Influence of missing data on analysis and power calculation in bioequivalence studies

Henrike Häbel

Ulm University
in cooperation with
Boehringer Ingelheim Pharma GmbH & Co. KG

May 4th, 2011

Outline

Outline

Bioequivalence studies (BE studies)

- Test if test- and reference formulations differ marginally with respect to pharmacokinetic (PK) characteristics
- Endpoints: PK-parameters

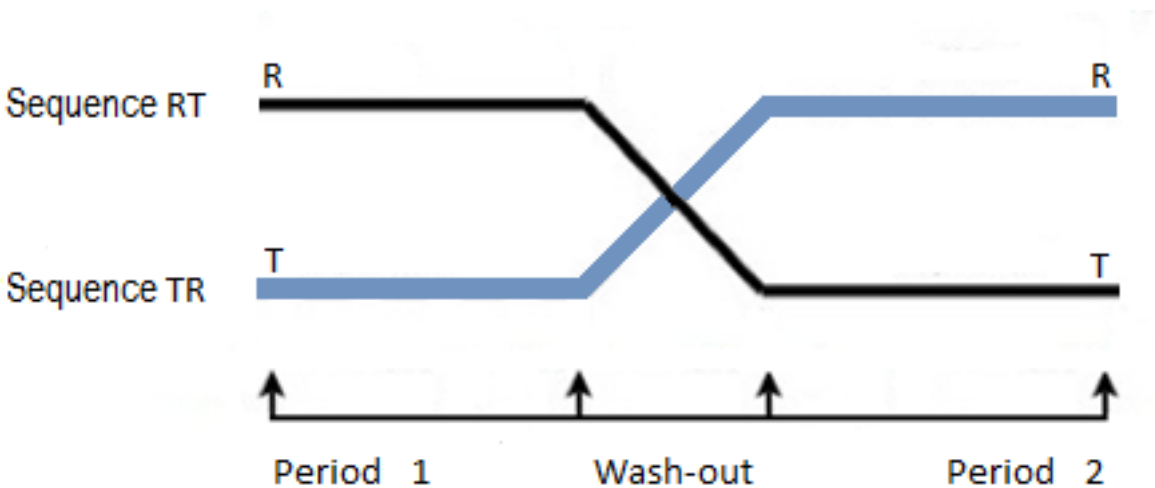
  Primary: AUC and C_{max}

Problem concerning analysis and power calculation:

At how many missing values should further subjects be recruited?

Typical design of BE studies: 2x2-Crossover

- Test- (T) and reference (R) formulation are given in two periods
- N subjetcs are allocated randomly evenly to one of the following sequences: RT and TR
- Treatment according to sequence and period



Outline

Model (on log-scale)

$$Y_{ijk} = S_{ik} + P_j + F_{jk} + C_k + e_{ijk}$$

Y_{ijk} logarithm of response (PK parameter) measured on subject i in sequence k in period j, i $= 1, \ldots, N$, j $= 1, 2$, k $= 1, 2$

S_{ik} random i-th subject effect in sequence k, $N(0, \tau^2)$ i.i.d.

P_j fixed effect in period j

F_{jk} fixed treatment effect in the k-th sequence in period j

C_k fixed effect in the k-th sequence

e_{ijk} residual (random) error associated with the i-th subject in sequence k in period j, $N(0, \sigma^2)$ i.i.d.

Why random subject effect?

EMA: All effects should be considered fixed
⇒ Subjects with one missing value excluded from analysis
⇒ If applicable, further subjects have to be recruited!

FDA: All avaible data should be included in the analysis
⇒ Properties of REML can be used
(e.g. in SAS® with PROC MIXED)
⇒ When should further subjects be recruited?

Model assumptions

Number of subjects: N = 24

$$Y_{ijk} = S_{ik} + P_j + F_{jk} + C_k + e_{ijk}$$

S_{ik} : subject $S_{ik} \sim N(0, \tau^2)$ with $\tau^2 = 0.25$

P_j : period $P_1 = 0,\ P_2 = 0.2$

F_{jk} : treatment $F_{11} = F_{22} = 0,\ F_{12} = F_{21} := F_2 = 0.1$

C_k : sequence $C_1 = C_2 = 0$

e_{ijk} : residual error $e_{ijk} \sim N(0, \sigma^2)$ with $\sigma^2 = 0.06$

Modeling of response per sequence and period

Sequence	Period 1	Period 2
RT	$Y_{i11} = S_{i1} + e_{i11}$	$Y_{i21} = S_{i1} + 0.2 + 0.1 + e_{i21}$
TR	$Y_{i12} = S_{i2} + 0.1 + e_{i12}$	$Y_{i22} = S_{i2} + 0.2 + e_{i22}$

Outline

Randomly missing values

With N = 24 subjects:

- Number of missing values: 1, 2, 3, 4, 8, 12, 16, 20
- Three alternatives to distribute missing values:
  - for both treatments and in both periods (case I)
  - only in period 2 (case II)
  - only for treatment R (case III)

Pseudo code

DO 1 TO 1000

1. Simulate complete dataset for 24 subjects
2. Analyse complete dataset with PROC MIXED
3. FOR i = 1, 2, 3, 4, 8, 12, 16, 20 DO

   PROC SURVEYSELECT to simulate incomplete datasets with i missing values for case I - III

   END
4. Analyse incomplete dataset with PROC MIXED

END

One missing value 'for both treatments' (case I)

PTNO	TPATT	PKPER	PKTRT	lnPK	STRA
1	RT	1	R	X	1
2	RT	1	R	X	1
3	RT	1	R	X	1
4	RT	1	R	X	1
5	RT	1	R	X	1
6	RT	1	R	X	1
7	RT	1	R	X	1
8	RT	1	R	X	1
9	RT	1	R	X	1
10	RT	1	R	X	1
11	RT	1	R	X	1
12	RT	1	R	X	1
13	TR	1	T	X	2
14	TR	1	T	X	2
15	TR	1	T	X	2
16	TR	1	T	X	2
17	TR	1	T	X	2
18	TR	1	T	X	2
19	TR	1	T	X	2
20	TR	1	T	X	2
21	TR	1	T	X	2
22	TR	1	T	X	2
23	TR	1	T	X	2
24	TR	1	T	X	2

PTNO	TPATT	PKPER	PKTRT	lnPK	STRA
1	RT	2	T	X	3
2	RT	2	T	X	3
3	RT	2	T	X	3
4	RT	2	T	X	3
5	RT	2	T	X	3
6	RT	2	T	X	3
7	RT	2	T	X	3
8	RT	2	T	X	3
9	RT	2	T	X	3
~~10~~	~~RT~~	~~2~~	~~T~~	~~X~~	~~3~~
11	RT	2	T	X	3
12	RT	2	T	X	3
13	TR	2	R	X	4
14	TR	2	R	X	4
15	TR	2	R	X	4
16	TR	2	R	X	4
17	TR	2	R	X	4
18	TR	2	R	X	4
19	TR	2	R	X	4
20	TR	2	R	X	4
21	TR	2	R	X	4
22	TR	2	R	X	4
23	TR	2	R	X	4
24	TR	2	R	X	4

Two missing values 'for both treatments' (case I)

PTNO	TPATT	PKPER	PKTRT	lnPK	STRA
1	RT	1	R	X	1
2	RT	1	R	X	1
3	RT	1	R	X	1
4	RT	1	R	X	1
5	RT	1	R	X	1
6	RT	1	R	X	1
7	RT	1	R	X	1
8	RT	1	R	X	1
9	RT	1	R	X	1
10	RT	1	R	X	1
11	RT	1	R	X	1
12	RT	1	R	X	1
13	TR	1	T	X	2
14	TR	1	T	X	2
15	TR	1	T	X	2
16	TR	1	T	X	2
17	TR	1	T	X	2
18	TR	1	T	X	2
19	TR	1	T	X	2
20	TR	1	T	X	2
21	TR	1	T	X	2
22	TR	1	T	X	2
23	TR	1	T	X	2
24	TR	1	T	X	2

PTNO	TPATT	PKPER	PKTRT	lnPK	STRA
1	RT	2	T	X	3
2	RT	2	T	X	3
3	RT	2	T	X	3
4	RT	2	T	X	3
5	RT	2	T	X	3
6	RT	2	T	X	3
7	RT	2	T	X	3
8	RT	2	T	X	3
9	RT	2	T	X	3
10	RT	2	T	X	3
11	RT	2	T	X	3
12	RT	2	T	X	3
13	TR	2	R	X	4
14	TR	2	R	X	4
15	TR	2	R	X	4
16	TR	2	R	X	4
17	TR	2	R	X	4
18	TR	2	R	X	4
19	TR	2	R	X	4
20	TR	2	R	X	4
21	TR	2	R	X	4
22	TR	2	R	X	4
23	TR	2	R	X	4
24	TR	2	R	X	4

Three or more missing values 'for both treatments' (case I)

PTNO	TPATT	PKPER	PKTRT	lnPK	STRA
1	RT	1	R	X	1
2	RT	1	R	X	1
3	RT	1	R	X	1
4	RT	1	R	X	1
5	RT	1	R	X	1
6	RT	1	R	X	1
7	RT	1	R	X	1
8	RT	1	R	X	1
9	RT	1	R	X	1
10	RT	1	R	X	1
11	RT	1	R	X	1
12	RT	1	R	X	1
13	TR	1	T	X	2
14	TR	1	T	X	2
15	TR	1	T	X	2
16	TR	1	T	X	2
17	TR	1	T	X	2
18	TR	1	T	X	2
19	TR	1	T	X	2
20	TR	1	T	X	2
21	TR	1	T	X	2
22	TR	1	T	X	2
23	TR	1	T	X	2
24	TR	1	T	X	2

PTNO	TPATT	PKPER	PKTRT	lnPK	STRA
1	RT	2	T	X	3
2	RT	2	T	X	3
3	RT	2	T	X	3
4	RT	2	T	X	3
5	RT	2	T	X	3
6	RT	2	T	X	3
7	RT	2	T	X	3
8	RT	2	T	X	3
9	RT	2	T	X	3
10	RT	2	T	X	3
11	RT	2	T	X	3
12	RT	2	T	X	3
13	TR	2	R	X	4
14	TR	2	R	X	4
15	TR	2	R	X	4
16	TR	2	R	X	4
17	TR	2	R	X	4
18	TR	2	R	X	4
19	TR	2	R	X	4
20	TR	2	R	X	4
21	TR	2	R	X	4
22	TR	2	R	X	4
23	TR	2	R	X	4
24	TR	2	R	X	4

Missing values 'in period 2' (case II) or 'for treatment R' (case III)

PTNO	TPATT	PKPER	PKTRT	lnPK	STRA
1	RT	1	R	X	1
2	RT	1	R	X	1
3	RT	1	R	X	1
4	RT	1	R	X	1
5	RT	1	R	X	1
6	RT	1	R	X	1
7	RT	1	R	X	1
8	RT	1	R	X	1
9	RT	1	R	X	1
10	RT	1	R	X	1
11	RT	1	R	X	1
12	RT	1	R	X	1
13	TR	1	T	X	2
14	TR	1	T	X	2
15	TR	1	T	X	2
16	TR	1	T	X	2
17	TR	1	T	X	2
18	TR	1	T	X	2
19	TR	1	T	X	2
20	TR	1	T	X	2
21	TR	1	T	X	2
22	TR	1	T	X	2
23	TR	1	T	X	2
24	TR	1	T	X	2

PTNO	TPATT	PKPER	PKTRT	lnPK	STRA
1	RT	2	T	X	3
2	RT	2	T	X	3
3	RT	2	T	X	3
4	RT	2	T	X	3
5	RT	2	T	X	3
6	RT	2	T	X	3
7	RT	2	T	X	3
8	RT	2	T	X	3
9	RT	2	T	X	3
10	RT	2	T	X	3
11	RT	2	T	X	3
12	RT	2	T	X	3
13	TR	2	R	X	4
14	TR	2	R	X	4
15	TR	2	R	X	4
16	TR	2	R	X	4
17	TR	2	R	X	4
18	TR	2	R	X	4
19	TR	2	R	X	4
20	TR	2	R	X	4
21	TR	2	R	X	4
22	TR	2	R	X	4
23	TR	2	R	X	4
24	TR	2	R	X	4

PROC SURVEYSELECT

```
PROC SURVEYSELECT DATA = data
   METHOD = SRS N = 47 OUT = out1;
RUN;

PROC SORT DATA = data;
   BY tpatt;
RUN;

PROC SURVEYSELECT DATA = data
   METHOD = SRS N = 23 OUT = out2;
   STRATA tpatt;
RUN;

PROC SORT DATA = subdata;
   BY stra;
RUN;

PROC SURVEYSELECT DATA = subdata
   METHOD = SRS N = 6/*-k*/ OUT = out3;
   STRATA stra;
RUN;
```

Outline

Is there a difference between the formulations?

Test if:

- the 90% confidence interval, which covers the treatment quotient $\frac{\mu_T}{\mu_R}$, lies within the equivalence domain $[0.80, 1.25]$
- the 90% confidence interval, which covers the treatment difference $\mu_T - \mu_R$, lies within the equivalence domain $[-0.223, 0.223]$ (log-scale)

PROC MIXED

Theory: Mixed model $Y = \underbrace{X\beta}_{\text{fix}} + \underbrace{Z\gamma}_{\text{random}} + \epsilon$

```
PROC MIXED DATA = data ORDER = interval;
   CLASS ptno pktrt tpatt pkper;
   MODEL lnpk = pktrt pkper tpatt
      / HTYPE = 2 SOLUTION
        OUTPRED = outpred
        DDFM = KR;
   RANDOM ptno(tpatt)  / S;
   LSMEANS pktrt pkper / CL;
   ESTIMATE 'Diff T-R' pktrt -1 1 / CL alpha = 0.1;
RUN;
```

Value of interest: gCV

Die Prozedur MIXED

Covariance Parameter Estimates

Kov.Parm	Schätzwert
ptno(tpatt)	0.1888
Residual	0.04484

$\widehat{\tau^2}$

$MSE = \widehat{\sigma^2}$

Among others considered:

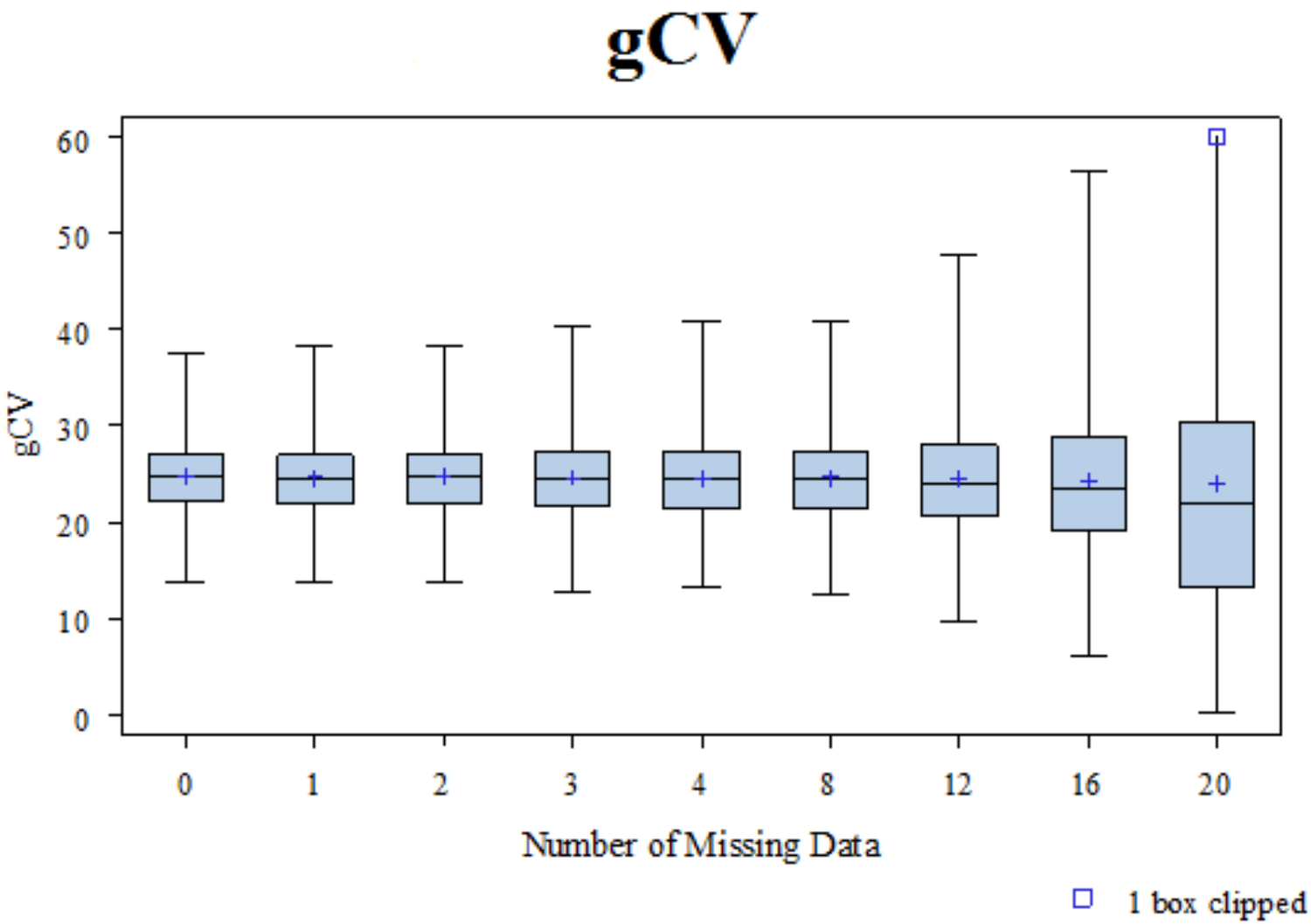
$$gCV := 100 * \sqrt{exp(MSE) - 1}$$

derived from CV of a log-normally distributed random variable X

$$CV := \frac{\sqrt{Var(X)}}{E(X)}$$

Evaluation with PROC BOXPLOT

$\sigma^2 = 0.06 \triangleright gCV = 25$, missing values for treatment R (case III)



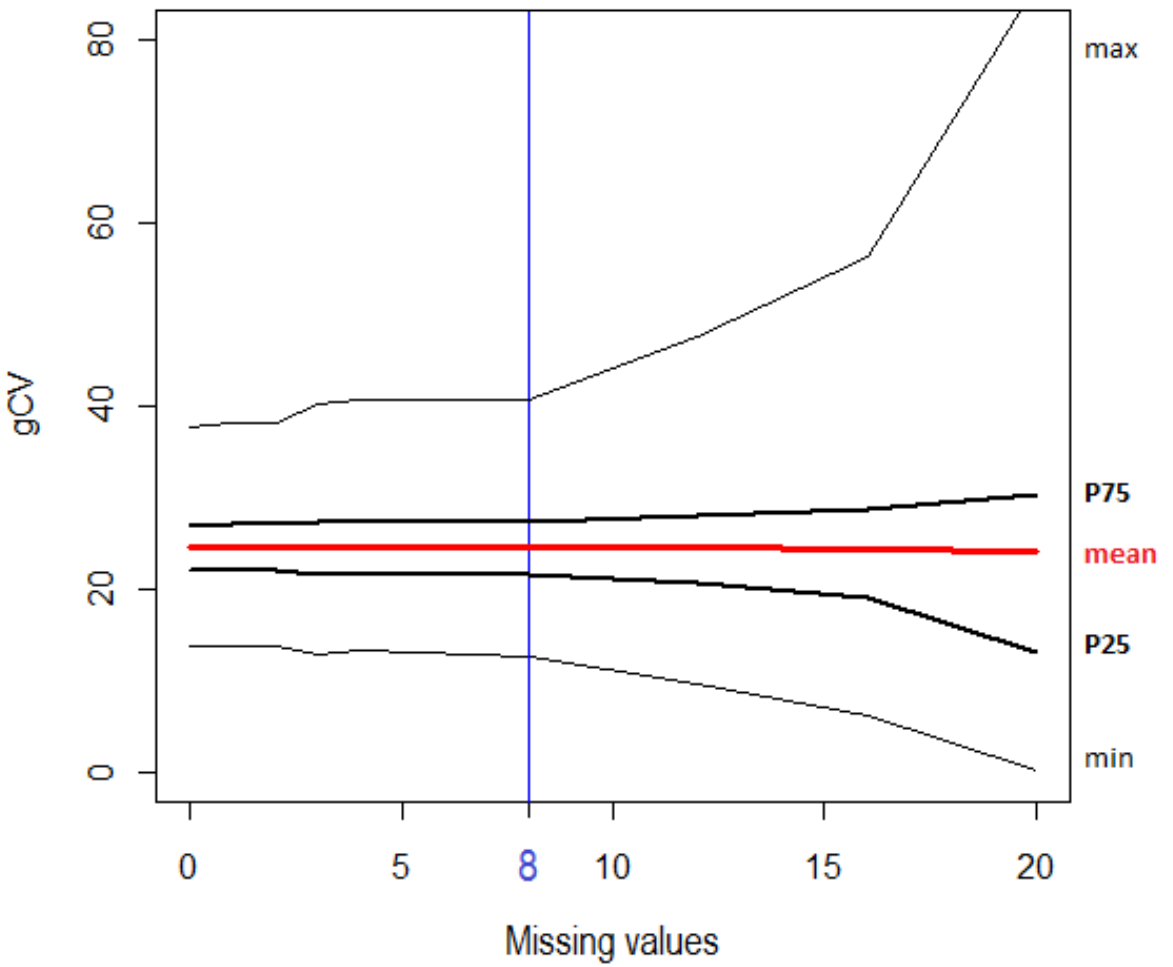
Summary statistics with PROC MEANS

$\sigma^2 = 0.06 \triangleright gCV = 25$, missing values for treatment R (case III)

Missing values	Number of simulations	Min	P25	Mean	P75	Max
0	1000	13.765	22.001	24.569	26.828	37.540
1	1000	13.887	22.001	24.594	26.917	40.143
2	1000	13.887	21.820	24.536	27.045	39.052
3	1000	14.258	21.662	24.561	27.138	39.950
4	1000	5.301	19.552	24.317	28.520	48.217
8	1000	7.589	19.137	24.417	29.182	53.245
12	1000	4.040	16.892	24.018	29.659	68.274
16	1000	0.599	13.904	23.498	30.804	68.017
20	1000	0.0001	0.011	20.338	36.371	87.275

Summary statistics graphic

$\sigma^2 = 0.06 \triangleright gCV = 25$, missing values for treatment R (case III)



Power

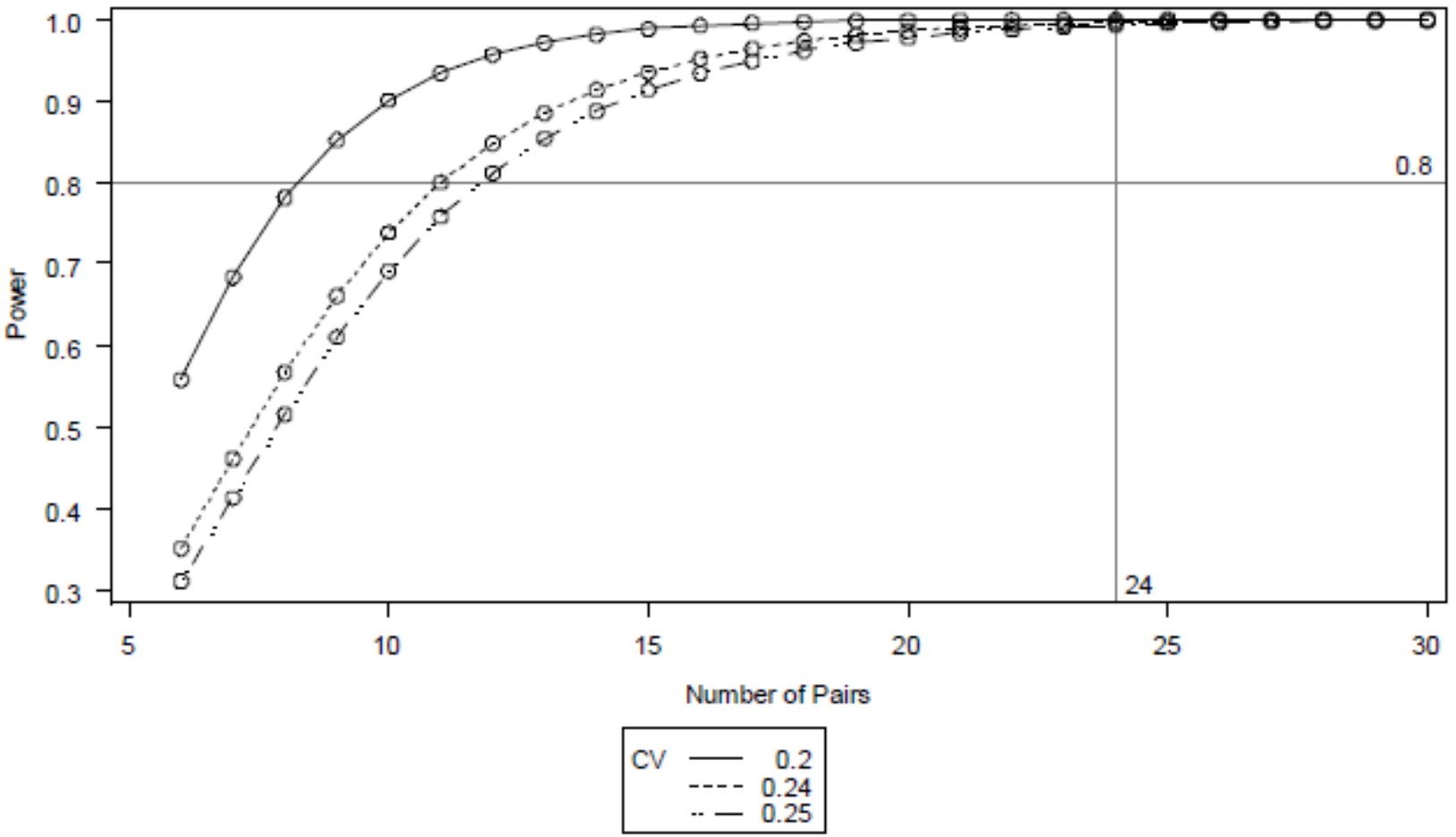
- Goal:
  The probability to reject the hypothesis that the formulations are different when there is no relevent difference should lie e.g. between **80**% and **90**%
- This probability is called Power
- gCV is a decisive factor here

PROC POWER

```
PROC POWER;
    PAIREDMEANS TEST = Equiv_Mult
        LOWER = 0.8
        UPPER = 1.25
        ALPHA = 0.05
        MEANRATIO = 1
        CV =          0.20    0.24    0.25
/*No. of missing values: 20;  16,12,8,4;  0-3*/
        CORR = 0.5
        POWER = .
        NPAIRS = 24
        ;
    PLOT X = n MIN = 6 MAX = 30 VARY(linestyle,symbol)
        YOPTS = (ref = 0.8) XOPTS = (ref = 24)
        /DESCRIPTION = 'Plot of power by sample size';
RUN;
```

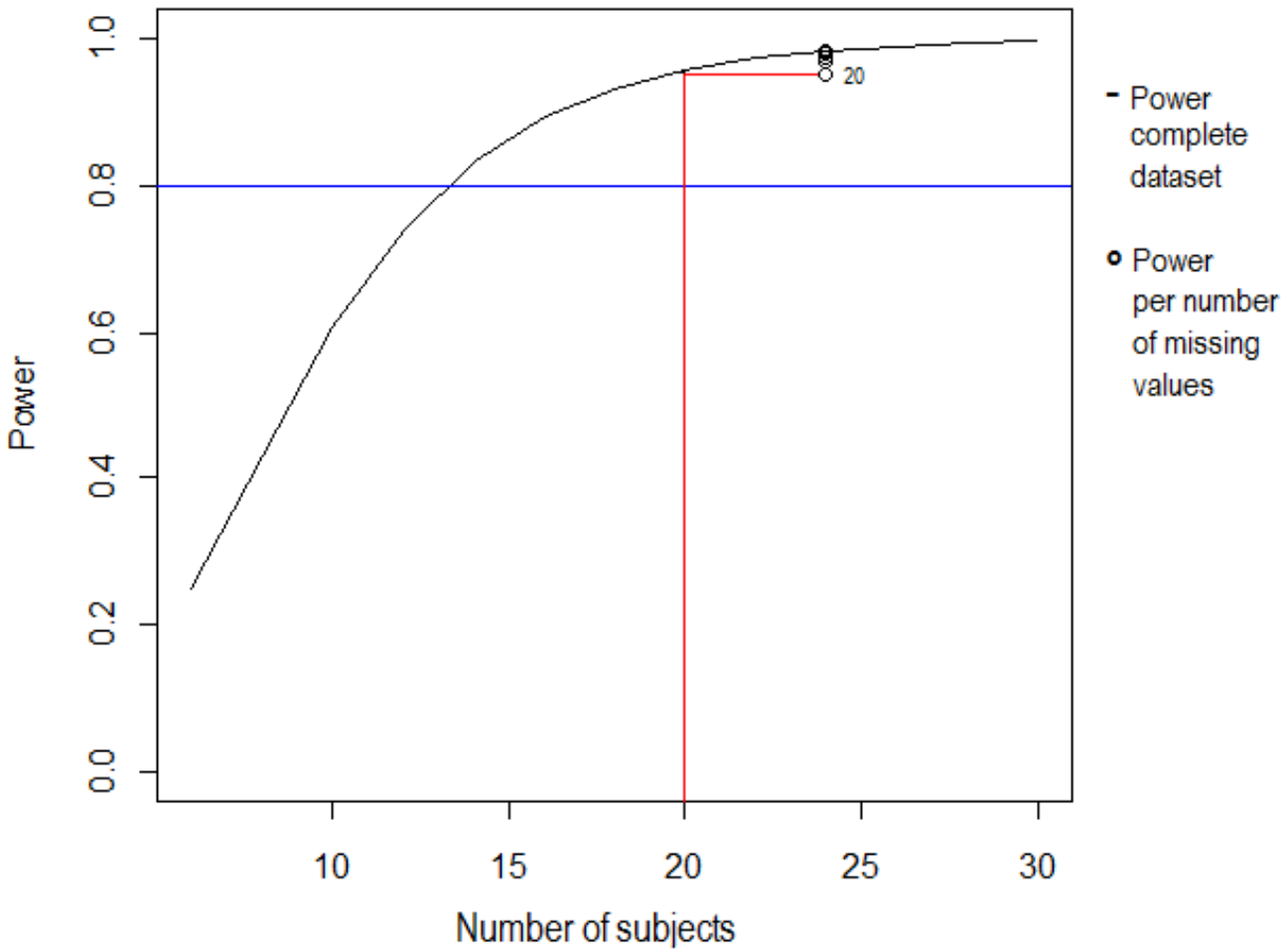
Power graphic: Mean of simulations

$\sigma^2 = 0.06 \triangleright gCV = 25$, missing values for treatment R (case III)



Power graphic: P75 of simulations

$\sigma^2 = 0.06 \triangleright gCV = 25$, missing values for treatment R (case III)



Outline

When should further subjects be recruited?

For N = 24 subjects and 1000 simulations each (per number of missing values and case):

- In general, no further subjects have to be recruited
- Based on the observations in the 75% percentile:
  - at 10 to 12 missing values further subjects should be recruited
  - especially if the values are missing mainly for both treatments and in both periods

Outlook: Further questions

- cohorts and other covariates
- Variability of the rest variance σ^2

References

[1] SAS/Stat 9.2 User's Guide: The Mixed Procedure (Book Excerpt). SAS Publishing, 2008

[2] S.C. Chow, J.P. Liu. Design and Analysis of Bioavailability and Bioequivalence Studies. Marcel Dekker, Inc., New York, 1992

[3] S. Patterson, B. Jones. Bioequivalence and Statistics in Clinical Pharmacology. Chapman & Hall / CRC, Boca Raton, 2006

[4] G.A. Milliken, D.E. Johnson. Analysis of Messy Data Volume I: Designed Experiments. Wadsworth, Inc., Belmont, 1984

Thank you for your attention!

Special thanks to Julia Habeck and Michaela Mattheus from Boehringer Ingelheim Pharma GmbH & Co. KG