

SECCIÓN ESPECIAL

LA MUESTRA EN ENSAYOS CLÍNICOS
ALEATORIZADOS CON ANÁLISIS INTERINOMichelle Saaibi Meléndez^{1,2,a}, Felipe Botero-Rodríguez^{1,2,a},
Carlos Javier Rincón Rodríguez^{1,2,b}¹ Semillero de Bioestadística, Facultad de Medicina, Pontificia Universidad Javeriana, Bogotá, Colombia.² Departamento de Epidemiología Clínica y Bioestadística, Pontificia Universidad Javeriana, Bogotá, Colombia.^a Médico general; ^b estadístico, magister en Epidemiología Clínica.

RESUMEN

En este artículo se introducen los ensayos clínicos aleatorizados y conceptos básicos de la inferencia estadística. Se presenta como calcular el tamaño de muestra por tipo de desenlace e hipótesis a probar, junto con el código en el lenguaje de programación R para realizar su aplicación. Se presentan cuatro métodos para realizar el ajuste del tamaño de muestra original, cuando se planean análisis interinos. De una manera sencilla y concreta se busca realizar una introducción a estos temas, considerando las expresiones matemáticas que soportan los resultados y su implementación en programas estadísticos disponibles. Con el fin de acercar a los estudiantes de áreas de la salud a la estadística y al uso de programas estadísticos, aspectos poco considerados en su formación.

Palabras clave: Tamaño de la Muestra; Ensayos Clínicos; Pruebas de Hipótesis (fuente: DeCS BIREME).

SAMPLES IN RANDOMIZED CLINICAL TRIALS WITH
INTERIM ANALYSIS

ABSTRACT

This article introduces randomized clinical trials and basic concepts of statistical inference. We present methods for calculating the sample size by outcome type and the hypothesis to be tested, together with the code in the R programming language. We describe four methods for adjusting the original sample size for interim analyses. We sought to introduce these topics in a simple and concrete way, considering the mathematical expressions that support the results and their implementation in available statistical programs; therefore, bringing health students closer to statistics and the use of statistical programs, which are aspects that are rarely considered during their training.

Keywords: Sample Size; Clinical Trials; Hypothesis Tests (source: MeSH NLM).

Citar como: Saaibi Meléndez M, Botero-Rodríguez F, Rincón Rodríguez CJ. La muestra en ensayos clínicos aleatorizados con análisis interino. Rev Peru Med Exp Salud Publica. 2023;40(2):220-8. doi: [10.17843/rpmesp.2023.402.12217](https://doi.org/10.17843/rpmesp.2023.402.12217).

Correspondencia:
Michelle Saaibi Meléndez;
msaaibi@javeriana.edu.co

Recibido: 11/10/2022
Aprobado: 26/04/2023
En línea: 30/06/2023



Esta obra tiene una licencia de Creative Commons Atribución 4.0 Internacional

INTRODUCCIÓN

El enfoque de la medicina ha cambiado desde una mirada inicial paternalista hasta el reduccionismo pragmático. Este cambio ocurrió por la búsqueda de mejorar la calidad de la atención, disminuir los incentivos económicos individuales y priorizar la importancia de la investigación para mejorar la calidad de la evidencia ⁽¹⁾. La medicina basada en la evidencia surge como un nuevo paradigma en la década de los 90 como sustento científico para la toma de decisiones clínicas y se fundamenta en una jerarquía de tres afirmaciones: a) los ensayos clínicos aleatorizados (ECA) o revisiones sistemáticas de muchos experimentos suelen aportar mayor evidencia que los estudios observacionales; b) los estudios clínicos analíticos aportan mejor evidencia que la sola racionalidad fisiopatológica; y c) los estudios clínicos analíticos aportan más evidencia que el juicio de expertos ⁽²⁾.

Obtener resultados válidos de los ECA depende de la calidad de los datos, la cual debe ser suficiente para abordar la pregunta planteada. Para obtener estos datos de calidad, se debe tener un tamaño de muestra lo suficientemente grande para obtener una estimación precisa del efecto de la intervención. Los errores aleatorios no van a afectar la interpretabilidad de los resultados, siempre y cuando, la muestra sea lo suficientemente grande, sin embargo, un error sistemático puede invalidar un estudio ⁽³⁾.

El análisis interino consiste en plantear un(os) punto(s) de observación, donde se realizará una evaluación del comportamiento de la muestra hasta ese punto. Dependiendo de los resultados, el comité podrá determinar la pertinencia de seguir con el estudio o finalizarlo en ese momento⁽³⁾. En este artículo se busca realizar una introducción al cálculo del tamaño de muestra por tipo de desenlace e hipótesis. Además de su ajuste por análisis interino, considerando las fórmulas matemáticas y su implementación en programas estadísticos disponibles como el lenguaje de programación R. El objetivo es acercar al personal de la salud a la estadística y al uso de programas, aspectos poco considerados en su formación. Aunque, ya existen diversas fuentes donde se desarrollan los temas anteriores, no existen muchos documentos donde se integre tanto la teoría como la práctica, incluyendo todos los aspectos mencionados anteriormente en los ECA en paralelo. Contar con artículos de revisión en el idioma español, permite a los jóvenes investigadores y profesionales del área de la salud realizar una primera aproximación a estos temas, sin que se genere un rechazo inicial por su complejidad, en un diseño frecuentemente requerido. Conectar las expresiones matemáticas con su implementación en un programa estadístico busca evitar que, nuevamente jóvenes investigadores ejecuten funciones preestablecidas como por ejemplo *TwoSampleMean.Equality* o *TwoSampleMean.NIS* (incluidas en el paquete «TrialSize»⁽⁴⁾) sin entender de donde se obtienen los resultados que «arroja» el programa, el efecto de los parámetros en el tamaño de muestra o la necesidad de escoger parámetros con valores coherentes con el tipo de hipótesis que se está evaluando. Lo anterior, busca fomentar la comprensión sobre la ejecución mecánica de tareas solo para cumplir el requerimiento de un comité evaluador.

Ensayos clínicos aleatorizados

El principio de equipoise corresponde a un estado de incertidumbre frente a los resultados terapéuticos de un tratamiento, donde se justifica la realización de un ECA⁽⁵⁾. Los ECA con grupo control son estudios prospectivos que comparan los desenlaces de alguna(s) intervención(es) con la mejor alternativa disponible. En estos estudios siempre debe primar la seguridad del paciente, por lo que se les debe explicar los posibles beneficios, daños y las alternativas de tratamiento que existen para su condición. Aunque puede tener limitaciones, se considera la mejor alternativa para evaluar la eficacia o seguridad de una intervención^(6,7). Se caracteriza por: a) hay una intervención y se compara con un grupo control que puede ser placebo o el tratamiento habitual, b) hay asignación aleatorizada de las intervenciones en la población para reducir un posible sesgo de confusión al obtener grupos homogéneos y un posible sesgo de selección al evitar prever el grupo al cual se asigna el paciente c) el cegamiento de los grupos del tratamiento se puede realizar

tanto para los investigadores, pacientes o analistas y esto minimiza posibles sesgos de información^(6,7).

Los ECA se han dividido en cuatro fases. La fase I busca determinar posibles efectos tóxicos, absorción, distribución y metabolismo del fármaco en un grupo de 20-80 personas sanas. La fase II se realiza en población enferma para determinar la seguridad y eficacia del fármaco, a partir de marcadores biológicos y evaluando reacciones adversas. La fase III se realiza cuando hay evidencia sobre la seguridad y eficacia de la intervención y se busca información adicional sobre la seguridad y efectividad del fármaco en un mayor número de participantes. Se compara la intervención con la terapia habitual o placebo en un seguimiento a largo plazo para poder identificar posibles efectos secundarios. En la fase IV, después de que la molécula ha sido aprobada para su comercialización, se compara con otros productos existentes en población general; se realiza farmacovigilancia buscando eventos adversos no identificados en la fase III debido a su baja incidencia o a prolongados periodos de aparición^(3,6).

En este artículo nos enfocaremos en los estudios de fase III y IV, donde se requiere un cálculo del tamaño de muestra. Adicionalmente, trabajaremos los ECA en paralelo caracterizados por un seguimiento simultáneo a cada grupo al cual fueron asignados⁽³⁾.

Estadística inferencial

La estadística inferencial permite estimar el comportamiento de toda la población a partir de los resultados obtenidos en una muestra. Este comportamiento, se resume en medidas como medias, proporciones o varianzas, que de ser obtenidas en toda la población, se denominarían parámetros⁽⁷⁾. Se tienen dos alternativas: los intervalos de confianza y las pruebas de hipótesis; con resultados consistentes, la primera busca un rango de valores que, con un grado de confianza, contenga el parámetro de interés, mientras que la segunda, evalúa una afirmación sobre el parámetro de interés tomando la decisión de rechazarla o no.

Dado que este documento presenta el cálculo del tamaño de muestra en los ECA en paralelo para evaluar afirmaciones sobre parámetros, describiremos el proceso de las pruebas de hipótesis. Inicialmente se plantean dos hipótesis, la hipótesis nula (H_0) que es una afirmación sobre el parámetro, y la hipótesis alternativa (H_a) que es su negación; casi siempre se busca comprobar la hipótesis alternativa⁽⁸⁾ que se relaciona con la pregunta de investigación; al final se decide rechazar o no la H_0 . Teniendo en cuenta que esta decisión depende de los resultados obtenidos únicamente de una muestra, existe la probabilidad de cometer dos errores, el error tipo I o nivel de significancia (α) que ocurre al rechazar la H_0 cuando es verdadera y el error tipo II (β), ocurre al no rechazar la H_0 siendo esta falsa⁽⁹⁾. Lo opuesto al error tipo I es el nivel de confianza ($1-\alpha$) y corresponde a la probabilidad de no rechazar la H_0 cuando esta es verdadera, y el opuesto del error tipo

II ($1-\beta$), que es el poder, es la probabilidad de rechazar la H_0 cuando está es falsa ⁽⁹⁾. Al realizar una prueba de hipótesis, se busca que la probabilidad de cometer el error tipo I y II sea baja, lo cual, implica que el nivel de confianza y poder tengan una probabilidad alta (típicamente: $\alpha=0,05$ y $\beta=0,1$ o $0,2$). Para poder garantizar estos valores se requiere realizar el cálculo del tamaño de muestra.

Para tomar la decisión de rechazar la H_0 , se realiza una operación a partir de los valores observados en la muestra (estadístico de prueba) y se contrasta con el comportamiento que debería presentar si la H_0 fuese verdadera. Si el valor observado por el estadístico de prueba es poco probable, esto evidencia que la H_0 es falsa y se rechaza en favor de la H_a , de lo contrario se dice que no hay evidencia para rechazar la H_0 . La probabilidad que refleja esta evidencia «a favor» o «en contra» de la H_0 se denomina valor de p ^(10,11), y es igual a la probabilidad, asumiendo que la hipótesis nula es verdadera, de obtener un valor del estadístico de prueba «...tan extremo o más (en la dirección adecuada de la H_a) que el valor calculado en realidad» ^(11,12); finalmente se rechaza la H_0 bajo la premisa de un valor de $p < \alpha$. La significancia estadística, comúnmente evaluada mediante el valor de p , no da cuenta de la significancia clínica; hablamos de significancia estadística cuando se cumple la premisa de un valor de $p < \alpha$, mientras que la significancia clínica es definida por aquellos resultados que mejoren la funcionalidad física, mental y social del paciente, que puede llegar a ocasionar una mejoría en su calidad o cantidad de vida, dependiendo del contexto ⁽¹⁴⁾.

Tipos de hipótesis

En un ECA, se pueden tener diferentes preguntas de investigación que se relacionan con cuatro formas distintas de plantear la H_0 . El cálculo del tamaño de muestra depende del tipo de hipótesis a probar, por lo tanto, en la Tabla 1 se presentan sus definiciones junto con un ejemplo.

Tamaño de muestra

Generalmente, no es posible observar a toda la población, por lo tanto, se requiere de una muestra de tamaño específico (n) que represente su comportamiento. A medida que el tamaño de muestra va aumentando, los resultados se aproximan al de la población, a tal fin que a partir de un tamaño específico, los resultados no van a presentar grandes cambios haciendo innecesario seguir recolectando a participantes ⁽¹⁵⁾. Reclutar más sujetos de lo necesario aumenta tanto la complejidad en la operación logística como los costos, y plantea un dilema ético al asignar innecesariamente sujetos a un tratamiento que no ha probado su beneficio. Por otro lado, definir un tamaño de muestra muy pequeño, implica un alto riesgo de que se presenten el error tipo II mencionado anteriormente. El cálculo del tamaño de muestra permite determinar si un estudio es factible basado en supuestos *a priori*, dados por el poder, la significancia y antecedentes de

estudios previos que abordan la misma pregunta de investigación, teniendo en cuenta las consideraciones éticas de someter personas a un experimento ^(13,16).

Adicionalmente, en la realización de un ECA surge la posibilidad de ir observando los resultados obtenidos a medida que se va recolectando la muestra. Lo anterior se denomina «análisis interinos», los cuales deben ser planeados desde el inicio de la investigación en la elaboración del protocolo. Estos análisis adicionales aumentan la posibilidad de que se presenten los errores tipo I y II, por tal motivo, se debe ajustar el tamaño de muestra para mantener un nivel de confianza y poder global en toda la realización del ECA. Lo anterior, refleja la importancia del cálculo del tamaño de muestra, por lo tanto, en este artículo se presenta como realizar el cálculo de tamaño de muestra en los ECA mostrando la expresión de donde se obtiene y su aplicación utilizando el lenguaje de programación R ⁽¹⁷⁾. Adicionalmente, se presenta como realizar el ajuste por análisis interinos junto con un ejemplo.

MATERIALES Y MÉTODOS

A partir de la revisión del libro «*Sample size calculations in clinical research*» realizado por Chow *et al.* ⁽¹³⁾, en este artículo se presenta como calcular el tamaño de muestra para un ECA en paralelo, por: 1) tipo de desenlace (dicotómico, continuo) y 2) tipo de hipótesis a evaluar (de igualdad, no inferioridad, superioridad y equivalencia). Se incluyen las expresiones matemáticas correspondientes y el código para crear una función en los programas R ⁽¹⁷⁾ y RStudio ⁽¹⁸⁾. Para el uso de este código, se requiere que el lector tenga un conocimiento básico sobre el uso de estos programas, en donde debe copiar la función y ejecutarla; ya creada la función se puede utilizar incluyendo los parámetros requeridos que se describen en la sección de resultados. Para cada escenario, se incluye un ejemplo con datos ficticios y se mencionan consideraciones específicas relacionadas con los parámetros de las funciones.

A continuación, se describen los métodos de Pocock, O'Brien y Fleming, Wang y Tsiatis e Inner Wedge para realizar el ajuste del tamaño de muestra original, obtenido a partir de las funciones creadas anteriormente, para realizar el análisis interino. El ajuste consiste en multiplicar el tamaño de muestra original por los coeficientes incluidos en los Anexos 1 a 4 dependiendo del método que se utilice, y considerando el número de evaluaciones planeadas (R), el poder y nivel de significancia definido para el estudio. Adicionalmente, se incluye la expresión del estadístico de prueba que se utiliza en cada evaluación por tipo de desenlace, con base en la información de los participantes que van ingresando al estudio. En resumen, de cada método se presentan: 1) los valores críticos que corresponden a los valores de la distribución normal estándar que determinan la zona de rechazo

Tabla 1. Tipos de hipótesis en los ensayos clínicos aleatorizados.

Tipo	Definición ⁽²⁴⁾	Hipótesis ^(25,26)	Ejemplo	
			Hipótesis	Interpretación
Igualdad	Evalúa si hay diferencias entre el grupo tratamiento y grupo control.	H_0 : No hay una diferencia entre las dos terapias	La presión sobre la proyección esternal estimada de la válvula aórtica en el esternón no se asocia con un cambio en los parámetros hemodinámicos del paciente hipotenso.	Los pacientes a quienes se les realizó una presión de 6 mm de profundidad sobre la proyección esternal estimada de la válvula aórtica en el esternón, mantenida por 90 s, determina una evolución decreciente de forma homogénea en los parámetros de tensión arterial y frecuencia cardíaca ⁽²⁷⁾ .
		H_a : Hay una diferencia entre las dos terapias	La presión sobre la proyección esternal estimada de la válvula aórtica en el esternón se asocia con un cambio en los parámetros hemodinámicos del paciente hipotenso.	
No inferioridad	Evalúa si el efecto de un tratamiento nuevo (cuyo efecto es inferior del tratamiento convencional, pero mayor que el placebo), se encuentra dentro de un rango aceptado y, se establece a partir de la mejor evidencia disponible. Esta diferencia se justifica por efectos secundarios o facilidades para su consecución.	H_0 : El efecto de la nueva intervención es menor o igual al placebo	El nuevo anticomicial presenta la misma efectividad que el placebo.	El nuevo anticomicial, aunque es mejor tolerada que la terapia convencional, es menos efectiva clínica y estadísticamente, por lo que no puede ser recomendado como primera línea ⁽²⁸⁾ .
		H_a : El efecto de la nueva intervención es mayor que el placebo	El nuevo anticomicial presenta mayor efectividad que el placebo.	
Superioridad	Busca evaluar si una nueva intervención genera mejores desenlaces clínicos respecto a una terapia bien establecida o placebo	H_0 : La nueva intervención no es superior a la terapia establecida	El voluntariado no reduce el aislamiento social ni impacta en mejores desenlaces en salud mental.	El voluntariado no demostró ser superior comparado con el grupo control respecto a desenlaces en salud mental o aislamiento ⁽²⁹⁾ .
		H_a : La nueva intervención es superior a la terapia establecida	El voluntariado reduce el aislamiento social e impacta en mejores desenlaces en salud mental.	
Equivalencia	Busca evaluar si el efecto del tratamiento es idéntico al de otra terapia.	H_0 : Las terapias no son equivalentes	La inclusión en el tratamiento para el síndrome de ovario poliquístico de metformina, asociada con anticonceptivos orales, no es igual de efectivo que la monoterapia con solo anticonceptivos orales.	Se evidenció una mayor efectividad en la terapia combinada dado por una remisión ecográfica más corta, así como menores síntomas, y una tasa de recurrencia a los 3 meses inferior en comparación al grupo de estudio que recibió monoterapia ⁽³⁰⁾ .
		H_a : Las terapias son equivalentes.	La administración de monoterapia con anticonceptivos orales es igual de efectivo que la terapia de anticonceptivos orales más metformina para el tratamiento del síndrome de ovario poliquístico.	

H_0 : hipótesis nula, H_a : hipótesis alternativa

para evaluar la hipótesis nula en cada momento, y 2) los coeficientes para el ajuste del cálculo del tamaño de muestra.

Cálculo de tamaño de muestra desenlace dicotómico

Como ejemplo, asumimos que se van a comparar dos tratamientos y el desenlace de interés es la proporción de fallecidos. Para todas las expresiones que se presentan a continuación, se denota p_T y p_C como la proporción de fallecidos en el grupo de tratamiento y en el grupo de control, respectivamente; ϵ es la diferencia esperada entre estas dos proporciones ($\epsilon=p_T-p_C$), δ es el margen de tolerancia o superioridad definido por los investigadores y k es la razón entre el tamaño de muestra del grupo de tratamiento y el grupo control ($k=n_T/n_C$), es decir que $n_T=kn_C$. Finalmente, denotamos α y β como el error tipo I y II, respectivamente; y $z_{(q)}$ como el percentil q de la función de distribución normal estándar. En la Tabla 2, se presentan las expresiones para obtener n_C , y, en el recuadro, el código en el lenguaje de programación R que crea una función para su implementación, junto con un ejemplo donde $\alpha=0,05$, $\beta=0,2$ y $k=1$.

En las cuatro hipótesis, a menor diferencia esperada (ϵ) y proporciones cercanas a 0,5, mayor es el tamaño de muestra. Al probar una hipótesis de no inferioridad, si a mayor proporción del evento mayor efectividad, entonces $\delta < 0$; si a menor proporción del evento mayor efectividad entonces $\delta > 0$. Al probar una hipótesis de superioridad, si a mayor proporción del evento mayor efectividad entonces $\delta > 0$; si a menor proporción del desenlace mayor efectividad entonces $\delta < 0$. Al probar una hipótesis de equivalencia, siempre $\delta > 0$.

Cálculo de tamaño de muestra desenlace continuo

Como ejemplo, asumimos que se van a comparar dos tratamientos donde el desenlace es la presión arterial sistólica en mmHg (PAS). Para todas las expresiones que se presentan a continuación, se denota μ_T y μ_C como la media de la PAS en el grupo de tratamiento y en el grupo de control, respectivamente; ϵ es la diferencia esperada entre las dos medias ($\epsilon=\mu_T-\mu_C$) y s es la desviación estándar de las dos muestras en conjunto. δ , k , α , β y $z_{(q)}$ representan los mismos valores que en la sección anterior. En la Tabla 3, se presentan las expresiones para obtener el código en el lenguaje de programación R para su implementación con un ejemplo donde $\alpha=0,05$, $\beta=0,2$ y $k=1$.

En las cuatro hipótesis, a mayor s y a menor ϵ se requiere mayor tamaño de muestra. Al probar una hipótesis de no inferioridad, si a mayor μ mayor efectividad entonces $\delta < 0$; si a menor μ mayor efectividad entonces $\delta > 0$. Al probar una hipótesis de superioridad, si a mayor μ mayor efectividad entonces $\delta > 0$; si a menor μ mayor efectividad entonces $\delta < 0$. Al probar una hipótesis de equivalencia, siempre $\delta > 0$.

RESULTADOS

Análisis interino

En un ECA se puede ir probando la hipótesis del estudio de manera secuencial a medida que se va recolectando la muestra, dando la posibilidad de suspender la recolección si se identifica tempranamente un claro beneficio de la intervención. Dependiendo el número de evaluaciones (R) que se programen, se requiere ajustar el tamaño de muestra inicial para mantener el nivel de significancia global del estudio, y establecer los valores críticos sobre la distribución del estadístico de prueba para rechazar o no la hipótesis nula en cada evaluación. Las R evaluaciones se realizan a medida que se van acumulando $\frac{n}{R}$ sujetos, y el estadístico de prueba z_r ($r=1,2,\dots,R$) para un desenlace dicotómico es igual a:

$$z_r = \frac{\sqrt{n_r}(\hat{p}_{Tr} - \hat{p}_{Cr})}{\sqrt{\hat{p}_{Tr}(1-\hat{p}_{Tr}) + \hat{p}_{Cr}(1-\hat{p}_{Cr})}}$$

donde n_r , \hat{p}_{Tr} y \hat{p}_{Cr} son el tamaño de muestra por grupo de intervención y las proporciones estimadas del desenlace al momento de la evaluación r del grupo de tratamiento y el grupo control, respectivamente.

Para un desenlace continuo, el estadístico de prueba es igual a:

$$z_r = \frac{1}{\sqrt{n_r(s_{Tr}^2 + s_{Cr}^2)}} \left(\sum_{j=1}^{n_r} x_{Tj} - \sum_{j=1}^{n_r} x_{Cj} \right)$$

donde n_r , s_{Tr}^2 y s_{Cr}^2 son el tamaño de muestra en cada grupo de intervención, y las varianzas estimadas al momento de la r -ésima evaluación del grupo de tratamiento y el grupo control, respectivamente. x_{Tj} y x_{Cj} son los valores observados del desenlace en cada sujeto recolectado hasta el momento r .

Se presentan cuatro métodos que permiten realizar el ajuste del tamaño de muestra dependiendo del número de evaluaciones programadas, el nivel de significancia y el poder establecido en una hipótesis de igualdad. Primero, el método de *Pocock*, donde el ajuste del tamaño de muestra se realiza multiplicando el tamaño de muestra obtenido inicialmente de las expresiones presentadas en la sección anterior, por los coeficientes incluidos en el Anexo 1, según el número de evaluaciones y los niveles de significancia y poder establecido. Ahora, si $|z_r| > CP_{(r,\alpha)}$ se rechaza la H_0 y se suspende la recolección, en caso contrario, la recolección continua. Los valores críticos $CP_{(r,\alpha)}$ se presentan en el Anexo 1 para R y α definidos. El segundo método es el de *O'Brien y Fleming* y los coeficientes para realizar el ajuste del tamaño de muestra inicial se presentan en el Anexo 2. En esta propuesta se rechaza

Tabla 2. Tipos de hipótesis con desenlace dicotómico y su código en el lenguaje de programación R.

Tipos de hipótesis	Código en lenguaje de programación R
<p>Hipótesis de igualdad</p> $n_c = \frac{\left(z_{(1-\alpha/2)} + z_{(1-\beta)} \right)^2}{\epsilon^2} \left[\frac{p_T(1-p_T)}{k} + p_C(1-p_C) \right]$ <p>Entonces, si se espera que la proporción de fallecidos en el grupo de tratamiento y en el grupo control, son iguales a $p_T = 0,15$ y $p_C = 0,2$, $n_C = n_T = 903$.</p>	<pre>n.2prop.igual<-function(alpha,beta,k,pT,pC){ nC<-(qnorm(1-alpha/2)+qnorm(1-beta))^2/(pT-pC)^2*(pT*(1-pT)/ k+pC*(1-pC)) nT<-k*nC Grupo<-c("Tratamiento=", "Control=") n<-ceiling(c(nT,nC)) n<-data.frame(Grupo,n) print(n) }</pre>
<p>Hipótesis de no inferioridad</p> $n_c = \frac{(z_{(1-\alpha)} + z_{(1-\beta)})^2}{(\epsilon-\delta)^2} \left[\frac{p_T(1-p_T)}{k} + p_C(1-p_C) \right]$ <p>Entonces, si se espera que la proporción de fallecidos en el grupo de tratamiento y en el grupo control, son $p_T=0,2$ y $p_C=0,22$ y se tolera un aumento en la mortalidad del $\delta=0,03$, $n_C = n_T = 821$.</p>	<pre>n.2prop.noinf<-function(alpha,beta,k,pT,pC,delta){ nC<-(qnorm(1-alpha)+qnorm(1-beta))^2/((pT-pC)-delta)^2*(pT*(1- pT)/k+pC*(1-pC)) nT<-k*nC Grupo<-c("Tratamiento=", "Control=") n<-ceiling(c(nT,nC)) n<-data.frame(Grupo,n) print(n) }</pre>
<p>Hipótesis de superioridad</p> $n_c = \frac{(z_{(1-\alpha)} + z_{(1-\beta)})^2}{(\epsilon-\delta)^2} \left[\frac{p_T(1-p_T)}{k} + p_C(1-p_C) \right]$ <p>Entonces, si se espera que la proporción de fallecidos en el grupo de tratamiento y en el grupo control son $p_T=0,18$ y $p_C=0,25$, y se considera superior si disminuye la mortalidad al menos en $\delta=0,01$, $n_C = n_T = 576$.</p>	<pre>n.2prop.sup<-function(alpha,beta,k,pT,pC,delta){ nC<-(qnorm(1-alpha)+qnorm(1-beta))^2/((pT-pC)-delta)^2*(pT*(1- pT)/k+pC*(1-pC)) nT<-k*nC Grupo<-c("Tratamiento=", "Control=") n<-ceiling(c(nT,nC)) n<-data.frame(Grupo,n) print(n) }</pre>
<p>Hipótesis de equivalencia</p> $n_c = \frac{(z_{(1-\alpha)} + z_{(1-\beta/2)})^2}{(\delta- \epsilon)^2} \left[\frac{p_T(1-p_T)}{K} + p_C(1-p_C) \right]$ <p>Entonces, si se espera que la proporción de fallecidos en el grupo de tratamiento y en el grupo control son $p_T=0,22$ y $p_C=0,18$ y se define que son equivalente si no se diferencias en más del $\delta =0,1$, $n_C = n_T = 760$.</p>	<pre>n.2prop.equivalencia<-function(alpha,beta,k,pT,pC,delta){ nC<-(qnorm(1-alpha)+qnorm(1-beta/2))^2/(delta-abs(pT- pC))^2*(pT*(1-pT)/k+pC*(1-pC)) nT<-k*nC Grupo<-c("Tratamiento=", "Control=") n<-ceiling(c(nT,nC)) n<-data.frame(Grupo,n) print(n) }</pre>

la H_0 en cada evaluación si $|z_r| > COF_{(r,\alpha)} \sqrt{(R/r)}$, de lo contrario continua. Los valores críticos $COF_{(r,\alpha)}$ se presenta en el Anexo 2 según el número de evaluaciones y el nivel de significancia. El tercer método es el de Wang y Tsiatis que incluye un nuevo parámetro Δ ; en el Anexo 3 se incluyen los coeficientes para el ajuste del tamaño de muestra para $\alpha=0,05$. En este método se rechaza H_0 si $|z_r| > CWT_{(r,\alpha,\Delta)} \left(\frac{r}{R}\right)^{\Delta-0,5}$, en caso contrario, continua. Los valores críticos $CWT_{(r,\alpha,\Delta)}$ se presentan en el Anexo 3 para $\alpha=0,05$. Los métodos de Pocock y O'Brien y Fleming son casos particulares del método de Wang y Tsiatis cuando $\Delta=0,5$ y $\Delta=0$, respectivamente, por lo tanto, los valores críticos para estos valores de Δ se obtienen de los Anexos

1 y 2. Finalmente, se presenta el método de Inner Wedge; en esta propuesta a diferencia de las tres anteriores, se propone dos valores críticos de tal forma que si $|z_r| \geq b_r$ rechaza la H_0 y se suspende la recolección, concluyendo que, se encontró un efecto significativo del tratamiento, si $|z_r| < a_r$ no rechaza la H_0 y se suspende la recolección, concluyendo que, no se van a encontrar diferencias entre el tratamiento y el control, de lo contrario, continua la recolección. Los valores críticos a_r y b_r son iguales a:

$$a_r = [Cw1_{(r,\alpha,\beta,\Delta)} + Cw2_{(r,\alpha,\beta,\Delta)}] \sqrt{\frac{r}{R}} - Cw2_{(r,\alpha,\beta,\Delta)} \left(\frac{r}{R}\right)^{(\Delta-0,5)}, \text{ si } a_r < 0 \Rightarrow a_r = 0$$

$$\text{y } b_r = Cw1_{(r,\alpha,\beta,\Delta)} \left(\frac{r}{R}\right)^{(\Delta-0,5)}$$

Tabla 3. Tipos de hipótesis por desenlace continuo y código en lenguaje de programación R.

Tipo de hipótesis	Código en lenguaje de programación R
<p>Hipótesis de igualdad</p> $n_c = \frac{(z_{(1-\alpha/2)} + z_{(1-\beta)})^2 s^2 (1+1/k)}{\epsilon^2}$ <p>Entonces, si la media $\mu_T=150$ y $\mu_C=160$ y la $s=28$, $n_c=n_T=124$.</p>	<pre>n.2mu.igual<- function(alpha,beta,k,muT,muC,s){ nC<-(qnorm(1-alpha/2)+qnorm(1- beta))^2*s^2*(1+1/k)/(muT-muC)^2 nT<-k*nC Grupo<-c("Tratamiento =", "Control =") n<-ceiling(c(nT,nC)) n<-data.frame(Grupo,n) print(n) }</pre>
<p>Hipótesis de no inferioridad</p> $n_c = \frac{(z_{(1-\alpha)} + z_{(1-\beta)})^2 s^2 (1+1/k)}{(\epsilon-\delta)^2}$ <p>Entonces, si $\mu_T=155$, $\mu_C=160$ y $s=28$, y se define que no es inferior si el máximo aumenta en $\delta=5$, $n_c=n_T=97$.</p>	<pre>n.2mu.noinf<- function(alpha,beta,k,muT,muC,s,delta){ nC<-(qnorm(1-alpha)+qnorm(1- beta))^2*s^2*(1+1/k)/((muT-muC)-delta)^2 nT<-k*nC Grupo<-c("Tratamiento =", "Control =") n<-ceiling(c(nT,nC)) n<-data.frame(Grupo,n) print(n) }</pre>
<p>Hipótesis de superioridad</p> $n_c = \frac{(z_{(1-\alpha)} + z_{(1-\beta)})^2 s^2 (1+1/k)}{(\epsilon-\delta)^2}$ <p>Entonces, si $\mu_T=145$, $\mu_C=160$ y $s=28$, y se considera superior si por lo menos disminuye en $\delta=-10$, $n_c=n_T=388$.</p>	<pre>n.2mu.sup<- function(alpha,beta,k,muT,muC,s,delta){ nC<-(qnorm(1-alpha)+qnorm(1- beta))^2*s^2*(1+1/k)/((muT-muC)-delta)^2 nT<-k*nC Grupo<-c("Tratamiento =", "Control =") n<-ceiling(c(nT,nC)) n<-data.frame(Grupo,n) print(n) }</pre>
<p>Hipótesis de equivalencia</p> $n_c = \frac{(z_{(1-\alpha)} + z_{(1-\beta/2)})^2 s^2 (1+1/k)}{(\delta- \epsilon)^2}$ <p>Entonces, si $\mu_T=150$, $\mu_C=160$ y $s=28$, y se define que son equivalentes si no se diferencian en más de $\delta =5$, $n_c=n_T=538$.</p>	<pre>n.2mu.equivalencia<-function(alpha,beta,k,muT,muC,s,delta){ nC<-(qnorm(1-alpha)+qnorm(1-beta/2))^2*s^2*(1+1/k)/((delta-abs(muT- muC))^2) nT<-k*nC Grupo<-c("Tratamiento =", "Control =") n<-ceiling(c(nT,nC)) n<-data.frame(Grupo,n) print(n) }</pre>

En el Anexo 4 se presentan los valores de $Cw1$ y $Cw2$ para un $\alpha=0,05$ y un poder de 0,8 y 0,9, y en las columnas *coef.ajuste* se incluyen los coeficientes, por los cuales se debe multiplicar la muestra original para realizar las *R* evaluaciones.

A manera de ejemplo, considere que se quiere comparar el medicamento A vs. placebo y el desenlace es la proporción de muertes al final del seguimiento. En una hipótesis de igualdad, asumiendo $\alpha=0,05$, $\beta=0,1$, $p_T=0,1$ y $p_C=0,2$ (es decir $\epsilon=0,1$), para dos grupos del mismo tamaño, $k=1$, el tamaño de muestra requerido en cada grupo es de 263 sujetos. Si se plantea realizar $R=5$ evaluaciones, tenemos que por el método de *Pocock*, el tamaño de muestra ajustado es $263 \times 1,207=318$ por cada grupo y el valor crítico en cada evaluación es $CP_{0,05,r}=2,413$. Por el método de *O'Brien y Fleming*, el tamaño de muestra ajustado es $263 \times 1,026=270$ para cada grupo y los valores críticos para cada evaluación son $COF_{(r,0,05)}=4,562; 3,226; 2,634; 2,281$ y $2,040$. Por el método de *Wang y Tsiatis*, el tamaño de muestra ajustado para cada grupo,

con $\Delta=0,25$, es $263 \times 1,066=281$ y los valores críticos en cada evaluación son $CWT_{(r,0,05;0,25)}=3,194; 2,686; 2,427; 2,259$ y $2,136$. Finalmente, por el método de *Inner Wedge*, el tamaño de muestra ajustado por cada grupo, con $\Delta=0,25$, es $263 \times 1,199=316$ y los valores críticos para cada evaluación son $a_r=0; 0,388; 1,072; 1,613$ y $2,073$ y $b_r=3,1; 2,607; 2,355; 2,192$ y $2,073$.

DISCUSIÓN

En este artículo presentamos una aproximación al ajuste del tamaño de muestra por análisis interino en ECA en paralelo, partiendo del cálculo del tamaño de muestra original para su posterior ajuste por uno de los cuatro métodos descritos. El documento está dirigido a estudiantes y jóvenes investigadores, principalmente en el área de la salud, en el que encontrarán un contexto inicial sobre los ECA y una revisión sobre los principales conceptos de la inferencia estadística a partir de las pruebas de hipótesis. Buscamos de una manera sencilla y concreta, realizar

una introducción a este tema, integrando los diferentes aspectos como las expresiones matemáticas que soportan los resultados y su implementación en programas estadísticos disponibles. Aunque, existen otros recursos disponibles para el cálculo de tamaño de muestra como páginas de internet⁽¹⁹⁾ o paquetes en el lenguaje de programación R⁽⁴⁾, principalmente en idiomas distintos al español, encontramos que brindar la posibilidad de utilizar programas estadísticos que les permitan aplicar la teoría a los estudiantes, da una mayor comprensión de estos temas, a diferencia de seguir una secuencia de pasos de forma mecánica, muchas veces sin entender lo que se genera por los distintos programas o recursos disponibles. Lo anterior, permite acercarse a los estudiantes de las áreas de la salud, a la estadística y a la utilización de programas estadísticos, aspecto muchas veces considerado poco necesario en su formación.

Este artículo permite al lector, la planeación de un ECA en paralelo definiendo el tamaño de muestra y permitiendo la monitorización de los resultados durante el transcurso del estudio. En este punto, se recomienda continuar con la revisión de métodos adicionales que dan mayor flexibilidad al momento de realizar estas evaluaciones, por ejemplo, permitiendo planear las evaluaciones intermedias en fechas específicas y no cuando se van completando un número fijo de participantes en ambos grupos, el cual es la principal restricción de los cuatro métodos presentados en este artículo. El método propuesto por *Lan y DeMets*⁽²⁰⁾ y paquetes del lenguaje de programación R como *gsDesign*⁽²¹⁾ serán material de interés para seguir profundizando en estos temas.

Finalmente, la decisión de interrumpir la ejecución de un ECA, ya sea porque se observan grandes beneficios, se presentan potenciales daños o es muy improbable encontrar beneficios (futilidad), debe ser tomada por un grupo independiente al de los investigadores, conformado por expertos en el área clínica bajo estudio, en aspectos metodológicos como epidemiólogos o bioestadísticos y en aspectos éticos^(3,22,23). En esta decisión, se deben considerar todos los aspectos necesarios y no sólo el resultado de la evaluación de una prueba estadística. Realizar la planeación de un análisis interino ajustando el tamaño de muestra desde la elaboración del protocolo, permitirá soportar, en mayor medida, el valor de este criterio en la toma de la decisión.

Contribuciones de autoría. Todos los autores declaran que cumplen los criterios de autoría recomendados por el ICMJE.

Roles según CRediT. MSM: Conceptualización. Metodología. Investigación. Redacción – borrador original. Redacción–revisión y edición. Visualización. FBR: Metodología. Investigación. Redacción – borrador original. Visualización. CJR: Conceptualización. Metodología. Software. Investigación. Redacción – borrador original. Redacción–revisión y edición.

Financiamiento. Autofinanciado.

Conflictos de interés. Los autores declaran no tener conflicto de intereses.

Material suplementario. Disponible en la versión electrónica de la *RPMESSP*.

REFERENCIAS BIBLIOGRÁFICAS

1. F Berwick DM. Era 3 for Medicine and Health Care. *Obstet Gynecol Surv.* 2016;315(13):1329–30. doi:10.1001/jama.2016.1509.
2. Gordon G, Drummond R, Maureen OM, Deborah JC. *Users' Guides to the Medical Literature, 3rd ed* [Internet]. McGraw Hill; 2008 [citado el 11 de octubre de 2022]. Disponible en: <https://jamaevidence.mhmedical.com/content.aspx?bookid=847§ionid=69030714>.
3. Friedman LM, Furberg CD, DeMets DL, Reboussin DM, Granger CB. *Fundamentals of clinical trials.* In: *Fundamentals of Clinical Trials* [Internet]. Spinger; 2015 [citado el 11 de octubre de 2022]. Disponible en: <https://link.springer.com/book/10.1007/978-3-319-18539-2>.
4. Zhang E, Wu VQ, Chow S-C, Zhang HG. *TrialSize: R Functions for Chapter 3,4,6,7,9,10,11,12,14,15 of Sample Size Calculation in Clinical Research.* R package version 1.4 [Internet]. 2020 [citado el 11 de octubre de 2022]. Disponible en: <https://cran.r-project.org/package=Trialsize>.
5. Miller FG, Joffe S. Equipose and the Dilemma of Randomized Clinical Trials. *N Engl J Med.* 2011;364(5):476–80. doi: 10.1056/nejmsb1011301.
6. Lazcano-Ponce E, Salazar-Martínez E, Gutiérrez-Castrellón P, Angeles-Llerenas A, Hernández-Garduño A, Viramontes JL. Ensayos clínicos aleatorizados: variantes, métodos de aleatorización, análisis, consideraciones éticas y regulación. *Salud Publica Mex.* 2004;46(6):559–84.
7. Diana MF. Conceptos básicos sobre bioestadística descriptiva y bioestadística inferencial. *Rev Argentina Anestesiol.* 2006;64(6):241–51.
8. Flight L, Julious SA. Practical guide to sample size calculations: An introduction. *Pharm Stat.* 2016;15(1):68–74. doi: 10.1002/pst.1709.
9. Cohen HW. P values: Use and misuse in medical literature. *Am J Hypertens.* 2011;24(1):18–23. doi: 10.1038/ajh.2010.205.
10. Prel J-B du, Hommel G, Röhrig B, Blettner M. Confidence interval or P value. *Dtsch Arztebl.* 2009;106(19):335–9. doi: 10.3238/arztebl.2009.0335.
11. Goodman S. A Dirty Dozen: Twelve P-Value Misconceptions. *Semin Hematol.* 2008;45(3):135–40. doi: 10.1053/j.seminhematol.2008.04.003.
12. Wayne DW. *Bioestadística. Base para el análisis de las ciencias de la salud.* 4ta ed. LIMUSA WILEY. 2006.
13. Chow S-C, Shao J, Wang H, Lokhnygina Y. *Sample size calculations in clinical research.* 2nd ed. Chapman & Hall; 2008.
14. Armijo-Olivo S, Warren S, Fuentes J, Magee DJ. Clinical relevance vs. statistical significance: Using neck outcomes in patients with temporomandibular disorders as an example. *Man Ther.* 2011;16:563–72. doi: 10.1016/j.math.2011.05.006.
15. Chan A, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-jezić K, et al. Declaración SPIRIT 2013: definición de los elementos estándares del protocolo de un ensayo clínico. *Rev Panam salud pública.* 2015;38(6):506–14.
16. Gowda GS, Komal S, Sanjay TN, Mishra S, Kumar CN, Math SB. Sociodemographic, legal, and clinical profiles of female forensic inpatients in Karnataka: A retrospective study. *Indian J Psychol Med.* 2019;41(2):138–43. doi: 10.4103/IJPSYM.IJPSYM_152_18.
17. R Core Team. R: A language and environment for statistical computing [Internet]. R Foundation for Statistical Computing, Vienna, Austria; 2022 [11 de octubre 2022]. Disponible en: <https://www.r-project.org/>.
18. RStudio Team. RStudio: Integrated Development Environment for R [Internet]. RStudio, PBC, Boston, MA; 2022 [citado el 11 de octubre de 2022]. Disponible en: <http://www.rstudio.com/>.
19. Sample Size Calculator [Internet]. Cleveland Clinic; 2022 [citado el 11 de octubre de 2022]. Disponible en: <https://riskcalc.org/sample-size/>.
20. Demets DL, Lan KKG. Interim analysis: the alpha spending approach. *Stat Med.* 1994;13:1341–52. doi: 10.1002/sim.4780131308.

21. Anderson K. gsDesign: Group Sequential Design [Internet]. R package version 3.4.0; 2022 [citado el 11 de octubre de 2022]. Disponible en: <https://CRAN.R-project.org/package=gsDesign>.
22. Ellenberg S, Fleming T, DeMets D. Data Monitoring Committees in Clinical Trials: A Practical Perspective. Wiley; 2003.
23. Fisher M, Roecker E, DeMets D. The role of an independent statistical analysis center in the industry-modified National Institutes of Health model. *Drug Inf J*. 2001;(35):115–29. doi: [10.1177/009286150103500113](https://doi.org/10.1177/009286150103500113).
24. Kumbhare D, Alavinia M, Furlan J. Hypothesis Testing in Superiority, Noninferiority, and Equivalence Clinical Trials. *Am J Phys Med Rehabil*. 2019;98(3):226–30. doi: [10.1097/PHM.0000000000001023](https://doi.org/10.1097/PHM.0000000000001023).
25. Services H. Non-Inferiority Clinical Trials to Establish Effectiveness Guidance for Industry [Internet]. FDA; 2016 [citado el 11 de octubre de 2022]. Disponible en: <https://www.fda.gov/media/78504/download>.
26. Flight L, Julious SA. Practical guide to sample size calculations: an introduction. *Pharm Stat*. 2016;15(1):68–74. doi: [10.1002/pst.1709](https://doi.org/10.1002/pst.1709).
27. Benito MM, Marín RC. Cambios en la presión arterial y frecuencia cardíaca después de una presión sobre la válvula aórtica en sujetos con hipertensión arterial esencial. *Osteopat Cient*. 2008;3(3):100–7. doi: [10.1016/S1886-9297\(08\)75758-8](https://doi.org/10.1016/S1886-9297(08)75758-8).
28. Chadwick D. Safety and efficacy of vigabatrin and carbamazepine in newly diagnosed epilepsy: a multicentre randomised double-blind study. *Lancet*. 1999;354:13–9. doi: [10.1016/S0140-6736\(98\)10531-7](https://doi.org/10.1016/S0140-6736(98)10531-7).
29. Priebe S, Chevalier A, Hamborg T, Golden E, King M, Pistrang N. Effectiveness of a volunteer befriending programme for patients with schizophrenia: Randomised controlled trial. *Br J Psychiatry*. 2019;1–7. doi: [10.1192/bjp.2019.42](https://doi.org/10.1192/bjp.2019.42).
30. Bascope EL, Ortiz YM, Llanos GRL, Lizbeth MHA, Lazo L. Metformina en el tratamiento del síndrome de ovarios poliquísticos. Un ensayo clínico aleatorizado. *Rev Cient Cienc Med*. 2017;20(2):45–52.