

Elementos esenciales a considerar en los ensayos de no inferioridad

Rolando Uranga¹, Mayelin Mirabal²

¹. Centro Nacional Coordinador de Ensayos Clínicos. Calle 23, esq. 200. Reparto Atabey. Playa. Ciudad de La Habana, Cuba.
E-mail: rolando@cencec.sld.cu

². Instituto Finlay. Centro de Investigación-Producción de Vacunas. Ave.27 No. 19805. La Lisa. Ciudad de La Habana, Cuba.
E-mail: mmirabal@finlay.edu.cu

En algunas situaciones es común comparar una vacuna experimental con una vacuna estándar sin pretender mostrar superioridad. La hipótesis nula usual de igualdad es inapropiada y conlleva a dificultades lógicas. Con este artículo se pretende esclarecer la problemática al respecto, revisando la metodología desarrollada para estos efectos y presentando un valor conocido como límite de no inferioridad. Se discuten dos métodos para la realización de una prueba de hipótesis de no inferioridad. Se introducen además dos conceptos útiles en este tipo de estudio: preservación y descuento.

Palabras clave: límite de no inferioridad, diseño de superioridad, preservación, descuento.

Introducción

En la práctica clínica es común proponerse ensayos con el objetivo de evaluar la eficacia de una vacuna experimental contra un control activo, con el fin de mostrar que la nueva vacuna no es inferior desde el punto de vista clínico o estadístico a una vacuna comercial. Ensayos con objetivos como el anterior son conocidos como de equivalencia unilateral o de no inferioridad.

La idea de una correcta formulación de las hipótesis en un ensayo de no inferioridad ha sido frecuentemente obviada en el diseño de tales estudios. En 1982 Blackwelder (1) alertó sobre esta problemática y fundamentó que podía estar dada por dos razones principales:

1. El desconocimiento por parte de los investigadores de las desventajas del uso de las pruebas de hipótesis clásicas para este tipo de estudio
2. La imposibilidad de definir *a priori* una cantidad que marque una diferencia clínicamente aceptable entre los tratamientos que se comparan.

Con el fin de incentivar el uso de los diseños de no inferioridad, en este trabajo se explican las desventajas de las pruebas de hipótesis usuales cuando se compara una vacuna experimental con un control activo y se proponen opciones para la selección de un límite de no inferioridad adecuado. Se introducen, además, dos conceptos característicos de los diseños de no inferioridad: preservación y descuento, y se ilustran numéricamente.

Motivaciones, prueba de hipótesis e intervalo de confianza para un ensayo de no inferioridad

Si se aplicara la formulación usual para comparar una vacuna experimental con una vacuna estándar o control activo, se obtendría la prueba de hipótesis siguiente:

$$H_0 : \pi_s \leq \pi_e \quad H_1 : \pi_s > \pi_e \quad (I)$$

donde: π_s , π_e son las proporciones de éxito en los grupos de la terapia estándar y experimental respectivamente.

Desde el punto de vista de la teoría de las pruebas de hipótesis, tal formulación es incorrecta. Esto se debe a que lo que se desea probar aparece especificado en la hipótesis nula; o sea, se pretende “probar” una hipótesis nula. El carácter contradictorio de este proceder se justifica por el hecho de que el nivel de significación o error de tipo I es el único parámetro que puede ser rigurosamente controlado en toda prueba de hipótesis, y sólo garantiza confiabilidad de los resultados en el caso en que se rechace la hipótesis nula.

Con el fin de vencer esta dificultad se pudiera pensar en sustituir la prueba de hipótesis anterior por la siguiente:

$$H_0 : \pi_s \geq \pi_e \quad H_1 : \pi_s < \pi_e \quad (II)$$

Se observa entonces otro inconveniente: estamos en presencia de una prueba de superioridad. Nuestro objetivo es demostrar que la terapia experimental no es inferior a la estándar, y no que es superior.

Existe una solución curiosa a este dilema. La idea comienza con la introducción de una nueva cantidad, que denotaremos por la letra griega δ , y que se conoce como límite de equivalencia o no inferioridad. Se sugiere entonces aplicar la siguiente prueba:

$$H_0 : \pi_s \geq \pi_e + \delta \quad H_1 : \pi_s < \pi_e + \delta \quad (III)$$

Estamos en presencia ahora de un diseño de no inferioridad. Blackwelder (1) mostró el estadígrafo de esta prueba y el tamaño muestral necesario para cada grupo cuando se usan las hipótesis especificadas.

Con el uso de este diseño surge un nuevo problema: la elección de un valor adecuado para δ . Este aspecto es uno de los más difíciles de definir cuando se plantea un estudio de este tipo.

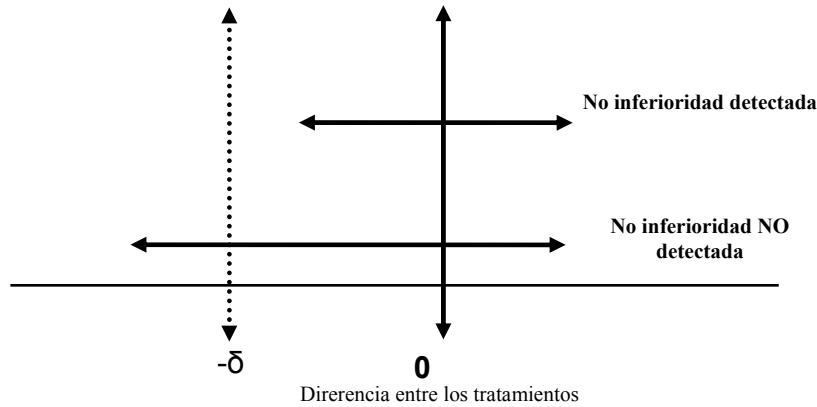
Es útil reportar los resultados de un ensayo de no inferioridad en términos de intervalos de confianza porque, además de tomar una

decisión dicotómica, permite estimar la cantidad desconocida, en este caso la diferencia entre las proporciones de éxito de la terapia experimental y estándar.

La fórmula para dicho intervalo de confianza sería:

$$IC(\pi_e - \pi_s) = [p_e - p_s - Z_{1-\alpha}SE; 1]$$

Figura 1. Relación entre intervalos de confianza y prueba de hipótesis para un diseño de no inferioridad.



Estrategia para seleccionar un límite de no inferioridad

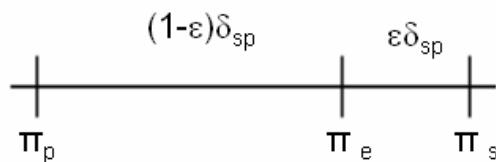
En el marco de los ensayos clínicos y con el fin de seleccionar un límite de no inferioridad adecuado es preciso haber realizado con anterioridad al menos un ensayo de superioridad +de la terapia estándar con respecto al placebo.

Para la realización de una prueba de no inferioridad es necesario la especificación, explícita o implícitamente, del límite de no inferioridad δ .

El enfoque explícito sería estimar δ como una fracción de la magnitud del efecto estándar-placebo.

intuitivamente, como se muestra en la Figura 2, el límite de no inferioridad δ debe ser una fracción del efecto de la terapia estándar sobre el placebo (δ_{sp}), es decir, $\delta = \varepsilon\delta_{sp}$, donde $0 < \varepsilon \leq 1$ y $\delta_{sp} = \pi_s - \pi_p$, ya que en una prueba de no inferioridad es lógico esperar que la respuesta en el grupo experimental se encuentre en una posición intermedia entre la respuesta del placebo y la del comparador activo (2).

Figura 2. Interpretación de δ como una fracción de δ_{sp} .



En este caso, el rechazo de la hipótesis nula equivale al hecho de que la cantidad $-\delta$ esté a la izquierda del límite inferior del intervalo de confianza. Por otro lado, si el intervalo de confianza contiene al valor $-\delta$, entonces no se rechaza la hipótesis nula (ver Figura No. 1. Además, el intervalo de confianza estima a la cantidad $\pi_e - \pi_s$, lo cual brinda más información.

Tie-Hua Ng (3) justifica como para un ensayo de no inferioridad $0 < \varepsilon < 0,5$. Esto es lógico pues para concluir no inferioridad, la respuesta de la terapia experimental debe quedar más próxima al estándar que al placebo.

Algunos autores (3-6) sugieren elegir:

$$\delta = 0,2\delta_{sp}$$

En este caso la terapia experimental preserva el 80% del efecto δ_{sp} en caso de concluirse equivalencia. Ver Figura No. 2. En este sentido el criterio es razonable. En general a la cantidad $(1 - \varepsilon) \times 100\%$ se le conoce como por ciento de preservación del efecto estándar-placebo.

Un mayor δ puede estar justificado si el objetivo es establecer la eficacia del tratamiento experimental comparado con placebo a través de su comparación con la terapia estándar, y no concluir equivalencia o no inferioridad con respecto a esta terapia. Brian L. Wiens (7) sugiere elegir $\varepsilon = 0,5$ para este caso. De hecho, ε puede llegar a ser uno cuando se quiere comprobar superioridad de la terapia en estudio con respecto al placebo.

Si $\varepsilon = 1$ y si se tiene $H_0 : \pi_e - \pi_s \leq -\delta$ (Hipótesis nula para la prueba de no inferioridad.), se toma $\delta = \varepsilon\delta_{sp}$, $\delta_{sp} = \pi_s - \pi_p$, y se sustituye, se obtiene la hipótesis nula de la prueba de superioridad (de la terapia en estudio con respecto al placebo):

$$H_0 : \pi_e \leq \pi_p .$$

Un método común de estimación del efecto estándar-placebo (δ_{sp}) es construir, a partir de datos históricos, un intervalo de confianza para la diferencia $\pi_s - \pi_p$ y elegir el límite inferior de este intervalo de confianza como un estimador conservador de la verdadera diferencia δ_{sp} (3). Es decir, $\hat{\delta}_{sp} = p_s - p_p - z_\alpha .SE$, donde p_s y p_p son respectivamente estimaciones puntuales de las proporciones de éxito para las terapias estándar y placebo obtenidas de estudios anteriores.

El método de estimación de δ_{sp} que se ha descrito, y por tanto la elección de δ , asume que la diferencia esperada entre las respuestas de los tratamientos es invariable a través de los diversos estudios. Esta condición se conoce como supuesto de invariabilidad (constancy assumption).

Existe un segundo enfoque (enfoque implícito) para la realización de una prueba de no inferioridad que no exige la estimación directa del efecto estándar-placebo ni la selección directa de delta (3, 8). En este enfoque se construye una expresión que describe la hipótesis nula de la prueba, un estadígrafo para ella y una regla de decisión. El estadígrafo se calcula a partir de una estimación de la expresión mencionada, utilizando los datos del ensayo actual y de estudios históricos. La estructura de delta como una fracción del efecto estándar-placebo queda incluida en la expresión para la hipótesis nula de la prueba, y en la estimación utilizada por el estadígrafo intervienen los estimadores puntuales de las respuestas estándar-histórica, placebo-histórica, estándar-actual y experimental.

Un concepto asociado a situaciones en las que no se cumple el supuesto de invariabilidad entre los diferentes estudios es el concepto de "Descuento". A continuación se introduce este concepto en dos escenarios diferentes, uno en el que se asume que la preservación del efecto estándar-placebo es nula (superioridad con respecto al placebo) y otro más general en el que se admite un porcentaje de preservación distinto de cero (no inferioridad).

Primer escenario: Descuento sin preservación

Supongamos que $\varepsilon = 1$, es decir, se pretende evaluar indirectamente la eficacia de la terapia experimental con respecto al placebo mediante su comparación con una terapia estándar ya probada en el supuesto extremo.

La hipótesis nula de eficacia, $H_0 : \pi_e \leq \pi_p$, se puede

reescribir como: $H_0 : (\pi_e - \pi_s) + (\pi_s - \pi_p) \leq 0$ que, bajo la suposición $\pi_s - \pi_p = \pi_s^* - \pi_p^*$, donde

$\pi_s^* - \pi_p^*$ proviene de datos históricos (o sea, bajo el supuesto

de invariabilidad), se describe como:

$$H_0 : (\pi_e - \pi_s) + (\pi_s^* - \pi_p^*) \leq 0 .$$

Supóngase que, realmente, $\pi_s - \pi_p = \gamma(\pi_s^* - \pi_p^*)$, donde $0 < \gamma \leq 1$ (situación de descuento del tamaño del efecto estimado a partir de estudios previos, donde en general no se cumple el supuesto de invariabilidad). $\gamma = 1$ es la situación de no descuento. Si hay un $\gamma\%$ de descuento, entonces $\gamma = 1 - \gamma\%$.

Se tiene:

$$\delta = \varepsilon \delta_{sp} = \pi_s - \pi_p = \gamma(\pi_s^* - \pi_p^*)$$

para una situación en que se evalúa eficacia respecto al placebo (o sea, $\varepsilon = 1$), y se incluye un factor de descuento $1 - \gamma$.

Para la realización de la prueba de hipótesis de no inferioridad, dos posibles enfoques serían:

1. Calcular $\hat{\delta} = \gamma \hat{\delta}_{sp} = \gamma [p_s - p_p - z_\alpha .SE]$ y realizar la prueba de no inferioridad usando $\hat{\delta}$.
2. Realizar la prueba de la hipótesis nula $H_0 : \pi_t - \pi_s + \gamma(\pi_s^* - \pi_p^*) \leq 0$ que resulta al sustituir la expresión teórica hallada de δ . El miembro izquierdo puede ser estimado a partir de los datos del presente ensayo y de los datos históricos. Puede definirse un estadístico de la prueba correspondiente.

Segundo escenario: Descuento con preservación

Una situación en la cual hay preservación de cierto porcentaje $1 - \varepsilon$ del efecto estándar-placebo, y que considera además un factor de descuento $1 - \gamma$, se formaliza del siguiente modo:

$$\delta = \varepsilon \delta_{sp} = \varepsilon(\pi_s - \pi_p) = \varepsilon \gamma(\pi_s^* - \pi_p^*)$$

La cantidad $1 - \varepsilon$ se denomina por ciento de preservación.

Los dos métodos de realización de la prueba de hipótesis de no inferioridad varían del modo obvio:

1. Calcular $\hat{\delta} = \varepsilon \gamma \hat{\delta}_{sp} = \gamma \varepsilon [p_s - p_p - z_\alpha .SE]$, y luego probar no inferioridad usando $\hat{\delta}$.
2. Prueba de la hipótesis $H_0 : \pi_t - \pi_s + \varepsilon \gamma(\pi_s^* - \pi_p^*) \leq 0$.

Si la hipótesis nula se rechaza en cualquiera de estos enfoques, se puede afirmar que el tratamiento experimental preserva más del $100(1 - \varepsilon)\%$ del efecto con un $100(1 - \gamma)\%$ de descuento.

Ejemplo (Tomado de Tie Hua Ng (4))

Supongamos que $\varepsilon\gamma = 0,72$. Usando el primer método se calcula

$\hat{\delta} = 0,72\hat{\delta}_{sp} = 0,72 [p_s - p_p - z_\alpha \cdot SE]$; usando el

segundo método realizamos la prueba de la hipótesis nula para

$H_0 : \pi_t - \pi_s + 0,72(\pi_s^* - \pi_p^*) \leq 0$. Si esta hipótesis

nula es rechazada con cualquiera de los dos enfoques anteriores, se puede afirmar que la terapia experimental preserva más del 28% del efecto estándar-placebo con cero descuento ($\varepsilon = 0,72$, $\gamma = 1$).

O bien se puede afirmar que la terapia experimental es eficaz (con cero preservación) con un 28% de descuento ($\varepsilon = 1$, $\gamma = 0,72$).

O bien se puede afirmar que la terapia experimental preserva más del 10% del efecto con un 20% de descuento ($\varepsilon = 0,9$, $\gamma = 0,8$).

Consideraciones finales

Una interpretación válida de un estudio de no inferioridad depende de los supuestos de que el control activo es efectivo (sensibilidad), y de que el tamaño del efecto es el mismo a través de los estudios (invariabilidad).

Existen asimismo diseños de equivalencia bilaterales que necesitan de la especificación de los límites, uno inferior (no inferioridad) y uno superior. Este tipo de estudio no ha sido tratado en este artículo.

La prueba de hipótesis de no inferioridad puede efectuarse no sólo para comparar proporciones y diferencias entre ellas, sino también para comparar medias y cocientes (9,10).

Se puede realizar un diseño con tres grupos: estándar, experimental y placebo. En este caso no se necesita del supuesto de invariabilidad porque no se requiere de datos históricos pues no se necesita la especificación del límite de no inferioridad.

Se puede hacer previamente un estudio pequeño para estimar el efecto estándar-placebo y posteriormente, realizar el diseño de equivalencia o no inferioridad.

Essential elements to consider in non-inferiority studies

Abstract

In some situations, it is usual to compare an experimental vaccine to a standard one without the objective of showing superiority. The common null hypothesis of equality is inappropriate and leads to logical difficulties. The aim of this work is to clarify some issues about this problem, reviewing the methodology developed for these purposes and presenting a value known as non-inferiority limit. Two methods for conducting a hypothesis test of non-inferiority are discussed and two useful concepts: Preservation and Discounting are also introduced.

Keywords: Non-inferiority limit, superiority design, preservation, discounting.

Matemáticamente, la preservación y el descuento son indistinguibles, pero el uso del descuento como una forma de preservación puede dar lugar a confusiones.

El primer método descrito para realizar la prueba de no inferioridad requiere la determinación de delta utilizando el límite inferior de confianza del efecto calculado a partir de datos históricos mientras el segundo método no precisa la determinación explícita de delta.

Referencias

1. William C. Blackwelder. "Proving the null hypothesis" in clinical trials. *Controlled Clinical Trials* 1982; 3:345-353.
2. Ralph B. D'Agostino, Joseph M. Massaro, and Lisa M. Sullivan Non-inferiority trials. Design concepts and issues – the encounters of academic consultants in statistics. *Statistics in Medicine* 2003; 22:169–186.
3. Tie-Hua Ng. Choice of delta in equivalence testing. *Drug Information Journal* 2001; 35: 1517-1527.
4. Tie-Hua Ng. A Specification of treatment difference in the design of clinical trials with active controls. *Drug Information Journal* 1993;27:705-719.
5. Tie-Hua Ng. Active control equivalence studies. *Proceedings of the biopharmaceutical section. American Statistical Association* 1997; 124-128
6. Tie-Hua Ng. statistical issues in equivalence testing-FDA reviewer's perspectives. *Proceedings of the biopharmaceutical section. American Statistical Association* 1999;209-213.
7. Brian L. Wiens. Choosing an equivalence limit for non-inferiority or equivalence studies. *Controlled Clinical Trials* 2002; 23:2-14.
8. Hauck WW, Anderson S. Some issues in the design and analysis of equivalence trials. *Drug Information Journal* 1999; 33:109-118.
9. Larry L. Laster, Mary F. Johnson. Non-inferiority trials: the 'at least as good as' criterion. *Statistics in Medicine* 2003; 22:187–200.
10. H. M. James Hung, Jane Wang, Yi Tsong, John Lawrence and Robert T. O'Neil. Some fundamental issues with non-inferiority testing in active controlled trials. *Statistics in Medicine* 2003; 22:213–225.