Revista Latinoamericana de Metodología de la Investigación Social.
Nº29. Año 15. Abril - Septiembre 2025. Argentina. ISSN 1853-6190. Pp. 7-21.

Delimitación del tamaño muestral en estudios piloto

Sample size delimitation in pilot studies

Rodrigo Vargas Salomón

Resumen

Una idea bastante difundida en las investigaciones, principalmente cuantitativas, es que las muestras deben ser muy amplias, asumiendo que, a mayor tamaño, menores posibilidades de error en los datos, y mejor representatividad estadística de los análisis inferenciales. El objetivo de este trabajo es enfatizar que este principio no es aplicable a todas las investigaciones, pues en algunas como los estudios piloto, el tamaño muestral, además de la representatividad estadística, requiere considerar otros criterios que pueden ser relevantes para un estudio en particular. Por ello, se analizan dos reglas generales al momento de determinar el tamaño muestral en los estudios piloto: las que establecen un número fijo de participantes; y las que sugieren delimitar el tamaño muestral a través de una serie de pasos, diseñados con base en las características propias de la investigación a realizar. Al final, ninguna de estas reglas es mejor que la otra, pues el tamaño muestral deberá ser una decisión ampliamente razonada y crítica por parte del o lo(a)s investigadore(a)s, en relación con las características y fines de su investigación.

Palabras clave: pruebas piloto; estudios piloto; tamaño muestral; muestreo plano; muestreo escalonado.

Abstract

An idea that is quite widespread in research, mainly quantitative, is that the samples must be very large, assuming that the larger the size lower the chances of error in the data, and a better statistical representativeness of the inferential analyses. The objective of this work is to emphasize that this principle does not apply to all research since in some cases such as pilot studies, the sample size, in addition to statistical representativeness, requires considering other criteria that may be relevant for a particular study. Therefore, two general rules are analyzed when determining the sample size in pilot studies: those that establish a fixed number of participants; and those that suggest delimiting the sample size through a series of steps, designed based on the characteristics of the research to be carried out. After all, none of these rules is better than the other, because the size of a sample must be a widely reasoned and critical decision made by the researcher(s), related to the characteristics and purposes of their research.

Keywords: pretests; pilot studies; sample size; flat rules; stepped rules.

Introducción

Una perspectiva clásica en investigación, aunque no es la única, es que el muestreo debe utilizarse cuando se va a someter hipótesis a contrastación empírica, la cual se puede probar en un número reducido de datos relevantes para la propia hipótesis, elegidos sin tendencia ni criterio a partir de una infinitud potencial de datos (Bunge, 1983), que conformen la totalidad del estudio. No obstante, delimitar el tamaño de una muestra para una investigación no es tarea sencilla, y es probablemente uno de los primeros grandes problemas a los que se enfrentan lo(a)s investigadore(a)s al desarrollar un proyecto, especialmente los nóveles, quienes al tener que definir cuántas personas o casos deben incluir en su investigación para que ésta sea representativa del total de la población, suelen encontrarse con una serie de fórmulas y procedimientos complejos para delimitar el tamaño muestral. También suelen toparse con datos que no siempre tienen al alcance o que desconocen cómo calcularlos, como el de población finita, los errores alfa (tipo I) y beta (tipo II), el margen de error absoluto permitido y los intervalos de confianza esperados, las puntuaciones z, la probabilidad de ocurrencia, el poder estadístico (potencia), el tamaño del efecto estandarizado, entre otros.

No obstante, la dificultad que pueda representar este cálculo, se convierte en una actividad primordial para la mayoría de las investigaciones, que pretenden utilizar estadísticas paramétricas, principalmente por los motivos que se resumen en la siguiente cita:

> El tamaño de muestra permite a los investigadores saber cuántos individuos son necesarios estudiar, para poder estimar un parámetro determinado con el grado de confianza deseado, o el número necesario para poder detectar una determinada diferencia entre los grupos de estudio, suponiendo que existiese realmente. El cálculo del tamaño de la muestra es una función matemática que expresa la relación entre las variables, cantidad de participantes y poder estadístico (García-García *et al*., 2013: 217).

Sin embargo, las opciones para calcular el tamaño muestral adecuado para una investigación son muy diversas. Actualmente existen incluso en internet calculadoras de tamaños de muestra gratuitas, que utilizan precisamente fórmulas estadísticas ya establecidas para tal efecto, principalmente la denominada "fórmula de Cochran"[1], que con algunas variantes, es la más utilizada para el cálculo de tamaños de muestras proporcionales y en términos porcentuales. La diferencia es que estas calculadoras en línea realizan los cálculos de manera automática, y no de forma manual como tradicionalmente se hacía, además de que facilitan la realización de cálculos previos inherentes a las fórmulas, como el valor de la puntuación z o probabilidad de ocurrencia, con lo que minimizan el número de datos requeridos para calcular el tamaño de una muestra.

Un ejemplo de este tipo de calculadoras es la del sitio web QuestionPro (2023), donde con tres sencillos pasos se puede calcular el tamaño muestral "de forma rápida y efectiva":

1. Señalar el nivel de confianza (que puede ser 95% o 99%, aunque el sitio sugiere el 95%);
2. Seleccionar el margen de error (donde se sugiere 5% o menos); y
3. Asignar el total de la población (el cual dependerá del dato de población finita que se tenga para la investigación).

Ingresando esta información, lo único que resta hacer es dar *clic* en una casilla denominada "calcular muestra", y se obtiene un número entero como resultado, que corresponde al tamaño muestral sugerido. Otros sitios similares son el de SurveyMonkey (2023) y el de Qualtrics (2023), en los cuales se piden exactamente los mismos tres datos para calcular, con la misma facilidad, el tamaño de una muestra. Cabe resaltar que el resto de los valores requeridos para aplicar la fórmula de Cochran, estos sitios los tienen predeterminados, por lo que no se solicitan a los usuarios.

---

[1] Existen varias versiones de esta fórmula, pero en general utiliza los mismos parámetros en su planteamiento cuando se trata de poblaciones finitas, y aunque se le conoce así porque está basada en las aportaciones de William G. Cochran, lo cierto es que en su obra principal *Sampling Techniques*, publicada originalmente en 1953, no establece directamente la fórmula que hoy se utiliza ampliamente, sino que realiza diversas propuestas dependiendo del tipo de estudio o de muestreo que se pretende realizar (Cochran, 1977).

Además, existen aplicaciones como GRANMO, diseñada por Marrugat *et al*. (1998) hace más de 25 años, precisamente considerando lo complicado que es aplicar manualmente algoritmos para el cálculo de muestras. Desde sus inicios se trató de un software de libre difusión (*shareware*), que evolucionó hasta contar actualmente con una calculadora muy avanzada de uso gratuito en línea, que permite calcular tamaños muestrales según diversos tipos o diseños de investigación (medidas apareadas, equivalencias, con dos proporciones independientes, estudios de casos y controles, estudios de cohortes, entre otros). Además de los datos básicos ya señalados, GRANMO permite otros como las desviaciones estándar, los riesgos de error alfa y beta, proporciones de grupos, estimaciones poblacionales, etc., según el diseño elegido.

Otro ejemplo similar destacable es Epi Info, desarrollada a finales del pasado siglo por el Centro para el Control y Prevención de Enfermedades de Atlanta, Estados Unidos, y distribuida desde entonces de manera gratuita. Esta herramienta también evolucionó de un programa en MS-DOS, hasta contar actualmente con un software para computadoras y una aplicación para teléfonos inteligentes –ambos de uso libre– que también permiten calcular de manera automatizada tamaños de muestra con distintos diseños e indicadores estadísticos según se requiera CDC, 2022).

No obstante la existencia de estas herramientas en línea, u otras como los softwares estadísticos tipo SPSS, ECHANT, PASS, STATS, EPIDAT, etc., de los que se puede hacer uso en la actualidad, y que son muy efectivos porque se apegan a los procedimientos tradicionales de cálculos de tamaño muestral, lo cierto es que tanto en los sitios en línea como en los softwares, y actualmente incluso con el uso de sistemas de Inteligencia Artificial como el ChatGPT, Claude o Gemini, no se consideran características particulares que pueden tener algunos tipos o situaciones de investigación, en los cuales no es tan simple definir cuál es tamaño de muestra más conveniente.

En algunos casos, apegarse a una fórmula no es la mejor opción, menos aún si se toman en cuenta aspectos como los que atinadamente advierten García-García *et al*. (2013), para quienes calcular el tamaño de una muestra implica mucho más que realizar una simple operación aritmética que proporcione un valor, pues se trata más bien de una función matemática, donde cambiar el valor de una variable de la ecuación, afecta a otros valores, teniendo un efecto no sólo sobre el número obtenido como tamaño muestral, sino también sobre la potencia[2] de la misma, y por tanto, sobre la calidad de los datos que se obtengan.

Un ejemplo de investigaciones donde el tamaño muestral no es tan simple de definir mediante una fórmula matemática, son las investigaciones cualitativas, donde más que buscar la representatividad en función de un valor paramétrico e inferencial, lo que se busca es la representatividad teórica (no paramétrica), en función del tema o fenómeno de estudio. Esto hace que no sea tan relevante a cuántas personas incluir, sino más bien a quiénes, de modo que sean lo(a)s más idóneo(a)s para la investigación, generando un sesgo deliberado en la selección de informantes, al que se le suele denominar como muestreo teórico, intencional, crítico o por juicio. No obstante, si bien algunos estudios cualitativos podrían ser útiles como prueba piloto en estudios cuantitativos e híbridos, esta alternativa no es abordada en este artículo, dada la complejidad y espacio que requiere su discusión.

En el campo de las investigaciones cuantitativas también suele haber estudios donde tampoco es tan simple establecer un tamaño muestral idóneo. Al respecto, Kerlinger & Lee (2002) señalan que "un [sic] regla dura pero eficaz, que se enseña a estudiantes principiantes de investigación es: utilizar una muestra tan grande como sea posible" (p. 157). Según estos autores, esto no es sólo un asunto de que los números grandes sean mejores en sí mismos, sino que las muestras grandes permiten que realmente funcione el principio de aleatorización, para que sean realmente representativas de la población estudiada ya que, en las muestras pequeñas, la posibilidad de generar sesgos en la selección de casos o informantes es mayor, y con ello, también la posibilidad de errores en los análisis inferenciales. Por tanto, otro principio básico al respecto es

[2] Potencia o poder estadístico, es la probabilidad de que una prueba aplicada, identifique de forma correcta un resultado o efecto de interés para el estudio. Se vincula de manera directa con la sensibilidad de los instrumentos, la efectividad de los resultados, y a su vez, con la probabilidad de no cometer un error de tipo II o beta.

que, "a menor tamaño de la muestra, mayor será el error, y que, a mayor tamaño de la muestra, menor será el error resultante" (Kerlinger & Lee, 2002: 157).

No obstante, estos mismos autores señalan que esta regla de tener muestras grandes, no es necesariamente apropiada para todas las investigaciones, pues mientras en algunos casos puede ser muy simple obtenerlas, en otros no lo es, de modo que la determinación estadística del tamaño muestral, debe estimarse en relación con la situación particular de la investigación a realizar (Kerlinger & Lee, 2002). Además, las muestras grandes tampoco son sinónimo de ausencias de errores o sesgos en los resultados, en algunas investigaciones, "una muestra demasiado grande hará que una diferencia muy pequeña resulte estadísticamente significativa, sin tener necesariamente una significancia práctica" (Kerlinger & Lee, 2002: 266), aspecto en el que también coinciden Babbie (1990) y Stevens (2009). De tal modo que, aun usando fórmulas o procedimientos para simplificar el cálculo de un tamaño de muestra, esta tarea rara vez es algo sencillo.

Los estudios piloto

Un tipo de estudios que suelen presentar problemas al delimitar el tamaño de una muestra, son los estudios piloto, los cuales suelen llevarse a cabo para "probar en menor escala aspectos logísticos de la ejecución del estudio, lo que evitará cometer errores en los estudios posteriores y de mayor envergadura" (Díaz-Muñoz, 2020: 100). No obstante, como bien señala este mismo autor, este tipo de estudios, a pesar de ser publicables en algunas revistas, por lo general reciben poca o nula importancia cuando se enseña investigación, lo que genera prácticas inadecuadas en su realización y diseño, a pesar de que pueden ser muy útiles para evitar la aparición de defectos que podrían ser graves y costosos en términos de recursos, si es que se replican en el estudio final para el cual se diseñó el piloto.

Es tan poca la relevancia que se le suele dar a los estudios piloto, que por lo general se les considera sólo como una fase del apartado metodológico de los proyectos de investigación, con la única y exclusiva finalidad de probar instrumentos o procedimientos que no cuentan con adecuados criterios de confiabilidad y validez, o que fueron diseñados específicamente para el estudio final que se pretende realizar, usándolos erróneamente como sinónimo de las pruebas piloto o pretest. Este poco valor otorgado a los estudios piloto, puede observarse por ejemplo en los libros o manuales generales de metodología de la investigación que se utilizan con mayor frecuencia en México, y probablemente también en otros países de Latinoamérica, donde ni siquiera suele hablarse de estudios piloto, sino únicamente de pruebas piloto.

Por ejemplo, para Hernández *et al*. (2014) y Hernández y Mendoza (2018), la prueba piloto es sólo una fase para construir instrumentos de medición, en la cual, se administra el instrumento a una muestra pequeña [aunque no se señala qué tan pequeña], y a partir de los hallazgos se puede probar la confiabilidad y validez iniciales de los instrumentos, así como otros aspectos como sus instrucciones, condiciones de aplicación y procedimientos involucrados. Para McMillan y Schumacher (2005), las pruebas piloto también son necesarias para identificar cambios en los instrumentos y sus instrucciones, pero además pueden servir para proporcionar una idea de los patrones de respuesta más probables, de los tiempos de aplicación, así como de la claridad y el formato de aplicación. Aunque para esto consideran muy útil que se cuente con al menos diez personas dispuestas a completar estas pruebas y proporcionar sugerencias para mejorarlas, y que se cuente no con una, sino con dos pruebas piloto: una inicial, y otra más formal con una muestra mayor.

Dieterich (2021) también señala algo similar sobre la relevancia de las pruebas piloto (las cuales el autor denomina erróneamente como estudios piloto), para recabar información previa que ayude a planificar adecuadamente una encuesta. Aunque este autor tampoco señala un número específico de personas a incluir, y su recomendación va más bien en un sentido cualitativo, que implica incluir a personas que se consideren como representativas de la población a estudiar, o que puedan proporcionar datos o pautas relevantes que ayuden a mejorar el diseño de los instrumentos definitivos. Kerlinger & Lee (2002) abordan incluso menos el tema, pues no lo incluyen en lo que exponen sobre muestreo y aleatoriedad, y únicamente se refieren a las pruebas

piloto de manera muy general cuando hablan de las investigaciones por encuesta, señalando que una vez que se han construido los instrumentos a utilizar, "se prueban con antelación en una muestra pequeña representativa del universo" (Kerlinger & Lee, 2002: 548), para luego revisarlos y darles la forma final.

Como puede advertirse en los ejemplos señalados, existen diversas posturas sobre la relevancia de los estudios piloto y la forma en que deben desarrollarse, aunque se relacionan especialmente con poner a prueba instrumentos o procedimientos que se incluirán en una investigación más amplia. No obstante, sobre lo cual no hay realmente un consenso, es sobre la determinación de su tamaño muestral, o bien, la delimitación de cuántos casos, personas o sujetos de estudio se deben incluir en un pilotaje.

Al respecto, ya se ha señalado que sí existen diversas fórmulas y herramientas para dicha tarea, lo curioso es que, en muchas de estas fórmulas y herramientas, se señala la importancia de definir el error de estimación o la variabilidad esperada de los datos, no de forma arbitraria, sino de estudios previos similares, o bien, a partir de los datos obtenidos mediante una prueba piloto; sin aclarar cómo se obtendría el tamaño de la muestra para llevar a cabo ese pilotaje. En el fondo parece generarse una especie de error recursivo, donde el cálculo del tamaño de una muestra requiere un dato de una prueba o un estudio piloto, en la que a su vez hay que calcular un tamaño de muestra. Además, los criterios también suelen diferir dependiendo del campo disciplinar desde el cual se realicen las investigaciones.

Tamaño de muestra para estudios piloto en la investigación médica

Lewis *et al*. (2021) señalan que en las investigaciones que incluyen Pruebas o Ensayos Controlados Aleatorizados (RCTs, por sus siglas en inglés), que son experimentos científicos usados principalmente en el área médica, definir el tamaño de una muestra para los estudios piloto es uno de los aspectos que aún no quedan claros, pues las recomendaciones al respecto varían desde 10-12, hasta 60-75 personas por grupo. Esto depende sobre todo del objetivo principal de la investigación, o de criterios como la viabilidad, un tipo de parámetro clínico, la tasa de eventos esperados, el tamaño del efecto estandarizado[3], si los datos se requieren para validar una escala, o incluso si se trata de un estudio piloto interno o externo al ensayo principal.

Whitehead *et al*. (2016), al respecto de la estimación del tamaño de la muestra en estudios piloto, que son por lo general parte de todo protocolo de investigación médica o de estudios clínicos, señalan que existen dos reglas generales que suelen utilizarse: las de tamaño de muestra plano o fijo (*flat rules),* y las de tamaño de muestra escalonada o establecida mediante una serie de pasos (*stepped rules).* De estas dos, lo más simple, según estos autores, es recurrir a una regla general donde el investigador tiene que elegir un tamaño de muestra para su estudio piloto, de entre las propuestas ya existentes al respecto, como si se tratara de un número fijo [*flat rule*].

Sobre este punto, existen diversas propuestas (todas ellas citadas por Whitehead *et al.,* 2016), como la de Browne (1995), quien sugiere al menos 30 sujetos para poder estimar un parámetro; la de Julious (2005), quien propone un mínimo de 12 sujetos por cada tratamiento empleado; la de Teare *et al*. (2014), quienes recomiendan mínimo 70 sujetos para lograr reducir la imprecisión en la estimación de la desviación estándar; la de Kieser & Wassmer (1996), para quienes 20 a 40 sujetos serían adecuados en estudios cuya muestra final incluya entre 80 a 250 sujetos; y la de Sim & Lewis (2011), quienes propusieron que 55 o más sujetos minimizaría el tamaño de efecto en poblaciones pequeñas a medianas.

No obstante, estas reglas planas tienen sus limitaciones, ya que suelen aplicarse de manera independiente al tamaño de la población final de la prueba que se está diseñando, por lo que el riesgo en la simplicidad de este enfoque es generar una muestra mayor [o no correspondiente] que el tamaño de muestra final requerido para estudios grandes o pequeños

[3] El tamaño del efecto estandarizado, es una medida estadística referida a la magnitud o fuerza que tienen los resultados obtenidos de un instrumento o una intervención experimental, muy útil para la estadística inferencial, pues permite medir la fuerza de relación existente, por ejemplo, entre dos o más variables o entre distintos grupos de estudio.

(Whitehead *et al*., 2016). Además, existen evidentes diferencias entre los diversos estimados, los cuales no sólo difieren en el número plano que establecen como idóneo, sino también en los planteamientos en los que se basan para su establecimiento, que en algunos casos el número estimado parece haberse calculado de manera un tanto arbitraria.

La segunda regla a la que se refieren Whitehead *et al*. (2016), sobre la estimación del tamaño de la muestra en estudios piloto, es la de tamaño de muestra escalonada [*stepped rules*] con posibilidades de mejorarse, la cual consideran una mejor opción que la elección de un número plano. Para este tipo de muestra, es recomendable tomar en consideración al menos un estimado sobre la probabilidad de que el tamaño del efecto estandarizado que se espera en el estudio principal, pueda ser pequeño, mediano o grande dentro de un rango preestablecido. Para ello, estos autores retoman la propuesta de los estándares de Cohen, que varían dependiendo de si el tamaño del efecto estandarizado para el estudio principal, se estima como pequeño, mediano o grande (a la cual Whitehead *et al*. agregan la categoría de extra pequeño), con niveles de potencia estadística ya sea del 80% o 90%. A partir de estos parámetros que pueden estimarse al menos como probables, sería más factible determinar un tamaño de muestra para un estudio piloto con base en lo establecido en la Tabla 1.

Tabla 1. Estimación de reglas generales escalonadas (*stepped rules*) para calcular el tamaño muestral de una prueba piloto, requerido por cada tratamiento o grupo que se utilizará en el estudio principal

Diferencia estandarizada	Prueba principal con 80% de potencia esperada	Prueba principal con 90% de potencia esperada
Extra pequeña ($\delta < 0.1$)	50	75
Pequeña ($0.1 = \delta < 0.3$)	20	25
Mediana ($0.3 = \delta < 0.7$)	10	15
Grande ($\delta = 0.7$)	10	10

Tabla retomada de Whitehead et al (2016, p. 1071). Traducción propia.
El símbolo δ se refiere al tamaño del efecto estandarizado según la clasificación de Cohen.

Este tipo de cálculos son especialmente relevantes en los estudios clínicos y las investigaciones médicas, que es precisamente el foco general de discusión del trabajo de Whitehead *et al.* (2016), y en los cuales también son ampliamente recomendados los estudios experimentales y cuasi experimentales con al menos un grupo control, ya que el tamaño del efecto estandarizado, permite medir el efecto de una intervención o tratamiento a partir de la diferencia de las medias de los grupos de estudio. Este valor estadístico se puede calcular mediante pruebas como la "d" de Cohen o la "g" de Hegdes, y de manera general permitiría estimar si un tratamiento, medicamento o cualquier tipo de intervención, es o no efectiva, a partir de las medias estadísticas de los grupos estudiados.

Tamaño de muestra para estudios piloto en ciencias sociales

De manera similar a los tipos de muestreo sugeridos en las investigaciones médicas, en las ciencias sociales, donde los estudios piloto suelen utilizarse también para probar algunos procedimientos, pero con mayor frecuencia para verificar las propiedades psicométricas de algún instrumento a partir de una prueba piloto, es posible encontrar algunas reglas similares de tamaños fijos de muestras. Una regla de este tipo es la propuesta de McMillan y Schumacher (2005), quienes señalan que aun cuando el tamaño de la muestra para un estudio piloto debería ser mayor de 20, lo importante es que en toda investigación se realice un estudio piloto, aunque para ello se cuente con al menos 10 sujetos.

Otros autores como García-García *et al*. (2013) y Chavarria y Pulgarín (2020), señalan que el tamaño de muestra para pruebas piloto debe ser entre 30 y 50 participantes, quienes además deben poseer los atributos que se pretende medir en la población objetivo; estimado que supuestamente retoman de las recomendaciones de Earl Babbie, particularmente de su obra *Fundamentos de investigación social*. No obstante, Babbie (2000) en dicha obra, aunque sí desarrolla un capítulo completo sobre la lógica del muestreo, donde expone diversos cálculos de tamaños muestrales dependiendo del tipo o diseño de investigación, lo cierto es que nunca se refiere directamente a los estudios piloto. La referencia más cercana a este aspecto es una anotación que hace este autor en relación de la prueba previa (pretest) de los cuestionarios, señalando que "no es esencial que los sujetos de la prueba previa comprendan una muestra representativa, pero debe acudir a personas para quienes el cuestionario sea por lo menos pertinente" (Babbie, 2000: 138), mas no establece de manera directa ninguna cifra estimada para tal efecto.

En otra de sus obras clásicas, específicamente sobre los instrumentos de medición, Babbie (1990) resalta la relevancia de las pruebas previas como una estrategia útil para identificar posibles errores antes de un estudio final, ahorrando dinero y esfuerzos innecesarios, en caso de que se considere que no es posible lograr los objetivos esperados. Babbie (1990) sí distingue aquí entre las pruebas previas de los instrumentos (*pretests*) y los estudios piloto (*pilot estudies*), señalando que los primeros deben ser flexibles y variados, lo que implica ignorar por completo el muestreo controlado y la representatividad estricta, para seleccionar sujetos que estén relacionados con el diseño de investigación y los instrumentos bajo consideración, de modo que puedan contribuir a su mejora. En cambio, los estudios piloto, que para este autor son una versión en miniatura del diseño del estudio final, sí deben incluir una muestra representativa de la población objetivo, seleccionada de la misma forma en que se pretende seleccionar la muestra final, mediante una estrategia que contemple o no la superposición de individuos, según lo permita el estudio planteado.

Aunque Babbie tampoco especifica en esta obra cómo determinar el tamaño de muestra para un estudio piloto, es posible inferir que al considerar a este tipo de estudios con la misma rigurosidad que a los estudios finales, se debe utilizar un procedimiento de determinación del tamaño de la muestra, similar al que se utilizará dependiendo el diseño de la investigación final. Para ello se debería utilizar el mismo principio que este autor denomina como PPS (*Probability Proportionate to Size Sampling*), para resaltar la importancia de que, a menor tamaño de la muestra, mayor probabilidad de error muestral y viceversa (Babbie, 1990). Con esta acotación, se regresa a dos reglas generales ya discutidas en este trabajo a partir de los planteamientos de Kerlinger & Lee (2002) –sobre los que sin duda coincide Babbie (1990, 2000)– respecto de que muestras grandes son necesariamente mejores en cualquier tipo de estudio y que, por tanto, entre más grande sea la muestra, menores serán los errores resultantes, principios que al menos en ciencias sociales, pueden ser muy cuestionados.

No obstante, sí existen autores que consideran algunas reglas o principios a seguir para determinar el tamaño de una muestra, ya sea en las pruebas piloto o en los estudios piloto. Recientemente Roco *et al*. (2021), elaboraron un breve, pero muy interesante, escrito donde enfatizan un problema similar al antes planteado, que es precisamente la falta de fundamentación que respalde la delimitación del tamaño muestral en las investigaciones que se realizan para la validación de instrumentos, aun cuando dichas investigaciones que se publican, sí cuenten con un análisis detallado de validación a partir de los resultados obtenidos de las muestras.

Lo que se cuestiona, por tanto, no es la calidad de los análisis realizados para validar instrumentos, sino la falta de fundamento sobre la determinación del tamaño de la muestra a partir de la cual se elaboran dichos análisis. Si se es muy crítico al respecto, esto sí implica cuestionar los resultados, debido a las inconsistencias o lagunas en el origen y selección de lo(a)s sujetos de estudio, que como ya ha sido señalado, en muchas ocasiones se realizan de forma arbitraria. Bajo este cuestionamiento, Roco *et al*. (2021: 877) aclaran que: "Si bien no existe un único criterio para determinar el tamaño muestral en el proceso de validación de un cuestionario, sí existen una serie de recomendaciones que creemos que es importante tener siempre a la vista". Entre las cuales, usando las aportaciones de otros autores, destacan las siguientes propuestas, cuyas fuentes fueron cotejadas y revisadas para este escrito:

1. Si se pretenden realizar análisis estadísticos de los ítems a partir de la media, desviación típica y coeficiente de correlación, puede ser suficiente contar con una muestra de entre 50 y 100 sujetos para una primera aplicación, previendo que se contará con una segunda aplicación ya con el instrumento mejorado, donde se pueden incluir entre 5 y 10 sujetos por ítem, con un mínimo de 300 sujetos, para así poder tener mejores garantías sobre la validez del instrumento (Osterlind, 2002 y Martínez-Arias, 1995).
2. Si se van a utilizar procedimientos como el análisis factorial confirmatorio, para evaluar la estructura interna de los instrumentos, la recomendación es incluir entre 5 y 10, o incluso entre 2 y 3 participantes por ítem, siempre y cuando el número total no sea inferior a 200 (Stevens, 2009 y Kline, 1986).
3. Snook & Gorsuch (1989), proponen a 300 participantes como regla general, para obtener soluciones fiables en cualquier tipo de análisis factorial realizado.
4. Si se pretende estimar la fiabilidad de un instrumento mediante estadísticos test-retest, formas paralelas o alfa de Cronbach; el estimado de participantes puede ser de entre 200 y 300 (Clark & Watson, 1995).
5. Muestras más pequeñas podrían ser aceptables si la metodología [o diseño] del estudio de validación, tiene contemplado replicar la medición en diferentes grupos, donde el estudio final considere incluir al menos al doble de sujetos que el número de ítems del instrumento a validar, con un mínimo de 100 participantes por grupo (Kline, 1986).

A partir de esta revisión de autores, Roco *et al*. (2021) también concuerdan en la existencia de diversos criterios de tamaños muestrales para validar instrumentos, concluyendo que antes de decidir un número:

> Debemos primero asegurarnos de que el tamaño que estimemos permita que todos los segmentos de la población que sean de nuestro interés estén correctamente representados, evitando sesgos por exceso o deficiencia. Si esta consideración no se cumple, el instrumento no logrará su finalidad, a pesar de que se nos señale estadísticamente que estamos en presencia de un instrumento fiable (Roco *et al*., 2021: 878).

No obstante lo categóricas que parezcan algunas de las propuestas mencionadas, es importante señalar que incluso los mismos autores citados por Roco *et al*. (2021), tampoco son tan dogmáticos al proponer un número fijo o un valor estimado para el tamaño muestral [*flat rule*], como si siempre fuera ésta la mejor opción. Por ejemplo, Osterlind (2002), señala que, al momento de construir un instrumento, no existe un número específico de personas que deban participar en la determinación de una respuesta "correcta" para los ítems de una prueba. Para algunos ítems, sólo dos o tres personas pueden ser suficientes, mientras que, en otros contextos o ítems, se deben incluir más personas en el consenso. Aunque sí se recomienda que para programas de evaluación a gran escala se incluya en el consenso de cada ítem, un mínimo de tres personas, y preferiblemente cuatro o incluso cinco. Esto no implica que este número deba multiplicarse por el total de los ítems de la prueba, sino más bien considerar las respuestas de estas personas en conjunto, buscando establecer un consenso a partir las mismas.

Osterlind (2002) también señala que hay que tomar en cuenta que la validez no se consigue a través del instrumento en sí mismo, sino que es la interpretación de las puntuaciones obtenidas luego de su aplicación, lo que le da significado. En tal sentido, la validez de una prueba se construye como un proceso escalonado de reunir evidencia, para generar una interpretación específica de las puntuaciones obtenidas [*stepped rule*]. Éstas deberán evaluarse de manera concreta a partir de diferentes métodos que justifiquen las inferencias particulares realizadas a partir de los datos recabados, tomando en cuenta que algunas pruebas pueden ser apropiadas para ciertos tipos de inferencias, pero no para otras.

Diversos autore(a)s como Fox *et al*. (2009), Lewis *et al*. (2021) y Stevens (2009), enfatizan la importancia del tamaño de la muestra en términos paramétricos, por su relación directa con la potencia estadística y los errores alfa y beta o de tipo I y II (la probabilidad de rechazar una hipótesis nula cuando es verdadera, o de aceptarla cuando es falsa), lo cual según estos

autore(a)s, en muestras muy grandes (>200) no genera un problema. No obstante, aclara Stevens, si el tamaño de la muestra no es lo suficientemente grande, se puede recurrir a un agrupamiento lógico basado en fundamentos teóricos o sustantivos, por ejemplo en los análisis factoriales, donde a pesar de que muchos autores han recomendado tamaños de muestra mínimos de 200, lo cierto es que muestras más pequeñas también pueden ser adecuadas para modelos con menos parámetros a estimar, e incluso un tamaño de muestra grande puede derivar en la retención de demasiados factores: "un tamaño de muestra grande puede llevar a la retención de demasiados factores" (Stevens, 2009: 329; la traducción es del autor).

Stevens (2009) enfatiza además que, entre los factores retenidos con muestras demasiado amplias, pueden incluirse algunos que en realidad son triviales, como también lo advierte Goodwyn: "Los mismos problemas asociados con el tamaño de la muestra en las pruebas de significancia estadística tradicionales, se presentan cuando estas pruebas se aplican a la retención de factores. Pues utilizar muestras grandes en el análisis factorial, es una estrategia que a menudo suele considerar factores triviales como estadísticamente significativos" (2012: 5; la traducción es del autor).

Como puede observarse, tanto en los estudios piloto, como en las pruebas piloto o pruebas previas de los instrumentos a utilizar en una investigación con aplicación final, no existe una fórmula preestablecida ni un número "mágico" que represente el tamaño ideal de una muestra paramétrica o no paramétrica, por lo que tampoco son totalmente aplicables las reglas generales de números fijos o *flat rules.* Más bien, se plantean diversas opciones de muestras, generalmente reducidas, que pueden funcionar dependiendo del propio diseño y fines de la investigación.

En la literatura existente al respecto en ciencias sociales, parece predominar la sugerencia de que la selección de los casos no se realice de manera aleatoria, sino por su pertinencia en relación con los atributos o características que se pretendan medir, o bien, por los elementos que se pretendan probar con los instrumentos que se estén piloteando, sugiriendo que la muestra de participantes, sea seleccionada de forma no necesariamente paramétrica, sino más bien intencional, deliberada o a juicio de lo(a)s investigadore(a)s, lo que corresponde con tamaños de muestra que se pueden ir mejorando desde una perspectiva más escalonada, tipo *stepped rules.*

Otro aspecto a considerar son las diferencias entre los estudios piloto internos y externos, ya que según lo propuesto por Herbert *et al*. (2019, citado por Lewis *et al*., 2021), esta distinción es muy importante al momento de definir el tamaño de una muestra, debido a que los estudios piloto externos suelen ser ligeramente mayores y diferentes a los internos, que se pueden basar en un porcentaje del tamaño total de la muestra y ser además una parte ya programada de un estudio mayor.

A pesar de que los estudios piloto internos pueden ser muy atractivos, porque es posible incluir los datos de lo(a) participantes en el mismo análisis que se realice para el estudio principal, de modo que el cálculo de la muestra del estudio piloto interno puede no ser necesario, pues se incluirían a una porción de los sujetos que conformarían la muestra total del estudio, lo cierto es que los estudios piloto internos presentan también sus propios riesgos. Basta con citar lo señalado por In (2017), quien aclara que, para realizar un estudio piloto interno, éste debe planificarse minuciosamente como parte de la etapa de diseño del estudio principal, e incluirlo como parte de su procedimiento, de modo que al menos una de las muestras del estudio piloto interno, debe ser idéntica a la del estudio principal, y por tanto se tienen que utilizar los mismos criterios de inclusión y exclusión, que en ocasiones no se conocen bien desde el estudio piloto.

Hay que considerar también como lo señala este autor (In, 2017), que no sería posible realizar cambios sugeridos a partir del estudio piloto, si es que estos afectan a otras categorías del estudio principal. Además, el empleo de estudios piloto internos genera una probabilidad ligeramente mayor de cometer errores tipo I (o alfa), debido a que las hipótesis que se establecen tanto para el estudio piloto como para el estudio principal, suelen ser independientes entre sí. También por el hecho de que los estudios piloto no tienen como finalidad el probar hipótesis, sino comprender mejor la viabilidad del reclutamiento de los participantes o del diseño del estudio principal [o de los instrumentos], por lo que el tamaño de la muestra o muestras, debe establecerse con base en ello, y no con la finalidad de probar hipótesis.

Conclusiones

Ante esta diversidad de propuestas respecto de un tamaño idóneo de muestras, tanto para las pruebas piloto como para los estudios piloto; o bien, ante la posible aplicación de reglas generales para definir dicho tamaño, prácticamente en cualquier campo de investigación, pero especialmente en las ciencias sociales, parece ser que la única certeza, es que no existe tal certeza. Por eso es que en los libros clásicos de metodología de la investigación, se abordan fórmulas y procedimientos específicos para delimitar el tamaño muestral dependiendo el tipo o diseño de investigación, pero rara vez se incluyen a los estudios piloto o a las pruebas previas, que suelen realizarse para validar instrumentos, determinar errores muestrales, generar una idea previa del comportamiento que podrían tener los resultados en un estudio mayor, determinar la necesidad de ajustes en los procedimientos, estimar el costo o viabilidad de un estudio, o cualquier otra finalidad para la cual, se considere necesario realizar un estudio piloto.

La respuesta en estos casos no está en la tendencia general de trabajar siempre con muestras grandes. Autores como Babbie (1990), Stevens (2009), Goodwyn (2012) o Kerlinger & Lee (2002), aunque resaltan las ventajas estadísticas de las muestras grandes, también señalan que éstas no están necesariamente libres de sesgos, por lo que no deberían ser una regla fija para todos los tipos de estudios. Esto es especialmente aplicable a las pruebas o estudios piloto internos, donde la muestra utilizada es en realidad una pequeña porción de la muestra global del estudio principal, la cual necesariamente será más grande, debiendo ser calculada en relación con la población total de estudio con la cual se trabaje.

La respuesta parece ir más bien en el sentido planteado por Lewis *et al*. (2021), quienes destacan precisamente que no existe un consenso respecto de la determinación del tamaño de muestra requerido en los estudios piloto, pues las recomendaciones suelen ser muy variadas dependiendo incluso de quien las realice. Ante estas variaciones, Lewis *et al*. (2021) advierten que, antes de cualquier requisito de muestra fija, el tamaño de la misma puede basarse en el nivel de precisión de un parámetro particular a utilizar para calcular la viabilidad, en la tasa de eventos a incluir en el estudio principal, así como en el tamaño del efecto esperado.

A estos, se podrían agregar otros criterios no tan mencionados, como la posibilidad de aceptación de participar de los sujetos seleccionados, la fidelidad del tratamiento o instrumentos a aplicar, y la retención esperada de lo(a)s participantes. También se sugieren opciones complementarias para disminuir los errores de estimación en los análisis estadísticos, que pueden ser especialmente útiles para los estudios piloto, entre las que se destacan principalmente dos: 1) una adecuada y exhaustiva revisión bibliográfica sobre los temas o variables en los que se va a fundamentar el estudio piloto y, 2) la consulta previa de opiniones de expertos; que pueden complementar la determinación de criterios teóricos y estadísticos para asignar un tamaño de muestra más apropiado, contribuyendo a una mejor confiabilidad y validez.

Es importante mencionar algunos otros supuestos previos, que pueden complicar el empleo de muestras grandes en los estudios piloto, entre los que se destacan:

1. Cuando la población a investigar no es muy extensa, y una muestra amplia puede originar que en el estudio final se dupliquen muchos casos que ya participaron en el pilotaje, sobre todo si éste se realiza como un estudio piloto externo. Ante esta situación, aunque tampoco existe un criterio respecto del tiempo que debe pasar entre la aplicación de una prueba y otra (test-retest), las estimaciones fluctúan entre seis y doce meses cuando se trata de encuestas o pruebas psicológicas, debido a que se asume que existe aprendizaje o habituación a los reactivos, que puede afectar los resultados de la segunda aplicación.
2. Cuando cada caso o sujeto que se estudie supone un costo alto en términos de recursos, generalmente financieros, pero también de otros tipos como humanos o técnicos. Un ejemplo sería el pilotaje de un procedimiento médico innovador altamente costoso, como los derivados de la ingeniería genética o la ingeniería de células madre, donde además de los conflictos éticos que puede suponer su desarrollo, los procedimientos pueden ser muy complejos, por lo que su pilotaje suele hacerse generalmente en especies animales, o con grupos muy reducidos de personas.

3. Cuando los casos o sujetos de estudio son muy reducidos, debido a ciertas características especiales o muy particulares. Por ejemplo, si se realizara una investigación con gemelos monocigóticos, que además por cualquier circunstancia hayan sido separados en sus primeras semanas o meses de vida y criados de manera distinta, lo cual podría ser muy útil para profundizar sobre la diferencia entre lo biológico (genética) y lo social (aprendizaje) en la formación de la personalidad. Las posibilidades de detectar este tipo de casos serían de una entre miles, por lo que habría que conformarse tal vez con algunos pocos casos tanto para los estudios piloto que se requieran (si es que se requieren), como para el estudio final.
4. Cuando se realizarán estudios destructivos y las muestras agotan la posibilidad de experimentación, por lo que no hay forma de replicar los procedimientos con el resto de la población, o bien, su replicación puede ser muy compleja. Aquí se incluye el empleo de procedimientos químicos, físicos o mecánicos que suelen emplearse en ingenierías para probar la duración o calidad de algunos materiales, aunque podrían existir casos similares en otras disciplinas, donde un procedimiento implique dañar o destruir las muestras elegidas para los ensayos.
5. En tipos o diseños de estudio particulares, generalmente de corte cualitativo y no paramétrico, donde el diseño mismo no requiere incluir muchos casos para el estudio final, y en ocasiones tampoco es necesaria la selección de una muestra, como podrían ser las historias de vida, los estudios etnográficos, los diseños fenomenológicos y algunos estudios de casos cualitativos. Estos estudios son especialmente útiles en ciencias sociales, y los estudios piloto en este ámbito, en muchas ocasiones ni siquiera son necesarios, aunque también existen perspectivas teóricas distintas sobre el muestreo cualitativo y no paramétrico, que deberían ser revisadas a detalle cuando se desarrollen este tipo de estudios.
6. Cuando la disponibilidad de sujetos para integrar las muestras es muy limitada, como podría ocurrir en investigaciones que abordan temas muy sensibles donde puede ser difícil conseguir informantes, no por sus características personales, sino por las del tema a investigar que puede generar un muy bajo nivel de participación. Ejemplos de estos temas pueden ser las situaciones vinculadas con el narcotráfico, la tortura, la discriminación, los crímenes violentos, las violaciones graves a los derechos humanos, entre otros, donde una muestra pequeña se puede justificar a partir de la relevancia teórica de los sujetos de estudio incluidos, y no necesariamente por su representatividad estadística.

En estos casos u otras situaciones de investigación particulares, donde no es posible cumplir la tendencia general de trabajar con muestras grandes, la decisión no puede depender de un criterio plano de muestreo o *flat rule,* sino que el tamaño de la muestra, si es que es requerida, deberá ser una decisión ampliamente razonada de manera crítica, por parte de quienes dirijan la investigación. Esta decisión se deberá construir desde una perspectiva sistemática de muestreo escalonado tipo *stepped rule*, la cual considere principalmente las características de la población a incluir en el estudio, así como el diseño y los objetivos de la investigación a realizar, y del estudio piloto en caso de que se requiera, pues como se ha señalado en algunos de los ejemplos anteriores, existen situaciones de investigación donde los estudios piloto pueden no ser necesarios.

Algunas consideraciones sobre el tamaño muestral, también serán distintas en un estudio piloto realizado para validar un instrumento, donde se requerirá un número más amplio de casos para poder efectuar ciertos tipos de análisis estadísticos (como los análisis factoriales exploratorios y confirmatorios), que cuando se realiza simplemente para calcular un error estándar necesario para una muestra final, o bien, para conocer otros aspectos de los instrumentos, como la claridad en sus instrucciones, estructura de ítems, opciones de respuesta, tiempo de aplicación, etc., para lo cual puede no ser necesaria una muestra tan amplia en el pilotaje, y podría incluso ser recomendable alguna forma de muestreo no paramétrico.

Por último, todo proceso de determinación del tamaño muestral, debería considerar las características propias de cada investigación, por lo que antes de tomar una decisión sobre el tamaño de una muestra en un estudio piloto o una prueba previa, es importante analizar de

manera detallada también otros aspectos como la disponibilidad de recursos, el tiempo que se tiene para realizar la investigación, la accesibilidad a los sujetos de estudio, el tipo de datos a recabar para su análisis, entre otros. Una reflexión profunda de estos y otros aspectos propios de las investigaciones, no sólo serviría para determinar de mejor manera un tamaño muestral, sino que podría evitar posibles errores o sesgos en el estudio final, lo cual, es precisamente una de las finalidades principales de los estudios piloto.

Referencias

BABBIE, E. (2000). *Fundamentos de investigación social*. México: International Thomson Editores.

_____ (1990). *Survey Research Methods*. Belmont, California: Wadsworth Publishing Company.

BROWNE, R. H. (1995). On the use of a pilot sample for sample size determination. *Stat Med,* 14(17), 1933-1940. DOI: 10.1002/sim.4780141709

BUNGE, M. (1983). *La investigación científica. Su estrategia y su filosofía*. México: Editorial Ariel S. A.

Centers for Disease Control and Prevention (CDC). (2022). *The Epi Info™ Story.* Recuperado de https://www.cdc.gov/epiinfo/story.html

CHAVARRIA, C. T. M. y PULGARÍN, M. S. A. (2020). Construcción y validación de un instrumento para caracterizar el nivel de innovación en instituciones prestadoras de servicios de salud. *Revista de métodos cuantitativos para la economía y la empresa*, (30), 258-278. https://doi.org/10.46661/revmetodoscuanteconempresa.3519

CLARK, L. A. & WATSON, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment,* 7(3), 309-319. https://doi.org/10.1037/1040-3590.7.3.309

COCHRAN, W. G. (1977). *Técnicas de muestreo*. México: Compañía Editorial Continental.

Díaz-Muñoz, G. (2020). Metodología del estudio piloto. *Revista chilena de radiología*, 26(3), 100-104. http://dx.doi.org/10.4067/S0717-93082020000300100

DIETERICH, H. (2021). *Nueva guía para la investigación científica*. México: Grupo Editor Orfila Valentini.

FOX, N., HUNN, A. & MATTERS, N. (2009). *Sampling and Sample Size Calculation*. National Institute for Health Research. Recuperado de: https://www.studocu.com/row/document/haramaya-university/community-based-training-program/sampling-and-sample-size-calculation/107278420

GARCÍA-GARCÍA, J. A.; REDING-BERNAL, A. y LÓPEZ-ALVARENGA, J. C. (2013). Cálculo del tamaño de la muestra en investigación en educación médica. *Investigación en Educación Médica*, 2 (8), 217-224.

GOODWYN, F. (2012). Question Number Two: How Many Factors? *Annual meeting of the Southwest Educational Research Association*, 1-16.

HERNÁNDEZ, S. R., FERNÁNDEZ, C. C. y BABTISTA, L. M. P. (2014). *Metodología de la investigación*. México: McGraw-Hill.

HERNÁNDEZ, S. R. y MENDOZA, T. C. P. (2018). *Metodología de la investigación: las rutas cuantitativa, cualitativa y mixta*. México: McGraw-Hill.

IN, J. (2017). Introduction of a pilot study. *Korean Journal of Anesthesiology,* 70(6), 601-605. https://doi.org/10.4097/kjae.2017.70.6.601

JULIOUS, S. A. (2005). Sample size of 12 per group rule of thumb for a pilot study. *Pharmaceutical Statistics.* 4(4), 287-291. https://doi.org/10.1002/pst.185

KERLINGER, F. N. & LEE, H. B. (2002). *Investigación del comportamiento. Métodos de investigación en ciencias sociales.* México: McGraw-Hill.

KIESER, M. & WASSMER, G (1996). On the use of the upper confidence limit for the variance from a pilot sample for sample size determination. *Biometrical Journal*, 38(8), 941-949. https://doi.org/10.1002/bimj.4710380806

KLINE, P. (1986). *A Handbook of test Construction.* New York: Methuen.

LEWIS, M.; BROMLEY, K.; SUTTON, C. J.; MCGRAY, G.; MYERS, H. L. & LANCASTER, G. A. (2021). Determining sample size for progression criteria for pragmatic pilot RCTs: the hypothesis test strikes back! *Pilot and Feasibility Studies*, 7(40), 1-14. https://doi.org/10.1186/s40814-021-00770-x

MARRUGAT, J.; VILA, J.; PAVESI, M. y SANZ, F. (1998). Estimación del tamaño de la muestra en la investigación clínica y epidemiológica. *Medicina clínica*, 111, 267-76.

MARTÍNEZ-ARIAS, R. (1995). *Psicometría: Teoría de los test psicológicos y educativos*. Madrid: Síntesis.

MCMILLAN, J. H. y SCHUMACHER, S. (2005). *Investigación educativa. Una introducción conceptual*. Madrid: Pearson.

OSTERLIND, S. J. (2002). *Constructing Test Items: Multiple-Choice, Constructed-Response, Performance, and Other Formats.* New York: Kluwer Academic Publishers.

QUALTRICS (2023). *Sample size calculator*. Recuperado de: https://www.qualtrics.com/blog/calculating-sample-size/

ROCO V., A.; HERNÁNDEZ O., M. y SILVA G., O. (2021) ¿Cuál es el tamaño muestral adecuado para validar un cuestionario? *Nutrición Hospitalaria*, 38(4), 877-878. https://dx.doi.org/10.20960/nh.03633

SIM, J. & LEWIS, M. (2011). The size of a pilot study for a clinical trial should be calculated in relation to considerations of precision and efficiency. *Journal of Clinical Epidemiology,* 65(3) 301-308. https://doi.org/10.1016/j.jclinepi.2011.07.011

SNOOK, S. C. & GORSUCH, R. L. (1989). Component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin,* 106(1), 148-154. https://doi.org/10.1037/0033-2909.106.1.148

SOFTWARE PARA ENCUESTAS QUESTIONPRO (2023). *Calculadora de muestras*. Recuperado de: https://www.questionpro.com/es/calculadora-de-muestra.html#calculadora_de_muestra

SURVEY MONKEY (2023). *Calculadora del tamaño de muestra*. Recuperado de: https://es.surveymonkey.com/mp/sample-size-calculator/

STEVENS, J. P. (2009). *Applied multivariate statistics for the social sciences*. New York: Routledge.

TEARE, M. D.; DIMAIRO, M.; SHEPHARD, N.; HAYMAN, A.; WHITEHEAD, A. & WALTERS, S. J. (2014). Sample size requirements to estimate key design parameters from external pilot randomised controlled trials: a simulation study. *Trials*, 15(264), 1-13. DOI: 10.1186/1745-6215-15-264

WHITEHEAD, A. L.; JULIOUS, S. A.; COOPER, C. L. y CAMPBELL, M. J. (2016). Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. *Statistical Methods in Medical Research*, 25(3), 1057-1073.

Autor.

Rodrigo Vargas Salomón

Departamento de Ciencias de la Salud del Centro Universitario de Los Altos, Universidad de Guadalajara, México.

Doctor en Ciencias Sociales por el Colegio de Jalisco y Maestro en Psicología por la Universidad de Guadalajara. Profesor-Investigador Asociado C de la Universidad de Guadalajara.

E-mail: rodrigo.vargas0848@academicos.udg.mx

Citado.

VARGAS SALOMÓN, Rodrigo (2025). Delimitación del tamaño muestral en estudios piloto. *Revista Latinoamericana de Metodología de la Investigación Social – ReLMIS,* Nº29, Año 15, pp. 7-21.

Plazos.

Recibido: 02/06/2024. Aceptado: 01/11/2024.