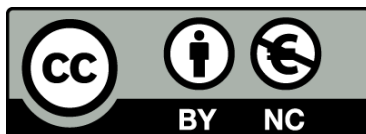




UNIVERSITAT DE
BARCELONA

Bioequivalencia promedio escalada

Joel Alejandro Muñoz Gutiérrez



Aquesta tesi doctoral està subjecta a la llicència **Reconeixement- NoComercial 3.0. Espanya de Creative Commons.**

Esta tesis doctoral está sujeta a la licencia **Reconocimiento - NoComercial 3.0. España de Creative Commons.**

This doctoral thesis is licensed under the **Creative Commons Attribution-NonCommercial 3.0. Spain License.**

Bioequivalencia promedio escalada

Joel Alejandro Muñoz Gutiérrez

Barcelona, mayo 2017



UNIVERSITAT DE
BARCELONA

Bioequivalencia promedio escalada

Memoria presentada por Joel Alejandro Muñoz Gutiérrez para optar al grado de Doctor
por la Universitat de Barcelona

Programa: Estadística.

Departamento: Genética, Microbiología y Estadística (Sección Estadística), Facultad de
Biología.

Fecha: Barcelona, 31 de mayo 2017.

Doctorando:
Joel Alejandro Muñoz Gutiérrez

Director y Tutor:
Jordi Ocaña Rebull

Agradecimientos y dedicatorias

A mis padres, hermanos y a mi esposa, por el apoyo en todo momento, en especial, a mi hija que es una de las felicidades más grande que me ha dado la vida. A mi director Jordi Ocaña por toda la dedicación, enseñanza y apoyo en mi etapa de formación e investigación para realizar esta memoria.

Índice general

Índice general	IV
Índice de figuras	v
Índice de tablas	VII
Notación	IX
1. Introducción	1
1.1. Diseños crossover	3
1.1.1. Diseño crossover RT/TR	5
1.1.2. Diseño parcialmente replicado	8
1.1.3. Diseño replicado	10
1.1.4. Método de estimación requerido por la EMA	10
1.2. Bioequivalencia promedio y BE escalada	10
1.2.1. Parámetros farmacocinéticos	11
1.2.2. Principio de inclusión de intervalos	12
1.2.3. Bioequivalencia promedio	14
1.2.4. Bioequivalencia promedio escalada(RSABE)	14
1.3. Método EMA	14
1.3.1. Intervalo de confianza para ϕ	15
1.3.2. Criterio de decisión de RSABE, EMA	16
1.4. Método FDA	16
1.4.1. Criterio de decisión de BE, FDA	17
1.5. Método de Howe	18

2. Objetivos	21
3. Métodos propuestos	23
3.1. Método HoweEMA	23
3.2. Métodos de límites continuos	24
3.3. Método LO	24
3.3.1. Criterio de decisión de RSABE, LO	26
3.4. Método Howe-LO	27
3.4.1. Corrección del sesgo en la estimación de los límites de bioequivalencia escalados	30
3.5. Ajuste del nivel de significación	33
4. Resultados y Discusión	35
4.1. Resultados	35
4.1.1. Simulaciones	35
4.1.2. Análisis de datos reales	45
4.2. Discusión	48
5. Conclusiones	51
Bibliografía	55
A. Anexos	57
A.1. Packages <code>simcrossover</code> y <code>adjustalpha</code>	57
A.1.1. Funciones <code>simCrossover</code> y <code>adjAlpha</code>	57
A.2. Artículo N°1	61

Índice de figuras

1.1. Gráfico concentración-tiempo	12
1.2. Perfil de la concentración en la sangre o plasma.	13
1.3. Límites de bioequivalencia según EMA.(2010) y la FDA.(2010)	20
3.1. Límites de BE lineales y suaves, respectivamente	25
3.2. Gráfico conjunto donde $n = 12$	31
4.1. Probabilidad de declarar BE	37
4.2. Potencia de los métodos escalados ajustados de bioequivalencia, cuando $\phi = 0$	38
4.3. Potencia de los métodos escalados ajustados de bioequivalencia, cuando $\lambda = 0.8$	39
4.4. Probabilidad de declarar BE, con límites suaves	41
4.5. Probabilidad de declarar BE, con límites suaves y corrección del sesgo	42
4.6. Probabilidad de declarar BE, con límites suaves y corrección del sesgo	43
4.7. Probabilidad de declarar BE, en los distintos métodos	44

Índice de tablas

1.1. Diseño RT/TR.	5
1.2. Diseño TRR/RTR/RRT.	8
1.3. Diseño TRTR/RTRT.	10
4.1. Pérdida de potencia para los niveles de significancia ajustados	40
4.2. Resultados de la aplicación de los enfoques RSABE para el parámetro farmacocinético $\log(C_{max})$	47
4.3. Resultados de la aplicación de los enfoques RSABE para el parámetro farmacocinético $\log(AUC)$	47

Notación

AUC	Área bajo la curva de la concentración-tiempo
C _{max}	Concentración plasmática máxima
T _{max}	Tiempo en alcanzar la concentración plasmática máxima
CV _{WR}	Coficiente de variación del tratamiento de referencia
EMA	European Medicine Agency
FDA	Food and Drug Administration
ABE	Bioequivalencia promedio
BE	Bioequivalencia
HVD	Highly Variable Drugs
RSABE	Bioequivalencia promedio escalada de referencia
IC	Intervalo de confianza
BA	Biodisponibilidad
CIIP	Principio de inclusión de intervalos (Confidence Interval Inclusion Principle)
ISC	Contrastes intrasujeto (Intra Subject Contrasts)

Capítulo 1

Introducción

Para el desarrollo de un fármaco generalmente los primeros estudios que se realizan son en animales. En ellos pueden surgir indicios que una nueva molécula puede ser efectiva y segura para enfrentar una enfermedad. La suposición de que la situación puede ser la misma en seres humanos puede iniciar el largo y costoso proceso que conocemos como “ensayo clínico”. Se entiende por ensayo clínico a toda evaluación experimental de una sustancia o medicamento, a través de su administración o aplicación a los seres humanos, el cual permite demostrar la eficacia y seguridad de la molécula. El periodo de desarrollo clínico de un producto farmacéutico se divide en 4 fases [1] :

- Fase I: Son los primeros estudios que se realizan en seres humanos con la nueva droga y el objetivo fundamental es probar la eficacia y tolerancia del compuesto.
- Fase II: Proporciona información preliminar sobre la eficacia del producto y establece la relación dosis-respuesta, son estudios terapéuticos exploratorios.
- Fase III: Evalúa la eficacia y seguridad del tratamiento experimental en las condiciones de uso habituales y con respecto a las alternativas terapéuticas disponibles para la indicación estudiada. Se trata de estudios terapéuticos de confirmación.
- Fase IV: Se realiza después de la comercialización del fármaco para estudiar condiciones de uso distintas de las autorizadas, como nuevas indicaciones, la efectividad y seguridad en la utilización clínica diaria.

Es importante señalar que no es lo mismo la molécula en estudio que la marca comercial del medicamento que la contiene, ya que la marca comercial del medicamento contiene la molécula más excipientes.

Por otra parte, cuando la patente de la molécula ha expirado, otros laboratorios pueden producir medicamentos genéricos que contengan dicha molécula y lo único

que tienen que demostrar es que la biodisponibilidad de la molécula, en el lugar de acción, es similar a la asociada al medicamento de marca, ahorrándose el largo y costoso proceso del ensayo clínico que condujo a demostrar la eficacia y seguridad de la molécula.

El concepto de biodisponibilidad es propio de la farmacocinética. (La farmacocinética es la rama de la farmacología que se ocupa de los procesos a los que un fármaco será sometido en su paso por el organismo. Estudia qué sucede con un fármaco desde el momento en el que es administrado hasta su total eliminación del cuerpo). Como medida de la cantidad de fármaco absorbido se utiliza AUC , el área bajo la curva de la concentración del fármaco en la sangre o plasma a través del tiempo, como indicador de la velocidad de absorción C_{max} la concentración máxima que se alcanza en un periodo de tiempo y T_{max} , el tiempo en alcanzar dicha concentración máxima. Estas variables o parámetros farmacocinéticos se describen con más detalle en los siguientes capítulos.

En 21 CFR 320.1[12] se define la bioequivalencia (BE) como “la ausencia de una diferencia significativa en la velocidad y la medida en que el ingrediente activo o la fracción activa de equivalentes farmacéuticos o alternativas farmacéuticas se hace disponible en el sitio de acción farmacológico cuando se administran en la misma dosis molar bajo condiciones similares en un estudio diseñado apropiadamente”.

Los estudios de bioequivalencia se utilizan para demostrar que un medicamento genérico o formulación de prueba (test)(T) tiene la misma eficacia terapéutica que el medicamento referencia o formulación de referencia (marca) (R). Así estas dos formulaciones pueden ser totalmente intercambiables [25].

Existen tests de BE, bien establecidos, aprobados por las agencias y usados de forma habitual, pero que no son suficientemente potentes para fármacos de alta variabilidad (HVD) definidos como aquellos cuyo coeficiente de variación intrasujetos para la formulación de referencia (CV_{WR}) es superior o igual a 30 %, lo que hace muy difícil aprobar BE con los tamaños muestrales habituales en ensayos de BE, se requerirían tamaños mucho mayores, inasumibles por razones éticas (exponer innecesariamente muchos sujetos sanos a la medicación) y que serían muy costosos.

La agencia europea del medicamento EMA, y la estadounidense FDA, propusieron límites de bioequivalencia BE “escalada”, para facilitar la declaración de BE de genéricos de fármacos de alta variabilidad (HVD). Ambos enfoques entran dentro de lo que se denomina bioequivalencia promedio escalada de referencia (Reference Scaled Average Bioequivalence, RSABE) Patterson and Jones(2012)[26]. Los tests de BE asociados a dichos límites escalados, en parte debido a que los límites están basados en funciones algo abruptas de la variabilidad, tienen problemas de control del error de tipo I. Algunos autores como Karalis et al.(2011)[16, 19] han propuesto límites más suaves.

Sus propiedades se han estudiado mediante estudios de simulación sobre los diseños crossover más favorecidos en las regulaciones, el diseño “parcialmente replicado”, TRR/RTR/RRT y el diseño replicado, “TRTR/RTRT”, donde se consideran

dos niveles de tratamiento, T (Test) y R (Referencia). Además se han considerado diferentes escenarios para conocer el comportamiento del error tipo I y la potencia en los contrastes de hipótesis realizados para determinar la bioequivalencia promedio escalada de referencia.

En el capítulo 1 se presenta una introducción general en donde se define un modelo lineal general para de los diseños crossover en particular se estudian los dos tipos de diseños mencionados anteriormente, parcialmente replicado y replicado. Seguidamente se presenta una sección donde se plantea el problema estadístico de demostrar BE (con los métodos habituales, TOST -Two One Sided Tests- y el principio de inclusión de intervalos). A continuación se plantea con más detalle la problemática HVD y se definen los distintos métodos para probar en este caso bioequivalencia, propuestos por las agencias del medicamento EMA y FDA, publicados en el año 2010. En el capítulo 2 se presentan los objetivos de esta tesis. En el capítulo 3 se proponen tres métodos basados en límites de aceptación que tienen en consideración alguna de las objeciones realizadas para cada criterio definidas más adelante, concretamente en EMA se propone una mejora (método que se denomina “HoweEMA”) basado en límites constantes al aplicar el principio de inclusión de intervalos y no aleatorios. Para el método de la FDA se consideran dos métodos con límites de BE continuos, “contFDA” y “contFDA2”. También se consideran métodos donde los límites de BE son suaves en todos sus puntos definiendo así el método LO y HoweLO, además se considera una corrección del sesgo en la aplicación de método HoweLO y se propone un método de ajuste del nivel de significación. En el capítulo 4 se describen unas simulaciones y sus resultados. En el capítulo 5 se exponen las conclusiones de esta memoria.

1.1. Diseños crossover

Los diseños crossover son un tipo de experimento de medidas repetidas, donde distintos tratamientos son aplicados en períodos de tiempo sucesivos a las unidades experimentales. Estos han demostrado ser eficaces en una amplia gama de aplicaciones, y en particular, en ensayos clínicos.

En el caso de ensayos clínicos de bioequivalencia las unidades experimentales son los sujetos, y si consideramos el factor formulación con dos niveles, es decir, tratamiento R y T , estos son asignados en periodos de tiempo sucesivos generando una secuencia de tratamientos, donde el número de secuencias de tratamiento es prefijado por el investigador y el proceso de aleatorización indica que se asignan aleatoriamente sujetos a las secuencias de tratamientos.

El washout son los tiempos entre cada tratamientos administrados y es importante que sea un tiempo adecuado para que el cuerpo se limpie del efecto del tratamiento aplicado en el periodo anterior.

El modelo del diseño crossover tiene en cuenta esta necesidad de limpieza y considera un factor fijo llamado carryover, este efecto carryover puede generar sesgo

en las estimaciones en el caso de que dicho efecto exista. Por esta razón es adecuado estudiar como punto de inicio en el análisis de bioequivalencia si existen indicios de este efecto carryover [22].

En esta tesis se utiliza la misma formulación general que usan Chow y Liu(2009)[4] del modelo lineal asociado a un diseño crossover:

$$Y_{ijk} = \mu + G_k + S_{i(k)} + P_j + F_{(j,k)} + C_{(j-1,k)} + e_{ijk} \quad (1.1)$$

donde $i = 1, 2, \dots, n_k$, $j = 1, \dots, J$, $k = 1, \dots, K$, i -ésimo sujeto en la k -ésima secuencia, en el período j -ésimo, n_k es el número de sujetos en la secuencia k y $N = n_1 + \dots + n_K$ es el número total de sujetos en el estudio. Además el efecto formulación se define como $\phi = F_T - F_R$ y Y_{ijk} corresponde al logaritmo de la biodisponibilidad (C_{max} en la regulación EMA, C_{max} o AUC en la regulación FDA).

- Y_{ijk} :Es el logaritmo de la respuesta farmacocinética observada (biodisponibilidad) en el i -ésimo sujeto en la secuencia k -ésima, en el j -ésimo período.
- μ :Es la media general de la variable respuesta correspondiente.
- G_k :Es el efecto fijo de la k -ésima secuencia, donde $\sum_{k=1}^K G_k = 0$
- $S_{i(k)}$:Es el efecto aleatorio del sujeto i dentro de la secuencia k -ésima.
- P_j :Es el efecto fijo del j -ésimo periodo, donde $\sum_{j=1}^J P_j = 0$.
- $F_{(j,k)}$:Es el efecto fijo del tratamiento administrado en el j -ésimo periodo de la k -ésima secuencia, donde $\sum_{j=1}^J \sum_{k=1}^K F_{(j,k)} = 0$
- $C_{(j-1,k)}$:Es el efecto fijo residual o “de acarreo” (carryover) del $j - 1$ -ésimo período de la k -ésima secuencia, este efecto ocurre a partir del segundo periodo. Además $\sum_{j=1}^J \sum_{k=1}^K C_{(j-1,k)} = 0$
- e_{ijk} :Es el error aleatorio o residuo al observar Y_{ijk} .

Además se deben cumplir los siguientes supuestos:

- $S_{i(k)} \stackrel{iid}{\sim} N(0, \sigma_S^2)$: *iid* significa que las variables aleatorias son independiente e idénticamente distribuidas, la varianza σ_S^2 representa la variabilidad intersujetos (variabilidad entre los sujetos).
- $e_{ijk} \stackrel{iid}{\sim} N(0, \sigma_t^2)$: Donde $t = 1, \dots, L$ (el número de formulaciones o tratamientos a comparar en nuestro caso $L = 2$ y $\sigma_1^2 = \sigma_{WR}^2$, $\sigma_2^2 = \sigma_{WT}^2$. Estas varianzas representan la variabilidad intrasujetos (variabilidad dentro de los sujetos). Como se puede observar, en general, no es necesario considerar homocedasticidad entre formulaciones o tratamientos en cuanto a la varianza residual.
- $\{S_{i(k)}\}$ y $\{e_{ijk}\}$: Sean mutuamente independientes.

Observación 1 (Kuehl, 2000)[18]

- i) Las pruebas crossover se diseñan para evitar confundir los efectos de periodo y tratamiento.
- ii) La principal ventaja de estos diseños es el aumento de la eficacia y el ahorro de casos.

Este tipo de diseños no considera interacción entre los factores, solo considera los efectos principales fijos, aleatorio, anidado y cruzado, evitando la presencia de confusión en el diseño factorial, es decir, se asume que las interacciones no tienen un efecto relevante.

1.1.1. Diseño crossover RT/TR

En general, las agencias reguladoras([10]) recomiendan obtener datos usando un diseño crossover (más específicamente, un diseño crossover RT/TR de 2×2 , 2 periodos y 2 secuencias.), considerando $J = K = 2$ en la (1.1) se tiene este modelo en particular.

Secuencia	Periodo	
	1	2
1	R	T
2	T	R

Tabla 1.1: Diseño RT/TR.

En un diseño crossover de 2×2 , cada sujeto experimental recibe una sola dosis de ambas formulaciones, R y T en solo uno de dos posibles secuencias de tratamiento

RT o TR, donde una muestra de $N = n_1 + n_2$ sujetos son aleatorizados de manera que n_1 corresponden a la secuencia RT y n_2 a la secuencia TR. En general todas las agencias del medicamento recomiendan realizar el análisis para una variable Y dada en escala logarítmica, es decir, $Y = \ln(C_{max})$ o $Y = \ln(AUC_{0-t})$ donde Y_{ijk} es designada como la observación del i -ésimo sujeto, en el j -ésimo periodo en la k -ésima secuencia $i = 1, \dots, n_k$; $k = j = 1, 2$.

Se considera el modelo definido en (1.1) como se mencionó anteriormente, donde μ es la media global, P_j el efecto fijo de la administración en el periodo j , $F(j, k)$ es el efecto fijo de la formulación administrada en la k -ésima secuencia y j -ésimo periodo y $C_{(j-1, k)}$ corresponde al efecto fijo carryover; el posible efecto carryover de la formulación de referencia desde el primer periodo a la del segundo periodo en la secuencia 1, se denota por C_R y el efecto equivalente de la formulación T en la secuencia 2 es definido por C_T , es decir:

$$C_{(j-1, k)} = \begin{cases} C_R & \text{si } j = 2 \text{ y } k = 1 \\ C_T & \text{si } j = 2 \text{ y } k = 2 \\ 0 & \text{e.o.c} \end{cases} \quad (1.2)$$

donde $C_R = -C_T = C$ ya que para los efectos fijos asumiremos que suman cero, se cumple que $C_R + C_T = 0$.

Similarmente,

$$F_{(j, k)} = \begin{cases} F_R & \text{si } j = k \\ F_T & \text{si } j \neq k \end{cases} \quad (1.3)$$

con $F_R = -F_T = F$ y $P_1 = -P_2 = P$ dado que se considera $F_R + F_T = 0$ y $P_1 + P_2 = 0$. Luego se designa el efecto formulación $\phi = F_T - F_R = -2F$, el efecto carryover como $\kappa = C_T - C_R = -2C$ y el efecto periodo $\pi = P_2 - P_1$.

$S_{i(k)} \stackrel{iid}{\sim} N(0, \sigma_S^2)$ representa el efecto aleatorio del i -ésimo sujeto anidado en la k -ésima secuencia. σ_S^2 es la varianza inter-sujeto y $e_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$ por simplicidad se asume varianza residual constante, σ^2 , y además $\{S_{i(k)}\}$ y $\{e_{ijk}\}$ son mutuamente independientes.

La inferencia del efecto formulación se basa en el contraste de diferencia de cada sujeto i dentro de cada secuencia k :

$$d_{ik} = 0.5(Y_{i2k} - Y_{i1k}).$$

Las diferencias entre tratamientos dentro de cada sujeto definen el concepto de “contraste intrasujeto” (intra subject contrast, ISC). En esta sección (1.1) describiremos con cierto detalle el enfoque ISC de estimación, para distintos

diseños crossover. Como veremos más adelante, no es el único enfoque posible, y cuál de ellos escoger depende en gran parte de lo requerido por las distintas agencias del medicamento.

Volviendo al diseño 2x2 RT/TR, la esperanza y la varianza de los contrastes intrasujeto d_{ik} es:

$$E(d_{ik}) = \begin{cases} \frac{1}{2}(\pi + \phi + C_R) & \text{si } k = 1 \\ \frac{1}{2}(\pi - \phi + C_R) & \text{si } k = 2 \end{cases} \quad (1.4)$$

$$Var(d_{ik}) = \frac{1}{2}\sigma^2. \quad (1.5)$$

Si $\bar{d}_{.k} = \frac{\sum_{i=1}^{n_k} d_{ik}}{n_k}$ son las medias muestrales de las diferencias de periodos, la diferencia:

$$\hat{\phi} = \bar{D} = \bar{d}_{.1} - \bar{d}_{.2} = \bar{Y}_T - \bar{Y}_R \quad (1.6)$$

donde $\bar{Y}_T = \frac{1}{2}(\bar{Y}_{.21} + \bar{Y}_{.12})$, $\bar{Y}_R = \frac{1}{2}(\bar{Y}_{.11} + \bar{Y}_{.22})$, es un estimador insesgado del efecto formulación ϕ siempre que no este presente el efecto carryover, es decir, $\kappa = 0$ ya que en general \bar{D} es un estimador sesgado de ϕ , es decir,

$$E(\bar{D}) = \phi - \frac{1}{2}(C_T - C_R) = \phi - \frac{1}{2}\kappa.$$

La varianza del contraste de la semidiferencia d_{ik} puede ser estimado como:

$$\hat{\sigma}_d^2 = \frac{1}{N-2} \sum_{k=1}^2 \sum_{i=1}^{n_k} (d_{ik} - \bar{d}_{.k})^2 = \frac{1}{2}\hat{\sigma}^2$$

y el error estándar de \bar{D} puede ser estimado independiente por

$$\hat{s}e_{\hat{\phi}} = \hat{s}e_{\bar{D}} = \hat{\sigma}_d \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \hat{\sigma} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}. \quad (1.7)$$

Este diseño tiene el problema de que no es posible estimar la varianza intrasujeto de la formulación de referencia, σ_{WR}^2 , por separado, ya que cada secuencia de tratamientos no tiene réplicas del tratamiento R o réplicas del tratamiento T , la cual es necesaria para estimar la varianza intra-sujeto de la formulación de referencia o test σ_{WR}^2 o σ_{WT}^2 dependiendo del diseño de orden superior que se estudie, por ejemplo el diseño parcialmente replicado en estudio y recomendado por las agencias reguladoras, la estimación de este parámetro esta dada por (1.10), el cual es necesario para el caso de HVD.

Observación 2 *Notar que el tema del carryover en los estudios de bioequivalencia basados en diseños crossover, la presencia de carryover es muy problemática [2]. Es importante mencionar que en esta memoria, se asume un modelo lineal donde el carryover es nulo.*

1.1.2. Diseño parcialmente replicado

Particularizando el modelo general en este caso se tiene que $K = 3$ y $J = 3$ en (1.1). Esto define un diseño crossover de 3×3 .

Secuencia	Periodo		
	1	2	3
1	T	R	R
2	R	T	R
3	R	R	T

Tabla 1.2: Diseño TRR/RTR/RRT.

El estimador del efecto formulación $\hat{\phi}$ obtenido a partir del método de los momentos, es dado por [5]:

$$\begin{aligned}
 \hat{\phi} &= \frac{1}{6}[(\bar{Y}_{.11} - \bar{Y}_{.21}) + (\bar{Y}_{.11} - \bar{Y}_{.31}) + (\bar{Y}_{.22} - \bar{Y}_{.12}) + (\bar{Y}_{.22} - \bar{Y}_{.32}) + (\bar{Y}_{.33} - \bar{Y}_{.13}) + (\bar{Y}_{.33} - \bar{Y}_{.23})] \\
 &= \frac{1}{6}[(2\bar{Y}_{.11} + 2\bar{Y}_{.22} + 2\bar{Y}_{.33}) - (\bar{Y}_{.12} + \bar{Y}_{.13} + \bar{Y}_{.21} + \bar{Y}_{.23} + \bar{Y}_{.31} + \bar{Y}_{.32})] \\
 &= \frac{1}{3}(\bar{Y}_{.11} + \bar{Y}_{.22} + \bar{Y}_{.33}) - \frac{1}{6}(\bar{Y}_{.12} + \bar{Y}_{.13} + \bar{Y}_{.21} + \bar{Y}_{.23} + \bar{Y}_{.31} + \bar{Y}_{.32})
 \end{aligned}$$

y la varianza del estimador es:

$$\begin{aligned}
 Var(\hat{\phi}) &= Var\left(\frac{1}{6}[(2\bar{Y}_{.11} + 2\bar{Y}_{.22} + 2\bar{Y}_{.33}) - (\bar{Y}_{.12} + \bar{Y}_{.13} + \bar{Y}_{.21} + \bar{Y}_{.23} + \bar{Y}_{.31} + \bar{Y}_{.32})]\right) \\
 &= \frac{1}{6^2}[Var(2\bar{Y}_{.11} - \bar{Y}_{.12} - \bar{Y}_{.13}) + Var(2\bar{Y}_{.22} - \bar{Y}_{.21} - \bar{Y}_{.23}) + Var(2\bar{Y}_{.33} - \bar{Y}_{.31} - \bar{Y}_{.32})] \\
 &= \frac{\sigma_e^2}{6} \left(\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3} \right). \tag{1.8}
 \end{aligned}$$

En el resultado anterior se debe tener en cuenta lo siguiente:

$$\begin{aligned}
 Var\left(\frac{1}{6}[Var(2\bar{Y}_{.11} - \bar{Y}_{.12} - \bar{Y}_{.13})]\right) &= \frac{1}{6^2} (4Var(\bar{Y}_{.11}) + Var(\bar{Y}_{.12}) + Var(\bar{Y}_{.13})) \\
 &= \frac{1}{6^2} \left(4\frac{\sigma_e^2}{n_1} + \frac{\sigma_e^2}{n_1} + \frac{\sigma_e^2}{n_1} \right) \\
 &= \frac{1}{6} \frac{\sigma_e^2}{n_1}
 \end{aligned}$$

por lo cual,

$$\text{Var}(\bar{Y}_{.11}) = \text{Var}\left(\frac{\sum_{i=1}^{n_1} Y_{i11}}{n_1}\right) = \frac{1}{n_1^2} \sum_{i=1}^{n_1} \text{Var}(Y_{i11}) = \frac{1}{n_1^2} n_1 \sigma_e^2 = \frac{\sigma_e^2}{n_1}.$$

y así el error estándar de $\hat{\phi}$ puede ser estimado:

$$\widehat{se}_{\hat{\phi}} = \sqrt{\widehat{\text{Var}}(\hat{\phi})}. \quad (1.9)$$

Observación 3 *Notar que en este caso se está asumiendo homocedasticidad de varianzas intrasujetos, es decir, $\sigma_{WR}^2 = \sigma_{WT}^2 = \sigma_e^2$ en la fórmula (1.8), cuya estimación se define en (1.10).*

Por otra parte, la estimación de la varianza intrasujetos de la formulación de referencia es extraída del documento sobre progesterona de la FDA [11] para RSABE, es dada por:

$$\hat{\sigma}_{WR}^2 = \frac{\sum_{k=1}^K \sum_{i=1}^{n_k} (d_{ik} - \bar{d}_{.k})^2}{2(N - K)} \quad (1.10)$$

en el cual se define lo siguiente:

- $d_{ik} = y_{Rik1} - y_{Rik2}$ (donde 1 y 2 representan las réplicas del tratamiento de referencia)

- $\bar{d}_{.k} = \frac{\sum_{i=1}^{n_k} d_{ik}}{n_k}$

- $N = \sum_{k=1}^K n_k$

1.1.3. Diseño replicado

Particularizando el modelo general en este caso se tiene que $K = 2$ y $J = 4$ en (1.1). Luego se define un diseño crossover de 4×2 cuatro periodos y dos secuencias.

Secuencia	Periodo			
	1	2	3	4
1	T	R	T	R
2	R	T	R	T

Tabla 1.3: Diseño TRTR/RTRT.

Observación 4 *De manera similar se pueden obtener los parámetros estimados de interés para este diseño replicado y en este caso se pueden estimar por separado los parámetros σ_{WR}^2 y σ_{WT}^2 .*

1.1.4. Método de estimación requerido por la EMA

Tal como se indicó, en los apartados anteriores se describió la estimación de los principales parámetros del modelo (1.1) en términos del enfoque basado en los contrastes intrasujeto (ISC).

Los métodos de estimación basados en los contrastes intrasujeto ISC son admisibles desde el punto de vista de las regulaciones FDA.

Pero la EMA, en un documento de Questions & Answers de 2015 [9], aclaró que las estimaciones se deben basar en ajustes ANOVA, es decir, las estimaciones se deben basar en los cuadrados medios obtenidos en la tabla ANOVA del diseño.

De manera que los métodos de estimación que se comentan en este capítulo serían los aplicables a FDA. En el caso de RT/TR no hay problema, ya que ambos enfoques coinciden, pero para diseños de orden superior puede haber diferencias. También en Q&A(2015) se admite, con algunos reparos, la utilización del semireplicado pero con sólo dos secuencias: TRR/RTR.

1.2. Bioequivalencia promedio y BE escalada

A continuación se definen los conceptos necesarios para definir los métodos de bioequivalenci promedio (ABE) y RSABE, tales como los parámetros farmacocinéticos y el principio de inclusión de intervalos.

1.2.1. Parámetros farmacocinéticos

Los parámetros farmacocinéticos en la evaluación estadística de bioequivalencia promedio para la formulación de una sola dosis son: AUC , C_{max} y T_{max} . Éstos son calculados a partir de las curvas de concentración-tiempo, es decir, gráficas de los pares ordenados $(t_0, c_0), (t_1, c_1), \dots, (t_k, c_k)$. En la figura (1.1), se ilustra un ejemplo donde cada uno de los 11 sujetos en estudio presenta su curva de concentración-tiempo, es decir, para cada tiempo del estudio se determinó la concentración de la cantidad de fármaco en la sangre o plasma. A partir de estas curvas se obtienen los parámetros farmacocinéticos, según las siguientes expresiones, (Peace y Chen 2010)[7]:

1. AUC : el área bajo la curva concentración-tiempo. Existen dos variantes de dicho parámetro farmacocinético, el área bajo la curva hasta el último instante observado en la curva de concentración-tiempo, que se denota como AUC_{0-k} o más comúnmente AUC_{0-t} si $t = t_k$ que se calcula mediante la regla del trapecio como:

$$\sum_{i=1}^k \frac{(c_i + c_{i-1}) \times (t_i - t_{i-1})}{2}$$

y el área bajo la curva hasta infinito, $AUC_{0-\infty}$:

$$AUC_{0-k} + \frac{c_k}{\beta_t}$$

donde β_t es la tasa de eliminación terminal.

2. C_{max} : la máxima concentración definida:

$$C_{max} = \max\{c_0, c_1, \dots, c_k\}$$

3. T_{max} : el tiempo requerido para alcanzar la máxima concentración.

Como medida de la cantidad de fármaco absorbido se utiliza AUC y como indicador de la velocidad de absorción C_{max} y T_{max} , que se ilustran en la figura 1.2.

Observación 5 *En esta memoria se estudia principalmente C_{max} y AUC_{0-k} , ya que las metodologías presentes son aplicables a estas medidas farmacocinéticas y son los exigidos habitualmente por las agencias reguladoras.*

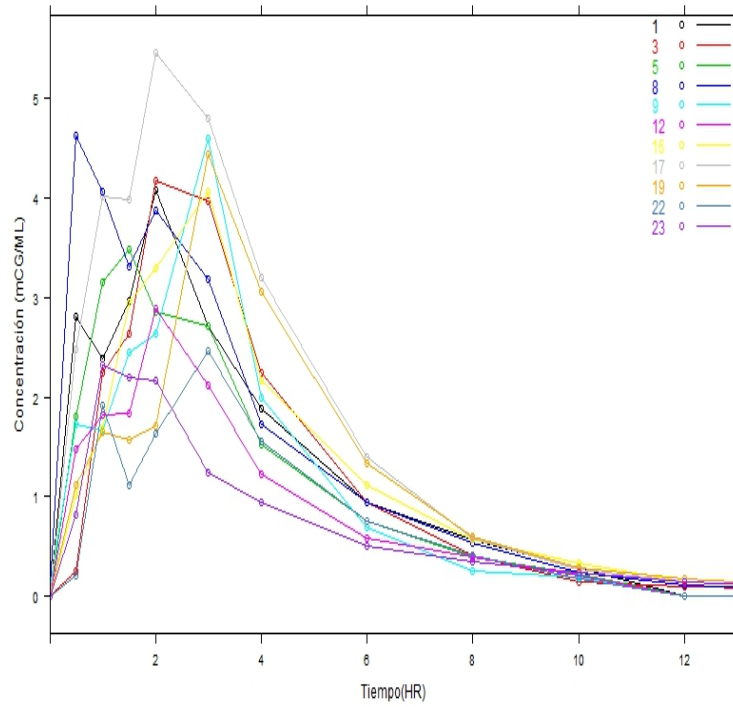


Figura 1.1: Gráfico concentración-tiempo para la primera secuencia y segundo periodo.

1.2.2. Principio de inclusión de intervalos

Uno de los métodos más frecuentemente utilizados para la evaluación de pruebas de equivalencia[3] (1.11) se basa en el principio de inclusión de intervalos de confianza. Sea ϕ el efecto formulación, para demostrar bioequivalencia se deben contrastar las siguientes hipótesis:

$$\begin{aligned} H_0 : \phi &\leq -\epsilon_1 \quad \vee \quad \phi \geq \epsilon_2 \\ H_1 : &-\epsilon_1 < \phi < \epsilon_2 \end{aligned} \quad (1.11)$$

Para el parámetro de interés ϕ se procede a calcular la estimación de su intervalo de confianza ya sea en forma exacta o aproximada, así por ejemplo: si el IC de nivel de confianza $(1 - \alpha)100\%$ es (L, U) entonces basándose en el principio de inclusión intervalos se rechaza H_0 si

$$(L, U) \subset (-\epsilon_1, \epsilon_2)$$

Observación 6 (Munk y Pfluger 1999)[24] La propiedad general que asegura este principio es que si el intervalo tiene nivel de confianza mayor o igual a $1 - \alpha$, el procedimiento anterior asegura un test de nivel de significación menor o igual a α .

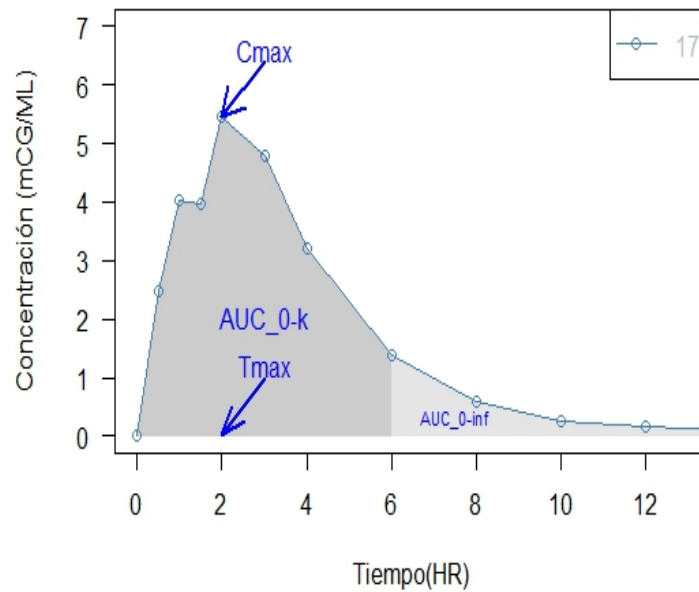


Figura 1.2: Perfil de la concentración en la sangre o plasma.

Para el intervalo de confianza paramétrico normal habitual, simétrico alrededor de la estimación puntual de ϕ , basado en la distribución t de Student, si se utiliza un nivel de confianza $1 - 2\alpha$ se obtiene exactamente un nivel de significación α .

La propiedad anterior se debe a la exigencia de una propiedad de equivariancia (esencialmente, la simetría del IC), para intervalos que no necesariamente deben cumplir lo anterior, puede ocurrir perfectamente que un intervalo $1 - \alpha$ tenga asociado un test α .

Observación 7 *Con este método de evaluación de una prueba de hipótesis, basta calcular el IC_ϕ y comprobar que esté totalmente incluido en $(-\epsilon_1, \epsilon_2)$ donde estos límites son constantes, así lo más importante es calcular adecuadamente el IC_ϕ .*

Observación 8 *La agencia europea del medicamento, EMA, y la estadounidense FDA, propusieron límites de bioequivalencia (BE) “escalada”, función de la variabilidad, para facilitar la declaración de BE de genéricos de fármacos de alta variabilidad que se describen en las siguientes secciones.*

1.2.3. Bioequivalencia promedio

Un procedimiento para demostrar bioequivalencia promedio (ABE) se expresa en términos de una prueba de equivalencia para el efecto formulación ϕ .

$$\begin{aligned} H_0 : \phi &\leq -\varepsilon_1 \text{ o } \phi \geq \varepsilon_2 \\ H_1 : -\varepsilon_1 &< \phi < \varepsilon_2. \end{aligned} \quad (1.12)$$

donde $-\varepsilon_1 = \varepsilon_2 = \varepsilon = 0.223$, así $\pm\varepsilon = \pm 0.223$.

Para realizar este contraste, se utiliza la propuesta de Schuirmann (1987)[27] que basa en la realización de dos tests unilaterales y el principio de unión-intersección a un nivel α (TOST: Two One Sided Tests):

$$\begin{aligned} H_{01} : \phi &\leq -\varepsilon \quad vs \quad H_{11} : \phi > -\varepsilon \\ H_{02} : \phi &\geq \varepsilon \quad vs \quad H_{12} : \phi < \varepsilon \end{aligned}$$

El procedimiento basado en rechazar la hipótesis nula en (1.12) si se rechazan a un nivel α ambas hipótesis nulas en los contrastes unilaterales anteriores, y por lo tanto declarar bioequivalencia, define un test de nivel α para (1.12).

Otro procedimiento equivalente es aplicar el principio de inclusión de intervalos, es decir, si el intervalo de confianza de nivel $1 - 2\alpha$ está incluido dentro de los límites de equivalencia $(-0.223; 0.223)$ entonces rechazamos la hipótesis nula de bioinequivalencia.

El mantenimiento de las hipótesis estándar (1.12) en un contexto de bioequivalencia escalada, discutida en la sección siguiente conduce a enfoques que tienen poco control del error tipo I [23].

1.2.4. Bioequivalencia promedio escalada(RSABE)

La determinación de la BE para un fármaco de alta variabilidad (HVD) fue una cuestión ampliamente debatida en los años finales del siglo XX y la primera década del siglo XXI (p.e. Shah et al. 1996, Tothfalusi et al, 2001, Tothfalusi et al. 2003). A principios de la presente década, tanto la EMA como la FDA presentaron regulaciones, que se basan en criterios escalados, sobre HVD [21, 28, 13, 16, 17].

1.3. Método EMA (Sólo para C_{max})

Sea ϕ el efecto formulación para demostrar bioequivalencia escalada se ponen a prueba las siguientes hipótesis [8]:

$$\begin{aligned} H_0 : \phi &\leq -\phi_{EMA}(\sigma_{WR}) \quad \vee \quad \phi \geq \phi_{EMA}(\sigma_{WR}) \\ H_1 : -\phi_{EMA}(\sigma_{WR}) &< \phi < \phi_{EMA}(\sigma_{WR}) \end{aligned} \quad (1.13)$$

en la cual el límite de BE superior según el criterio EMA, para el parámetro farmacocinético C_{max} se define:

$$\phi_{EMA}(\sigma_{WR}) = \begin{cases} 0.223 & \text{if } \sigma_{WR} < \sigma_{CV=30} \\ k_{EMA}\sigma_{WR} & \text{if } \sigma_{CV=30} \leq \sigma_{WR} < \sigma_{CV=50} \\ 0.359 & \text{if } \sigma_{WR} \leq \sigma_{CV=50} \end{cases}$$

y además las constantes k_{EMA} , $\sigma_{CV=30}$ y $\sigma_{CV=50}$ son típicamente redondeadas a $k = 0.76$, $\sigma_{CV=30} = 0.294$ y $\sigma_{CV=50} = 0.472$. Estos valores provienen de la relación

$$CV = 100\sqrt{(e^{\sigma_{WR}^2} - 1)} \text{ entonces } \sigma_{WR} = \sqrt{\ln\left(\left(\frac{CV_{WR}}{100}\right)^2 + 1\right)}$$

y $k_{EMA} = \frac{\ln(1.25)}{\sigma_{CV=30}} = 0.76$.

1.3.1. Intervalo de confianza para ϕ

El intervalo de confianza para el efecto formulación ϕ , cuyo nivel de confianza es $(1 - 2\alpha)100\%$ esta dado por:

$$IC_{\phi} = \hat{\phi} \pm t_{1-\alpha, \nu} \sqrt{Var(\hat{\phi})} \quad (1.14)$$

donde ν son los grados de libertad que varía de acuerdo al diseño crossover en estudio y además dependen del método inferencial que se está aplicando.

De acuerdo a las indicaciones del documento EMA (2015) sobre Q&A [9] los grados de libertad se obtienen a través de la tabla ANOVA asociada al modelo lineal (1.1) pero considerando sólo los efectos tratamiento (o formulación en el caso concreto de los estudios de BE), periodo, secuencia y sujeto jerárquicamente dentro de secuencia, es decir, habiendo eliminado el efecto carryover, que se asume nulo en un experimento bien realizado [9]. Por lo tanto, en el caso del diseño semireplicado en estudio serían $\nu = 2N - 3$ siendo N el número total de sujetos. Estos sólo afectaría al método EMA.

Cabe mencionar que para los otros métodos mencionados en esta memoria el método de estimación que se utiliza es el ISC. De esta manera, sus grados de libertad serían $\nu = N - 2$ para el diseño semireplicado.

Observación 9 *Un IC_{ϕ} al 90% implica un nivel de significancia del 5%.*

1.3.2. Criterio de decisión de RSABE, EMA

1. Obtener la estimación del efecto formulación, $\hat{\phi}$.
2. (Restricción sobre la estimación puntual). Si $\hat{\phi}$ esta fuera de ± 0.223 no se declara BE y el proceso termina.
3. Obtener la estimación de σ_{WR} , $\hat{\sigma}_{WR}$.
4. Calcular el intervalo de confianza del 90 % para el efecto formulación ϕ , es decir, IC_{ϕ} , definido en (1.14).
5. Si IC_{ϕ} al 90 % está completamente incluido en los límites $\pm \phi_{EMA}(\hat{\sigma}_{WR})$, se declara BE (rechazar H_0), en otro caso, no se declara BE.

Observación 10 *El IC_{ϕ} es bilateral y se obtiene en forma exacta.*

Observación 11 (*Posibles Objeciones, Método EMA*)

- *Este criterio no está bien fundamentado teóricamente ya que emplea el principio de inclusión de intervalos pero con límites de BE aleatorios (función de la estimación de σ_{WR}) y no constantes.*
- *Función que define el límite RSABE poco regular, no derivable en dos puntos.*
- *Criterio bietápico:*
 - *Estimar σ_{WR} , $\hat{\sigma}_{WR}$.*
 - *Si $\hat{\sigma}_{WR} < \sigma_{CV=30\%}$ se utilizan límites constantes.*
Si $\hat{\sigma}_{WR} \geq \sigma_{CV=30\%}$ se utilizan límites escalados.

1.4. Método FDA (C_{max} y AUC)

El criterio escalado FDA es más riguroso estadísticamente que el criterio escalado de la EMA (aunque basado en unos muy inadecuados límites de bioequivalencia que presentan una discontinuidad) y como se mencionó anteriormente, ambos métodos son aplicables en situaciones de fármacos de alta variabilidad, es decir, cuando el $CV_{WR} \geq 30\%$. El **documento sobre progesterona**[11] es lo que se toma como directriz de la FDA para RSABE, que se describe a continuación. Sea ϕ el efecto formulación para demostrar bioequivalencia escalada se ponen a prueba las siguientes hipótesis:

$$\begin{aligned} H_0 &: \phi \leq -\phi_{FDA}(\sigma_{WR}) \quad \vee \quad \phi \geq \phi_{FDA}(\sigma_{WR}) \\ H_1 &: -\phi_{FDA}(\sigma_{WR}) < \phi < \phi_{FDA}(\sigma_{WR}) \end{aligned} \tag{1.15}$$

en la cual el límite de BE superior según el criterio FDA, para el parámetro farmacocinético C_{max} y AUC se define:

$$\phi_{FDA}(\sigma_{WR}) = \begin{cases} 0.223 & \text{if } \sigma_{WR} < \sigma_{CV=30} \\ k_{FDA}\sigma_{WR} & \text{if } \sigma_{WR} \geq \sigma_{CV=30} \end{cases} \quad (1.16)$$

donde $k_{FDA} = \frac{\ln(1.25)}{0.25} = 0.892$, notar que $k_{FDA} > k_{EMA}$. Además en $\sigma_{CV=30} = 0.2935604$, ϕ_{FDA} presenta una discontinuidad.

1.4.1. Criterio de decisión de BE, FDA

La parte escalada de las hipótesis definidas en (1.15) y (1.16) pueden ser reexpresada:

$$\begin{aligned} H_0 : \frac{\phi}{\sigma_{WR}} &\leq -k_{FDA} \quad \vee \quad \frac{\phi}{\sigma_{WR}} \geq k_{FDA} \\ H_1 : -k_{FDA} &< \frac{\phi}{\sigma_{WR}} < k_{FDA} \end{aligned} \quad (1.17)$$

De esta forma, las hipótesis se redefinen a un nuevo parámetro de interés $\frac{\phi}{\sigma_{WR}}$ con límites fijos $\pm k_{FDA}$. Posteriormente, se aplica el principio de inclusión de intervalos para este parámetro. En concreto, el procedimiento establecido por la FDA se basa en expresar la hipótesis alternativa en (1.17) como: $\frac{\phi^2}{\sigma_{WR}^2} < k_{FDA}^2$ ó $\phi^2 - k_{FDA}^2 \sigma_{WR}^2 < 0$, de manera que el problema es reformulado en términos del nuevo parámetro $\theta = \phi^2 - k_{FDA}^2 \sigma_{WR}^2$ con $k_{FDA}^2 = 0.796$.

$$\begin{aligned} H_0 : \theta &\geq 0 \\ H_1 : \theta &< 0 \end{aligned} \quad (1.18)$$

Luego para declarar BE, la hipótesis nula en (1.18) debe ser rechazada. Según el principio de inclusión de intervalos, esto ocurre cuando el límite superior de un intervalo de confianza unilateral para θ es negativo.

Concretando, el criterio de decisión de BE puede ser resumido en los siguientes pasos:

1. Obtener la estimación del efecto formulación, $\hat{\phi}$.
2. Si $\hat{\phi}$ esta fuera de ± 0.223 no se declara BE y el proceso termina.
3. Obtener la estimación de σ_{WR} , $\hat{\sigma}_{WR}$.
4. Si $\hat{\sigma}_{WR} < \sigma_{CV=30}$ entonces,

- a) Obtener el intervalo de confianza del 90 % para el efecto formulación ϕ , es decir, IC_ϕ , definido en (1.14).
- b) Si IC_ϕ está completamente incluido en los límites ± 0.223 , se declara BE, en otro caso, concluimos que la BE no puede ser declarada y se termina el procedimiento.

5. Otro caso

- a) Obtener el límite superior del intervalo de confianza unilateral del 95 % para el parámetro $\theta = \phi^2 - k_{FDA}^2 \sigma_{WR}^2$ ($IC_\theta = (-\infty, \hat{\theta}_U]$).
- b) Si el límite superior $\hat{\theta}_U < 0$, se declara BE, en otro caso, concluimos que la BE no puede ser declarada.

El límite superior $\hat{\theta}_U$ es calculado usando el método de Howe, que se describe en [14, 15] y adaptado a situaciones previas [29]. El algoritmo es descrito en detalle en [28], el cual es presentado y aplicado en la siguiente sección.

Observación 12 *El IC_ϕ es bilateral y se obtiene en forma exacta mediante el estadístico t -Student, pero el IC_θ es unilateral y aproximadamente válido (Howe, 1974).*

1.5. Método de Howe para construir $IC_\theta = (-\infty, \hat{\theta}_U]$

Dicho intervalo de confianza se obtiene mediante el método de Howe, ya que el parámetro de interés $\theta = \phi^2 - k_{FDA}^2 \sigma_{WR}^2$ es la suma de los parámetros ϕ^2 y $-k_{FDA}^2 \sigma_{WR}^2$, y además, bajo el enfoque paramétrico normal que se analiza en esta memoria, se dispone de estimadores puntuales e intervalos de confianza unilaterales (exactos o aproximados) $1 - \alpha$ $(-\infty, C_m]$ para ϕ^2 y $[C_s, +\infty)$ para $-k_{FDA}^2 \sigma_{WR}^2$.

Cabe mencionar, que el método de Howe, se aplica en la construcción de un intervalo de confianza unilateral de las sumas de dos variables aleatorias cuyo procedimiento lo realiza en forma aproximada. En esta aplicación del método de Howe, las variables aleatorias son los estimadores puntuales.

Por lo mencionado anteriormente, se tiene que el extremo superior de un intervalo de confianza unilateral izquierdo exacto de nivel $1 - \alpha$ para ϕ^2 es:

$$C_m = (|\hat{\phi}| + t_{N-K}^{1-\alpha} \widehat{SE}_{\hat{\phi}})^2$$

donde $t_{N-K}^{1-\alpha}$ corresponde al cuantil $1 - \alpha$ de la distribución t de student con $N - K$ grados de libertad. Siendo K el número de secuencias del diseño crossover y el número total de sujetos, $N = Kn$ en un diseño balanceado, $\widehat{SE}_{\hat{\phi}}$ corresponde a la estimación del error estándar de ϕ .

Por otra parte, para $-k_{FDA}^2\sigma_{WR}^2$ el extremo inferior de un intervalo de confianza unilateral derecho $[C_s, +\infty)$ exacto de nivel $1 - \alpha$ viene dado por:

$$C_s = \frac{-k_{FDA}^2(N - K)\hat{\sigma}_{WR}^2}{\chi^2(1 - \alpha, N - K)}.$$

Luego se definen:

- $E_m = \hat{\phi}^2$
- $E_s = -k_{FDA}^2\hat{\sigma}_{WR}^2$
- $L_m = (C_m - E_m)^2$
- $L_s = (C_s - E_s)^2$

donde E_m y E_s son los estadísticos que corresponden a los estimadores puntuales de ϕ^2 y $-k_{FDA}^2\sigma_{WR}^2$, respectivamente. Mientras que los indicados con una C corresponden a extremos de intervalos de confianza unilaterales para dichos parámetros, y los indicados con una L representan una longitud al cuadrado de cada límite del intervalos unilateral y su estimación puntual.

Así, el extremo superior de un intervalo de confianza unilateral izquierdo de nivel $1 - \alpha$ para el parámetro θ será:

$$\hat{\theta}_U = (E_m + E_s) + \sqrt{L_m + L_s}.$$

Criterio de decisión:

$\hat{\theta}_U$ extremo superior del intervalo de confianza unilateral $(-\infty, \hat{\theta}_U]$ de nivel $1 - \alpha$. Rechazamos H_0 si $\hat{\theta}_U < 0 \Rightarrow$ que el test es de nivel α (teóricamente).

Observación 13 *La figura 1.3 muestra los límites escalados EMA y FDA en escala logarítmica, pero en función del coeficiente de variación en la escala original para mayor claridad (ya que posiblemente 30% es más fácil de recordar que 0.294).*

Observación 14 *(Posibles Objeciones, Método FDA)*

- *Función que define el límite RSABE poco regular, no derivable en un punto y además es discontinuo en ese punto.*
- *Criterio bietápico:*

- Estimar σ_{WR} , $\hat{\sigma}_{WR}$.
 - Si $\hat{\sigma}_{WR} < \sigma_{CV=30\%}$ se utilizan límites constantes.
Si $\hat{\sigma}_{WR} \geq \sigma_{CV=30\%}$ se utilizan límites escalados.
- A favor: criterio de inclusión de intervalos aplicado correctamente inferencia sobre θ aplicada sobre límites constantes.

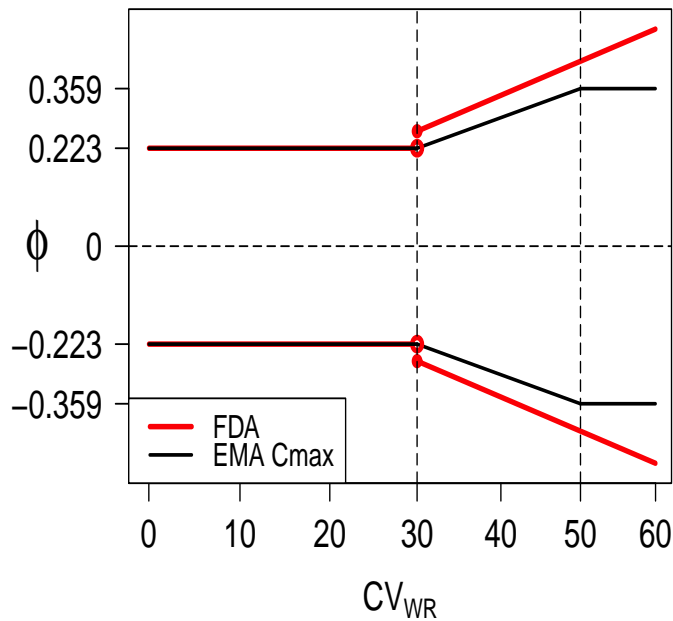


Figura 1.3: Límites de bioequivalencia según EMA.(2010) y la FDA.(2010), escalado en función de la variabilidad dentro del sujeto de la formulación de referencia R , σ_{WR} . Esta formulación se refiere a $Cmax$ y AUC en el reglamento de la FDA, y sólo a $Cmax$ en el reglamento EMA, siempre considerado en escala logarítmica. En ambas regulaciones el escalamiento comienza en $\sigma_{WR} = 0.294$ ($CV_{WR} = 30\%$) y, sólo en la regulación EMA, finaliza en $\sigma_{WR} = 0.472$ ($CV_{WR} = 50\%$).

Capítulo 2

Objetivos

- Estudiar las propiedades inferenciales de distintos métodos escalados para HVD, tanto para la FDA como para EMA. Con el objetivo de comparar dichos métodos y establecer mejoras en los métodos establecidos.
- Reformular mediante límites más regulares matemáticamente los métodos mencionados anteriormente, aportando una definición más gradual de HVD, en lugar del criterio de sí/no basado en el umbral 30 %.
- Se propone un ajuste del nivel de significancia en un test de bioequivalencia promedio escalada de referencia (RSABE) que se basa en datos de un diseño crossover. El objetivo es controlar el error de Tipo I.

Capítulo 3

Métodos propuestos

3.1. Método HoweEMA

La Observación 11 realizada al método EMA, en particular la que indica que el criterio de inclusión de intervalos, CIIP no esta bien aplicado, tiene una solución que consiste en utilizar parte del método de la FDA que se describe a continuación:

Criterio de decisión de RSABE, HoweEMA

1. Obtener la estimación del efecto formulación, $\hat{\phi}$.
2. Si $\hat{\phi}$ está fuera de ± 0.223 no se declara BE y el proceso termina.
3. Obtener la estimación de σ_{WR} , $\hat{\sigma}_{WR}$.
 - Si $\hat{\sigma}_{WR} < \sigma_{CV=30} \vee \hat{\sigma}_{WR} \geq \sigma_{CV=50}$ entonces,
 - a) Obtener el intervalo de confianza del 90% para el efecto formulación ϕ , IC_{ϕ} .
 - b) Si IC_{ϕ} al 90% está completamente incluido en los límites $\pm \phi_{EMA}(\hat{\sigma}_{WR})$ (es decir, ± 0.223 si $\hat{\sigma}_{WR} < \sigma_{CV=30} \vee \pm 0.359$ si $\hat{\sigma}_{WR} \geq \sigma_{CV=50}$), se declara BE, en otro caso, no se declara BE y el proceso termina.
4. otro caso,
 - a) Obtener límite superior $\hat{\eta}_U$ del intervalo de confianza para $\eta = \phi^2 - k_{EMA}^2 \sigma_{WR}^2$, $IC_{\eta} = (-\infty, \hat{\eta}_U]$.
 - b) Si $\hat{\eta}_U < 0$ declara BE, en otro caso, no se declara BE.

3.2. Métodos: límites continuos en el enfoque de la FDA, ContFDA y ContFDA2

Otra de las soluciones que se proponen es considerar límites de BE continuos en el método de la FDA. De esta manera, si el $CV_{WR} = 30\%$ y se considera una constante de escala definida por $k_{FDA} = k_{EMA} = 0.76$, se logra la continuidad en el punto. Esto describe un nuevo método que se llama “ContFDA”. Alternativamente, si el factor de escala $k_{FDA} = 0.893$ debe ser redefinido, $\sigma_0 = 0.25$ ($CV_0 = 25.40\%$). Se designa éste último enfoque como “ContFDA2”.

Observación 15 *Podemos definir de modo general ϕ_{RSABE} el cual representa la función que define los límites de bioequivalencia de RSABE, así las hipótesis a probar pueden ser expresadas:*

$$\begin{aligned} H_0 : \phi &\leq -\phi_{RSABE}(\sigma_{WR}) \quad \vee \quad \phi \geq \phi_{RSABE}(\sigma_{WR}) \\ H_1 : -\phi_{RSABE}(\sigma_{WR}) &< \phi < \phi_{RSABE}(\sigma_{WR}) \end{aligned} \quad (3.1)$$

Luego en el caso de EMA o FDA, se tiene que $\phi_{RSABE}(\sigma_{WR}) = \phi_{EMA}(\sigma_{WR})$ o $\phi_{RSABE}(\sigma_{WR}) = \phi_{FDA}(\sigma_{WR})$ definidas anteriormente.

3.3. Método LO

Las pruebas de bioequivalencia asociadas a los límites escalados (RSABE) definidos anteriormente, unas de las objeciones que presentan es que se basan en límites de bioequivalencia poco regulares. Esto quizás es una posible causa de problemas de control de error tipo I.

Algunos autores como Karalis et al.(2011) y Kytariolos J et al.(2006)[16, 19] han propuesto límites más regulares que se basa en la función Sigmoidea, Weibull y otras. En particular, en esta memoria se utiliza la propuesta de los autores Karalis et al.(2011), porque emulan muy fielmente los límites EMA, son muy comparables a ellos.

Si redefinimos las hipótesis RSABE pero ahora considerando límites suaves se tiene que:

$$\begin{aligned} H_0 : \phi &\leq -\phi_{LO}(\sigma_{WR}) \quad \vee \quad \phi \geq \phi_{LO}(\sigma_{WR}) \\ H_1 : -\phi_{LO}(\sigma_{WR}) &< \phi < \phi_{LO}(\sigma_{WR}) \end{aligned} \quad (3.2)$$

donde ϕ_{LO} representa la función que define los límites suaves leveling-off

$$\phi_{LO}(\sigma) = \log(s(\sigma)) \quad (3.3)$$

y s representa la función Sigmoidea en [16] que define los límites de la escala original:

$$s(\sigma) = \phi_0 + \frac{\delta}{1 + \exp\left\{-\frac{\sigma-\tau}{\gamma}\right\}} \quad (3.4)$$

con:

$$\begin{aligned} \phi_0 &= 1.25 \\ \delta &= 1.4319 - 1.25 \\ \tau &= 0.3853 \\ \gamma &= 0.0336. \end{aligned} \quad (3.5)$$

En la figura 3.1 se puede apreciar la gran similitud entre los límites de BE de la EMA y los límites de BE suaves.

Cabe mencionar que las estimaciones de τ y γ se obtienen del artículo [16] a través de métodos empíricos.

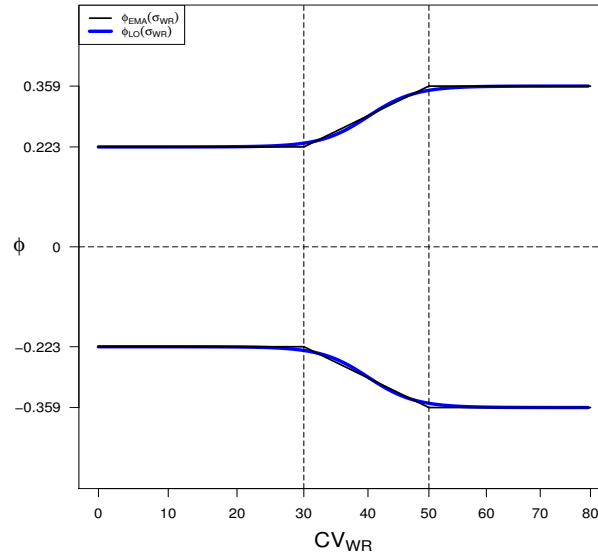


Figura 3.1: Límites de bioequivalencia promedio escalada (RSABE) EMA [8] y límites RSABE suaves leveling-off (LO) [16] basado en una función suave que emula de cerca los límites de EMA. Ambos límites son definidos en función de σ_{WR} en escala logarítmica pero representado aquí en función del coeficiente de variación CV_{WR} en escala original. En los límites EMA, la escala lineal comienza en $\sigma_{WR} = 0.294$ ($CV_{WR} = 30\%$) y termina en $\sigma_{WR} = 0.472$ ($CV_{WR} = 50\%$) donde se vuelven constantes. Los límites LO se basan en la función 3.3. La condición de BE verdadera se supone si el valor verdadero pero desconocido valor del efecto de formulación ϕ se encuentra entre los límites correspondientes, de hecho, el concepto de BE difiere según los límites considerados. El efecto de formulación se refiere C_{max} en la regulación EMA, siempre considerado en escala logarítmica.

3.3.1. Criterio de decisión de RSABE, LO

1. Obtener la estimación del efecto formulación, $\hat{\phi}$.
2. Si $\hat{\phi}$ está fuera de ± 0.223 no se declara BE y el proceso termina.
3. Obtener la estimación de σ_{WR} , $\hat{\sigma}_{WR}$.
4. Si CI_{ϕ} está completamente incluido en los límites $\pm \phi_{LO}(\hat{\sigma}_{WR})$, se declara BE (Rechazar H_0), en otro caso no se declara BE.

El criterio de decisión que se ha indicado anteriormente (método LO) tiene uno de los posibles inconvenientes indicados en [23] como causas potenciales de no validez de las pruebas de BE, es decir, el uso del principio de inclusión de intervalos (CIIP) con límites de BE aleatorios (que son una función de la cantidad aleatoria $\hat{\sigma}_{WR}$) en lugar de ser límites constantes previamente fijados.

Observación 16 *Notar que el criterio mencionado anteriormente no requiere de un pretest para decidir si se aplica o no escalamiento.*

Por otra parte, obteniendo la primera y segunda derivada de la función 3.4 :

$$s'(\sigma) = \frac{\delta \exp\left(\frac{-\sigma-\tau}{\gamma}\right)}{\gamma \left(1 + \exp\left(\frac{-\sigma-\tau}{\gamma}\right)\right)^2}$$

$$s''(\sigma) = \frac{\delta \exp\left(\frac{-\sigma-\tau}{\gamma}\right) \left(\exp\left(\frac{-\sigma-\tau}{\gamma}\right) - 1\right)}{\gamma^2 \left(1 + \exp\left(\frac{-\sigma-\tau}{\gamma}\right)\right)^3}.$$

Por lo tanto, las dos primeras derivadas de $\phi_{LO}(\sigma)$ se pueden expresar como:

$$\phi'_{LO}(\sigma) = \frac{s'(\sigma)}{s(\sigma)}$$

$$\phi''_{LO}(\sigma) = \frac{s'(\sigma)s''(\sigma) - (s'(\sigma))^2}{(s(\sigma))^2}.$$

Estos resultados serán utilizados en las siguientes secciones y en el capítulo de resultados donde se va a explicar en forma más exhaustiva.

3.4. Método Howe-LO

En esta sección se aplica el método de Howe pero considerando los límites suaves propuestos, definidos en la sección anterior, redefiniendo las hipótesis de RSABE de tal manera de construir una suma de variables aleatorias con distribuciones conocidas, en la cual una de estas variables aleatorias es el estimador de la función de límites suaves que depende de σ_{WR} .

Definiendo el parámetro de interés

$$\theta = \frac{\phi}{\phi_{LO}(\sigma_{WR})}$$

y considerando el contraste de hipótesis RSABE con límites más regulares se tiene:

$$\begin{aligned} H_0 : \phi &\leq -\phi_{LO}(\sigma_{WR}) \quad \vee \quad \phi \geq \phi_{LO}(\sigma_{WR}) \\ H_1 : -\phi_{LO}(\sigma_{WR}) &< \phi < \phi_{LO}(\sigma_{WR}) \end{aligned} \quad (3.6)$$

reemplazando el parámetro de interés en el contraste (3.6) y elevando al cuadrado se pueden definir equivalentemente:

$$\begin{aligned} H_0 : \theta^2 &\geq 1 \\ H_1 : \theta^2 &< 1. \end{aligned} \quad (3.7)$$

Despejando se define el siguiente contraste de hipótesis:

$$\begin{aligned} H_0 : \eta &\geq 0 \\ H_1 : \eta &< 0 \end{aligned} \quad (3.8)$$

donde el nuevo parámetro de interés es:

$$\eta = \phi^2 - \phi_{LO}^2(\sigma_{WR}).$$

Para resolver esta prueba de hipótesis utilizamos el método inferencial del CIIP. La construcción del intervalo de confianza unilateral se realiza mediante el método de Howe que consiste en tener los intervalos de confianza $1 - \alpha$ unilateral $(-\infty, \hat{\phi}_U^2]$ para ϕ^2 y unilateral $[\hat{\sigma}_L, +\infty)$ para σ_{WR} .

La construcción de los intervalos de confianza unilaterales que se definirán a continuación, se obtienen por el método del pivote:

- Intervalo de confianza para σ_{WR} : se sabe que la variable aleatoria $\hat{\sigma}_{WR}^2$ tiene la siguiente distribución:

$$\hat{\sigma}_{WR}^2 \sim \frac{\sigma_{WR}^2}{\nu} \chi^2(\nu).$$

Luego despejando se obtiene la cantidad pivotal que se distribuye ji-cuadrado con N-K grados de libertad:

$$\frac{\nu \hat{\sigma}_{WR}^2}{\sigma_{WR}^2} \sim \chi^2(\nu).$$

De esta manera, se construye el intervalo de confianza para la varianza intra-sujeto del efecto de formulación de referencia, σ_{WR}^2 , y para su desviación estándar σ_{WR} . Considerando un nivel de confianza de $(1 - \alpha)100\%$ se tiene que:

$$P \left\{ \frac{\nu \hat{\sigma}_{WR}^2}{\sigma_{WR}^2} \leq \chi_{1-\alpha, \nu}^2 \right\} = 1 - \alpha$$

equivalentemente

$$P \left\{ \hat{\sigma}_L^2 = \frac{\nu \hat{\sigma}_{WR}^2}{\chi_{1-\alpha, \nu}^2} \leq \sigma_{WR}^2 \right\} = 1 - \alpha$$

es el extremo inferior del intervalo de confianza $1 - \alpha$ $[\hat{\sigma}_L^2, +\infty)$ unilateral para σ_{WR}^2 .

Luego por la monotonicidad de la raíz cuadrada:

$$P \left\{ \sqrt{\frac{\nu \hat{\sigma}_{WR}^2}{\chi_{1-\alpha, \nu}^2}} \leq \sqrt{\sigma_{WR}^2} \right\} = 1 - \alpha$$

$$\hat{\sigma}_L = \sqrt{\frac{\nu \hat{\sigma}_{WR}^2}{\chi_{1-\alpha, \nu}^2}}$$

es el extremo inferior de intervalo de confianza unilateral para σ_{WR} $[\hat{\sigma}_L, +\infty)$.

Por otra parte, por la monotonicidad de ϕ_{LO} se tiene que:

$$P \left\{ -\phi_{LO}^2(\hat{\sigma}_{WR})_U = -\phi_{LO}^2 \left(\sqrt{\frac{\nu \hat{\sigma}_{WR}^2}{\chi_{1-\alpha, \nu}^2}} \right) \geq -\phi_{LO}^2(\sigma_{WR}) \right\} = 1 - \alpha$$

es el extremo superior de un intervalo de confianza unilateral $1 - \alpha$ $(-\infty, -\phi_{LO}^2(\hat{\sigma}_{WR})_U]$ para $-\phi_{LO}^2(\sigma_{WR})$.

- Intervalo de confianza $1 - \alpha$ unilateral $(-\infty, \hat{\phi}_U^2]$ para ϕ^2 :

$$\hat{\phi}_U^2 = (|\hat{\phi}| + \widehat{SE}_{\hat{\phi}} t_{1-\alpha, \nu})^2$$

Luego cuando la inferencia se realiza sobre el parámetro compuesto

$$\eta = \phi^2 - \phi_{LO}^2(\sigma_{WR})$$

se declara BE cuando se verifica que $\hat{\eta}_U < 0$ donde $\hat{\eta}_U$ es el límite superior de un intervalo de confianza unilateral $(-\infty, \hat{\eta}_U]$ de nivel de confianza (nominal) $1 - \alpha$ para θ . En el método “Howe-LO” el intervalo de confianza aproximado se construye mediante el método de Howe. Entonces $\hat{\eta}_U$ se construye exactamente:

$$\hat{\eta}_U = E_m - E_s + \sqrt{L_m + L_s}$$

donde:

$$E_m = \hat{\phi}^2$$

$$E_s = \phi_{LO}^2(\hat{\sigma}_{WR})$$

$$C_m = (|\hat{\phi}| + \widehat{SE}_{\hat{\phi}} t_{1-\alpha, \nu})^2$$

$$C_s = \phi_{LO}^2 \left(\sqrt{\frac{\nu \hat{\sigma}_{WR}^2}{\chi_{1-\alpha, \nu}^2}} \right)$$

$$L_m = (C_m - E_m)^2 = \left\{ (|\hat{\phi}| + \widehat{SE}_{\hat{\phi}} t_{1-\alpha, \nu})^2 - \hat{\phi}^2 \right\}^2$$

$$L_s = (C_s - E_s)^2 = \left\{ \phi_{LO}^2 \left(\sqrt{\frac{\nu \hat{\sigma}_{WR}^2}{\chi_{1-\alpha, \nu}^2}} \right) - \phi_{LO}^2(\hat{\sigma}_{WR}) \right\}^2.$$

Así se tiene que:

$$\hat{\eta}_U = \hat{\phi}^2 - \phi_{LO}^2(\hat{\sigma}_{WR}) + \sqrt{\left\{ (|\hat{\phi}| + \widehat{SE}_{\hat{\phi}} t_{1-\alpha, \nu})^2 - \hat{\phi}^2 \right\}^2 + \left\{ \phi_{LO}^2 \left(\sqrt{\frac{\nu \hat{\sigma}_{WR}^2}{\chi_{1-\alpha, \nu}^2}} \right) - \phi_{LO}^2(\hat{\sigma}_{WR}) \right\}^2} \quad (3.9)$$

Los estadísticos indicados con E_m y E_s corresponden a estimadores puntuales de ϕ^2 y $\phi_{LO}(\sigma_{WR})$ respectivamente, mientras que los indicados como “C” corresponden a extremos de intervalos de confianza unilaterales para dichos parámetros. Para C_s es correcto dada la monotonicidad de la función ϕ_{LO} , pero E_s proporciona una estimación sesgada de $\phi_{LO}^2(\sigma_{WR})$ con sesgo muy dependiente de la segunda derivada de ϕ_{LO} .

3.4.1. Corrección del sesgo en la estimación de los límites de bioequivalencia escalados

Se define la función $h(\sigma^2) = \phi_{LO}^2((\sigma^2)^{\frac{1}{2}})$. Su primera derivada respecto a σ^2 es:

$$h'(\sigma^2) = (\sigma^2)^{-\frac{1}{2}} \phi_{LO}((\sigma^2)^{\frac{1}{2}}) \phi'_{LO}((\sigma^2)^{\frac{1}{2}})$$

y la segunda derivada:

$$h''(\sigma^2) = \frac{\phi_{LO}((\sigma^2)^{\frac{1}{2}}) \phi''_{LO}((\sigma^2)^{\frac{1}{2}})}{2((\sigma^2)^{\frac{1}{2}})((\sigma^2)^{\frac{1}{2}})} - \frac{\phi'_{LO}((\sigma^2)^{\frac{1}{2}}) \phi_{LO}((\sigma^2)^{\frac{1}{2}})}{2(\sigma^2)^{\frac{3}{2}}} + \frac{(\phi'_{LO}((\sigma^2)^{\frac{1}{2}}))^2}{2(\sigma^2)^{\frac{1}{2}}(\sigma^2)^{\frac{1}{2}}}.$$

Agrupando los términos semejantes y factorizando se tiene:

$$h''(\sigma^2) = \frac{1}{2} \left(\frac{\left(\phi'_{LO}((\sigma^2)^{\frac{1}{2}}) \right)^2 + \phi_{LO}((\sigma^2)^{\frac{1}{2}}) \phi''_{LO}((\sigma^2)^{\frac{1}{2}})}{\sigma^2} - \frac{\left(\phi_{LO}((\sigma^2)^{\frac{1}{2}}) \right) \phi'_{LO}((\sigma^2)^{\frac{1}{2}})}{(\sigma^2)^{\frac{3}{2}}} \right).$$

Por lo tanto, desarrollando h en serie de Taylor hasta el grado 2 y considerando $\sigma^2 = \sigma_{WR}^2$:

$$\underbrace{h(\hat{\sigma}_{WR}^2)}_{=\phi_{LO}^2(\hat{\sigma}_{WR})} = \underbrace{h(\sigma_{WR}^2)}_{=\phi_{LO}^2(\sigma_{WR})} + h'(\sigma_{WR}^2)(\hat{\sigma}_{WR}^2 - \sigma_{WR}^2) + \frac{h''(\sigma_{WR}^2)}{2}(\hat{\sigma}_{WR}^2 - \sigma_{WR}^2)^2 + R_2(h)$$

luego aplicando el operador esperanza y despreciando el residuo quedaría:

$$E(\phi_{LO}^2(\hat{\sigma}_{WR})) - \phi_{LO}^2(\sigma_{WR}) = \text{bias}(\phi_{LO}^2(\hat{\sigma}_{WR})) = h'(\sigma_{WR}^2) \underbrace{E(\hat{\sigma}_{WR}^2 - \sigma_{WR}^2)}_{=0} + \frac{h''(\sigma_{WR}^2)}{2} E((\hat{\sigma}_{WR}^2 - \sigma_{WR}^2)^2)$$

$$\text{bias}(\phi_{LO}^2(\hat{\sigma}_{WR})) \simeq \frac{1}{4} C \text{var}(\hat{\sigma}_{WR}^2) \quad (3.10)$$

donde C es un valor constante que depende de σ_{WR} y de la función que define los límites escalados ϕ_{LO} con su primera y segunda derivada:

$$C = \frac{[\phi'_{LO}(\sigma_{WR})]^2 + \phi_{LO}(\sigma_{WR}) \phi''_{LO}(\sigma_{WR})}{\sigma_{WR}^2} - \frac{\phi_{LO}(\sigma_{WR}) \phi'_{LO}(\sigma_{WR})}{(\sigma_{WR})^3}. \quad (3.11)$$

Al comparar las escalas de valores tomados por la función ϕ_{LO} y sus distintas derivadas figura 3.2 se comprende el peso que podría tener la segunda derivada:

$$\begin{aligned} \frac{[\phi'_{LO}(\sigma_{WR})]^2}{\sigma_{WR}^2} &< \left| \frac{\phi_{LO}(\sigma_{WR})\phi''_{LO}(\sigma_{WR})}{\sigma_{WR}^2} \right| \Leftrightarrow [\phi'_{LO}(\sigma_{WR})]^2 < |\phi_{LO}(\sigma_{WR})\phi''_{LO}(\sigma_{WR})| \\ \frac{\phi_{LO}(\sigma_{WR})\phi'_{LO}(\sigma_{WR})}{(\sigma_{WR})^3} &< \left| \frac{\phi_{LO}(\sigma_{WR})\phi''_{LO}(\sigma_{WR})}{\sigma_{WR}^2} \right| \Leftrightarrow \frac{\phi'_{LO}(\sigma_{WR})}{\sigma_{WR}} < |\phi''_{LO}(\sigma_{WR})| \end{aligned} \quad (3.12)$$

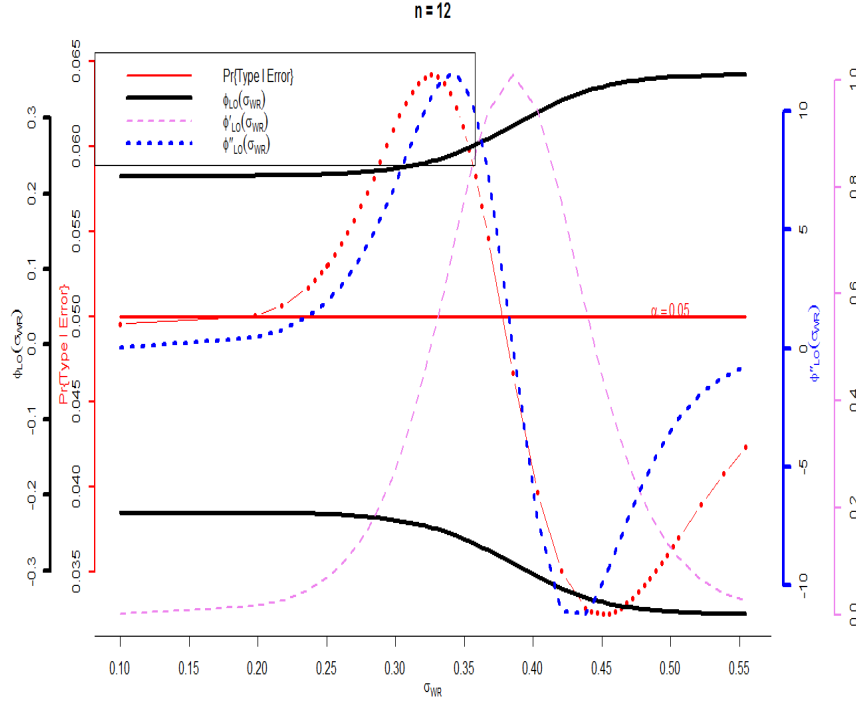


Figura 3.2: Representación conjunta de la probabilidad de declarar BE para el método Howe-LO (para el caso $n = 12$; $N = 36$) y la función que define los límites $\phi_{LO}(\sigma_{WR})$ y sus dos primeras derivadas, en un diseño parcialmente replicado TRR/RTR/RRT.

A partir de las implementaciones de estas funciones $\phi_{LO}(\sigma)$, $\phi'_{LO}(\sigma)$, $\phi''_{LO}(\sigma)$, se puede ver que las oscilaciones de la probabilidad de error de tipo I siguen de una forma considerablemente fiel a las oscilaciones de la segunda derivada para el método de Howe-LO (ver figura 3.2).

En la situación paramétrica-normal que se está considerando en esta tesis la varianza de $\text{var}(\hat{\sigma}_{WR}^2)$ tiene distribución ji-cuadrado con $\nu = N - K$ grados de libertad multiplicado por el factor $\frac{\sigma_{WR}^2}{\nu}$. Por lo tanto,

$$\text{var}(\hat{\sigma}_{WR}^2) = \left(\frac{\sigma_{WR}^2}{\nu}\right)^2 2\nu = \frac{2(\sigma_{WR}^2)^2}{\nu}$$

como consecuencia de (3.10) y (3.11) el sesgo puede ser estimado por:

$$\begin{aligned} \widehat{\text{bias}}(\phi_{LO}^2(\hat{\sigma}_{WR})) &= \frac{1}{4} \left(\frac{[\phi'_{LO}(\hat{\sigma}_{WR})]^2 + \phi_{LO}(\hat{\sigma}_{WR}) \phi''_{LO}(\hat{\sigma}_{WR})}{\hat{\sigma}_{WR}^2} - \frac{\phi_{LO}(\hat{\sigma}_{WR}) \phi'_{LO}(\hat{\sigma}_{WR})}{(\hat{\sigma}_{WR})^3} \right) \frac{2(\hat{\sigma}_{WR}^2)^2}{\nu} \\ &= \frac{\hat{\sigma}_{WR}^2}{2\nu} \left([\phi'_{LO}(\hat{\sigma}_{WR})]^2 + \phi_{LO}(\hat{\sigma}_{WR}) \phi''_{LO}(\hat{\sigma}_{WR}) - \frac{\phi_{LO}(\hat{\sigma}_{WR}) \phi'_{LO}(\hat{\sigma}_{WR})}{\hat{\sigma}_{WR}} \right). \end{aligned} \quad (3.13)$$

Luego esta expresión se tendría que restar al término indicado con E_s en el método “HoweLO” definiéndose así el enfoque “bcHoweLO” que se analizará con más detalle en el capítulo de resultados.

Otras posibles correcciones en el método HoweLO

A pesar que $\hat{\sigma}_{WR}^2$ es un estimador insesgado de σ_{WR}^2 , su cuadrado no es un estimador insesgado, pues si se considera la versión exacta de un desarrollo de Taylor de grado 2 se tiene:

$$(\hat{\sigma}_{WR}^2)^2 = (\sigma_{WR}^2)^2 + 2\sigma_{WR}^2(\hat{\sigma}_{WR}^2 - \sigma_{WR}^2) + (\hat{\sigma}_{WR}^2 - \sigma_{WR}^2)^2$$

luego tomando esperanzas se obtiene:

$$E((\hat{\sigma}_{WR}^2)^2) - (\sigma_{WR}^2)^2 = \text{bias}(\hat{\sigma}_{WR}^2)^2 = \text{var}(\hat{\sigma}_{WR}^2)$$

$$\widehat{\text{bias}}\left((\hat{\sigma}_{WR}^2)^2\right) = \widehat{\text{var}}(\hat{\sigma}_{WR}^2) = \frac{2(\hat{\sigma}_{WR}^2)^2}{\nu}.$$

El sesgo de la estimación de $\text{var}(\hat{\sigma}_{WR}^2)$ se podría mejorar mediante:

$$\widetilde{\text{var}}(\hat{\sigma}_{WR}^2) = \frac{2 \left[(\hat{\sigma}_{WR}^2)^2 - \widehat{\text{bias}}\left((\hat{\sigma}_{WR}^2)^2\right) \right]}{\nu} = \frac{2 \left[(\hat{\sigma}_{WR}^2)^2 - 2(\hat{\sigma}_{WR}^2)^2/\nu \right]}{\nu} = \frac{2(\nu - 2)}{\nu^2} (\hat{\sigma}_{WR}^2)^2.$$

Así, se tendría la siguiente expresión para estimar el sesgo de $\phi_{LO}^2(\hat{\sigma}_{WR})$

$$\widetilde{\text{bias}}(\phi_{LO}^2(\hat{\sigma}_{WR})) = \frac{(\nu - 2)\hat{\sigma}_{WR}^2}{2\nu^2} \left([\phi'_{LO}(\hat{\sigma}_{WR})]^2 + \phi_{LO}(\hat{\sigma}_{WR}) \phi''_{LO}(\hat{\sigma}_{WR}) - \frac{\phi_{LO}(\hat{\sigma}_{WR}) \phi'_{LO}(\hat{\sigma}_{WR})}{\hat{\sigma}_{WR}} \right). \quad (3.14)$$

Análogamente a lo realizado en (3.13), se tendría que restar a E_s el término (3.14) en el método “HoweLO”.

De la misma manera $\phi_{LO}^2(\hat{\sigma}_{WR})$ es un estimador sesgado de $\phi_{LO}^2(\sigma_{WR})$ y se puede corregir con las expresiones anteriores, a la estimación de ϕ^2 , $\hat{\phi}^2$, se le puede añadir una corrección para el sesgo. A partir del mismo razonamiento que se ha hecho para el cuadrado de la varianza, tenemos que:

$$\hat{\phi}^2 = \phi^2 + 2\phi(\hat{\phi} - \phi) + (\hat{\phi} - \phi)^2$$

aplicando esperanza y considerando efecto carryover nulo se tiene que:

$$E(\hat{\phi}^2) = E(\phi^2) + 2\phi \underbrace{E((\hat{\phi} - \phi))}_0 + E((\hat{\phi} - \phi)^2)$$

$$E(\hat{\phi}^2) - \phi^2 = \text{bias}(\hat{\phi}^2) = \text{var}(\hat{\phi})$$

$$\widehat{\text{bias}}(\hat{\phi}^2) = \widehat{\text{var}}(\hat{\phi}) = \widehat{SE}_{\hat{\phi}}^2$$

Luego al término E_m se le resta $\widehat{SE}_{\hat{\phi}}^2$.

Observación 17 *La comparación de estos métodos suavizados y corregidos se presentan en el capítulo de resultados.*

3.5. Ajuste del nivel de significación

Bajo el modelo (1.1) y suponiendo un diseño bien llevado a cabo, sin efecto carryover (es decir, con un adecuado washout tiempos entre cada tratamientos administrados), la probabilidad de declarar la bioequivalencia es una función p , que depende: del verdadero efecto formulación que se fije ϕ , de las variabilidades dentro de los sujetos σ_{WR} y σ_{WT} , del diseño crossover elegido D , de los tamaños muestrales en cada secuencia $\mathbf{n} = (n_1, \dots, n_K)$ y el preespecificado nivel de significancia, α .

$$p(\phi, \sigma_{WR}, \sigma_{WT}, D, \mathbf{n}, \alpha)$$

Idealmente, esta probabilidad debe ser igual a α para valores de ϕ en el límite de bioequivalencia, es decir, $\phi = \phi_{RSABE}(\sigma_{WR})$ y monótonamente aumentar para valores de ϕ progresivamente dentro de la región de bioequivalencia $-\phi_{RSABE}(\sigma_{WR}) < \phi < \phi_{RSABE}(\sigma_{WR})$ y monótonamente decreciente para valores progresivamente fuera de la región de bioequivalencia. Esta condición de monotonicidad se cumple, pero en general:

$$p(\phi = \phi_{RSABE}(\sigma_{WR}), \sigma_{WR}, \sigma_{WT}, D, \mathbf{n}, \alpha) \neq \alpha.$$

Todos los parámetros fijados, están en función σ_{WR} . Por otra parte, en todos los métodos aplicados tienen un máximo absoluto en el límite HVD, es decir σ_0 , que coincide con $\sigma_{WR} = 0.294$ ($CV_{WR} = 30\%$ en la escala original), excepto $\sigma_{WR} = 0.25$ para el método contFDA2.

En resumen, el método consiste en resolver, con respecto a la variable α_{adj} , la ecuación:

$$\alpha = p(\phi = \phi_{RSABE}(\sigma_{WR}), \sigma_{WR} = \sigma_{WT} = \sigma_0, D, \mathbf{n}, \alpha_{adj}). \quad (3.15)$$

Si se requiere un control total del error de tipo I, un posible enfoque sería ajustar el nivel de significación para evitar sobrepasar el valor nominal anteriormente expresado ($\alpha = 0.05$). Un similar enfoque se consideró en el [20] e implementó en el paquete R PowerTOST por medio de su función `scABEL.ad`, como se discute posteriormente.

En primer lugar, se realizaron simulaciones exhaustivas (1.000.000 de repeticiones de simulación) proporcionando la verdadera probabilidad máxima de error de tipo I de diferentes niveles de confianza nominal y tamaños de muestra por cada secuencia balanceada. A partir de estos resultados de simulación empíricamente se ajustaron las funciones lineales para cada uno de los cinco métodos considerados y los diseños (TRR/RTR/RRT o TRTR/RTRT). Dado un tamaño de muestra y un nivel de significación objetivo, éstas funciones devuelven un nivel de significancia ajustado para igualar aproximadamente este α .

Los valores aproximados obtenidos se utilizan como valores iniciales para un proceso iterativo basado en $m = 1.000.000$ réplicas de simulación obteniendo el nivel de significancia ajustado α_{adj} que verifica la ecuación (3.15).

Para un diseño dado y vector de tamaños de muestra, no necesariamente balanceado, se debe tener en cuenta que este proceso implica repetir repetidamente sobre los posibles valores de α_{adj} para calcular la proporción de las declaraciones de bioequivalencia, pero esto implica que los procesos consumen más tiempo, es decir, la generación de los m conjuntos de datos y el posterior cálculo de las estimaciones relevantes ($\hat{\phi}, \hat{\sigma}_{WR}^2$), sobre los conjuntos de datos generados se realizan sólo una vez. Este método se implementa en la función `adjAlpha` en el paquete R `adjustalpha`.

En todas las simulaciones y ejemplos presentados aquí, el ajuste del nivel de significación se realizó usando la función `adjAlpha`, excepto para el método de regulación 2010 de la EMA anterior que se indica en EMA(2010), para este método, el ajuste se realizó mediante la función `scABEL.ad`, del package `PowerTOST` [20].

Capítulo 4

Resultados y Discusión

4.1. Resultados

4.1.1. Simulaciones

El objetivo de las simulaciones realizadas en esta memoria es comparar distintos métodos (RSABE) y proponer un ajuste del nivel de significancia en un test de bioequivalencia promedio escalada de referencia (RSABE) que se basa en datos de un diseño crossover con el fin de controlar el error de Tipo I.

Así al considerar límites constante en lugar de límites aleatorios en el principio de inclusión de intervalos, se propone el método HoweEMA y se compara con el comportamiento del criterio regulatorio RSABE EMA 2010. Además se proponen dos métodos alternativos donde los límites de BE son continuos y se comparan con el comportamiento del criterio regulatorio RSABE FDA 2010.

Se estudiaron los diseños parcialmente replicados TRR/RTR/RRT balanceados, con idéntico número de sujetos asignados a cada secuencia en forma aleatoria. Se consideraron los siguientes tamaños muestrales por secuencia: $n = 8, 12, 16$ y 24 (es decir, $N = 24, 36, 48$ y 72 tamaños totales de las tablas de datos generadas), comunes en estudios de BE reales. El otro diseño estudiado fue el diseño replicado TRTR/RTRT balanceado, para tamaños muestrales por secuencia de $n = 8, 12, 16$ y 24 y se consideró válida la típica suposición de normalidad de los datos en la escala logarítmica y el supuesto de homocedasticidad en las varianzas que representan la variabilidad intrasujetos.

Como se comentó anteriormente, un fármaco es considerado de alta variabilidad si su coeficiente de variación intrasujeto para la formulación de referencia es superior o igual a 30%. Sin embargo, se simularon también valores inferiores para ver su comportamiento frente a las metodologías estudiadas.

En las simulaciones de los diseños crossover sólo se considera presente el efecto del tratamiento. Los otros efectos fueron considerados nulos, es decir, se definen con un vector de ceros en cada caso. La suposición de carryover nulo (igual para ambas formulaciones, que bajo la restricción de que los efectos fijos sumen cero es equivalente a que valga cero, es decir, suponer que se ha aplicado un período washout suficiente) es muy importante, ya

que en caso contrario, los estimadores del efecto formulación serán sesgados. La suposición de que los restantes efectos son nulos es puramente convencional, hace que los datos estén centrados en valores posiblemente no realistas, pero sin ningún efecto especial.

Los escenarios a simular más interesantes son aquellos en los que el efecto formulación simulado está justo en los límites de BE y donde, idealmente, la probabilidad de declarar BE tendría que igualar el nivel de significación nominal, mientras que valores superiores serán indicativos de un test no válido. Otros posibles escenarios corresponderán a situaciones en las que el efecto formulación fuese una fracción λ (mayor o menor que 1) del valor que estará sobre el límite de BE de manera que, si $\lambda < 1$ simulamos valores pequeños (H_0 falsa, poca diferencia), en particular, se simulan escenarios donde $\lambda = 0.8$ figura 4.3.

La figura 4.1 muestra la probabilidad de declarar BE en los mismos escenarios de la Figura 2 presentados en el artículo [23] donde se estudian los métodos en el diseño replicado TRTR/RTRT, pero ajustando el nivel de significación para respetar el nivel 0,05. Esta probabilidad corresponde al caso en que el valor real del efecto formulación se encuentra exactamente en los límites de bioequivalencia (y en este caso declarar la bioequivalencia es un error Tipo I).

La figura 4.2 corresponde al caso más favorable de bioequivalencia: efecto formulación nulo, excepto para tamaños muy pequeño (por ejemplo para $n = 8$ sujetos en cada secuencia en un diseño parcialmente replicado) el ContFDA ajustado, con la misma constante de escalada que el método EMA, se comporta de forma muy similar a la HoweEMA, excepto para variabilidades altas. Para tamaños de muestra medianos (por ejemplo $n = 12$ por secuencia) todos los métodos son suficientemente potentes y para tamaños de muestra grande todos los métodos son indistinguibles, con potencias muy altas.

La figura 4.3 corresponde al caso de verdadera bioequivalencia pero una situación desfavorable para la posibilidad de declarar bioequivalencia: el verdadero efecto de la formulación está dentro de los límites pero cerca de ellos, es decir, cuando toma el valor 80 % todos los métodos son muy similares potencias bajas, siendo muy difícil de demostrar bioequivalencia en todos estos escenarios.

Con el fin de cuantificar la potencia global a lo largo de los valores en estudio de CV_{WR} , se calcula el área bajo la curva de la potencia vs CV_{WR} (estos valores se designarán auc_{pow}).

Para todos los métodos y sus versiones ajustadas y no ajustadas, se define el porcentaje de pérdida conjunta en términos de auc_{pow} , como

$$100 \cdot \frac{auc_{pow}(unadjusted) - auc_{pow}(adjusted)}{auc_{pow}(unadjusted)}$$

y además se calculó la máxima pérdida de potencia para cada método y el valor del CV_{WR} que lo indicaba. Esto se resume en la tabla 4.1 que corresponde al escenario más favorable de declaración de BE, es decir, efecto formulación nulo $\phi = 0$.

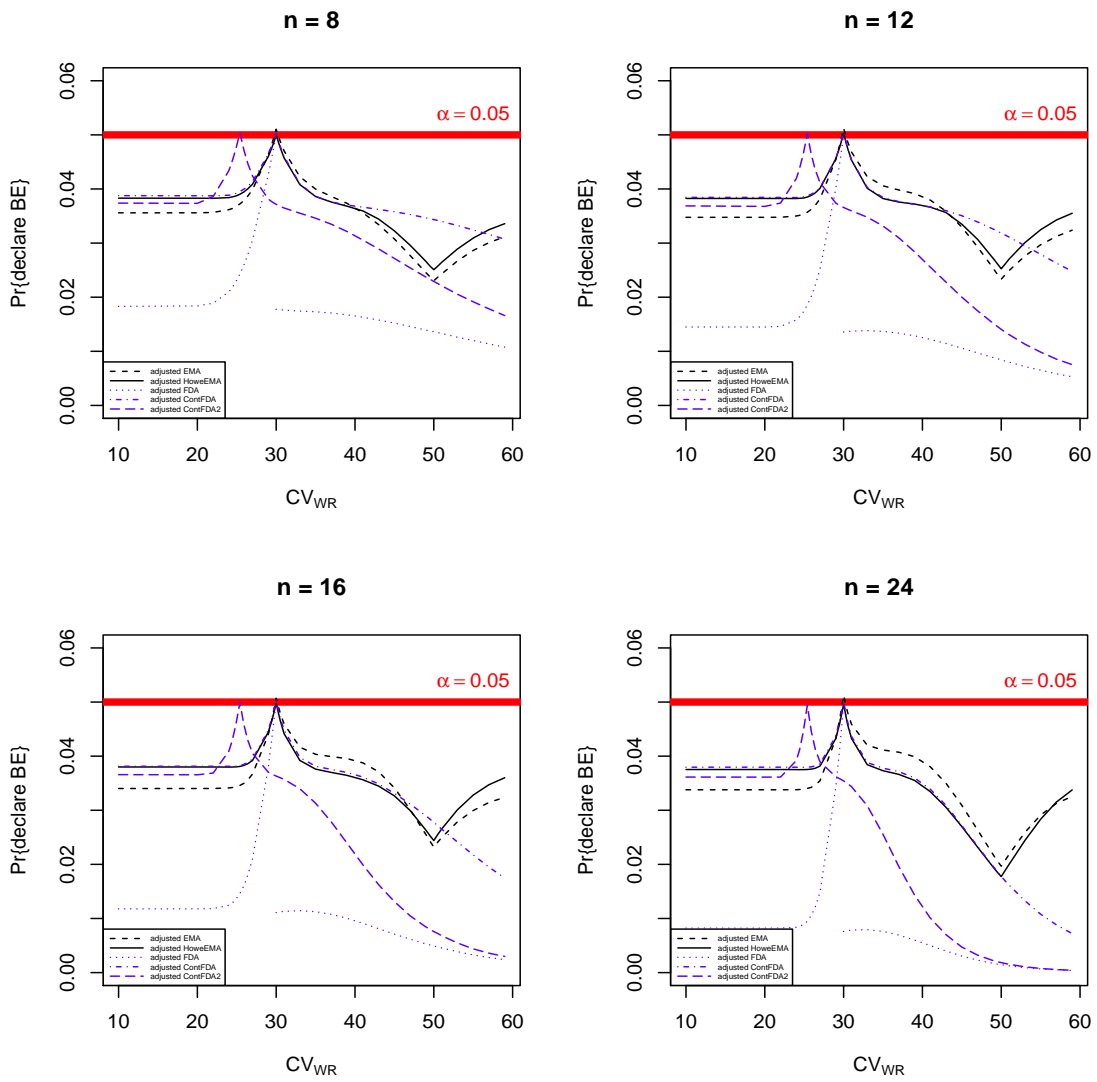


Figura 4.1: Esta probabilidad corresponde al caso en que el valor verdadero del efecto formulación se encuentra exactamente en los límites de bioequivalencia, en este caso la probabilidad de declarar BE es un error de tipo I.

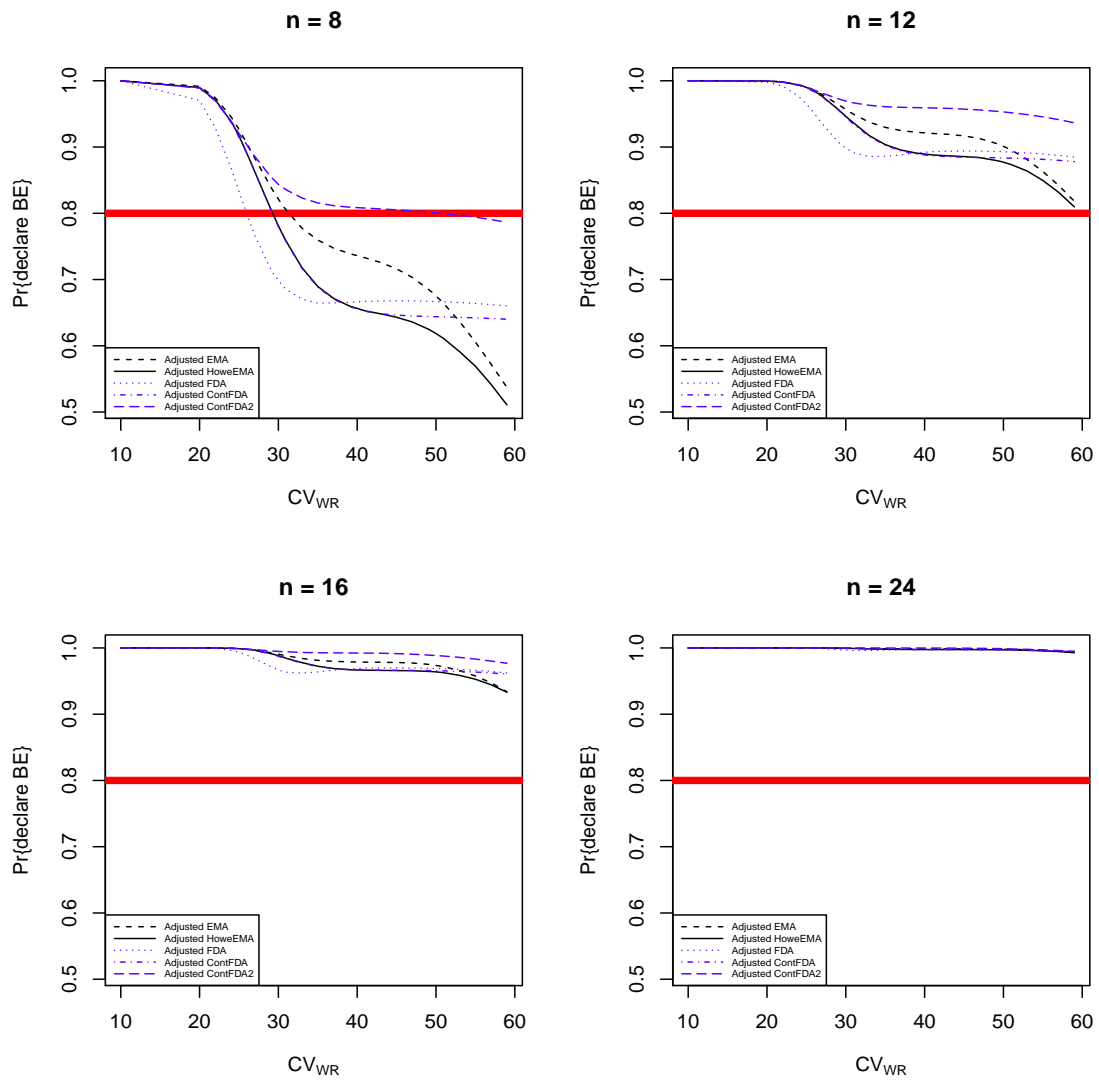


Figura 4.2: Potencia de los métodos escalados ajustados de bioequivalencia cuando el efecto formulación verdadera es nulo. Como referencia, también se representa una línea de horizontal del 80 %.

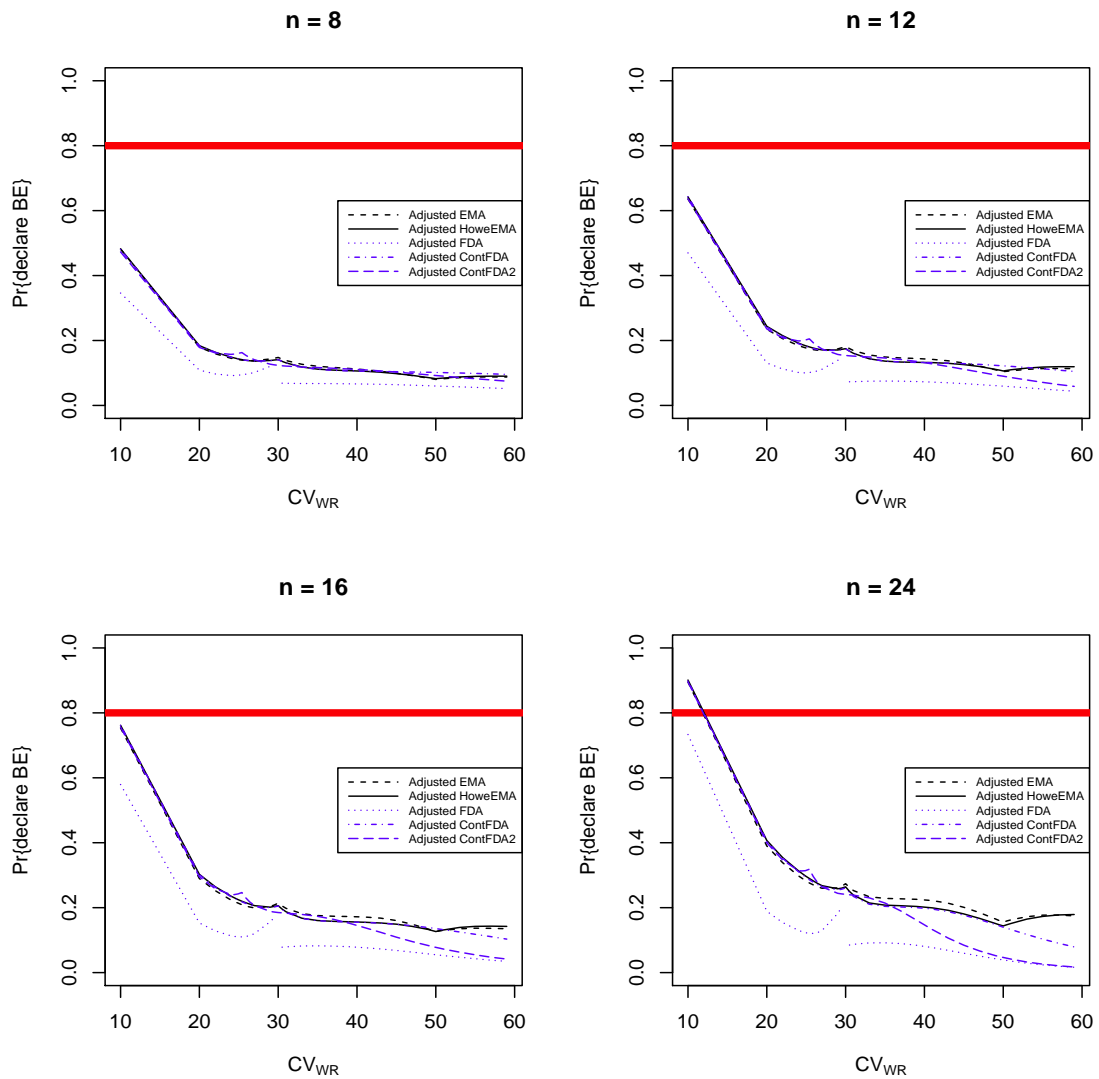


Figura 4.3: Potencia de los métodos escalados ajustados de bioequivalencia cuando el efecto formulación verdadera está dentro de la región de bioequivalencia pero cerca (una fracción $\lambda = 0.8$) de la bioequivalencia límite de EMA y enfoque escalado de la FDA. Como referencia, un 80 % también está representado en la figura.

	Method	auc_{pow}		% power loss in terms of auc_{pow}		Max. power loss and CV_{WR} where it is reached	
		Unadjusted	Adjusted	Absolute	Percentage	CV_{WR}	Percentage
$n = 8$	EMA	41.7	39.4	2.3	5.5%	59	14.8%
	HoweEMA	39.7	37.5	2.2	5.5%	59	13.5%
	FDA	43.9	37.5	6.4	14.6%	35	21.8%
	ContFDA	40.3	38.2	2.1	5.2%	43	9.1%
	ContFDA2	44.1	42.5	1.6	3.6%	39	5.4%
$n = 12$	EMA	47	46.1	0.9	1.9%	59	5.8%
	HoweEMA	46.3	45.4	0.9	1.9%	59	5.0%
	FDA	47.9	45.4	2.5	5.2%	33	8.4%
	ContFDA	46.5	45.7	0.8	1.7%	43	3.0%
	ContFDA2	47.9	47.6	0.3	0.6%	39	1.3%
$n = 16$	EMA	48.5	48.2	0.3	0.6%	59	2.4%
	HoweEMA	48.3	47.9	0.4	0.8%	59	1.9%
	FDA	48.7	47.9	0.8	1.6%	33	3.1%
	ContFDA	48.3	48	0.3	0.6%	43	0.9%
	ContFDA2	48.7	48.7	0	0.0%	37	0.3%
$n = 24$	EMA	49	48.9	0.1	0.2%	59	0.2%
	HoweEMA	48.9	48.9	0	0.0%	59	0.2%
	FDA	49	48.9	0.1	0.2%	31	0.4%
	ContFDA	48.9	48.9	0	0.0%	39	0.1%
	ContFDA2	49	49	0	0.0%	29	0.0%

Tabla 4.1: Pérdida de potencia para los niveles de significancia ajustados, como una medida de potencia global sobre el rango de valores de CV_{WR} se utilizó el área bajo la curva de la potencia vs CV_{WR} , auc_{pow} .

A continuación se presentan los resultados de las simulaciones considerando los métodos: EMA, HoweEMA, LO, HoweLO y considerando además la corrección del sesgo presentada en el capítulo 3, cuyo método se define como: bcHoweLO.

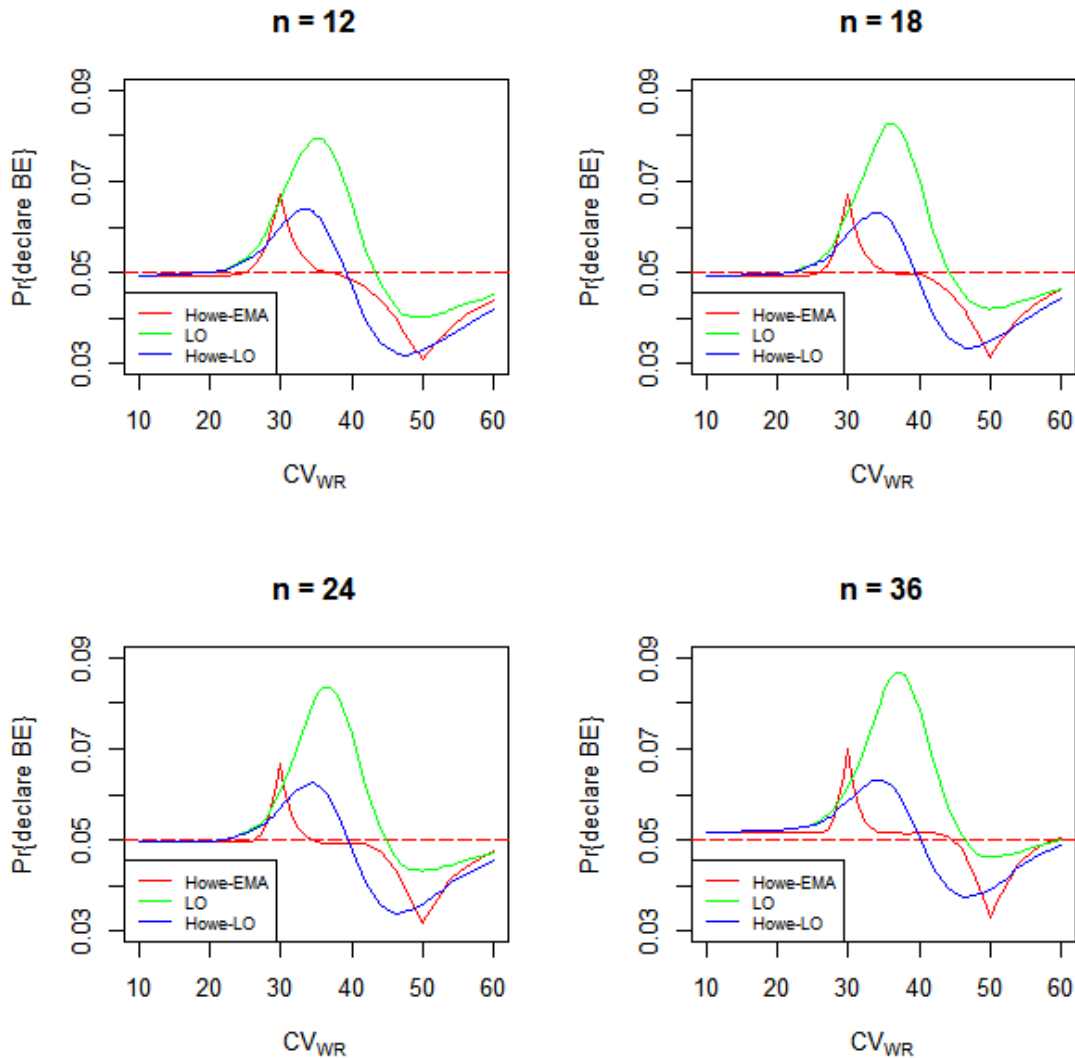


Figura 4.4: Probabilidad de declarar BE, para el método HoweEMA, LO y Howe-LO, para el método Howe-LO no se aplica ninguna corrección al método de Howe.

Como se puede apreciar, la probabilidad de error de tipo I sigue teniendo un comportamiento algo inadecuado. El método LO presenta una gráfica suave pero controla pobremente el error de tipo I. El método Howe-LO presenta una gráfica más suave y mejora algo la probabilidad máxima de error de tipo I, pero no se puede decir que domine al método Howe-EMA en este aspecto, aunque es manifiestamente mejor que LO.

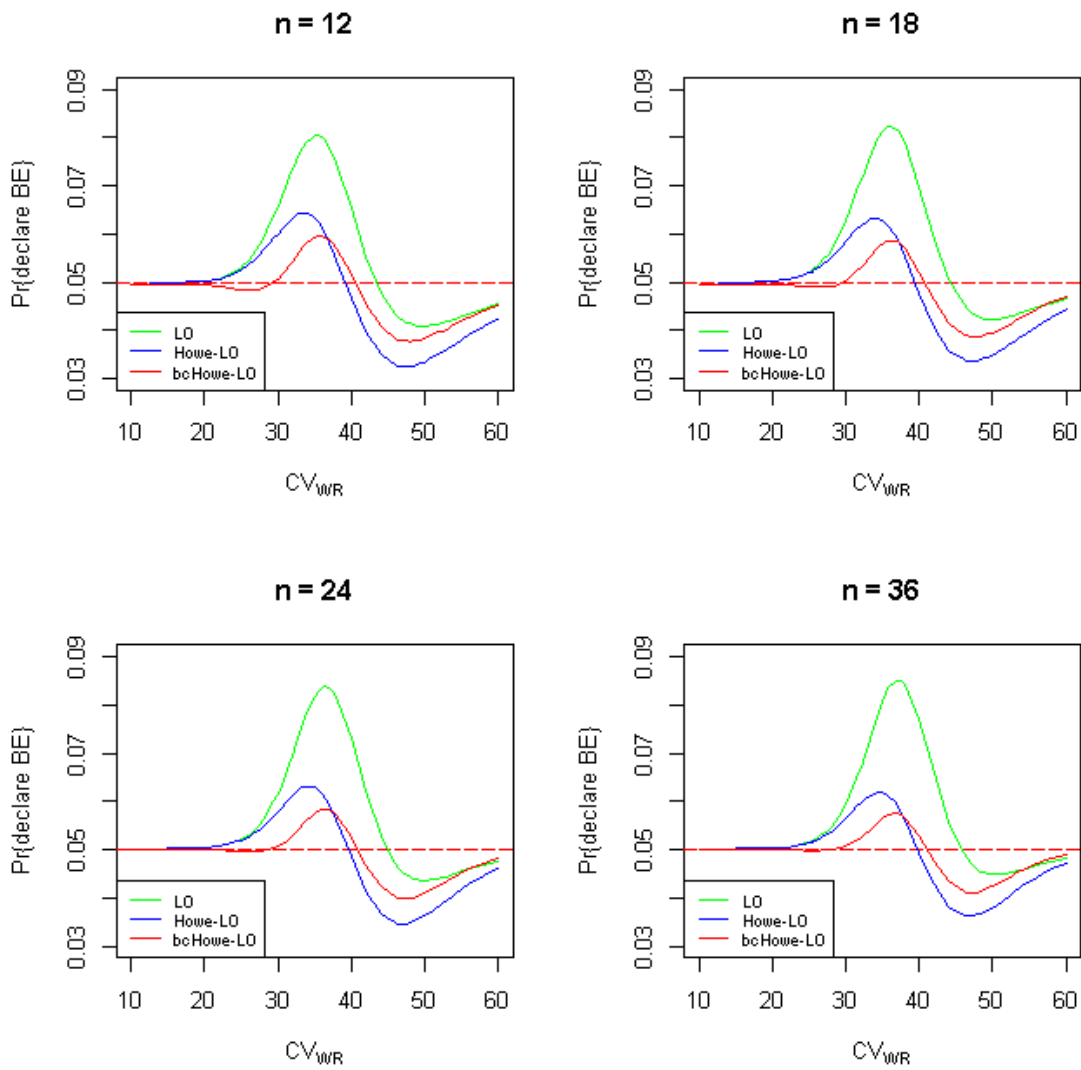


Figura 4.5: Probabilidad de declarar BE, para el método Howe-LO , LO y bcHoweLO.

Como se puede observar en este caso el método que mejor controla el error de tipo I es bcHoweLO.

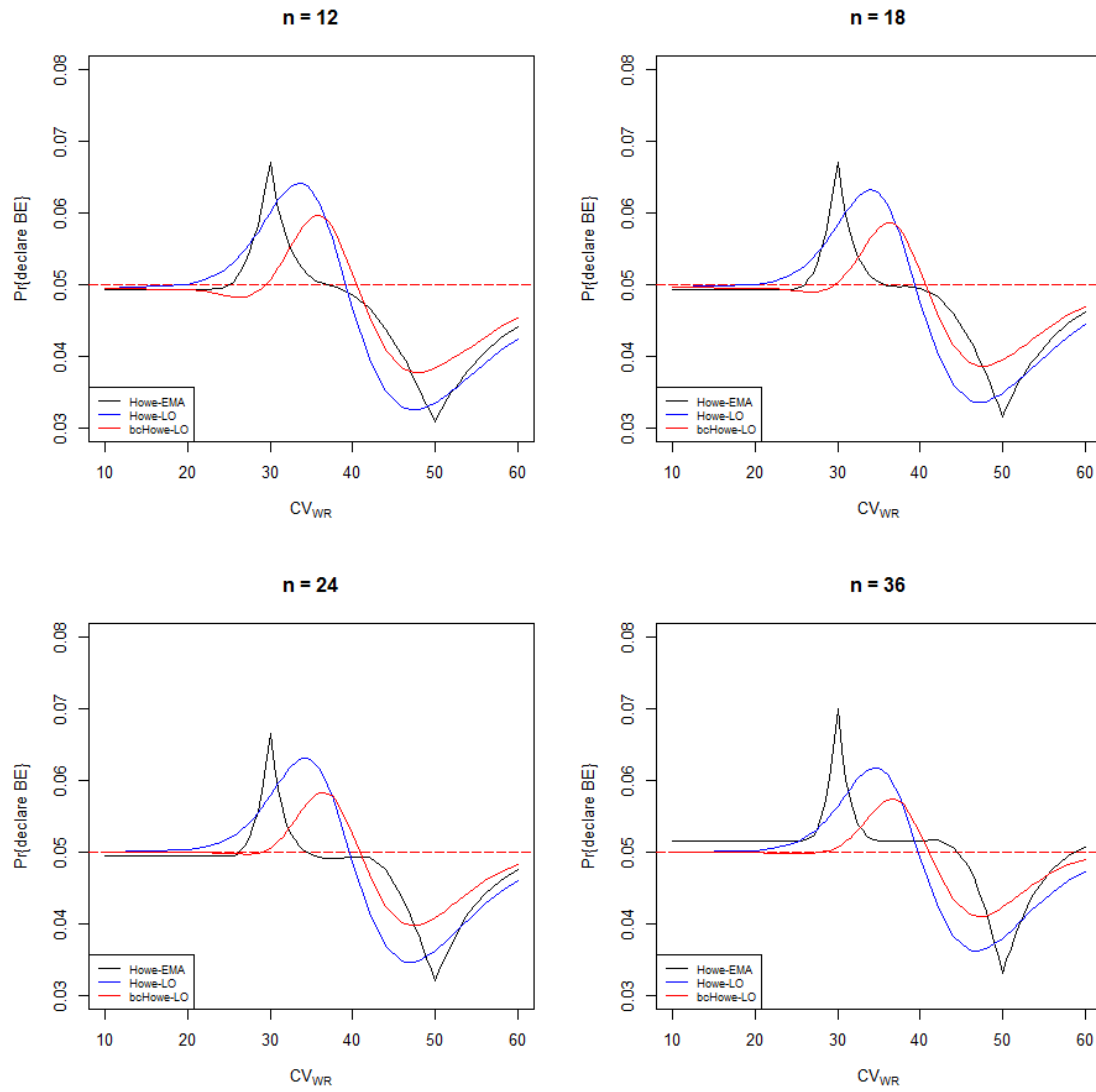


Figura 4.6: Probabilidad de declarar BE, para el método HoweEMA, Howe-LO , LO y bcHoweLO.

En este caso se concluye que el método corregido es el que mejor controla el error de tipo I, coincidiendo con la conclusión de la figura 4.5.

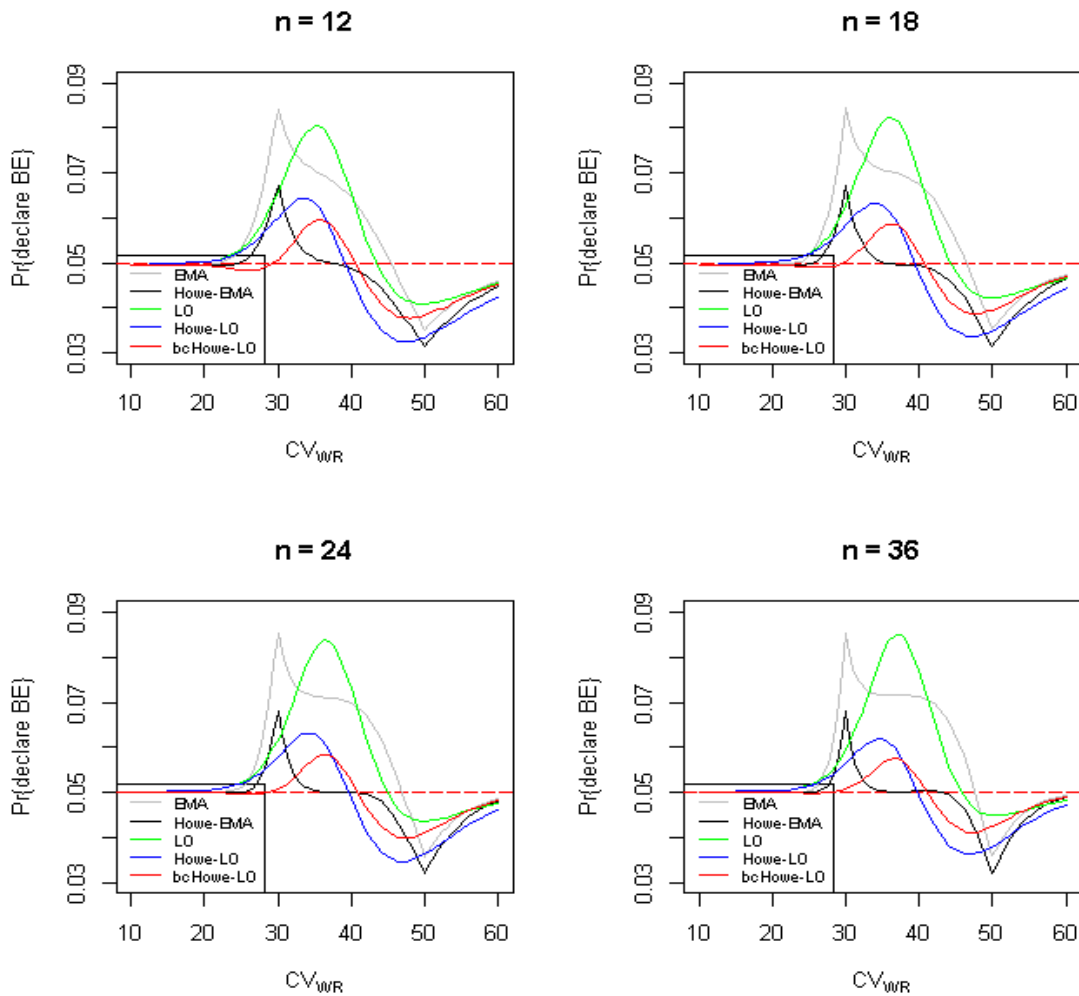


Figura 4.7: Probabilidad de declarar BE, para el método HoweEMA, Howe-LO , LO , bcHoweLO y EMA.

La siguiente figura 4.7 ilustra el error de tipo I asociado al método RSABE propuesto por la EMA en 2010 (EMA) y el método basado en los límites "Leveling-off" (LO), ambos basados en el principio de inclusión de intervalos (a pesar de ser los límites aleatorios), la hipotética mejora de ambos enfoques basada en el método de Howe (Howe-EMA y Howe-LO, respectivamente), y finalmente, el método Howe-LO corregido para el sesgo en el sentido descrito anteriormente (bcHowe-LO).

Como se puede apreciar, el método bcHowe-LO consigue el mejor control del error de tipo I, aunque persiste cierta oscilación en el mismo. Queda como una cuestión pendiente el desarrollo de un método de ajuste del nivel de significación para estos métodos "LO" basados en límites regulares. En principio, especialmente para el método que presenta menor inflación de la probabilidad de error de tipo I, "bcHoweLO", ajustar el nivel de significación conduciría a un test totalmente válido y con el menor grado de pérdida de potencia asociada a dicho ajuste.

4.1.2. Análisis de datos reales

Los métodos estudiados en esta memoria se aplican al conjunto de datos reales procedentes de un diseño crossover parcialmente replicado de 51 sujetos sanos, asignados a cada una de las tres secuencias de tratamientos TRR,RTR,RRT, en forma balanceada y aleatoria, es decir, 17 sujetos en cada secuencia extraído del artículo [26]. En las tablas 4.2 y 4.3 están resumidos los resultados de los métodos aplicados con y sin ajuste del nivel de significación para los parámetros farmacocinéticos C_{max} y AUC del conjunto de datos.

En la tabla 4.2, se presentan los resultados para el parámetro $\log(C_{max})$, para ambas regulaciones EMA y FDA, considerando la posibilidad de escalamiento de la variable.

La estimación del efecto formulación es $\hat{\phi} = 0.32$ y la desviación estándar de la formulación de referencia intra-sujeto estimada es $\hat{\sigma}_{WR} = 0.57$ en escala logarítmica, o en escala original $\widehat{CV}_{WR} = 61.96\%$ cuando estos parámetros son estimados a partir de la diferencia de contraste intrasujeto, el cual es usado en los métodos HoweEMA, FDA, ContFDA y ContFDA2. Estas estimaciones difieren ligeramente de acuerdo al documento Q&A(2015)[9], en particular, el método A de dicho documento de donde se obtiene $\hat{\sigma}_{WR} = 0.56$ y $\widehat{CV}_{WR} = 61.22\%$; estos resultados se aplican en el método EMA.

Con el fin de comparar, lo primero que se aplica son los métodos sin considerar ningún ajuste del nivel de significancia. Si se considera el método A de EMA 2015, se tiene que $\widehat{CV}_{WR} = 61.22\%$ es mayor que el 50% (o $0.56 > 0.472$ en escala logarítmica), de acuerdo con el criterio EMA, los límites de bioequivalencia son fijos cuyos valores son ± 0.359 . Por lo tanto, el método A de la EMA indica que el intervalo del efecto formulación al 90% de confianza es $CI_{\phi} = [0.1647, 0.4681]$. Este intervalo está fuera de los límites de BE, por lo que no debe declararse Bioequivalencia.

Para el método HoweEMA, se tiene que $\widehat{CV}_{WR} = 61.96\%$ es mayor a 50% lo cual implica aplicar los límites de bioequivalencia ± 0.359 . Luego el intervalo de confianza para ϕ es $CI_{\phi} = [0.1711, 0.4617]$, por lo que nuevamente no es posible declarar bioequivalencia, pues el intervalo de confianza CI_{ϕ} no está totalmente incluido en ± 0.359 .

El enfoque FDA, no considera la posibilidad de estabilización de los límites de BE, por lo que los límites escalados para los métodos FDA y contFDA2 son $\pm k_{FDA} 0.57 = \pm 0.5087$, sin embargo, recordando el método FDA estos límites escalados se traducen en un criterio de declarar bioequivalencia si la estimación del límite superior del intervalo de confianza del 95% unilateral $\hat{\phi}_U$ del parámetro compuesto $\theta = \phi^2 - k_{FDA}^2 \sigma_{WR}^2$ es negativo. En este caso $\hat{\theta}_U = -0.027 < 0$, indicaría inicialmente que ambos enfoques FDA y ContFDA2 se declara bioequivalencia. No obstante, la estimación puntual del efecto formulación $\hat{\phi} = 0.32$ está fuera de los límites ± 0.223 . Luego aplicando el requisito de restricción sobre la estimación puntual, se concluye que no es posible declarar bioequivalencia.

Por otro lado, para el método ContFDA siempre que la constante sea la misma que el método EMA, se tiene que $\theta_U = 0.0358 > 0$. Esto indica que no es posible declarar BE, incluso antes de considerar la restricción sobre la estimación puntual.

En la tabla 4.2, se considera además el nivel de significancia de estos métodos mencionados, pues en ellos no se controlan del todo el error de tipo I. Como se mencionó anteriormente, el grado del ajuste depende del tamaño muestra, del diseño crossover y del método RSABE.

Para este conjunto de datos se considera un diseño balanceado TRR/RTR/RRT con 17 sujetos en cada secuencia, con el fin de mantenerla probabilidad del error de tipo I en el valor nominal 0.05, los niveles de significancia ajustados, son para el método EMA: 0.0341 (ajuste realizada mediante la función `scABEL.ad`). Para los métodos HoweEMA y `contFDA` son 0.0381 y 0.0368 respectivamente, y para `contFDA2` es 0.0368.

Para el método FDA también puede ser calculado el ajuste del nivel de significancia del cual se obtiene una considerable corrección de 0.011.

Por otra parte, independiente del nivel de significancia, recordemos que para $\log(C_{max})$ el efecto formulación estimado está fuera de los límites ± 0.223 , por lo tanto, en todos los métodos no sería posible declarar bioequivalencia. Esto implicaría que no es necesario aplicar los niveles de significancia ajustados, pero se presentan los resultados con fines informativos. Así los intervalos de confianza modificados que serían obtenidos con estos niveles de significancia son:

En el método EMA, se tiene un nivel de confianza del 93.18 %, es decir, $1 - 2 * 0.0341 = 0.9318$ donde el intervalo de confianza que se obtiene es $[0.148, 0.485]$.

Para el método HoweEMA el intervalo de confianza del 92.38 % es $[0.159, 0.473]$. Para el método FDA el límite superior estimado del intervalo de confianza unilateral de θ , es $\theta_U = 0.034$ al nivel de confianza del 98.87 % (es decir $1 - 0.011$).

Para el método `contFDA` el límite superior $\theta_U = 0.047$ al nivel de confianza del 96.19 % (es decir, $1 - 0.0381$).

Finalmente, para el `contFDA2` el límite superior es $\hat{\theta}_U = -0.0135$ al nivel de confianza del 96.32 %, notar que éste sería el único escenario posible de declarar bioequivalencia, en ausencia, de la restricción sobre la estimación puntual.

Los resultados para $\log(AUC)$ se resumen en la tabla 4.3, notar que en este caso la restricción sobre la estimación puntual se cumple para este parámetro farmacocinético, ya que la estimación puntual del efecto formulación es $\hat{\phi} = 0.056$ el cual está dentro de los límites ± 0.223 .

La estimación de la desviación estándar de la formulación de referencia $\hat{\sigma}_{WR} = 0.345$. Si no se ajusta el nivel de significancia entonces de acuerdo al método FDA el límite superior del intervalo de confianza unilateral del 95 % construido con el método de Howe es $\hat{\theta}_U = -0.0603 < 0$, es decir, la bioequivalencia es declarada, este límite superior y la conclusión también son aplicables al método `contFDA`.

Para el método `contFDA`, el límite superior estimado del intervalo de confianza de θ , es $\theta_U = -0.0393 < 0$, por lo tanto, también se declara BE.

Recordando que el criterio regulatorio EMA, del parámetro farmacocinético $\log(AUC)$ no puede ser considerado para escalamiento, pero aplicando el método estándar con límites constantes de ± 0.223 y sin ajustar el nivel de significancia de 0.05, el método A de Q&A(2015) el intervalo de confianza del efecto formulación del 90 % es $CI_{\phi} = [-0.0405, 0.1520]$ lo que implica que también es declarada la BE.

Ahora si se considera el ajuste del nivel de significancia, se tiene que para el método FDA el límite superior estimado por el método de Howe al nivel de confianza de 98.87 % es $\theta_U = -0.0461 < 0$, es decir, se declara bioequivalencia.

Para los métodos `ContFDA` y `ContFDA2`, los límites superiores estimados del in-

tervalo de confianza unilateral son -0.0368 y -0.0573, respectivamente. Luego en ambos métodos se concluye BE.

Method	$\alpha = 0.05$ (fixed)				Adjusted significance level: α_{adj}				
	CI_ϕ	$\hat{\theta}_U$	Criterion	Decision	α_{adj}	CI_ϕ	$\hat{\theta}_U$	Criterion	Decision
EMA	[0.1647, 0.4681]	-	$CI_\phi \not\subseteq \pm 0.359$	no BE	0.0341	[0.1479, 0.4848]	-	$CI_\phi \not\subseteq \pm 0.359$	no BE
HoweEMA	[0.1711, 0.4617]	-	$CI_\phi \not\subseteq \pm 0.359$	no BE	0.0381	[0.1594, 0.4734]	-	$CI_\phi \not\subseteq \pm 0.359$	no BE
FDA	-	-0.0223	$\hat{\theta}_U < 0$	BE(if no PEC)	0.0113	-	0.0339	$\hat{\theta}_U > 0$	no BE
ContFDA	-	0.0358	$\hat{\theta}_U > 0$	no BE	0.0381	-	0.0471	$\hat{\theta}_U > 0$	no BE
ContFDA2	-	-0.0267	$\hat{\theta}_U < 0$	BE (if no PEC)	0.0368	-	-0.0135	$\hat{\theta}_U < 0$	BE (if no PEC)

Tabla 4.2: Resultados de la aplicación de los enfoques RSABE para el parámetro farmacocinético $\log(Cmax)$. el CI_ϕ estándar intervalo de confianza shortest para el efecto formulación, θ_U límite superior del intervalo de confianza unilateral para el parámetro $\theta = \phi^2 - k^2\sigma_{WR}$ (donde k corresponde a la constante escalada de EMA o FDA), BE(bioequivalencia) y PEC(restricción de estimación puntual).

Method	$\alpha = 0.05$ (fixed)				Adjusted significance level: α_{adj}				
	CI_ϕ	$\hat{\theta}_U$	Criterion	Decision	α_{adj}	CI_ϕ	$\hat{\theta}_U$	Criterion	Decision
FDA	-	-0.0603	$\hat{\theta}_U < 0$	BE	0.0057	-	-0.0461	$\hat{\theta}_U < 0$	BE
ContFDA	-	-0.0603	$\hat{\theta}_U < 0$	BE	0.0191	-	-0.0368	$\hat{\theta}_U < 0$	BE
ContFDA2	-	-0.0393	$\hat{\theta}_U < 0$	BE	0.0184	-	-0.0573	$\hat{\theta}_U < 0$	BE

Tabla 4.3: Resultados de la aplicación de los enfoques RSABE para el parámetro farmacocinético $\log(AUC)$.

4.2. Discusión

- Para evaluar la bioequivalencia en fármacos de alta variabilidad, se deben escalar los límites de bioequivalencia en función de la variabilidad. Este enfoque ha sido recomendado por las agencias del medicamento presentadas en esta memoria y ha sido ampliamente estudiado en la literatura, pero todos estos métodos sufren en general de no controlar el error de tipo I, es decir, un mal control en cercanías de las variabilidades intrasujeto de la formulación de referencia donde los límites de BE cambian bruscamente de constante a escalado. Hasta cierto punto, esto es preocupante ya que estos valores de variabilidad parecen ser bastante frecuentes en estudios de bioequivalencia real para HVD [6].
- En esta memoria se proponen mejoras para los métodos RSABE de las regulaciones FDA y EMA (2010), ya que en estos métodos propuestos en [23] (capítulo 3) controlan de una forma mejor el error de tipo I. Estos enfoques se aproximan de mejor manera al nivel de significancia nominal de 0.05, pero a pesar de esto no controlan del todo bien el error de tipo I.
- Como todavía persiste algún grado de error de tipo I, se propone el enfoque de ajuste del nivel de significancia que se basa en el hecho de que la verdadera curva de probabilidad de error de tipo I de estos métodos cruza la línea de nivel de significación nominal 0.05. De la figura 4.1 se observa que todos los métodos tienen un máximo absoluto precisamente en el umbral de alta variabilidad, es decir, un $CV_{WR} = 30\%$ pero no sobrepasan el nivel de significancia prefijado en 0.05.
- Ajustar el nivel de significación puede ser una manera de hacer el enfoque escalado estadísticamente correcto, probado en todos los métodos escalados, incluso con las correcciones sugeridas en el artículo [23]. Por otra parte, queda como futuro trabajo ver cómo se comporta el ajuste del nivel de significación, para el método bcHoweLO, pues al considerar límites suaves y aplicar el método de Howe sobre éstos y además corregir el sesgo que presentan los estadísticos de dicho método, mejora aún más el control del error de tipo I.
- Para el ajuste del método EMA 2010, una adecuada opción es aplicar el método implementado en la función `scABEL.ad` del package `PowerTOST`, ya que el método de estimación de parámetro está basada en ajuste ANOVA que es el método que se plantea a utilizar en el documento Q&R 2015 por la EMA.
- Los resultados no son totalmente correctos para variabilidades dentro del sujeto R y T desiguales, es decir, bajo heterocedasticidad de las varianzas intrasujeto, importante cuestión que sigue abierta y merece estudio posterior.
- Una forma posible de solución puede consistir en hacer que el ajuste del nivel de significación dependa de las estimaciones de estas variabilidades dentro de los sujetos para ambas formulaciones.
- También en esta memoria se propone considerar límites de BE más regulares, aplicarles el método Howe y corregir el sesgo. Como se puede observar en los resultados de las simulaciones anteriores éstos controlan mucho mejor el error de tipo I.

- La segunda derivada de estos límites más regulares $\phi_{LO}(\sigma_{WR})$ influyen en el comportamiento de la probabilidad de declarar BE, esto se puede apreciar en la figura 3.2 donde siguen de una forma considerablemente fiel a las oscilaciones de la segunda derivada para el método de Howe-LO en el diseño parcialmente replicado TRR/RTR/RRT.

Capítulo 5

Conclusiones

- Considerar las objeciones de cada método RSABE planteadas y proponer mejoras sobre ellas se obtienen resultados cada vez más precisos como por ejemplo: considerar límites constantes y no aleatorios en el CIIP, proponer límites BE continuos y otra posibilidad es utilizar límites de BE regulares.
- Al aplicar los métodos RSABE ajustando el nivel de significancia nominal y posteriormente realizar estudios sobre la pérdida global de potencia entre los métodos ajustados y sin ajustar, se produce una pérdida de potencia global pero ésta pérdida convierte los test potencialmente inválidos en test que preservan el error de tipo I.
- Entre los métodos propuestos: “ContFDA” , “ContFDA2” y “HoweEMA” , mejoran los resultados al considerar el método de ajuste del nivel de significancia ya que son los que menor pérdida de potencia presentan. Queda como futuro trabajo ver qué se obtendría con el método “bcHoweLO” , pues como se puede ver en los resultados de las simulaciones anteriores sin ajustar el nivel de significancia presentan un mejor control del error de tipo I.
- Independientemente del método que se está aplicando, el ajuste del nivel de significancia es una manera adecuada de tener pruebas RSABE estadísticamente correctas, con respecto al error de tipo I, y en general, con una penalización aceptable en la potencia.
- Todos los resultados anteriores también son válidos en escenarios desbalanceados, es decir, para los diseños crossover con tamaños de muestra por secuencia desiguales.
- El ajuste del nivel de significación puede proporcionar métodos RSABE estadísticamente correctos, pero no garantiza su validez clínica.

Bibliografía

- [1] Abad F, Martínez E, and Galvez MA. Las distintas fases del ensayo clínico.
- [2] Ocaña J and Sanchez M and Carrasco J. Carryover negligibility and relevance in bioequivalence studies. *Pharmaceutical Statistics*, 14:400–408, 2015.
- [3] Berger R and Hsu J. Bioequivalence trials, Intersection-Union Tests and Equivalence Confidence Sets. *Statistical Science*, 11(4):283–319, 1996.
- [4] Liu JP. Chow SC. *Design and Analysis of Bioavailability and Bioequivalence Studies*. Chapman & Hall/CRC Press: Boca Raton, 2009.
- [5] Alcaide D. Métodos de bioequivalencia escalada. Estudio de sus propiedades inferenciales. Master’s thesis, 2012.
- [6] Davit BM, Conner DP, Fabian-Fritsch B, Haidar SH, Jiang X, Patel DT, Seo PRH, Suh K, Thompson CL, and Yu LX. Highly variable drugs: Observations from bioequivalence data submitted to the FDA for new generic drug applications. *The AAPS Journal*, 10(1):148–156, 2008.
- [7] Karl E. Peace. Din Chen. *Clinical Trial Data Analysis - Using R*. 2010.
- [8] EMA. Guideline on the Investigation of Bioequivalence. European Medicines Agency, CPMP/EWP/QWP/1401/98 Rev. 1/Corr. London., 2010.
- [9] EMA. Questions & answers: positions on specific questions addressed to the pharmacokinetics working party (pkwp). European Medicines Agency, EMA/618604/2008 Rev. 12. London., 2015.
- [10] FDA. Food and drug administration. guidance for industry: Statistical approaches to establishing bioequivalence. CDER, Department of Health and Human Services. Rockville, MD., 2001.
- [11] FDA. Draft guidance on progesterone. The United States Food and Drug Administration: Rockville, Maryland., 2010.
- [12] FDA. Food and drugs. 2016.
- [13] Haidar SH, Makhoul F, Schuirmann DJ, Hyslop T, Davit BM, Conner DP, and Yu LX. Evaluation of a Scaling Approach for the Bioequivalence of Highly Variable Drugs. *The AAPS Journal*, 10(3):450–454, 2008.

- [14] Howe WG. Approximate confidence limits on the mean of $X + Y$ where X and Y are two tabled independent random variables. *Journal of the American Statistical Association*, 69(347):789–794, 1974.
- [15] Hyslop T, Hsuan F, and Holder DJ. A small sample confidence interval approach to assess individual bioequivalence. *Statistics in Medicine*, 19(20):2885–2897, 2000.
- [16] Karalis V, Symillides M, and Macheras P. On the leveling-off properties of the new bioequivalence limits for highly variable drugs of the EMA guideline. *European Journal of Pharmaceutical Sciences*, 44(4):497–505, 2011.
- [17] Karalis V, Symillides M, and Macheras P. Bioequivalence of Highly Variable Drugs: A Comparison of the Newly Proposed Regulatory Approaches by FDA and EMA. *Pharmaceutical Research*, 29(4):1066–1077, 2012.
- [18] R. O. Kuehl. *Diseño de experimentos Principios estadísticos de diseño y análisis de investigación*. 2000.
- [19] Kytariolos J, Karalis V, Macheras P, and Symillides M. Novel scaled bioequivalence limits with leveling-off properties. *Pharmaceutical Research*, 23(11):2657–2664., 2006.
- [20] Labes D and Schütz H. Inflation of Type I Error in the Evaluation of Scaled Average Bioequivalence, and a Method for its Control. *Pharmaceutical Research*, 35 (12):1933–1943, 2016.
- [21] Midha KK, Shah VP, Singh GJP, and Patnaik R. Conference report: Bio-international 2005. In *Journal of Pharmaceutical Sciences*, volume 96 (4), pages 747–754, 2007.
- [22] Muñoz J. Bioequivalencia escalada basada en límites suaves y remuestreo. Master’s thesis, UPC-UB, 2014.
- [23] Muñoz J, Alcaide D, and Ocaña J. Consumer’s risk in the EMA and FDA regulatory approaches for bioequivalence in highly variable drugs. *Statistics in Medicine*, 35 (12):1933–1943, 2016.
- [24] Munk A and Pflugger R. 1- α Equivariant Confidence Rules for Convex Alternatives are $\alpha/2$ -Level Tests-With Applications to the Multivariate Assessment of Bioequivalence. *Journal of the American Statistical Association*, 94(448):1311–1319, 1999.
- [25] Ocaña J, Sánchez M and Sánchez A, and Carrasco J. On equivalence and bioequivalence testing. *Sort*, 32(2):151–176, 2008.
- [26] Patterson SD and Jones B. Viewpoint: observations on scaled average bioequivalence. *Pharmaceutical Statistics*, 11(1):1–7, 2012.
- [27] Schuirmann D. A comparison of the one sided test procedure and the power approach for assessing the equivalence of average bioavailability. *Pharmacometrics and Biopharmaceutics*, 15(6), 1987.
- [28] Tothfalusi L, Endrenyi L, and Garcia Arieta, A. Evaluation of bioequivalence for highly variable drugs with scaled average bioequivalence. *Clinical Pharmacokinetics*, 48(11):725–743, 2009.

- [29] Tothfalusi L, Endrenyi L, Midha K, Rawson M, and Hubbard J. Evaluation of the bioequivalence of highly-variable drugs and drug products. *Pharmaceutical Research*, 18(6):728–733, 2001.

Apéndice A

Anexos

A.1. Packages simcrossover y adjustalpha

A.1.1. Funciones simCrossover y adjAlpha

Para instalar los packages se puede hacer con los siguientes códigos:

```
install.packages("E:/simcrossover_0.1.0.zip", repos = NULL, type = "win.binary")
install.packages("E:/adjustalpha_0.2.0.zip", repos = NULL, type = "win.binary")
```

Para cargar packages necesarios:

```
library(pryr)
library(PowerTOST)
library(simcrossover)
library(adjustalpha)
```

Descripción de las funciones y ejemplos:

```
>?simCrossover
>?adjAlpha
```


Descripción función `simCrossover`

Generación de datos de un diseño crossover (como objetos `data.frame`)

Descripción

Función que genera “`nsim`” `data.frames` según el diseño crossover y los valores del modelo lineal que se le indique.

Uso

```
simCrossover(nsim = 1, design = matrix(c("R", "T", "T", "R"), nrow = 2,
  byrow = TRUE), ssize = 1, mu = 0, periodEff = NULL, treatEff = NULL,
  carryEff = NULL, seqEff = NULL, sigmaSubj = 0, sigma = 0,
  simplify = TRUE, as.factors = FALSE, as.ordereds = FALSE)
```

Significado de los argumentos:

<code>nsim</code>	Número de réplicas de simulación (número de <code>data.frames</code> o “bases de datos” que va a generar)
<code>design</code>	Matriz que define el diseño crossover, por defecto es el diseño 2x2 RT/TR
<code>ssize</code>	Tamaño muestral en cada secuencia. Si “ <code>ssize</code> ” es un escalar, se generarán bases de datos balanceadas, con “ <code>ssize</code> ” sujetos en cada secuencia, en total $N = \text{ssize} \times (\text{número de secuencias})$
<code>mu</code>	Media general en el modelo lineal, por defecto cero.
<code>periodEff</code>	Efecto período en el modelo lineal. Tiene que ser un vector de longitud igual al número de períodos en el diseño, excepto si solamente hay dos períodos; en este caso si “ <code>periodEff</code> ” es escalar, se hace <code>periodEff = c(-periodEff, periodEff)</code> . Por defecto es un vector de ceros (ausencia de efecto período)
<code>treatEff</code>	Efecto de tratamiento. Tiene que ser un vector de longitud igual al número de tratamientos posibles (2 en la mayoría de aplicaciones de diseños crossover). Puede ser un escalar si sólo hay 2 tratamientos, entonces <code>treatEff = c(-treatEff, treatEff)</code> . Por defecto es un vector de ceros (ausencia de efecto tratamiento)
<code>carryEff</code>	Efecto carryover de primer orden. Tiene que ser un vector de la misma longitud que el número de tratamientos posibles. Codifica el efecto carryover o residual de haber administrado en el período anterior determinado tratamiento. Al vector, automáticamente, se le añade por la izquierda el valor 0, que corresponde al efecto carryover en el primer período. Si solamente hay dos tratamientos, puede ser un escalar; en este caso se convierte en <code>c(-carryEff, carryEff)</code> (y finalmente en <code>c(0, -carryEff, carryEff)</code>). Vector de ceros por defecto (ausencia de carryover).
<code>seqEff</code>	Efecto secuencia. Comentarios similares al efecto período, para las secuencias.
<code>sigmaSubj</code>	Desviación típica intersujetos. Tiene que ser un escalar. Cero por defecto. El efecto aleatorio de sujeto se genera según una distribución normal $N(0, \text{sigmaSubj})$.
<code>sigma</code>	Desviación típica residual. Puede ser un escalar o bien un vector con la misma longitud que el número de tratamientos. En este caso se simula distinta variabilidad residual para cada tratamiento (e.g. <code>sigmaR</code> , <code>sigmaT</code>). El residuo se genera según una distribución normal de media cero y la desviación estándar indicada.
<code>simplify</code>	Valor booleano, por defecto TRUE. Ver “Comentarios”.
<code>as.factors</code>	Valor booleano, por defecto FALSE. Ver “Comentarios”.
<code>as.ordereds</code>	Valor booleano, por defecto FALSE. Ver “Comentarios”.

Comentarios:

- Si `nsim > 1`, se genera una lista de longitud “`nsim`” en la que cada elemento es un “`data.frame`”.
Si `nsim == 1` entonces se genera un `data.frame`, no una lista formada por un solo “`data.frame`”, excepto si se indica `simplify = FALSE`.
- El argumento “`as.factors`” (por defecto a `FALSE`), estas columnas se codifican internamente como números enteros, no como valores de clase “`factor`”.
- El argumento “`as.ordereds`” si se establece en `TRUE` (por defecto a `FALSE`), estas columnas se codifican internamente como valores de clase “`ordenados`”.

Descripción función `adjAlpha`

Ajuste del nivel de significancia en un test de bioequivalencia promedio escalada de referencia (RSABE) basado en datos de un diseño crossover con el objetivo de controlar su error de tipo I.

Descripción

Ajustar el nivel de significación en una prueba bioequivalencia promedio escalada de referencia (RSABE) basada en datos de diseño de crossover con el fin de controlar su error de tipo I.

Uso

```
adjAlpha(n, nominalAlphas = 0.05, method = "EMA", design = "TRR-RTR-RRT",
  constraint = TRUE, approximate = FALSE, minAlpha = 0.005,
  deltaAlpha = 0.5 * minAlpha, steps = 100, precis = 1e-04,
  nsim = 1e+06, extraInfo = FALSE, trace = TRUE, seed = 15727)
```

Arguments

Significado de los argumentos:

<code>n</code>	Tamaño de muestra. Un vector numerico especificando el tamaño de muestra para cada secuencia en un diseño crossover o, alternativamente, un escalar especificando un tamaño de muestra balanceado
<code>nominalAlphas</code>	Un vector numérico con los niveles de significancia objetivos, es decir, la verdadera probabilidad del error de tipo I que no se sobrepase
<code>method</code>	String especificando el método RSABE. Valores posibles son: "EMA" (o "EMA2010"), "FDA" (o "FDA2010"), "HoweEMA", "ContFDA", "ContFDA2", "LO", "HoweLO" y "bcHoweLO". "EMA" y "FDA" correspondiente a los metodos regulatorios RSABE emitido por a EMA y la FDA, respectivamente, en el 2010. "HoweEMA", "ContFDA" y "ContFDA2" son variantes mejoradas de estos métodos, descritas en Muñoz et al. <i>Statistics in Medicine</i> , 2016, 35 1933-1943. Los métodos restantes "LO" corresponden a diferentes variantes del método RSABE introducidos en Karalis et al. <i>European Journal of Pharmaceutical Sciences</i> 2011, 44 497-505, todavía en estudio y sin método iterativo exacto disponible, por el momento.
<code>design</code>	String especificando el diseño crossover. Posibles valores son: "TRR-RTR-RRT" el diseño parcialmente replicado y el diseño replicado "TRTR-RTRT", o cualquier posible sinónimos: <code>executando matchedDesigns()</code> .
<code>constraint</code>	Booleano, activa el punto estimado de restricción, es decir, el requerimiento para declarar BE, es que el efecto formulación estimado esté dentro de los límites ± 0.223 límites (en escala logarítmica) o dentro de los límites 0.8/1.25 (en la escala original).
<code>approximate</code>	Booleano, Si es TRUE el ajuste es más rápido pero sólo aproximado, de lo contrario este valor aproximado se utiliza como el valor inicial para un método de simulación iterado más exacto.
<code>minAlpha</code>	A veces no se puede encontrar el valor inicial de iteración aproximado. entonces, este valor es usado para empezar la iteración.
<code>deltaAlpha</code>	Incremento o decrecimiento inicial en la búsqueda iterativa del nivel de significancia ajustado.
<code>steps</code>	Máximo número de iteraciones.
<code>precis</code>	Máxima diferencia permitida entre <code>nominalAlphas</code> y la verdadera probabilidad de error tipo I simulado.
<code>nsim</code>	Número de réplicas de simulación en el método iterado.
<code>extraInfo</code>	Boolean, muestra información adicional sobre el tamaño de la muestra, el nivel de significación nominal, el método y el diseño como el nombre del resultado.
<code>trace</code>	Boolean, si no es aproximado, obtiene lo pasos del proceso iterativo.
<code>seed</code>	Iniciando semilla aleatoria (<code>set.seed</code>) para el enfoque iterativo de Montecarlo, no se considera si es aproximado igual TRUE.

Observación 18 *Los niveles de significancia ajustados que se utilizarán en la prueba RSABE, el objetivo es de no sobrepasar `nominalAlphas` la probabilidad de error tipo I.*

A.2. Artículo N°1

Consumer's risk in the EMA and FDA regulatory approaches for bioequivalence in highly variable drugsJoel Muñoz,^a Daniel Alcaide^b and Jordi Ocaña^{a,*†}

The 2010 US Food and Drug Administration and European Medicines Agency regulatory approaches to establish bioequivalence in highly variable drugs are both based on linearly scaling the bioequivalence limits, both take a 'scaled average bioequivalence' approach. The present paper corroborates previous work suggesting that none of them adequately controls type I error or consumer's risk, so they result in invalid test procedures in the neighbourhood of a within-subject coefficient of variation of 30% for the reference (*R*) formulation. The problem is particularly serious in the US Food and Drug Administration regulation, but it is also appreciable in the European Medicines Agency one. For the partially replicated TRR/RTR/RRT and the replicated TRTR/RTRT crossover designs, we quantify these type I error problems by means of a simulation study, discuss their possible causes and propose straightforward improvements on both regulatory procedures that improve their type I error control while maintaining an adequate power. Copyright © 2015 John Wiley & Sons, Ltd.

Keywords: scaled average bioequivalence; interval inclusion principle; point estimate constraint

1. Introduction

The main application of the bioequivalence (BE) concept arises when comparing the bioavailabilities (BA) of a test *T* (generic) formulation and a reference *R* (brand) formulation, both containing equal quantity of the same drug or active principle. BA is defined in terms of the concentration of the drug on its site of action, after a single administration of *T* or *R*. For each subject participating in a BE study, BA is usually defined from his or her concentration versus time curve, by means of parameters like the area under the curve (AUC) or the maximum concentration reached, *C*_{max}. Currently, the regulatory agencies like the US Food and Drug Administration (FDA) or the European Medicines Agency (EMA) establish that BE holds when the ratio of the true (but unknown) geometric means of the bioavailabilities of *T* and *R* falls in the interval 0.80 to 1/0.80 = 1.25. This criterion is usually expressed in logarithmic scale (say for variables like $Y = \log(\text{AUC})$ or $\log(\text{Cmax})$) stating that the true formulation effect ϕ , the mean difference between the logarithmic bioavailabilities, should lie inside the limits ± 0.223 , where $0.223 = \log(1.25) = -\log(0.80)$. (ϕ and other parameters appearing in this introductory section will be defined more precisely in the next sections, when dealing with the crossover designs considered in this paper.)

The most commonly used (and recommended in the regulations, e.g. [1]) criterion to statistically test for BE is based on the 'interval inclusion principle', say to declare BE if the 90% standard symmetric confidence interval for ϕ based on the Student's distribution assuming data normality, is fully included in the constant BE limits ± 0.223 . This criterion defines a test of size 0.05 for the contrast of a null hypothesis of bioequivalence versus an alternative of bioequivalence:

$$\begin{aligned} H_0 &: \phi \leq -0.223 \quad \vee \quad \phi \geq 0.223 \\ H_1 &: -0.223 < \phi < 0.223. \end{aligned} \quad (1)$$

^aDepartment of Statistics, Universitat de Barcelona, Av. Diagonal 643, 08028 Barcelona, Spain

^bBioClever S.L. Barcelona, Spain

*Correspondence to: Jordi Ocaña, Department of Statistics, Faculty of Biology, Av. Diagonal 643, 08028 Barcelona, Spain.

†E-mail: jocana@ub.edu

The recommended experimental design to cope with the previous problem is a crossover design. In a well-performed experiment under a crossover design, the power of the procedure based on the interval inclusion principle only depends on the ‘within-subject’ or ‘residual’ variance σ_w^2 (again to be more precisely defined later) and on the true value of the formulation effect ϕ of the corresponding log-bioavailability.

Broadly defined, highly variable drugs (HVD) are characterized by high levels of within-subject variance. The main problem with HVD is the large sample sizes that would be required in order to declare BE, when it really holds, by means of the previous procedure. The determination of BE for HVD has been a controversial issue for many years, debated in many papers and conferences, for example, [2, 3]. In 2010, the EMA [4] issued a regulation for the study of BE in HVD. Also in 2010, the FDA published a draft guidance on progesterone, [5], that seems to have become a *de facto* standard, in view of subsequent applications and discussions, for example, [6, 7]. Although different in some details, both regulations recommended an approach based on linearly scaling the BE limits in function of the estimated ‘within’ standard deviation of the *R* formulation, σ_{WR} (scaled average bioequivalence, SABE) and include a ‘point estimate constraint’, say to declare BE only if the point estimate of the formulation effect is included inside the ± 0.223 limits.

The main goal of the present paper is to study the user’s risk, say the probability of declaring as bioequivalent a truly (according to the underlying, more or less conventional, accepted BE definition) bioinequivalent drug. With this objective in mind, in the next section, we try to clarify the hypotheses under consideration in the 2010 SABE limits and to separate them from the statistical criteria used to take a decision. Section 3 is devoted to some concerns that may appear when considering these hypotheses and statistical methods and to suggest straightforward improvements to them. Section 4 illustrates the use of these methods on a specific data set. Section 5 presents a simulation study trying to quantify the type I error probability and power of the regulatory methods and their (possibly) improved versions. Finally, section 6 discusses the main results and some ways of continuing this research.

2. European Medicines Agency and US Food and Drug Administration definitions and testing methods for scaled average bioequivalence

The EMA and FDA approaches to SABE have been studied in several papers like [8–10] but mainly from the point of view of the probability of declaring BE, without considering whether this decision was a type I error, that is, declaring as bioequivalent a truly bioinequivalent drug or not. A recent exception in the case of the EMA criterion is [11].

In the literature dealing with the SABE approach (except in some papers like [12]), little or null attention is paid to exactly defining the hypotheses under consideration. Sometimes, the overall impression is that the underlying BE definition still is that the true value of the formulation effect lies inside the standard fixed ± 0.223 limits, independently of using a SABE testing criterion. Then, the hypotheses under question still are the usual, defined previously in (1). This approach poses potential risks to the public health. Even with the additional safeguard of requiring that the *estimated* formulation effect must lie between the ± 0.223 limits, the probability of type I error uncontrolledly soars for growing variability values starting near a coefficient of variation in the original scale (CV) near 30%, as is graphically illustrated in the Supporting Information PDF, section S1.

Type I error study requires a precise definition of the hypotheses under consideration and to clearly distinguish them from the corresponding testing criteria. Here, we continue the approach taken in [12]. The hypotheses (i.e. the confronted concepts of ‘true non-SABE’ and ‘true SABE’) are previous to data; they must be expressed in terms of *population* parameters, never in terms of quantities depending on the sample. On the other hand, the testing procedures are based on data, that is, on the estimated values of these parameters. The definitions of SABE and the decision criteria differ in both regulatory approaches:

2.1. Hypotheses under the European Medicines Agency’s approach (only for C_{max})

In the EMA approach, the BE limits depend on the true (and unknown in practice) value of the reference residual standard deviation (in the logarithmic scale), σ_{WR} . If σ_{WR} is less than 0.294 (corresponding to a coefficient of variation in the original scale of $CV = 30\%$, $0.294 = \sigma_{CV=30}$), the BE limits are the usual ± 0.223 limits. If σ_{WR} lies between 0.294 and 0.472 (corresponding to a coefficient of variation in the original scale of 50%, $0.472 = \sigma_{CV=50}$), the BE limits grow as a linear function of σ_{WR} , $\pm 0.76\sigma_{WR}$. Otherwise, from $\sigma_{WR} = 0.472$, the BE limits stay constant at ± 0.359 . A true situation of bioequivalence

occurs if the true (and unknown in practice) value of the formulation effect lays between these limits. Thus, the null (non-SABE) and alternative (SABE) hypotheses under consideration are:

$$\begin{aligned} H_0 &: \phi \leq -\phi_{EMA}(\sigma_{WR}) \quad \vee \quad \phi \geq \phi_{EMA}(\sigma_{WR}) \\ H_1 &: -\phi_{EMA}(\sigma_{WR}) < \phi < \phi_{EMA}(\sigma_{WR}) \end{aligned} \quad (2)$$

where:

$$\phi_{EMA}(\sigma_{WR}) = \begin{cases} 0.223 & \text{if } \sigma_{WR} < \sigma_{CV=30} \\ k_{EMA}\sigma_{WR} & \text{if } \sigma_{CV=30} \leq \sigma_{WR} < \sigma_{CV=50} \\ 0.359 & \text{if } \sigma_{WR} \geq \sigma_{CV=50} \end{cases} \quad (3)$$

(or, in other words, $\phi_{EMA}(\sigma_{WR}) = \min\{\max\{0.223, k_{EMA}\sigma_{WR}\}, 0.359\}$). The constants defining the hypotheses, k_{EMA} , $\sigma_{CV=30}$ and $\sigma_{CV=50}$, come from the basic relation:

$$CV = 100\sqrt{e^{\sigma_{WR}^2} - 1} \quad \text{or} \quad \sigma_{WR} = \sqrt{\log\left\{\left(\frac{CV}{100}\right)^2 + 1\right\}} \quad (4)$$

and thus

$$k_{EMA} = \frac{\log 1.25}{\sigma_{CV=30}}, \quad \sigma_{CV=30} = \sqrt{\log\left\{\left(\frac{30}{100}\right)^2 + 1\right\}}, \quad \sigma_{CV=50} = \sqrt{\log\left\{\left(\frac{50}{100}\right)^2 + 1\right\}}.$$

These values are typically rounded to $k_{EMA} = 0.76$, $\sigma_{CV=30} = 0.294$ and $\sigma_{CV=50} = 0.472$.

2.2. European Medicines Agency's scaled average bioequivalence decision criterion

To declare or not bioequivalence is a decision subject to uncertainty that must be taken in view of data and the estimated values. A short statement of the EMA testing decision criterion is:

- (1) Obtain the formulation effect estimate, $\hat{\phi}$.
- (2) (Point estimate constraint) if $\hat{\phi}$ is outside the ± 0.223 limits, do not declare BE (do not reject H_0) and stop the process.
- (3) Obtain the estimate of σ_{WR} , $\hat{\sigma}_{WR}$.
- (4) Obtain the standard shortest confidence interval for ϕ , $CI_{\hat{\phi}}$.
- (5) If $CI_{\hat{\phi}}$ is fully included in the $\pm\phi_{EMA}(\hat{\sigma}_{WR})$ limits, declare BE (reject H_0), otherwise do not declare BE.

2.3. Hypotheses under the US Food and Drug Administration's Approach (Cmax and AUC)

Under the FDA approach, the hypotheses may be stated in a similar way:

$$\begin{aligned} H_0 &: \phi \leq -\phi_{FDA}(\sigma_{WR}) \quad \vee \quad \phi \geq \phi_{FDA}(\sigma_{WR}) \\ H_1 &: -\phi_{FDA}(\sigma_{WR}) < \phi < \phi_{FDA}(\sigma_{WR}) \end{aligned} \quad (5)$$

but the function defining the limits, ϕ_{FDA} , is:

$$\phi_{FDA}(\sigma_{WR}) = \begin{cases} 0.223 & \text{if } \sigma_{WR} < \sigma_{CV=30} \\ k_{FDA}\sigma_{WR} & \text{if } \sigma_{WR} \geq \sigma_{CV=30} \end{cases} \quad (6)$$

Thus, the scaling region is not right-bounded. k_{FDA} is taken as $\log(1.25)/0.25 = 0.892$, which produces a discontinuity at $\sigma_{CV=30}$. Figure 1 shows the EMA and FDA scaled limits, in the logarithmic scale, but in function of the coefficient of variation in the original scale.

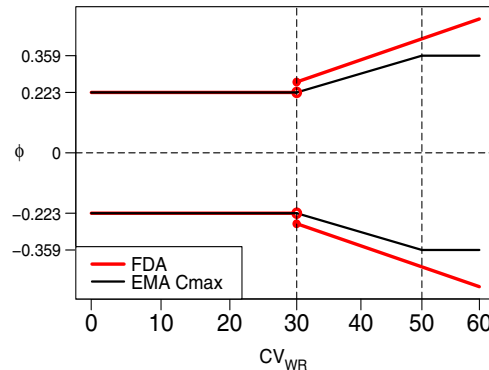


Figure 1. Bioequivalence limits according to the European Medicines Agency (EMA) [4] and the US Food and Drug Administration [5] regulations, scaled in function of the within-subject variability of the reference R formulation, σ_{WR} . In both regulations, scaling starts at $\sigma_{WR} = 0.294$ ($CV_{WR} = 30\%$) and, only in the EMA regulation, ends at $\sigma_{WR} = 0.472$ ($CV_{WR} = 50\%$). True condition of bioequivalence is assumed if the true but unknown formulation effect value ϕ lies between these limits. This formulation effect refers to C_{max} and AUC in the US Food and Drug Administration regulation, and only to C_{max} in the EMA regulation, always considered in logarithmic scale.

2.4. US Food and Drug Administration's scaled average bioequivalence decision criterion

In the scaled region, the FDA testing procedure takes a different approach. Provided that $-k_{FDA}\sigma_{WR} < \phi < k_{FDA}\sigma_{WR}$ is equivalent to $\phi^2 - k_{FDA}^2\sigma_{WR}^2 < 0$, inference is not performed on the ϕ parameter but on the compound parameter $\theta = \phi^2 - k_{FDA}^2\sigma_{WR}^2$:

- (1) Obtain the formulation effect estimate, $\hat{\phi}$.
- (2) If $\hat{\phi}$ is outside the ± 0.223 limits, do not declare BE (do not reject H_0) and stop.
- (3) Obtain the estimate of σ_{WR} , $\hat{\sigma}_{WR}$.
- (4) If $\hat{\sigma}_{WR} < \sigma_{CV=30}$ then,
 - (a) Obtain the standard shortest confidence interval for ϕ , CI_{ϕ} .
 - (b) If CI_{ϕ} is fully included in the ± 0.223 limits, declare BE (reject H_0), otherwise do not declare BE. Stop the process.
- (5) Otherwise,
 - (a) Obtain the upper limit $\hat{\theta}_U$ of a one-sided confidence interval for θ , $(-\infty, \hat{\theta}_U]$
 - (b) If $\hat{\theta}_U < 0$ declare BE (reject H_0), otherwise do not declare BE.

The upper confidence bound $\hat{\theta}_U$ is computed according to the Howe's method, described in [13, 14] and adapted to the previous situation in [15]. The algorithm is described in detail in [3].

3. Some concerns and possible improvements on the scaled average bioequivalence regulations

Possible concerns over the previous concepts and methods definitions are:

- **Non-regular functions of within-subject variability.** In fact, the BE limits are a concatenation of two or three different functions: a constant function continued by a linear function (and, again, by a constant function for the EMA case). The resulting limits are non-derivable at the juncture points and (possibly worse) there is a discontinuity in the FDA case.
- **Stepwise decision criteria.** Both criteria are based on a first decision according to the *estimated* value of the reference within-subject variability. By chance, in the neighbourhood of 30%, for a *true* CV below 30% (deserving constant limits), the estimation may be above 30% and thus the widened scaled limits applied (and vice versa). A priori, one may think that the severity of the preceding effect may depend on how abrupt is the change from constant to scaled limits. Thus, it would be worse for the FDA limits because of their discontinuity: data in fact deserving constant standard

limits may be evaluated under much more wider, less strict limits, possibly resulting in a type I error inflation. This drawback was pointed in [16], pp. 180–181 in a different context, in reference to the possibility of abruptly jumping from the standard 80/125 limits to widened 75/133 limits at $CV = 30\%$. The undesirable consequences of this discontinuity in the FDA SABE limits were also pointed in [17].

- **Interval inclusion principle with random limits.** This possible objection is only applicable to the EMA criterion. The confidence interval inclusion principle is strictly valid for constant, previously fixed, BE limits. But in fact, at the scaled region of the BE limits, the EMA limits are a function of the estimated (random) within-subject variability for the reference formulation. This is not the case for the FDA criterion where the parameter of interest is the compound parameter $\theta = \phi^2 - k_{FDA}^2 \sigma_{WR}^2$ tested against a fixed limit.

3.1. Howe's method on the European Medicines Agency scaled limits

The last objection on the EMA decision criterion has a quite obvious solution: just substitute its steps 4 and 5 by and adaptation of the corresponding steps in the FDA criterion:

- (1) Obtain the formulation effect estimate, $\hat{\phi}$.
- (2) If $\hat{\phi}$ is outside the ± 0.223 limits, do not declare BE (do not reject H_0) and stop the process.
- (3) Obtain the estimate of σ_{WR} , $\hat{\sigma}_{WR}$.
 - If $\hat{\sigma}_{WR} < \sigma_{CV=30}$ or $\hat{\sigma}_{WR} \geq \sigma_{CV=50}$ then,
 - (a) Obtain the standard shortest confidence interval for ϕ , CI_ϕ .
 - (b) If CI_ϕ is fully included in the limits $\pm \phi_{EMA}(\sigma_{WR})$ (say ± 0.223 if $\hat{\sigma}_{WR} < \sigma_{CV=30}$ or ± 0.359 if $\hat{\sigma}_{WR} \geq \sigma_{CV=50}$) declare BE (reject H_0), otherwise do not declare BE. Stop the process.
- (4) Otherwise,
 - (a) Obtain the upper limit $\hat{\eta}_U$ of a one-sided confidence interval for $\eta = \phi^2 - k_{EMA}^2 \sigma_{WR}^2$, $(-\infty, \hat{\eta}_U]$.
 - (b) If $\hat{\eta}_U < 0$ declare BE (reject H_0), otherwise do not declare BE.

In step 5, the upper confidence limit is obtained again as is described in [3], simply changing the constant k_{FDA} ($= 0.892$) by k_{EMA} ($= 0.76$). In the next sections, we will designate this third approach as 'Howe-EMA'.

3.2. Continuous US Food and Drug Administration limits

The solution to the concern on the lack of continuity of the FDA limits is still more direct: just adequately adjust the slope value k_{FDA} and/or the coefficient of variation CV_0 (or the reference standard deviation σ_0) value where the limits change from constant to scaled. Possible choices have been discussed in the literature and summarized in [3]. To maintain the high variability (HV) threshold at $CV_0 = 30\%$ take $k_{FDA} = k_{EMA} = 0.76$. We will designate this approach as 'Cont-FDA', and the limits function as ϕ_{CFDA} . Alternatively, if the scaling factor $k_{FDA} = 0.893$ should be preserved, $\sigma_0 = 0.25$ ($CV_0 = 25.40\%$) should be taken. We will designate this last approach as 'Cont-FDA2', and the function defining these limits as ϕ_{CFDA2} .

4. An example

To illustrate the use of the previous methods, we use the same data considered in [12], corresponding to a partially replicated crossover design (TRR/RTR/RRT). The R functions implementing these methods are available as Supporting Information. For $\log(AUC)$, the estimate of the formulation effect is 0.056 and the estimated within-subject standard deviation for the reference formulation is 0.345. For $\log(Cmax)$, the same estimates are 0.32 and 0.57, respectively.

For $\log(AUC)$, the point estimate constraint is fulfilled as the estimated formulation effect is 0.056, inside the ± 0.223 limits. According to the FDA regulatory criterion, as $0.345 > \sigma_0 = 0.294$, the BE limits must be scaled. With $k_{FDA} = 0.893$, the upper limit for the one-sided confidence interval for the compound parameter $\theta = \phi^2 - k_{FDA}^2 \sigma_{WR}^2$ is $\hat{\eta}_U = -0.06 < 0$, and BE is declared. With the continuous limits designated as Cont-FDA2, scaling is also necessary as $0.345 > \sigma_0 = 0.25$. The scaling

J. MUÑOZ, D. ALCAIDE AND J. OCAÑA

occurs if the true (and unknown in practice) value of the formulation effect lays between these limits. Thus, the null (non-SABE) and alternative (SABE) hypotheses under consideration are:

$$\begin{aligned} H_0 : \phi &\leq -\phi_{EMA}(\sigma_{WR}) \quad \vee \quad \phi \geq \phi_{EMA}(\sigma_{WR}) \\ H_1 : -\phi_{EMA}(\sigma_{WR}) &< \phi < \phi_{EMA}(\sigma_{WR}) \end{aligned} \quad (2)$$

where:

$$\phi_{EMA}(\sigma_{WR}) = \begin{cases} 0.223 & \text{if } \sigma_{WR} < \sigma_{CV=30} \\ k_{EMA}\sigma_{WR} & \text{if } \sigma_{CV=30} \leq \sigma_{WR} < \sigma_{CV=50} \\ 0.359 & \text{if } \sigma_{WR} \geq \sigma_{CV=50} \end{cases} \quad (3)$$

(or, in other words, $\phi_{EMA}(\sigma_{WR}) = \min\{\max\{0.223, k_{EMA}\sigma_{WR}\}, 0.359\}$). The constants defining the hypotheses, k_{EMA} , $\sigma_{CV=30}$ and $\sigma_{CV=50}$, come from the basic relation:

$$CV = 100\sqrt{e^{\sigma_{WR}^2} - 1} \quad \text{or} \quad \sigma_{WR} = \sqrt{\log\left\{\left(\frac{CV}{100}\right)^2 + 1\right\}} \quad (4)$$

and thus

$$k_{EMA} = \frac{\log 1.25}{\sigma_{CV=30}}, \quad \sigma_{CV=30} = \sqrt{\log\left\{\left(\frac{30}{100}\right)^2 + 1\right\}}, \quad \sigma_{CV=50} = \sqrt{\log\left\{\left(\frac{50}{100}\right)^2 + 1\right\}}.$$

These values are typically rounded to $k_{EMA} = 0.76$, $\sigma_{CV=30} = 0.294$ and $\sigma_{CV=50} = 0.472$.

2.2. European Medicines Agency's scaled average bioequivalence decision criterion

To declare or not bioequivalence is a decision subject to uncertainty that must be taken in view of data and the estimated values. A short statement of the EMA testing decision criterion is:

- (1) Obtain the formulation effect estimate, $\hat{\phi}$.
- (2) (Point estimate constraint) if $\hat{\phi}$ is outside the ± 0.223 limits, do not declare BE (do not reject H_0) and stop the process.
- (3) Obtain the estimate of σ_{WR} , $\hat{\sigma}_{WR}$.
- (4) Obtain the standard shortest confidence interval for ϕ , CI_{ϕ} .
- (5) If CI_{ϕ} is fully included in the $\pm\phi_{EMA}(\hat{\sigma}_{WR})$ limits, declare BE (reject H_0), otherwise do not declare BE.

2.3. Hypotheses under the US Food and Drug Administration's Approach (Cmax and AUC)

Under the FDA approach, the hypotheses may be stated in a similar way:

$$\begin{aligned} H_0 : \phi &\leq -\phi_{FDA}(\sigma_{WR}) \quad \vee \quad \phi \geq \phi_{FDA}(\sigma_{WR}) \\ H_1 : -\phi_{FDA}(\sigma_{WR}) &< \phi < \phi_{FDA}(\sigma_{WR}) \end{aligned} \quad (5)$$

but the function defining the limits, ϕ_{FDA} , is:

$$\phi_{FDA}(\sigma_{WR}) = \begin{cases} 0.223 & \text{if } \sigma_{WR} < \sigma_{CV=30} \\ k_{FDA}\sigma_{WR} & \text{if } \sigma_{WR} \geq \sigma_{CV=30}. \end{cases} \quad (6)$$

Thus, the scaling region is not right-bounded. k_{FDA} is taken as $\log(1.25)/0.25 = 0.892$, which produces a discontinuity at $\sigma_{CV=30}$. Figure 1 shows the EMA and FDA scaled limits, in the logarithmic scale, but in function of the coefficient of variation in the original scale.

constant is the same as before, and thus, $\hat{\theta}_U$ is again -0.06 , so BE is declared. For Cont-FDA limits, we return to the FDA threshold for HV. Again $0.345 > \sigma_0 = 0.294$, so scaling is required. Now the scaling constant is 0.76 , and the upper confidence limit for θ becomes $\hat{\theta}_U = -0.039$, which also results in a BE declaration.

According to the EMA regulatory criterion, $\log(AUC)$ should not be considered for scaling. But we discuss also the case of this variable, only for illustrative purposes. Then, scaling should be performed as $\sigma_{CV=30} = 0.294 < 0.345 < \sigma_{CV=50} = 0.472$. The scaled limits would be $\pm k_{EMA} 0.345 = \pm 0.76 \times 0.345 = \pm 0.262$, fully including the 90% confidence interval for the formulation effect, $(-0.038, 0.15)$. Applying the Howe-EMA criterion, $k_{EMA} = 0.76$ and the upper confidence limit for θ is the same than for Cont-FDA, -0.039 . In all cases, BE would be declared.

Strictly following the regulatory recommendations, for $\log(Cmax)$, BE cannot be declared as the formulation effect estimate is outside the standard ± 0.223 limits. But the possible convenience of removing the point estimate constraint (say, jumping step 1 in the previously stated decision criteria for BE) is considered in the next sections. In absence of this constraint, under the FDA method (and under Cont-FDA), the upper limit of the 95% confidence interval for the compound parameter $\theta = \phi^2 - k_{FDA}^2 \sigma_{WR}^2$ is negative, -0.027 , and thus, BE would be declared. Under the Cont-FDA2 approach this limit is positive, 0.036 , and the opposite decision would be taken. Finally, under the EMA and Howe-EMA approaches, the extended but constant ± 0.359 limits apply, as $\hat{\sigma}_{WR} = 0.57 > 0.472$. The 90% standard confidence interval for the formulation effect, $(0.1711, 0.4617)$, is not fully included in these limits, and it would not be possible to declare BE.

5. A comparative study of type I error and power

Simulations were performed in order to study the extent in which the previously stated concerns over the regulatory SABE methods might influence their desirable test properties, and the extent to what the proposed solutions may correct undesirable behaviour. For each simulation scenario, 100,000 data sets were generated under a replicated crossover design, TRTR/RTRT, or a partially-replicated crossover design, TRR/RTR/RRT.

The linear model under consideration was the same stated in [18], p.38:

$$Y_{ijk} = \mu + G_k + S_{i(k)} + P_j + F_{(j,k)} + C_{(j-1,k)} + e_{ijk} \quad (7)$$

where Y_{ijk} corresponds to the logarithm of the bioavailability ($Cmax$ in the EMA regulation, $Cmax$ or AUC in the FDA regulation), i stands for the subject, $i = 1, \dots, n_k$, nested in sequence k , $k = 1, \dots, K$ ($K = 2$ in the replicated case, $K = 3$ in the partially replicated case) and j stands for the period, $j = 1, \dots, J$, ($J = 4$ in the replicated case, $J = 3$ in the partially replicated case). μ corresponds to the general mean; G_k to the fixed effect of sequence k ; $S_{i(k)}$ to the random effect of subject i in sequence k ; P_j to the fixed effect of period j ; $F_{(j,k)}$ to the fixed effect of the formulation administered in the j -th period of the k -th sequence ($F_{(1,1)} = F_T, F_{(2,1)} = F_R$, etc.), $C_{(j-1,k)}$ to the fixed carryover effect of period $j-1$ on period j , $j = 2, \dots, J$, in sequence k and e_{ijk} to the within subject residual. The fixed effects are assumed to add up to zero, and the random effects are assumed independent and normally distributed, $S_{i(k)} \sim N(0, \sigma_S^2)$. In model (7), the precise definition of the formulation effect is $\phi = F_T - F_R$. The residual or 'within' variances under different formulations may differ: $e_{ijk} \sim N(0, \sigma_{WR}^2)$ if $F_{(j,k)} = F_R$, and $e_{ijk} \sim N(0, \sigma_{WT}^2)$ if $F_{(j,k)} = F_T$.

Each simulation scenario was characterized by the crossover design (TRTR/RTRT or TRR/RTR/RRT), the regulatory criterion (FDA, Cont-FDA, Cont-FDA2, EMA or Howe-EMA) and a combination of σ_{WR} , ϕ and sample size values. Additionally, simulations were performed under the point estimate constraint and also in absence of it. Simulated data sets were always balanced, with sample sizes per sequence $n = 12, 18, 24$ or 36 for the TRTR/RTRT design and $n = 8, 12, 16$ or 24 for the TRR/RTR/RRT design – and always a total sample size $N = 24, 36, 48$ or 72 . The σ_{WR} values ranged from $\sigma_{CV=10\%} (= 0.09975135)$ to $\sigma_{CV=60\%} (= 0.554513)$ in CV increments of 2%, except in the neighbourhood of the possibly critical values $\sigma_{CV=30\%}$ and $\sigma_{CV=50\%}$ (and $\sigma_{CV=25.4\%}$ for the Cont-FDA2 case) where the increments were of 0.5%. Homoscedasticity in the within-subject variabilities (say, $\sigma_W = \sigma_{WR} = \sigma_{WT}$) was assumed. The simulated ϕ values depended of the corresponding σ_{WR} value, as is explained later. A sufficient washout period, and thus null carryover effect, was assumed. This assumption is important because, otherwise, the estimators

J. MUÑOZ, D. ALCAIDE AND J. OCAÑA

of the formulation effect would be biased. The values of the remaining model parameters were irrelevant in the final conclusions.

All simulations were performed with 64 bits R 3.1.2 under different machines running Windows 7 and Windows 8. The simulations are reproducible with the R scripts available as Supporting Information.

5.1. Simulations under true H_0 (no bioequivalence)

Under H_0 true, that is, for a bioequivalence scenario with the true value of the formulation effect, $\phi = F_T - F_R$, outside the SABE limits, a valid test should have a user's risk, say a probability of declaring BE (rejecting H_0), no larger than the nominal significance level (e.g. $\alpha = 0.05$). In other words, the probability of type I error should not exceed α .

A crucial series of scenarios corresponds to the case where the values of ϕ are just over the SABE limits. Ideally, in these cases, the type I error should be exactly equal to the significance level α (and increasingly greater than α for ϕ values progressively inside the SABE limits $-H_1$ true-; decreasingly smaller than α for ϕ values outside the SABE limits $-H_0$ true).

In correspondence to the previous considerations, for each simulated σ_{WR} value, the sample data sets were generated according to a ϕ value exactly equal to $\phi_{EMA}(\sigma_{WR})$ or to $\phi_{FDA}(\sigma_{WR})$, depending on the regulatory criterion under study, or exactly equal to $\phi_{CFDA}(\sigma_{WR})$ or $\phi_{CFDA2}(\sigma_{WR})$ if the suggested alternative FDA limits were considered.

Figure 2 shows the simulated proportion of BE declarations in function of the within-subject coefficient of variation for the reference formulation and the replicated design TRTR/RTRT, under the point estimate constraint. Very similar results for the partially replicated design TRR/RTR/RRT are available in the Supporting Information PDF, section S2. With 100,000 simulation replicates, the uncertainty of these estimates of the probability of declaring BE (here a type I error) is relegated to the third decimal place, if expressed as a 95% confidence interval for the true probability.

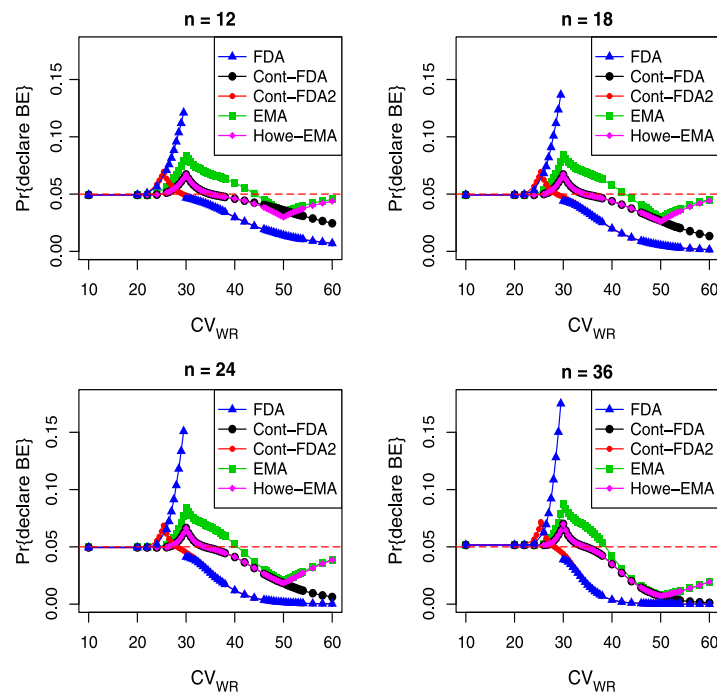


Figure 2. Probability of declaring bioequivalence (BE) (and in this case committing a type I error) for different BE criteria, when the true value of the formulation effect exactly lies on the BE limits and the point estimate constraint applies, in function of the true (but unknown) within-subject variability of the reference R formulation. US Food and Drug Administration (FDA) stands for the regulatory FDA decision criterion and limits, Cont-FDA and Cont-FDA2 stand for applying the same FDA procedure but for two alternative continuous limits, European Medicines Agency (EMA) stands for the regulatory EMA procedure and limits, and Howe-EMA stands for using the EMA limits but in conjunction with the statistical decision criterion proposed by the FDA.

With respect to the user's risk, the conclusions seem clear:

The FDA criterion conducts to an inadmissible test when the reference within-subject variability approaches a 30% coefficient of variation in the original scale, with type I probabilities overly above the nominal 0.05. In this point, the function relating variability and rejection probability presents a wide discontinuity. This problem increases with growing sample sizes. For within-subject reference variabilities over $CV = 50\%$, its type I error probability falls to nearly 0.

The EMA criterion performs better but still with a too large type I probability (e.g. values near 0.1 for a nominal significance level of 0.05).

The suggested improvements on both regulatory criteria (Howe's approximation over the EMA limits and both continuous versions of FDA limits) greatly palliate these drawbacks, although a peak with (slightly) too large type I error probabilities in the neighbourhood of the arbitrarily chosen threshold for HV (either 25% or 30%) still persists. In fact, as may be expected, the Howe-EMA and Cont-FDA methods (sharing the same scaling factor) perform equally, except for within-subject reference variabilities over $CV = 50\%$, where the unbounded scaling nature of the FDA approach matters. For these high CV values, the type I error probability in both methods falls to values clearly below the 0.05 level, although this probability remains closer to 0.05 in the Howe-EMA approach, with greater values than those corresponding to the Cont-FDA method. Similarly, FDA and Cont-FDA2, sharing the same scaling factor, perform equally except in the neighbourhood of 25% to 30% CV values where Cont-FDA2 avoids the discontinuity in the probability curve while maintaining a peak of undesirably large type I error probability.

At the juncture point $CV = 50\%$ where the EMA limits are also non-derivable, the probability curve associated with the regulatory EMA and the Howe-EMA testing methods has also a non-derivable peak associated with a tendency change, from decreasing to increasing values, but always below 0.05. In any case, all methods share a clear decline in the type I error probability for large CV values, which is markedly below the nominal 0.05 level.

As is illustrated by Figure 3, the point estimate constraint, say to declare BE only if the formulation effect point estimate remains inside the ± 0.223 limits, is not essential for type I error control: with or without this constraint, type I error probabilities are too large in the same scenarios (in the neighbourhood of the HV threshold). Additionally, for large CV values, it dominates the testing criterion: the simulated true value of $|\phi|$ (just over the SABE limits or in their neighbourhood) is much greater than 0.223, thus

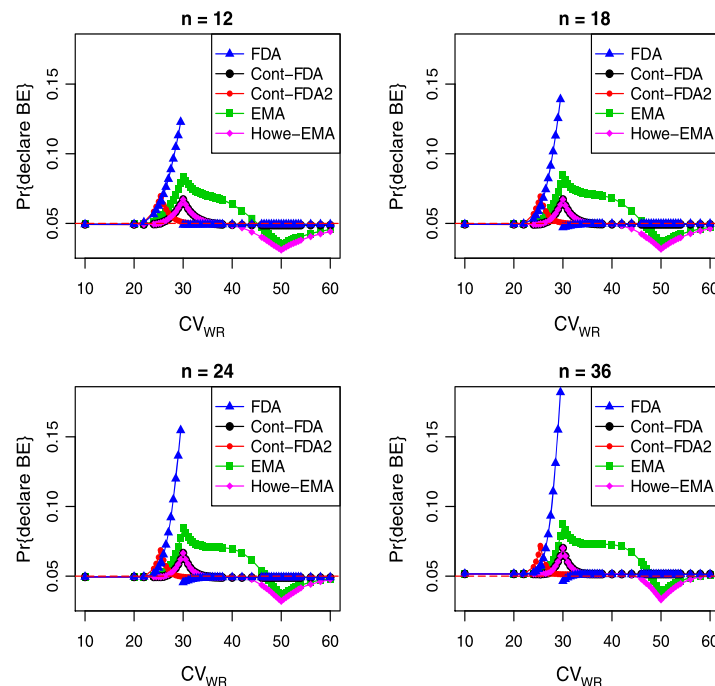


Figure 3. Probability of declaring bioequivalence in the same scenarios than Figure 2 but removing the restriction on the point estimate of the formulation effect lying inside the limits ± 0.223 .

rarely its estimate may lie inside ± 0.223 in order to declare BE. In absence of the constraint, the type I error probability for large CV values is closer to the nominal 0.05 and still defines a valid test – and without the penalty in power mentioned in the next section.

5.2. Simulations under true H_1 (bioequivalence)

In order to study the power performance of the regulatory criteria, and the extent to what the corrections suggested in previous sections may affect this power, simulations were performed in the same conditions than in the previous sections but generating data for ϕ values progressively inside the SABE region. This was done by simulating ϕ values over the line defined by a fraction λ of the SABE limits, say $\lambda\phi_{FDA}(\sigma_{WR})$, $\lambda\phi_{EMA}(\sigma_{WR})$, $\lambda\phi_{CFDA}(\sigma_{WR})$ or $\lambda\phi_{CFDA2}(\sigma_{WR})$, depending on the scaled limit under consideration. The simulated values were $\lambda = 0.8, 0.5, 0.2$ and 0 . $\lambda = 0$ corresponds to complete equality in means between R and T , $\phi = 0$, and the other values define different degrees of difference, from less difference (0.2) to more difference (0.8) but always inside the SABE limits. Obviously $\lambda = 1$ would correspond to the scenarios considered in the previous subsection, that is, formulation effect values exactly over the SABE limits, and $\lambda > 1$ to values progressively inside the non-SABE (H_0 true) region.

In general, the SABE methods are powerful for the sample sizes recommended in the regulations, especially in the most favourable case for BE, that is, when $\phi = 0$. As may be expected, for values of ϕ still in the SABE region but near the bioequivalence limit (e.g. when $\lambda = 0.8$) there is a decline in power, especially for large within-subject variabilities. This decline is more pronounced under the point estimate constraint, which seems to impose an unnecessary (almost from the point of view of type I error control) penalty on power. These results confirm the findings in previous studies (e.g. [19]) but under our partially new setting. As the main focus of the present paper is type I error or consumer's risk, and not power, they are relegated to section S3 of Supporting Information.

6. Discussion

The preceding results on a poor type I error control of current SABE regulatory criteria for HVD (and especially in the case of the FDA approach) for reference coefficients of variation around 30% may potentially create an issue in public health, as a non-negligible proportion of generic HVD drug applications for BE seem to correspond to variabilities around these values. As an example, according to the study in [20], more than half of HVD applications reviewed at FDA in 2003–2005 correspond to σ_w estimates between 0.3 and 0.35, and a 10% correspond to borderline values, slightly smaller or larger than 0.3. So, correcting these issues on consumer's risk, in the line of the straightforward alternatives proposed here (or future more elaborated ones), may be of crucial interest. In any case, although not providing a full correction, the conjunction of EMA limits and a more rigorous decision criterion (Howe-EMA) or a continuous version of the FDA limits, greatly improve the type I error control of the regulatory approaches issued in 2010. It seems also clear that these corrections do not seriously penalize power, as there is not clear power dominance of one method over the others.

All other considerations aside except type I error control, the point estimate constraint do not appear to be essential, and it imposes some penalty on power.

Despite the previous comments (and perhaps being too finicky), the behaviour of the improved EMA and FDA criteria still is penalized by the non-regular character of scaled limits on the variability values where limits abruptly (although continuously) change from constant to scaled or again from scaled to constant in the EMA regulation. A possible further improvement may be based on more regular limits based in softer functions anywhere derivable, for example, those stated in [21, 22] and subsequent literature produced by these authors. Soft limits closely emulating the present EMA limits are proposed in [9]. These limits are more in agreement with the appealing idea that, possibly, the change from non-high variability to high variability is more adequately described as a progressive, fuzzy, concept rather than by a yes/no criterion and imposing a quite unnatural threshold definition (25%, 30%?) for high variability. On the other hand, a possible criticism on the way that these soft limits are integrated in BE testing relates to the last concern stated in Section 4, that is, using the interval inclusion principle with random BE limits. To solve this issue, standard techniques like the Howe's method may be adapted to be used with these non-linear limits, as will be discussed in a forthcoming paper.

Finally, it is worth pointing that two-stage designs have been considered also acceptable from a regulatory point of view (e.g. [4, 23–25]) for dealing with HVD. It would be interesting to compare this approach with SABE in terms of criteria like statistical adequacy, exposure of healthy subjects or

economic cost. Such a comparison poses some methodological problems, like the fact that the own concept of bioequivalence is not exactly the same in both approaches: while in the two-stage approach it is associated with the standard ± 0.223 region, in the SABE approach it is associated with an expanding region of possible formulation effect values, which requires additional arguments in terms of a wider therapeutic range associated with HVD in order to make it justifiable.

Acknowledgements

This research was supported by the Spanish Ministerio de Economía y Competitividad with grant MTM2011-15849-E and by grant 2014 SGR 464, Generalitat de Catalunya.

References

1. FDA. *Food and drug administration. guidance for industry: Statistical approaches to establishing bioequivalence*. CDER, Department of Health and Human Services. Rockville, MD, 2001. Available from: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070244.pdf> [Accessed on 29 November 2014].
2. Midha KK, Shah VP, Singh GJP, Patnaik R. Conference report: bio-international 2005. *Journal of Pharmaceutical Sciences* 2007; **96**(4):747–754.
3. Tothfalusi L, Endrenyi L, Garcia Arieta A. Evaluation of bioequivalence for highly variable drugs with scaled average bioequivalence. *Clinical Pharmacokinetics* 2009; **48**(11):725–743.
4. EMA. *Guideline on the investigation of bioequivalence*. European Medicines Agency, CPMP/EWP/QWP/1401/98 Rev. 1/Corr.London, 2010. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/01/WC500070039.pdf [Accessed on 29 November 2014].
5. FDA. *Draft guidance on progesterone*. The United States Food and Drug Administration: Rockville, Maryland, 2010. Available from: www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm209294.pdf [Accessed on 29 November 2014].
6. Davit BM, Chen ML, Conner DP, Haidar SH, Fabian-Fritsch B, Kim S, Lee CH, Lionberger RA, Makhlof FT, Nwakama PE, Patel DT, Schuirmann DJ, Yu LX. Implementation of a reference-scaled average bioequivalence approach for highly variable generic drug products by the US food and drug administration. *The AAPS Journal* 2012; **14**(4):915–921.
7. Davit BM, Brady AC, Conner DP, Yu LX. International guidelines for bioequivalence of systemically available orally administered generic drug products: a survey of similarities and differences. *The AAPS Journal* 2013; **15**(4):974–990.
8. Haidar SH, Makhlof FT, Schuirmann DJ, Hyslop T, Davit, BM author=Conner DP, Yu LX. Evaluation of a scaling approach for the bioequivalence of highly variable drugs. *The AAPS Journal* 2008; **10**(3):450–454.
9. Karalis V, Symillides M, Macheras P. On the leveling-off properties of the new bioequivalence limits for highly variable drugs of the EMA guideline. *European Journal of Pharmaceutical Sciences* 2011; **44**(4):497–505.
10. Karalis V, Symillides M, Macheras P. Bioequivalence of highly variable drugs: A comparison of the newly proposed regulatory approaches by FDA and EMA. *Pharmaceutical Research* 2012; **29**(4):1066–1077.
11. Wonnemann M, Frömke C, Koch A. Inflation of the type I error: investigations on regulatory recommendations for bioequivalence of highly variable drugs. *Pharmaceutical Research* 2015; **32**(1):135–143.
12. Patterson SD, Jones B. Viewpoint: observations on scaled average bioequivalence. *Pharmaceutical Statistics* 2012; **11**(1): 1–7.
13. Howe WG. Approximate confidence limits on the mean of $X + Y$ where X and Y are two tabled independent random variables. *Journal of the American Statistical Association* 1974; **69**(347):789–794.
14. Hyslop T, Hsuan F, Holder DJ. A small sample confidence interval approach to assess individual bioequivalence. *Statistics in Medicine* 2000; **19**(20):2885–2897.
15. Tothfalusi L, Endrenyi L, Midha K, Rawson M, Hubbard J. Evaluation of the bioequivalence of highly-variable drugs and drug products. *Pharmaceutical Research* 2001; **18**(6):728–733.
16. EMA. *Overview of comments received on draft guideline on the investigation of bioequivalence cpmp/ewp/qwp/1401/98 rev. 1*. London, 2010. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Other/2010/02/WC500073572.pdf [Accessed on 1 December 2014].
17. Endrenyi L, Tothfalusi L. Regulatory conditions for the determination of bioequivalence of highly variable drugs. *Journal of Pharmacy & Pharmaceutical Sciences* 2009; **12**(1):138–149.
18. Chow SC, Liu JP. *Design and Analysis of Bioavailability and Bioequivalence Studies*. Chapman & Hall/CRC Press: Boca Raton, 2009.
19. Tothfalusi L, Endrenyi L. Sample size for designing bioequivalence studies for highly variable drugs. *Journal of Pharmacy & Pharmaceutical Sciences* 2011; **15**(1):73–84.
20. Davit BM, Conner DP, Fabian-Fritsch B, Haidar SH, Jiang X, Patel DT, Seo PRH, Suh K, Thompson CL, Yu LX. Highly variable drugs: observations from bioequivalence data submitted to the FDA for new generic drug applications. *The AAPS Journal* 2008; **10**(1):148–156.
21. Karalis V, Macheras P, Symillides M. Geometric mean ratio-dependent scaled bioequivalence limits with leveling-off properties. *European Journal of Pharmaceutical Sciences* 2005; **26**(1):54–61.
22. Kytariolos J, Karalis V, Macheras P, Symillides M. Novel scaled bioequivalence limits with leveling-off properties. *Pharmaceutical Research* 2006; **23**(11):2657–2664.

J. MUÑOZ, D. ALCAIDE AND J. OCAÑA

23. EMA. *Questions & answers: positions on specific questions addressed to the pharmacokinetics working party (PKWP)*. European Medicines Agency, EMA/618604/2008 Rev. 12. London, 2015. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002963.pdf [Accessed on 3 September 2015].
24. FDA. *ANDA submissions refuse-to-accept standards. guidance for industry*. U.S. Department of Health and Human Services. Food and Drug Administration. Center for Drug Evaluation and Research (CDER), 2015. Available from: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM370352.pdf> [Accessed on 3 September 2015].
25. Polli JE, Cook JA, Davit BM, Dickinson PA, Argenti D, Barbour N, Garcia-Arieta A, Geoffroy JM, Hartauer K, Li S, Mitra A, Muller FX, Purohit V, Sanchez-Felix M, Skoug JW, Tang K. Summary workshop report: facilitating oral product development and reducing regulatory burden through novel approaches to assess bioavailability/bioequivalence. *The AAPS Journal* 2012; **14**(3):627–638.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site.