Bioequivalence: The Case for Scaling

Robert Schall
University of the Free State
Bloemfontein, South Africa

20 October 2009

Disclaimer

The views and opinions expressed in the following PowerPoint slides are those of the individual presenter and should not be attributed to Drug Information Association, Inc. (“DIA”), its directors, officers, employees, volunteers, members, chapters, councils, Special Interest Area Communities or affiliates, or any organization with which the presenter is employed or affiliated.

These PowerPoint slides are the intellectual property of the individual presenter and are protected under the copyright laws of the United States of America and other countries. Used by permission. All rights reserved. Drug Information Association, DIA and DIA logo are registered trademarks or trademarks of Drug Information Association Inc. All other trademarks are the property of their respective owners.
Acknowledgements

Laszlo Endrenyi, University of Toronto
Laszlo Tothfalusi, Semmelweis University, Budapest
Arne Ring, Boehringer Ingelheim, Germany

for their published and to be published work on the topic, for providing other relevant information, and for discussions.

However, all errors of judgement and fact are mine.

Average Bioequivalence

Assessment of bioequivalence:
• Compare two products: Test (T) and Reference (R)

Compare T and R w.r.t. bioavailability (BA) metrics:
• AUC, $C_{\text{max}}$, etc

Assessment of average bioequivalence:
• Compare geometric means (GM) of BA metrics of T and R
**Average Bioequivalence**

**Criterion for average bioequivalence:**
- Geometric mean ratio (GMR) of BA metrics of T and R falls in acceptance range

**Statistical decision rule:**
- Declare average bioequivalence if 90% CI for GMR falls in acceptance range

**Acceptance range:**

\[
[0.8, 1.25]
\]

---

**Average Bioequivalence**

Log scale: Criterion for average bioequivalence

\[
\log(0.8) \leq \mu_T - \mu_R \leq \log(1.25)
\]

- Confidence interval for \( \mu_T - \mu_R \):

\[
(m_T - m_R) \pm t \times (2s^2/n)^{1/2}
\]

- Bioequivalence can be declared if \( [\theta_A=\log(1.25)] \):

\[
-\theta_A \leq (m_T - m_R) \pm t \times (2s^2/n)^{1/2} \leq \theta_A
\]
Problem of Highly Variable Drugs

Statistical decision rule:
- Declare average bioequivalence if 90% CI for GMR fits into acceptance range
- Narrow CI: Easy to fit
- Wide CI: Difficult to fit

- Width of confidence interval:
  - Increases with increasing variability
  - Decreases with increasing sample size

Problem of Highly Variable Drugs

- Bioequivalence can be declared if:

  $$-\theta_A \leq (m_T-m_R) \pm t \times (2s^2/n)^{\frac{1}{2}} \leq \theta_A$$

- Thus: CI must fit into the acceptance range $$[-\theta_A, \theta_A]$$

- Width of confidence interval proportional to: $$s^2/n$$
  - Increases with increasing variability: $$s$$
  - Decreases with increasing sample size: $$n$$
Problem of Highly Variable Drugs

To show bioequivalence:
• CI must fit into the acceptance range (with high probability / power)

Therefore:
If variability of drug is high, sample size must be high

Problem of Highly Variable Drugs

Sample size requirements (power = 90%):
• Bioequivalence study of two identical drug products:

<table>
<thead>
<tr>
<th>CV(%)</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>20</td>
<td>28</td>
<td>40</td>
<td>66</td>
<td>100</td>
<td>136</td>
<td>176</td>
</tr>
</tbody>
</table>

• Highly variable drugs mean high sample sizes
• Very Highly variable drugs mean very high sample sizes
• Highly variable drug (HVD): CV at least 30%
Problem of Highly Variable Drugs

Sample sizes seen in recent years:
• Numerous studies with n=30 to n=60
• Studies with n=150
• Study planned with n>200?
• All studies with 4-period replicate designs!

Problem of Highly Variable Drugs

Concerns with high sample sizes / HVDs:
• Cost: Impedes both generic and new drug development
• Ethics: Expose too many subjects to drug
• Paradox: Cannot show bioequivalence of drug product to itself
Potential Solutions

Problem:
- Must fit CI into acceptance range \([-\theta_A, \theta_A]\)
- Width of CI proportional to \(s^2/n\)

Potential solutions for HVDs:
1. Decrease variability (usually not practical)
2. Increase sample size (problematic as discussed)
3. Widen the acceptance range

Widening the Acceptance Range

Conventional acceptance range / criterion:

\[-\theta_A \leq \mu_T - \mu_R \leq \theta_A, \quad \theta_A = \log(1.25)\]

Proposals:
1. “Direct” widening
2. “Scaling” the limit
3. “Scaling” the criterion

(Haidar et al 2008)
1. Direct Widening

Conventional acceptance range / criterion:

$$\log(0.80) \leq \mu_T - \mu_R \leq \log(1.25)$$

“Direct” widening: Use

$$\log(0.75) \leq \mu_T - \mu_R \leq \log(1.33)$$

$$\log(0.70) \leq \mu_T - \mu_R \leq \log(1.43)$$

Problems:

- Rather arbitrary: which range for which variability?
- Discontinuous: small increase in variability might pay!
- Regulatory fairness: Which range to use for which drug?
2. Scaling the Limit

Conventional acceptance range / criterion:

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

Scaling the limit: (Boddy et al 1995)

- Scale acceptance limit proportional to variability of drug
- Highly variable drug: Wide acceptance limits

Problems:

- Regulatory limit \(\theta_S \sigma_{WR}\) depends on unknown \(\sigma_{WR}\)
- Statistically unsatisfactory (in that form)
3. Scaling the Criterion

Conventional acceptance range / criterion:

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

Scaling the criterion:

- Scale mean difference inversely proportional to variability of drug
- Highly variable drug: Small scaled mean difference

Among authors who prefer scaling, “scaling the criterion” is preferred approach (incl. FDA authors)
Scaled Criterion

Scaled criterion:

\[-\theta_s \leq \frac{(\mu_T-\mu_R)}{\sigma_{WR}} \leq \theta_s\]

Why is it a good idea?
1. Equivalent to “scaling the limit” but statistically sound
2. Regulatory constant \(\theta_s\) easily determined
3. Special case of switchability/IBE criterion
4. Sensible criterion for T/R distance between distributions of BA metrics

1. Equivalence to “Scaling the Limit”

Scaling the criterion:

\[-\theta_s \leq \frac{(\mu_T-\mu_R)}{\sigma_{WR}} \leq \theta_s\]

Multiply all sides by \(\sigma_{WR}\):

\[-\theta_s \sigma_{WR} \leq \mu_T-\mu_R \leq \theta_s \sigma_{WR}\]

• “Scaling the criterion” equivalent to “scaling the limit”
• But: “Scaling the criterion” is statistically sound
1. Equivalence to “Scaling the Limit”

Scaled criterion: \[-\theta_S \leq (\mu_T - \mu_R) / \sigma_{WR} \leq \theta_S\]
Equivalent to: \[-\theta_S \sigma_{WR} \leq \mu_T - \mu_R \leq \theta_S \sigma_{WR}\]

2. Determination of Regulatory Limit

Acceptance range (in terms of \(\mu_T - \mu_R\)) of scaled criterion:

- **Wider** than conventional \((\theta_A)\) for HVDs
- **Same** as conventional \((\theta_A)\) for non-HVDs (at a suitable cut-off)
2. Determination of Regulatory Limit

\[ \theta_S \sigma_0 = \theta_A \]

HVD: CV=30% or \( \sigma_0 = 0.294 \):

Therefore:

\[ \theta_S = \theta_A / \sigma_0 \]
\[ = \log(1.25)/0.294 \]
\[ = 0.760 \]
3. Special Case of IBE Metrics

**General principle:**
- Test is bioequivalent to Reference if Test is “close” to Reference in some sense

**Question:** What is “close”?  

**Basic idea:**
- Test is bioequivalent to Reference if Test is as close to Reference, or “nearly” as close, as the Reference is to itself  
(Schall and Luus1993, Schall 1995b)

(Almost) all PBE/IBE metrics can be derived using above principle  
(Schall 1995b)

**Under reasonable assumptions:**
- Scaled average bioequivalence emerges as special case
3. Special Case of IBE Metrics

For example: IBE criterion

\[
\frac{[(\mu_T - \mu_R)^2 + (\sigma_{WT}^2 - \sigma_{WR}^2) + \sigma^2_D]}{\sigma_{WR}^2} \leq \theta_S^2
\]

If: \( \sigma_{WT}^2 - \sigma_{WR}^2 = 0 \) and \( \sigma_D^2 = 0 \)

\[
(\mu_T - \mu_R)^2 / \sigma_{WR}^2 \leq \theta_S^2
\]

Equivalent to:

\[-\theta_S \leq (\mu_T - \mu_R) / \sigma_{WR} \leq \theta_S\]

4. Distance between Distributions

Average bioequivalence criterion:

- Measures distance between T and R in terms of mean bioavailability of T and R

“Better” criterion:

- Consider more than just mean BA metrics
- ... Distributions of BA metrics should be close
4. Distance between Distributions

Goal:
- Bioavailability $Y_T$ of T and $Y_R$ of R are “close” in distribution

What is “close” in distribution?
- Large probability that difference $Y_T - Y_R$ is small

Under reasonable assumptions:
- Scaled average bioequivalence emerges as special case

Bioequivalence criterion:
\[
\text{Prob}(|Y_T - Y_R| < \gamma \cdot \sigma_W) > \text{MINP}
\]

Equivalent to:
\[
-\theta_S \leq (\mu_T - \mu_R)/\sigma_W \leq \theta_S
\]

Therefore the following are equivalent:
- T and R are scaled average bioequivalent
- Guarantee a minimum probability that bioavailabilities of T and R are close
4. Distance between Distributions

(Tothfalusi, Endrenyi and Arieta 2009)

Summary

Scaled average bioequivalence
- Provides potential solution to HVD problem
- Equivalent to all (most?) reasonable proposals in this regard
- Emerges as special case of IBE/Switchability metric
- Regulatory constant easily determined
- Sensible criterion for T/R distance between distributions of BA metrics
- Can also be motivated as “effect size” metric
- Statistical methods / study designs available and practical
Practical Implementation Issues

**Designs:**
- 3- or 4-period replicate design cross-over studies
- (I prefer 4-period design)

**Choice of scaled vs unscaled approach:**
- Based on observed intra-subject CV of drug
- Based on prior knowledge of variability of drug and pre-specified in study protocol (I prefer that)

Potential Problems

**Scaled average bioequivalence**

1. May allow for differences in mean bioavailability of more than 30%, 40% etc.
2. Outliers
1. Large Mean Differences

Scaled average bioequivalence
• May allow for differences in mean bioavailability of more than 30%, 40% etc.

Counter-counter argument
• So what?
• A CV of 40%, 50%, 60% implies that bioavailabilities of T and R typically differ by 40%, 50%, 60%
• Does not matter whether difference of 40% arises from mean difference or because of high variability
1. Large Mean Differences

Therapeutic Range

<table>
<thead>
<tr>
<th>Variability</th>
<th>Narrow</th>
<th>Wide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>High</td>
<td>Not possible</td>
<td>✓</td>
</tr>
</tbody>
</table>

(Tothfalusi, Endrenyi and Arieta 2009)
1. Large Mean Differences

**Scaled average bioequivalence**
- May allow for differences in mean bioavailability of more than 25%, 30% etc.

**No problem for the “single” dose situation**

**Chronic use / accumulation?**
- Might have “long-term” differences of > 30% etc.
- Potentially require steady-state studies: Measure AUC over 2-4 consecutive days (say)

2. Outliers

**Average bioequivalence:**
- Outliers tend to lower power of showing bioequivalence
- Certain robustness
- In the sponsors’ interest to decrease variability of drug products and “prevent” outliers

**Scaled average bioequivalence:**
- Outliers (in R product) may increase power of showing bioequivalence
- “Anti-conservative” property
2. Outliers

Scaled average bioequivalence criterion:

\[-\theta_s \leq (\mu_T - \mu_R) / \sigma_{WR} \leq \theta_s\]

- Outlier in R product will inflate $\sigma_{WR}$ and cause “unjustified” scaling
- Might increase consumer’s risk

Further Research

Robust statistical methods
- Outlier problem

Need for steady-state studies?
- For drugs to be used chronically
- Long-term differences in exposure of 40%, 50% etc might be problematic
- But: Long-term drug exposure might be less variable than single-dose exposure
- Clinical implications of various patterns of variance components
Some References


Schall R (1995a). Assessment of individual and population bioequivalence using the probability that bioavailabilities are similar. Biometrics, 51: 615-626


Schall R, Luus HG (1993). On population and individual bioequivalence. Statistics in Medicine, 12:1109-1124
