The posterior probability of in vitro dissolution equivalence

Stan Altan, Janssen R&D
Hans Coppenolle, Janssen R&D
*David.LeBlond@sbcglobal.net
Areti Manola, Janssen R&D
Linas Mockus, Purdue University
Steve Novick, MedImmune
John Peterson, GSK
Yan Shen, Janssen R&D
Jyh-Ming Shoung, Janssen R&D
Harry Yang, MedImmune
Acknowledgements

This presentation is drawn from the following joint efforts:


We are grateful to Gianluca Baio University College London, Bruno Boulanger Arlenda, and Emmanuel Lesaffre, KU Leuven for helpful suggestions.
Outline

- Frame the dissolution equivalency question
- Normative requirements for a Bayesian approach
- Univariate vs multivariate criterion
- Model dependent approach (Weibull)
- Concept of minimum posterior probability of equivalence
- Importance of inference space
- Example
  - Compare 2 processes
  - Multiple lots/process & multiple profiles/lot
  - Hierarchical nonlinear model
  - Priors
  - MCMC inference
  - Univariate & Multivariate criteria
- Discussion points
Not considered here...

- Is IV dissolution even predictive of efficacy, safety, or quality?
- Is pairwise equivalence testing the best way?
- What is the “best” criterion of equivalence?
- Frequentist vs Bayesian
Is B equivalent to A?
Normative Requirements
(Bayesian version$^1$)

1. A specified function, $\phi$, of the population parameters is used to define a fixed region of equivalence $\rho$.

2. The test of equivalence...
   - Makes an inference to the specific process under study
   - Is conditional on the justified population model (including prior knowledge), and observed data
   - Is based $solely^2$ on the posterior probability, $\hat{P}$, that $\phi \subset \rho$

3. Equivalence is accepted iff $\hat{P} \geq P_{MIN}$
   - Requires a choice of $P_{MIN}$

---


$^2$Calibration of frequentist performance of the test is useful, but is not the basis for an equivalence decision about the process under study.
Normative Requirements
(Bayesian version)

1. A specified function, $\phi$, of the population parameters is used to define a fixed region of equivalence $\rho$.

Possible candidates for $\phi$ and $\rho$ (p time points):

$$f^2 = 50 \cdot \log_{10} \left( \frac{100}{1 + \frac{1}{p} \sum_{n=1}^{p} (\bar{y}_{B,n} - \bar{y}_{A,n})^2} \right) > 50 \ ?$$

No. $f^2$ is a function of data, not population parameters

$$F^2 = 50 \cdot \log_{10} \left( \frac{100}{1 + \frac{1}{p} \sum_{n=1}^{p} (w_{B,n} - w_{A,n})^2} \right) > 50 \ ?$$

Yes. $F^2$ is a function of population parameters

$\phi = F^2$

$\rho$

$50$

$100$

1. A specified function, $\phi$, of the population parameters is used to define a fixed region of equivalence $\rho$.

Possible candidates for $\phi$ and $\rho$ ($p=3$ time points):

$$
\rho = \begin{cases} 
- R < \Delta w_1 < + R & \text{&} \\
- R < \Delta w_2 < + R & \\
- R < \Delta w_3 < + R
\end{cases}
$$

Yes. Both are functions of population parameters and constants, not dependent on measurement results when $R = 99p$, $\phi \equiv F2$
Normative Requirements
(Bayesian version)

1. A specified function, $\phi$, of the population parameters is used to define a fixed region of equivalence $\rho$.

What about “model dependent” equivalence regions?

$$w_{m,n} = \text{Weibull}(\mu, t)$$

$$= \mu_1 \left[ 1 - \exp \left( -\left( \frac{t_n}{\exp(\mu_{2,n})} \right) \right) \right]; m = A, B$$

1. Dimensional reduction
2. $A:B$ mapping is 1:1
3. Yes. Define $\phi$ & $\rho$ in terms of $\mu$ in place of $w$
4. $\mu$’s have mechanistic interpretations
Normative Requirements
(Bayesian version)

2. The test of equivalence...
   • Makes an inference to the specific process under study
   • Is conditional on the justified population model (including prior knowledge), and observed data
   • Is based solely on the posterior probability, $\hat{P}$, that $\phi \subset \mu$
   • Requires a choice of $P_{\text{MIN}}$

3. Equivalence is accepted iff $\hat{P} \geq P_{\text{MIN}}$
Defining the inference space

<table>
<thead>
<tr>
<th>Study type</th>
<th>Model of Lot(s) required?</th>
<th>Model of Process(es) required?</th>
<th>Inference space</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enumerative</td>
<td>Yes</td>
<td>No</td>
<td>Sampled Lot(s) only</td>
</tr>
<tr>
<td>Analytic</td>
<td>Yes</td>
<td>Yes</td>
<td>Lot(s) &amp; Process(es)</td>
</tr>
</tbody>
</table>

Should IV dissolution similarity studies be enumerative or analytic?

- Often a model of the process(es) is absent, yet...
- Inferences are applied to the process(es)

---

1Deming E (1975) On probability as a basis for action, The American Statistician 29(4), 146-152
Example IV dissolution study

\[ y_{m,l,u,n} \sim N\left(w_{m,l,u,n}, \sigma^2\right) \]
\[ \varepsilon_{m,l,u} \sim N_3(\mathbf{0}, \Sigma^{Lot}_m) \]
\[ w_{m,l,u,n} = \text{Weibull}\left(\mu_m + \varepsilon_{m,l,u} + \delta_{m,u}, t_n\right) \]
\[ \delta_{m,u} \sim N_3(\mathbf{0}, \Sigma^{Unit}_m) \]

**Index** | **Variable** | **Levels**
---|---|---
\( m \) | Process | A,B
\( l \) | Lot | 1,2,...,13
\( u \) | Unit | 1,2,...,156
\( n \) | \( t \) (min) | 1,2,...,7

12 tablets(units) per batch (lot)
- Circles = measured %dissolution (\( y \))
- Lines = predicted %dissolution (\( w \))

5 lots made using process B
8 lots made using process A

Batches 1-8 are from Site 1 and Batches 9-1

Batch numbers are shown above each panel.
Grand average observed dissolution

What is the probability that A and B are “Equivalent”?

Grand Mean % Dissolution

Minutes

0 10 20 30 40

0 20 40 60 80
Priors used

\[ y_{m,l,u,n} \sim N\left(w_{m,l,u,n}, \sigma^2\right) \]

\[ w_{m,l,u,n} = \text{Weibull}\left(\mu_m + \mathbf{e}_{m,l} + \mathbf{d}_{m,u}, t_n\right) \]

\[ \mathbf{e}_{m,l} \sim N_3\left(0, \Sigma^{\text{Lot}}_m\right) \]

\[ \mathbf{d}_{m,u} \sim N_3\left(0, \Sigma^{\text{Unit}}_m\right) \]

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Minimally informative(^1)</th>
<th>Prior distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\mu)'s (3/process)</td>
<td>Uniform((-\infty, +\infty)) (^2)</td>
<td>(\mu_m)</td>
</tr>
<tr>
<td>Common (\sigma)</td>
<td>Uniform((0, +\infty)) (^2)</td>
<td>(\sigma_m)</td>
</tr>
<tr>
<td>(\sigma^{\text{Lot}})'s and (\sigma^{\text{Unit}})'s (6/process)</td>
<td>Uniform((0, +\infty)) (^2)</td>
<td>(\sigma_m)</td>
</tr>
<tr>
<td>Unit correlation matrix (3 (\rho)'s/process)</td>
<td>Uniform density over all 3x3 correlation matrices</td>
<td>(\rho_{1,2}, \rho_{1,3}, \rho_{2,3})</td>
</tr>
</tbody>
</table>

\(^1\)If scientifically justified, informative priors would be preferable

\(^2\)Use of these improper priors led to posterior distributions indistinguishable from those obtained using wide proper prior distributions
MCMC inference

Calculate quantities of interest for each draw → 16,000 draw sample from joint posterior of quantities of interest

4 chains
- 4,000 draw warmup/chain
- 4,000 draw posterior sample/chain

16,000 draw sample from the joint posterior of all model parameters

Due diligence
- Convergence?
- Mixing?
- Prior impact?
- Effective sample size?
- MC std dev low?

OK

not OK

1\(^{http://mc-stan.org/users/documentation/}\)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Minimally informative(^1) Prior distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \eta ) (1 process)</td>
<td>Uniform(–( \infty ),( \infty )) (^2)</td>
</tr>
<tr>
<td>Common ( \mu )</td>
<td>Uniform(0, ( \infty )) (^2)</td>
</tr>
<tr>
<td>( \eta ) (1 process) ( \rho )</td>
<td>Uniform(0, ( \infty )) (^2)</td>
</tr>
<tr>
<td>Unit correlation matrix (3 ( y )/process)</td>
<td>Uniform density over all 3x3 correlation matrices</td>
</tr>
</tbody>
</table>
\[
F^{2[i]} = 50 \cdot \log_{10} \left( \frac{100}{1 + \frac{1}{N_{\text{max}}} \sum_{n=1}^{N_{\text{max}}} (w_{B,n}^{[i]} - w_{A,n}^{[i]})^2} \right) ; i = 1, \ldots, 16000
\]

\[
w_{m,n}^{[i]} = \text{Weibull} \left( \mu_m^{[i]}, t_n \right) ; m = A, B ; i = 1, \ldots, 16000
\]

Posterior sample: \( \mu_m^{[i]} ; i = 1, \ldots, 16000 \)
\[
\rho = \left\{ \phi = \Delta \mu : -3S_1 < \Delta \mu_1 < +3S_1 \land -3S_2 < \Delta \mu_2 < +3S_2 \land -3S_3 < \Delta \mu_3 < +3S_3 \right\}^1
\]

\[
\Pr[\Delta \mu \in \rho] \approx \frac{1}{16000} \sum_{i=1}^{16000} \prod_{p=1}^{3} I (-3S_p \leq \Delta \mu_p^{[i]} \leq +3S_p)
\]

\[
= 0.895
\]

\[
\Delta \mu_p^{[i]} = \mu_B^{[i]} - \mu_A^{[i]}, i = 1, \ldots, 16000
\]

Posterior sample: \( \mu_p^{[i]} ; i = 1, \ldots, 16000; m = A, B \)

\(^1\text{Sathe, Tsong & Shah (1996) suggest } S_p \text{ be obtained from reference process total standard deviation, ignoring estimation error} \)
\[
\rho = \left( \phi = \Delta \mu : \Delta \mu \hat{\Sigma}^{-1} \Delta \mu \leq \chi^2_{3,0.99} \right)
\]

\[
\frac{1}{16000} \sum_{i=1}^{16000} \left( \Sigma[i]_{Lot,A} + \Sigma[i]_{Unit,A} \right)
\]

\[
\Pr[\Delta \mu \in \rho] \approx \frac{1}{16000} \sum_{i=1}^{16000} I\left(\Delta \mu[i] \hat{\Sigma}^{-1} \Delta \mu[i] \leq \chi^2_{3,0.99}\right)
\]

\[
= 0.880
\]

\[
\Delta \mu[i] = \mu_B[i] - \mu_A[i] ; i = 1, \ldots, 16000
\]

Posterior sample: \( \mu_m[i] ; i = 1, \ldots, 16000 ; m = A, B \)

\(^1\)An arbitrary choice for illustration. Should be fixed based on product specific requirements and scientific knowledge.
For discussion...

• Does this Bayesian paradigm provide a useful framework for IV dissolution profile (and/or other) similarity comparisons?

• What would regulators think?

• What would sponsors think?

• How could a link to current practice (e.g., f2, MSD) be assured?

• How should the parameter function, $\phi$, and the equivalence region, $\rho$, be chosen?

• Should the equivalence region, $\rho$, be fixed across studies or be estimated from individual study data?

• If scientifically justifiable prior knowledge is available, should it be used?