Statistical aspects of reference-scaled studies

Helmut Schütz
BEBAC

In vitro in vivo Correlation (IVIVC), Biowaivers & Statistical Aspects of Bioequivalence in Drug Product Development | Mumbai, 29 January 2012
### Hierarchy of Designs

- The more ‘sophisticated’ a design is, the more information (in terms of $\sigma^2$) we may obtain.

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

**Variances which can be estimated:**
- **Parallel:** total variance (between + within)
- **$2 \times 2$ Xover:** + between, within subjects
- **Partial replicate:** + within subjects (reference)
- **Full replicate:** + within subjects (reference, test)
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.
Variance

Power to show BE with 40 subjects for \( CV_{\text{intra}} = 30\text{–}50\% \)

- \( \frac{\mu_T}{\mu_R} = 0.95, CV_{\text{intra}} = 30\% \)  
  \( \rightarrow \) power 0.816

- \( \frac{\mu_T}{\mu_R} = 1.00, CV_{\text{intra}} = 45\% \)  
  \( \rightarrow \) power 0.476 < \textbf{Roulette} 0.486 (!)

- \( \frac{\mu_T}{\mu_R} = 0.95, CV_{\text{intra}} = 50\% \)  
  \( \rightarrow \) \( n=98 \) (power 0.803)
Counterintuitive concept of BE: Two formulations with a large difference in means are declared bioequivalent if variances are low, but not bioequivalent – even if the difference is quite small – due to high variability.

Modified from Fig. 1
L Tóthfalusi, L Endrenyi and A García Arieta
Evaluation of Bioequivalence for Highly Variable Drugs with Scaled Average Bioequivalence
Replicate 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: Same overall number of individual treatments!
Replicate designs

- Required if reference-scaled average bioequivalence (RSABE) is targeted or widening of the AR for $C_{\text{max}}$ (for countries following the ‘old’ EU guideline).

- Advantages
  - Some experience from FDA’s initiative on Population Bioequivalence (PBE) and Individual Bioequivalence (IBE).
  - Mentioned in RSA’s GL; FDA’s API GLs and EMA.
  - RSABE of different metrics acceptable in some countries (FDA, RSA AUC/$C_{\text{max}}$, EMA $C_{\text{max}}$, TGD AUC).
  - Handling of outliers (Subject-by-Formulation Interaction may be ruled out).
Replicate 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 (!)
- Handling of outliers. For the EMA it has to be shown that $\text{CV}_{WR} > 30\%$ is not caused by outliers. Method?
Replicate designs

- Designs
  - Two-sequence three-period
    T R T
    R T R
    Sample size to obtain the same power as a 2×2×2 study: 75%
  - Two-sequence four-period
    T R T R
    R T R T
    Sample size to obtain the same power as a 2×2×2 study: 50%
  - and many others… (FDA: TRR|RTR|RRT aka ‘partial replicate’)
  - The statistical model is quite complicated – and dependent on the actual design!

\[ X_{ijkl} = \mu \cdot \pi_k \cdot \Phi_l \cdot s_{ij} \cdot e_{ijkl} \]
Application: HVDs/HVDPs

- Highly Variable Drugs / Drug Products
  \( \text{(CV}_{WR} > 30 \% \) }

  - USA  Recommended in product specific guidances. GMR 0.80 – 1.25. Minimum sample size 24.

  - CAN 2010 draft GL. Scaling for AUC only. No restriction on GMR.

  ± EU  Widening of acceptance range (for \( \text{C}_{\text{max}} \) only: to maximum 69.84% – 143.19%), if \( \text{CV}_{WR} \) in the study >30%. GMR 0.80 – 1.25. Demonstration that \( \text{CV}_{WR} > 30\% \) not caused by outliers.
Application: HVDs/HVDPs

- All (!) ANDAs submitted to FDA/OGD 2003 – 2005 (1010 studies, 180 drugs)
  - 31% (57/180) highly variable (CV ≥30%)
  - Of these HVDs/HVDPs,
    - 60% due to PK (e.g., first pass metabol.)
    - 20% formulation performance
    - 20% unclear

Highly Variable Drugs: Observations from Bioequivalence Data Submitted to the FDA for New Generic Drug Applications
The AAPS Journal 10/1, 148–56 (2008)
http://www.springerlink.com/content/51162107w327883r/fulltext.pdf

In vitro in vivo Correlation (IVIVC), Biowaivers & Statistical Aspects of Bioequivalence in Drug Product Development | Mumbai, 29 January 2012
HVDPs (US/EU)

- Advisory Committee for Pharmaceutical Sciences (ACPS) to FDA (10/2006) on HVDs
- Follow-up papers in 2008 (ref. in API-GLs)
  - Replicate study design [TRR–RTR–RRT]
  - Reference Scaled Average Bioequivalence (RSABE)
  - Minimum sample size 24 subjects
  - GMR restricted to [0.80, 1.25]

*Bioequivalence Approaches for Highly Variable Drugs and Drug Products*
http://www.springerlink.com/content/u503p62056413677/fulltext.pdf

Haidar SH, Makhlof F, Schuirmann DJ, Hyslop T, Davit B, Conner D, and LX Yu
*Evaluation of a Scaling Approach for the Bioequivalence of Highly Variable Drugs*
The AAPS Journal, 10/3, (2008) DOI: 10.1208/s12248-008-0953-4

*In vitro in vivo* Correlation (IVIVC), Biowaivers & Statistical Aspects of Bioequivalence in Drug Product Development | Mumbai, 29 January 2012
HVDs/HVDPs

- Replicate designs
  - 4-period replicate designs:
    sample size = \( \frac{1}{2} \) of 2×2 study’s sample size
  - 3-period replicate designs:
    sample size = \( \frac{3}{4} \) of 2×2 study’s sample size
  - Reminder: number of treatments (and biosamples) identical to the conventional 2×2 cross-over.
  - Allow for a safety margin – expect a higher number of drop-outs due to the additional period(s).
  - Consider increased blood loss (ethics!)
    Eventually bioanalytics has to be improved.
Tóthfalusi et al. (2009), Fig. 3
Simulated (n=10000) three-period replicate design studies (TRT-RTR) in 36 subjects;
GMR restriction 0.80–1.25. (a) CV=35%, (b) CV=45%, (c) CV=55%.
ABE: Conventional Average Bioequivalence, SABE: Scaled Average Bioequivalence,
0.76: EU criterion, 0.89: FDA criterion.
**HVDPS (US/EU)**

- FDA’s and EMA’s approaches differ; FDA’s leads to a discontinuity of the acceptance range at CV=30%, because FDA’s scaling CV is 25.83% ($\sigma_{WR} 0.294$) – but to be *applied* at CV≥30%.

*In vitro in vivo* Correlation (IVIVC), Biowaivers & Statistical Aspects of Bioequivalence in Drug Product Development | Mumbai, 29 January 2012
HVDPS (No Global Harmonization!)

In vitro in vivo Correlation (IVIVC), Biowaivers & Statistical Aspects of Bioequivalence in Drug Product Development | Mumbai, 29 January 2012
HVDs/HVDPs (Reg. models)

- Common to FDA and EMA
  ABE model
  \[-\theta_A \leq \mu_T - \mu_R \leq +\theta_A\]
  SABE model
  \[-\theta_S \leq \frac{\mu_T - \mu_R}{\sigma_W} \leq +\theta_S\]
  Regulatory regulatory switching condition \(\theta_S\) is derived from the regulatory standardized variation \(\sigma_0\) (proportionality between acceptance limits in ln-scale and \(\sigma_W\) in the highly variable region).

Tóthfalusi et al. (2009)

*In vitro in vivo* Correlation (IVIVC), Biowaivers & Statistical Aspects of Bioequivalence in Drug Product Development | Mumbai, 29 January 2012
HVDs/HVDPs (Reg. models)

- Differences between FDA and EMA
  FDA: Regulatory regulatory switching condition $\theta_S$ is set to 0.893, which would translate into

  \[ CV_{WR} = 100\sqrt{e^{\left(\frac{\ln(1.25)}{0.893}\right)^2}} - 1 \approx 25.83\% \]

  RSABE is allowed only if $CV_{WR} \geq 30\% \ (s_{WR} \geq 0.294)$, which explains to the discontinuity at 30\%.
HVDs/HVDPs (Reg. models)

- Differences between FDA and EMA
  
  EMA: Regulatory regulatory switching condition $\theta_S$ avoids the discontinuity.

$$CV_w = 0.30$$

$$\sigma_0 = \sqrt{\ln(CV_w^2 + 1)} = 0.2935603792085...$$

$$\theta_S = \frac{\ln(1.25)}{\sigma_0} = -\frac{\ln(0.80)}{\sigma_0} \approx 0.760$$
HVDs/HVDPs (FDA)

- Haidar et al. (2008), progesterone guid. (2010)

Starting from the SABE model

\[-\theta_S \leq \frac{\mu_T - \mu_R}{\sigma_W} \leq +\theta_S\]

Rearrangement leads to a linear form

\[\left(\mu_T - \mu_R\right)^2 - \theta_S^2 \cdot \sigma_W^2 \leq 0\]

Since we don’t have the true parameters, we use estimates

\[E_m = \left(\mu_T - \mu_R\right)^2\]

\[E_s = \theta_S^2 \cdot \sigma_W^2\]
HVDs/HVDPs (FDA)


Distributions of $E_m$ and $E_s$ are known and their upper confidence limits can be calculated

\[ C_m = \left( |m_T - m_R| + t_{\alpha,N-S} \cdot SE \right)^2 \]

\[ C_s = \frac{\theta^2_S \cdot (N - S) \cdot s_W^2}{\chi^2_{\alpha,N-S}} \]

$t$ and $\chi^2$ are the inverse cumulative distribution functions at $\alpha 0.05$ and $N - S$ degrees of freedom ($N$ subjects, $S$ sequences). $SE$ is the standard error of the difference between means.
HVDs/HVDPs (FDA)

- Haidar et al. (2008), progesterone guid. (2010)
- Howe method gets the CI from individual CIs

\[
L_m = (C_m - E_m)^2
\]
\[
L_s = (C_s - E_s)^2
\]
\[
CI = E_m - E_s + \sqrt{L_m + L_s}
\]

The CI of the rearranged SABE criterion (slide 20) is evaluated at the 95% level. If the upper 95% is positive RSABE is rejected, and accepted otherwise.
HVDs/HVDPs (EMA)

- EU GL on BE (2010)
  - Average Bioequivalence (ABE) with Expanding Limits (ABEL)
    - The regulatory switching condition $\theta_S$ at $CV_{WR} 30\%$ would be 0.760128297680…
    - According to the GL (2010) and the Q&A document (2011) use $k (\equiv \theta_S)$ with 0.760 (not the exact value).
EU GL on BE (2010)

- If you have $\sigma_{WR}$ (the *intra*-subject standard deviation of the reference formulation) go to the next step; if not, calculate it from $CV_{WR}$

$$
\sigma_{WR} = \sqrt{\ln(CV_{WR}^2 + 1)}
$$

- Calculate the scaled acceptance range based on the regulatory constant $k$ ($\theta_s = 0.760$)

$$
[L,U] = e^{\mp k \cdot \sigma_{WR}}
$$
HVDs/HVDPs (EMA)

- Q&A document (March 2011)
  - Two methods proposed (Method A preferred)
    - **Method A**: All effects fixed; assumes equal variances of test and reference, and no subject-by-formulation interaction; only a common within (*intra*)-subject variance is estimated.
    - **Method B**: Similar to A, but random effects for subjects. Common within (*intra*)-subject variance and between (*inter*)-subject variance are estimated.
  - Outliers: Boxplots (of model residuals?) suggested.

*Questions & Answers on the Revised EMA Bioequivalence Guideline*
*Summary of the discussions held at the 3rd EGA Symposium on Bioequivalence*
June 2010, London
http://www.egagenerics.com/doc/EGA_BEQ_Q&A_WEB_QA_1_32.pdf

*In vitro in vivo* Correlation (IVIVC), Biowaivers & Statistical Aspects of Bioequivalence in Drug Product Development | Mumbai, 29 January 2012
HVDs/HVDPPs (EMA)

- At higher CVs only the GMR is of importance!
- At CVs > 50% still large sample sizes required.
- No commercial software for sample size estimation can handle the GMR restriction.
- Recently sample size tables were published.
- Expect a solution from the community soon…

L Tóthfalusi and L Endrenyi
Sample Sizes for Designing Bioequivalence Studies for Highly Variable Drugs

*In vitro in vivo* Correlation (IVIVC), Biowaivers & Statistical Aspects of Bioequivalence in Drug Product Development | Mumbai, 29 January 2012
Example datasets (EMA)

- Q&A document (March 2011)
  - Data set I
    - RTRT | TRTR full replicate, 77 subjects, imbalanced, incomplete
      - FDA
        - $s_{WR} \geq 0.294 \rightarrow$ apply RSABE ($CV_{WR} 46.96\%$)
        - a. critbound $-0.0921 \leq 0$ and
          b. $80.00\% \leq \text{pointest} 115.46\% \leq 125.00\%$
      - EMA
        - $CV_{WR} 46.96\% \rightarrow$ apply RSABE ($> 30\%$)
        - Scaled Acceptance Range: $71.23\% - 140.40\%$
        - A: $71.23\% \leq 107.11\% - 124.89\% \leq 140.40\%, \text{PE} 115.66\%$
        - B: $71.23\% \leq 107.17\% - 124.97\% \leq 140.40\%, \text{PE} 115.73\%$
Example datasets (EMA)

- Q&A document (March 2011)
  - Data set II
    - TRR | RTR | RRT partial replicate, 24 subjects, balanced, complete
      - FDA
        - $s_{WR} \ 0.114 < 0.294 \rightarrow \text{apply ABE (} CV_{WR} \ 11.43\%)$
        - $80.00\% \leq 97.05 - 107.76 \leq 125.00\% \ (CV_{\text{intra}} \ 11.55\%)$  ✔
      - EMA
        - $CV_{WR} \ 11.17\% \rightarrow \text{apply ABE (} \leq 30\%)$
        - A: 90% CI 97.32% – 107.46%, PE 102.26%  ✔
        - B: 90% CI 97.32% – 107.46%, PE 102.26%  ✔
        - A/B: $CV_{\text{intra}} \ 11.86\%$
Outliers (EMA)

- EU GL on BE (2010), Section 4.1.10
  - The applicant should justify that the calculated intra-subject variability is a reliable estimate and that it is not the result of outliers.

- EGA/EMA Q&A (2010)
  - Q: How should a company proceed if outlier values are observed for the reference product in a replicate design study for a Highly Variable Drug Product (HVDP)?
Outliers (EMA)

EGA/EMA Q&A (2010)

A: The outlier cannot be removed from evaluation [...] but should not be taken into account for calculation of within-subject variability and extension of the acceptance range. An outlier test is not an expectation of the medicines agencies but outliers could be shown by a box plot. This would allow the medicines agencies to compare the data between them.
Outliers (EMA)

- Data set I (full replicate)
  - $\text{CV}_{WR}$ 46.96%
  - ABEL 71.23% – 140.40%
    - Method A: 107.11% – 124.89%
    - Method B: 107.17% – 124.97%
  - But there are two outliers!
    - Excluding subjects 45 and 52
    - $\text{CV}_{WR}$ drops to 32.16%
    - ABEL 78.79% – 126.93%
    - Almost no more gain compared to conventional limits.
Thank You!

Statistical aspects of reference-scaled studies

Open Questions?

(EMA’s and FDA’s SAS code and EMA’s datasets in the handouts)

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*
SAS code (EMA)

Method A

```sas
proc glm data=replicate;
  class formulation subject period sequence;
  model logDATA= sequence subject(sequence) period formulation;
  estimate "test-ref" formulation -1+1;
  test h=sequence e=subject(sequence);
  lsmeans formulation / adjust=t pdiff=control("R") CL alpha=0.10;
run;
```

Method B

```sas
proc mixed data=replicate;
  class formulation subject period sequence;
  model logDATA= sequence period formulation;
  random subject(sequence);
  estimate "test-ref" formulation -1 1 / CL alpha=0.10;
run;
```

$CV_{WR}$ (both methods)

```sas
data var;
  set replicate;
  if formulation='R';
run;
proc glm data=var;
  class subject period sequence;
  model logDATA= sequence subject(sequence) period;
run;
```
SAS code (FDA)

Partial reference-replicated 3-way design

```sas
data test;
  set pk;
  if trt='T';
  latt=lauct;
run;

data ref1;
  set ref;
  if (seq=1 and per=2) or (seq=2 and per=1) or (seq=3 and per=1);
  lat1r=lauct;
run;

data ref2;
  set ref;
  if (seq=1 and per=3) or (seq=2 and per=3) or (seq=3 and per=2);
  lat2r=lauct;
run;
```
Partial reference-replicated 3-way design (cont’d)

```sas
proc glm data=scavbe;
   class seq;
   model ilat=seq/clparm alpha=0.1;
   estimate 'average' intercept 1 seq 0.3333333333 0.3333333333 0.3333333333 0.3333333333 0.3333333333 0.3333333333 0.3333333333 0.3333333333 0.3333333333 0.3333333333 0.3333333333;
   ods output overallanova=iglm1;
   ods output Estimates=iglm2;
   ods output NObs=iglm3;
   title1 'scaled average BE';
run;

pointest=exp(estimate);
x=estimate**2-stderr**2;
boundx=(max((abs(LowerCL)),(abs(UpperCL))))**2;

proc glm data=scavbe;
   class seq;
   model dlat=seq;
   ods output overallanova=dglm1;
   ods output NObs=dglm3;
   title1 'scaled average BE';
run;

dfd=df;
s2wr=ms/2;
```
SAS code (FDA)

Partial reference-replicated 3-way design (cont’d)

\[
\theta = \left( \frac{\log(1.25)}{0.25} \right)^2;
\]

\[
y = -\theta \cdot s_{2wr};
\]

\[
bondy = y \cdot dfd/cinv(0.95, dfd);
\]

\[
s_{WR} = \sqrt{s_{2wr}};
\]

\[
\text{critbound} = (x+y) + \sqrt{((\text{boundx}-x)^2 + ((\text{boundy}-y)^2));}
\]

Apply RSABE if \(s_{WR} \geq 0.294\)

RSABE if

a. \(\text{critbound} \leq 0 \text{ and}\)

b. \(0.8000 \leq \text{pointest} \leq 1.2500\)

If \(s_{WR} < 0.294\), apply conventional (unscaled ABE), mixed effects model.

ABE if 90% CI within \(0.8000\) and \(1.2500\).
SAS code (FDA)

Fully replicated 4-way design

```sas
data test1;
   set test;
   if (seq=1 and per=1) or (seq=2 and per=2);
   lat1t=lauct;
run;

data test2;
   set test;
   if (seq=1 and per=3) or (seq=2 and per=4);
   lat2t=lauct;
run;

data ref1;
   set ref;
   if (seq=1 and per=2) or (seq=2 and per=1);
   lat1r=lauct;
run;

data ref2;
   set ref;
   if (seq=1 and per=4) or (seq=2 and per=3);
   lat2r=lauct;
run;
```
SAS code (FDA)

Fully replicated 4-way design (cont’d)

```sas
data scavbe;
  merge test1 test2 ref1 ref2;
  by seq subj;
  ilat=0.5*(lat1t+lat2t-lat1r-lat2r);
  dlat=lat1r-lat2r;
run;

proc mixed data=scavbe;
  class seq;
  model ilat =seq/ddfm=satterth;
  estimate 'average' intercept 1 seq 0.5 0.5/e cl alpha=0.1;
  ods output CovParms=iout1;
  ods output Estimates=iout2;
  ods output NObs=iout3;
  title1 'scaled average BE';
  title2 'intermediate analysis - ilat, mixed';
run;

pointest=exp(estimate);
x=estimate**2-stderr**2;
boundx=(max((abs(lower)),(abs(upper))))**2;
```

*In vitro in vivo* Correlation (IVIVC), Biowaivers & Statistical Aspects of Bioequivalence in Drug Product Development | Mumbai, 29 January 2012
SAS code (FDA)

Fully replicated 4-way design (cont’d)

```sas
proc mixed data=scavbe;
   class seq;
   model dlat=seq/ddfm=satterth;
   estimate 'average' intercept 1 seq 0.5 0.5/e cl alpha=0.1;
   ods output CovParms=dout1;
   ods output Estimates=dout2;
   ods output NObs=dout3;
   title1 'scaled average BE';
   title2 'intermediate analysis - dlat, mixed';
run;

s2wr=estimate/2;
dfd=df;

theta=((log(1.25))/0.25)**2;
y=-theta*s2wr;
boundy=y*dfd/cinv(0.95,dfd);
sWR=sqrt(s2wr);
critbound=(x+y)+sqrt(((boundx-x)**2)+((boundy-y)**2));
```

---

In vitro in vivo Correlation (IVIVC), Biowaivers & Statistical Aspects of Bioequivalence in Drug Product Development | Mumbai, 29 January 2012
**SAS code (FDA)**

Unscaled 90% BE confidence intervals (applicable if critbound>0)

```
PROC MIXED
  data=pk;
  CLASSES SEQ SUBJ PER TRT;
  MODEL LAUCT = SEQ PER TRT/ DDFM=SATTERTH;
    RANDOM TRT/TYPE=FA0(2) SUB=SUBJ G;
  REPEATED/GRP=TRT SUB=SUBJ;
  ESTIMATE 'T vs. R' TRT 1 -1/CL ALPHA=0.1;
ods output Estimates=unsc1;
title1 'unscaled BE 90% CI - guidance version';
title2 'AUCt';
run;

data unsc1;
  set unsc1;
  unscabe_lower=exp(lower);
  unscabe_upper=exp(upper);
run;
```

**Note:** Lines marked with an arrow are missing in FDA's code!
Example datasets (EMA)

- Q&A document (March 2011)
  - Data set I
    4-period 2-sequence (RTRT | TRTR) full replicate, imbalanced (77 subjects), incomplete (missing periods: two periods in two cases, one period in six cases).
  - Data set II
    3-period 3-sequence (TRR | RTR | RRT) partial replicate, balanced (24 subjects), complete (all periods).