.
- Avoid M$-Excel! Almost impossible to validate; tricky for imbalanced datasets – a nightmare for higher-order Xovers. Replicates impossible.
- Suitable software: SAS, Phoenix/WinNonlin, Kinetica and EquivTest/PK (both only 2×2 Xover), S+, Package *bear* for R (freeware).

Cross-over design: Example

subject	T	R
1	28.39	35.44
2	39.86	49.42
3	32.75	36.78
4	33.36	33.40
5	34.97	34.81
6	24.29	24.65
7	28.61	31.77
8	45.44	45.54
9	59.49	65.29
10	27.87	28.23
11	24.26	25.71
12	42.30	37.01

	sequence RT			sequence TR	
subject	P I	P II	subject	P I	P II
2	39.86	49.42	1	28.39	35.44
3	32.75	36.78	4	33.36	33.40
5	34.97	34.81	6	24.29	24.65
8	45.44	45.54	7	28.61	31.77
10	27.87	28.23	9	59.49	65.29
11	24.26	25.71	12	42.30	37.01

Ordered by treatment sequences (RT|TR)

ANOVA on log-transformed data →

Cross-over design: Example

Sequence	Period 1		Period 2		Sequence mean	
1	1R = $X_{.11}$	3.5103	1T = $X_{.21}$	3.5768	$X_{..1}$	3.5436
2	2T = $X_{.12}$	3.5380	2R = $X_{.22}$	3.5883	$X_{..2}$	3.5631
Period mean		$X_{.1.}$ 3.5241		$X_{.2.}$ 3.5826	$X_{...}$	3.5533

RT = n_1 = 6

TR = n_2 = 6 $1/n_1+1/n_2$ 0.3333

balanced n = 12 1/n 0.0833 n_1+n_2-2 10

Analysis of Variance

Source of variation	df	SS	MS	F	P-value	CV
Inter-subjects						
Carry-over	1	0.00230	0.00230	0.0144	0.90679	
Residuals	10	1.59435	0.15943	29.4312	4.32E-6	28.29%
Intra-subjects						
Direct drug	1	0.00040	0.00040	0.0733	0.79210	
Period	1	0.02050	0.02050	3.7844	0.08036	
Residuals	10	0.05417	0.00542			7.37%
Total	23	1.67172				29.31%

δ_{ML} **1.0082** *MLE (maximum likelihood estimator) of Delta-ML*

X_R 3.5493 *LS (least squares mean for the reference formulation)* $\exp(X_R)$ 34.79

X_T 3.5574 *LS (least squares mean for the test formulation)* $\exp(X_T)$ 35.07

Cross-over design: Example

Classical (Shortest) Confidence Interval

± x rule: 20 [100 - x; 1 / (100 - x)]

θ_L	-0.2231	θ_U	+0.2231	α 0.0500	p=1-2·α 0.9000
δ_L	**80%**	δ_U	**125%**	$t_{2\cdot\alpha,df}$	1.8125
L_1	-0.0463	U_1	0.0626	*difference within Theta-L AND Theta-U; bioequivalent*	
L_2	**95.47%**	U_2	**106.46%**	***difference within Delta-L AND Delta-U; bioequivalent***	
δ_{ML}	**100.82%**			*MLE; maximum likelihood estimator*	
δ_{MVUE}	100.77%			*MVUE; minimum variance unbiased estimator*	
δ_{RM}	100.98%			*RM; ratio of formulation means*	
δ_{MIR}	101.44%			*MIR; mean of individual subject ratios*	

Cross-over design: Example

- Calculation of 90% CI (2-way cross-over)
  - Sample size (n) 12, Point Estimate (PE) 100.82%, Residual Mean Squares Error (MSE) from ANOVA (*ln*-transformed values) 0.005417, $t_{\alpha,n-2}$ 1.8125
    - Standard Error (SE_Δ) of the mean difference

$$SE_\Delta = \sqrt{MSE}\sqrt{\frac{2}{n}} = \sqrt{0.005417}\sqrt{\frac{2}{12}} = 0.030047$$

    - Confidence Interval

$$CL_L = e^{\ln PE - t_{2\alpha,df}\cdot SE_\Delta} = e^{0.0081349 - 1.8125\times 0.030047} = 95.47\%$$

$$CL_H = e^{\ln PE + t_{2\alpha,df}\cdot SE_\Delta} = e^{0.0081349 + 1.8125\times 0.030047} = 106.46\%$$

R code / result

```
#Cross-over 12 subjects
T1     <- c(28.39,33.36,24.29,28.61,59.49,42.30)
T2     <- c(49.42,36.78,34.81,45.54,28.23,25.71)
R1     <- c(39.86,32.75,34.97,45.44,27.87,24.26)
R2     <- c(35.44,33.40,24.65,31.77,65.29,37.01)
RT     <- log(R1) - log(T2)
TR     <- log(R2) - log(T1)
n1     <- length(RT)
mRT    <- mean(RT)
vRT    <- var(RT)
n2     <- length(TR)
mTR    <- mean(TR)
vTR    <- var(TR)
mD     <- mean(log(c(T1,T2))) - mean(log(c(R1,R2)))
MSE    <- (((n1-1)*vRT + (n2-1)*vTR)/(n1+n2-2))/2
alpha  <- 0.05
lo     <- mD - qt(1-alpha,n1+n2-2)*sqrt(MSE)*
            sqrt((1/(2*n1) + 1/(2*n2)))
hi     <- mD + qt(1-alpha,n1+n2-2)*sqrt(MSE)*
            sqrt((1/(2*n1) + 1/(2*n2)))
result <- paste(
          paste(" Back transformed (raw data scale)\n",
                "Point estimate:“,
                round(100*exp(mD), digits=2),"%\n"),
          paste("90 % confidence interval:"),
          paste(round(100*exp(lo), digits=2), "% to"),
          paste(round(100*exp(hi), digits=2),"%\n",
          paste("CVintra:",round(100*sqrt(exp(MSE)-1),
            digits=2),"%\n")))
cat(result)
```

```
Back transformed (raw data scale)
Point estimate: 100.82 %
90 % confidence interval: 95.47 % to 106.46 %
CVintra: 7.37 %
```

Comparison of designs

- Most important in an ANOVA table is the residual mean error ($\rightarrow$ CI, CV_{intra} for future studies).
  - Carry-over (aka sequence effects) can not be handled! Must be excluded by design (long enough washout).
  - Period effects are accounted for (significant *p*-values are not important). Example: all values in P2 ×100…

Dependent	Hypothesis	DF	SS	MS	F_stat	P_value
Ln(AUCinf)	sequence	1	0.002299563	0.002299563	0.014423232	0.90678513
Ln(AUCinf)	sequence*subject	10	1.5943472	0.15943472	29.431239	4.3211352E-0
Ln(AUCinf)	treatment	1	0.000397058	0.000397058	0.07329589	0.79210291
Ln(AUCinf)	period	1	0.020500963	0.020500963	3.784425	0.080364101
Ln(AUCinf)	Error	10	0.054171935	0.005417193		

CI_90_Lower	CI_90_Upper
95.471828	106.46102

Dependent	Hypothesis	DF	SS	MS	F_stat	P_value
Ln(AUCinf)	sequence	1	0.002299563	0.002299563	0.014423232	0.90678513
Ln(AUCinf)	sequence*subject	10	1.5943472	0.15943472	29.431239	4.3211352E-0
Ln(AUCinf)	treatment	1	0.000397058	0.000397058	0.07329589	0.79210291
Ln(AUCinf)	period	1	130.49632	130.49632	24089.286	0
Ln(AUCinf)	Error	10	0.054171935	0.005417193		

CI_90_Lower	CI_90_Upper
95.471828	106.46102

Reading ANOVA tables

Analysis of Variance

Source of variation	df	SS	MS	F	P-value	CV
Between subjects						
Carry-over	1	0.00230	0.002300	0.0144	0.90679	
Residuals	10	1.59435	0.159435	29.4312	4.32E-6	28.29%
Within subjects						
Direct drug	1	0.00040	0.000397	0.0733	0.79210	
Period	1	0.02050	0.020501	3.7844	0.08036	
Residuals	10	0.05417	0.005417			7.37%
Total	23	1.67172				29.31%

Should not be tested: Design – washout!

$$CV_{inter} = \sqrt{e^{\frac{MSE_B - MSE_W}{2}} - 1}$$

$$CV_{intra} = \sqrt{e^{MSE_W} - 1}$$

$$CV_{total} = \sqrt{e^{\frac{MSE_B + MSE_W}{2}} - 1}$$

Not surprising: different subjects!

Not important: Significant value would old mean that 100% is not included in the CI.

Not important: Both formulations would be affected in the same way.

balanced: $n_1 = n_2; n = n_1 + n_2$

$$CI = e^{\ln PE \pm t_{\alpha,n-2} \cdot \sqrt{MSE_W} \sqrt{\frac{2}{n}}}$$

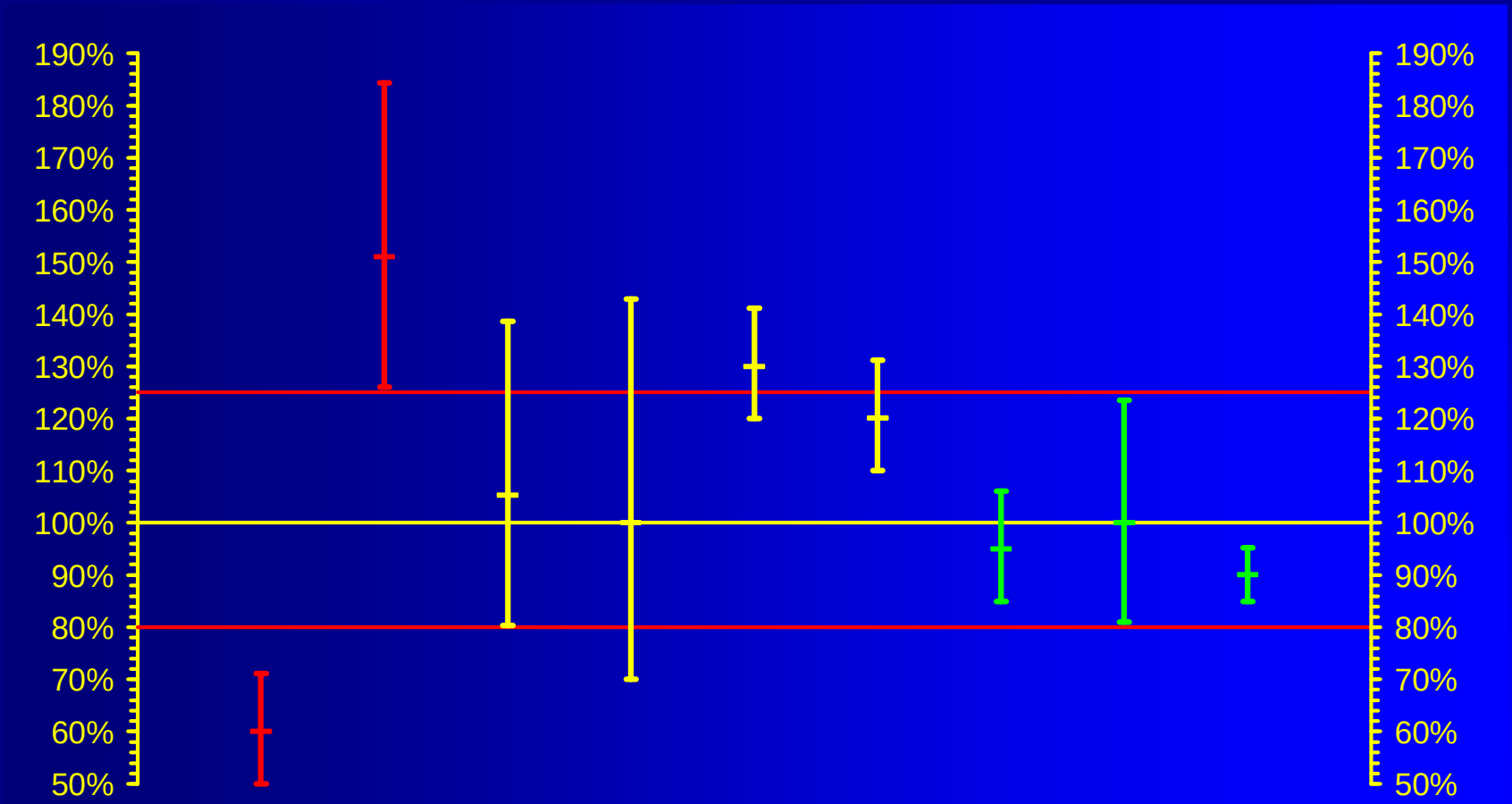
imbalanced: $n_1 \neq n_2$

$$CI = e^{\ln PE \pm t_{\alpha,n_1+n_2-2} \cdot \sqrt{MSE_W} \sqrt{\frac{1}{2n_1} + \frac{1}{2n_2}}}$$

A note on BE assessment

- The *width* of the confidence interval depends on the variability observed in the study.
- The *location* of the confidence interval depends on the observed test/reference-ratio.
- Decision rules:
  - Confidence Interval (CI) entirely outside the Acceptance Range (AR): Bioinequivalence proven.
  - CI overlaps the AR, but is not entirely within the AR: Bioequivalence not proven.
  - CI entirely within the AR: Bioequivalence proven.

A note on BE assessment



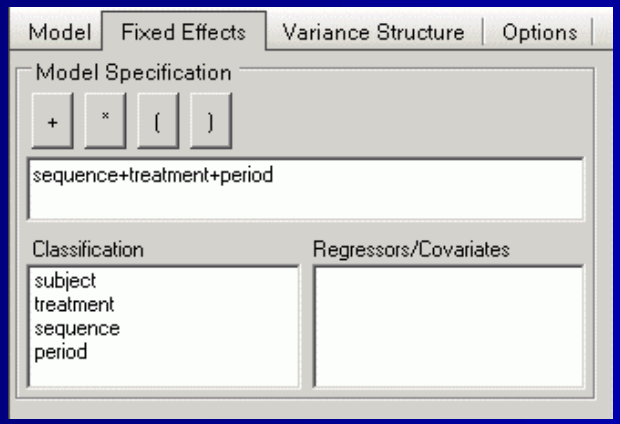
EMA *vs.* Rest of the World

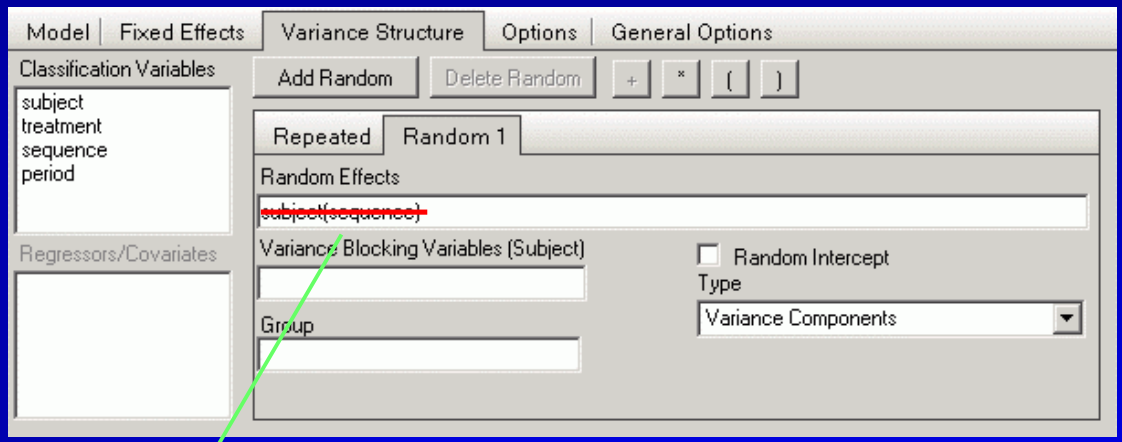
- EMA BE GL (2010), 4.1.8 Evaluation / Statistical analysis:

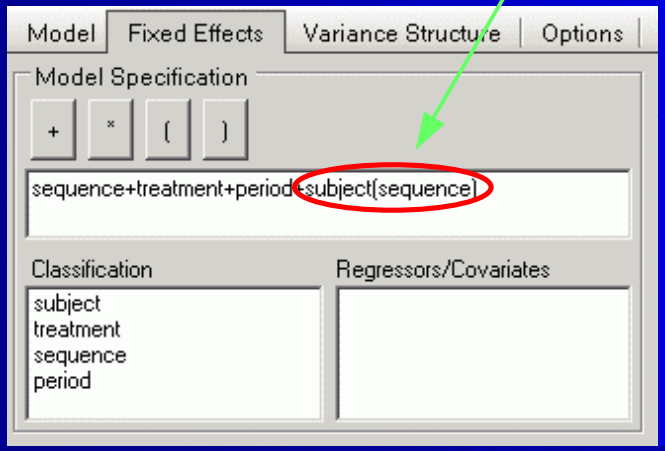
  The terms to be used in the ANOVA model are usually sequence, subject within sequence, period and formulation. Fixed effects, rather than random effects, should be used for all terms.

- Adapt your standard setup:
  - SAS: Proc GLM instead of Proc MIXED (*i.e.* incomplete data are dropped).
  - Phoenix/WinNonlin: Don't use the defaults!

EMA *vs.* Rest of the World







Balanced, incomplete, imbalanced data…

- Balanced: Identical numbers of subjects in both sequences ($n_{RT} = n_{TR}$)
- Incomplete: $n_{RT} = n_{TR}$, one period is missing
- Imbalanced: $n_{RT} \neq n_{TR}$
- A mixed (random effects) model might recover some information from missing periods, but the differences are small and rarely used in 2×2 Xovers (although default in Phoenix / WinNonlin).
- Proc GLM drops incomplete data.

Random *vs.* fixed

- Dataset from Chow & Liu
  Design and Analysis of Bioavailability and Bioequivalence Studies, Chapman & Hall/CRC Press, Boca Raton, 3rd ed. 2008, p71
  - Complete data; n=24
  - Incomplete data; n=24, subject 24's period 2 (T) removed
  - Imbalanced data; n=23, subject 24 removed

Random *vs.* fixed

- Dataset from Chow & Liu
  Design and Analysis of Bioavailability and Bioequivalence Studies, Chapman & Hall/CRC Press, Boca Raton, 3rd ed. 2008, p71

dataset	model	df	PE	90% CI		CV_{intra}
balanced	random	22.	97.18%	88.31%	106.93%	19.47%
	fixed	22.	97.18%	88.31%	106.93%	19.47%
incomplete	random	20.82	96.47%	87.61%	106.22%	19.22%
	fixed	21.	95.61%	86.86%	105.23%	19.06%
imbalanced	random	21.	95.61%	86.86%	105.23%	19.06%
	fixed	21.	95.61%	86.86%	105.23%	19.06%

Cross-over designs (cont'd)

- Special case: Evaluation of t_{max}
  - Since t_{max} is sampled from discrete values, a nonparametric method must be applied
  - Estimation of differences (linear model)
  - Wilcoxon Two-Sample Test (available in SAS 9.2+ Proc NPAR1way, Proc StatXact, Phoenix / WinNonlin, EquivTest/PK, R package *coin*)
  - Since based on a discrete distribution, generally $\alpha<0.05$ (*e.g.*, n=12: 0.0465, 24: 0.0444, 32: 0.0469, 36: 0.0485, 48: 0.0486,…)

Hauschke D, Steinijans VW and E Diletti
A distribution-free procedure for the statistical analysis of bioequivalence studies
Int J Clin Pharm Ther Toxicol 28/2, 72–78 (1990)

Cross-over designs (cont'd)

Sequence 1 (RT)				Sequence 2 (TR)			
Subject	Period I	Period II	P.D.	Subject	Period I	Period II	P.D.
2	3.0	1.5	-1.5	1	2.0	2.0	±0.0
4	2.0	2.0	±0.0	3	2.0	2.0	±0.0
6	2.0	3.0	+1.0	5	2.0	3.0	+1.0
8	2.0	3.0	+1.0	7	2.0	1.5	-0.5
10	1.5	2.0	+0.5	9	3.0	2.0	-1.0
12	3.0	2.0	-1.0	11	2.0	1.5	-0.5
14	3.0	3.0	±0.0	13	3.0	1.5	-1.5

Cross-over designs (cont'd)

ADDITIVE (raw data) MODEL

metric: t_{max}

Sequence	Period 1		Period 2	
1	R_{L1} =	65	R_{U1} =	46
2	R_{L2} =	36	R_{U2} =	55
RT =	n_1 =	7		
TR =	n_2 =	7		
balanced	n =	14	$n_1 \cdot n_2$	49

d_1 0.0000 d_2 -0.1786 *(mean period difference in sequence 1 / 2)*

$\tilde{Y}_R$ 2.000 *median of the reference formulation*

$\tilde{Y}_T$ 2.000 *median of the test formulation*

Distribution-Free Confidence Interval (Moses)

± x rule : 20

θ_L -0.429 θ_U +0.429 α 0.0487 $p = 1-2\cdot\alpha$ 0.9026

δ_L 80% δ_U 120%

L_W -0.250 U_W +0.750 ***difference outside Theta-L AND/OR Theta-U; not bioequivalent***

$\tilde{\theta}$ **+0.250** ***Hodges-Lehmann estimate (median of paired differences)***

Wilcoxon-Mann-Whitney Two One-Sided Tests Procedure (Hauschke)

W_L 37 W_U 18

$W_{0.95,n1,n2}$ 38 $W_{0.05,n1,n2}$ 12 *H0(1): diff. <= Theta-L AND H0(2): diff. => Theta-U; not bioequivalent*

p_1 >0.0487 and p_2 >0.0487

Thank You!

Statistical aspects of two-way cross-over studies

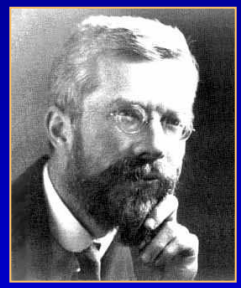
Open Questions?

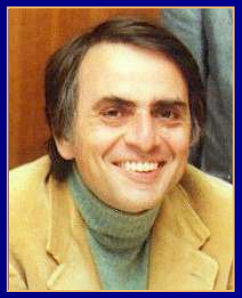
Helmut Schütz
BEBAC
Consultancy Services for
Bioequivalence and Bioavailability Studies
1070 Vienna, Austria
helmut.schuetz@bebac.at

To bear in Remembrance...

To call the statistician after the experiment is done may be no more than asking him to perform a *post-mortem* examination: he may be able to say what the experiment died of.

Ronald A. Fisher





[The] impatience with ambiguity can be criticized in the phrase:
absence of evidence is not evidence of absence.

Carl Sagan

[…] our greatest mistake would be to forget that data is used for serious decisions in the very real world, and bad information causes suffering and death.

Ben Goldacre

Subject	Period	Sequence	Formulation	AUC
1	1	RT	Reference	74.675
1	2	RT	Test	73.675
2	1	TR	Test	74.825
2	2	TR	Reference	37.35
3	1	TR	Test	86.875
3	2	TR	Reference	51.925
4	1	RT	Reference	96.4
4	2	RT	Test	93.25
5	1	RT	Reference	101.95
5	2	RT	Test	102.125
6	1	RT	Reference	79.05
6	2	RT	Test	69.45
7	1	TR	Test	81.675
7	2	TR	Reference	72.175
8	1	TR	Test	92.7
8	2	TR	Reference	77.5
9	1	TR	Test	50.45
9	2	TR	Reference	71.875
10	1	TR	Test	66.125
10	2	TR	Reference	94.025
11	1	RT	Reference	79.05
11	2	RT	Test	69.025
12	1	RT	Reference	85.95
12	2	RT	Test	68.7

Subject	Period	Sequence	Formulation	AUC
13	1	TR	Test	122.45
13	2	TR	Reference	124.975
14	1	TR	Test	99.075
14	2	TR	Reference	85.225
15	1	RT	Reference	69.725
15	2	RT	Test	59.425
16	1	RT	Reference	86.275
16	2	RT	Test	76.125
17	1	TR	Test	86.35
17	2	TR	Reference	95.925
18	1	TR	Test	49.925
18	2	TR	Reference	67.1
19	1	RT	Reference	112.675
19	2	RT	Test	114.875
20	1	RT	Reference	99.525
20	2	RT	Test	116.25
21	1	TR	Test	42.7
21	2	TR	Reference	59.425
22	1	TR	Test	91.725
22	2	TR	Reference	114.05
23	1	RT	Reference	89.425
23	2	RT	Test	64.175
24	1	RT	Reference	55.175
24	2	RT	Test	74.575