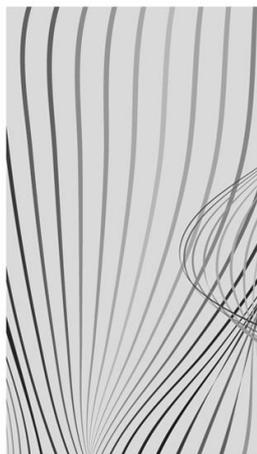


Estudios clínicos:

Algunos conceptos estadísticos

Allan Ramos Esquivel



WB141

R175e Ramos Esquivel, Allan

Estudios clínicos: algunos conceptos estadísticos / Allan Ramos
Esquivel. – San José, CR. : EDNASSS-CCSS, 2019.
144 p. : ilustraciones ; 14 x 21 centímetros.

ISBN: 978-9968-916-74-5

1. ESTUDIO CLÍNICO. 2. ESTADÍSTICA 3. MEDICINA BASADA EN
EVIDENCIA. I. Título.

Esta publicación fue aprobada por el Consejo Editorial de EDNASSS,
en la sesión N° 156, del 14 de agosto de 2018.

Diseño de portada: OMAI comunicación y publicidad.

Artes e impresión: Ofiprinte Comercial M.B. S.A.

© Editorial Nacional de Salud y Seguridad Social (EDNASSS) 2019.
Centro de Desarrollo Estratégico e Información en Salud y
Seguridad Social.
Caja Costarricense de Seguro Social.
Teléfono: 2221-6193. Fax: 2233-8359
Correo electrónico: ednasss@binasss.sa.cr

EDNASSS: una editorial al servicio de la salud y la seguridad social

TABLA DE CONTENIDOS

AUTOR.....	5
PRÓLOGO.....	7
CAPÍTULO 1. CONCEPTOS DEL DISEÑO DE ESTUDIOS.....	9
Medicina basada en evidencia.....	9
Tipos de diseños.....	11
¿Qué es y cómo se mide la incidencia?.....	24
Trabajando con la muestra: los errores de la estimación...	30
CAPÍTULO 2. CONCEPTOS DEL ANÁLISIS ESTADÍSTICO.....	39
Medidas de asociación.....	39
Contabilizando los eventos: análisis por intención a tratar y análisis por protocolo.....	47
Midiendo la precisión.....	51
El valor de p	62
El objetivo primario: ¿simple o compuesto?.....	70
Análisis múltiples.....	77
Estudios de no inferioridad.....	86
Análisis de regresión.....	91
Análisis de supervivencia.....	97
Revisiones sistemáticas y metaanálisis.....	107
¿Cómo se calcula la muestra en un ensayo clínico?.....	113
Consideraciones bioéticas.....	123

ANEXOS	133
Anexo 1. ¿Cómo buscar la mejor evidencia clínica?.....	135
Anexo 2. Definición de conceptos claves.....	141

AUTOR

Allan Ramos Esquivel nació en San José, Costa Rica, en 1983, y obtuvo su título de Licenciatura en Medicina y Cirugía en el año 2007, en la Universidad de Costa Rica (UCR). Realizó una Maestría en Epidemiología Clínica en la Universidad Erasmus de Rotterdam, Países Bajos, y una Maestría Académica en Farmacología en la UCR. Además, obtuvo una especialidad en Medicina Interna y en Oncología Médica en la UCR.

Desde el año 2015 se desempeña como Médico Asistente Especialista en el Departamento de Oncología Médica del Hospital San Juan de Dios, de la Caja Costarricense de Seguro Social (CCSS). Asimismo, es Profesor Asociado en los Departamentos de Farmacología y Fisiopatología de la Escuela de Medicina de la UCR y actualmente es Investigador Asociado en el Centro de Investigación en Hematología y Trastornos Afines (CIHATA) de dicha universidad.

El autor ha mostrado su interés en el análisis de la información científica, en particular en el diseño y el escrutinio de ensayos clínicos, y es revisor para diversas revistas científicas internacionales. En la actualidad es profesor del curso de Análisis de Información Clínica Científica, en donde desarrolla los contenidos de esta obra.

PRÓLOGO

Los estudiantes y profesionales en salud se enfrentan a una relativa escasez de literatura que respalde su formación en el área de la estadística médica. Muchos de los libros disponibles emplean fórmulas matemáticas extensas o carecen de una aplicabilidad clínica que facilite la comprensión de los estudios clínicos. Por esa razón se creó esta obra, que constituye un recurso valioso para introducir conceptos estadísticos básicos, los cuales son frecuentemente utilizados en el diseño, la interpretación y el análisis de estudios clínicos.

Este libro, de fácil lectura y con valiosos ejemplos prácticos, contiene un resumen de una variedad de temas que difícilmente se encuentran en un solo tomo, ya que el autor ha realizado una búsqueda de diferentes fuentes bibliográficas, para facilitarle al lector la comprensión de elementos básicos incluidos en los ensayos clínicos.

CAPÍTULO 1. CONCEPTOS DEL DISEÑO DE ESTUDIOS

MEDICINA BASADA EN EVIDENCIA

En la práctica de la medicina nos enfrentamos a diario con varias preguntas formuladas tanto por el paciente como por el médico. De estas, las cuatro interrogantes básicas son:

1. ¿Cuál es el diagnóstico más probable?
2. ¿Cuál es la etiología de la enfermedad diagnosticada o los factores de riesgo para la aparición de una patología?
3. ¿Cuál es el pronóstico de dicha condición?
4. ¿Cuál es el mejor tratamiento o intervención para esta enfermedad?

Estas cuatro preguntas pueden ser solucionadas mediante varios métodos. Por ejemplo, la experiencia de un oncólogo puede influir en la respuesta que un paciente con cáncer de pulmón espera obtener sobre su pronóstico de vida. Similarmente, un médico puede emplear su intuición o raciocinio fisiopatológico para suponer el mejor tratamiento de una patología específica. Sin embargo, la mejor forma de contestar cada una de las interrogantes previas es aplicando el método científico, a través de la medicina basada en evidencia.

¿Por qué la medicina basada en la fisiopatología de la enfermedad no es suficiente para tomar decisiones terapéuticas? Un fatal ejemplo

Para el año 1979, Bernard Lown describió que muchos de los pacientes que morían en las primeras horas después de un infarto de miocardio lo hacían como consecuencia de trastornos del ritmo (arritmias).

Unos años después, se publicaron en la revista *New England Journal of Medicine* (1) los efectos del antiarrítmico flecainida en la prevención de contracciones preauriculares en pacientes cardiopatas. Los resultados de la investigación fueron contundentes a favor del uso de la flecainida para la reducción de arritmias *versus* placebo, lo cual condujo a la aprobación de dicho medicamento por parte de la *Food and Drug Administration (FDA)* y su inclusión como componente del manejo estándar de los pacientes con infarto agudo al miocardio. Sin embargo, años después, otros investigadores describieron que si bien la flecainida reducía los eventos de arritmias, el uso de esta droga produjo más muertes que en el grupo control. De tal forma, casi una década después de su aprobación, la FDA tuvo que prohibir el uso de esta droga en ese contexto, debido al efecto deletéreo en la supervivencia de pacientes con infarto agudo al miocardio, pese a una reducción en el número de arritmias.

Este ejemplo ilustra uno de los desafíos de la medicina basada en evidencia: la validez de los datos. ¿Fue la información proveniente del primer estudio falsa? ¿Se pudo haber evitado el uso de flecainida con una adecuada lectura de los datos del primer ensayo clínico?

Definición

La medicina basada en evidencia surgió como un cambio en el paradigma de la práctica de la profesión médica a finales del siglo XX. Este concepto prioriza los hallazgos provenientes de estudios científicos, como fuente primordial para contestar cada uno de los cuatro escenarios clínicos antes descritos (2).

Así pues, la medicina basada en evidencia representa la *“integración de la experiencia clínica, los valores o necesidades del paciente y la incorporación de la mejor evidencia disponible en el proceso de decisión relacionado con el cuidado de la salud”* (3). Dicha evidencia proviene de diversos tipos de estudios clínicos, los cuales, según su diseño y metodología,

permiten contestar con veracidad y precisión las preguntas de diagnóstico, etiología, tratamiento y pronóstico, que tanto médicos como pacientes nos formulamos a diario.

La medicina basada en evidencia tiene una estrecha relación con la epidemiología clínica, ya que emplea métodos estadísticos y epidemiológicos para responder cada uno de los escenarios clínicos.

Surgen entonces varias preguntas que constituyen el objeto de este libro: ¿Cuál es la mejor evidencia disponible? ¿Dónde encontrarla? ¿Cómo interpretarla? ¿Cómo aplicarla a la práctica clínica diaria?

TIPOS DE DISEÑOS

Existen diferentes métodos por medio de los cuales la medicina basada en evidencia puede brindarnos las respuestas a las interrogantes básicas de etiología, diagnóstico, tratamiento y pronóstico de una condición clínica particular. Sin embargo, no todos los diseños pueden brindarnos evidencia suficiente para responder con certeza tales preguntas.

Lo anterior surge del problema epidemiológico de comprobar la causalidad de un evento a partir de la presencia de un factor de riesgo o de la exposición a una intervención. Si bien en términos epidemiológicos se prefiere hacer referencia a la asociación entre un factor de riesgo y un evento de interés, algunas características de estas asociaciones pueden implicar que tales asociaciones son, en efecto, causales. Dentro de este contexto, los criterios de causalidad de Bradford Hill (Cuadro N° 1) constituyen elementos clásicos que permiten establecer la probable causalidad, a partir de una asociación entre un factor de riesgo y el evento de interés. De tal forma, los diferentes diseños de estudios clínicos tendrán mayor o menor fortaleza para detectar causalidad según cumplan estos criterios.

Cuadro N° 1. Criterios de causalidad de Bradford Hill

1. Fuerza de asociación.
2. Consistencia.
3. Especificidad.
4. Temporalidad.
5. Gradiente biológico.
6. Plausibilidad biológica.
7. Coherencia.
8. Evidencia experimental.
9. Analogía.

Fuente: Hill AB. The environment and disease: association or causation. *Proc Royal Soc Med.* 1965; 58: 295-300.

Aplicando los criterios de causalidad de Bradford Hill antes mencionados, se construye una pirámide que ilustra el grado de evidencia científica, según el diseño del estudio utilizado para explicar la causalidad de un evento a partir de ciertas variables (Figura N° 1).



Figura N° 1. Diseño de los estudios clínicos según el grado de evidencia.

Es importante mencionar que no siempre un metaanálisis provee la mejor evidencia científica en un contexto clínico determinado, ya que si el metaanálisis carece de rigor estadístico e incluye sesgos en su diseño, las asociaciones que concluya serán inválidas. De igual forma, existen interrogantes particulares que no pueden responderse mediante ensayos controlados aleatorizados, ya sea por la rareza de la condición clínica o por cuestiones éticas. En tales casos, la evidencia proveniente de estudios de cohorte o de estudios caso-control deberá aportar al lector la suficiente evidencia científica para apoyar o rechazar una hipótesis.

Estudio transversal

El estudio transversal es el típico diseño que se compara con una fotografía de un momento específico y consiste en la observación analítica o descriptiva de las características de un grupo determinado de sujetos. Tal como se deriva de su nombre, consiste en una descripción de los datos de una población o muestra en un momento determinado, por lo que adolece de una relación causal.

En este tipo de estudios se trata de recabar una única vez por cada sujeto incluido, la información sobre el factor de riesgo o la intervención (información de exposición) y, al mismo tiempo, determinar el número de casos del evento de interés.

Dentro de estos estudios se encuentran los estudios de prevalencia de enfermedades en una población, los estudios de asociación cruzada, las encuestas, los estudios de evaluación de pruebas diagnósticas y los estudios de concordancia o “*agreement*”, los cuales suelen tener especial relevancia en el campo de la salud pública.

Tienen la ventaja de ser económicos y relativamente rápidos de ejecutar y analizar, pero carecen de la secuencia temporal que imposibilita determinar la relación causa-efecto de un factor de riesgo con el evento final, además de que son difíciles de

realizar para investigar enfermedades raras, ya que requerirían de muestras de sujetos excesivamente amplias. A pesar de esto, los estudios transversales son útiles porque brindan nuevas hipótesis y determinan la prevalencia de enfermedades.

En este tipo de diseños es fundamental especificar con claridad la pregunta de investigación, la población de estudio y la técnica de muestreo, así como determinar el factor de riesgo y el evento de interés.

Caso para analizar

Un estudio publicado reporta la prevalencia simultánea de osteoartritis de rodilla (gonartritis) y obesidad. Dentro de las conclusiones se reporta una alta prevalencia de ambas condiciones. ¿Se puede concluir de ese estudio que la obesidad es un factor de riesgo de gonartritis? Evidentemente, el diseño descriptivo del estudio impide establecer asociación causal; por lo tanto, no puede concluirse que la obesidad es la causa de la osteoartritis o que sea la consecuencia del reducido ejercicio físico producto del dolor articular.

Estudios longitudinales

En contraste con los estudios transversales, los diseños longitudinales toman en cuenta el tiempo de seguimiento como una variable más del análisis investigativo. Este tipo de estudios se divide en dos: estudios prospectivos y estudios retrospectivos.

Estudios prospectivos

Este diseño se caracteriza por dos aspectos:

1. La variable de interés u *outcome* no ha ocurrido al inicio del desarrollo del estudio.
2. La variable de exposición se verifica en el presente.

Un ejemplo del diseño prospectivo sería determinar el efecto del tabaquismo (variable de exposición) sobre el cáncer de pulmón (variable de interés). Para que este estudio sea prospectivo, ninguno de los sujetos que se incluyan en el seguimiento debe tener cáncer de pulmón. Además, debe definirse al inicio del estudio cuál será la variable de exposición (tabaquismo). Los participantes serán seguidos a lo largo de un período de tiempo, durante el cual se determinará cuántos pacientes llegan a desarrollar cáncer.

Los estudios prospectivos tienen la desventaja de que con frecuencia sufren pérdidas de los individuos participantes durante el seguimiento, ya que este suele ser prolongado, o bien, el participante decide retirarse del estudio. Estos hechos pueden introducir sesgos de información que invaliden las conclusiones finales. No obstante, este diseño de análisis es útil porque permite recabar información prospectiva.

Estudios retrospectivos

De forma opuesta, los estudios retrospectivos se caracterizan por:

1. La variable de interés ya ocurrió al momento en que se realiza la recolección de los datos.
2. La exposición a un determinado factor de riesgo o intervención se hace de forma retrógrada.

En el mismo ejemplo anterior, un diseño retrospectivo sería aquel que identifica en el presente a sujetos con cáncer de pulmón y de forma retrospectiva recaba información sobre los factores de riesgo o exposición al tabaco por parte de cada sujeto.

Los estudios retrospectivos carecen muchas veces de variables de interés, las cuales deben ser recabadas en registros médicos o consultadas directamente al participante.

Tipos de estudios longitudinales

Estudios caso-control

Los estudios caso-control son diseños retrospectivos por definición. Surgen a partir de la identificación de casos; es decir, sujetos con el evento de interés. Una vez definido el caso, este se compara con un sujeto “control”. Es de esperarse que el control sea en todo similar al “caso”, excepto por no presentar la patología o el evento de interés. Después de esto, el investigador se dedicará a identificar y a comparar los factores de riesgo o de exposición en los casos y los controles.

Estos estudios son particularmente útiles cuando la enfermedad o la variable de interés es rara, ya que estos análisis surgen de la identificación del evento primario y no de la exposición.

En los estudios caso-control se construye una tabla de contingencia como la presentada en el Cuadro N° 2, de la cual se calcula el *odds ratio*.

Cuadro N° 2. Posibles resultados de un estudio caso-control		
	Enfermo: caso	No enfermo: control
Expuesto	a	b
No expuesto	c	d

Ejemplo de un estudio caso-control

Yusuf y colaboradores (4) realizaron un estudio de casos y controles, con el fin de determinar la asociación de varios factores de riesgo cardiovascular (tabaquismo, hipertensión arterial, diabetes, entre otros) con la ocurrencia de un infarto agudo al miocardio. El estudio publicado empleó cerca de 15 mil casos e igual número de controles, procedentes de 52

países. Los casos fueron seleccionados considerando si tenían síntomas y cambios electrocardiográficos típicos de un nuevo infarto, mientras que los controles fueron elegidos sin diagnóstico de enfermedad cardiovascular. Al final, este estudio confirmó en diferentes regiones del mundo varios de los factores de riesgo cardiovascular.

Estudios de cohorte

Los estudios de cohorte, como todo estudio longitudinal, incluyen el tiempo de seguimiento como una variable más del análisis. La incorporación de la temporalidad permite que a partir de los estudios de cohorte se pueda establecer la incidencia de una enfermedad en un grupo de sujetos expuestos a un factor de riesgo o sometidos a una intervención determinada. Además, esto permite que se cumpla con el cuarto de los criterios de causalidad establecidos por Bradford Hill, acercándose así este diseño a la definición causal de una asociación.

La cohorte clásicamente representa un grupo de sujetos que comparten características comunes y que “marchan” a lo largo del tiempo de seguimiento hasta que presentan o no el evento de interés. Por lo general, hay dos cohortes: una expuesta (cohorte expuesta) y otra no expuesta (cohorte de referencia) a uno o más factores de riesgo.

Para poder determinar el efecto promedio de la exposición de una cohorte a un factor de riesgo, es necesario que los sujetos de investigación sean expuestos a dicho factor durante un período de seguimiento (“*follow-up*”). Sin embargo, cuando hay pérdida de sujetos durante este período de tiempo se generan “datos perdidos” (“*loss to follow-up*”), los cuales requerirán de un análisis estadístico especial para poder estimar con validez el riesgo de presentar un determinado evento posterior a dicha exposición. En cada cohorte se calculará la incidencia del evento de interés y se establecerá el riesgo relativo entre cada grupo (Cuadro N° 3).

Cuadro N° 3. Posibles resultados de un estudio de cohorte		
	Con el evento de interés	Sin el evento de interés
Cohorte expuesta	a	b
Cohorte de referencia	c	d

Clasificación de las cohortes según la exposición de los sujetos al factor de riesgo

Las cohortes se pueden clasificar de la siguiente forma, según ocurra la exposición de los sujetos en estudio al factor de riesgo:

1. *Cohorte fija*: en este caso, la exposición de la cohorte al factor de riesgo está definida y es “fija” desde que se inicia el seguimiento, y no está permitido el movimiento de un sujeto entre los grupos de exposición y no exposición. Una cohorte fija, que no sufre pérdidas durante el seguimiento del estudio, se conoce también como una cohorte cerrada.
2. *Cohorte dinámica o abierta*: en este caso, el número de integrantes de la cohorte se modifica con el tiempo. Lo anterior surge de que muchas veces la exposición a un factor de riesgo no es fija durante el seguimiento del participante, además, en ocasiones, los sujetos suelen “entrecruzarse” entre grupos de exposición y no exposición (“*cross-over*”).

Análisis de caso

Suponga que se quiere comparar el riesgo de mesotelioma en sujetos expuestos laboralmente al silicio. Para ello, se utiliza un estudio de cohorte prospectiva. Como la exposición al silicio no es permanente durante el período de seguimiento, esta cohorte

será clasificada como abierta, ya que habrá momentos donde el individuo esté expuesto al silicio y otros momentos donde no tendrá dicha exposición.

Clasificación de las cohortes según la forma en que se recopilen los datos

Por otro lado, las cohortes también pueden clasificarse en *cohortes prospectivas* y *retrospectivas* (o *históricas*), según la forma en que se recopilen los datos. En el estudio prospectivo se recolecta la información del evento de interés en el futuro, en espera del acontecimiento de este; mientras que en el diseño retrospectivo se recaba información del pasado, ya que el evento de interés se determina en el presente (Figura N° 2).

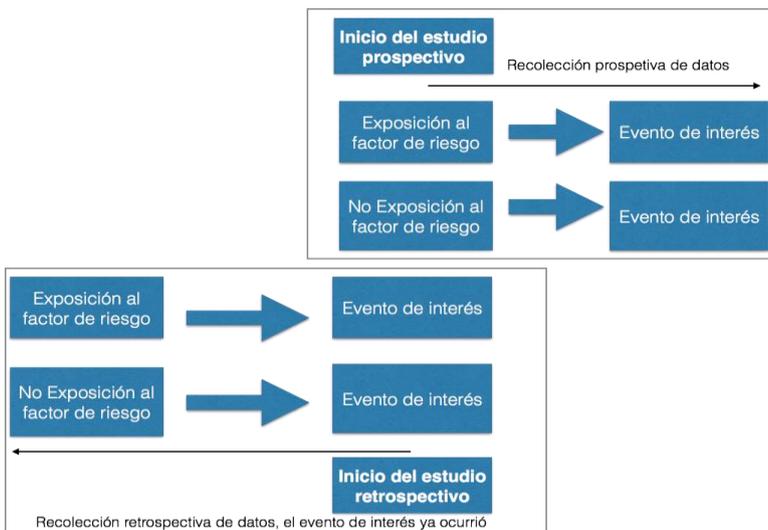


Figura N° 2. Esquema de un estudio de cohorte prospectivo (arriba) y retrospectivo (abajo).

Ejemplo de cohorte prospectiva

En 1951, la Asociación de Médicos Británicos hizo un estudio de cohorte prospectiva, respecto a los hábitos de tabaquismo

en 34.439 profesionales en salud. El seguimiento a 20 años de esta cohorte de médicos demostró que alrededor de la mitad de los tabaquistas morían como consecuencia de su hábito, con un exceso de muertes relacionadas al cáncer y a la enfermedad cardiovascular. Los resultados de esta cohorte indujeron a un cambio en los hábitos de tabaquismo en la historia reciente.

Estudio controlado aleatorizado

El estudio o ensayo controlado aleatorizado involucra a una cohorte de sujetos cuya exposición al factor de riesgo no es definida arbitrariamente, sino mediante el azar; es decir, de forma aleatoria: donde cada sujeto tiene la misma probabilidad de estar expuesto a un medicamento (o a una intervención) y en donde la selección de cada individuo es independiente de la participación del otro. Además, se caracteriza por “controlar” quien entra o no al estudio, mediante estrictos criterios de inclusión y exclusión, al mismo tiempo que se “controla” la determinación del evento de interés.

El proceso de aleatorización permite que los grupos expuestos y no expuestos al factor de riesgo sean en todo comparables, ya que sus únicas diferencias son debidas al azar. Estos ensayos pueden tener a su vez varios diseños, como por ejemplo: el diseño paralelo clásico, en el cual dos o más grupos son asignados a las intervenciones o a placebo; el diseño entrecruzado, en el que los sujetos son asignados de forma aleatoria a la secuencia de dos o más tratamientos en diferente orden; y factoriales, en donde los participantes se asignan a más de un tratamiento.

Clasificación de los ensayos clínicos según la fase de desarrollo de un fármaco

La industria farmacéutica ha adoptado una clasificación de ensayos clínicos basada en el desarrollo de un medicamento. La fase I hace referencia a los estudios experimentales en

sujetos sanos o en pacientes que no suelen responder a otras terapias disponibles, con el objetivo de determinar parámetros farmacocinéticos y/o farmacodinámicos, mientras determinan la seguridad y la tolerancia de dicho fármaco. En la fase II se examinan las curvas de dosis-respuesta y se identifica la actividad terapéutica del nuevo agente farmacológico. La fase III es aquella en la cual suelen emplearse los ensayos controlados y aleatorizados, con el fin de comparar el nuevo producto con un estándar o placebo. Esta es la fase crucial del desarrollo del medicamento, ya que establece la eficacia terapéutica y suele ser el requerido por agencias regulatorias para la comercialización del nuevo fármaco. La última fase (IV), suele realizarse para obtener información a partir de un grupo amplio de pacientes en el contexto de la práctica clínica usual y brinda importantes hallazgos para la farmacovigilancia. Estas fases se resumen en la Figura N° 3.

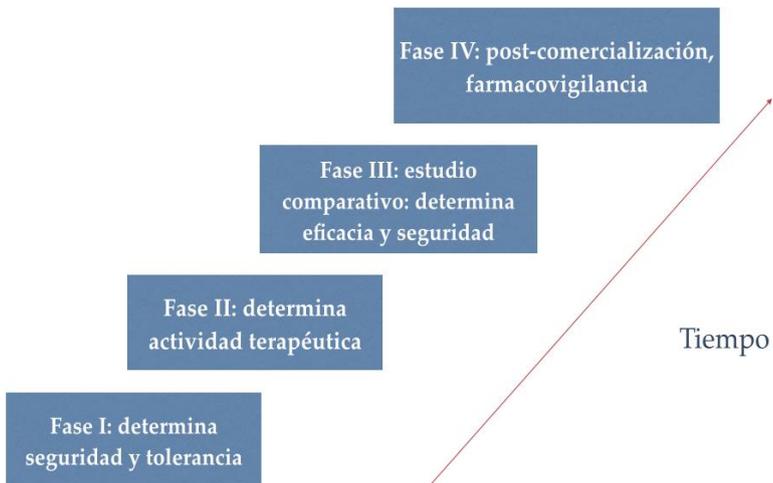


Figura N° 3. Fases de los ensayos clínicos experimentales.

Revisiones sistemáticas y metaanálisis

Las revisiones sistemáticas consisten en revisiones bibliográficas de diversas fuentes, las cuales son definidas *a priori* por el investigador de forma sistematizada, de manera que cualquier otra persona pueda replicarla con los criterios de inclusión y exclusión específicos. Esta revisión debe contar, por lo tanto, con todas las especificaciones necesarias para incluir los estudios en un área específica. En ocasiones se pretende también obtener el efecto combinado de varios estudios; cuando esto ocurre, empleando un método estadístico específico, se habla de un metaanálisis.

Este tipo de estudios son útiles cuando se quieren conciliar resultados disímiles de varios ensayos, o bien, cuando se quieren resumir en un solo estudio los hallazgos de varios artículos publicados.

Los diseños de todos los estudios antes descritos también tienen relación directa con cada una de las preguntas de diagnóstico, etiología, terapéutica y pronóstico. En el Cuadro N° 4 se detallan las características básicas de los principales diseños de estudios, de acuerdo al tipo de pregunta que intentan responder.

Cuadro N° 4. Diseños de estudios clínicos		
Pregunta	Diseño del estudio	Características
Diagnóstico	Estudio transversal	El diseño de este estudio implica idealmente una comparación independiente y cegada con el comparador estándar o "Gold standard".

Cuadro N° 4. Diseños de estudios clínicos		
Pregunta	Diseño del estudio	Características
Etiología/ factores de riesgo	Estudios de cohorte	Los eventos de interés son comparados en un grupo expuesto <i>versus</i> un grupo no expuesto al factor de riesgo, de forma retrospectiva o prospectiva.
	Estudios caso-control	De forma retrospectiva se comparan los sujetos con el evento de interés (casos) <i>versus</i> un grupo sin tal evento (controles).
Intervención/ tratamiento	Estudio controlado aleatorizado	Los sujetos se asignan de forma aleatoria a un grupo con una intervención o a un grupo control y se comparan prospectivamente, en tanto se contabiliza el evento de interés en cada grupo.
Pronóstico	Estudio de cohorte	Seguimiento a largo plazo de un grupo de sujetos (cohorte) expuestos o no a potenciales factores de riesgo.

¿QUÉ ES Y CÓMO SE MIDE LA INCIDENCIA?

La mayoría de preguntas clínicas pueden contestarse mediante la determinación de la incidencia de un evento específico bajo ciertas condiciones. Este evento de interés u “*outcome*” se conoce también como objetivo primario. Por ejemplo, para responder la pregunta ¿cuál es el riesgo de una mujer embarazada infectada con el virus del Zika de tener un producto con microcefalia?, es necesario conocer la incidencia de este evento de interés (microcefalia) en esa población específica. Similarmente, si se quiere saber la eficacia de la aspirina para prevenir infartos, se debe calcular la incidencia de estos eventos en una población que consume el antiagregante plaquetario.

Por lo tanto, es necesario comprender qué es la incidencia y cómo se calcula. La definición más simple indica que es el número de nuevos casos que ocurren en un tiempo determinado. En general, existen dos formas de definir la incidencia: la incidencia acumulada y la tasa de incidencia.

Incidencia acumulada

La incidencia acumulada se calcula fácilmente (Fórmula 1). Será siempre un número entre cero y uno, por lo que suele expresarse en porcentaje (%). Dado que varía entre cero y uno, también se le llama riesgo (en inglés: *risk = R*).

- Fórmula 1:

$$\text{Incidencia acumulada} = \text{casos} / \text{total de pacientes}$$

Tasa de incidencia

Otra forma de expresar los casos que ocurren en un período determinado es mediante la tasa de incidencia (en inglés: *hazard = h*). Para ello, es necesario saber el número de casos, así como el tiempo que cada individuo estuvo en riesgo de

presentar el evento de interés (Fórmula 2). La tasa de incidencia, tasa de riesgo o *hazard*, es particularmente útil cuando los eventos ocurren a lo largo del tiempo, como por ejemplo: la mortalidad global o específica por patología, los eventos cardiovasculares (angina, infartos fatales y no fatales, eventos cerebrovasculares isquémicos o hemorrágicos) o los eventos de progresión en pacientes con cáncer.

- Fórmula 2:

$$\text{Tasa de incidencia} = \text{casos} / \sum \text{personas} * \text{tiempo}$$

El uso de estos indicadores dependerá muchas veces de si el evento de interés ocurre en condiciones agudas o crónicas, tal como lo ejemplifica el Cuadro N° 5.

Cuadro N° 5. Características de los eventos clínicos en condiciones agudas y crónicas		
	Condiciones agudas	Condiciones crónicas
Tipo de intervención	Procedimiento/ medicamento	Usualmente un medicamento
Duración de la intervención	Usualmente fijo/ dosis única	Usualmente variable
Ocurrencia del evento de interés	Los eventos se agrupan en un tiempo corto	Los eventos se distribuyen a lo largo del seguimiento
Duración de seguimiento	Fijo y corto, hasta que se resuelva la condición	Arbitrario, hasta que ocurra un evento relacionado con la enfermedad de estudio
Tipo de incidencia reportada	Incidencia acumulada	Tasa de incidencia

Para ilustrar el cuadro anterior, se pueden considerar, por ejemplo, las fases aguda y crónica del infarto del miocardio. Los pacientes en la fase aguda del infarto suelen recibir fibrinolíticos (como la estreptoquinasa) o anticoagulantes durante un período fijo de tiempo (horas o pocos días); en esta fase predominan los eventos agudos, como las arritmias fatales o los sangrados por tratamiento, los cuales pueden determinarse en un período de seguimiento relativamente corto, de pocos días o semanas. En estas circunstancias es preferible el reporte de incidencias acumuladas o riesgos de tales eventos.

En contraste, los supervivientes de un infarto de miocardio entran en una fase crónica, donde pueden morir por causas cardiovasculares (infarto, insuficiencia cardiaca, arritmia fatal) o no cardiovasculares (cáncer, accidentes). Estos eventos se denominan “competitivos”, debido a que cualquiera puede ocurrir durante el período de seguimiento. La duración del seguimiento determinará cuántos eventos ocurrirán. En estos casos, la intervención médica (por ejemplo, antiagregantes plaquetarios y vasodilatadores) suele administrarse de forma continua hasta que ocurran nuevos eventos cardiovasculares, hasta que se presente algún evento adverso o hasta la muerte del paciente. En estas condiciones crónicas, es recomendable el uso de las tasas de incidencia, ya que lo que interesa saber es cuándo ocurrirá el evento de interés.

Diferencia entre incidencia acumulada y tasa de incidencia

A continuación se muestran dos ejemplos para comprender mejor la diferencia entre la incidencia acumulada (*risk*) y la tasa de incidencia (*hazard*).

- Ejemplo 1:

El objetivo del estudio es determinar la incidencia de mortalidad en pacientes con cáncer de pulmón que inician tratamiento con una quimioterapia. Para ello, se escoge una cohorte de ocho

pacientes, representados por las líneas horizontales en la Figura N° 4, a quienes se les da seguimiento hasta por 24 meses con el tratamiento citotóxico; los pacientes que fallecen son representados en la Figura N° 4 con flechas. A la par de cada paciente se describe el tiempo de seguimiento a lo largo del ensayo clínico.

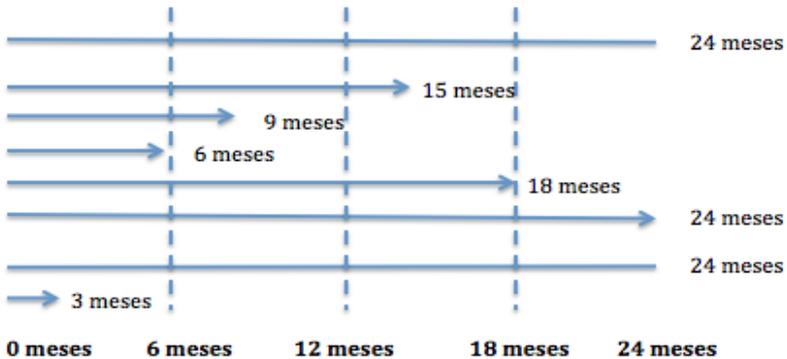


Figura N° 4. Esquema del desarrollo de un estudio de cohorte.

El cálculo de la incidencia acumulada de mortalidad a lo largo del plazo total del estudio, según la fórmula correspondiente (Fórmula 1) sería:

$$\text{Incidencia acumulada} = \text{casos} / \text{total de pacientes}$$

$$\text{Incidencia acumulada} = 6 / 8 = 0,75 = 75 \%$$

Es decir, la incidencia acumulada del evento de interés (mortalidad) es de 75 % durante los dos años de seguimiento. Se debe recordar que debido a que las incidencias acumuladas siempre son números entre cero y uno, suelen reportarse como porcentajes y con frecuencia se hace referencia a este valor como el riesgo del evento. En este sentido, es importante tener presente que un riesgo es igual a una probabilidad, ya que su valor mínimo es cero y su valor máximo es uno. De tal forma,

en el ejemplo anterior, es correcto afirmar que el riesgo de muerte o la probabilidad de muerte de los pacientes con cáncer de pulmón tratados con quimioterapia es del 75 % durante los 24 meses de seguimiento.

Por otra parte, para conocer la tasa de incidencia (en este caso la tasa de mortalidad), se emplea la fórmula correspondiente (Fórmula 2), donde el numerador es también el número de casos del evento de interés, es decir, seis sujetos; sin embargo, el denominador es la sumatoria del tiempo en riesgo de toda la población (ver Figura N° 4). De tal manera, la tasa de incidencia sería:

$$\text{Tasa de incidencia} = \text{casos} / \sum \text{personas} * \text{tiempo}$$

$$\text{Tasa de incidencia} = 6 / (\Sigma = 24 + 15 + 9 + 6 + 18 + 24 + 24 + 3)$$

$$\text{Tasa de incidencia} = 6 / 123 = 0,049 \text{ eventos/persona*mes}$$

Al transformar los meses a años (123 meses = 10,25 años) se obtienen 0,58 eventos/personas*año, que equivale a decir que ocurrieron 58 muertes por cada 100 personas*año. En otras palabras, si se diera tratamiento quimioterapéutico a 100 personas con cáncer de pulmón durante un año, 58 de ellas morirían.

De acuerdo con lo mencionado, se puede observar que aunque en muchas ocasiones los términos riesgo (incidencia acumulada) y tasa de incidencia (*hazard*) se emplean de forma indistinta en la literatura médica, ambos conceptos no son sinónimos; un riesgo corresponde a un valor porcentual entre cero y uno, mientras que el valor de las tasas de incidencia varía desde cero hasta infinito y siempre deben reportarse tomando en cuenta la dimensión temporal.

- Ejemplo 2:

En un estudio clínico suele ocurrir que no todos los pacientes son seguidos a lo largo del período de seguimiento planeado, ya que muchas veces ocurren pérdidas. Además, no todos los pacientes ingresan al estudio al mismo tiempo, pudiendo ocurrir un esquema como el de la Figura N° 5, en donde el primer y el penúltimo paciente se perdieron durante el seguimiento (representados por un círculo al final de la línea); el último paciente se retiró del estudio momentáneamente, pero luego regresó; y el cuarto y el séptimo paciente ingresaron al ensayo seis meses después de haberse iniciado el reclutamiento. Este ejemplo ilustra, por lo tanto, una cohorte dinámica.

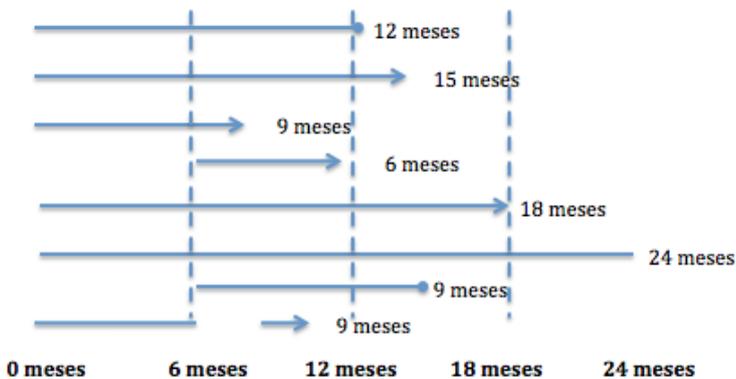


Figura N° 5. Esquema del desarrollo de una cohorte dinámica.

En este caso se desea nuevamente calcular la tasa de mortalidad, pero al hacerlo surgen dos preguntas: ¿Cuál tiempo debo incluir en el denominador de la fórmula? ¿Debo incluir al primer y al séptimo paciente que se salieron del estudio? La respuesta a ambas interrogantes es que se debe incluir el tiempo de todos los pacientes que estuvieron en riesgo de tener el evento, así se hubieran salido del estudio. Lo anterior porque se tiene conocimiento que durante dicho tiempo el paciente estuvo vivo o en riesgo de morir. Por lo tanto, aplicando la Fórmula 2, la tasa de mortalidad sería:

$$\text{Tasa de incidencia} = \text{casos} / \sum \text{personas} * \text{tiempo}$$

$$\text{Tasa de incidencia} = 5 / (\Sigma = 12 + 15 + 9 + 6 + 18 + 24 + 9 + 9)$$

$$\text{Tasa de incidencia} = 5 / 102 = 0,049 \text{ eventos/persona} * \text{mes}$$

$$0,049 * 12 = 0,588 \text{ eventos/persona} * \text{año} = 59 \text{ eventos} / 100 \text{ personas} * \text{año}$$

La tasa calculada de 59 eventos/100 personas*año implica que si se siguiesen a 100 personas por un año, 59 de estas fallecerían. Este número es muy distinto del riesgo de morir, el cual, aplicando la Fórmula 1, sería del 62,5 %.

Queda claro, entonces, que las tasas de incidencia (*hazards*) y los riesgos o incidencias acumuladas (*risks*) son dos medidas distintas. Sin embargo, si las tasas permanecen constantes a lo largo del tiempo de seguimiento, es posible relacionar ambas mediciones a través de la Fórmula 3:

$$\text{Tasa de incidencia} = 1 - e^{(\text{incidencia acumulada} * \text{tiempo})}$$

TRABAJANDO CON LA MUESTRA: LOS ERRORES DE LA ESTIMACIÓN

En la mayoría de las investigaciones biomédicas es imposible conocer el comportamiento de una población entera, ya que usualmente es infinita, o bien, existen impedimentos técnicos que no permiten comprender con certeza lo que ocurre en el total de individuos que la constituyen. En otras palabras, se podría considerar que lo que pasa en una población es prácticamente desconocido, a raíz de la imposibilidad de estudiar cada uno de los individuos que la componen. Por eso, es necesario trabajar con una parte de la población, denominada muestra, la cual debería reflejar a plenitud las características de la población que representa. De esta forma, al estudiar lo que ocurre en la muestra, se podría inferir lo que ocurriría en la

población general. Este proceso de establecer conclusiones de una población a partir del comportamiento de una muestra, es la base de la estadística inferencial. Sin embargo, para que tales conclusiones sean válidas, es necesario que la muestra sea una representación fiel de la población; esto puede conseguirse con el proceso de muestreo aleatorio, el cual se basa en la premisa de que cada individuo de la población tiene las mismas probabilidades de ser elegido en la muestra y que dicha elección es independiente una de la otra.

Dentro de este contexto, la población de referencia es aquel conjunto de individuos para quienes se desea extrapolar la información de un estudio muestral. A partir de esta población puede describirse una categoría de sujetos elegibles (población elegible), en quienes se aplican criterios de inclusión y exclusión para ser aceptados en el ensayo, constituyéndose finalmente la muestra (Figura N° 6).



Figura N° 6. Representación gráfica de la población de referencia (N) y de la muestra (n).

Es importante tener presente que siempre se debe analizar con cuidado la manera en que se elige la muestra de cada estudio clínico. De igual forma, se debe considerar que la selección de los individuos en una muestra debe estar basada en el objetivo del estudio, para garantizar la validez externa.

Cuadro N° 6. Definiciones importantes para comprender los errores de la estimación

- **Estadístico:** medida cuantitativa derivada de un conjunto de datos.
- **Parámetro:** estadístico poblacional; se describe en letras griegas. Ejemplos: media (μ), desviación estándar (σ), proporción (π).
- **Estimado:** estadístico muestral; se describe en letras minúsculas. Ejemplos: media (x), desviación estándar (sd), proporción (p).

Errores de la estimación

A pesar de que existen diversos métodos estadísticos de muestreo, resulta fácil suponer que la inferencia de los parámetros de la población, a partir de la muestra, no siempre es completamente exacta, y que existirán errores hipotéticos que deben ser delimitados a la hora de seleccionar dicha muestra. Estos errores pueden ser de dos tipos:

1. **Errores de precisión.**
2. **Errores de validez (exactitud).**

Para ilustrar estos dos conceptos se puede utilizar la Figura N° 7.

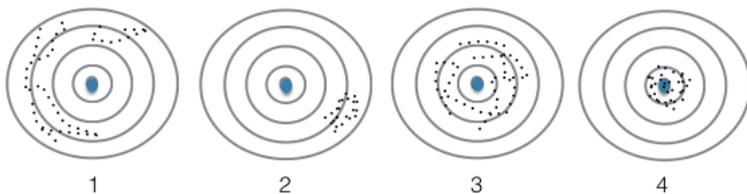


Figura N° 7. Diferencias entre precisión y exactitud. La precisión hace referencia a la dispersión de los puntos, mientras que la exactitud hace referencia al "blanco" o verdadero valor.

Si se supone que el valor central corresponde al verdadero valor del parámetro de la población, como la media de edad, pero este valor es desconocido en la población y no es posible conocerlo (por las características infinitas de la población o por limitaciones técnicas), se deben realizar varios muestreos para calcular, en este caso, la media de la edad de la muestra, de forma tal que esta sea representativa de la verdadera edad de la población.

Cada muestreo (denotado por cada punto al blanco) puede alejarse del verdadero valor de la población, en cuyo caso el cálculo estimado será inexacto (casos 1 y 2). Sin embargo, también puede ocurrir que el valor medio de la edad calculado en una muestra sea muy distinto al valor medio que hubiera sido calculado en otra muestra de la misma población, en cuyo caso el valor será impreciso (casos 1 y 3). Lo ideal es que la inferencia del parámetro de la población sea exacto y preciso (caso 4). En este último caso, los valores de cada muestra se aproximan al verdadero parámetro de la población y la variabilidad de tal dato entre muestras es pequeña.

De esta forma, a la hora de calcular el estimado del parámetro de la población en la muestra, se debe hacer referencia a su precisión y exactitud. Matemáticamente, la precisión puede medirse y reportarse en términos numéricos, no así la exactitud de un estimado, precisamente porque el verdadero valor de la población es desconocido. Por lo tanto, la estimación siempre tendrá consigo un error inherente. Estos tipos de errores pueden deberse al desarrollo del estudio mismo (error sistemático) o pueden ser secundarios a la propia estimación (error aleatorio).

Error sistemático

Los errores sistemáticos son también denominados sesgos. Corresponden a errores que limitan la validez interna del estudio, ya que sobrestiman o subestiman los estimados calculados en la muestra.

Estos errores se agrupan en tres categorías principales: sesgos de selección, sesgos de clasificación y sesgos de confusión, y ninguno de ellos se relaciona con el tamaño de la muestra.

Sesgos de selección

Suponiendo que se quiere estudiar el peso medio de la población de estudiantes de una escuela y para ello se toma una muestra de estudiantes de último año, evidentemente, el cálculo de la media del peso estará sesgado por haber elegido una muestra no representativa de la población estudiantil. Por lo tanto, en este caso, el sesgo surge de la inadecuada selección de participantes, a la vez que confiere una violación a la validez externa del estudio.

Sesgos de clasificación

Si en el ejemplo anterior se le pregunta el peso a cada participante, y tanto hombres como mujeres mienten al reportar el verdadero peso corporal, se estaría calculando un peso medio distinto al que presenta la población. Lo mismo ocurriría si a unos sujetos se les determina el peso a partir de una balanza electrónica y a otros mediante una entrevista personal. Ambos casos constituyen un sesgo de clasificación, ya que durante el estudio las personas se clasificarían erróneamente.

Sesgos de confusión

Un sesgo de confusión se presenta, por ejemplo, cuando se quieren estudiar las causas de cáncer de laringe y para tal efecto se les pregunta a varios individuos sobre los hábitos de consumo diario de café. Después de desarrollado el estudio, se concluye que el consumo de café se asocia al cáncer de laringe. Sin embargo, en dicho estudio se omitió preguntar por el hábito de tabaquismo de los sujetos, lo cual era relevante, ya que cuando se toma en cuenta la historia de fumado previo, se observa que la asociación entre el consumo de café y el cáncer de laringe ya no existe, por cuanto aquellos sujetos que

tomaban café también eran tabaquistas. Es decir, una variable (consumo de café) está asociada a otra (tabaquismo), la cual sí está relacionada con la variable de interés (cáncer de laringe). En este ejemplo se dice que la asociación del café con el cáncer de laringe está *confundida* por el hábito del tabaquismo.

En general, los errores sistemáticos pueden corregirse al inicio de la investigación, mediante la inclusión adecuada de individuos en la muestra, la adecuada clasificación del evento y la adición de potenciales variables confusoras. Al respecto, cabe destacar que los sesgos de confusión pueden corregirse con análisis estadísticos, no así los otros dos tipos de errores.

Error aleatorio

El error aleatorio es aquel debido al azar, que se produce por el hecho de trabajar con muestras de individuos y no con toda la población.

Cuando en un estudio clínico se trabaja con una muestra de una población específica, al momento de la interpretación se pueden alcanzar dos conclusiones básicas: considerar que hay diferencias entre las intervenciones o bien, concluir que no existen tales discrepancias. Dado que nunca se conocerá cuál de ambas conclusiones es la verdadera en la población, se pueden presentar cuatro escenarios:

1. Concluir que las intervenciones en la muestra son distintas, cuando en la población también existían diferencias (verdadero positivo).
2. Concluir que las intervenciones en la muestra no son distintas, cuando en la población tampoco existían diferencias (verdadero negativo).
3. Concluir que las intervenciones en la muestra son distintas, cuando en la población no existían tales diferencias (falso positivo).

4. Concluir que las intervenciones en la muestra no son distintas, cuando en la población sí existían diferencias (falso negativo).

En el Cuadro N° 6 se resumen esos posibles escenarios.

Cuadro N° 6. Descripción de los posibles resultados de un estudio muestral con referencia a la población		
	Hay diferencias entre las intervenciones en la muestra	No hay diferencias entre las intervenciones en la muestra
Hay diferencias entre las intervenciones en la población	Verdadero positivo 😊	Falso negativo 😞
No hay diferencias entre las intervenciones en la población	Falso positivo 😞	Verdadero negativo 😊

Así pues, hay dos errores intrínsecos a la estimación de parámetros de la población a partir de la muestra. El error tipo I (error α) que hace referencia a los falsos positivos y el error tipo II (error β) que corresponde a los falsos negativos. Como estos errores son inherentes al proceso de estimación y no es posible evitarlos, se deben tratar de mantener al nivel más bajo que se pueda. En general, en la literatura médica se acepta un error α del 5 % y un error β del 20 %. El porcentaje de estos errores viene definido sobre todo por el tamaño de la muestra: entre mayor sea el número de individuos en la muestra menores los errores tipos α y β . En ocasiones, la literatura científica no hace referencia explícita al error tipo II (β), y solo menciona el “poder estadístico” del estudio, siendo este calculado a partir de la siguiente fórmula (Fórmula 4):

$$\text{Poder estadístico} = 1 - \beta$$

El poder estadístico es la probabilidad que tiene el estudio de estimar diferencias en la muestra que son ciertas en la población. Nótese que para la mayoría de los estudios clínicos este valor alcanza el 80 %.

En general, el poder del estudio dependerá de:

1. El tamaño de la muestra: a mayor muestra, mayor poder. De forma análoga, una muestra pequeña tendrá alta probabilidad de presentar falsos negativos.
2. La variabilidad de las observaciones: a mayor variabilidad menor poder.
3. El efecto de interés: el poder será mayor si los efectos de la intervención son mayores.
4. El error tipo I: entre mayor sea el error α , mayor será el poder estadístico.

BIBLIOGRAFÍA

1. Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH, Arensberg D, Baker A, Friedman L, Greene HL, *et al.* Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. *N Engl J Med.* 1991; 324(12): 781-788.
2. Anderson JL, Stewart JR, Perry BA, Van Hamersveld D, Johnson T, Conard G, Chang S, Kvam D, Pitt B. Oral flecainide acetate for the treatment of ventricular arrhythmias. *N Engl J Med.* 1981; 305(9): 473-477.
3. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA.* 1992; 268(17): 2420-2425.
4. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanus F, McQueen M, Budaj A, Pais P, Varigos J, Lisheng L. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet.* 2004; 364(9438): 11-17.
5. Adebisi AO. Bias: a review of current understanding. *Afr J Med Sci.* 2010; 39: 241-248.
6. Altman D. *Practical statistics for medical research.* New York, NY: Chapman & Hall; 1991.

7. Bowers D. *Medical statistics from scratch: An introduction for health professionals*. 3 ed. Leeds, UK: Wiley Blackwell; 2014.
8. Colditz GA. Overview of the epidemiology methods and applications: strengths and limitations of observational study designs. *Crit Rev Food Sci Nutr*. 2010; 50(Suppl.1): 10-12.
9. Daniel W. *Bioestadística. Base para el análisis de las ciencias de la salud*. 4 ed. México: Limusa Wiley; 2004.
10. Doll R, Peto R, Wheatley K, Gray R, Sutherland I. Mortality in relation to smoking: 40 years' observations on male British doctors. *BMJ*. 1994; 309(6959): 901-911.
11. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. 2006; 60(7): 578-586.
12. Hill AB. The environment and disease: association or causation. *Proc Royal Soc Med*. 1965; 58(5): 295-300.
13. Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. *Designing clinical research*. 3 ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2007.
14. Kuller LH. Relationship between acute and chronic disease epidemiology. *Yale J Biol Med*. 1987; 60(4):363-370.
15. Mathews DE, Farewell VT. *Using and understanding medical statistics*. 4 ed. Basel: Karger; 2007.
16. Petrie A, Sabin C. *Medical statistics at a glance*. 3 ed. London: Wiley-Blackwell; 2009.
17. Sackett DL, Rosenberg W, Mc Gray JA, Haynes RB, Richardson WS. Evidence-based medicine: what it is and what it isn't. *BMJ*. 1996; 312(7023): 71-72.
18. Spruance SL, Reid JE, Grace M, Samore M. Hazard ratio in clinical trials. *Antimicrob Agents Chemother*. 2004; 48(8): 2787-2792.
19. Suzuki E, Mitsuhashi T, Tsuda T, Yamamoto E. A typology of four notions of confounding in epidemiology. *J Epidemiol*. 2017; 27(2): 49-55.
20. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Selection bias and information bias in clinical research. *Nephron Clin Pract*. 2010; 115(2): 94-99.
21. Vetter TR, Mascha EJ. Bias, confounding, and interaction: lions and tigers, and bears, oh my! *Anesth Analg*. 2017; 125(3): 1042-1048.
22. Wang D, Bakhai A. *Clinical trials: a practical guide to design, analysis, and reporting*. Chicago, IL: Remedica; 2006.

CAPÍTULO 2. CONCEPTOS DEL ANÁLISIS ESTADÍSTICO

MEDIDAS DE ASOCIACIÓN

Tal como se discutió en el capítulo anterior, en la mayoría de investigaciones que pretenden contestar las cuatro preguntas básicas de diagnóstico, etiología, tratamiento y pronóstico, es posible calcular la incidencia acumulada y la tasa de incidencia del evento de interés.

La comparación de dichas medidas entre grupos permitirá establecer si hay diferencias entre el grupo intervención y el grupo control. Para ello se emplean las medidas de asociación: riesgo relativo, riesgo absoluto, radio de tasas o *hazard ratio* y *odds ratio*.

Riesgo relativo

Se debe recordar que las incidencias acumuladas corresponden a probabilidades o riesgos de un evento de interés, los cuales se calculan dividiendo el número de eventos entre el total de individuos de la muestra (Fórmula 1). Así pues, el riesgo relativo es la división del riesgo de presentar un evento en un grupo *versus* el riesgo del otro grupo (Fórmula 5). Por razones prácticas, suele colocarse en el numerador el grupo intervención. Al ser una división, un riesgo relativo igual a uno implicará que ambos grupos comparten el mismo riesgo de eventos.

- Fórmula 5:

$$\text{Riesgo relativo} = \frac{\text{Riesgo grupo 1}}{\text{Riesgo grupo 2}}$$

Usualmente, en la literatura se reporta también la reducción del riesgo relativo, para lo cual se emplea la Fórmula 6.

- Fórmula 6:

$$\text{Reducción del Riesgo relativo} = 1 - \frac{\text{Riesgo grupo 1}}{\text{Riesgo grupo 2}}$$

Riesgo relativo = 1	→	No hay diferencias entre grupos.
Riesgo relativo < 1	→	El grupo intervención (numerador) tiene menos riesgo de eventos que el grupo control (denominador).
Riesgo relativo > 1	→	El grupo intervención (numerador) tiene más riesgo de eventos que el grupo control.

Riesgo absoluto

Si en vez de dividir los riesgos de los dos grupos estos se restan, es posible obtener una medida absoluta de la diferencia entre ambos riesgos. Un riesgo absoluto de cero significaría que no hay diferencia entre los grupos.

- Fórmula 7:

$$\text{Riesgo absoluto} = \text{Riesgo grupo 1} - \text{Riesgo grupo 2}$$

Riesgo absoluto = 0	→	No hay diferencias entre grupos.
Riesgo absoluto estrecho	→	Las diferencias son marginales.

Riesgo absoluto amplio	→	La intervención difiere en alto grado del grupo control.
------------------------	---	--

Razón de tasas (*Hazard ratio*)

En este caso, la división que se efectúa (Fórmula 8) es entre las tasas de incidencia (calculadas con la Fórmula 2). Nuevamente, al tratarse de una división, un resultado del *Hazard ratio* de uno significa que no hay diferencia en las incidencias del evento de interés entre los grupos comparados.

- Fórmula 8:

$$\text{Razón de tasas (Hazard ratio)} = \frac{\text{Tasa de incidencia grupo 1}}{\text{Tasa de incidencia grupo 2}}$$

<i>Hazard ratio</i> = 1	→	No hay diferencias entre grupos.
<i>Hazard ratio</i> < 1	→	El grupo intervención (numerador) tiene menos eventos que el grupo control (denominador).
<i>Hazard ratio</i> > 1	→	El grupo intervención (numerador) tiene más eventos que el grupo control.

Odds ratio

Esta última medida de asociación surge a partir de la definición de *odds* (en castellano: “momios”), que es la razón de eventos opuestos; por ejemplo, pacientes fallecidos/pacientes vivos.

Una definición matemática de *odds* se presenta en la Fórmula 9:

$$\text{Odds} = \frac{\text{Pacientes con el evento de interés}}{\text{Pacientes sin el evento de interés}}$$

El *odds* puede calcularse a partir de la probabilidad de un evento, debido a que la probabilidad de no ocurrencia del evento es complementaria (es decir, ambas suman uno). De ahí la Fórmula 10:

$$\text{Odds} = \frac{\text{Probabilidad del evento}}{1 - \text{probabilidad del evento}}$$

Despejando dicha fórmula se obtiene la Fórmula 11:

$$\text{Probabilidad del evento} = \frac{\text{Odds}}{1 + \text{Odds}}$$

De esta manera, el *odds ratio* será el *odds* de un evento en el grupo uno (usualmente el grupo intervención) entre el *odds* del mismo evento del grupo control, tal como se describe en la Fórmula 12:

$$\text{Odds ratio} = \frac{\text{Odds del grupo 1}}{\text{Odds del grupo 2}}$$

$\text{Odds ratio} = 1$	→	No hay diferencias entre grupos.
$\text{Odds ratio} < 1$	→	El grupo intervención (numerador) tiene menos eventos que el grupo control (denominador).
$\text{Odds ratio} > 1$	→	El grupo intervención (numerador) tiene más eventos que el grupo control.

Número necesario a tratar (*Number Needed to Treat: NNT*)

Debido a que muchas intervenciones no son 100 % efectivas en todos los pacientes, con frecuencia suele plantearse la siguiente pregunta: ¿cuántos sujetos hay que tratar para que uno de ellos obtenga un beneficio? Por ejemplo, si una intervención da un beneficio absoluto del 20 % sobre placebo, también se puede interpretar que, al tratar a cinco pacientes, solo uno de ellos tendría un beneficio. De ahí la importancia de calcular el NNT. La fórmula general para esta medida de asociación sería:

- Fórmula 13:

$$NNT = \frac{1}{\text{Riesgo absoluto}}$$

Este cálculo permite formarse una idea de la efectividad de la intervención. Un NNT de uno es el valor más efectivo, ya que implica que cada paciente del grupo intervención recibe un beneficio, mientras que ningún sujeto del grupo control lo recibe.

Tal como lo evidencia la Fórmula 13, el NNT solo puede calcularse a partir de riesgos o probabilidades, no así a partir de tasas.

Es importante aclarar que a la hora de reportar el NNT de una determinada intervención, se debe hacer énfasis en el tiempo de seguimiento, ya que los riesgos de muchos eventos no son constantes en el tiempo.

Número necesario para hacer daño (*Number Needed to Harm: NNH*)

El NNH tiene las mismas implicaciones e interpretaciones que el NNT, pero tratándose de efectos adversos de alguna intervención.

- Fórmula 14:

$$NNH = \frac{1}{\text{Riesgo absoluto de eventos adversos}}$$

Ejemplos de medidas de asociación

- **Ejemplo 1:**

El estudio ISIS-2 comparó la mortalidad de sujetos con infarto agudo al miocardio que recibieron estreptoquinasa o placebo (1). A los 35 días, la incidencia acumulada de muertes fue de 12,1 % para placebo y 9,3 % para estreptoquinasa.

A la vez, se reportó un exceso absoluto de 0,3 % sangrados que requirieron transfusión en el grupo de pacientes que recibió estreptoquinasa. El cálculo de las medidas de asociación para este estudio sería:

Riesgo relativo de mortalidad

- Fórmula 5:

$$\text{Riesgo relativo} = \frac{\text{Riesgo grupo 1}}{\text{Riesgo grupo 2}}$$

$$\text{Riesgo relativo} = \frac{9,3}{12,1} = 0,77$$

El riesgo relativo de muerte (estreptoquinasa *versus* placebo) es de 0,77. Dado que el riesgo relativo de uno implica que no hay diferencias entre grupos, un riesgo relativo de 0,77 implica una reducción del 23 % del riesgo de muerte en el grupo asignado a estreptoquinasa.

Riesgo absoluto de mortalidad

- Fórmula 7:

$$\text{Riesgo absoluto} = \text{Riesgo grupo 1} - \text{Riesgo grupo 2}$$

$$\text{Riesgo absoluto} = 12,1 - 9,3 = 2,8$$

Como se observa, hubo una reducción del 2,8 % en el riesgo absoluto de muerte, favoreciendo al grupo asignado a estreptoquinasa. Nótese que se trata siempre de términos absolutos.

Odds ratio de mortalidad

Dado que no se conoce el *odds* de ninguno de los dos grupos, ni el número de pacientes sin el evento en cada grupo, se debe emplear la Fórmula 10 para calcular el *odds* en ambos grupos, para así luego calcular el ratio o división entre ellos (Fórmula 12).

- Fórmula 10:

$$\text{Odds} = \frac{\text{Probabilidad del evento}}{1 - \text{probabilidad del evento}}$$

$$\text{Odds grupo control} = \frac{0,121}{1 - 0,121} = 0,14$$

$$\text{Odds grupo estreptoquinasa} = \frac{0,093}{1 - 0,093} = 0,10$$

- Fórmula 12:

$$\text{Odds ratio} = \frac{\text{Odds del grupo 1}}{\text{Odds del grupo 2}}$$

$$\text{Odds ratio} = \frac{0,10}{0,14} = 0,71$$

Tras aplicar ambas fórmulas, se obtiene como resultado que el *odds ratio* de muerte es de 0,71. Es decir, hubo una reducción del 29 % en el *odds* de fallecer, favoreciendo al grupo que recibió estreptoquinasa.

Número necesario a tratar (NNT)

Para calcular el NNT se utiliza la Fórmula 13, como se observa a continuación:

$$NNT = \frac{1}{\text{Riesgo absoluto}}$$

$$NNT = \frac{1}{0,121 - 0,093} = 35,7$$

En este caso, el NNT es de 35,7, lo cual significa que es necesario tratar a 36 pacientes en cada grupo para producir una muerte menos en el grupo de estreptoquinasa posterior a 35 días de seguimiento.

Número necesario para hacer daño (NNH)

Para calcular el NNH se utiliza la Fórmula 14:

$$NNH = \frac{1}{\text{Riesgo absoluto de eventos adversos}}$$

$$NNH = \frac{1}{0,003} = 333,3$$

En este caso, es posible afirmar que es necesario tratar a 333 pacientes con estreptoquinasa en comparación con placebo, para que uno de ellos tenga un sangrado que amerite transfusión.

- **Ejemplo 2:**

En un estudio se incluyeron 2.000 pacientes. A la mitad se le dio tratamiento activo y al resto placebo. El evento de interés u objetivo final fue determinar la mortalidad total después de cinco años de seguimiento. El Cuadro N° 1 muestra los eventos acumulados por año de tratamiento.

Cuadro N° 1. Resultados de un estudio clínico				
Año	Muertes en cada grupo de tratamiento		Riesgo relativo	Odds ratio
	Tratamiento activo (n= 1.000)	Placebo (n= 1.000)		
1	113	148	0,76	0,73
2	213	274	0,78	0,72
3	302	381	0,79	0,70
4	381	473	0,81	0,69
5	451	551	0,82	0,67

Las últimas dos columnas del cuadro anterior corresponden al riesgo relativo y al *odds ratio* calculados al final de cada año. Nótese que ambas mediciones cambian durante el tiempo de seguimiento. Al final del estudio, se reportó una reducción del riesgo relativo de muerte del 18 %, mientras que en el primer año la reducción fue del 24 %. Este ejemplo ilustra que tanto el riesgo relativo, como el *odds ratio* y el riesgo absoluto (y por ende, el número necesario a tratar) varían con el tiempo. Por el contrario, el *hazard ratio* es una medida de asociación que suele ser constante a lo largo del tiempo.

CONTABILIZANDO LOS EVENTOS: ANÁLISIS POR INTENCIÓN A TRATAR Y ANÁLISIS POR PROTOCOLO

En los estudios clínicos se pueden contabilizar los eventos de interés de diferentes maneras. Dado que el estudio clínico se desarrolla en el transcurso de un tiempo específico, es frecuente que durante el seguimiento ocurran pérdidas de sujetos (“*lost to follow-up*”), o bien, que los sujetos asignados a recibir un tratamiento específico terminen recibiendo el tratamiento del otro grupo (entrecruzamiento o “*cross-over*”). En

ambos casos, el estimado final de la muestra carecerá de validez interna. Por ello, existen dos formas de analizar los eventos; estos son el análisis por intención a tratar (ITT) y el análisis por protocolo (PP).

Análisis por intención a tratar (ITT)

En este tipo de análisis se toman en cuenta todos los sujetos que estaban destinados a recibir el tratamiento propuesto, sin importar si lo recibieron o no. Para el cálculo de la incidencia acumulada del evento de interés o de la tasa de incidencia, todos los sujetos se incluyen en el denominador.

Análisis por protocolo (PP)

En este análisis únicamente se consideran los individuos que completaron el tratamiento asignado. Por tanto, si hubiera pérdidas de sujetos o entrecruzamientos, los cálculos se harán con base en los individuos que efectivamente recibieron el medicamento.

¿Cuál es el mejor tipo de análisis?

En general, la mayoría de autores sugieren que el ITT es el que debe reportarse, ya que permite conservar la comparabilidad de los grupos tal y como fueron aleatorizados. Si bien el análisis PP pareciera ser intuitivamente más fidedigno, no hay certeza de que al final del período de seguimiento los grupos sean comparables. En cambio, el ITT preserva las fortalezas de la aleatorización. Además, este tipo de análisis semeja lo que pasa en condiciones usuales de la vida diaria, ya que muchos sujetos no son adherentes al tratamiento. Sin embargo, pese a estas ventajas, el ITT asume que las pérdidas de seguimiento y entrecruzamientos son iguales en ambos grupos y ocurren de forma aleatoria, situación que no siempre se cumple. Además, dado que este análisis mezcla información de los sujetos adherentes y no adherentes al tratamiento, diluye el efecto final de la intervención. De hecho, en ciertos casos, el análisis por

ITT puede no mostrar diferencias entre grupos, mientras que el análisis PP podría detectarlas. Por otra parte, el análisis PP reporta el máximo beneficio posible de una intervención en condiciones de excelente adherencia y es el método preferido en estudios de no inferioridad.

Independientemente de lo mencionado, es importante resaltar que lo preferible en cualquier estudio clínico es que no ocurran ni pérdidas ni entrecruzamientos y que tanto el análisis PP como el ITT brinden exactamente el mismo valor.

Ejemplo de la aplicación de ambos análisis

Se realiza un estudio clínico con 10 sujetos, quienes se aleatorizan a recibir dos intervenciones distintas. En la Figura N° 1 se ilustran dos casos partiendo de esa premisa.

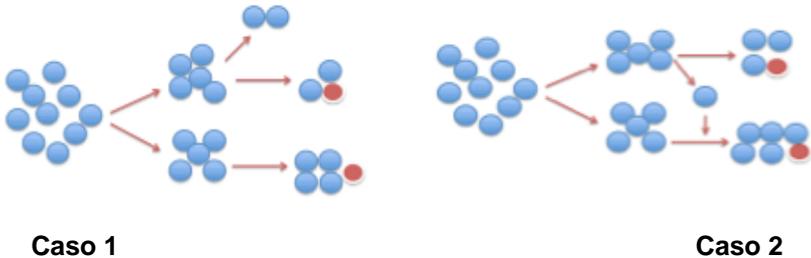


Figura N° 1. Casos hipotéticos de comportamiento de un estudio clínico. En el caso 1 hubo pérdida de sujetos durante el seguimiento, mientras que en el caso 2 hubo entrecruzamiento de pacientes entre grupos. En rojo se observan los sujetos que tuvieron el evento de interés.

En el caso 1, dos de los cinco sujetos asignados a recibir una terapia se “perdieron” al salirse del estudio. En el caso 2, un sujeto al que se le había asignado recibir terapia en un grupo, se “entrecruzó” hacia el grupo comparador.

Suponiendo que al final del tiempo de seguimiento, un individuo de cada grupo tuvo el evento de interés, se procede a calcular el riesgo del evento de interés en cada grupo, según el análisis por intención a tratar y por protocolo.

En el caso 1, según el análisis por protocolo, en el primer grupo el riesgo de tener el evento es de $1/3$, mientras que empleando el análisis por intención a tratar, dicho riesgo es de $1/5$.

En el caso 2, el uso de un análisis por protocolo genera un riesgo del evento de interés en el primer grupo de $1/4$, en tanto que el análisis por intención a tratar vuelve a revelar un riesgo de $1/5$ para dichos sujetos.

Si se calcula el riesgo relativo en el caso 1, según el análisis por intención a tratar, se obtiene el siguiente resultado:

- Fórmula 5:

$$\text{Riesgo relativo} = \frac{\text{Riesgo grupo 1}}{\text{Riesgo grupo 2}}$$

$$\text{Riesgo relativo} = \frac{1/5}{1/5} = 1$$

Si se calcula el riesgo relativo en el mismo caso 1, pero según el análisis por protocolo, se obtiene el siguiente resultado:

$$\text{Riesgo relativo} = \frac{1/3}{1/5} = 1,67$$

Como puede verse, las conclusiones de ambos análisis son opuestas, mientras que el análisis por intención a tratar reporta no diferencias, el estudio por protocolo revela un 67 % más de riesgo del evento de interés en el primer grupo.

De manera análoga, en el ejemplo 2, el análisis por intención a tratar generó un riesgo relativo de 1, en tanto que en el análisis por protocolo el riesgo relativo fue de 1,67.

MIDIENDO LA PRECISIÓN

En todo estudio se debe partir del hecho de que todos los estimados que se calculan en la muestra deben ser precisos y exactos, para ser adecuadamente extrapolables a la población. La exactitud no tiene una forma matemática de definirse, debido a que nunca se sabrá el verdadero valor del parámetro de la población. No obstante, la precisión sí puede definirse matemáticamente, ya que entre mayor sea la muestra, más cerca se estará del verdadero parámetro de la población. Dicho de otra forma, la precisión depende directamente del tamaño muestral.

Para entender cómo medir la precisión, se deben comprender primero las medidas básicas de dispersión de los datos continuos de una muestra o población: la desviación estándar de la media y la varianza.

La media

La media corresponde a la suma de todos los valores individuales de una variable continua, dividido por el número total de individuos.

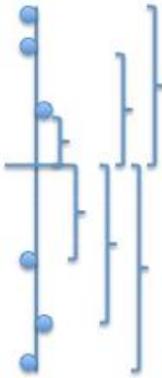
La varianza

La varianza corresponde al promedio de las diferencias cuadráticas de cada dato individual, con respecto al valor medio (Fórmula 15).

- Fórmula 15:

$$\text{Varianza} = \frac{1}{n} \sum_{i=1}^n (x_i - X)^2$$

En la Figura N° 2 se muestra una representación gráfica de la varianza.



Para obtener la varianza:

1. Calcule la diferencia entre cada dato individual y valor medio $(x - X)$
2. Eleve cada diferencia al cuadrado $(x - X)^2$
3. Sume cada término $\sum_{i=1}^n (x - X)^2$
4. Calcule el promedio de esa sumatoria:
 $\frac{1}{n} \sum_{i=1}^n (x - X)^2$

Figura N° 2. Representación gráfica del cálculo de la varianza.

La desviación estándar de la media

La desviación estándar de la media corresponde a la raíz cuadrada de la varianza (Fórmula 16).

- Fórmula 16:

$$\text{Desviación estándar (Sd)} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - X)^2}$$

Campana de Gauss

Cuando se grafica la distribución de frecuencias de la mayoría de valores de una variable continua, esta suele seguir la distribución normal o paramétrica, clásicamente llamada “campana de Gauss”, tal como se muestra en la Figura N° 3. En el eje x de esta gráfica se anotan los valores de las presiones arteriales de una muestra de sujetos sanos, mientras que en el eje y se anotan las frecuencias porcentuales de esos valores. El valor medio (que en este ejemplo es de 120 mmHg)

coincide con la mediana (punto en el cual se encuentra el 50 % de los datos por encima y por debajo del valor medio) y con la moda (valor más frecuente de la muestra). Lo anterior, porque en una distribución normal tanto la media, como la moda y la mediana corresponden al mismo número.

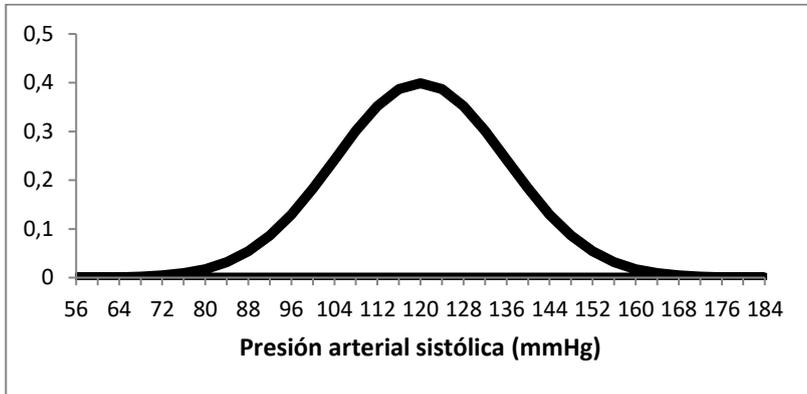


Figura N° 3. Distribución normal de una variable x (presión arterial sistólica en mmHg).

Debido a que los valores de x pueden tomar cualquier número dependiendo de la variable continua que se esté midiendo (por ejemplo: presión arterial, edad, peso, índice de masa corporal, entre otros), estos valores se deben estandarizar, realizando una conversión matemática, de acuerdo a la siguiente fórmula, donde Sd es la desviación estándar, x_i corresponde al valor de la variable x a transformar; y X es la media:

- Fórmula 17:

$$z = \frac{x_i - X}{Sd}$$

Cuando se hace esta transformación matemática, los valores de x ahora se convierten en valores de z , tal como se muestra en la Figura N° 4, donde la media, la moda y la mediana son ahora el valor cero.

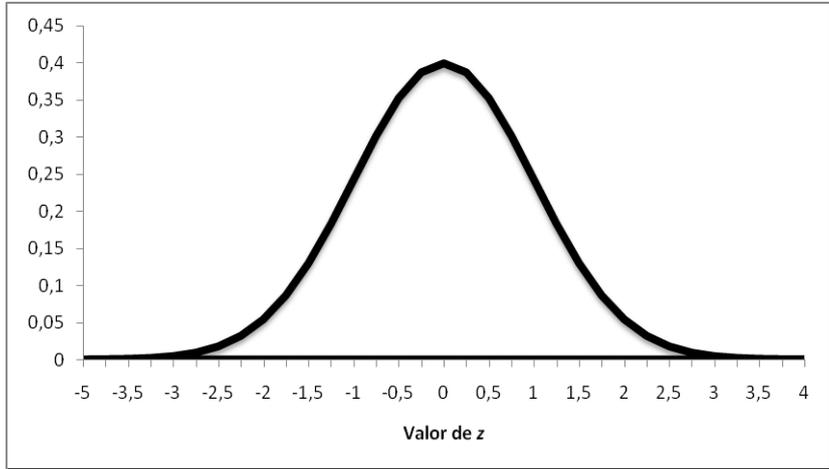


Figura N° 4. Distribución normal estándar.

En el ejemplo anterior, suponiendo una desviación estándar de 10 mmHg, el valor de 120 mmHg al ser transformado por la Fórmula 17 se convierte en cero. Análogamente, el valor de 110 y 130 (que correspondían al valor medio \pm 1 desviación estándar) serán ahora -1 y +1, respectivamente.

- Fórmula 17:

$$z = \frac{x_i - X}{Sd}$$

$$\frac{120-120}{10} = 0; \frac{110-120}{10} = -1; \frac{130-120}{10} = +1$$

Dentro de las características que posee la curva normal, destaca la capacidad de conocer dónde se ubica un porcentaje del total de los datos de la muestra, a partir de cierto valor de z. Por ejemplo, si se toma el área bajo la curva comprendida entre los valores de $z = -1$ y $z = +1$, se podrá obtener el 68,2 % del total de datos alrededor del valor medio. Similarmente, entre los valores de $z = -1,96$ y $z = +1,96$ estarán el 95 % de los datos alrededor de la media.

Para determinar con mayor facilidad el área bajo la curva según un determinado valor de z , se dispone de tablas estadísticas, como la que se muestra en la Figura N° 5.

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,500	0,504	0,508	0,512	0,516	0,520	0,524	0,528	0,532	0,536
0,10	0,540	0,544	0,548	0,552	0,556	0,560	0,564	0,567	0,571	0,575
0,20	0,579	0,583	0,587	0,591	0,595	0,599	0,603	0,606	0,610	0,614
0,30	0,618	0,622	0,626	0,629	0,633	0,637	0,641	0,644	0,648	0,652
0,40	0,655	0,659	0,663	0,666	0,670	0,674	0,677	0,681	0,684	0,688
0,50	0,691	0,695	0,698	0,702	0,705	0,709	0,712	0,716	0,719	0,722
0,60	0,726	0,729	0,732	0,736	0,739	0,742	0,745	0,749	0,752	0,755
0,70	0,758	0,761	0,764	0,767	0,770	0,773	0,776	0,779	0,782	0,785
0,80	0,788	0,791	0,794	0,797	0,800	0,802	0,805	0,808	0,811	0,813
0,90	0,816	0,819	0,821	0,824	0,826	0,829	0,831	0,834	0,836	0,839
1,00	0,841	0,844	0,846	0,848	0,851	0,853	0,855	0,858	0,860	0,862
1,10	0,864	0,867	0,869	0,871	0,873	0,875	0,877	0,879	0,881	0,883
1,20	0,885	0,887	0,889	0,891	0,893	0,894	0,896	0,898	0,900	0,901
1,30	0,903	0,905	0,907	0,908	0,910	0,911	0,913	0,915	0,916	0,918
1,40	0,919	0,921	0,922	0,924	0,925	0,926	0,928	0,929	0,931	0,932
1,50	0,933	0,934	0,936	0,937	0,938	0,939	0,941	0,942	0,943	0,944
1,60	0,945	0,946	0,947	0,948	0,949	0,951	0,952	0,953	0,954	0,954
1,70	0,955	0,956	0,957	0,958	0,959	0,960	0,961	0,962	0,962	0,963
1,80	0,964	0,965	0,966	0,966	0,967	0,968	0,969	0,969	0,970	0,971
1,90	0,971	0,972	0,973	0,973	0,974	0,974	0,975	0,976	0,976	0,977
2,00	0,977	0,978	0,978	0,979	0,979	0,980	0,980	0,981	0,981	0,982
2,10	0,982	0,983	0,983	0,983	0,984	0,984	0,985	0,985	0,985	0,986
2,20	0,986	0,986	0,987	0,987	0,987	0,988	0,988	0,988	0,989	0,989
2,30	0,989	0,990	0,990	0,990	0,990	0,991	0,991	0,991	0,991	0,992
2,40	0,992	0,992	0,992	0,992	0,993	0,993	0,993	0,993	0,993	0,994
2,50	0,994	0,994	0,994	0,994	0,994	0,995	0,995	0,995	0,995	0,995
2,60	0,995	0,995	0,996	0,996	0,996	0,996	0,996	0,996	0,996	0,996
2,70	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,997
2,80	0,997	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,998
2,90	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,999	0,999	0,999
3,00	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999

Figura N° 5. Valores de z . Distribución normal. Probabilidad contenida de $-\infty$ a z .

Error estándar

Volviendo al problema de definir la precisión de un estimado, se debe recordar que es poco probable que cada estimado de una muestra aleatoria corresponda al verdadero parámetro de dicha población. Es evidente que una muestra diferente de la misma población brindará un estimado distinto. Por lo tanto, el error estándar corresponderá a una medida de la precisión del estimado de la muestra.

Para ilustrar el concepto del error estándar, se puede emplear el caso de un estudio clínico que se va a llevar a cabo ininidad de veces y no una única vez, por lo que se tomarán múltiples muestras de la misma población. En este ejemplo, la distribución de los valores del estimado que se calcula en cada muestra seguirá una distribución normal (denominada en este caso distribución muestral). Se puede demostrar que esta distribución tiene una media que corresponde al verdadero parámetro de la población, y una nueva “desviación estándar”, que en realidad corresponde al error estándar, el cual mide cuán preciso es el estimado de la muestra con respecto al parámetro de la población.

Existen varias fórmulas para calcular el error estándar según el tipo de variable. Por ejemplo, hay una fórmula para calcular el error estándar de una proporción (Fórmula 18) y una fórmula para calcular el error estándar de una media (Fórmula 19).

- Fórmula 18:

$$\text{Error estándar de una proporción } (p) = \sqrt{\frac{p(1-p)}{n}}$$

- Fórmula 19:

$$\text{Error estándar de una media} = \frac{Sd}{\sqrt{n}}$$

Ejemplos del cálculo del error estándar

- *Ejemplo 1:*

Se desea estimar la prevalencia de diabetes mellitus en los pacientes que asisten a un hospital universitario. Se toma una muestra de 210 pacientes y se determina que el 7 % de la muestra posee dicha enfermedad. Aplicando la Fórmula 18, se obtiene un error estándar de 0,018.

$$\begin{aligned} \text{Error estándar (Se) de una proporción} &= \sqrt{\frac{0,07 (1 - 0,07)}{210}} \\ &= 0,018 \end{aligned}$$

- *Ejemplo 2:*

Se calcula la edad de una muestra de 100 personas. La media de edad fue de 50 años, con una desviación estándar de 15. Aplicando la Fórmula 19, se obtiene que el error estándar es de 1,5.

$$\text{Error estándar (Se) de una media} = \frac{15}{\sqrt{100}} = 1,5$$

¿Cómo interpretar el error estándar?

Para interpretar el error estándar se hace uso análogo de la desviación estándar. Aplicando las propiedades de la distribución normal a la distribución muestral, se puede concluir nuevamente que entre -1,96 y +1,96 veces el error estándar del valor medio del estimado calculado, se obtiene el 95 % de los valores posibles del parámetro de la población. A este intervalo de valores se le conoce con el nombre de intervalo de confianza del 95 %. En otras palabras, el intervalo de confianza del 95 % corresponde a un rango de valores donde hay una probabilidad del 95 % de encontrar el verdadero parámetro de la población.

En general, para el cálculo del intervalo de confianza del 95 % se utiliza la Fórmula 20, donde X corresponde al estimado calculado en la muestra. Conviene recordar que si se desea obtener un intervalo de confianza del 90 % o del 99 %, únicamente deberá cambiarse el valor de 1,96 por el correspondiente valor de z dentro del cual se incluya el área deseada ($z = \pm 1,64$ para un área del 90 % y $z = \pm 2,58$ para un área del 99 %).

- Fórmula 20:

$$\text{Intervalo de confianza del 95 \%} = X \pm 1,96 Se$$

- Ejemplo:

¿Cuál sería el intervalo de confianza del 95 % de la media de edad de la población del ejemplo 2?

$$\text{Intervalo de confianza del 95 \%} = X \pm 1,96 Se$$

$$\text{Intervalo de confianza del 95 \%} = 50 \pm 1,96 * (1,5)$$

$$= 50 \pm 2,94 = [47,06 - 52,94]$$

Esto significa que entre los 47,06 y los 52,94 años de edad se encuentra, con un 95 % de probabilidad, el valor real de la media de la población. Nótese que a pesar de esta alta probabilidad, existe un riesgo de que el verdadero parámetro no esté incluido en el rango.

El concepto del intervalo de confianza puede aplicarse también a las medidas de asociación, específicamente al riesgo absoluto, *hazard ratio*, riesgo relativo y *odds ratio*. En tales casos, cuando el intervalo de confianza de un riesgo relativo, *hazard ratio* u *odds ratio* incluya el valor de uno, o en el caso de un riesgo absoluto el valor de cero, se infiere que el verdadero parámetro de la población no difiere de manera significativa entre los grupos comparados.

- *Ejemplo:*

El estudio ISIS-2 comparó la mortalidad de sujetos con infarto agudo al miocardio que recibieron estreptoquinasa ($n= 8.592$) o placebo ($n= 8.595$) (1). A los 35 días, la incidencia acumulada de muertes fue de 12,1 % ($n= 1.029$) para placebo y 9,3 % ($n= 791$) para estreptoquinasa. Se reportó un riesgo absoluto del 2,8 % y un riesgo relativo de 0,77. Para calcular los intervalos de confianza de estas dos medidas de asociación, se deben efectuar los siguientes procedimientos:

1. *Cálculo del intervalo de confianza del 95 % para el riesgo absoluto:*

Se debe recordar que hay que emplear la Fórmula 20 para obtener el intervalo de confianza del 95 %, pero para ello se requiere primero calcular el error estándar del riesgo absoluto, lo cual se logra con la siguiente fórmula:

- *Fórmula 21:*

Error estándar (Se) del riesgo absoluto($p_1 - p_2$)

$$= \sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}}$$

Si se sustituyen los valores con los datos del ensayo clínico, el error estándar del riesgo absoluto sería:

Error estándar del riesgo absoluto($0,121 - 0,093$)

$$= \sqrt{\frac{0,121(1 - 0,121)}{8.592} + \frac{0,093(1 - 0,093)}{8.595}}$$

$$= \sqrt{\frac{0,121(0,879)}{8.592} + \frac{0,093(0,907)}{8.595}} = 0,0047$$

Ya con estos datos, se procede a calcular el intervalo de confianza del 95 %:

- Fórmula 20:

$$\text{Intervalo de confianza del 95 \%} = X \pm 1,96 Se$$

$$\text{Intervalo de confianza del 95 \%} = 0,028 \pm 1,96 (0,0047)$$

$$= 0,028 \pm 0,0092$$

$$\text{Intervalo de confianza del 95 \%} = [0,0188 - 0,0372]$$

De acuerdo con los resultados obtenidos, se puede afirmar, entonces, que la reducción del riesgo absoluto de muerte fue de un 2,8 %, con un intervalo de confianza del 95 % de 1,88 % a 3,72 %. En otras palabras, si se repitiera el estudio 100 veces, 95 de esas veces el riesgo absoluto calculado estará entre estos dos valores.

Dado que el intervalo no incluye al cero, se concluye que los riesgos de muerte son significativamente distintos entre ambos grupos (estreptoquinasa vs placebo).

2. *Cálculo del intervalo de confianza del 95 % para el riesgo relativo:*

Como se indicó, el riesgo relativo del estudio ISIS-2 fue de 0,77. Para determinar el intervalo de confianza del 95 %, se debe calcular primero el error estándar del riesgo relativo, utilizando la Fórmula 22.

- Fórmula 22:

$$\text{Error estándar (Se) del riesgo relativo} \left(\frac{p_1}{p_2} \right)$$

$$= \sqrt{\frac{p_1(1-p_1)}{p_1 * n_1} + \frac{p_2(1-p_2)}{p_2 * n_2}}$$

Si se sustituyen los valores con los datos del ensayo clínico, el error estándar del riesgo relativo sería:

$$\begin{aligned} & \text{Error estándar del riesgo relativo} \left(\frac{0,093}{0,121} \right) \\ &= \sqrt{\frac{0,093(1-0,093)}{0,093 * 8.595} + \frac{0,121(1-0,121)}{0,121 * 8.592}} \\ &= \sqrt{\frac{0,093(0,907)}{799,3} + \frac{0,121(0,879)}{1039,6}} \\ &= 0,0144 \end{aligned}$$

Una vez calculado el error estándar, se procede a determinar el intervalo de confianza del 95 %, utilizando en este caso la Fórmula 23, donde X puede ser un riesgo relativo, *odds ratio* o *hazard ratio*.

- Fórmula 23:

$$\text{Intervalo de confianza del 95 \%} = X * e^{\pm 1,96 (Se)}$$

$$\text{Intervalo de confianza del 95 \%} = 0,77 * e^{\pm 1,96 (0,0144)}$$

$$= 0,77 * e^{\pm 0,028}$$

$$\text{Intervalo de confianza del 95 \%} = [0,75 - 0,79]$$

Es decir, el riesgo relativo de la población tratada está con un 95 % de certeza entre 0,75 y 0,79. En este caso, como el

intervalo no incluye el valor uno (que implicaría la no diferencia entre grupos), se concluye que sí hay diferencias significativas en los riesgos de mortalidad de los sujetos que recibieron estreptoquinasa y los de placebo.

A continuación se presentan otras fórmulas útiles para el cálculo de los errores estándar del *odds ratio* y del *hazard ratio*.

- Fórmula 24:

$$\text{Error estándar del odds ratio} \left(\frac{Odds_1}{Odds_2} \right) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

- Fórmula 25:

$$\begin{aligned} \text{Error estándar del hazard ratio} & \left(\frac{Hazard_1}{Hazard_2} \right) \\ &= \sqrt{\frac{1}{\# \text{ eventos}_1} + \frac{1}{\# \text{ eventos}_2}} \end{aligned}$$

Como el lector puede suponer, cada estimado que se calcule en una muestra posee una fórmula correspondiente para el cálculo de su error estándar. Tal medida cobra valor cuando se calcula el intervalo de confianza como un elemento útil para describir la precisión de cada estimado.

EL VALOR DE p

El objetivo de los estudios clínicos es determinar si las observaciones encontradas en la muestra corresponden a verdaderos hallazgos en la población, o bien, si corresponden a efectos debidos al azar. El primer paso para contestar esta interrogante consiste en la formulación de las hipótesis estadísticas sobre el comportamiento hipotético de los parámetros de la población. Posteriormente, en un segundo

paso, se procede a analizar si los estimados de la muestra se comportan como la hipótesis planteada formula.

Formulación de hipótesis estadísticas

Así pues, el primer paso corresponde a la definición de dos hipótesis estadísticas: la hipótesis nula y la hipótesis alternativa.

Hipótesis nula (H_0)

Esta hipótesis formula que el valor de un parámetro es fijo, o bien, que no hay diferencias entre los parámetros de las poblaciones comparadas en el estudio.

Hipótesis alternativa (H_a)

Esta hipótesis propone que los parámetros son distintos a los formulados por la hipótesis nula, o bien, que existen diferencias entre los parámetros de las poblaciones comparadas en el estudio.

- *Ejemplos:*
 1. Se quiere determinar la prevalencia de una enfermedad de transmisión sexual en un grupo de estudiantes de último año de medicina. La hipótesis nula para este caso sería: “la prevalencia de enfermos en la población de estudiantes de último año de medicina es del 20 %”. Nótese que acá se escogió un valor del parámetro “fijo”, seleccionado de forma intencional por los investigadores. La hipótesis alternativa sería “la prevalencia de enfermos en la población de estudiantes de último de año de medicina es distinta al 20 %”.
 2. Se desea comparar las tasas de muerte por cáncer de pulmón en tabaquistas y no tabaquistas. Para tal efecto se escoge una muestra aleatoria de sujetos expuestos o no al tabaco. La hipótesis nula de este estudio sería: “no hay

diferencias entre las tasas de muerte por cáncer de pulmón en la población de tabaquistas *versus* no tabaquistas”. La hipótesis alternativa sería “hay diferencias entre las tasas de mortalidad por cáncer de pulmón en la población de tabaquistas *versus* no tabaquistas”.

Nótese, en ambos ejemplos, que para las hipótesis alternativas existen dos opciones. En el primer caso, si la proporción de enfermos no es del 20 % en la población de estudiantes de último año de medicina, puede ser que la prevalencia sea menor o mayor a este valor. Similarmente, si hay diferencias entre las proporciones del segundo ejemplo, estas pueden ser mayores o menores en un grupo con respecto al otro. Por lo tanto, se dice que estas hipótesis poseen “dos colas”. Son pocos los estudios en los que se emplean hipótesis de “una cola”; es decir, donde se desea saber si un medicamento o intervención posee un efecto igual o menor a cierto valor preestablecido (*ver apartado “Estudios de no inferioridad”*).

Definición del valor de p

Una vez formuladas ambas hipótesis estadísticas, se procede con el segundo paso, que consiste en definir cuál es la probabilidad de que los datos obtenidos en la muestra reflejen lo que la hipótesis nula plantea. Para ello, se procede a calcular el **valor de p** , por medio de las denominadas pruebas estadísticas.

Para explicar el valor de p se puede utilizar el primer ejemplo señalado anteriormente. En dicho estudio se plantearon las siguientes hipótesis:

H_0 = La proporción de la enfermedad en la población es de un
20 % ($\pi=20\%$)

$H_a = \pi \neq 20\%$

Suponiendo que en dicho estudio se escogió una muestra de 50 personas y que después de la recolección de datos se determinó que la prevalencia de enfermos fue del 42 % ($n= 21$), se procede a determinar cuál es la probabilidad de que teniendo una proporción de la enfermedad del 20 % en la población y seleccionando una muestra de 50 sujetos, 21 de ellos estén enfermos (42%). Para el cálculo de esta probabilidad, es necesario conocer cuál sería la distribución de probabilidades bajo el supuesto de que la hipótesis nula sea cierta (es decir, que la prevalencia sea del 20 % en una muestra de 50 personas). En este caso específico, dicha distribución de probabilidades viene dada por la llamada “distribución binomial”, que se grafica en la Figura N° 6.

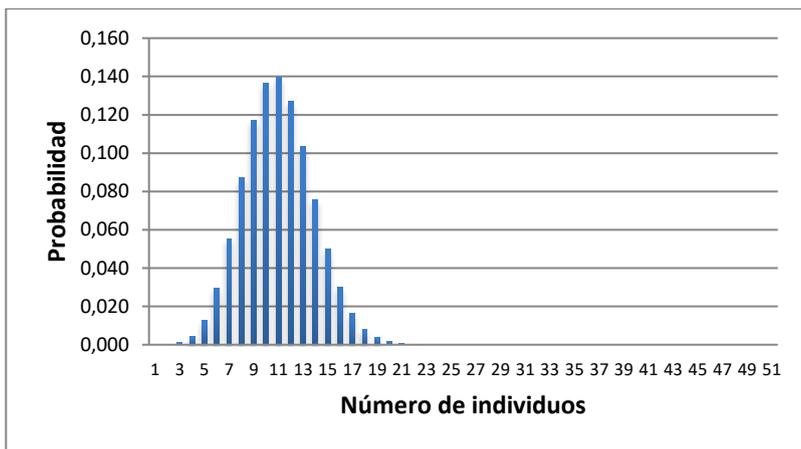


Figura N° 6. Distribución de probabilidad binomial para una proporción del 20 % en 50 individuos.

Para entender el gráfico anterior, se pueden imaginar los casos más extremos que pueden ocurrir en ese estudio; por ejemplo, al escoger la muestra de 50 individuos, en la cual la prevalencia de enfermos es del 20 %, podría ocurrir el caso hipotético de que los 50 sujetos seleccionados tengan la

enfermedad de interés. En el caso extremo, también podría ocurrir que ningún individuo seleccionado presente tal enfermedad. No obstante, estos dos escenarios son muy poco probables en el contexto de la hipótesis nula cierta.

Nótese que la probabilidad de escoger 10 sujetos enfermos en una muestra de 50 personas es el evento más probable cuando la hipótesis nula de este ejemplo es cierta.

Observando ese mismo gráfico, ¿podría usted calcular cuál sería la probabilidad de detectar 42 pacientes enfermos? Como puede suponer, dicha probabilidad es muy baja (ni siquiera aparece graficada); por lo que se puede concluir que es muy poco probable que la hipótesis nula sea cierta.

Esa probabilidad de obtener un resultado al menos tan extremo como el que realmente se ha obtenido, suponiendo que la hipótesis nula sea cierta, es lo que se conoce como el valor de p . Cuando el valor de p se encuentra por debajo de cierto valor, también llamado nivel de significancia, que suele ser de 0,05, se rechaza la hipótesis nula. Si por el contrario, el valor de p se encuentra por encima del nivel de significancia, no se puede rechazar la hipótesis nula. Al respecto, es importante dejar claro que la hipótesis nula nunca puede ser aceptada, ya que el verdadero parámetro de la población es desconocido.

En general, el proceso de definición del valor de p es matemáticamente complejo y exige la utilización de las pruebas de hipótesis y del conocimiento de varias distribuciones de probabilidad, tales como la distribución de chi-cuadrado o de t -student. Sin embargo, todas las fórmulas de las pruebas de hipótesis pueden resumirse en una sola fórmula:

- Fórmula 26:

$$\text{Valor de } p = \frac{\text{Error estándar de las diferencias entre grupos}}{\text{Diferencia del efecto entre grupos}}$$

De la Fórmula 26 se puede inferir que entre más grande sea el tamaño de la muestra (y por ende, menor el error estándar), menor será el valor de p . Similarmente, si hay diferencias muy amplias entre los grupos a comparar, el valor de p será pequeño. En ambos casos, es probable que el valor de p sea menor que el valor de significancia usual (0,05) y, por ende, que se concluyan diferencias significativas entre grupos.

Variables

Debido a que las pruebas de hipótesis empleadas dependen ante todo de la naturaleza de las variables, conviene conocer, entonces, las dos grandes categorías de variables: variables categóricas y variables continuas.

Variables categóricas

Los posibles valores de estas variables son “categorías”, como el sexo (masculino/femenino; dos categorías) o la escala de Glasgow (que varía de tres a 15; 13 categorías). Pueden dividirse en dos: nominales (sexo, por ejemplo) y ordinales (como la escala de Glasgow). Las variables nominales poseen categorías arbitrarias. Por el contrario, las variables ordinales tienen categorías cuyo orden es previamente definido por los valores dados.

Un tipo especial de variables categóricas son las variables binomiales, en las que solo existen dos categorías posibles de clasificación; por ejemplo, la presencia o la ausencia de una enfermedad.

Variables continuas

En este caso los valores posibles de cada variable son números continuos. Pueden subdividirse a su vez en variables discretas (como el número de partos) o variables continuas métricas (como el peso en kilogramos al nacer). También se les denomina variables numéricas.

Pruebas pareadas y no pareadas

Otro punto decisivo en la elección de la prueba de hipótesis, es determinar si se desea comparar al mismo sujeto antes y después de una intervención, en cuyo caso deben emplearse las pruebas “pareadas” (utilizadas para hacer comparaciones sobre una misma unidad experimental); o bien, si se comparará un grupo *versus* otro grupo, en cuyo caso deben aplicarse las “pruebas no pareadas”.

Pruebas post hoc

En general, las pruebas estadísticas comparan un grupo *versus* otro. Sin embargo, en otros casos se trata de comparar más de dos conjuntos de datos. En este escenario, la hipótesis alternativa suele ser que al menos un grupo difiere de los demás, sin poder determinar fehacientemente cuál de todos los grupos tiene diferencia con otro. En estos casos, se acude a las pruebas estadísticas denominadas *post hoc*, en las cuales se analiza, por separado, la relación de cada par. Un ejemplo de estos análisis *post hoc* es la Prueba de Bonferroni (2).

Pruebas paramétricas o no paramétricas

Un último determinante para la selección de la prueba de hipótesis es establecer si la variable a comparar entre grupos posee una distribución de frecuencias normal (paramétrica) o no paramétrica. Esto permitirá seleccionar entre la aplicación de una prueba paramétrica, utilizada para analizar datos numéricos distribuidos de forma normal, o una prueba no paramétrica, que se utiliza para analizar datos con variables nominales y ordinales, que no poseen una distribución específica.

Las figuras N° 7 y N° 8 constituyen una guía para elegir de manera adecuada cuál prueba de hipótesis debe emplearse según los criterios previamente mencionados. Cabe destacar que en estas figuras no aparecen otras pruebas de hipótesis

empleadas con frecuencia en el desarrollo de estudios clínicos, como las prueba de Log-rank y la prueba de Wilcoxon, las cuales se utilizan para comparar tasas de incidencia entre dos o más grupos.



Figura N° 7. Pruebas de hipótesis para el análisis de variables numéricas (Adaptado de: du Prel JB, Röhrig B, Hommel G, Blettner M. Choosing statistical tests: part 12 of a series on evaluation of scientific publications. *Dtsch Arztebl Int.* 2010; 107(19): 343-348).

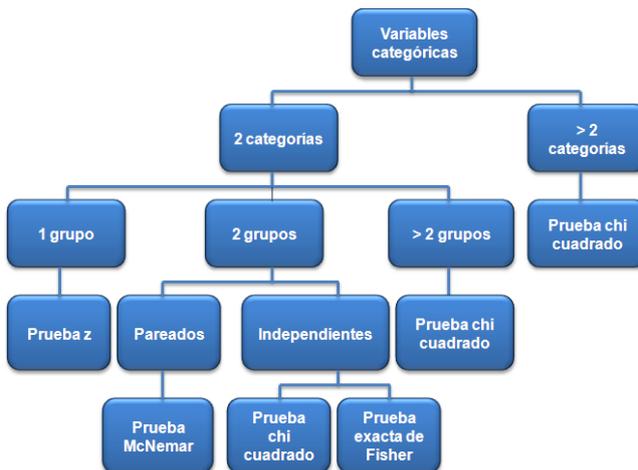


Figura N° 8. Pruebas de hipótesis para el análisis de variables categóricas (Adaptado de: du Prel JB, Röhrig B, Hommel G, Blettner M. Choosing statistical tests: part 12 of a series on evaluation of scientific publications. *Dtsch Arztebl Int.* 2010; 107(19): 343-348).

EL OBJETIVO PRIMARIO: ¿SIMPLE O COMPUESTO?

El evento de interés es la variable sobre la cual se establece la hipótesis del estudio clínico y para la cual se calcula el tamaño de la muestra. Los eventos de interés están en relación directa con el objetivo principal del estudio, en tanto los objetivos secundarios surgen como preguntas anexas al objetivo principal. Estos objetivos secundarios pretenden generar nuevas hipótesis, las cuales deben ser formalmente probadas en nuevos ensayos clínicos, que las incluyan como objetivos primarios. En general, la mayoría de estudios clínicos carecen de suficiente poder estadístico para establecer conclusiones válidas con respecto a los objetivos secundarios; de ahí que las conclusiones sobre estos objetivos sean denominadas “conclusiones generadoras de hipótesis”.

Idealmente, el objetivo principal debe medirse con un único evento de interés, que permita responder de manera fácil la pregunta de investigación (objetivo primario simple). Sin embargo, en varias ocasiones el objetivo de interés es compuesto; es decir, está conformado por varios eventos que, en conjunto, constituyen el “gran evento de interés”, también denominado “objetivo primario compuesto” o “*composite primary end point*”. Por ejemplo, en el área cardiovascular con frecuencia se emplea el término “eventos cardiovasculares mayores” para referirse al conjunto de muertes de causa cardiovascular, infartos fatales y no fatales, eventos cerebrovasculares y episodios de angina, los cuales constituyen el “objetivo primario compuesto”, donde la ocurrencia de cualquiera de estos desenlaces será considerado como un evento de interés. El principal problema es que la ocurrencia de uno de los eventos no implica la no ocurrencia de los otros, ya que no son mutuamente excluyentes. Estos desenlaces se conocen como eventos “competitivos”. En otras palabras, una persona puede tener angina y ser contabilizada de la misma forma que otro individuo que murió de causa cardiovascular. Es evidente que no es lo mismo tener angina a morir de un infarto, pero para

efectos del estudio se contabilizan ambos desenlaces dentro del mismo objetivo.

El análisis de eventos de interés compuestos y competitivos es dificultoso, ya que no se pueden hacer conclusiones válidas para cada componente del evento, sino solo para el total de desenlaces. Todavía es más problemático cuando el evento compuesto no difiere significativamente entre grupos, pero sí lo hace un componente específico de dicho combinado; en este caso, debido a que el estudio clínico se diseña con un poder estadístico para detectar diferencias en el total de eventos compuestos, no es posible llegar a conclusiones válidas estadísticamente para cada componente en particular.

A pesar de lo anterior, un estudio con objetivo compuesto tiene la gran ventaja de emplear una menor cantidad de sujetos, ya que el tamaño de la muestra depende de la cantidad de eventos esperados; por lo tanto, entre más elementos se incorporen al evento primario compuesto, menos sujetos se necesitarán en la muestra. Esta es la razón por la cual el evento primario compuesto suele estar conformado por una mezcla de desenlaces frecuentes (por ejemplo: angina inestable) e infrecuentes (por ejemplo: muerte de causa cardiovascular), los cuales, en la mayoría de los casos, comparten un fondo fisiopatológico común.

De forma análoga, el uso de eventos compuestos permite probar el efecto de la intervención en varias medidas de eficacia, como por ejemplo, la muerte de causa cardiovascular, el infarto no fatal, la ocurrencia de eventos cerebrovasculares y los episodios de angina.

Si para el médico y el paciente los componentes del evento primario compuesto son igual de importantes, la conclusión del estudio puede ser clínicamente relevante. Por el contrario, el uso de eventos compuestos de distinta relevancia práctica puede generar conclusiones de escasa notabilidad.

Análisis con eventos primarios compuestos
<p>Ventajas:</p> <ul style="list-style-type: none">✓ Reduce el tamaño de la muestra.✓ Estima el beneficio neto de una terapia.✓ Evita hacer análisis secundarios.
<p>Desventajas:</p> <ul style="list-style-type: none">✓ Interpretación práctica puede ser problemática, ya que los componentes pueden ser diferentes en importancia.✓ El efecto de la intervención puede ser diferente entre componentes✓ Sesgos de análisis por “eventos competitivos”.

Figura N° 9. Ventajas y desventajas de los análisis con eventos primarios compuestos.

Características de los componentes del evento primario compuesto

En general, los componentes del evento primario compuesto deben presentar las siguientes características: coherencia, equivalencia y homogeneidad.

Coherencia

Idealmente, los componentes del evento compuesto deben compartir un trasfondo fisiopatológico común y ser consistentes entre sí. Por ejemplo, en pacientes diabéticos tendría coherencia contabilizar como eventos comunes las complicaciones crónicas de esta patología (retinopatía, enfermedad cardiovascular, nefropatía).

Equivalencia

En la mayoría de análisis de eventos compuestos, solo uno de ellos es contabilizado para el análisis de los datos, y suele ser el primero que ocurre durante el seguimiento; de ahí la importancia de que todos los componentes sean equivalentes. Suponiendo que un evento de interés compuesto es la ocurrencia de hospitalizaciones por angina o la incidencia acumulada de mortalidad cardiovascular, el ideal sería que tanto la hospitalización como la mortalidad sean equivalentes, es decir, que tengan el mismo peso clínico. Sin embargo, para muchos clínicos (y pacientes) una hospitalización no tiene la misma equivalencia que la muerte. Otros investigadores pueden argumentar que una hospitalización por angina tiene alta mortalidad y por lo tanto, ser equivalente con el otro evento.

Homogeneidad

Como ya se mencionó, los componentes del evento de interés deben seguir la misma tendencia al momento del análisis. En otras palabras, las medidas de asociación no deberían tener una variabilidad significativa entre cada desenlace.

Aunque estas situaciones solo pueden evaluarse hasta obtener los resultados del estudio clínico, será más sencillo explicar el beneficio de todos los desenlaces particulares, que explicar el beneficio de un evento particular y la inferioridad de otro.

- *Ejemplos de estudios con objetivo primario compuesto*

El estudio JÚPITER (*Rosuvastatin to prevent vascular events in men and women with elevated C-Reactive Protein*) (3), evaluó el uso de la rosuvastatina en sujetos sin historia de enfermedad cardiovascular, con una concentración de LDL colesterol menor de 130 mg/dL y una proteína C reactiva de alta sensibilidad mayor o igual a 2 mg/l. El evento de interés compuesto fue la ocurrencia del primer evento cardiovascular (definido como

infarto de miocardio no fatal, evento cerebrovascular no fatal, hospitalización por angina inestable, un procedimiento de revascularización arterial o muerte por causas cardiovasculares). Los principales resultados del estudio se resumen en el Cuadro N° 2.

Cuadro N° 2. Objetivo final compuesto en el Estudio Júpiter

Variable	Rosuvastatina (n=8.901)		Placebo (n=8.901)		Hazard ratio	Valor de p
	Número	Tasa (por 100 personas *año)	Número	Tasa (por 100 personas *año)		
Objetivo primario compuesto	142	0,77	251	1,36	0,56	< 0,001
Infarto de miocardio no fatal	22	0,12	62	0,33	0,35	< 0,001
Cualquier infarto de miocardio	31	0,17	68	0,37	0,46	< 0,001
Evento cerebrovascular no fatal	30	0,16	58	0,31	0,52	0,003
Cualquier evento cerebrovascular	33	0,18	64	0,34	0,52	0,002
Revascularización arterial	71	0,38	131	0,71	0,54	< 0,001
Hospitalización por angina inestable	16	0,09	27	0,14	0,59	0,09
Muerte por causa cardiovascular, infarto de miocardio y evento cerebrovascular	83	0,45	157	0,85	0,53	< 0,001

Tras analizar el Cuadro N° 2, y partiendo de la definición del evento final compuesto para este estudio, se observa que la suma final de cada desenlace no corresponde al valor reportado como “eventos del objetivo primario compuesto” (en negrita). Es decir, la suma de 31 (cualquier infarto de miocardio) + 33

(cualquier evento cerebrovascular) + 71 (procedimientos de revascularización arteriales) + 16 (hospitalizaciones por angina inestable) da como resultado 151 eventos, no los 142 reportados. Este ejemplo ilustra que muchas veces el manejo de los eventos compuestos no suele ser uniforme, lo cual favorece la interpretación errónea de los datos.

Por otra parte, el estudio AGI-1067 (*Effects of succinobucol after an acute coronary syndrome: a randomised, double-blind, placebo-controlled trial*) (4) reportó el siguiente cuadro de eventos primarios, en el cual, la sumatoria de los componentes individuales (42+4+113+16+105+250) sí suma el total de los 530 eventos descritos por los autores.

Cuadro N° 3. Objetivo final compuesto en el Estudio AGI-1067				
Variable	Succinobucol (n)	Placebo (n)	Hazard ratio (intervalo de confianza 95 %)	Valor de p
Objetivo primario compuesto	530	529	1,00 (0,89-1,13)	0,955
Muerte cardiovascular	42	48		
Paro cardíaco con resucitación	4	5		
Infarto de miocardio (no fatal)	113	130		
Evento cerebrovascular (no fatal)	16	28		
Hospitalización por angina inestable	105	90		
Revascularización coronaria	250	228		

Otro ejemplo de un análisis que empleó un evento primario compuesto fue el estudio CAPRIE (5). En este ensayo clínico controlado y aleatorizado, se asignaron 19.185 pacientes a recibir aspirina (325 mg/d) o clopidogrel (75 mg/d). El objetivo primario compuesto fue la ocurrencia de evento cerebrovascular isquémico, infarto de miocardio o muerte vascular. Estos tres eventos independientes comparten los tres criterios antes citados: coherencia, equivalencia y homogeneidad. Son coherentes porque los componentes del evento compuesto corresponden a una misma fisiopatología; son equivalentes en cuanto a la importancia clínica tanto para el paciente como para la mayoría de los clínicos; y, finalmente, son homogéneos porque la dirección del efecto obtenido es similar para cada componente del evento primario compuesto.

CLASIFICACIÓN DE LOS OBJETIVOS PRIMARIOS POR RANGO DE JERARQUÍA

Los objetivos primarios también pueden clasificarse en un rango de prioridad o jerarquía. Por ejemplo, el principal objetivo en la mayoría de las enfermedades sería evitar la muerte del paciente; lo cual constituiría el llamado “punto fuerte” (del inglés: “*hardpoint*”). Otros desenlaces, como hospitalizaciones, infartos o procedimientos de revascularización, no suelen ser tan relevantes como la muerte, por lo que tendrían un rango de prioridad menor. Otros objetivos finales tienen todavía menos relevancia clínica, como algunos valores de laboratorio (colesterol LDL, niveles de proteína C reactiva, negativización de cargas virales) o biomarcadores de una determinada enfermedad; en ocasiones, estos marcadores son conocidos como objetivos subrogados o indirectos, y reemplazan los eventos clínicos. Los biomarcadores ideales deben medir objetivamente los procesos biológicos o patogénicos de la enfermedad, o bien, predecir las respuestas farmacológicas a una intervención farmacéutica. Si bien los biomarcadores suelen ser más fáciles y baratos de medir que los desenlaces clínicos, se debe tener cautela al momento de su interpretación,

ya que no necesariamente las modificaciones de los niveles de un biomarcador implicarán la reducción de eventos clínicos.

Cuando los estudios clínicos prueban la correlación de una variable con los desenlaces o eventos que una intervención provee, se habla de que tales variables son **predictivas** de la eficacia de un tratamiento.

ANÁLISIS MÚLTIPLES

El diseño de cualquier estudio clínico implica no solo la recolección de la variable primaria (ya sea única o múltiple), sino también el análisis de otra serie de variables de eventual relevancia clínica. Por ejemplo, en un estudio clínico de una droga para una enfermedad cardiovascular, no solo es necesario conocer si el fármaco o intervención reduce el número de infartos, sino también si la reducción de tales eventos varía con la edad, el sexo, la presencia o la ausencia de comorbilidades, entre otras características. Por esa razón, en la mayoría de estudios clínicos se reportan varios análisis adicionales al establecido en el objetivo primario. No obstante, conviene recordar que el poder y el diseño del estudio deberán estar dirigidos a contestar la pregunta de investigación que establece el objetivo primario.

Análisis de subgrupos

Los análisis de subgrupos son un ejemplo de análisis múltiples, en los cuales se busca identificar alguna variable particular en la que la eficacia de la intervención sea comparada en diversas categorías y con ello generar hipótesis que deberán probarse formalmente en posteriores estudios.

Los análisis de subgrupos **no** deben considerarse en ningún momento definitivos ni contundentes para tomar decisiones clínicas. Por error, en muchos casos se recurre al análisis de subgrupos para identificar alguna categoría particular en la cual la intervención sea eficaz, aún cuando en el total de la muestra

estudiada no se demostró ningún beneficio. Del mismo modo, suelen realizarse análisis de subgrupos para identificar sujetos con determinadas características que sugieran la ausencia de beneficio de una intervención, aún cuando la muestra total de individuos analizados sí obtuvo un beneficio con el tratamiento. En ambos casos es imposible llegar a conclusiones válidas científicamente, más si se considera que en este tipo de análisis hay un alto riesgo de falsos negativos y falsos positivos. El riesgo de falsos negativos surge a partir de la poca representación que tiene un subgrupo particular del total de la muestra estudiada; de forma tal que entre menor es la cantidad de sujetos que componen el subgrupo, mucho mayor es el riesgo de falsos negativos. El riesgo de falsos positivos, por su parte, es elevado, porque con la repetición de cada análisis estadístico es muy probable que solo por azar alguno de los subgrupos sea favorecido sobre otro; además, subgrupos con pocos sujetos tendrán estimados de eficacia del tratamiento muy imprecisos, en los cuales el efecto de dicha terapia puede sobredimensionarse.

Otro argumento en contra de estos análisis es que en muchos de los casos los subgrupos comprenden categorías que no son mutuamente excluyentes entre sí, por lo que un individuo puede pertenecer a varias categorías, con distintos efectos en la intervención.

Si a pesar de los argumentos anteriores se desea recurrir a un análisis de subgrupos, debe tomarse en cuenta que el correcto estudio estadístico debe ir encaminado a determinar si el efecto del tratamiento difiere entre cada categoría, y no a determinar si el efecto del tratamiento es positivo para tal subgrupo. Para ello, debe recurrirse a una prueba estadística de interacción o de heterogeneidad, en la cual se establecen las siguientes hipótesis:

*Hipótesis nula = **NO hay diferencia** en el efecto de la intervención **ENTRE** categorías*

Hipótesis alternativa = *Sí hay diferencias de la intervención en al menos una de las categorías*

En el Cuadro N° 4 se describen otras recomendaciones para la adecuada interpretación de los estudios de subgrupos.

Cuadro N° 4. Recomendaciones para la adecuada interpretación de los estudios de subgrupos
<ol style="list-style-type: none">1. Los subgrupos deben definirse <i>a priori</i>, es decir, antes de realizar el estudio.2. Los análisis de subgrupos deben reportar el valor de <i>p</i> de la prueba de interacción que evidencia heterogeneidad del efecto entre subgrupos.3. Las conclusiones de un estudio no deben nunca hacer énfasis en subgrupos.4. Las evidencias aportadas de los estudios de subgrupos solo pueden servir para generar nuevas hipótesis.

Es relativamente usual que los investigadores realicen un sinnúmero de análisis *post hoc* y que solo sean reportados los resultados que fueron positivos. Esta técnica, también conocida como *fishing data*, consiste en la evaluación de muchas variables que puedan asociarse unas con otras después de desarrollado un estudio clínico. Evidentemente, el efecto del azar puede explicar que alguna de estas asociaciones sea estadísticamente significativa, por lo que el uso de esta práctica debe ser desestimado. Lo que se aconseja es la definición *a priori* de los subgrupos y de la estratificación en el proceso de muestreo, así como de un plan estadístico preplaneado antes del análisis de los datos.

Análisis múltiples “*end points*”

En clara relación con el análisis de subgrupos, se encuentra el estudio de análisis múltiples “*end points*”. En este caso también

hay un alto riesgo de encontrar asociaciones secundarias al azar como consecuencia de la multiplicidad de pruebas. Matemáticamente, la probabilidad de encontrar un hallazgo falso positivo puede calcularse con la Fórmula 27, en donde alfa (α) corresponde al error tipo I del estudio original, y n se refiere al número de análisis múltiples o subgrupos.

- Fórmula 27:

$$\text{Probabilidad de falsos positivos} = 1 - (1 - \alpha)^n$$

Una representación gráfica de la Fórmula 22 viene dada por la Figura N° 10. En el eje y se representa la probabilidad de al menos un falso positivo en el análisis múltiple, mientras que en el eje x se anota el número de análisis cuando el valor de alfa es 0,05. Nótese cómo con tan solo cinco análisis múltiples, la probabilidad de un falso positivo es del 22,6 %

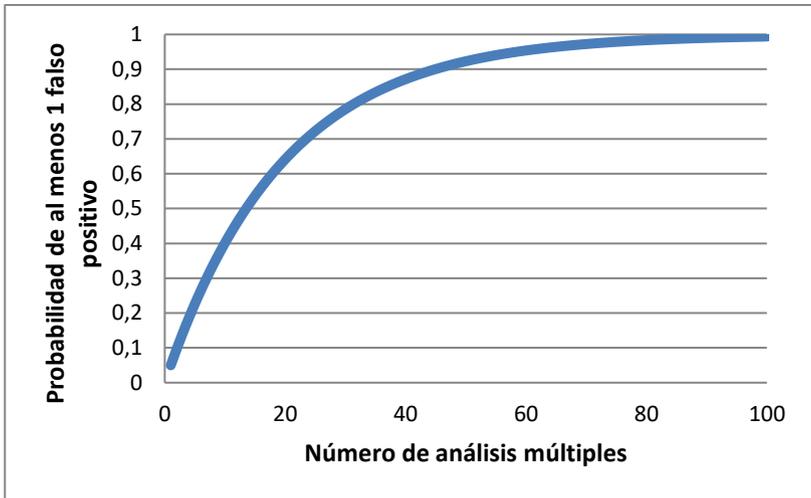


Figura N° 10. Probabilidad de falsos positivos de acuerdo al número de análisis múltiples.

Ejemplo de análisis múltiples: análisis de subgrupo

Un ejemplo claro de análisis de subgrupo es el estudio PARADIGM-HF (*Angiotensin-neprilysin inhibition versus enalapril in heart failure*) (6), un ensayo clínico controlado aleatorizado, en el que se asignó a 8.442 pacientes con insuficiencia cardiaca a recibir enalapril o LCZ696, para determinar, como objetivo primario compuesto, el efecto sobre la mortalidad de causa cardiovascular y las hospitalizaciones por insuficiencia cardiaca. Los autores concluyeron afirmativamente que el fármaco experimental LCZ696 fue superior al enalapril en todos los subgrupos, exceptuando la categoría de individuos con un estadio funcional I y II de la escala *New York Heart Association* (NYHA), donde la prueba de interacción para el objetivo primario fue estadísticamente significativa ($p = 0,003$), lo que sugiere un mayor beneficio del fármaco experimental en pacientes menos sintomáticos, pero no se reportó tal interacción para la muerte cardiovascular ($p = 0,76$) (ver Figura N° 11). Sin embargo, debido a que se analizaron 18 subgrupos, la probabilidad de falsos positivos es cercana al 60 % al aplicar la Fórmula 22. Además, algún lector podría argumentar que los individuos con un estadio funcional NYHA III y IV no se benefician del uso del fármaco LCZ696; pero este subgrupo solo incluye al 24,7 % del total de la muestra estudiada, por lo que el riesgo de un falso negativo es igualmente alta. Los lectores no familiarizados con la prueba de interacción podrían concluir de manera errónea que otras categorías de pacientes tampoco derivan beneficio del fármaco experimental, por ejemplo: adultos mayores de 75 años, sujetos de etnia negra, asiáticos o nativos norteamericanos, pacientes con fracción de eyección superior al 35 % y aquellos sin uso previo de un inhibidor de la enzima convertidora de angiotensina (IECA), ya que en todos estos casos el intervalo de confianza del 95 % del *hazard ratio* incluye al uno. No obstante, los valores de p de la prueba de interacción no son significativos para estos subgrupos.

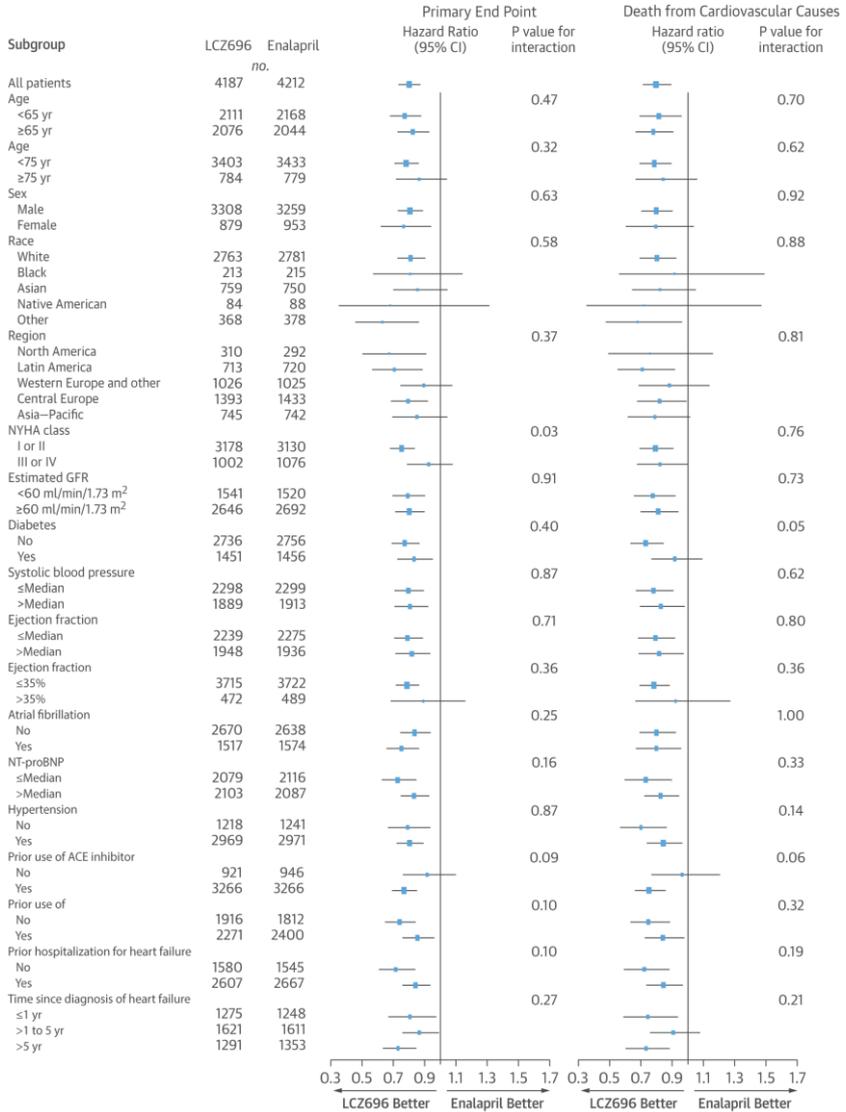


Figura N° 11. Resultados del análisis de subgrupos del estudio PARADIGM-HF (reproducido con permiso de: Pocock SJ, McMurray JJV, Collier TJ. Statistical controversies in reporting of clinical trials: Part 2 of a 4-Part series on statistics for clinical trials. *J Am Coll Cardiol.* 2015; 66(23): 2648-2662).

Análisis interino de datos

Otro ejemplo de análisis múltiples surge cuando un estudio clínico es sometido a monitoreo interino de datos. En la práctica, casi todos los ensayos clínicos están siendo monitorizados por un comité, usualmente independiente del investigador principal y del patrocinador, el cual se reúne periódicamente, con el fin de determinar la continuidad del estudio, ya sea porque el fármaco o la intervención a probar es muy eficaz, o bien, porque el nuevo tratamiento está generando peores resultados que el comparador, sea este placebo o estándar. En este último caso, conocido como futilidad, es evidente que las razones éticas suponen la finalización del ensayo. Quizás menos lógico resulte el terminar un estudio clínico por razones positivas a favor del nuevo fármaco; sin embargo, también las razones éticas obligan a detener un estudio para que los resultados sean rápidamente publicados y más pacientes se vean beneficiados de la nueva intervención. El problema estadístico surge cuando se realiza cada análisis interino y existe posibilidad de que por azar alguno de estos análisis sea positivos. De hecho, esta posibilidad es más alta al principio del desarrollo del ensayo clínico, ya que al contarse con menos datos la precisión de los estimados calculados es baja, y una aparente asociación positiva a favor de una intervención puede no llegar a ser real si el estudio se hubiera desarrollado en la totalidad del tiempo previamente planeado.

Para ilustrar el concepto anterior, se puede tomar como supuesto que se quiere calcular la media de edad de una población de estudiantes de medicina, para lo cual se obtiene una muestra representativa y aleatoria de 50 estudiantes. Es esperable que con los primeros tres sujetos la media de edad diste mucho de la verdadera media obtenida al terminar de recoger los datos de las 50 personas. En este sentido, se debe tener presente que con cada observación se estima mejor la verdadera media de edad. Este fenómeno, denominado regresión a la media, sugiere que cualquier análisis interino de datos debe interpretarse cautelosamente.

Esa es la razón por la cual han surgido diversos métodos o técnicas que corrigen el riesgo de cometer un error tipo I, o falso positivo, con cada análisis.

Técnicas estadísticas para la detención precoz de un estudio

Estas técnicas estadísticas creadas para corregir el riesgo de error tipo I son conocidas en conjunto como funciones de mantenimiento del error alfa, y las más empleadas son el método de Pocock, el método de Peto, el método de O'Brien-Fleming y el método de Lan-DeMets.

El Cuadro N° 5 muestra, por cada uno de esos métodos, cuál debe ser el valor de significancia mínimo a partir del cual debe rechazarse o no la hipótesis nula en caso de que se quiera mantener un alfa de 0,05.

Cuadro N° 5. Valores de alfa para la detención de un estudio clínico, según el número de análisis interinos planeados y la técnica estadística empleada				
Número planeado de análisis interinos	Análisis interino	Pocock	Peto	O'Brien-Fleming
2	1	0,029	0,001	0,005
	2	0,029	0,05	0,048
3	1	0,022	0,001	0,0005
	2	0,022	0,001	0,014
	3	0,022	0,05	0,045
4	1	0,018	0,001	0,0001
	2	0,018	0,001	0,004
	3	0,018	0,001	0,019
	4	0,018	0,5	0,043
5	1	0,016	0,001	0,00001
	2	0,016	0,001	0,0013
	3	0,016	0,001	0,008
	4	0,016	0,001	0,023
	5	0,016	0,05	0,041

De acuerdo a lo indicado en el Cuadro N° 5, si un investigador desea realizar cinco análisis interinos de sus datos, deberá detener su estudio en el primero de estos análisis solo cuando el valor de p sea menor a 0,00001 según el método de O'Brien-Fleming, o bien, si el valor de p fuese menor a 0,016 o 0,001 por el método de Pocock o Peto, respectivamente.

Cabe destacar que el método de Pocock fue de los primeros descritos, pero cuenta con la desventaja de que el valor final de alfa se reduce drásticamente conforme aumenta el número de análisis interinos. Así pues, un estudio que al final de su desarrollo genera un valor de p de 0,04 es considerado estadísticamente NO significativo si antes se realizaron tres análisis interinos, ya que el valor de p requerido para rechazar la hipótesis nula debe ser menor a 0,022 en cada análisis.

En general, la mayoría de estudios utiliza el método de O'Brien-Fleming, con el cual se emplean criterios muy estrictos para la detención del estudio clínico en fases tempranas del análisis, a la vez que se preserva el poder del estudio.

En todo caso, el autor deberá definir *a priori* cuál de estos métodos estadísticos será empleado para detener el estudio; al mismo tiempo, deberá determinar cuándo realizará cada análisis interino, que puede ser cada cierto período de tiempo o, mejor aún, con cada cierto número o proporción de los eventos primarios.

Conviene recordar que, aunque el estudio clínico sea detenido prematuramente con las reglas estadísticas mencionadas, los estimados calculados no están exentos de sesgos. De hecho, como se explicó en el ejemplo del cálculo de la media de edad, es muy probable que al inicio del desarrollo de cada ensayo clínico se obtengan estimados que usualmente sobreestiman la magnitud real del efecto del tratamiento.

ESTUDIOS DE NO INFERIORIDAD

Con relativa frecuencia surgen en la búsqueda bibliográfica estudios que comparan una intervención con otra, pero no en términos de superioridad, sino en términos de “no inferioridad”; es decir, estudios que tratan de demostrar que un tratamiento no es inferior (o peor) que otro. Pero, ¿para qué quiere un autor demostrar que su intervención no es peor que otra? En primer lugar, se debe tener presente que en muchas ocasiones ya existe una terapia estándar bien establecida para cada enfermedad, por lo cual no es posible administrar placebo a un grupo de sujetos. En estos casos, intentar demostrar que una nueva molécula es no inferior al estándar, supone la introducción al mercado de un tratamiento novedoso que tendrá alguna ventaja particular. Por ejemplo, muchos de los nuevos anticoagulantes no se comparan *versus* placebo en estudios de superioridad, ya que está bien documentada la eficacia de la warfarina en estos contextos clínicos. Además, administrar placebo a un sujeto con alto riesgo de tromboembolismo no es éticamente posible. Sin embargo, la nueva molécula tendrá al menos una teórica ventaja sobre la warfarina. En este mismo ejemplo, algunos autores han documentado que los nuevos anticoagulantes no presentan tantas interacciones farmacológicas con otras drogas o alimentos como sí se producen con el uso de warfarina, o bien, que no es necesario implementar una estricta medida de la eficacia de la anticoagulación con estudios rutinarios de sangre.

Otra ventaja estadística de estos estudios, es que si bien están diseñados para demostrar la no inferioridad de una intervención, pueden en ocasiones demostrar que la nueva droga es, de hecho, superior al comparador.

Por otro lado, debe aclararse que nunca se debe concluir que una intervención es igual a otra a partir de estudios de no inferioridad. Dicho en otras palabras, la no inferioridad no es sinónimo de igualdad entre dos intervenciones. Ni siquiera los denominados estudios de bioequivalencia pueden demostrar

que una sustancia es exactamente igual que otra. Aun estos estudios plantean un margen en el cual puede asegurarse la bioequivalencia de dos moléculas. De hecho, ningún estudio clínico que trabaje con muestras de una población puede concluir la equivalencia de una terapia con otra, ya que siempre existirá incertidumbre sobre el verdadero valor de la población.

El margen de no inferioridad

Si un estudio clínico plantea en su hipótesis que la terapia innovadora no es inferior a un estándar, debe especificarse en primer lugar cuál es el margen (Δ) de no inferioridad. Por ejemplo, suponiendo que una persona se quiere comparar con un corredor olímpico en una carrera de 100 metros planos, en un estudio usual de superioridad se debería contemplar cuál competidor llega primero a la meta; en cambio, un estudio de no inferioridad plantearía en primer lugar el margen por el cual, aun llegando después del corredor olímpico, la persona puede considerarse como no inferior al ganador. Si el margen escogido es amplio, por ejemplo 50 segundos, la persona podría ser no inferior al campeón olímpico si él hace un tiempo de nueve segundos y ella 59 segundos, ya que llegó dentro del rango establecido. De tal forma, aunque la persona no es igual que el campeón, se puede concluir válidamente que no es peor (o no inferior) que él.

Así pues, este margen o “delta” de no inferioridad debe establecerse con adecuado conocimiento clínico y estadístico. Por ejemplo, un estudio clínico puede establecer que la terapia innovadora será hasta 33 % peor que el estándar y aún así considerarse no inferior. En tal caso, el clínico y el paciente deberán estar dispuestos a perder eficacia, con tal de ganar alguna otra ventaja, como menos efectos adversos o una aplicación más sencilla de la droga.

La parte matemática de establecer el margen de no inferioridad debe contemplar un delta que asegure, como mínimo, una eficacia de la droga innovadora al menos superior que el

placebo. Por ejemplo, imaginando que una terapia estándar es superior a placebo por un margen de 20 %, y el estudio de no inferioridad de la nueva terapia *versus* el estándar emplea un margen del 33 %, el fármaco innovador podría ser no inferior que el comparador, pero podría ser inferior al placebo. Por tanto, la decisión estadística del margen de no inferioridad usualmente comprende la elección de al menos un 50 % de eficacia que tenía el estándar sobre el placebo, determinado a partir de los estudios pivotaes de la molécula original, o bien, a partir de metaanálisis que hayan determinado un “promedio” de la misma medida de asociación que empleará el estudio de no inferioridad. Tomando esto en cuenta, en el caso anterior, el máximo margen de no inferioridad correspondería a un 10 %.

Como se puede entender, la decisión del margen de no inferioridad es arbitraria y caprichosa según los deseos del investigador. Sin embargo, para llegar a conclusiones válidas, este margen debe ser siempre reportado y justificado a partir de estudios previos en donde se conserve el beneficio del nuevo tratamiento.

Prueba de hipótesis e intervalo de confianza en estudios de no inferioridad

Las pruebas de hipótesis que se emplean en los análisis de no inferioridad se denominan en conjunto pruebas de “una cola”, ya que evalúan específicamente si la intervención novedosa produce un efecto que no traspase el margen de no inferioridad.

Las siguientes son, por lo tanto, las hipótesis estadísticas que se emplean cuando se quiere probar la no inferioridad de una intervención *versus* otra:

$$\text{Hipótesis nula} = HR \geq \delta$$

$$\text{Hipótesis alternativa} = HR < \delta$$

En este ejemplo se utilizó como medida de asociación el *hazard ratio*, aunque también se pueden emplear otras medidas, como el riesgo relativo o el riesgo absoluto.

Conviene recordar que el reporte del estimado de la medida de asociación (riesgo relativo, riesgo absoluto, *odds ratio* o *hazard ratio*), debe acompañarse siempre de su respectivo intervalo de confianza. En el caso de los estudios de no inferioridad, interesa que uno de los extremos de dicho intervalo (el extremo más inferior) no supere el margen preestablecido. De esta manera, si el extremo del intervalo de confianza no sobrepasa el límite de la no inferioridad, se concluirá, en efecto, la NO inferioridad de la intervención.

La Figura N° 12 ilustra las posibles conclusiones de este tipo de estudios.

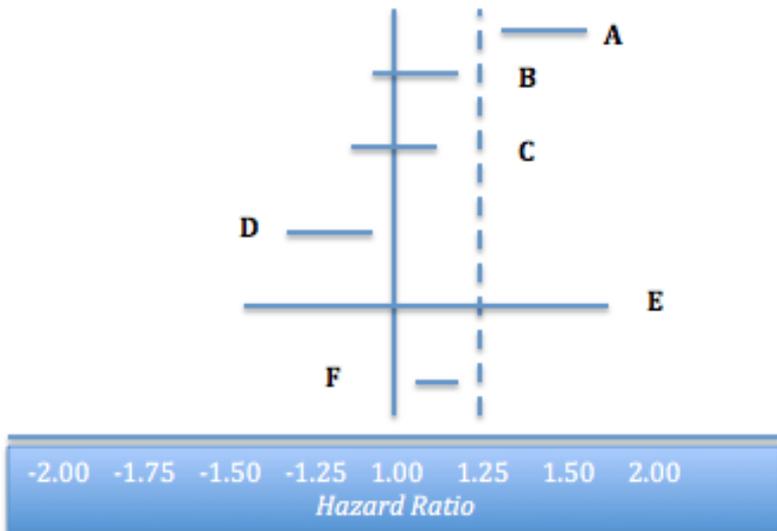


Figura N° 12. Posibles conclusiones de un estudio de no inferioridad (la línea punteada corresponde al margen de no inferioridad ($\Delta=1,25$)). Caso A: inferior. Caso B y C: no inferior. Caso D: no inferior y superior. Caso E: indeterminado. Caso F: no inferior.

- **Ejemplo:**

Se desea establecer el *hazard ratio* (HR) entre dos intervenciones, con el objetivo de determinar si existe la no inferioridad de un fármaco novedoso sobre el estándar. Tomando de base la Figura N° 12, se puede determinar lo siguiente: la línea continua indica un HR de 1, lo que significa que no hay diferencia entre ambos medicamentos. La línea punteada, por su parte, denota el margen de no inferioridad establecido, que es del 25 %. En el caso A, el nuevo medicamento es claramente inferior al estándar, ya que todo el intervalo de confianza se encuentra por fuera del margen de no inferioridad. En los casos B y C se puede concluir la no inferioridad del nuevo fármaco (nótese que el extremo derecho del intervalo de confianza NO cruza el delta del 25 %). En el caso D, se concluye que el nuevo medicamento no es inferior al estándar; más bien, es superior, ya que la totalidad del intervalo de confianza se encuentra por fuera del margen de no inferioridad y tampoco incluye el valor de $HR=1$. En el caso E, no puede llegarse a ninguna conclusión, por cuanto el intervalo de confianza incluye valores inferiores al margen predefinido, pero también comprende valores superiores o incluso iguales a un HR de 1. En el último ejemplo (F), el nuevo fármaco resulta NO ser inferior al estándar, pero sí inferior al comparador, ya que la totalidad del intervalo de confianza se encuentra dentro de dicho margen.

Análisis de los datos

Tipo de análisis

Otras cuestiones metodológicas son igual de importantes a la hora de establecer conclusiones válidas a partir de un estudio de no inferioridad. Dentro de estas, destaca el tipo de análisis de los datos. Como ya se había mencionado, el conteo de los eventos de interés puede realizarse por dos métodos: análisis por intención a tratar (ITT) y análisis por protocolo (PP). A pesar de que el análisis ITT es el preferido en los estudios de

superioridad, en los análisis de no inferioridad se recomienda utilizar los dos tipos y, por ende, reportar ambos resultados, ya que el análisis ITT tiende a generar sesgos hacia la no inferioridad. Por ejemplo, si existiera un entrecruzamiento de los sujetos durante el desarrollo del estudio, es probable que el efecto final del tratamiento se diluya en un grupo. Lo mismo puede ocurrir cuando se presentan pérdidas durante el seguimiento y en casos de no adherencia.

Poder estadístico

Igual cautela debe tenerse al momento de establecer el poder estadístico. Un estudio con poco poder estadístico tiende a sesgar hacia la no inferioridad, por lo cual se prefiere al menos un poder del 90 %. En el caso de la tasa de errores tipo I (alfa), se aconseja un valor de 0,025, ya que se trabajará con una prueba de hipótesis de una cola.

ANÁLISIS DE REGRESIÓN

Una de las preguntas básicas de la epidemiología clínica es el pronóstico o la predicción de una variable de interés. Es común que el médico y el paciente quieran saber cuáles resultados esperar con una terapia o posterior a un diagnóstico. En tales casos, los análisis de regresión surgen como el método estadístico preferido para contestar esos cuestionamientos.

Los análisis de regresión comúnmente empleados son:

- **La regresión lineal:** parte del supuesto de que la relación entre dos variables tiene una forma lineal.
- **La regresión logística:** en la cual se pretende relacionar una variable binaria con otras.
- **La regresión de COX:** detallada en el apartado “Análisis de supervivencia”.

En todos estos casos se pretende predecir una variable y (variable dependiente) a partir de una o más variables x (variables independientes).

En general, las regresiones pueden clasificarse también en univariadas o multivariadas, dependiendo de si la variable independiente es única o múltiple.

El tipo de regresión a emplear dependerá de la naturaleza de la variable y , tal como se observa en el Cuadro N° 6.

Cuadro N° 6. Tipos de regresión	
Tipo de variable dependiente (y)	Tipo de regresión
Continua	Lineal
Categoría binomial	Logística
Continua (temporal)	Modelo proporcional de Cox

Además de lo indicado, los modelos de regresión pueden utilizarse para determinar la asociación entre dos variables, para lo cual se emplean los coeficientes de correlación (r). Asimismo, estos modelos matemáticos son útiles para “ajustar” la relación de una variable dependiente con la variable independiente bajo la influencia de una o más variables confusoras.

Todos los tipos de regresiones son modelos matemáticos que deben cumplir con ciertos supuestos para ser válidos, entre los cuales se encuentran: la bondad de ajuste, la homocedasticidad y, en algunos casos, la normalidad de alguna de las variables del modelo.

Ejemplos de análisis de regresión

- *Ejemplo 1:*

Se quiere saber si el peso de una persona (y) depende de la edad (x). En este caso, la regresión correspondería a un modelo lineal, ya que el peso del sujeto (variable dependiente) es una variable continua. Además, dicha regresión sería univariada o simple.

- *Ejemplo 2:*

Se quiere determinar si el peso de una persona (y) varía no solo por la edad, sino también por el sexo. En este caso, la regresión se clasificará como lineal (y = variable continua) y múltiple (dos variables independientes = edad y sexo).

El modelo matemático de regresión lineal múltiple también permite conocer cuál es la asociación entre peso y edad bajo la influencia del sexo, o como suele describirse más frecuente: ¿cuál es la asociación entre peso y edad ajustada por el sexo del individuo?

- *Ejemplo 3:*

Se desea conocer la probabilidad de que una persona tenga diabetes mellitus tipo 2 según su peso. En este caso, la regresión sería logística simple, ya que la variable dependiente (ocurrencia de diabetes mellitus) es binomial, es decir, un sujeto tendrá o no diabetes mellitus tipo 2.

Si a este modelo se le agrega la influencia de otras variables como la edad, el diámetro abdominal y los antecedentes familiares, se estará hablando, entonces, de una regresión logística múltiple.

Regresión lineal simple

El caso más sencillo de regresión corresponde a la regresión lineal simple, en donde se busca relacionar dos variables, siendo la variable dependiente continua, como en el ejemplo 1.

Suponiendo que se desea predecir el peso de la población a partir de las edades de unos individuos, los resultados de la investigación se deben ilustrar mediante un gráfico de dispersión, como el que se muestra en la Figura N° 13.

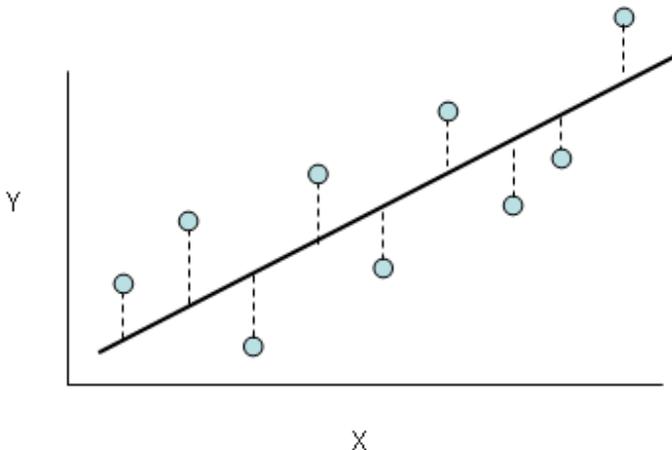


Figura N° 13. Línea de mejor ajuste para una regresión lineal.

En este gráfico cada punto corresponde a un individuo y cada línea discontinua corresponde al denominado “residuo”; es decir, a la distancia entre el valor observado y el valor predicho por la línea continua. Esta línea de “mejor ajuste”, que atraviesa los puntos observados, generando la mínima suma de residuos (método de mínimos cuadrados), corresponderá al modelo matemático de la regresión lineal simple.

Puede demostrarse que esta línea sigue lo indicado en la Fórmula 28, en donde y corresponde a la variable dependiente (peso) y x corresponde a la variable independiente (edad).

- Fórmula 28:

$$y = \beta x + \alpha$$

El valor de beta (β) se conocerá como el coeficiente de regresión y el valor de alfa (α) como el intercepto. Nótese que el coeficiente de regresión se interpretará como el incremento medio de y por unidad de x .

Si se trabaja con una muestra, el valor calculado de β corresponde a un estimado del verdadero β de la población. Por lo tanto, dicha variable posee una precisión específica. Similarmente, toda la línea de regresión corresponde a un estimado de la verdadera línea que explicaría la asociación entre ambas variables en la población. Si el valor de β es distinto de cero, se puede concluir que x predice de forma estadísticamente significativa la variable y .

El modelo anterior puede “complicarse” fácilmente agregando cuantas variables desee el investigador, para predecir de forma significativa el valor de y . No obstante, el uso de múltiples variables en un modelo de regresión debe evitarse, ya que produciría un “*overfitting*” o sobreajuste, de forma tal que el modelo no será fácilmente replicable en otra muestra o población.

Las variables a emplear deben escogerse *a priori* y con sentido fisiológico o clínico.

Regresión logística

Otros modelos de predicción son más complejos de graficar, como los de regresión logística. Un ejemplo sería intentar hacer un gráfico de dispersión cuando la variable dependiente es binomial. Suponiendo que en el ejemplo 3 se quiere predecir la presencia (1) o ausencia (0) de diabetes mellitus tipo 2 de acuerdo al peso de un grupo de sujetos, la gráfica sería similar a la que se presenta en la Figura N° 14.

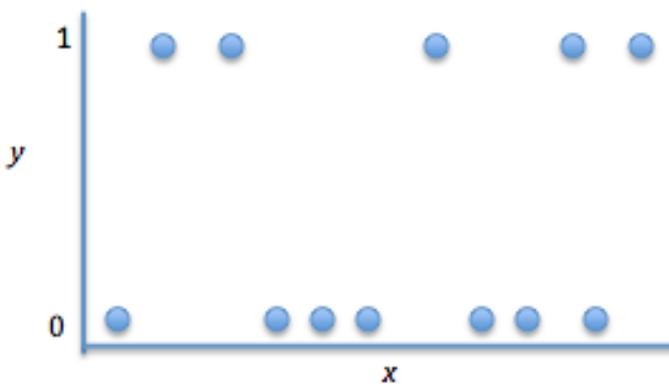


Figura N° 14. Distribución de los datos de x si la variable dependiente y es binomial.

Como se puede inferir, al tener la variable dependiente solo dos valores posibles (ausencia o presencia de la enfermedad), no puede trazarse fácilmente una línea recta que correlacione ambas variables. Por tal razón, se prefiere convertir o transformar la variable dependiente y en una variable continua. Se ha demostrado que la mejor forma para convertir una variable dicotómica en una variable continua es mediante el cálculo del logaritmo del *odds* de y . En este caso, aplicando la Fórmula 10, el *odds* de y sería la probabilidad de presentar la enfermedad entre la probabilidad de no presentarla.

- Fórmula 10:

$$\text{Odds} = \frac{\text{Probabilidad del evento}}{1 - \text{probabilidad del evento}} = \frac{p}{1 - p}$$

De tal forma, la ecuación de la regresión logística sería:

- Fórmula 29:

$$\text{Ln}\left(\frac{p}{1 - p}\right) = \beta x + \alpha$$

De dicha fórmula se deriva la siguiente (Fórmula 30), que tiene como objetivo eliminar el término logarítmico.

- Fórmula 30:

$$Odds = e^{\beta x + \alpha} = \frac{p}{1 - p}$$

Si la variable x es dicotómica ($x=1$ o $x=0$), puede calcularse el *odds ratio* mediante la Fórmula 31:

$$Odds\ ratio = \frac{e^{\beta x + \alpha}}{e^{\alpha}} = e^{\beta x + \alpha - \alpha} = e^{\beta x}$$

Entonces, la regresión logística también sirve para calcular el *odds ratio* de una variable dicotómica, y como cualquier estimado, también puede calcularse su respectivo intervalo de confianza.

ANÁLISIS DE SUPERVIVENCIA

Muchas de las variables de interés dependen del tiempo, por ejemplo: la supervivencia global, el tiempo libre de progresión de un tumor, el tiempo libre de enfermedad cardiovascular, entre otros. En todos estos casos importa no solo conocer el número de eventos, sino también el momento en el cual ocurrieron, tal como se observa en la Figura N° 15.

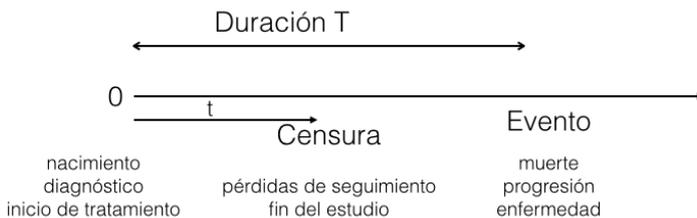


Figura N° 15. Posibles resultados de un estudio de supervivencia.

En dicha figura se muestran las variables imprescindibles para un análisis de supervivencia:

- **El momento cero (0):** suele hacer referencia al inicio del tratamiento o a la fecha del diagnóstico de la enfermedad de interés, ya que marca el inicio del tiempo de supervivencia.
- **La censura (*censoring*):** este término hace referencia a todos aquellos eventos que suponen la finalización del tiempo de seguimiento por algún otro evento que no sea el de interés, por ejemplo: cuando finaliza el estudio y el sujeto no presentó el evento, la pérdida de seguimiento o la salida del estudio por parte del participante. En general, la censura supone que no hay información sobre cuál es el tiempo de supervivencia para dicho individuo, ya que este no presentó el evento de interés.
- **La duración T:** se refiere al tiempo transcurrido desde el momento cero hasta la ocurrencia del evento de interés o censura.

Todos aquellos métodos estadísticos que involucren el uso de datos censurados se denominan en conjunto análisis de supervivencia. Si bien el término supervivencia se asocia a mortalidad, cabe destacar que estos análisis pueden realizarse para múltiples tipos de eventos de interés que ocurren a lo largo del tiempo.

Función de supervivencia y función condicional de fallo

La distribución probabilística que caracteriza el tiempo de supervivencia se denomina función de supervivencia (*survival function*), o en el caso opuesto, función condicional de fallo (*hazard function*). La función de supervivencia $S(t)$ brinda la probabilidad de que una persona sobreviva más allá del período t . Por el contrario, la función condicional de fallo $h(t)$ brinda el potencial instantáneo por unidad de tiempo de que el

evento ocurra en caso de que el individuo haya sobrevivido al momento t .

En la Figura N° 16 se ilustra un ejemplo de una función de supervivencia.

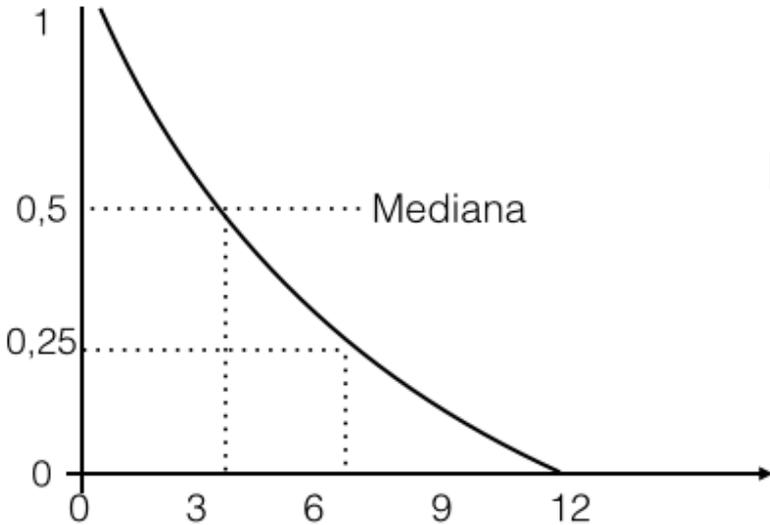


Figura N° 16. Función de supervivencia según el tiempo.

A partir de las curvas de supervivencia, es posible calcular un *hazard rate* “promedio” (utilizando la Fórmula 32) para todo el grupo de sujetos que conforman la curva, tomando el número de eventos y el tiempo T de cada individuo.

- Fórmula 32:

$$\text{Hazard rate promedio} = \frac{\# \text{ de eventos}}{\sum_{i=1}^n t_i}$$

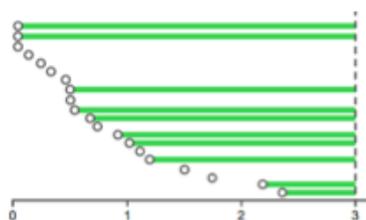
Métodos para estimar la función de supervivencia

Método de Kaplan-Meier

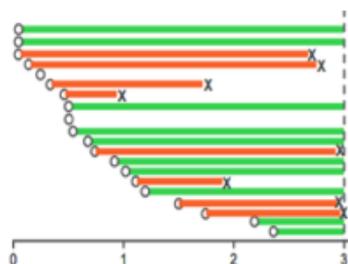
Existen diferentes métodos para calcular la función de supervivencia; el más empleado es el conocido como Kaplan-Meier, cuyo funcionamiento matemático se desarrolla en cinco pasos:

1. Ordenar en forma ascendente los períodos de seguimiento de todos los sujetos, incluyendo los individuos que tuvieron el evento de interés, así como aquellos cuyo seguimiento fue censurado.
2. Generar intervalos de tiempo cuyo límite superior sea el momento del evento de interés.
3. Suponer que las censuras ocurren exactamente al final del intervalo recién definido en el paso 2.
4. Estimar la probabilidad de supervivencia para el final de cada intervalo.
5. Multiplicar de forma aditiva todas las probabilidades de supervivencia para cada intervalo.

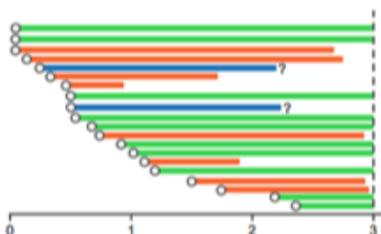
En la Figura N° 17 se ilustra el paso número uno; mientras que en el Cuadro N° 7 se presenta la llamada tabla de supervivencia, donde se resumen los pasos del dos al cinco.



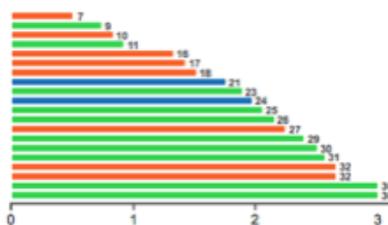
En primer lugar se añade el tiempo de los sujetos que no presentaron el evento de interés a lo largo del estudio.



A continuación se añaden los tiempos de seguimiento de los pacientes que si presentaron el evento de interés (X).



Posteriormente se añaden los períodos de seguimiento de los sujetos con datos censurados (?).



Finalmente, todos los tiempos de seguimiento son ordenados de menor a mayor, independientemente de si el paciente tuvo el evento de interés (X) o si el tiempo de seguimiento fue censurado (?).

Figura N° 17. Procedimiento gráfico del cálculo del período de seguimiento para el análisis de supervivencia, según el método de Kaplan-Meier.

Cuadro N° 7. Probabilidad de supervivencia por intervalo de tiempo						
Meses	Vivos al comienzo del intervalo	Censuras durante el intervalo	En riesgo del evento	Vivos al final del intervalo	Probabilidad de supervivencia durante el intervalo	Probabilidad acumulada de supervivencia
0-7	20	0	20	19	$19/20 = 0,95$	0,95
7-10	19	1	18	17	$17/18 = 0,94$	$0,95 \times 0,94 = 0,89$
10-16	17	1	16	15	$15/16 = 0,94$	$0,94 \times 0,89 = 0,84$
16-17	15	0	15	14	$14/15 = 0,93$	$0,93 \times 0,84 = 0,78$
17-18	14	0	14	13	$13/14 = 0,93$	$0,93 \times 0,78 = 0,73$
18-27	13	5	8	7	$7/8 = 0,87$	$0,87 \times 0,73 = 0,64$
27-32	7	3	4	2	$2/4 = 0,50$	$0,50 \times 0,64 = 0,32$
32-36	2	0	2	2	$2/2 = 1,00$	$1,00 \times 0,32 = 0,32$

Como se observa en el Cuadro N° 7, en la primera columna, el paso dos exige dividir el total del tiempo de seguimiento en intervalos, cada uno definido por el momento en el cual un paciente tiene el evento de interés. Por ejemplo, el primer período corresponde al comprendido entre el mes cero y el mes siete, cuando murió la primera persona. Para este período se calcula la probabilidad de permanecer vivo al mes siete (sexta columna). Esta probabilidad será calculada tomando el número de pacientes vivos al final del intervalo entre los sujetos en riesgo de morir. Nótese que el cálculo del número de sujetos en riesgo durante cada período no incluye los sujetos que tuvieron censura. Finalmente, se calcula la probabilidad acumulada de supervivencia para cada intervalo (última columna).

El estimado mediante el método de Kaplan-Meier está dado por la siguiente fórmula:

- Fórmula 33:

$$S(t) = \prod_{ti \leq t} \left[1 - \left(\frac{di}{ri} \right) \right]$$

Como la supervivencia y el hazard son eventos opuestos, la función de riesgo viene dada por la siguiente fórmula:

- Fórmula 34:

$$H(t) = -\ln(S)$$

De esa misma fórmula se deriva la siguiente:

- Fórmula 35:

$$S(t) = e^{-H*t}$$

Por ejemplo, si el *hazard ratio* es de 0,1 y se desea saber al año cuál es la probabilidad de supervivencia, se aplica la Fórmula 35; con lo cual se obtiene una probabilidad de supervivencia del 90,5 %, deduciéndose que un 9,5 % de los individuos han presentado un evento:

$$S(t) = e^{-H*t}$$

$$S(1) = e^{-0,1*1} = e^{-0,1} = 0,905$$

Las probabilidades de supervivencia pueden graficarse como se muestra en la Figura N° 18, también denominada curva de supervivencia.

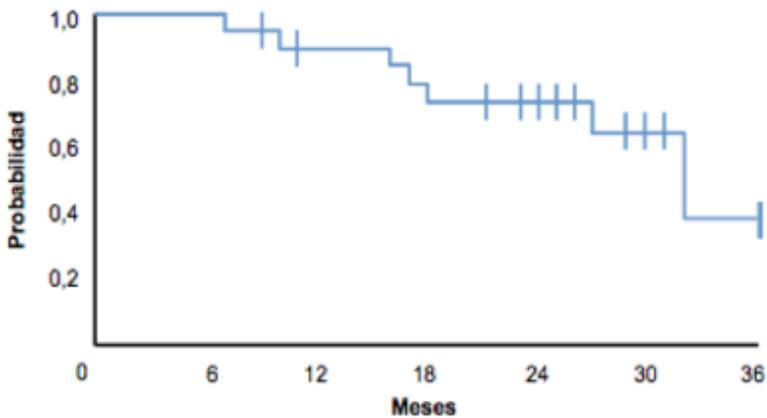


Figura N° 18. Curva de supervivencia según el método de Kaplan-Meier.

En esta gráfica se nota cómo al momento cero la probabilidad de supervivencia es de uno y desciende conforme ocurre cada evento. De tal manera, cada peldaño denota un evento de interés, mientras que cada raya vertical indica un paciente cuyo tiempo fue censurado.

Si se comparan los datos del Cuadro N° 7 con los de la Figura N° 18, se podrá observar cómo cada punto de la curva corresponde a las probabilidades de supervivencia calculadas a partir de los datos de la muestra. En este punto cabe recordar que cada dato posee su precisión, denotada por el intervalo de confianza.

Siguiendo con el análisis de la Figura N° 18, se puede afirmar, entonces, que toda la curva representa la función de supervivencia o riesgo del evento de interés. Si bien todos los puntos de la curva son importantes, conviene resaltar dos características de esta: la primera de ellas es el cálculo de la mediana de supervivencia, que corresponde al tiempo en el cual el 50 % de los sujetos posee la probabilidad del evento de

interés; este valor es fácil de obtener con la visualización del gráfico, buscando el valor de tiempo que corresponda a una probabilidad del 50 %. En este caso, la mediana de supervivencia está entre los 30 y los 36 meses. La segunda característica es que la precisión de la curva en su extremo final es poca, ya que dicho estimado se basó en pocos eventos y con un remanente de pacientes bajo. Por ende, las conclusiones de la función de supervivencia deben hacerse en relación con el comportamiento de toda la curva, sin enfatizar en algún punto al azar.

Por otra parte, se debe mencionar que el método de Kaplan-Meier puede utilizarse también para calcular las probabilidades de supervivencia de dos o más grupos, y de esta forma, comparar cuál grupo particular tiene una mayor o menor tasa de eventos. En la Figura N° 19 se muestra un ejemplo de esto.

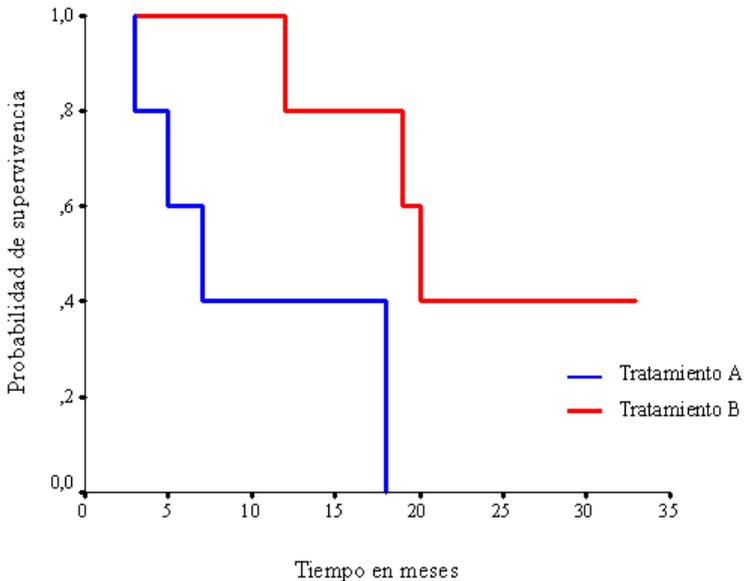


Figura N° 19. Curva de supervivencia según el método de Kaplan-Meier para el tratamiento A y B.

En este punto es posible comparar formalmente ambas curvas, mediante una prueba estadística. De nuevo, la hipótesis nula de tal prueba denotará la ausencia de diferencias entre ambas curvas, en tanto la hipótesis alternativa reportará lo contrario. El método utilizado con mayor frecuencia para tal fin se denomina prueba de Log-rank. Este método posee como principal supuesto matemático que los riesgos de ambas curvas son proporcionales, es decir, que no hay entrecruzamiento de curvas, y que una curva discurre paralela a la otra. En casos de no proporcionalidad de las tasas de riesgo, deben emplearse otros métodos estadísticos.

Regresión de Cox

El modelo de regresión de Cox es un método matemático que predice la probabilidad de supervivencia para un individuo en un tiempo T (variable dependiente), según la presencia de una o más variables independientes. Al igual que en el método anterior, se requiere una transformación logarítmica de la variable dependiente, para lograr una relación fácilmente interpretable. La fórmula general para este modelo es la siguiente, donde X_i corresponde a cada una de las variables explicativas o predictoras de h :

- Fórmula 36:

$$h(t, X) = h_0 \exp\left[\sum_{i=1}^p \beta_i X_i\right]$$

De forma análoga a lo que se describió para la regresión logística, cada variable X_i tiene un coeficiente de corrección β_i , cuya interpretación corresponde al *hazard ratio* de dicha variable.

Una aplicación práctica de este modelo es cuando se quiere explicar la probabilidad de muerte dependiendo de si un sujeto recibió el tratamiento X_i A o el tratamiento B.

REVISIONES SISTEMÁTICAS Y METAANÁLISIS

En muchas ocasiones se desea conocer el efecto de una intervención o medicamento en un tema específico. Sin embargo, la cantidad de información es abrumadora y con frecuencia las conclusiones de un estudio son opuestas a la de otro similar. En estos casos, surgen las denominadas revisiones sistemáticas, en las cuales se pretende resumir de forma concreta la información disponible en torno a dicho tema.

Estas revisiones deben contener criterios de inclusión y exclusión específicos, los cuales permitan reproducir los mismos hallazgos encontrados por otro investigador. Deben además contener información sobre el sitio de la búsqueda, así como la fecha y el período en que esta se efectuó y las palabras clave empleadas. Asimismo, deben definir el diseño de estudios que se desea incluir.

Las revisiones sistemáticas suelen ser realizadas por distintos autores, de forma tal que la búsqueda sea efectuada por separado por cada uno ellos, quienes deberían llegar a una inclusión similar de estudios.

Es importante mencionar que estas revisiones difieren de simples revisiones narrativas o editoriales, por cuanto incluyen criterios específicos para la inclusión de los estudios, de tal manera que la presentación de los datos no queda al criterio caprichoso del autor.

Por lo general, los metaanálisis acompañan a las revisiones sistemáticas para generar un “promedio ponderado” del efecto de la intervención. Además, los metaanálisis intentan responder alguna de las cuatro preguntas pilares de la epidemiología clínica, con mayor poder estadístico que un único ensayo clínico, ya que incluyen a todos los pacientes de cada estudio individual. Esto hace que el estimado calculado sea más preciso. Sin embargo, los metaanálisis también reproducen los errores de cada estudio y en ocasiones la suma de los

hallazgos específicos de los ensayos no es procedente, ya que cada estudio posee diferencias con respecto a otro. A la vez, puede existir un sesgo de publicación; es decir, pueden no incluirse los resultados de ciertos estudios que no han sido publicados o que no se identificaron en la búsqueda sistemática.

Los pasos de una revisión sistemática y metaanálisis se resumen en el Cuadro N° 8.

**Cuadro N° 8. Pasos de una revisión sistemática
con metaanálisis**

1. Definir los objetivos de la revisión.
2. Determinar las reglas de inclusión y exclusión de los estudios.
3. Identificar los estudios en bases de datos a través de una estrategia de búsqueda.
4. Aplicación de los métodos estadísticos para el cálculo ponderado del efecto.

En el desarrollo o lectura de un metaanálisis surgen dos preguntas básicas, que deben ser contestadas:

1. ¿Cómo determinar el efecto combinado? Es decir, ¿cómo se calcula el “promedio” ponderado?
2. ¿Cómo determinar si hay diferencia (heterogeneidad) en los estudios incluidos?

Cálculo del promedio ponderado

Para responder la primera pregunta, se debe calcular un promedio del efecto estimado en cada estudio. La manera más sencilla sería la suma de cada estimado individual entre el número total de estudios. No obstante, esta forma de cálculo toma en cuenta que todos los estudios son exactamente iguales y les concede el mismo peso a los estudios pequeños y a los grandes. Por lo tanto, es necesario “ponderar” cada

estudio, asignándole un peso específico. Esto es similar al cálculo de una nota de un curso: unos exámenes parciales tienen un valor porcentual (o peso) menor al de la nota de un examen final, por cuanto este último contempla más materia y preparación. De forma análoga, el cálculo del efecto combinado le dará un peso determinado a cada estudio individual. Existen dos métodos que permiten conceder este peso a los estudios clínicos incluidos en un metaanálisis: el método de efectos fijos y el método de efectos aleatorios.

Método de efectos fijos

En este análisis estadístico, el peso de cada estudio se determina por la Fórmula 37; es decir, calculando el inverso de la varianza:

$$\text{peso } (w) = \frac{1}{\text{varianza}}$$

Se debe recordar que la varianza está en relación inversa con el tamaño de la muestra. Esto implica que entre mayor el número de sujetos incluidos en un estudio, mayor será el peso en este modelo de efectos fijos. Dicho modelo asume que la única variación existente en todos los estudios incluidos proviene de la variabilidad dentro de cada estudio, lo cual es reflejado por la varianza. Así pues, el estimado final corresponderá a la Fórmula 38, en la cual se efectúa una ponderación de cada estimado individual según el peso que cada estudio aporte:

$$\text{Efecto combinado} = \frac{\Sigma \text{Efecto individual} * w}{\Sigma w}$$

Método de efectos aleatorios

El cálculo del peso a través de este método parte del supuesto de que los estudios incluidos en el metaanálisis proveen estimados diferentes. Por lo tanto, se asume que la variación

neta que se observa no es solo consecuencia de la variabilidad existente *dentro* de cada estudio individual, sino que también es la consecuencia de la variabilidad *entre* estudios. El peso, por ende, viene determinado tanto por la varianza individual de cada estudio, así como por otra medida denominada τ^2 , la cual mide la heterogeneidad entre estudios, tal como lo señala la Fórmula 39:

$$\text{peso } (w) = \frac{1}{\text{varianza} + \tau^2}$$

Este método es el de preferencia cuando se asume que los estudios incluidos son variables entre sí. Finalmente, el efecto combinado se calculará como lo indicado en la Fórmula 38.

Cálculo de la heterogeneidad

Un punto crítico en la elaboración o lectura de un metaanálisis es la comprobación de homogeneidad entre los diferentes estudios. De comprobarse homogeneidad, el cálculo del estimado combinado tendrá mayor validez interna.

Observe con detenimiento las figuras N° 20 y N° 21.

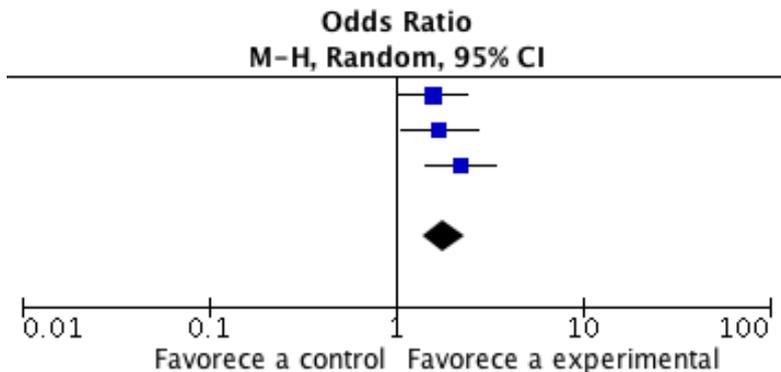


Figura N° 20. Simulación de los resultados de un metaanálisis con poca heterogeneidad entre estudios.

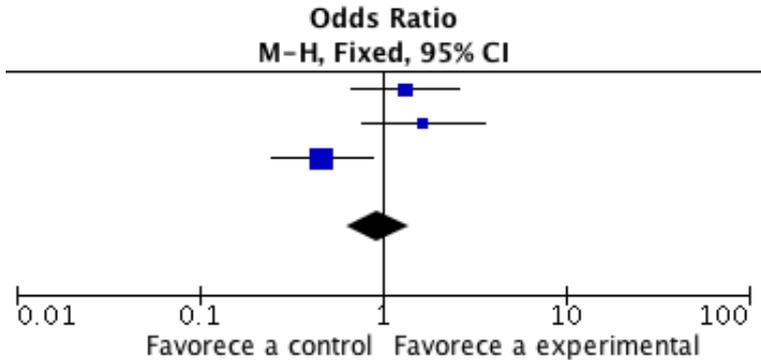


Figura N° 21. Simulación de los resultados de un metaanálisis con alta heterogeneidad entre estudios.

Resulta claro que la Figura N° 20 incluye tres estudios cuyo *odds ratio* es similar, en tanto la Figura N° 21 ilustra un caso de un metaanálisis con estudios heterogéneos. En este último caso, el estudio ubicado en la tercera línea no reproduce el efecto de los otros dos estudios incluidos; por lo tanto, este metaanálisis podría contener sesgos que dificulten su interpretación.

Este paso realizado anteriormente se conoce como inspección visual del gráfico o *forest plot* y constituye la primera forma para determinar heterogeneidad en los estudios que conforman el metaanálisis.

La segunda manera de identificar heterogeneidad, es a través de una prueba estadística y el cálculo de la *Q de Cochran* o simplemente *Q*. Esta prueba tiene las siguientes hipótesis:

Hipótesis nula = No hay heterogeneidad entre estudios

Hipótesis alternativa = Hay heterogeneidad entre estudios

La prueba estadística calculará un valor de p , el cual entre más bajo resulte, menor será la evidencia de la hipótesis nula cierta.

Para esta prueba en particular, un valor de p por debajo de 0,05 confirmará la heterogeneidad entre estudios. Sin embargo, las hipótesis nulas no pueden ser aceptadas con el cálculo del valor de p por encima de 0,05. Por ende, un valor de p mayor a 0,05 tampoco confirma homogeneidad de estudios. Además, el valor de la Q de Cochran tampoco brinda una idea de qué tanta heterogeneidad existe.

Una forma de estimar la heterogeneidad es mediante el cálculo de I^2 . Este valor corresponde a un porcentaje, que de ser mayor al 75 % implica una heterogeneidad sustancial. El cálculo de I^2 surge de la siguiente fórmula:

- Fórmula 40:

$$I^2 = \frac{Q - \text{número de estudios} - 1}{Q}$$

¿Qué se puede hacer al detectar heterogeneidad?

Una respuesta simplista sería no realizar metaanálisis, ya que siempre existirá algún grado de heterogeneidad entre estudios. Sin embargo, el principal valor de un metaanálisis es identificar cuáles serían las posibles causas de heterogeneidad; por ejemplo, si es a consecuencia de una intervención ligeramente distinta, una población de estudio diferente o una determinación desigual del estimado final obtenida a través de un diseño no homogéneo entre estudios.

En párrafos anteriores se mencionó que un análisis de efectos aleatorios intentará ajustar el peso de cada estudio según la variabilidad observada entre estos. No obstante, la mejor forma de lidiar con la heterogeneidad es realizando un análisis de metarregresión. En este sentido, se debe recordar que las regresiones corresponden a fórmulas matemáticas que intentan relacionar una variable con otra(s). En este caso, se pueden identificar variables predictivas que expliquen la heterogeneidad encontrada.

¿CÓMO SE CALCULA LA MUESTRA DE UN ENSAYO CLÍNICO?

Una pregunta básica en el diseño y lectura de cualquier estudio es la cantidad de individuos que deben ser incluidos en un ensayo para que los hallazgos sean válidos. Esta interrogante surge a partir de cuestiones éticas y estadísticas. Desde el punto de vista ético, es imperante que un estudio no incluya una cantidad extra de sujetos que eventualmente se expondrán a una terapia experimental sin asegurar un beneficio clínico.

Por otra parte, tampoco tiene sentido reclutar tan pocos participantes que los resultados carezcan de validez científica y que la exposición sea “en vano”. Desde el punto de vista estadístico, es necesario precisar la cantidad exacta de sujetos en quienes se puedan encontrar hallazgos significativos que reflejen lo que ocurriría en la población.

En la mayoría de los casos debe estimarse *a priori* el efecto que supondría una intervención o exposición específica, además de que debe establecerse la proporción deseada de falsos positivos y negativos.

Pasos para el cálculo de la muestra

Los pasos necesarios para el cálculo de la muestra son:

1. Definir el objetivo primario y el tipo de variable en la que se determinará tal resultado.
2. Plantear la hipótesis estadística del estudio.
3. Seleccionar la prueba de hipótesis indicada según el tipo de variable primaria.
4. Estimar el efecto planeado de la intervención y su variabilidad.
5. Seleccionar el error tipo I y tipo II.
6. Utilizar la fórmula apropiada.

Paso 1: Definición del objetivo primario y del tipo de variable

Al momento de definir el objetivo primario o hipótesis del estudio, se debe considerar que este debe ser muy claro y conciso, de forma tal que se eviten ambigüedades y desviaciones de la pregunta principal de la investigación. Este objetivo puede ser simple o compuesto, y puede corresponder a la determinación de eventos clínicos o a mediciones subrogadas de una condición específica.

Una vez definido el objetivo primario, se debe seleccionar y clasificar la variable con la cual se medirá el evento de interés. Por ejemplo, un ensayo clínico que pretenda comparar el efecto antihipertensivo de dos drogas, puede emplear como variable del objetivo primario los valores de presión arterial, o bien, eventos clínicos como: enfermedad renal crónica, infartos, episodios de angina o muerte. No obstante, la determinación del efecto antihipertensivo de ambos fármacos también puede hacerse mediante la medición de variables subrogadas, como la rigidez arterial o el aclaramiento de creatinina. Si se escogiera en este ejemplo una variable subrogada, como el aclaramiento de creatinina, el autor debe definir si esta variable la medirá en términos numéricos (mililitro/minuto) o en términos categóricos (aclaramiento menor o mayor de 30 ml/min).

Otro tipo de variables que pueden emplearse en la evaluación del objetivo primario son aquellas dependientes del tiempo, como por ejemplo: la supervivencia o el tiempo libre de enfermedad.

Por lo general, el tamaño de la muestra es más grande si se emplean variables binarias.

Paso 2: Planteamiento de la hipótesis estadística

La hipótesis estadística suele ser de superioridad (tratamiento A mejor que el tratamiento B); o bien, de no inferioridad.

Las muestras de los estudios de no inferioridad suelen ser de mayor tamaño que las que se necesitan para establecer la superioridad de un fármaco *versus* otro.

Paso 3: Selección de la prueba de hipótesis

Seguidamente, se seleccionará la prueba de hipótesis, de acuerdo al tipo de variable que medirá los eventos de interés.

El Cuadro N° 9 resume las pruebas estadísticas más empleadas en el diseño de estudios clínicos.

Cuadro N° 9. Prueba de hipótesis según los tipos de variable		
Variable predictiva	Variable de interés	
	Binaria	Continua
Binaria	Prueba de chi-cuadrado	Prueba t de Student
Continua	Prueba t de Student	Coefficiente de correlación

Paso 4: Estimación del efecto planeado y su variabilidad

Todo cálculo del tamaño de la muestra debe tener un estimado del tamaño del efecto. Así pues, la estimación de un efecto muy amplio de una intervención requerirá menor cantidad de participantes que cuando se anticipe un efecto limítrofe o marginal de un medicamento sobre otro.

Similarmente, debe considerarse la variabilidad de este estimado. Si el autor reconoce que la variable de interés es heterogénea en la población, deberá prever una mayor cantidad de individuos en la muestra, en comparación con un estudio que pretenda estimar una variable de comportamiento uniforme en la población.

Muchos estudios no logran demostrar la superioridad de una intervención sobre otra, simplemente por el hecho de haber previsto un efecto muy amplio (y ambicioso) de la intervención. En otros casos, los investigadores emplean una muestra muy grande y con ello demuestran que dos intervenciones son significativamente distintas entre sí, aun cuando el tamaño del efecto de la intervención es pequeño y, algunas veces, clínicamente insignificante.

Paso 5: Selección del error tipo I y error tipo II

Durante el diseño del estudio y de forma complementaria al planteamiento de la hipótesis estadística, deben seleccionarse los errores tipo I (alfa) y II (beta). Por conveniencia, el error tipo I suele establecerse en 0,05 y el error tipo II en 0,20. Con estos valores se evita ante todo obtener conclusiones “falsas positivas”.

En general, si el investigador desea mantener un margen bajo de errores estadísticos, deberá considerar un mayor tamaño muestral.

Paso 6: Utilización de la fórmula apropiada

El paso final consiste en emplear fórmulas matemáticas, que contemplen los elementos anteriores, para obtener el número de sujetos que deberán incluirse en el estudio. La elección de estas fórmulas dependerá de las variables que se deseen estimar. A continuación unos ejemplos.

Cálculo de la muestra para determinar una proporción

- Fórmula 41:

$$n = \frac{z_{\alpha}^2(4)(p)(1-p)}{d^2}$$

Donde: $z_{\alpha/2}$ corresponde al valor de z de la curva de distribución normal para un intervalo de confianza deseado (por ejemplo: $z = 1,96$ para un intervalo de confianza del 95 % o $z = 2,58$ para un intervalo de confianza del 99 %); p es la proporción esperada de la población (si no se conoce, se debe escoger el valor de 0,5); y d corresponde a la amplitud deseada del intervalo de confianza de la proporción a calcular.

- *Ejemplo:*

Se quiere estimar la prevalencia de diabetes mellitus en una población de adultos con una precisión de $\pm 0,05$ ($d = 0,05$) y un intervalo de confianza del 95 % ($z_{\alpha/2} = 1,96$). Estudios previos han demostrado que la prevalencia de diabetes mellitus es del 10 % (p).

$$n = \frac{z_{\alpha/2}^2(4)(p)(1-p)}{d^2}$$

$$n = \frac{1,96^2(4)(0,10)(1-0,10)}{0,05^2} = \frac{15,37(4)(0,10)(0,90)}{0,05^2}$$

$$n = 553$$

Aplicando la Fórmula 41, la muestra requerida sería de 553 personas.

Cálculo de la muestra para determinar una media

- Fórmula 42:

$$n = \frac{z_{\alpha/2}^2(4)Sd^2}{d^2}$$

Nuevamente, en este caso $z_{\alpha/2}$ adquiere el valor de 1,96 si se desea un intervalo de confianza del 95 %. Sd corresponde a la desviación estándar ya conocida de la variable continua

(Fórmula 16) y d es la amplitud deseada del intervalo de confianza.

- *Ejemplo:*

Se quiere saber cuál es la media de glicemia en ayunas de un grupo de pacientes diabéticos. Se desea obtener un intervalo de confianza del 95 %, con una amplitud de 10 mg/dl. Se ha reportado en estudios previos que la desviación estándar de una población similar es 12,2 mg/dl.

$$n = \frac{z_{\alpha/2}^2(4)Sd^2}{d^2} = \frac{1,96^2(4)12,2^2}{10^2} = \frac{2.287}{100} = 23$$

La muestra en este caso sería de 23 pacientes diabéticos.

Cálculo de la muestra para comparar medias de dos grupos:

- Fórmula 43:

$$n = \frac{2K * Sd^2}{d^2}$$

El valor de K surge del siguiente cuadro:

Cuadro N° 10. Valores de K para comparar dos medias o dos proporciones			
Poder (1-β)	Valor de α		
	5 %	1 %	0,1 %
80 %	7,8	11,7	17,1
90 %	10,5	14,9	20,9
95 %	13,0	17,8	24,3

- *Ejemplo*

Se desea comparar dos medicamentos antihipertensivos. El investigador principal quiere detectar una diferencia de 5 mmHg entre ambos medicamentos. La desviación estándar de la presión arterial sistólica se ha reportado en 15 mmHg. El autor desea conservar un error tipo I del 5 % y un poder del 80 %. Aplicando la Fórmula 43, se obtienen 140 sujetos en cada grupo.

$$n = \frac{2K * Sd^2}{d^2} = \frac{2(7,8) * 15^2}{5^2} = 140$$

Cálculo de la muestra para comparar dos proporciones:

- Fórmula 44:

$$n = \frac{K [P_1(1 - P_1) + P_2(1 - P_2)]}{(P_1 - P_2)^2}$$

- *Ejemplo:*

Se quiere estudiar si dos quimioterapias producen remisión en un tipo de leucemia. Se espera que una droga induzca remisión en un 5 % y la otra droga en un 15 % de los pacientes. Se desea mantener un error tipo I del 5 % y un poder del 90 %.

$$n = \frac{K [P_1(1 - P_1) + P_2(1 - P_2)]}{(P_1 - P_2)^2}$$

$$= \frac{10,5 [0,15(1 - 0,15) + 0,05(1 - 0,05)]}{(0,15 - 0,05)^2} = 183$$

En este ejemplo, se necesitan 183 participantes en cada brazo de tratamiento.

Cálculo de la muestra en análisis de supervivencia

En los análisis de supervivencia, el cálculo del tamaño muestral va dirigido a determinar el número de eventos necesarios para lograr un *hazard ratio* deseado. Para ello, se emplea la Fórmula 45:

$$\text{Eventos} = \frac{\left(\frac{z_{\alpha}}{2} + z_{\beta}\right)^2}{\pi_1\pi_2(\ln HR)^2}$$

El valor de $\frac{z_{\alpha}}{2}$ suele ser de 1,96 en caso de que se desee un error tipo I de 0,05; mientras que el valor de z_{β} sería de: 0,84; 1,28; y 1,64 si se quiere un poder del 80 %, 90 % o 95 %, respectivamente. Los valores $\pi_1\pi_2$ hacen referencia al porcentaje de individuos a asignar en cada grupo (usualmente el valor de esta multiplicación es 0,25 (0,5*0,5) en caso de que se desee una asignación 1:1).

- *Ejemplo:*

Se quiere efectuar un ensayo clínico con una duración de un año, para comparar la supervivencia de dos grupos asignados a una droga experimental o a placebo.

Se pretende obtener un *hazard ratio* de 0,8 en un ensayo que tenga un poder del 90 % y un error tipo I del 5 %. ¿Cuál será el número de eventos requeridos?

$$\begin{aligned} \text{Eventos} &= \frac{\left(\frac{z_{\alpha}}{2} + z_{\beta}\right)^2}{\pi_1\pi_2(\log HR)^2} \\ &= \frac{(1,96 + 1,28)^2}{0,5 * 0,5 (\ln 0,8)^2} = \frac{10,51}{0,25 * 0,0498} = 845 \end{aligned}$$

De acuerdo con el resultado obtenido tras aplicar la Fórmula 45, para este estudio se requieren en total 845 eventos.

La Figura N° 22 muestra cómo el número de eventos de interés aumenta exponencialmente de acuerdo al *hazard ratio* deseado y al poder estadístico del estudio.

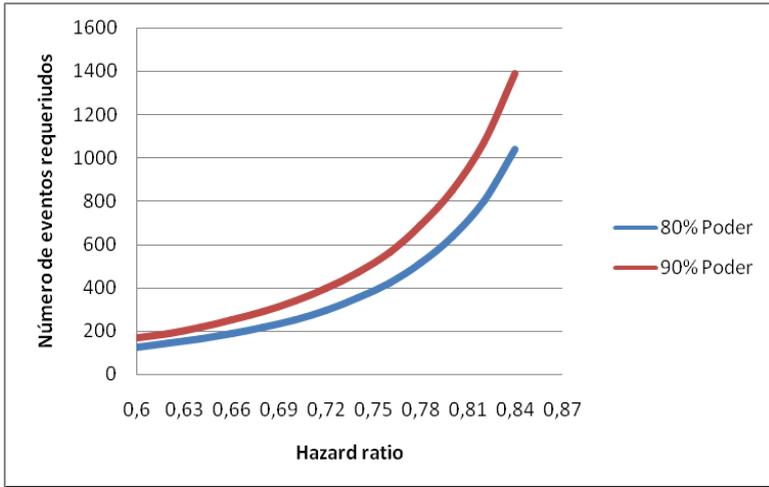


Figura N° 22. Número de eventos requeridos según el poder del ensayo y el *hazard ratio* deseado.

Una vez definido el número de eventos requeridos, se procede a calcular cuántos participantes se deben incluir en el estudio para obtener durante el tiempo de seguimiento tal cantidad de eventos. Para ello, es necesario determinar la probabilidad de sufrir un evento, según los estimados propuestos por el investigador. Este dato se obtiene con la Fórmula 46:

$$\text{Probabilidad de un evento} = 1 - \{\pi_1 S_1(T) + \pi_2 S_2(T)\}$$

Donde π_1 y π_2 hacen referencia a la proporción de individuos en el grupo control y experimental, respectivamente.

Para aplicar esta fórmula es importante tener presente que la probabilidad de supervivencia en un tiempo t se calcula con la Fórmula 35:

$$S(t) = e^{-H*t}$$

Una vez realizado este proceso, se aplica la Fórmula 47 para el cálculo específico de la muestra:

$$n = \frac{\text{eventos}}{\text{Probabilidad de un evento}}$$

- *Ejemplo:*

Se desea efectuar un ensayo clínico para comparar la supervivencia de dos grupos asignados a una droga experimental o a placebo, en un seguimiento de un año. Se pretende obtener un *hazard ratio* de 0,8 con un poder del 90 % y un error tipo I del 5 %. Ya se determinó del ejemplo previo que para este estudio es necesario que ocurran 845 eventos. Se sabe que el *hazard ratio* para el grupo control es de 0,5. Ahora se procede a calcular la probabilidad de un evento en cada grupo:

En el grupo experimental:

$$S(t) = e^{-H*t} = e^{-0,8(1)} = 0,45$$

En el grupo control:

$$S(t) = e^{-H*t} = e^{-0,5(1)} = 0,61$$

Aplicando la Fórmula 46:

$$\text{Probabilidad de un evento} = 1 - \{\pi_1 S_1(T) + \pi_2 S_2(T)\}$$

$$\text{Probabilidad de un evento} = 1 - \{0,5(0,45) + 0,5(0,61)\}$$

$$= 1 - (0,225 + 0,305)$$

$$= 1 - (0,53) = 0,47$$

Posteriormente, se utiliza la Fórmula 47:

$$n = \frac{\textit{eventos}}{\textit{Probabilidad de un evento}}$$

$$n = \frac{845}{0,47} = 1.798$$

De acuerdo a los resultados obtenidos, para este ejemplo es necesario reclutar a 1.798 pacientes (899 en cada grupo) y deben ocurrir 845 eventos para poder detectar un *hazard ratio* de 0,8 a favor del grupo experimental.

Otros aspectos a considerar para la definición de la muestra

Es importante resaltar que deben tomarse en cuenta imprevistos en el reclutamiento o seguimiento de los participantes de un estudio, que ameritan ajustar el tamaño de la muestra. Por ejemplo, si para un estudio clínico se va a recabar información retrospectiva de expedientes médicos, debe considerarse que un porcentaje de estos no tendrán todos los datos necesarios, siendo imperante agregar un número extra de casos para prevenir esta eventualidad. De la misma manera, si un ensayo clínico aleatorizado amerita un seguimiento amplio, se deberá incluir un número mayor de pacientes para anteponerse a una eventual pérdida de sujetos durante el período de estudio.

En la página <http://www.sample-size.net> de la Universidad de San Francisco, se encuentra una herramienta virtual para el cálculo de los tamaños muestrales, de acuerdo a lo estudiado en este apartado.

CONSIDERACIONES BIOÉTICAS

Las investigaciones con seres humanos exigen, además del rigor científico, el compromiso de los investigadores de cumplir con principios éticos y con regulaciones internacionales y nacionales, que pretenden evitar cualquier daño a la salud o integridad de los participantes.

Parte de la historia reciente de la humanidad está llena de violaciones a los participantes de ensayos clínicos, con el pretexto de generar datos clínicos valiosos para la comprensión y el tratamiento de ciertas enfermedades. Por ejemplo, durante la Segunda Guerra Mundial se generaron políticas de investigación en seres humanos vivos, a quienes se privaba de sus derechos fundamentales. Posterior a este conflicto bélico, el Código de Nüremberg estableció un modelo ético para la experimentación con seres humanos, en el que se buscó la protección de la autonomía de los participantes, por medio de un proceso de consentimiento informado. Sin embargo, posterior a esta declaratoria, continuaron ocurriendo casos serios, en los cuales se violentaba la integridad de los individuos participantes en un ensayo clínico. Por ejemplo, en Willowbrook, Estados Unidos, durante la década de 1960, se realizaron estudios para la prevención de hepatitis en sujetos vulnerables. Para la década de 1970 salió a la luz pública el caso de un estudio realizado en Tuskegee, Alabama, en la que una muestra de individuos con sífilis y de etnia negra no fueron tratados con antibióticos para observar el curso de la enfermedad. Ante estos y otros casos similares, surgieron diversas iniciativas encaminadas a garantizar el bienestar y el respeto de los participantes en un ensayo clínico; entre ellas sobresale el informe Belmont, publicado en 1978, en el cual se establecieron los principios éticos fundamentales para emplear seres humanos en un estudio clínico, a saber: respeto a las personas; beneficencia; no maleficencia; y justicia. Estos principios éticos se resumen en el Cuadro N° 11.

Otra de las iniciativas que surgieron en la segunda mitad del siglo XX, fue la Declaración de Helsinki, promulgada por la Asociación Médica Mundial por primera vez en 1964 y actualizada recientemente. Este documento establece nuevos principios éticos para la investigación ética en seres humanos, entre ellos la supremacía de los derechos e intereses del participante y el deber del investigador en proteger la vida, la salud, la dignidad, la integridad, la autodeterminación, la intimidad y la confidencialidad de las personas que participen

en la investigación. Según estos principios, las investigaciones deben ser realizadas por personal cualificado, con la formación ética y científica adecuada, y cada investigador debe reducir al mínimo los riesgos derivados de la participación en un ensayo clínico. La declaración de Helsinki también establece que toda investigación médica en seres humanos debe conformarse con principios científicos y éticos, los cuales deberán describirse con detalle en un protocolo de investigación. Cada protocolo será revisado por un Comité de Ética en Investigación para su consideración, comentario, aportes y aprobación previo al inicio del estudio.

Cuadro N° 11. Principios éticos fundamentales en la investigación con seres humanos	
Principio ético	Compromiso del investigador
Respeto a las personas	Los investigadores deben obtener el proceso de consentimiento informado libre y voluntario de los participantes.
Beneficencia	El diseño del estudio debe estar orientado a generar un beneficio directo al paciente o al conocimiento científico, de forma tal que se justifique su realización.
No maleficencia	El investigador debe asegurar que durante el desarrollo del estudio los participantes no sean objeto de ningún daño o perjuicio como consecuencia de su participación.
Justicia	La selección de los pacientes no debe basarse en condiciones de vulnerabilidad o discriminación. Del mismo modo, los aportes originados del estudio deben ser distribuidos con equidad.

De igual forma, los Principios de Buenas Prácticas Clínicas, que se muestran a continuación, definen también una serie de pautas para que los estudios clínicos en seres humanos sean diseñados de tal manera que se garanticen los derechos de los participantes:

Los ensayos clínicos solo deben realizarse si los beneficios previstos para cada sujeto y la sociedad superan los riesgos.

Aunque los beneficios de un ensayo clínico sean importantes, la consideración fundamental es el derecho, la seguridad y el bienestar de los participantes.

Todo estudio clínico se realizará de acuerdo a un protocolo sólido desde el punto de vista científico, y aprobado por un Comité de Ética en Investigación.

La aprobación de ensayos clínicos de productos experimentales estará respaldada por información preclínica e información clínica adecuada.

Se obtendrá de cada sujeto su autorización con conocimiento de causa previa a la participación de ensayos clínicos.

El estudio será llevado a cabo por un profesional en virtud de su educación, formación y experiencia.

El registro, la gestión y almacenamiento de la información derivada del estudio clínico debe ser adecuada para que la notificación, interpretación y verificación del ensayo sean precisas.

Se protegerá la confidencialidad de los participantes.

Los productos experimentales se fabricarán y almacenarán conforme a las buenas prácticas en fabricación.

Se implementarán sistemas con procedimientos que garanticen la calidad de cada aspecto del ensayo.

En Costa Rica, la Ley N° 9234 “Ley Reguladora de la Investigación Biomédica”, promulgada en el año 2014, establece la regulación de la investigación biomédica con seres humanos en materia de salud y promulga sanciones en caso de su infracción. Además, mediante esta ley se creó el Consejo Nacional de Investigaciones en Salud, como un órgano independiente, multidisciplinario, de carácter ético, técnico y científico, adscrito al Ministerio de Salud, cuya función primordial es garantizar la calidad de las investigaciones y su apego a los derechos humanos. Este ente se encarga, entre otras cosas, de la acreditación de los Comités Ético Científicos, de investigadores y de organizaciones de investigación o administración por contrato.

Además, existen en el país otras regulaciones nacionales y de la Caja Costarricense de Seguro Social (CCSS), las cuales se encuentran disponibles en la dirección electrónica <http://www.cendeiss.sa.cr/wp/index.php/regulaciones/>, en el apartado de Bioética.

Todas estas normas nacionales y principios éticos deben cumplirse fielmente, con el fin primordial de garantizar los derechos de los participantes.

BIBLIOGRAFÍA

1. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. *Lancet*. 1998; 2(8607): 349-360.
2. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ* 1995; 310(6973): 170.
3. Ridker PM, Danielson E, Fonseca FA, Genest J, Gotto AM Jr, Kastelein JJ, Koenig W, Libby P, Lorenzatti AJ, MacFadyen JG, Nordestgaard BG, Shepherd J, Willerson JT, Glynn RJ. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *N Engl J Med*. 2008; 359(21): 2195-2207.

4. Tardif JC, McMurray JJ, Klug E, Small R, Schumi J, Choi J, Cooper J, Scott R, Lewis EF, L'Allier PL, Pfeffer MA. Effects of succinobucol after an acute coronary syndrome: a randomised, double-blind, placebo-controlled trial. *Lancet*. 2008; 371(9626): 1761-1768.
5. CAPRIE Steering Committee. A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events. *Lancet*. 1996; 348(9038): 1329-1339.
6. McMurray J, Packer M, Desai AS, Gong J, Lefkowitz MP, Rizkala AR, Rouleau JL, Shi VC, Solomon SD, Swedberg K, Zile MR. Angiotensin-Nepriylsin inhibition versus enalapril in heart failure. *N Engl J Med*. 2014; 371(11): 993-1004.
7. Altman DG, Bland JM. Time to event (survival) data. *BMJ*. 1998; 317:468-469.
8. Asociación Médica Mundial. *Declaración de Helsinki de la AMM – Principios Éticos para las Investigaciones Médicas en Seres Humanos*. (Internet). Asociación Médica Mundial; 2017. Consultado el 19 de agosto de 2018, en: <https://www.wma.net/es/polices-post/declaracion-de-helsinki-de-la-amm-principios-eticos-para-las-investigaciones-medicas-en-seres-humanos/>
9. Bagley SC, White H, Golomb BA. Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. *J Clin Epidemiol*. 2001; 54(10): 979-985.
10. Barratt A, Wyer PC, Hatala R, McGinn T, Dans AL, Keitz S, Moyer V, For GG. Tips for learners of evidence-based medicine: 1. Relative risk reduction, absolute risk reduction and number needed to treat. *CMAJ*. 2004; 171(4): 353-358.
11. Bassler D, Montori VM, Briel M, Glasziou P, Guyatt G. Early stopping of randomized clinical trials for overt efficacy is problematic. *J Clin Epidemiol*. 2008; 61(3): 241-246.
12. Bender R. Introduction to the use of regression models in epidemiology. *Methods Mol Biol*. 2009; 471: 179-195.
13. Bland JM, Altman DG. Survival probabilities (the Kaplan-Meier method). *BMJ*. 1998; 317(7172): 1572.
14. Bland JM, Altman DG. The logrank test. *BMJ*. 2004; 328(7447): 1073.
15. Cook D, Sackett D. The number needed to treat: a clinically useful measure of treatment effect. *BMJ*. 1995; 310(6977): 452-454.
16. Cox DR. Regression models and life-tables. *J R Statist Soc B*. 1972; 34(2): 187-220.
17. Cox DR. Statistical significance tests. *Br J Clin Pharmacol*. 1982; 14(3): 325-331.
18. Currant-Everett D. Exploration in statistics: confidence intervals. *Adv Physiol Educ*. 2009; 33(2): 87-90.

19. DeMets DL. Statistical issues in interpreting clinical trials. *J Intern Med.* 2004; 255(5): 529-537.
20. du Prel JB, Hommel G, Röhrig B, Blettner M. Confidence Interval or P-Value? *Dtsch Arztebl Int.* 2009; 106(19): 335–339.
21. du Prel JB, Röhrig B, Hommel G, Blettner M. Choosing statistical tests: part 12 of a series on evaluation of scientific publications. *Dtsch Arztebl Int.* 2010; 107(19): 343–348.
22. Emanuel EJ, Wendler D, Grady C. What makes clinical research ethical? *JAMA.* 2000; 283(20):2701-2711.
23. Faguet GB, Davis HC. Regression analysis in medical research. *South Med J.* 1984; 77(6): 722-725.
24. Field AP, Gillett R. How to do a meta-analysis. *Br J Math Stat Psychol.* 2010; 63(Pt 3):665-694.
25. Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA.* 2003; 289(19): 2554-2559.
26. Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients' interests. *Lancet.* 2007; 370(9602): 1875-1877.
27. Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *Br Med J (Clin Res Ed).* 1986; 292(6522):746-750.
28. Gelman A, Stern HS. The difference between “significant” and “not significant” is not itself statistically significant. *Am Stat.* 2006; 60(4): 328-331.
29. Gelman A. P values and statistical practice. *Epidemiology.* 2013; 24(1): 69-72.
30. Goodman SN. Towards evidence-based medical statistics, I: the P-value fallacy. *Ann Intern Med.* 1999; 130(12): 995-1004.
31. Gordillo A, Medina UF, Pierdant M. *Manual de investigación clínica.* México: Editorial El Manual Moderno; 2012.
32. Greenland S, Senn SJ, Rothman KJ, Carlin J, Poole C, Goodman S. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol.* 2016; 31(4): 337-350.
33. Grieve AP. How to test hypotheses if you must. *Pharm Stat.* 2015; 14(2): 139-150.
34. Haybittle JL. Repeated assessment of results in clinical trials of cancer treatment. *Br J Radiol.* 1971; 44(526): 793-797.
35. Head SJ, Kaul S, Bogers AJ, Kappetein AP. Non-inferiority study design: lessons to be learned from cardiovascular trials. *Eur Heart J.* 2012; 33(11): 1318-1324.
36. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analysis. *BMJ.* 2003; 327(7414): 557-560.

37. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002; 21(11): 1539-1558.
38. Hung HM, Wang SJ, O'Neill R. A regulatory perspective on choice of margin and statistical inference in non-inferiority trials. *Biom J.* 2005; 47(1):28-36.
39. Ley N° 9234 Ley Reguladora de Investigación Biomédica. San José, Costa Rica: *Diario Oficial La Gaceta* N° 79 (25 de abril de 2014).
40. Kaul S, Diamond GA. Good enough: a primer on the analysis and interpretation of noninferiority trials. *Ann Intern Med.* 2006; 145(1): 62-69.
41. Khoshdel A, Attia J, Carney SL. Basic concepts in meta-analysis: a primer for clinicians. *In J Clin Pract.* 2006; 60(10): 1287-1294.
42. Kip KE, Hollabaugh K, Marroquin OC, Williams DO. The problem with composite end points in cardiovascular studies: The story of major adverse cardiac events and percutaneous coronary intervention. *J Am Coll Cardiol.* 2008; 51(7): 701-707.
43. Kleinbaum DG, Klein M. *Survival analysis. A self-learning text.* 3 ed. New York: Springer; 2012.
44. Land CE. Statistical limitations in relation to sample size. *Environ Health Perspect.* 1981; 42: 15-21.
45. Lang JM, Rothman KJ, Cann CI. That confounded P-value. *Epidemiology.* 1998; 9(1): 7-8.
46. Le Henanff A, Giraudeau B, Baron G, Ravaud P. Quality of reporting of noninferiority and equivalence randomized trials. *JAMA.* 2006; 295(10): 1147-1151.
47. Liberati A, Altman D, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLOS Medicine.* 2009; 6(7): e1000100.
48. Lubsen J, Hoes A, Grobbee D. Implications of trial results: the potentially misleading notions of number needed to treat and average duration of life gained. *Lancet.* 2000; 356(9243): 1757-1759.
49. Mantovani V, Lepore V, Mira A, Berglin E. Non-inferiority randomized trials, an issue between science and ethics: the case of the SINTAX Study. *Scand Cardiovasc J.* 2010; 44(6): 321-324.
50. Matthews J, Altman DG. Interaction 2: Compare effect sizes not P values. *Br Med J.* 1996; 313: 808.
51. Mendes D, Alves C, Batel-Marques F. Number needed to treat (NNT) in clinical literature: an appraisal. *BMC Medicine.* 2017; 15(1): 112.
52. Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, Lacchetti C, Leung TW, Darling E, Bryant DM, Bucher HC, Schünemann HJ, Meade MO, Cook DJ, Erwin PJ, Sood A, Sood R,

- Lo B, Thompson CA, Zhou Q, Mills E, Guyatt GH. Randomized trials stopped early for benefit: a systematic review. *JAMA*. 2005; 294(17): 2203-2209.
53. Morey RD, Hoekstra R, Rouder JN, Lee MD, Wagenmakers EJ. The fallacy of placing confidence in confidence intervals. *Psychon Bull Rev*. 2016; 23(1): 103-123.
54. Moyé LA. *Multiple analyses in clinical trials: fundamentals for Investigators (Statistics for Biology and Health)*. New York: Springer; 2003.
55. Mueller PS, Montori VM, Bassler D, Koenig BA, Guyatt GH. Ethical issues in stopping randomized trials early because of apparent benefit. *Ann Intern Med*. 2007; 146(12): 878-881.
56. Mulder PGH. Time-dependency in Cox's regression analysis. *J Chronic Dis*. 1987; 40(4): 365-366.
57. Nick TG, Campbell KM. Logistic regression. *Methods Mol Biol*. 2007; 404: 273-301.
58. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*. 1979; 35(3): 549-556.
59. Organización Panamericana de la Salud (OPS). *Buenas prácticas clínicas: documento de las Américas*. (Internet). OPS; 2005. Consultado el 19 de agosto de 2018, en: <http://apps.who.int/medicinedocs/documents/s18627es/s18627es.pdf>
60. Peacock JL, Peacock PJ. *Oxford handbook of medical statistics*. New York: Oxford University Press; 2011.
61. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG. Design and analysis of randomized clinical trials requiring prolonged observations of each patient. I. Introduction and design. *Br J Cancer*. 1976; 34(6): 585-612.
62. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA*. 2006; 295(10): 1152-1160.
63. Pocock S, White I. Trials stopped early: too good to be true? *Lancet*. 1999; 353: 943-944.
64. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Eng J Med*. 1987; 317(7): 426-432.
65. Pocock SJ, McMurray JJV, Collier TJ. Statistical controversies in reporting of clinical trials: Part 2 of a 4-Part series on statistics for clinical trials. *J Am Coll Cardiol*. 2015; 66(23): 2648-2662.
66. Pocock SJ. Current controversies in data monitoring for clinical trials. *Clin Trials*. 2006; 3(6): 513-521.

67. Ramos-Esquivel A. Overweight, obesity and all-cause mortality. *JAMA*. 2013; 309(16): 1680.
68. Rebaso P. Conceptos básicos del análisis de supervivencia. *Cir Esp*. 2005; 78(4): 222-230.
69. Schneider A, Hommel G, Blettner M. Linear regression analysis. *Dtsch Arztebl Int*. 2010; 107(44): 776-782.
70. Schulz KF, Grimes DA. Multiplicity in randomised trials. II: Subgroup and interim analyses. *Lancet*. 2005; 365(9471): 1657-1661.
71. Schumi J, Wittes JT. Through the looking glass: understanding non-inferiority. *Trials*. 2011; 12: 106.
72. Sim J, Reid N. Statistical inference by confidence intervals: issues of interpretation and utilization. *Phys Ther*. 1999; 79(2): 186-195.
73. Sinclair JC, Cook RJ, Guyatt GH, Pauker SG, Cook DJ. When should an effective treatment be used? Derivation of the threshold number needed to treat and the minimum event rate for treatment. *J Clin Epidemiol*. 2001; 54(3): 253-262.
74. Smith GD, Egger M, Phillips A. Meta-analysis: beyond the grand mean. *BMJ*. 1997; 315(7122): 1610-1614.
75. Snappin SM. Alternatives for discounting in the analysis of noninferiority trials. *J Biopharm Stat*. 2004; 14(2): 263-273.
76. Sterne JA, Davey Smith G. Sifting the evidence—what's wrong with significance tests? *BMJ*. 2001; 322(7280): 226–231.
77. Steyeberg EW. *Clinical prediction models. A practical approach to development, validation, and updating*. New York: Springer; 2009.
78. Tomlinson G, Detsky AS. Composite end points in randomized trials: there is no free lunch. *JAMA*. 2010; 303(3): 267-268.
79. Weinberg CR. It's time to rehabilitate the P-value. *Epidemiology*. 2001; 12(3): 288-290.
80. Zwiener I, Blettner M, Hommel G. Survival analysis: part 15 of a series on evaluation of scientific publications. *Dtsch Arztebl Int*. 2011; 108(10): 163-169.

ANEXOS

ANEXO 1. ¿CÓMO BUSCAR LA MEJOR EVIDENCIA CLÍNICA?

EL ACRÓNIMO PICOT

El esquema PICOT corresponde a un acrónimo que hace referencia a cinco componentes de cualquier cuestión de etiología, diagnóstico, tratamiento o pronóstico (Figura N° 1):

- **P:** población.
- **I:** intervención.
- **C:** comparador.
- **O:** objetivo, evento de interés u “*outcome*”.
- **T:** tiempo.

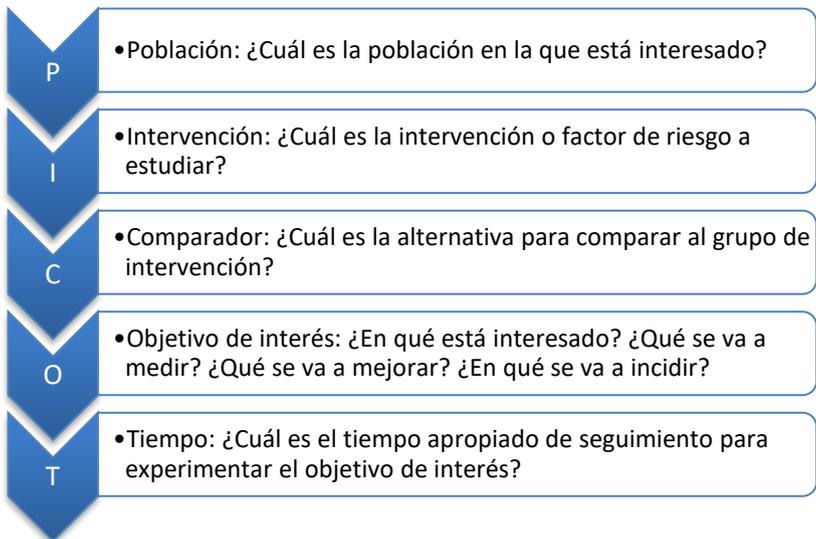


Figura N° 1. Esquema PICOT de la pregunta de investigación.

Cada uno de estos elementos posibilitará un acceso adecuado a la búsqueda bibliográfica y permitirá fácilmente ubicar el contexto específico del cual se desea obtener información. Además, se recomienda emplear estos cinco componentes al momento de plantear una pregunta de investigación para un ensayo clínico.

A continuación se presentan algunos ejemplos:

Ejemplo 1: Pregunta de etiología

Suponiendo que se desea determinar si la exposición al virus del Zika produce microcefalia en una población de mujeres embarazadas, en este caso el acrónimo PICOT correspondería a:

P: población de mujeres embarazadas.

I: exposición al virus del Zika.

C: no exposición al virus.

O: casos de microcefalia.

T: tiempo de exposición (nueve meses).

Ejemplo 2: Pregunta de diagnóstico

Suponiendo que se desea saber si una tomografía por emisión de positrones (PET) es superior a la resonancia magnética en el diagnóstico de metástasis a sistema nervioso central en pacientes con cáncer de pulmón, el esquema PICOT sería:

P: pacientes con cáncer de pulmón.

I: tomografía por emisión de positrones.

C: resonancia magnética.

O: detección de metástasis en sistema nervioso central.

T: en este caso el tiempo no es relevante. De hecho, los estudios transversales en los cuales el tiempo no es una variable determinante, son los empleados para responder la interrogante: ¿cuál es el mejor diagnóstico?

Ejemplo 3: Pregunta de tratamiento

Quizá esta pregunta es la más desarrollada en la práctica clínica diaria. Por ejemplo, si se quiere saber si el clopidogrel es superior a la aspirina en la prevención de infartos de miocardio en sujetos con cardiopatía isquémica previa cuando son tratados por al menos cinco años, la pregunta PICOT sería:

P: pacientes con cardiopatía isquémica.

I: clopidogrel.

C: aspirina.

O: infartos de miocardio.

T: cinco años.

Ejemplo 4: Pregunta de pronóstico

Una pregunta de pronóstico sería: ¿cuál es el riesgo de desarrollar epilepsia en un niño sano 10 años después de una convulsión febril? Utilizando el formato PICOT, esta pregunta tendría los siguientes elementos:

P: niños sanos.

I: convulsión febril.

C: no convulsión febril.

O: diagnóstico de epilepsia.

T: 10 años.

BÚSQUEDA DE INFORMACIÓN

Como se describió en los ejemplos anteriores, el uso del acrónimo PICOT es una herramienta útil para contextualizar y especificar cualquiera de las interrogantes clínicas básicas. Su utilidad radica en facilitar la búsqueda de información científica para intentar responder adecuadamente cada pregunta. Sin embargo, para que la búsqueda de información sea completa, es necesario acceder a bases de datos lo suficientemente exhaustivas. En el campo de la salud existen un sinnúmero de buscadores que realizan este proceso, mediante el uso de

palabras clave o descriptores, los cuales despliegan una serie de artículos referentes al tema. Dentro de las bases de datos preferidas por varios autores se incluyen: *PubMed*, *EMBASE*, *Cochrane Central Register of Clinical Trials* (CENTRAL), *Google Scholar*, entre otras. Si bien no son las únicas, estas engloban la mayoría de artículos científicos publicados en revistas de alta calidad, donde cada trabajo incluido ha pasado por un proceso de revisión por pares de alta confiabilidad, cumpliendo estrictos procesos de selección y rigurosidad científica.

Entre todas las bases de datos existentes, la de *PubMed* es quizás la más rigurosa y exhaustiva, y brinda resultados actualizados del tema de interés de forma eficiente. Accesando a través de www.pubmed.com, se encuentra una barra de búsqueda, en la cual se introducen las palabras claves de interés. Es conveniente que estas palabras claves o *Keywords* provengan de una fuente oficial, tal como MeSH (*Medical Subject Heading*).

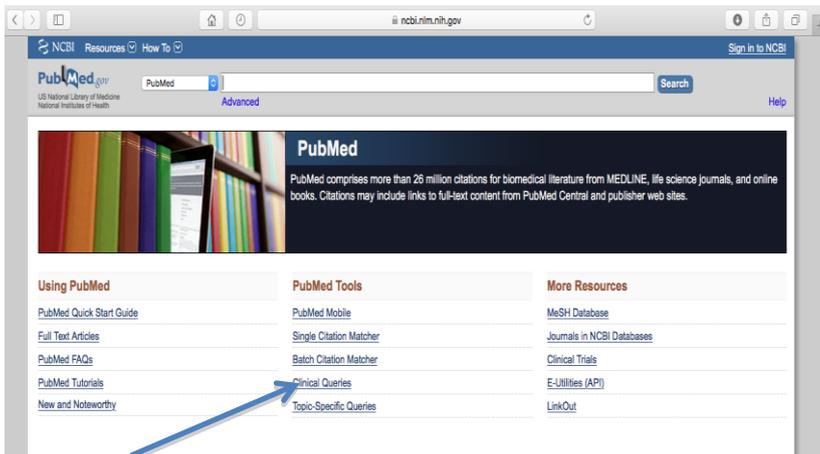


Figura N° 2. Pantalla de inicio de PubMed.gov (la flecha señala la opción de *Clinical Queries*, que facilita la búsqueda según la pregunta de investigación).

PubMed Clinical Queries

Results of searches on this page are limited to specific clinical research areas. For comprehensive searches, use [PubMed](#) directly.

diabetes mellitus Search

Clinical Etiology
Diagnosis
Category / Therapy
Prognosis
Scope Clinical prediction guides

Systematic Reviews

Medical Genetics
Topic: All

Figura N° 3. Pantalla de *PubMed* en la pestaña *Clinical Queries*, bajo la búsqueda de diabetes mellitus.

Esta fuente estándar posibilitará encontrar la palabra exacta o un sinónimo de esta (empleando el conector OR). Como se mencionó, tales descriptores deben ser digitados en el espacio de búsqueda. En este espacio cada componente del acrónimo PICOT será enlazado mediante el conector “AND”.

Así entonces, la estructura básica de una búsqueda sería:

(Población OR algún sinónimo) AND (Intervención OR algún sinónimo) AND (comparador OR algún sinónimo) AND (objetivo OR algún sinónimo).

Por ejemplo, si se desea encontrar información sobre la relación de infección del virus del Zika en mujeres embarazadas y el desarrollo de microcefalia, se formularía la búsqueda con los descriptores: pregnant women OR pregnancy AND Zika AND microcephaly.

The screenshot shows the PubMed search results for the query "(pregnant women OR pregnancy) AND (Zika) AND (microcephaly)". The search results are displayed in a list format, with the first result highlighted. The search interface includes filters for article types, text availability, and publication dates. The search results list includes titles, authors, journals, and publication dates.

Search results
Items: 1 to 20 of 253

1. [Prevalence and Clinical Attributes of Congenital Microcephaly - New York, 2013-2015.](#)
Graham KA, Fox DJ, Taliati A, Pantea C, Brady L, Carter SL, Friedenberg E, Vora NM, Browne M, Lee CT.
MMWR Morb Mortal Wkly Rep. 2017 Feb 10;66(5):125-129. doi: 10.15585/mmwr.mm6605a1.
PMID: 28182608 [Free Article](#)
[Similar articles](#)

2. [The Antigenic Structure of Zika Virus and Its Relation to Other Flaviviruses: Implications for Infection and Immunoprophylaxis.](#)
Heinz FX, Stiasny K.
Microbiol Mol Biol Rev. 2017 Feb 8;81(1). pii: e00055-16. doi: 10.1128/MMBR.00055-16. Review.
PMID: 28179396 [Free Article](#)
[Similar articles](#)

3. [High incidence of Zika virus infection detected in plasma and cervical cytology specimens from pregnant women in Guayaquil, Ecuador.](#)
Zambrano H, Waggoner J, León K, Pinsky B, Vera K, Schettino M, Rivera L, Landivar J, Granda M, Lee A, Mor G.
Am J Reprod Immunol. 2017 Feb;77(2). doi: 10.1111/aj.12630.
PMID: 28177195
[Similar articles](#)

Titles with your search terms
Zika Virus Infection in Pregnancy, Microcephaly, a [Arch Pathol Lab Med. 2017]
Estimating the Number of Pregnant Women Infected With Zika Virus [JAMA Pediatr. 2016]
Associated ultrasonographic findings in fetuses with microceph [Prenat Diagn. 2016]
[See more...](#)

Find related data
Database: [Select](#)
[Find items](#)

Search details
{("pregnant women"[MeSH Terms] OR ("pregnant"[All Fields] AND "women"[All Fields]) OR "pregnant women"[All Fields]) OR ("pregnancy"[MeSH Terms] OR

Figura N° 4. Resultado de la búsqueda de *PubMed* para el ejemplo 1.

BIBLIOGRAFÍA

1. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA*. 1992; 268(17): 2420–2425.
2. Helmer D, Savoie I, Green C, Kazanjian A. Evidence-based practice: extending the search to find material for the systematic review. *Bull Med Libr Assoc*. 2001; 89(4): 346-352.
3. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0*. The Cochrane Collaboration; 2011.
4. Shaw RL, Booth A, Sutton AJ, Miller T, Smith JA, Young B, Jones DR, Dixon-Woods M. Finding qualitative research: an evaluation of search strategies. *BMC Med Res Methodol*. 2004; 4: 5-10.

ANEXO 2. DEFINICIÓN DE CONCEPTOS CLAVES

A continuación se resumen algunos elementos claves tratados a lo largo de este libro:

- **Análisis ajustado:** se refiere a un tipo de regresión, en el cual se modifica la relación de la variable independiente con la dependiente, para determinar el efecto de las covariables que pueden afectar también la variable dependiente. Por ejemplo, la relación entre la mortalidad (variable dependiente) y la edad (variable independiente) puede ser ajustada por la covariable sexo cuando se cree que esta condición también afecta la mortalidad.
- **Análisis de intención a tratar:** método de análisis de los datos, basado en la intención original de tratamiento que a cada sujeto le fue asignado en el proceso de aleatorización, aun cuando dicho individuo no hubiera recibido dicha intervención. Por consiguiente, los sujetos asignados a recibir una terapia específica son evaluados en conjunto de acuerdo a esta asignación inicial, independientemente de la adherencia al tratamiento o al protocolo del estudio.
- **Análisis por protocolo:** método de análisis de los datos en el cual solo se analizan los sujetos participantes que cumplieron con el protocolo del ensayo y que estuvieron expuestos al tratamiento o intervención a lo largo de todo el período de seguimiento establecido.
- **Coefficiente de correlación:** medida de la asociación lineal entre dos variables continuas. El coeficiente de correlación varía entre los valores de -1 y +1. Entre más cercano se encuentre al cero, menor es la asociación. Si ambas variables se modifican en la misma dirección, el coeficiente obtiene un valor positivo, en tanto que los valores negativos reflejan una asociación inversa.

- **Covariables:** término que se utiliza para las variables predictivas o explicativas de un modelo de regresión que no suelen ser la variable de interés primario del investigador. Las covariables se miden al inicio de los ensayos clínicos, porque se cree que pueden influir en el objetivo final.
- **Error aleatorio:** consiste en una desviación no predecible del estimado calculado en la muestra, como consecuencia de la variabilidad de esta, y hace referencia a la probabilidad de realizar conclusiones inexactas en el proceso de estimación estadística. Comprende el error tipo I o alfa, que se refiere a la probabilidad de rechazar la hipótesis nula cuando esta es cierta, y el error tipo II o beta, que consiste en la probabilidad de no rechazar la hipótesis nula cuando esta es falsa. Se logra reducir el error aleatorio aumentando el tamaño de la muestra.
- **Error estándar:** medida de la variabilidad aleatoria de un estadístico (media, proporción, tasa de incidencia) que indica qué tan lejano está el estimado del verdadero valor poblacional (parámetro). Un error estándar pequeño implica un estimado más preciso del verdadero parámetro. Es imprescindible para el cálculo del intervalo de confianza y está inversamente relacionado con el tamaño de la muestra.
- **Error sistemático:** corresponde a las desviaciones del estimado de su verdadero valor poblacional, como consecuencia de un inadecuado diseño, selección, conducción, análisis o publicación de un estudio. Se agrupan en sesgos de selección, de información y confusión.
- **Hipótesis alternativa:** hipótesis estadística en la cual se establecen diferencias entre dos o más parámetros.
- **Hipótesis nula:** hipótesis estadística en la que se establece la “no asociación” o “no diferencia” entre dos o más parámetros.

- **Intervalo de confianza:** rango de valores dentro de los cuales se encuentra con cierta probabilidad el verdadero parámetro de la población. Usualmente, se emplea el intervalo de confianza del 95 %.
- **Poder estadístico:** corresponde a la probabilidad de rechazar la hipótesis nula cuando esta es falsa. Es la base para el cálculo del tamaño de la muestra requerido para detectar diferencias de una magnitud determinada por el investigador.
- **Prueba de hipótesis:** procedimiento estadístico para determinar si las diferencias observadas en un estudio son diferencias reales o sujetas a errores aleatorios; para ello se utiliza el cálculo del valor de p .
- **Valor de p :** probabilidad de que los datos obtenidos sean al menos tan extremos como hubiera ocurrido si la hipótesis nula fuera cierta.

