Reference-scaled Average Bioequivalence

Helmut Schütz
Study Designs

The more ‘sophisticated’ a design is, the more information can be extracted.

- Hierarchy of designs:
  - Full replicate (RTRT | TRTR or RTR | TRT)
  - Partial replicate (RRT | RTR | TRR)
  - $2 \times 2 \times 2$ crossover (RT | TR)
  - Parallel (R | T)

- Variances which can be estimated:
  - Parallel: total variance (pooled of between + within subjects)
  - $2 \times 2 \times 2$ crossover: + between, within subjects
  - Partial replicate: + within subjects (of R)
  - Full replicate: + within subjects (of R and T)
Highly Variable Drugs / Drug Products

Counterintuitive concept of BE:

Two formulations with a large difference in means are declared bioequivalent if variabilities are low, but not BE – even if the difference is quite small – due to high variability.

Modified from Tothfálusi et al. (2009), Fig. 1
HVD(P)s – Reference-scaling

It may be almost impossible to demonstrate ABE with a reasonable sample size.

- Reference-scaling (i.e., widening the acceptance range based on the variability of the reference) introduced in 2010 by the FDA and EMA and in 2016 by Health Canada.
  - Requires a replicate design, where at least the reference product is administered twice.
  - Smaller sample sizes compared to a standard $2\times2\times2$ design but outweighed by increased number of periods.
  - Similar total number of individual treatments.
  - Any replicate design can be evaluated for ‘classical’ (unscaled) Average Bioequivalence (ABE) as well. Switching $CV_{wr}$ 30%:
    - FDA: $AUC$ and $C_{max}$
    - EMA: $C_{max}$; MR products additionally: $C_{ss,min}$, $C_{ss,\tau}$, partial $AUC$s
    - Health Canada: $AUC$
HVD(P)s – Reference-scaling

Models (in log-scale).

• ABE Model:
  – A difference $\Delta$ of $\leq 20\%$ is considered to be clinically not relevant.
  – The limits $[L, U]$ of the acceptance range are fixed to
    $\log(1 – \Delta) = \log((1 – \Delta)^{-1})$ or $L \sim -0.2231$ and $U \sim +0.2231$.
  – The consumer risk ($\alpha$) is fixed with 0.05. BE is concluded if the $100(1 – 2\alpha)$ confidence interval lies entirely within the acceptance range.

\[ -\theta_A \leq \mu_T - \mu_R \leq +\theta_A \]

• SABEL Model:
  – Switching condition $\theta_S$ is derived from the regulatory standardized variation $\sigma_0$ (proportionality between acceptance limits in log-scale and $\sigma_{WR}$ in the highly variable region).

\[ -\theta_S \leq \frac{\mu_T - \mu_R}{\sigma_{WR}} \leq +\theta_S \]
HVD(P)s – Reference-scaling

Regulatory Approaches.

- Bioequivalence limits derived from $\sigma_0$ and $\sigma_{WR}$
  \[ \theta_s = \frac{\log(1.25)}{\sigma_0}, \quad [L, U] = e^{\pm \theta_s \cdot \sigma_{WR}} \]

- FDA
  - Scaling $\sigma_{WR}$ 0.25 ($\theta_s$ 0.893) but applicable at $CV_{WR} \geq 30\%$.
  - Discontinuity at $CV_{WR}$ 30\%.

- EMA
  - Scaling $\sigma_0$ 0.2936 ($\theta_s$ 0.760).
  - Upper cap at $CV_{WR}$ 50\%.

- Health Canada
  - Like EMA but upper cap at $CV_{WR}$ 57.4\%.
HVD(P)s – Reference-scaling

The EMA’s Approach.

- Average Bioequivalence with Expanding Limits – ABEL (crippled from Endrényi and Tóthfalusi 2009).
  - Justification that the widened acceptance range is clinically not relevant (important – different to the FDA).
  - Assumes identical variances of T and R [sic] like in a $2 \times 2 \times 2$.
  - All fixed effects model according to the Q&A-document preferred.
  - Mixed-effects model (allowing for unequal variances) is ‘not compatible with CHMP guideline’…
  - Scaling limited at a maximum of $CV_{wr}$ 50% (i.e., to 69.84 – 143.19%).
  - $GMR$ within 0.8000 – 1.2500.
  - Demonstration that $CV_{wr} >30\%$ is not caused by outliers (box plots of studentized intra-subject residuals?)…
  - $\geq 12$ subjects in sequence RTR of the 3-period full replicate design.
The EMA’s Approach.

- **Decision Scheme.**
  - The Null Hypothesis is *specified* in the face of the data.
  - Acceptance limits themselves become random variables.
  - Type I Error (consumer risk) might be inflated.

\[
CV_{wR} = 100 \sqrt{e^{s_{wR}^2} - 1}
\]

\[
\begin{align*}
100(1-2\alpha) CI \in [L,U] &= 80.00\%-125.00\% \\
100(1-2\alpha) CI \in [L,U] &= 100e^{20.760s_{wR}} \\
GMR \in [L,U] &= 80.00\%-125.00\%
\end{align*}
\]

- **Pass**
- **Fail**
Assessing the Type I Error (TIE).

- **TIE** = falsely concluding BE at the limits of the acceptance range.
- Due to the decision scheme direct calculation of the TIE at the scaled limits is not possible; → extensive simulations required (10^6 BE studies mandatory).
- Confirmed.
  - EMA’s ABEL
  - FDA’s RSABE
HVD(P)s – Reference-scaling

Example for ABEL

- RTRT | TRTR
  sample size 18 – 96
  $CV_{WR}$ 20% – 60%
  - $TIE_{max}$ 0.0837.
  - Relative increase of the consumer risk 67%! 
HVD(P)s – Reference-scaling

What is going on here?

- SABE is stated in model parameters ...
  \[-\theta_S \leq \frac{\mu_T - \mu_R}{\sigma_{wR}} \leq +\theta_S\]
  ... which are unknown.

  - Only their estimates (GMR, s_{wR}) are accessible in the actual study.
  - At CV_{wR} 30% the decision to scale will be wrong in ~50% of cases.
  - If moving away from 30% the chances of a wrong decision decrease and hence, the TIE.
  - At high CVs (>43%) both the scaling cap and the GMR-restriction help to maintain the TIE <0.05).
HVD(P)s – Reference-scaling

Outlook.

- **Utopia**
  - Agencies collect $CV_{wR}$ from submitted studies. Pool them, adjust for designs / degrees of freedom. The EMA publishes a fixed acceptance range in the product-specific guidance. No need for replicate studies any more. $2 \times 2 \times 2$ crossovers evaluated by ABE would be sufficient.

- **Halfbaked**
  - Hope [sic] that e.g., Bonferroni preserves the consumer risk. Still apply ABEL, but with a 95% CI ($\alpha 0.025$).
  - Drawback: Loss of power, substantial increase in sample sizes.

- **Proposal**
  - Iteratively adjust $\alpha$ based on the study’s $CV_{wR}$ and sample size – in such a way that the consumer risk is preserved (Labes and Schütz 2016).
ABEL (iteratively adjusted $\alpha$)

Previous example

- **Algorithm**
  - Assess the TIE for the nominal $\alpha = 0.05$.
  - If the TIE $\leq 0.05$, stop.
  - Otherwise adjust $\alpha$ (downwards) until the TIE $\cong 0.05$.
  - At $CV_{wR}$ 30% (dependent on the sample size) $\alpha_{adj}$ is $0.0273 - 0.0300$; → use a 94.00 – 94.54% CI.
Potential impact on the sample size.

- Example: RTRT | TRTR, $\theta_0$ 0.90, target power 0.80.
  - Moderate in the critical region (—).  
    - $CV_{wR}$ 30%: 36 → 42 (+17%);
    - $CV_{wR}$ 35%: 34 → 38 (+12%);
    - $CV_{wR}$ 40%: 30 → 32 (+7%).
  - None outside (—).
ABEL (iteratively adjusted $\alpha$)

Example (RTRT | TRTR, expected $CV_{wR}$ 35%, $\theta_0$ 0.90, target power 0.80); R package PowerTOST ($\geq 1.3$-3).

- **Estimate the sample size.**
  
  ```r
  sampleN.scABEL(CV=0.35, theta0=0.90, targetpower=0.80, design="2x2x4", 
                 details=FALSE, print=FALSE)["Sample size"]
  
  [1] 34
  ```

- **Estimate the empiric TIE for this study.**
  
  ```r
  UL <- scABEL(CV=0.35)["upper"] # scaled limit (1.2948 for CVwR 0.35)
  power.scABEL(CV=0.35, theta0=UL, n=34, design="2x2x4", nsims=1e6)
  
  [1] 0.065566
  ```

- **Iteratively adjust $\alpha$.**
  
  ```r
  scABEL.ad(CV=0.35, n=34, design="2x2x4")
  
  iteratively adjusted alpha
  
  iteratively adjusted alpha
  
  CVwR 0.35, n(1) 17|17 (N 34)
  Nominal alpha : 0.05
  Null (true) ratio : 0.9000
  Regulatory settings : EMA (ABEL)
  Empiric TIE for alpha 0.0500 : 0.06557
  Power for theta0 0.900 : 0.812
  Iteratively adjusted alpha : 0.03630
  Empiric TIE for adjusted alpha : 0.05000
  Power for theta0 0.900 : 0.773
  ```
Optionally compensate for the loss in power (0.812 → 0.773) by increasing the sample size:

```
sampleN.scABEL.ad(CV=0.35, theta0=0.90, targetpower=0.80, design="2x2x4")
```

Sample size estimation for iteratively adjusted alpha

Study design: 2x2x4 (RTRT|TRTR)

Expected CVwR 0.35
Nominal alpha : 0.05
Null (true) ratio : 0.9000
Target power : 0.8
Regulatory settings: EMA (ABEL)
Switching CVwR : 30%
Regulatory constant: 0.760
Expanded limits : 0.7723...1.2948
Upper scaling cap : CVwR 0.5
PE constraints : 0.8000...1.2500

\[ n = 38, \quad \text{adj. alpha: 0.03610 (power 0.8100), TIE: 0.05000} \]

\[ n = 34 \rightarrow 38 (+12\%), \text{ power 0.773} \rightarrow 0.810, \alpha_{adj} 0.0363 \rightarrow 0.0361. \]
Allowing ABEL only for $C_{\text{max}}$.

- Some drugs show high variability in $AUC$ as well.
  - Since in such a case the sample size is mandated by $AUC$, products with high deviations in $C_{\text{max}}$ will be approved.
  - Example: $CV_{wR}$ 90% ($C_{\text{max}}$), 60% ($AUC$), $\theta_0$ 0.90, target power 80% → the study is ‘overpowered’ for $C_{\text{max}}$; $C_{\text{max}}$-GMRs of [0.846 – 1.183] will pass BE. Really desirable?
  - With the FDA’s RSABE the study could be performed in only 34 subjects…
Thank You!

Open Questions?

Helmut Schütz
BEBAC
Consultancy Services for Bioequivalence and Bioavailability Studies
1070 Vienna, Austria
helmut.schuetz@bebac.at
The fundamental cause of trouble in the world today is that the stupid are cocksure while the intelligent are full of doubt.  

Bertrand Russell

100% of all disasters are failures of design, not analysis.  

Ronald G. Marks

My definition of an expert in any field is a person who knows enough about what’s really going on to be scared.  

Phillip J. Plauger
Backup

Example for the FDA’s RSABE

- **RTRT | TRTR**
  - sample size $18 - 96$
  - $CV_{wr} 20\% - 60\%$
  - $TIE_{\text{max}} 0.2245.$
  - Relative increase of the consumer risk 349%!
  - $TIE$ more dependent on the sample size than in ABEL.
  - However, no inflation of the $TIE$ for $CV_{wr} >30\%$; RSABE is very conservative for ‘true’ HVD(P)s.
FDA’s desired consumer risk model (Davit et al. 2012)

- Previous example
  - TIE assessed not at the scaled limits but
    - at 1.25 if $CV_{WR} \leq 25.4\%$
    - or
    - at $e^{0.893 \cdot \sigma_{WR}}$ otherwise.
  - $TIE_{\text{max}} = 0.0668$.
  - Lászlo Endrényi: “Hocus pocus!”
References


