

# DISTRIBUCIÓN DEL VALOR P PARA LA PRUEBA T EN ESTUDIOS DE NO-INFERIORIDAD CON MUESTRAS PARALELAS

Pedro A. Monterrey<sup>1</sup>

Departamento de Matemáticas, Universidad del Rosario, Bogotá, Colombia.

## ABSTRACT

In randomized controlled clinical trials, treatment efficacy is considered as its capability to produce beneficial effects. Efficacy of an experimental treatment is usually established comparing its effectiveness respect to a placebo or comparing with the effect of an active control treatment. The first stage to determine effectiveness of a new treatment is placebo-controlled trial, and, as a second stage, a superiority trial will compare new treatment with a treatment uniformly accepted as a standard one. Sometimes the use of a placebo control is considered unethical; in such a case, a non-inferiority trial may be appropriate. In non-inferiority the new treatment don't need to be superior to a control, it is enough that it doesn't be unacceptably worse and additionally to be better than placebo. This assert rules the construction of non-inferiority hypothesis, which is based in a non-inferiority margin; in its determination is necessary to consider both statistical reasoning and clinical judgement. The possibility to claim incorrectly non-inferiority is one of the methodological flaws in non-inferiority studies and is related with the determination of non-inferiority margin. Two samples t-test is the statistical criteria to analyze non-inferiority hypothesis when mean value is the parameter to measure treatments efficacy. The objective of present report was to obtain and to characterize the p-value probability distribution for t-test in non-inferiority designs and to establish the relationship between this distribution and test's power function, as a result a characterization of the effect of sample size and non-inferiority margin over test sensibility was obtained, in addition an application of p-value distribution to meta-analysis was done introducing a criteria to analyze the effect of a particular study in results integration.

**KEYWORDS:** Non-inferiority. Non-inferiority margin. P-values distribution. Test Power. T-test. Meta-analysys

**MSC:** 62F03

## RESUMEN

En los ensayos clínicos aleatorizados la eficacia de un tratamiento se define como su capacidad, al ser aplicado en una intervención, de producir resultados beneficiosos. La eficacia de un nuevo tratamiento se establece, usualmente, comprando su efectividad con un placebo o comparando con un tratamiento de control. La primera etapa en la determinación de la efectividad de un nuevo tratamiento es la realización de un ensayo clínico controlado con placebo, en una segunda etapa se realiza un estudio de superioridad en el que el nuevo tratamiento se compara con otro que es aceptado universalmente. En ocasiones el uso de placebos no es ético, en estos casos se considera apropiado realizar un ensayo clínico de no-inferioridad. En la no inferioridad el nuevo tratamiento no tiene que ser superior al control, basta que no sea inaceptablemente peor, adicionalmente si se deba garantizar que sea mejor que el placebo. Esto se refleja en la construcción de las hipótesis de no-inferioridad, las que se basan en la determinación de un umbral de no inferioridad, para su determinación se combinan aspectos estadísticos y clínicos. Uno de los errores metodológicos más comunes en los estudios de no-inferioridad es la posibilidad de detectar erróneamente la no-inferioridad de un tratamiento, este error está relacionado directamente con la determinación del umbral de no-inferioridad. La prueba t para dos muestras es la prueba estadística utilizada para analizar las hipótesis de no inferioridad cuando la media es la medida de efecto de los tratamientos. El objetivo del presente trabajo fue obtener y caracterizar la distribución del valor P de la prueba t para los diseños de no-inferioridad y establecer la relación de esa distribución con la función de potencia de la prueba; lo que permitió caracterizar la sensibilidad de la prueba frente a cambios en los valores de umbral y los tamaños de la muestra y caracterizar el comportamiento de los diferentes estudios que integran un meta-análisis, brindando un elemento de juicio adicional para valorar el aporte de cada estudio a la integración de resultados

## 1. INTRODUCCIÓN

En los ensayos clínicos aleatorizados la eficacia de un tratamiento se define como su capacidad, al ser aplicado en una intervención, de producir resultados beneficiosos. La eficacia de un tratamiento experimental se establece usualmente comparando su efectividad con un placebo, es decir un producto sin principio activo, o comparando con un tratamiento de reconocida eficiencia que sea de uso común. Para establecer la eficacia de un tratamiento se utilizan los estudios de superioridad, que son el principal tipo de ensayo clínico; sin embargo, en ocasiones, el interés al comparar dos tratamientos en una investigación clínica se reduce a establecer si los tratamientos son equivalentes en sus efectos,

<sup>1</sup> pedro.monterrey@urosario.edu.co

importando la magnitud de sus diferencias más que la dirección en que estas se presentan. La primera etapa para determinar la efectividad de un nuevo tratamiento son los ensayos clínicos controlados utilizando un control al que se le administra un placebo; en una segunda etapa se realiza un estudio de superioridad para comparar el nuevo tratamiento con el tratamiento que representa el estándar para atender el problema de salud. La segunda opción procede si existe ese tratamiento estándar. En ocasiones el uso de un placebo no es ético, en ese caso para determinar si el nuevo tratamiento tiene un efecto relevante se realiza un estudio de no-inferioridad. En los estudios de no-inferioridad el nuevo producto o procedimiento no necesita ser superior al control, es suficiente que no sea inaceptablemente peor que el procedimiento estándar en uso pero que sea mejor que el placebo. Para formular las hipótesis estadísticas a considerar en los estudios de no-inferioridad siguiendo el criterio anterior, se parte de un número real  $d > 0$  que representa el umbral de no inferioridad y según Tsong y Chen [2007]: si  $\mu_T$ ,  $\mu_R$  y  $\mu_P$  representan respectivamente la media poblacional de la distribución de la variable en estudio aplicando el tratamiento experimental, el de referencia o el placebo, las hipótesis de no inferioridad serían  $H_0: \mu_T - \mu_P \leq (\mu_R - \mu_P) - d$ ;  $H_A: \mu_T - \mu_P > (\mu_R - \mu_P) - d$ ; formulación que, de forma simplificada, puede escribirse como  $H_0: \mu_d \leq -d$ ;  $H_A: \mu_d > -d$  con  $\mu_d = \mu_T - \mu_R$ . Adicional a esta formulación básica Tsong y Chen [2007] analizan dos vías adicionales de construir las hipótesis de no inferioridad.

Los diseños de no-inferioridad fueron introducidos a mediados de los años 90 del siglo pasado y como fue señalado por Hann [2012] su importancia ha sido creciente porque sustituyen los ensayos clínicos basados en placebo y son convenientes en los casos en que el nuevo tratamiento está disponible, es de fácil aplicación o no tiene efectos adversos.

Uno de los elementos principales en la construcción de las hipótesis de no inferioridad es la determinación del umbral de no inferioridad, para su determinación, según señalan D'Agostino y cols [2003], se combinan argumentos de tipo estadístico, basados en datos históricos de ensayos clínicos para analizar la efectividad del tratamiento en uso, con juicios clínicos. Head y cols [2012] señalan los peligros de no determinar correctamente el margen de no inferioridad, concretamente establecen que fijar el margen de no-inferioridad de una manera muy conservadora pudiera disminuir los tamaños de la muestra pero incrementar a la vez el riesgo de cometer un error de tipo I y en consecuencia aceptar no-inferioridad cuando realmente no exista.

En el caso en que la medida del efecto de un tratamiento sea su valor medio, la prueba t para dos muestras independientes es la prueba estadística utilizada para analizar los resultados de un estudio de no inferioridad. En estos casos el conocimiento del comportamiento del valor P pudiera constituir un elemento fundamental en el diseño de los estudios de no inferioridad y mejorar la investigación clínica. Debido a la asimetría de la distribución de probabilidad del valor P su mediana es el valor de posición más adecuado para representar su distribución. Battacharya y Habtzghi [2002] obtuvieron expresiones explícitas para la mediana del valor P y la relacionaron con la potencia de la prueba en el caso en que las varianzas eran conocidas. Monterrey [2016] obtuvo las expresiones exactas y aproximaciones para la distribución de probabilidad del valor P en poblaciones normales. El objetivo del presente trabajo fue relacionar la función de potencia de la prueba t, en el caso en que las varianzas son desconocidas, con la función de distribución del valor P, obteniendo la relación entre percentiles y las tasas de error de tipo I y II de la prueba. Adicionalmente se analiza la relación entre el umbral de no inferioridad con la sensibilidad de la prueba de hipótesis según el tamaño de la muestra

## 2. EL VALOR P PARA LA PRUEBA T EN EL ANÁLISIS DE HIPÓTESIS DE NO INFERIORIDAD

Se considerarán dos poblaciones independientes  $X_T \sim n(\mu_T, \sigma_T^2)$  para representar el tratamiento de estudio o experimental y  $X_R \sim n(\mu_R, \sigma_R^2)$ , el tratamiento de referencia o control. En ambas poblaciones se considerarán muestras independientes de tamaño  $n$  y  $m$  respectivamente; en las que  $\bar{X}_T$ ,  $\bar{X}_R$ ,  $S_T^2$  y  $S_R^2$  representarán las medias y varianzas muestrales en cada uno de los casos. Para  $d \geq 0$ , umbral de no inferioridad, se considerarán las hipótesis de no inferioridad  $H_0: \mu_d \leq -d$ ;  $H_A: \mu_d > -d$  con  $\mu_d = \mu_T - \mu_R$ . Según las varianzas sean iguales ( $\sigma_T^2 = \sigma_R^2 = \sigma^2$ ) o diferentes ( $\sigma_T^2 \neq \sigma_R^2$ ) se utiliza para el análisis de las hipótesis el estadígrafo  $E_i = (\bar{X}_T - \bar{X}_R + d)$ .  $S_1^{-1}$ ,  $i=1$  cuando las varianzas son iguales,  $i=2$  para las varianzas diferentes, donde  $S_1^2 = S_D^2 (n^{-1} + m^{-1})$ , denotando  $S_D^2$  la varianza combinada:  $S_D^2 = (n+m-2)^{-1} [(n-1)S_T^2 + (m-1)S_R^2]$  y  $S_2^2 = n^{-1}S_T^2 + m^{-1}S_R^2$ . Cumpliéndose que, para  $i=1,2$ ;  $E_i \sim t_{v_i}^{\Delta_i}$ , denotando  $t_{v_i}^{\Delta_i}$  la distribución t no central con  $v_i$  grados de libertad y  $\Delta_i$  parámetro de no centralidad, con  $v_1 = n+m-2$ ,  $v_2 = (n^{-1}\sigma_T^2 + m^{-1}\sigma_R^2)[(n-1)^{-1}(n^{-1}\sigma_T^2)^2 + (m-1)^{-1}(m^{-1}\sigma_R^2)^2]^{-1}$ ,  $\Delta_1 = \sigma^{-1}(\mu_d + d) [n m (n+m)^{-1}]^{1/2}$  y  $\Delta_2 = (\mu_d + d) (n^{-1}\sigma_T^2 + m^{-1}\sigma_R^2)^{-1}$ .

Para  $i=1,2$ ; el valor P de la prueba correspondiente es una realización de la variable aleatoria  $\Pi_i=1-T_{v_i}^0(E_i)$  donde  $T_{v_i}^0$  representa la función de distribución de la distribución  $t_{v_i}^{\Delta}$ . En ese caso, según se estableció en Monterrey [2016], la función de distribución de  $\Pi_i$ , para cada  $i$ , es:

$$(0 < x < 1) \quad F_{\Pi_i / \mu_d, d}(x) = 1 - T_{v_i}^{\Delta_i}(t_{v_i}^{\Delta_i}(1-x)) \quad (1)$$

Denotando  $t_{v_i}^{\Delta_i}(1-x)$  el percentil de orden  $1-x$  de la distribución  $t_{v_i}^{\Delta_i}$ . Derivando se obtiene la función de densidad de la distribución del valor P para cada una de las pruebas:

$$(0 < x < 1) \quad f_{\Pi_i / \mu_d, d}(x) = \frac{t_{v_i}^{\Delta_i}(t_{v_i}^{\Delta_i}(1-x))}{t_{v_i}^{\Delta_i}(t_{v_i}^{\Delta_i}(1-x))} \quad (2)$$

Denotando  $t_{v_i}^{\Delta_i}$  la función de densidad de la distribución  $t_{v_i}^{\Delta_i}$ . La Figura 1 muestra las densidades de la distribución del valor P en el caso en que las varianzas son iguales,  $\sigma^{-1}(\mu_d + d) = 0,5$  y  $n=m=k$ , con  $k=10,50,150$ .

De (1) se obtiene que los percentiles de la distribución del valor P para cada una de las dos pruebas,  $i=1,2$ ,

$$\text{son: } (0 < \gamma < 1) \quad \tilde{\Pi}_i / \mu_d, d(\gamma) = 1 - T_{v_i}^0(t_{v_i}^{\Delta_i}(1-\gamma)) \quad (3)$$

Denotando  $\tilde{\Pi}_i / \mu_d, d(\gamma)$  el percentil de orden  $\gamma$  de la distribución de  $\Pi_i$ ,  $i=1,2$ . Como se observó en la Figura 1 la distribución del valor P es altamente asimétrica. La asimetría aumenta con el tamaño de muestra, lo que explica uno de los problemas de las Pruebas de Significación que han sido profusamente analizados en la literatura del tema y que fue discutido en detalle por Monterrey [2012], en el sentido de que para ciertos diferencias irrelevantes son detectadas como importantes o “significativas” por las pruebas.

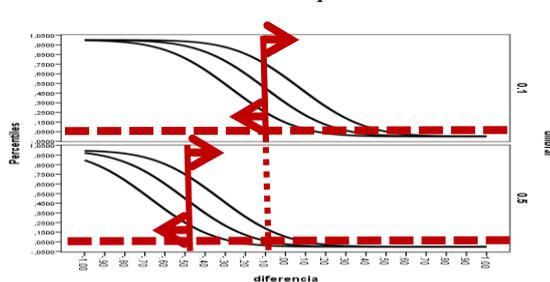
Figura 1: Densidades de la prueba para varianzas iguales, cuando  $\sigma^{-1}(\mu_d + d) = 0,5$

Una forma de cuantificar la sensibilidad de las pruebas, para un tamaño de muestra determinado, es cuantificar la asimetría de su distribución. Esto puede realizarse, a partir de los percentiles de la distribución, calculando, para  $i=1,2$ , el coeficiente de Galton,  $A_i(\mu_d, d)$ :

$$A_i(\mu_d, d) = \frac{2T_{v_i}^0(t_{v_i}^{\Delta_i}(0,5)) - T_{v_i}^0(t_{v_i}^{\Delta_i}(0,75)) - T_{v_i}^0(t_{v_i}^{\Delta_i}(0,25))}{T_{v_i}^0(t_{v_i}^{\Delta_i}(0,75)) - T_{v_i}^0(t_{v_i}^{\Delta_i}(0,25))} \quad (4)$$

Para el diseño de los estudios de no-inferioridad con muestras paralelas un elemento central, conjuntamente con la determinación del tamaño de la muestra, es la definición del umbral de no inferioridad. Lo usual para su determinación un conjunto de elementos clínicos y estadísticos, D’Agostino y cols [2003]; estos criterios podrían complementarse con un análisis de la sensibilidad de la prueba para detectar la no inferioridad en el marco de diferentes tamaños de la muestra. De esta forma se fortalecerían los criterios

para el diseño de este tipo de estudio. La Figura 2 muestra el comportamiento de la distribución del valor P mediante sus cuartilas en el caso en que las varianzas son iguales para dos valores de umbral (d) que



corresponden al 10 y 50% de la variabilidad de la medición en las poblaciones ( $\sigma$ ). Los valores de los percentiles se presentan en función de la diferencia entre las medias de las dos poblaciones ( $\mu_d$ ) representadas, como es usual, con valores relativos a la variabilidad de las poblaciones.

Figura 2. Cuartilas de la distribución del valor P para dos valores de umbral (d) en caso varianzas iguales, cuando  $n=m=30$

La Figura 2 muestra el caso  $n=m=30$ . Contrastando gráficos como el presentado en la Figura 2 permitirá identificar cómo el cambio en los valores de umbral modificaría el comportamiento del valor según los valores de las diferencias entre las poblaciones y el tamaño de la muestra. Adicionalmente al gráfico es posible utilizar (3) y (4) para construir tablas que permitan contrastar, para diferentes valores de umbral y tamaños de muestra, el comportamiento esperado del valor P. Esto podría ser utilizado, también, como un complemento de los procedimientos usuales para determinar el tamaño de la muestra.

### 3. APROXIMACIONES PARA LA DISTRIBUCIÓN DEL VALOR P

Las aproximaciones presentadas por Monterrey [2006] pueden ser adecuadas al caso de las pruebas de no inferioridad:

La aproximación de la distribución t no-central a la distribución central de Chmura Kraemer y Paik

[1979] establece que  $T_v^\Delta(a) \approx T_v^0(a(1 + \frac{\Delta^2}{v})^{\frac{1}{2}} - \Delta(1 + \frac{a^2}{v})^{\frac{1}{2}})$ . Así, para cada i, se obtiene de (1) que,

para  $0 < x < 1$ ,  $F_{\Pi_i/\mu_{d,d}}(x) \approx 1 - T_{v_i}^0(\kappa_{i,1}(x, \mu_{d,d}) - \kappa_{i,2}(x, \mu_{d,d}))$ , con  $\kappa_{i,1}(x, \mu_{d,d}) = t_{v_i}^0(1-x) \left(1 + \frac{\Delta_i^2}{v_i}\right)^{\frac{1}{2}}$  y

$\kappa_{i,2}(x, \mu_{d,d}) = \Delta_i \left(1 + \frac{(t_{v_i}^0(1-x))^2}{v_i}\right)^{\frac{1}{2}}$ . De donde:

$$(0 < x < 1) \quad F_{\Pi_i/\mu_{d,d}}(x) \approx 1 - T_{v_i}^0(t_{v_i}^0(1-x) - \Delta_i) \quad (5)$$

Por lo que  $\gamma \approx 1 - T_{v_i}^0(t_{v_i}^0(1 - \tilde{\Pi}_i/\mu_{d,d}(\gamma)) - \Delta_i)$ ;  $t_{v_i}^0(1-\gamma) \approx t_{v_i}^0(1 - \tilde{\Pi}_i/\mu_{d,d}(\gamma)) - \Delta_i$  y entonces para  $0 < \gamma < 1$  e  $i=1,2$ :

$$\tilde{\Pi}_i/\mu_{d,d}(\gamma) \approx 1 - T_{v_i}^0(t_{v_i}^0(1-\gamma) + \Delta_i) \quad (6)$$

En particular, las cuartilas quedan determinadas por:

$$\tilde{\Pi}_i/\mu_{d,d}(0,25) \approx 1 - T_{v_i}^0(t_{v_i}^0(0,75) + \Delta_i) \quad (7)$$

$$\tilde{\Pi}_i/\mu_{d,d}(0,50) \approx 1 - T_{v_i}^0(\Delta_i) \quad (8)$$

$$\tilde{\Pi}_i/\mu_{d,d}(0,75) \approx 1 - T_{v_i}^0(t_{v_i}^0(0,25) + \Delta_i) \quad (9)$$

Otro criterio de aproximación puede obtenerse utilizando la aproximación de la distribución t no central a la distribución normal de Johnson y Welch [1940]. Esta aproximación establece que:

$$T_v^\Delta(a) \approx \Phi\left(\frac{x-\Delta}{(1+\frac{a^2}{2v})^{1/2}}\right) \quad (10)$$

Denotando  $\Phi(\cdot)$  la función de distribución de la distribución normal estándar,  $n(0,1)$ .

De (1) y (10), para  $0 < x < 1$ , se obtiene, para  $i=1,2$ , que:

$$F_{\Pi_i/\mu_{d,d}}(x) \approx 1 - \Phi\left(\frac{t_{v_i}^0(1-x) - \Delta}{(1 + \frac{(t_{v_i}^0(1-x))^2}{2v_i})^{1/2}}\right) \quad (11)$$

En los ensayos clínicos los tamaños de muestral son relativamente elevados, por eso una primera aproximación de  $F_{\Pi_i/\mu_{d,d}}(x)$  utilizando la distribución normal se puede obtener sustituyendo en (11)

$t_{v_i}^0(1-x) \approx z_{1-x}$ , denotando  $z_{1-x}$  el percentil de orden  $1-x$  de la distribución  $n(0,1)$ , así para cada i:

$$F_{\Pi_i/\mu_{d,d}}(x) \approx 1 - \Phi\left(\frac{z_{1-x} - \Delta_i}{(1 + \frac{z_{1-x}^2}{2v_i})^{1/2}}\right) \quad (12)$$

Como para cada x y cada i,  $\frac{z_{1-x}^2}{2v_i} \approx 0$ ; para cada  $0 < \gamma < 1$   $\gamma \approx 1 - \Phi(z_{1-\tilde{\Pi}_i/\mu_{d,d}(\gamma)} - \Delta_i)$  y entonces

$$\tilde{\Pi}_i/\mu_{d,d}(\gamma) \approx 1 - \Phi(z_{1-\gamma} + \Delta_i) \quad (13)$$

Siendo las cuartilas  $\tilde{\Pi}_i/\mu_{d,d}(0,25) \approx 1 - \Phi(0,674 + \Delta_i)$ ,  $\tilde{\Pi}_i/\mu_{d,d}(0,50) \approx 1 - \Phi(\Delta_i)$  y

$\tilde{\Pi}_i/\mu_{d,d}(0,75) \approx 1 - \Phi(-0,674 + \Delta_i)$ .

Una aproximación más precisa a la distribución normal se puede obtener utilizando la aproximación de Jennett y Welch [1939] referida por Sahai y Ojeda [1998], en ella  $t_v^\Delta(\alpha) \approx \bar{z}_\alpha(\gamma, \Delta)$ , siendo  $\bar{z}_\alpha(v, \Delta) = (1 + (2v)^{-1} z_\alpha^2)^{-1} \Delta b_v \sqrt{1 + (2v)^{-1} (\Delta^2 + z_\alpha^2)}$  y  $b_v = \sqrt{2} \Gamma(\frac{v+1}{2}) / (\sqrt{v} \Gamma(\frac{v}{2}))$ , denotando  $\Gamma(\cdot)$  la función gamma.

Según esta aproximación de (1) y (11) se obtiene:

$$F_{\Pi_i/\mu_{d,d}}(x) \approx 1 - \Phi\left(\frac{\bar{z}_{1-x}(v_i, 0) - \Delta_i}{(1 + \frac{(\bar{z}_{1-x}(v_i, 0))^2}{2v_i})^{1/2}}\right) \quad (14)$$

### 4. RELACIÓN ENTRE EL VALOR P Y LA FUNCIÓN DE POTENCIA

Para una  $\alpha$ -prueba; la función de potencia para las hipótesis de no inferioridad,  $\beta_{i,\alpha}(\mu_{d,d})$  cumple que, para cada i:

$$\beta_{i,\alpha}(\mu_{d,d}) = P_{\mu_{d,d}}(\{E_i > t_{v_i}^0(1-\alpha)\}) = F_{\Pi_i/\mu_{d,d}}(\alpha) \quad (15)$$

Para la determinación del tamaño de muestra se utiliza la ecuación  $\beta_{i,\alpha}(\mu_{d,d}) = \beta$ , denotando  $\beta$  el error de tipo II de la prueba, utilizando (5) se obtiene, para cada i, que:

$$\Delta_i \approx t_{v_i}^0(1-\alpha) - t_{v_i}^0(1-\beta) \quad (16)$$

De (8) y (16) se obtiene la relación entre la mediana del valor P y las tasas de error de la prueba de hipótesis:

$$\tilde{\Pi}_{i/\mu_d,d}(0,50) \approx 1 - T_{v_i}^0(t_{v_i}^0(1-\alpha) - t_{v_i}^0(1-\beta)) \quad (17)$$

Esta relación permite caracterizar el comportamiento del valor P para diferentes tasas de error; lo que puede ser utilizado como una información adicional en la determinación del tamaño de la muestra.

Información que puede ser complementada con las otras cuartiles para disponer así de una idea de la distribución esperada del valor P para el tamaño de muestra seleccionado. Procediendo como en (17) es posible relacionar todas las cuartiles con las tasa de error:

$$\tilde{\Pi}_{i/\mu_d,d}(0,25) \approx 1 - T_{v_i}^0(t_{v_i}^0(0,75) + t_{v_i}^0(1-\alpha) - t_{v_i}^0(1-\beta)) \quad (18)$$

$$\tilde{\Pi}_{i/\mu_d,d}(0,75) \approx 1 - T_{v_i}^0(t_{v_i}^0(0,25) + t_{v_i}^0(1-\alpha) - t_{v_i}^0(1-\beta)) \quad (19)$$

Así en las tablas para el cálculo del tamaño de muestra para los estudios de no inferioridad se puede complementar el análisis con una estimación del comportamiento esperado del valor P utilizando (17), (18), (19) y la definición en términos de percentiles del coeficiente de Galton. Por ejemplo, en el caso en que las varianzas sean iguales, si  $\theta = (\mu_d + d)/\sigma = 0.5$  se puede utilizar la Tabla 3.2.1 en Chow y cols [2003] para calcular diferentes tamaños de muestra según las tasas de error de la prueba. La Tabla 1 muestra los tamaños de muestra en diferentes casos, así como las cuartiles y el coeficiente de asimetría que caracterizarían el comportamiento del valor P.

Tabla 1. Distribución del valor P para distintos tamaños de muestra

error tipo I	potencia	n=m	percentiles			asimetría
			25	50	75	
$\alpha=0.025$	$\beta=0.8$	64	0.0003	0.0028	0.0169	0.70
	$\beta=0.9$	86	0.0001	0.0007	0.0053	0.76
$\alpha=0.05$	$\beta=0.8$	51	0.0010	0.0069	0.0352	0.65
	$\beta=0.9$	70	0.0002	0.0020	0.0125	0.72

Adicionalmente a partir de (13) se obtienen las aproximaciones normales siguientes de las cuartiles:

$$\tilde{\Pi}_{i/\mu_d,d}(0,50) \approx 1 - \Phi(z_{1-\alpha} - z_{1-\beta}) \quad (20)$$

$$\tilde{\Pi}_{i/\mu_d,d}(0,25) \approx 1 - \Phi(0,674 + z_{1-\alpha} - z_{1-\beta}) \quad (21)$$

$$\tilde{\Pi}_{i/\mu_d,d}(0,75) \approx 1 - \Phi(-0,674 + z_{1-\alpha} - z_{1-\beta}) \quad (22)$$

El análisis anterior de la relación entre el comportamiento del valor P y las tasas de error de las pruebas pues complementarse con la determinación del tamaño de muestra para asegurar un comportamiento específico del valor P, así en el caso en que las varianzas sean iguales de (13) se sigue que, cuando  $m=kn$ ,  $n_i = \delta^{-2} (z_{1-m_i} - \theta_i)^2 (k+1)/k$ , denotando  $n_1$  el tamaño de muestra para que percentil 0,25 tenga el valor  $m_1$  y  $(n_2; m_2)$ ,  $(n_3, m_3)$  los valores correspondientes a la mediana y el percentil 0,75 respectivamente, siendo  $\theta_1=0,674$ ;  $\theta_2=0$  y  $\theta_3=0,674$ .

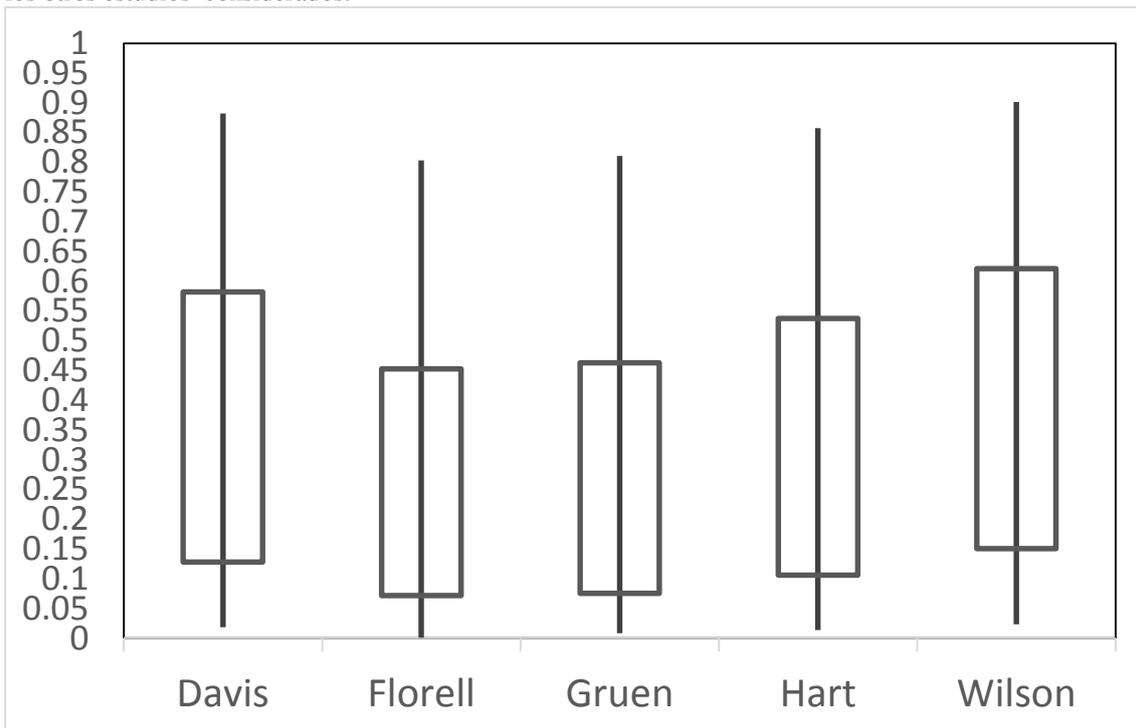
## 5. UNA APLICACIÓN AL META-ANÁLISIS

El conocimiento de la distribución del valor P puede servir para caracterizar el comportamiento de los diferentes estudios que integran un meta-análisis y así disponer de un elemento de juicio adicional para valorar el aporte de cada estudio a la integración de resultados. Tomando, por ejemplo algunos de los estudios del meta-análisis realizado por Munford y cols [1984] y que fue referido por Fleiss [1993]. Para determinar el efecto de la psicoterapia sobre el número de días de hospitalización de los pacientes. La Tabla 2 muestra los resultados de cinco de esos estudios.

Tabla 2. Resultados de los estudios del meta-análisis

Estudio	Psicoterapia			Control			efecto
	n	media	DE	n	media	DE	
Davis	13	5	4.7	13	8.5	3.8	3.50
Florell	30	4.9	1.7	50	6.1	2.3	1.20
Gruen	35	22.5	3.4	35	24.9	10.7	2.40
Hart	20	12.5	1.5	20	12.3	1.7	-0.20
Wilson	8	6.5	0.8	8	7.38	1.4	0.88

Utilizando esos estudios para ejemplificar la utilidad de los resultados que se presentan en la integración de evidencias de diferentes publicaciones (meta-análisis), se calculó el valor P para la prueba bilateral  $H_0: \theta = \mu_c - \mu_P \leq 0$   $H_0: \theta = \mu_c - \mu_P > 0$  ( $d=0$ ), donde la diferencia de medias es la medida del efecto de la psicoterapia. Utilizando los procedimientos de combinación de medidas de efecto para diferentes estudios, según un modelo de efectos fijos se estima  $\hat{\theta}=0.78$ . Al combinar las varianzas se obtiene  $\hat{\sigma}=4.2$ . Considerando estas estimaciones como los valores de los parámetros, se pueden determinar los percentiles de la distribución del valor P para cada uno de los estudios utilizando (6). La Figura 3 muestra los valores de esos percentiles para representar la distribución del valor P para cada uno de los estudios; en ella los bigotes representan los percentiles 5 y 95 y la caja las cuartiles. En cada distribución se inserta el valor P observado en cada estudio. Como se puede apreciar el estudio de Hart es altamente probable suponiendo que  $\theta=0.78$ , los demás estudios se encuentran en la cola inferior de las distribuciones, esto podría ser interpretado como un elemento en contra de que ese valor de  $\theta$  fuera su verdadero valor pues dichos estudios serían altamente improbables en ese caso. Adicionalmente, lo observado pudiera ser interpretado también como una medida de la influencia del estudio de Hart sobre la estimación del efecto y por tanto de lo discordante que serían los resultados de esta investigación respecto a lo encontrado en los otros estudios considerados.



Legenda: ● Valor P observado en el estudio

Figura 3. Representación de la distribución del valor P en cada uno de los estudios

## 6. CONCLUSIONES

El conocimiento explícito de la distribución de probabilidad de los valores P y su relación con la función de potencia permitió ampliar los criterios para el diseño de los estudios de no inferioridad mediante la determinación de la sensibilidad de la prueba frente a cambios en los valores de umbral y los tamaños de la muestra; adicionalmente esta distribución fue utilizada para caracterizar el comportamiento de los diferentes estudios que integran un meta-análisis, brindando un elemento de juicio adicional para valorar el aporte de cada estudio a la integración de resultados

RECEIVED: JUNE, 2016  
 REVISED: DECEMBER, 2016

## REFERENCIAS

- [1] BATTACHARYA B. and HABTZGHI D. (2002): Median of the P value under the alternative hypothesis. *Journal of the American Statistical Association*, 56, 202-206.

- [2] CHOW C.C., SHAO J. and WANG H. (2003): Sample Size Calculations in Clinical Research. Marcel Decker Inc. New York.
- [3] CHMURA KRAEMER H, PAIK M. (1979): A central t approximation to the non-central t distribution. **Technometric**, 21, 357-60.
- [4] D'AGOSTINO R., MASSARO J. and SULLIVAN L. (2003): Non-inferiority trials: design concepts and issues –the encounters of academic consultants in statistics. **Statistics in Medicine**, 22, 169-186.
- [5] FLEISS J.L. (1993): The statistical basis of meta-analysis. **Statistical Methods in Medical Research**, 2, 121-145.
- [6] HAHN S. (2012): Understanding non-inferiority trials. **Korean Journal of Pediatric**, 55, 403-407.
- [7] HEAD S.J., KAUL S., BOGERS J. and KAPPETEIN P. (2012): Non-inferiority study design: lessons to be learned from cardiovascular trials. **European Heart Journal**, doi:10.1093/eurheartj/ehs099.
- [8] JENNETT W. and WELCH B. L. (1939): The control of proportion defective as judged by a single quality characteristic varying on a continuous scale", **Journal of the Royal Statistical Society, Series B**, 6, 80-88.
- [9] JOHNSON N. and WELCH B. (1940): Applications of the non-central t distribution. **Biometrika**, 31, 362-89.
- [10] MONTERREY P. (2012):  $p < 0,05$ , ¿Criterio mágico para resolver cualquier problema o leyenda urbana? **Universitas Scientiarum. Universitas Scientiarum**, 17, 203-215.
- [11] MONTERREY P. (2016): Caracterización de los valores P de las pruebas paramétricas en poblaciones normales. **Investigación Operacional**, 37, 38-47.
- [12] MUMFORD E., SCHLESINGER H., GLASS G., PATRICK C. and CUERDON T: (1984) A new look at evidence about reduced cost of medical utilization following mental health treatment. **The American Journal of Psychiatry**, 141, 1145-1158 .
- [13] SAHAI H. and OJEDA M. (1998): Comparison of approximations to the percentiles of noncentral t,  $\chi^2$  and F distributions, Universidad Veracruzana, Xalapa, México. Disponible en <http://www.uv.mx/personal/mojeda/files/2012/04/Comparisons-of-approximations-to-the-percentiles-of-noncentral-t-ti-and-F-distributions-98.pdf>. Consultado: 06/12/2016 .
- [14] TSONG Y. and Chen W. (2007): Non-inferiority testing beyond simple two-sample comparison. **Journal of Biopharmaceutical Statistics**, 17, 289-308