Rational Statistical Analysis Practice in Dissolution Profile Comparison for Product Quality Assessment of Similarity through Real Case Studies: Industry Perspective

May 21, 2019

Yanbing Zheng, Jian-Hwa Han, James Reynolds, Mark Johnson, Karin Rosenblatt, Tzuchi R Ju, Yi Gao, Bei Chen, Hesham Fahmy
Disclosure

This presentation was sponsored by AbbVie. AbbVie contributed to the design, research, and interpretation of data, writing, reviewing, and approving the publication. The authors are employees of AbbVie, Inc.
Outline

• Model independent statistical methods
• Simulation studies
• Decision tree
• R Shiny tool
• Case studies
$f_2$ Rules (FDA 1997 Guidance)

- N=12 of (i) Reference (or prechange) and (ii) Test (or postchange) products
- Use the Mean values only for calculation
- Model Independent Method - most suitable for dissolution profile comparison when three to four or more dissolution time points are available
  - Same time points (minimally 3 times points)
  - Only one measurement should be considered after 85% dissolution of both the products
  - %RSD – NMT 20% at early points (e.g. 10 minutes); NMT 10% for all other points
What if $f_2$ assumptions are not satisfied?

- It is critical to identify a right tool/method in order to make meaningful assessment for product quality
  - Model independent statistical methods
    - $f_2$ bootstrap (Shah, et al. 1998)
    - Tsong’s MSD method (Tsong, et al. 1996)
    - SK method (Saranadasa and Krishnamoorthy 2005)
    - Saranadasa’s Hotelling’s $T^2$ based method (Saranadasa 2001)
    - Intersection union test (Berger and Hsu 1996)

- Simulation studies were performed to evaluate the power and type I error of different approaches.

- More than 250 cases were used for the establishment of decision tree and assessment.
Model Independent Statistical Methods

Methods are based on some function of the distance between the profiles at each time point

- $f_2$ – Euclidean distance (pythagorean theorem) based on equal weights $(1/p)$

- Tsong’s MSD and Hotelling’s $T^2$– Euclidean distance weighted by standard deviations and correlations

- SK – common distance weighted by complex function of standard deviations and correlations

- Intersection Union Test – maximum distance weighted by standard deviations
## Statistical Methods for Dissolution Profile Comparison

<table>
<thead>
<tr>
<th>Methods</th>
<th>Pros</th>
<th>Cons</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Similarity factor $f_2$</td>
<td>• <strong>Simple</strong></td>
<td>• Uses <strong>only the mean profile</strong></td>
<td>• FDA requirements: %CV $\leq$ 20% at the earlier time points and $\leq$ 10% at other time points.</td>
</tr>
<tr>
<td></td>
<td>• <strong>Common acceptable cutoff</strong>: 50</td>
<td>• Loses applicability when variability increases</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Lack of type I error control</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Unknown statistical distribution</td>
<td></td>
</tr>
<tr>
<td>$f_2$ bootstrap</td>
<td>• Considers profile mean and <strong>variation</strong></td>
<td>• Could be <strong>conservative</strong></td>
<td>• Recommended when $f_2$ usage requirements on variation are exceeded.</td>
</tr>
<tr>
<td></td>
<td>• <strong>Common acceptable cutoff</strong>: 50</td>
<td></td>
<td>• Strong regulatory connection.</td>
</tr>
<tr>
<td>Tsong’s Multivariate</td>
<td>• Considers profile mean and <strong>variation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>statistical distance (MSD)</td>
<td>• Real case studies suggest good statistical <strong>power</strong> of claiming similarity and <strong>type I error control</strong>.</td>
<td>• <strong>Cutoff is random and data dependent</strong></td>
<td>• No common acceptable cutoff.</td>
</tr>
<tr>
<td>method</td>
<td></td>
<td></td>
<td>• Strong regulatory connection.</td>
</tr>
</tbody>
</table>
## Statistical Methods for Dissolution Profile Comparison

<table>
<thead>
<tr>
<th>Methods</th>
<th>Pros</th>
<th>Cons</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saranadasa and Krishnamoorthy’s (SK) method</td>
<td>• Considers profile mean and <strong>variation</strong></td>
<td>• Assumes <strong>parallelism</strong> of the two dissolution profiles</td>
<td>• The assumption is usually not satisfied in practice.</td>
</tr>
<tr>
<td></td>
<td>• <strong>Cutoff 10%</strong> approximately corresponds to ( f^2_{50} )</td>
<td>• <strong>Cutoff value 6%</strong> was proposed.</td>
<td></td>
</tr>
<tr>
<td>Sarandasa's Hotelling ( T^2 )-based method</td>
<td>• Considers profile mean and <strong>variation</strong></td>
<td>• Assumes <strong>parallelism</strong> of the two dissolution profiles</td>
<td>• The assumption is usually not satisfied in practice.</td>
</tr>
<tr>
<td></td>
<td>• <strong>Cutoff value 6%</strong> was proposed.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intersection-Union Test</td>
<td>• Considers profile mean and <strong>variation</strong></td>
<td>• Time points are considered independently</td>
<td>• Too conservative</td>
</tr>
<tr>
<td></td>
<td>• Be able to <strong>identify the time-point(s) that does not show similarity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model-dependent approaches</td>
<td>• Measurements can be taken at <strong>different time points for reference and test batches.</strong></td>
<td>• <strong>Model selection</strong></td>
<td>• Appropriate when dissolution curves are sampled at many time points.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Spacing of time points may limit curve/model choices</td>
<td>Hard to have a common acceptable cutoff.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• <strong>Cutoff selection</strong></td>
<td></td>
</tr>
</tbody>
</table>
Simulation Study

• Mean for test profile = (35, 45, 70, 85) and compound symmetry covariance structure with correlation = 0.5.

• Assume equal covariance matrices.

• Assume parallelism between reference and test dissolution profiles ($\delta$: constant difference over time points between two profiles)

• Consider various variability
  • RSD% = (5.7, 4.4, 2.9, 2.4)% for test profile
  • RSD% = (14.3, 11.1, 7.1, 5.8)% for test profile
  • RSD% = (28.6, 22.2, 7.1, 5.8)% for test profile

• For each variability and $\delta$, 1000 simulated data sets were generated to assess probability of claiming equivalence.
Simulation Study – RSD%=(5.7, 4.4, 2.9, 2.4)% for test profile

True $f_2 \approx 50$

- All methods have high power to claim similarity for small $\delta$
- Bootstrapped $f_2$ and SK give probability of claiming equivalence close to 5% when $\delta=10\%$
Simulation Study – $\text{RSD\%} = (14.3, 11.1, 7.1, 5.8)\%$ for test profile

- MSD becomes relatively conservative.

![Graph showing the probability of claim equivalence against power, δ, and Type I error. The graph includes lines for $f_2$, bootstrapped $f_2$, Tsong's MSD, and SK.]
Simulation Study – $\text{RSD}\% = (28.6, 22.2, 7.1, 5.8)\%$ for test profile

True $f_2 \approx 50$

- $f_2$ assumptions are violated.
- Comparing to SK, $f_2$ bootstrap and MSD method are relatively conservative for highly variable cases.
Simulation Study – Method Comparison

- Assume equal covariance matrices and RSD% = (28.6, 22.2, 7.1, 5.8)% for test profile

<table>
<thead>
<tr>
<th>Similarity</th>
<th>Mean Diff= (28, 22, 10, 5)</th>
<th>Mean Diff= (18, 13, 8, 5)</th>
<th>Mean Diff= (12, 10, 9, 5)</th>
<th>Mean Diff= (10, 10, 3, 3)</th>
<th>Mean Diff= (5, 4, 3, 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>passing rate</td>
<td>f2=36.7</td>
<td>f2=45.8</td>
<td>f2=51.3</td>
<td>f2=56.4</td>
<td>f2=70.1</td>
</tr>
<tr>
<td>f2</td>
<td>0.001</td>
<td>0.203</td>
<td>0.586</td>
<td>0.822</td>
<td>0.982</td>
</tr>
<tr>
<td>Bootstrapped f2</td>
<td>0</td>
<td>0.014</td>
<td>0.094</td>
<td>0.257</td>
<td>0.728</td>
</tr>
<tr>
<td>MSD</td>
<td>0.005</td>
<td>0.054</td>
<td>0.073</td>
<td>0.373</td>
<td>0.623</td>
</tr>
<tr>
<td>f2&gt;=50 &amp; (Bootstrapped f2 or MSD)</td>
<td>0.001</td>
<td>0.041</td>
<td>0.137</td>
<td>0.445</td>
<td>0.804</td>
</tr>
<tr>
<td>SK</td>
<td>0.410</td>
<td>0.625</td>
<td>0.529</td>
<td>0.974</td>
<td>0.978</td>
</tr>
<tr>
<td>IUT</td>
<td>0</td>
<td>0</td>
<td>0.006</td>
<td>0.024</td>
<td>0.162</td>
</tr>
</tbody>
</table>

Good power and type I error control

Caution!
Summary/Remarks

- IUT is very conservative and has very low power to claim similarity.

- SK method has good power to detect similarity and control of type I error when the two dissolution profiles are parallel. But when the underlying assumption of parallelism fails, SK method could be too liberal with high type I error (pass similarity when dissimilar).

- Comparing to SK, f2 bootstrap and MSD method are relatively conservative for highly variable cases.

- MSD is inconsistent in its result comparing to bootstrapped f2. MSD method is likely to be less discriminating and sensitive in some scenarios (e.g. Paixão, et al. 2017 and Mangas-Sanjuan, et al. 2016). But on the other hand, MSD method can also have higher power to detect similarity in some scenarios when the two profiles are similar.

- f2 is a conservatively biased estimator. Although f2 and MSD are testing different hypotheses, comparisons may fail bootstrap and pass MSD in part because of the conservative bias of f2.
Three Methods are utilized in this practice: f2; f2 Bootstrapping; MSD (Tsong’s Method)

Scenario 1: f2 ≥ 50 & Variability met (Pass/ --- / ---)

Scenario 2: f2 < 50 (Fail/ --- / ---)

NOT met Variability requirements:

Scenario 3: (Pass/ Pass / ---) Confirmed by Bootstrapping!

Scenario 4: (Pass/ Fail / Fail) ➔ Cannot confirm similarity

Scenario 5: (Pass/ Fail / Pass) Confirmed by MSD method!
Basic Concepts for the Decision Tree

Three methods in series for analysis based on the f2 criteria

- f2 Calculation
- f2 Bootstrapping (more conservative than f2)
- Tsong’s MSD method (additional checking for borderline cases)

<table>
<thead>
<tr>
<th>Decision</th>
<th>3</th>
<th>5</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass!</td>
<td>P/P/-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>???</td>
<td>P/F/P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fail!</td>
<td></td>
<td>P/F/F</td>
<td></td>
</tr>
</tbody>
</table>

Scenario

1. Pass!
   - f2 > 50; %CV OK!

2. Fail!
   - P/F/P
   - F/P/P
   - F/P/F
   - F/F/P
   - F/F/F

- Confirmed Similarity!
- Failed to Confirm!
- May need a Second Look!
- f2 < 50; NOT trying to make it to Pass
R Shiny Web Application Tool established

Dissolution Profiles Comparison

Choose Ref lot to upload
Browse... Refxxx.csv

Choose Test lot to upload
Browse... Test.csv

Time Unit
minutes

All time points (eg, 10, 20, 30)
10, 15, 30, 45, 60, 90

Number of time points for comparison (>= 3)
4

Plot Title
Dissolution Profile Comparison

Submit

The study belongs to Scenario #3. Similarity is confirmed.

Note: If similarity is confirmed, then the recommended method for similarity assessment will be highlighted yellow with value 1 in the Recommend column.

Warning: The assumptions of Z2 method are not satisfied.

The CV for the latter time points exceeds 10%.
## Case Studies Summary

Prototype Formulations (Research Data, Total of 250 cases) – Confirmation rate is higher for lower release cases

<table>
<thead>
<tr>
<th>Categories</th>
<th>Pass!</th>
<th>Pass!</th>
<th>???</th>
<th>Fail!</th>
<th>Fail!</th>
<th>Confirmed</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;80% LA</td>
<td>11</td>
<td>66</td>
<td>16</td>
<td>9</td>
<td>28</td>
<td>130</td>
</tr>
<tr>
<td>50% ~80%</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>20% ~50%</td>
<td>0</td>
<td>30</td>
<td>3</td>
<td>1</td>
<td>6</td>
<td>40</td>
</tr>
<tr>
<td>&lt;20%</td>
<td>0</td>
<td>70</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>70</td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td><strong>12</strong></td>
<td><strong>170</strong></td>
<td><strong>21</strong></td>
<td><strong>10</strong></td>
<td><strong>37</strong></td>
<td><strong>250</strong></td>
</tr>
</tbody>
</table>

*The cases can be clearly identified as either Pass or Fail by the decision tree.*

→ More than 85% of cases can be confirmed by this decision tree practice
Bootstrapping Performance – Some Cases May Desire the MSD Analysis

![Graph showing Bootstrapping vs. f2](image)

- (<20% LA)
- (20~50% LA)
- (50~80% LA)
- (>80% LA)
EMA/810713/2017 – Q&A on Mahalanobis distance

Question and answer on the adequacy of the Mahalanobis distance to assess the comparability of drug dissolution profiles

<table>
<thead>
<tr>
<th>Draft agreed by Biostatistics Working Party</th>
<th>June 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adopted by CHMP</td>
<td>26 July 2018</td>
</tr>
</tbody>
</table>

| Keywords | Bioequivalence, dissolution profiles, f2, Mahalanobis distance, biowaiver |

“Based on these considerations, the MD metric cannot be supported as a preferred methodological approach to decide upon similar dissolution, ...............”
Example of Scenario (F/F/P) – EMA’s Concern on MD/MSD method

This case fails the similarity analysis according to our decision tree since $f_2 < 50$. MSD method may not be reliable if used alone.
Example of Scenario (P/F/P) – Very Similar Profiles Fail Bootstrapping due to high variability

Research data:
Reference and Test samples are both variable ➔ Similar! (MSD method is OK.) (May require N=12 to confirm!)
The Reference samples are very variable ➔ Data is not reliable! (May require New sample or Re-Test!)
The Reference samples are very consistent (low variability), but the Test samples showed different behavior ➔ Not Similar!
Low Release Case of Scenario (P/F/P):

The overall release is low and individual samples are not overlapping ➔ Not Similar!
Factors to Consider ➔ Obtain Reliable Data for Comparison

- Sample – Formulation Design, i.e. IR vs. ER
- Method –
  - Hydrodynamics:
    - Apparatus Types
    - RPM/DPM/Flow Rate
  - Medium pH – Physiological pH ranges

Statistical Tools (Methods) help us to understand the situation ➔ Identify the Root Cause & Fix it

Intention ➔ Improve and Assure Product Quality
(NOT just trying to Pass Similarity Analysis)
Performance Summary of Decision Tree

Attempting to let the science inform decision making.

   NOT trying to pass products which are “dissimilar”.

   Nor are we wanting to fail products which are “similar”.

The decision tree is not intended for use with every profile comparison situation. Check the science and the assumptions on the use of the statistical methods first.

If $f_2 < 50$, then no need to test further as this implies there is more than 10% difference between the means of the test and reference.

Some “similar” cases which fail bootstrap pass MSD

   $f_2$ is a conservatively biased statistic
References
