



MÉTODOS DE INVESTIGACIÓN CLÍNICA Y EPIDEMIOLOGÍA

4.ª EDICIÓN



Josep M.ª Argimon Pallás
Josep Jiménez Villa

Población de estudio

En la fase de diseño de una investigación se plantean tres cuestiones relacionadas con los participantes: definir la población de estudio (especificar los criterios de selección), calcular el tamaño de la muestra necesario y decidir el método para obtener la muestra (técnica de muestreo). Este capítulo se centra en la primera cuestión. Las otras dos se abordan en los capítulos siguientes.

La formulación del objetivo lleva implícita una definición genérica de la población que se desea estudiar, pero en la redacción del protocolo es necesario establecer con mayor precisión unos criterios de selección de los sujetos que se van a incluir en el estudio.

se selecciona una muestra. En otras ocasiones se incluyen sujetos consecutivamente hasta alcanzar el número necesario. En cualquier caso, la estrategia de inclusión de los sujetos debe intentar que la muestra sea representativa de la población de estudio.

Ejemplo 14.1. Se desea ensayar un nuevo diurético para el control de la hipertensión arterial (HTA) esencial leve-moderada en pacientes de la tercera edad. La *población diana* serían, pues, los sujetos de la tercera edad con HTA esencial leve-moderada. A continuación,

NIVELES DE POBLACIÓN

Pueden tomarse en consideración tres niveles de población (fig. 14.1):

1. **Población diana:** conjunto de individuos al que hace referencia la pregunta principal u objetivo del estudio. Es la población a la que se desearía generalizar los resultados. Se define principalmente por sus características clínicas y demográficas generales.
2. **Población de estudio:** subconjunto de la población diana al que se tiene la intención de estudiar. Se define por los criterios de selección establecidos en el protocolo y presenta determinadas características geográficas y temporales que la hacen accesible a los investigadores.
3. **Muestra:** conjunto de individuos realmente estudiados. En la mayoría de las ocasiones, el número de sujetos necesarios para la realización del estudio es mucho menor que el de candidatos que forman la población de estudio, por lo que, por razones de eficiencia y disponibilidad de recursos (viabilidad),

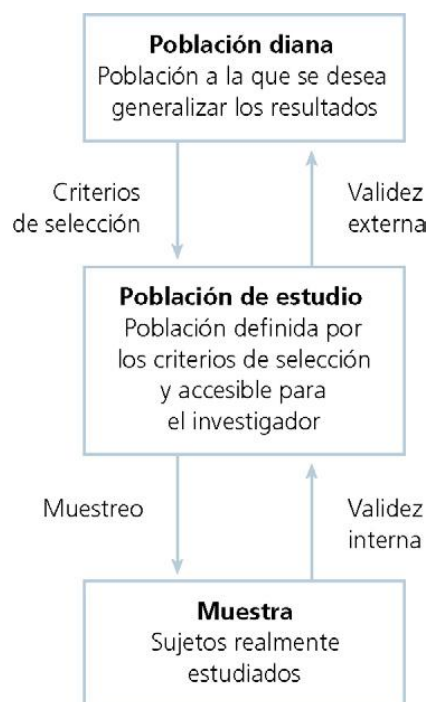


Figura 14.1 Selección de los sujetos de estudio.

deben establecerse criterios de inclusión y exclusión para definir la *población de estudio*: sujetos mayores de 65 años, con HTA esencial, con cifras de presión arterial diastólica (PAD) entre 90 y 109 mmHg, sin tratamiento previo, en los que se ha seguido un protocolo para descartar HTA secundaria, que no han recibido tratamiento antihipertensivo previamente, que no presentan otras patologías concomitantes y que han sido visitados y diagnosticados en los centros en que trabajan los investigadores. Se ha calculado que son necesarios 200 sujetos para la realización del estudio, por lo que los investigadores incluyen a los primeros 200 sujetos que cumplen los criterios de selección (*muestra*).

La consideración de estos niveles tiene repercusiones sobre la interpretación de los resultados. Un investigador desea que sus observaciones puedan generalizarse a todos los individuos que forman la población diana, independientemente de que se visiten o no en los centros en que se realiza el estudio. La posibilidad de realizar inferencias sobre otras poblaciones dependerá, en primer lugar, de cómo se haya diseñado y realizado la investigación, de si la muestra es representativa de la población de estudio, de si se han perdido sujetos que puedan comprometer esta representatividad, de si los datos han sido medidos y registrados correctamente, y de otros factores que condicionan la validez interna. Por otro lado, la validez externa, o capacidad de extrapolación de los resultados a poblaciones diferentes o su generalización a otras más amplias, depende de factores como las diferencias entre la población de estudio y la población diana o entre el tipo de sujetos que se visitan en los centros de estudio y la población diana.

Ejemplo 14.2. Van Spall et al (2007) revisaron sistemáticamente los ensayos clínicos publicados desde 1994 hasta 2006 en revistas con un alto factor de impacto, con el fin de conocer la naturaleza de los criterios de exclusión. Los autores observaron que el ser mujer, niño o anciano son criterios de exclusión frecuentes. También observaron que los criterios de inclusión son más restrictivos en los ensayos multicéntricos y en los que se evalúa un tratamiento farmacológico.

PRINCIPIOS DE REPRESENTATIVIDAD Y COMPARABILIDAD

La muestra de sujetos estudiada debe ser representativa de la población de estudio (*principio de representatividad*). Este principio puede verse comprometido

cuando la muestra inicial ha sido mal seleccionada o cuando, aun habiendo utilizado una técnica de muestreo adecuada, la variabilidad aleatoria (el azar) ha hecho que se obtenga una muestra no representativa, o cuando la muestra de sujetos finalmente analizados está sesgada debido a las no respuestas (sujetos de la muestra inicial de los que no se ha podido obtener la información deseada) o pérdidas durante el seguimiento (en estudios prospectivos).

Ejemplo 14.3. Supóngase que se desea conocer la opinión de los médicos de una provincia sobre un determinado tema y, para ello, se envía un cuestionario a profesionales que trabajan en equipos de atención primaria (EAP). La muestra resultante no es representativa de los médicos de la provincia, ya que no incluye a los médicos que no trabajan en EAP ni a los que sólo tienen consulta privada, por ejemplo.

Ejemplo 14.4. Supóngase que en el ejemplo anterior se selecciona una muestra aleatoria de médicos a partir del registro del Colegio de Médicos de la provincia, a la que se le envía un cuestionario. Responde el 40% de los profesionales. Si la opinión de los que no responden difiere de la de los que sí lo han hecho, el resultado basado únicamente en las respuestas obtenidas estará sesgado, ya que la muestra analizada no es representativa de la población de estudio.

En los estudios analíticos debe cumplirse, además, el *principio de comparabilidad* de los grupos. Estos estudios se basan en que los grupos son comparables por todos los factores pronóstico y en que se ha obtenido la información de la misma forma en todos los sujetos, de manera que las diferencias en los resultados observados pueden atribuirse al factor que se está estudiando.

Ejemplo 14.5. En un estudio se revisan las historias clínicas de los sujetos que han sufrido una enfermedad con la finalidad de comparar su pronóstico en función de si han recibido tratamiento médico o quirúrgico. Ya que la indicación del tipo de tratamiento depende de la gravedad de la enfermedad y de otros factores, los grupos no son comparables (p. ej., entre los sujetos que han recibido tratamiento médico pueden haber tanto sujetos con enfermedad leve como individuos que estén tan graves que en ellos esté contraindicada la cirugía).

El principio de comparabilidad se discute más ampliamente en el capítulo destinado a la formación de los grupos de estudio.

POBLACIÓN DE ESTUDIO

Su definición permitirá identificar en qué individuos se evaluará la presencia del efecto o la asociación de interés, evaluar su idoneidad para alcanzar el objetivo, así como la posibilidad de extrapolación de los resultados a otras poblaciones. Así mismo, permitirá que otros investigadores puedan utilizar criterios semejantes a la hora de diseñar estudios con objetivos similares.

Características generales

La población de estudio debe definirse con criterios precisos, teniendo en cuenta las siguientes consideraciones (cuadro 14.1):

- *Seleccionar pacientes idóneos para observar el efecto o la asociación de interés.* Para identificar las características de los sujetos en que es más plausible encontrar el efecto o la asociación de interés, hay que tener en consideración el mecanismo de acción del factor de estudio. Por ejemplo, en un estudio observacional sobre la asociación entre consumo de anticonceptivos orales y cáncer, se incluirían –obviamente– mujeres en edad fértil y se excluirían aquellas a las que se ha practicado una histerectomía.
- En los estudios experimentales sobre el efecto de una intervención, serían candidatos los pacientes que puedan beneficiarse de ella, y se excluirían los que presenten enfermedades o situaciones concomitantes (embarazo, edad avanzada, alergias, etc.) en las que la intervención pueda resultar peligrosa o causar efectos indeseables.
- *Seleccionar individuos en los que exista una alta probabilidad de detectar el efecto o la asociación de interés.* La selección de los sujetos debe hacerse de forma que se optimicen las posibilidades de detectar la diferencia o asociación que se busca.

Cuadro 14.1 Características generales de la población de estudio

- Pacientes idóneos para observar el efecto o asociación de interés
- Individuos en los que exista una alta probabilidad de detectar el efecto hipotético o la asociación de interés
- Sujetos que probablemente cumplirán el protocolo de estudio (en función del objetivo)
- Excluir aquellos sujetos en los que exista una alta probabilidad de que no pueda medirse la respuesta

Esta consideración es especialmente importante cuando se estudian fenómenos que se presentan con una baja frecuencia o con una latencia muy prolongada. Para reducir el número de sujetos necesarios o la duración del seguimiento, muchos estudios seleccionan poblaciones de alto riesgo.

Ejemplo 14.6. El *Heart Protection Study* (2002) tuvo como objetivo evaluar si un tratamiento hipolipemiente podría reducir la morbilidad y la mortalidad cardiovascular en pacientes de alto riesgo, independientemente de sus cifras de colesterol. Se incluyeron 20.536 hombres y mujeres de 40 a 80 años de edad, con unas concentraciones plasmáticas de colesterol total de al menos 3,5 mmol/l, siempre que se considerara que tenían, en un plazo de 5 años, un riesgo considerable de muerte a causa de unos antecedentes previos de enfermedad coronaria, enfermedad oclusiva de las arterias no coronarias, diabetes, HTA tratada o una combinación de estos factores. Esta muestra fue distribuida aleatoriamente en dos grupos: uno de ellos recibió un tratamiento hipolipemiente, y el otro un placebo. La intención de los investigadores era seleccionar un grupo de muy alto riesgo, ya que, si no, se hubiera necesitado un tamaño de muestra y un seguimiento mucho mayor para encontrar un efecto significativo.

- *Seleccionar sujetos que probablemente cumplirán el protocolo de estudio.* Se pretende evitar situaciones como omisiones de información o incumplimientos de la intervención que podrían comprometer la validez interna del estudio.

Ejemplo 14.7. En un estudio que evaluaba la efectividad de una técnica de educación sanitaria para disminuir las cifras tensionales de los pacientes con HTA esencial de un centro de salud (Mata et al, 1987), se escogieron 113 individuos que cumplieran los criterios de selección. De ellos, se excluyeron, entre otros, los siguientes: 10 pacientes que no habían acudido a las citaciones en los últimos 6 meses, ya que muy posiblemente tampoco cumplirían el protocolo del estudio; un paciente que presentaba hipoacusia importante, dado que el método de educación sanitaria consistía en charlas que difícilmente habría podido seguir; dos casos por enolismo crónico, ya que es conocido que por regla general son malos cumplidores, especialmente de medidas higienicodietéticas; un caso de analfabetismo, ya que se repartían documentos escritos de apoyo a la intervención; tres casos por incapacidad física para desplazarse al centro, y dos por demencia.

Sin embargo, la selección de sujetos muy cumplidores tiene el inconveniente de que aleja la población de estudio de la diana, limitando la generalización de resultados.

- *Excluir sujetos en los que exista una alta probabilidad de que no pueda ser medida la respuesta, ya que podrían dificultar la interpretación de los resultados.* Esta consideración es especialmente importante en los estudios con un seguimiento prolongado, durante los cuales los sujetos pueden presentar otros procesos que impidan su seguimiento o la medición adecuada de la respuesta.

Ejemplo 14.8. En el ejemplo 14.7, los autores excluyeron, además de los ya citados, a un paciente en situación administrativa de desplazado, ya que probablemente no completaría el estudio, y un caso de neoplasia terminal. A menudo, puede ser conveniente excluir pacientes embarazadas, ya que el parto puede interferir en el seguimiento.

Criterios de selección

Los criterios de selección deben establecerse en términos realistas que permitan la inclusión del número deseado de sujetos en el tiempo previsto. Aunque es posible que sean diferentes en estudios con objetivos similares, pueden agruparse en cuatro grandes categorías (cuadro 14.2):

1. *Características sociodemográficas.* La edad, el sexo, la profesión o los años de escolaridad pueden ser fundamentales para definir las características de los participantes y las limitaciones que puede haber para la extrapolación a otras poblaciones.
2. *Características de la enfermedad o exposición.* Es el criterio de inclusión más importante. El diagnóstico de una enfermedad o la identificación de una exposición deben establecerse con la mayor precisión posible para evitar que puedan existir dudas acerca de si un sujeto cumple o no este criterio. En la medida de lo posible es conveniente utilizar criterios estándar ampliamente aceptados o semejantes a los de otros estudios, con la finalidad de facilitar la comparación de los resultados.

Ejemplo 14.9. Marantz et al (1988) revisaron los ensayos clínicos aleatorios de fármacos frente a placebo en la insuficiencia cardíaca congestiva, realizados a doble ciego y publicados entre 1977 y 1985, con la finalidad de evaluar la homogeneidad de los criterios diagnósticos utilizados. Sólo 23 de los 51 estudios revisados (45%) especificaban dichos criterios, y éstos hacían referencia tanto a síntomas y signos clínicos como

Cuadro 14.2 Factores que se deben considerar en la definición de los criterios de selección de los sujetos

Características sociodemográficas de los pacientes

- Edad
- Sexo
- Raza
- Nivel cultural
- Estado socioeconómico
- Profesión
- Otras

Características de la enfermedad o exposición

- Definición de enfermedad o exposición
- Características: forma, tipo, estadio, curso clínico, etiología, complicaciones, duración, etc.
- Tratamiento actual
- Tratamientos previos
- Otras enfermedades o exposiciones
- Otros tratamientos

Otras características

- Embarazo o lactancia
- Consumo de tabaco o alcohol
- Abuso de drogas
- Hábitos dietéticos o nutricionales
- Factores de riesgo
- Limitaciones fisiológicas
- Características psicológicas
- Hipersensibilidad a tratamientos o exploraciones
- Factores ambientales
- Participación en otros estudios
- Incapacidades

Características de accesibilidad

- Lugar de residencia
- Visitado en el centro sanitario
- Inicio y final del período de inclusión

Esta lista es orientativa, y no siempre es preciso considerar todas las características. En algunas ocasiones deben valorarse otros factores.

a datos de laboratorio en combinaciones diferentes. Sólo cuatro estudios presentaban criterios idénticos, y todos ellos correspondían a estudios realizados por el mismo equipo investigador. Por otro lado, las diferencias en los criterios utilizados podrían explicar algunos de los resultados contradictorios observados.

Es conveniente tener presente que la inclusión de falsos positivos o falsos negativos suele conducir a una dilución del efecto, es decir, disminuye la probabilidad de detectarlo. También es importante establecer criterios relacionados con sus características clínicas de tipo, evolución, gravedad, complicaciones y tratamiento. Además, hay que considerar otras enfermedades, procesos o tratamientos concomitantes que puedan ser de interés como criterios de inclusión o de exclusión.

3. *Características de accesibilidad de la población.* No toda la población definida por las características de los apartados anteriores es accesible al investigador. Por ello, deben fijarse criterios de accesibilidad geográfica (lugar de residencia, atención en determinado centro sanitario, etc.) y temporal (inicio y final del período de inclusión).
4. *Otras características.* En determinadas situaciones, la existencia de embarazo o lactancia, o de alcoholismo u otras drogodependencias, o incluso la participación del paciente en otros estudios pueden aconsejar su exclusión.

De esta forma, los criterios establecidos permiten seleccionar un conjunto de sujetos. Este planteamiento es útil cuando se va a estudiar una única muestra de sujetos para estimar un parámetro o medir un efecto en un solo grupo de individuos (como ocurre en los estudios transversales que estiman una prevalencia o en los longitudinales que estiman una incidencia), o cuando se va a subdividir la muestra en dos grupos en función de su exposición a un factor de riesgo (como en los estudios de cohortes con control interno) o de forma aleatoria para comparar dos intervenciones (como en los ensayos clínicos aleatorios).

Sin embargo, otros diseños, como los estudios de casos y controles o los de cohortes con control externo, requieren la definición de criterios específicos para el grupo control. Por ejemplo, en los estudios de casos y controles hay que definir qué patologías se incluyen en el grupo control y, a la vez, excluir aquellos sujetos que puedan estar relacionados positiva o negativamente con el factor de estudio.

Criterios de selección amplios o restrictivos

A la hora de definir los criterios de selección, los investigadores pueden adoptar diferentes perspectivas. En un extremo se sitúa la utilización de criterios restrictivos, que definan una población de estudio muy homogénea y cumplidora con el protocolo, que sea fácil de estudiar y en la que existan mayores posibilidades

de detectar una diferencia o asociación, si existe. Esta postura tiene el inconveniente de que la población de estudio difiere enormemente de la diana, por lo que los resultados serán difícilmente generalizables o extrapolables. En el otro extremo se sitúa la definición de criterios amplios, que buscan facilitar la generalización de los resultados definiendo una población de estudio más representativa de la diana, aun a costa de que la muestra de sujetos sea muy heterogénea y se dificulte la detección de un efecto.

Ejemplo 14.10. Kennedy et al (2003) compararon los sujetos asmáticos tratados con corticoides inhalados que habían participado en ensayos clínicos con los que no lo habían hecho, y comprobaron que los primeros tomaban dosis superiores de corticoides y acudían menos a los servicios de urgencias, lo que demuestra la existencia de diferencias que pueden limitar la generalización de los resultados.

Al definir los criterios de selección, los investigadores deben tener en cuenta algunas consideraciones generales (cuadro 14.3) y sopesar la necesidad de definir una población lo suficientemente homogénea como para permitir detectar el efecto o la asociación que se busca con la necesidad de facilitar la generalización de los resultados, lo que habitualmente implica adoptar una posición intermedia entre ambos extremos.

Procedencia de los sujetos

La procedencia de los sujetos ha de estar en consonancia con el objetivo. Si el interés está en conocer la prevalencia de una enfermedad en la comunidad, deberá escogerse una muestra representativa de la

Cuadro 14.3 Consideraciones generales en la definición de criterios para la selección de la población de estudio

- Optimización de las posibilidades de detectar una diferencia o asociación, si existe
- Población suficientemente homogénea respecto a las características más relevantes del fenómeno de estudio
- Debe asemejarse, dentro de unos límites razonables, a la población diana a la que se desearía generalizar los resultados
- Criterios realistas que permitan la inclusión del número deseado de sujetos en el tiempo previsto
- Aspectos éticos

población de dicha comunidad. Si se optara por una muestra de individuos atendidos en un centro sanitario, se obtendría un resultado sesgado, ya que la demanda de servicios estaría influida, además de por las necesidades biológicas, por factores como el sistema sanitario, los propios profesionales o factores socioculturales, entre otros. En cambio, si el objetivo es conocer la prevalencia de esta misma enfermedad en la población demandante de un centro, la muestra deberá ser representativa de aquellos que se visitan en el centro.

Desde un punto de vista estricto, los resultados de un estudio son aplicables a la población de la que la muestra de sujetos estudiados sea realmente representativa. Por tanto, los resultados que se obtienen pueden variar en función de los criterios de selección utilizados y del ámbito o tipo de centro en que se realiza el estudio. De hecho, algunas de las diferencias observadas en la literatura entre estudios dependen más de la diferente procedencia de los sujetos que de otros aspectos del diseño.

Ejemplo 14.11. Ellenberg y Nelson (1980) revisaron los estudios publicados sobre la pertinencia del tratamiento preventivo de las recurrencias de las crisis convulsivas febriles en la infancia. Por un lado, en los estudios de base poblacional se encontraron tasas de recurrencia muy bajas, de entre el 1,5 y el 4,6%, mientras que si los niños estudiados procedían de consultas hospitalarias o unidades especializadas, las tasas eran muy variables, y en general más altas (desde el 2,6 hasta el 76,9%). Si se tuvieran en cuenta únicamente los resultados del último tipo de estudios, la necesidad de un tratamiento preventivo de las recurrencias parecería clara. Sin embargo, si se consideran los estudios de base poblacional, esta decisión no es tan evidente. Esto no significa que unos sean mejores que otros, sino simplemente que las poblaciones estudiadas son distintas en lo que se refiere al tipo de enfermedad, su gravedad, otras patologías concomitantes, etc., por lo que es importante que los investigadores utilicen criterios de selección coherentes con el objetivo concreto del estudio.

INCLUSIÓN DE LOS SUJETOS

El reclutamiento con éxito de los participantes es esencial en toda investigación. La estrategia de inclusión de sujetos busca obtener una muestra representativa de la población definida por los criterios de selección, del tamaño suficiente y de forma eficiente.

Ejemplo 14.12. Moore et al (2000) analizaron las características de los participantes en dos grandes ensayos sobre tratamiento antirretroviral en sujetos infectados por el VIH, y observaron que existía una infrarrepresentación de pacientes que se habían infectado por vía heterosexual o por el uso de drogas por vía parenteral, una mayor frecuencia de pacientes homosexuales y una menor probabilidad de pérdidas de seguimiento al año, en relación con los que cumplían los criterios de selección, pero que no fueron incluidos en el estudio.

El éxito del reclutamiento depende de diversos factores, entre los que destacan los relacionados con el propio fenómeno de estudio, las características socioculturales de la población y su relación con los profesionales o servicios sanitarios, y la forma en que se solicite el consentimiento para participar. Por ello, al solicitar la colaboración es importante presentar de forma comprensible el tema y los objetivos del estudio, la importancia de su participación, la institución que avala el proyecto, la garantía de confidencialidad de la información, los beneficios potenciales y los posibles riesgos y molestias que podría suponerles su participación, etc.

Cuando se diseña la estrategia de inclusión de los sujetos, deben preverse las posibles causas que puedan reducirla y establecer mecanismos para evitarlas. En estudios de larga duración es frecuente que la tasa de reclutamiento disminuya como consecuencia del cansancio de los investigadores. Suele ser útil monitorizar la tasa de inclusión a medida que el estudio progresa, para detectar la aparición de estos problemas y poder solucionarlos.

ESTRATEGIAS PARA DISMINUIR LAS NO RESPUESTAS Y LAS PÉRDIDAS DE SEGUIMIENTO

La representatividad de la muestra debe garantizarse a lo largo de todo el proceso de la investigación de forma que se eviten sesgos en el análisis de los resultados.

En los diseños transversales es importante conseguir el mayor porcentaje de respuestas posible. Si el reclutamiento es telefónico, deben preverse múltiples llamadas en diferentes franjas horarias, y se deben incluir también los fines de semana. Si es por carta, por ejemplo, puede enviarse un agradecimiento a los que han respondido, un recordatorio a la semana de que haya vencido el plazo de participación a aquellos que no han respondido, una nueva carta a las tres semanas si todavía no lo han hecho,

y, si es necesario, una tercera carta, aproximadamente a las siete u ocho semanas de la primera.

En los estudios que requieren seguimiento, deben preverse estrategias para reducir el número de pérdidas. En los criterios de selección, deben excluirse los individuos que es probable que se pierdan durante el seguimiento, por ejemplo, porque piensan cambiar de domicilio. Es conveniente recoger

información que permita localizar a los individuos en caso de que interrumpen el seguimiento (p. ej., la dirección y el teléfono del individuo, los amigos y los familiares), así como mantener contactos periódicos con ellos, telefónicamente o por correo, que también servirán para determinar con mayor exactitud el momento en que aparecen los desenlaces de interés.

BIBLIOGRAFÍA DE LOS EJEMPLOS

- Ellenberg JH, Nelson KB. Sample selection and the natural history of disease: studies of febrile seizures. *JAMA*. 1980;243:1337-40.
- Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial. *Lancet*. 2002;360:7-22.
- Kennedy WA, Laurier C, Malo JL, Ghezzi H, L'Archeveque J, Contandriopoulos AP. Does clinical trial subject selection restrict the ability to generalize use and cost of health services to «real life» subjects? *Int J Technol Assess Health Care*. 2003;19:8-16.
- Marantz PR, Alderman MH, Tobin JN. Diagnostic heterogeneity in clinical trials for congestive heart failure. *Ann Intern Med*. 1988;109:55-61.
- Mata M, Monraba M, Travería M, Guarner MA, Antó JM. Ensayo aleatorio controlado de educación sanitaria en pacientes hipertensos en asistencia primaria. *Aten Primaria*. 1987;4:189-94.
- Moore DA, Goodall RL, Ives NJ, Hooker M, Gazzard BG, Easterbrook PJ. How generalizable are the results of large randomized controlled trials of antiretroviral therapy? *HIV Med*. 2000;1:149-54.
- Van Spall HG, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. *JAMA*. 2007;297(11):1233-40.

BIBLIOGRAFÍA

- Akl EA, Gaddam S, Mustafa R, Wilson MC, Symons A, Grifasi A, et al. The effects of tracking responses and the day of mailing on physician survey response rate: three randomized trials. *PLoS One*. 2011;6(2):e16942.
- Blümle A, Meerpohl JJ, Rücker G, Antes G, Schumacher M, von Elm E. Reporting of eligibility criteria of randomised trials: cohort study comparing trial protocols with subsequent articles. *BMJ*. 2011;342:d1828.
- Edwards PJ, Roberts I, Clarke MJ, Diguseppi C, Wentz R, Kwan I, et al. Methods to increase response to postal and electronic questionnaires. *Cochrane Database Syst Rev*. 2009;(3):MR000008.
- Graffy J, Bower P, Ward E, Wallace P, Delaney B, Kinmonth AL, et al. Trials within trials? Researcher, funder and ethical perspectives on the practicality and acceptability of nesting trials of recruitment methods in existing primary care trials. *BMC Med Res Methodol*. 2010;10:38.
- Page MJ, French SD, McKenzie JE, O'Connor DA, Green SE. Recruitment difficulties in a primary care cluster randomised trial: investigating factors contributing to general practitioners' recruitment of patients. *BMC Med Res Methodol*. 2011;11:35.
- Treweek S, Mitchell E, Pitkethly M, Cook J, Kjeldstrøm M, Taskila T, et al. Strategies to improve recruitment to randomised controlled trials. *Cochrane Database Syst Rev*. 2010;(1):MR000013.
- Williamson MK, Pirkis J, Pfaff JJ, Tyson O, Sim M, Kerse N, et al. Recruiting and retaining GPs and patients in intervention studies: the DEPS-GP project as a case study. *BMC Med Res Methodol*. 2007;7:42.

Tamaño de la muestra

El cálculo del tamaño de la muestra permite responder a las siguientes preguntas: ¿cuántos individuos es necesario estudiar para poder estimar un parámetro determinado con el grado de confianza deseado? ¿Cuántos sujetos es necesario estudiar para tener las mínimas garantías de poder detectar una determinada diferencia entre los grupos de estudio, en el supuesto de que esta diferencia exista realmente?

El resultado del cálculo debe considerarse orientativo, ya que se basa en asunciones que pueden ser incorrectas. Por ello, es importante realizarlo cuidadosamente, ya que tiene implicaciones sobre la viabilidad del estudio y puede suponer modificaciones en su diseño (modificación de los criterios de selección, prolongación del período de inclusión, etc.).

La inclusión de un número excesivo de sujetos encarece el estudio tanto desde el punto de vista económico como de los recursos humanos y físicos. Además, puede ser considerado poco ético, ya que un mayor número de pacientes estará sometido innecesariamente a determinadas exploraciones o recibirá la intervención menos eficaz. Por otra parte, un estudio con un tamaño insuficiente de la muestra estimará un parámetro con poca precisión o será incapaz de detectar diferencias entre los grupos, conduciendo erróneamente a la conclusión de que no existen.

En general, puede considerarse que el objetivo de una investigación pertenece a uno de los dos tipos siguientes: *estimación de un parámetro* (p. ej., el porcentaje de pacientes hipertensos que presentan una determinada complicación) o *contraste de hipótesis* (p. ej., evaluar cuál de dos pautas terapéuticas consigue un mayor porcentaje de éxitos).

ESTUDIOS QUE ESTIMAN UN PARÁMETRO

Se trata de estudios en los que los investigadores, partiendo de los valores observados en la muestra estudiada, y utilizando la inferencia estadística, buscan estimar el valor de un parámetro en la población.

Consideraciones previas: intervalo de confianza

Supongamos que se desea estimar el valor medio de la colesterolemia en una población. Se escoge una muestra de 80 sujetos en los que se observa una media de 180 mg/100 ml, con una desviación estándar de 20 mg/100 ml. Si se hubiera estudiado una muestra diferente, seguramente se habrían obtenido cifras distintas aunque los criterios de selección hubiesen sido los mismos, si bien es probable que el valor observado en esta segunda muestra hubiera estado alrededor de 180 mg/100 ml. Por lo tanto, a partir del valor observado en una única muestra no puede conocerse exactamente el verdadero valor en la población de origen, ya que es tan sólo una aproximación (estimación puntual). Para conocer entre qué límites es más probable que se sitúe este verdadero valor, debe calcularse el *intervalo de confianza* (IC).

El método de cálculo del IC en las situaciones más frecuentes se desarrolla en los capítulos dedicados al análisis de los datos. La amplitud del IC, es decir, la precisión de la estimación, depende del grado de confianza, la variabilidad del parámetro de interés y el número de sujetos estudiados. Cuanto menor sea la variabilidad del parámetro y mayor el número

de sujetos, mayor precisión existirá en la estimación para un grado de confianza determinado. Cuanta más confianza se desee tener, más amplio será el IC, y menor la precisión obtenida.

Cálculo del tamaño de la muestra

Para calcular el tamaño de la muestra debe conocerse:

- La *variabilidad* del parámetro que se desea estimar. Si no se conoce, puede obtenerse una aproximación a partir de datos propios o de otras investigaciones, o un estudio piloto. En el caso de las variables cuantitativas se mide por la variancia, y en el de las cualitativas, por el producto $P \cdot (1 - P)$.
- La *precisión* con que se desea obtener la estimación, es decir, la amplitud del IC. Cuanto más precisa se desee, más estrecho deberá ser este intervalo, y más sujetos deberán ser estudiados. La precisión debe fijarse previamente, en función de la finalidad de la estimación. En algunos casos puede requerirse una gran precisión, mientras que en otros, si sólo se necesita conocer aproximadamente entre qué valores se encuentra el parámetro, se requerirá una menor precisión y, consecuentemente, menos sujetos.
- El *nivel de confianza* deseado, que habitualmente se fija en el 95%, correspondiente a un valor α de 0,05. Indica el grado de confianza que se tiene de que el verdadero valor del parámetro en la población se sitúa en el intervalo obtenido. Cuanta más confianza se desee, menor será el valor de α , y más elevado el número de sujetos necesario.

De estos tres elementos, sólo debe conocerse la variabilidad del parámetro, ya que tanto la precisión como el nivel de confianza se fijan a partir de los intereses del investigador.

Estimación de una proporción

La fórmula para el cálculo del número de sujetos para estimar una proporción se presenta en el cuadro 15.1. También puede utilizarse la tabla A que se encuentra al final del libro.

Ejemplo 15.1. Un estudio se propone estimar el porcentaje de pacientes hipertensos de un centro que están controlados. A partir de datos previos, se estima que debe estar situado alrededor del 40% ($p=0,40$). Se quiere realizar la estimación con una precisión de $\pm 4\%$ ($i=0,04$) y una confianza del 95% ($1 - \alpha = 0,95$; $\alpha = 0,05$; $Z\alpha = 1,96$). Aplicando

Cuadro 15.1 Fórmulas para el cálculo del número de sujetos necesarios para la realización de un estudio cuyo objetivo es la estimación de una media o una proporción

Estimación de una proporción (variable cualitativa)

$$N = \frac{Z\alpha^2 P(1-P)}{i^2}$$

Estimación de una media (variable cuantitativa)

$$N = \frac{Z\alpha^2 s^2}{i^2}$$

N: número de sujetos necesarios.

Z α : valor de Z correspondiente al riesgo α fijado (tabla 15.4).

P: valor de la proporción que se supone existe en la población.

s²: variancia de la distribución de la variable cuantitativa que se supone que existe en la población.

i: precisión con que se desea estimar el parámetro (2*i* es la amplitud del intervalo de confianza).

la fórmula correspondiente, se calcula que se precisan 576 sujetos. La tabla A proporciona el mismo resultado.

Esta estimación del número de sujetos realizada a priori no es más que una aproximación. Si en el ejemplo el porcentaje real de pacientes controlados es diferente del 40%, puede alterarse el grado de precisión de la estimación.

En la tabla 15.1 puede apreciarse cómo diferentes grados de precisión y diferentes valores del porcentaje esperado modifican el número de sujetos necesarios. Así, por ejemplo, para estimar el porcentaje del 40% del ejemplo 15.1 con el mismo nivel de confianza del 95%, los 576 sujetos se convierten en 9.220 cuando se desea una precisión muy alta ($i=0,01$), o en tan sólo 92 si se es menos exigente ($i=0,1$). Modificando cualquiera de estos valores, puede obtenerse un número de individuos que se aproxime al «deseado» o al disponible. Debe evitarse cualquier manipulación del cálculo, ya que, al reducir el número de sujetos que se van a estudiar, también disminuye el grado de precisión con que el parámetro va a ser estimado y aumenta la amplitud del IC.

En el cálculo del tamaño de la muestra también debe tenerse en cuenta la estrategia de análisis y cómo se presentarán los resultados. Así, si en el

Tabla 15.1 Influencia de la precisión de la estimación y del valor supuesto de la proporción que se desea estimar sobre el número de sujetos necesarios

Influencia de la precisión		
P	i	N
0,40	0,100	92
0,40	0,075	164
0,40	0,050	369
0,40	0,025	1.475
0,40	0,010	9.220
Influencia de la proporción		
P	i	N
0,10	0,05	138
0,20	0,05	246
0,30	0,05	323
0,40	0,05	369
0,50	0,05	384

Todos los cálculos se han realizado con el grado de confianza del 95%. P: valor supuesto de la proporción que se desea estimar. i: medida de la precisión de la estimación, que corresponde a la mitad del intervalo de confianza. N: número de sujetos necesarios para realizar la estimación de P con la precisión deseada y el 95% de confianza.

ejemplo 15.1 los investigadores presentan el resultado principal en función del sexo, la estratificación hará que la estimación se haya obtenido en un número menor de sujetos, por lo que la precisión será menor de la deseada.

En algunas ocasiones no se conoce el valor aproximado del parámetro que se está buscando. Si no existen datos de la literatura que resulten útiles, o si no puede realizarse una prueba piloto para obtener una primera aproximación a dicho valor, puede adoptarse la postura de la máxima indeterminación. Esta actitud consiste en suponer que el porcentaje que se quiere estimar se sitúa alrededor del 50%, ya que es el valor que requiere una mayor cantidad de individuos para una precisión determinada, de forma que el número obtenido supera cualquier otra combinación de parámetros. Esta actitud también es útil, por ejemplo, cuando se realizan encuestas en las que interesa estimar múltiples parámetros. Su inconveniente es que requiere el estudio de un número de sujetos a veces muy superior al realmente necesario.

Estimación de una media

El cálculo es similar al de la situación anterior (ver cuadro 15.1). También deben fijarse el nivel de confianza y la precisión de la estimación, que se traducen en la amplitud del intervalo alrededor de la media que se desea estimar. En las variables cuantitativas, la medida de la variabilidad viene proporcionada por la variancia de su distribución en la población. La tabla B del final del libro proporciona una fácil solución alternativa al cálculo matemático.

Ejemplo 15.2. Supongamos que se desea estimar la tensión arterial diastólica (TAD) de los pacientes diabéticos registrados en un centro de salud. Por estudios previos se conoce que la desviación estándar de la TAD en sujetos diabéticos es de 25 mmHg ($s = 25 \text{ mmHg}$; $s^2 = 625 \text{ mmHg}^2$). Se quiere realizar la estimación con una confianza del 95% ($1 - \alpha = 0,95$; $\alpha = 0,05$) y una precisión de 5 mmHg ($i = 5$). Aplicando la fórmula, se puede determinar que son necesarios 96 sujetos. Esta cifra debe aumentarse en función del porcentaje de sujetos en que se prevea que no podrá determinarse la TAD. El mismo resultado se obtiene consultando la tabla B en la línea correspondiente al valor $i/s = 5/25 = 0,200$.

Si no se dispone de una estimación de la variabilidad, puede utilizarse una regla práctica, que consiste en determinar la diferencia entre los valores máximo y mínimo esperables. Dado que se asume que esta variable es de distribución normal, el intervalo $m \pm (2s)$, siendo m la media y s la desviación estándar de la distribución, incluye aproximadamente el 95% de los valores posibles, por lo que, al dividir dicha amplitud de valores entre 4, puede obtenerse una cierta aproximación al valor de la desviación estándar s .

Corrección para poblaciones finitas

En los cálculos anteriores no ha intervenido el tamaño de la población, ya que en ellos se ha asumido que es infinito. Sin embargo, en muchas ocasiones es preciso obtener una muestra de una población de tamaño conocido (finito). En esta situación puede aplicarse la siguiente fórmula, que ajusta el número de individuos necesarios según el tamaño de la población:

$$n_a = n[1 + (n/N)]$$

donde n_a es el número de sujetos necesarios, n es el número de sujetos calculado para poblaciones infinitas y N es el tamaño de la población.

Ejemplo 15.3. En el ejemplo 15.1 se ha calculado que son necesarios 576 sujetos para estimar el

porcentaje de hipertensos controlados de un centro, realizando determinadas asunciones. Si la población de referencia es de 1.000 hipertensos, aplicando la fórmula anterior podría determinarse que son necesarios 365 pacientes hipertensos.

ESTUDIOS QUE COMPARAN DOS GRUPOS

Muchos estudios tienen por objetivo responder a preguntas del siguiente tipo: ¿es más eficaz el fármaco A que el B en el tratamiento de una enfermedad? ¿Existe un aumento del riesgo de malformaciones congénitas en los niños nacidos de madres que han padecido cierto trastorno durante el embarazo?, etc.

Consideraciones previas: pruebas de contraste de hipótesis

Bajo las preguntas que se han planteado en el párrafo anterior subyace una hipótesis, por ejemplo, la de que B es más eficaz que A en el tratamiento de los hipertensos esenciales de edad avanzada. En la realidad, esta hipótesis puede ser verdadera o falsa, y eso es lo que se trata de averiguar. Dado que el investigador no puede estudiar todos los hipertensos esenciales de edad avanzada, la hipótesis debe contrastarse en una muestra de esta población. A partir únicamente de los datos recogidos en un estudio, no puede aceptarse o rechazarse absolutamente la hipótesis que se contrasta. Existe cierto riesgo de error, sea cual sea la conclusión a la que se llegue.

Error α y error β

Desde el punto de vista estadístico, el investigador se enfrenta al problema del contraste de una

hipótesis mediante la asunción de que no existen diferencias en la eficacia de los dos fármacos A y B (*hipótesis nula*). A partir de los resultados observados en la muestra, el investigador utilizará las pruebas de significación estadística para evaluar si existe la suficiente evidencia que le permita rechazar esta hipótesis nula y, consecuentemente, aceptar la hipótesis alternativa de que sí existen diferencias entre ambas terapéuticas.

El *error tipo I*, conocido también como *error α* (tabla 15.2), se comete cuando el investigador rechaza la hipótesis nula, siendo ésta verdadera en la población. Es equivalente a encontrar un resultado falso positivo, ya que el investigador concluye que hay una diferencia, cuando en realidad no existe.

El *error tipo II*, o *error β* (ver tabla 15.2), se comete en la situación contraria, cuando el investigador no rechaza la hipótesis nula, siendo ésta falsa en la población. Es equivalente a un resultado falso negativo, ya que el investigador concluye que ha sido incapaz de encontrar una diferencia que existe en la realidad. Si β representa la probabilidad de un resultado falso negativo, su complementario $1 - \beta$, conocido como *poder o potencia estadística*, representa la probabilidad de observar en la muestra una determinada diferencia o un efecto, si existen en la población.

Para calcular el tamaño de la muestra, se fijan a priori los riesgos que se está dispuesto a asumir de cometer estos errores. Lo ideal sería poder reducirlos a cero, pero ello no es posible. Una de las formas para disminuirlos consiste en aumentar el número de sujetos estudiados.

Hipótesis unilaterales e hipótesis bilaterales

Desde el punto de vista estadístico, otro aspecto que afecta al cálculo del tamaño de la muestra es el de la formulación precisa de la hipótesis que se

Tabla 15.2 Tipos de error aleatorio en una prueba estadística de contraste de hipótesis

		Realidad (población)	
		Existe diferencia o asociación (Ho falsa)	No existe diferencia o asociación (Ho cierta)
Resultado de la prueba (muestra)	Diferencia o asociación significativa (rechazo de Ho)	No error	Error tipo I α
	Diferencia o asociación no significativa (no rechazo de Ho)	Error tipo II β	No error

Ho: hipótesis nula.

Muestreo

El principio de representatividad implica que, para que los resultados de una investigación tengan validez interna, la muestra de sujetos estudiados debe ser representativa de la población de estudio. En los estudios transversales que estiman un parámetro se emplean técnicas de muestreo probabilístico, que aumentan la probabilidad de obtener una muestra representativa. Cuando no se dispone de un listado con los miembros que forman la población de estudio o cuando ésta es dinámica, y la muestra de sujetos se selecciona por métodos como la inclusión consecutiva de los pacientes que acuden a la consulta y cumplen los criterios de selección, para poder realizar inferencias, también debe asumirse que la muestra seleccionada es representativa de la población de estudio.

VENTAJAS DEL MUESTREO

La razón principal para estudiar una muestra y no toda la población es el ahorro en tiempo, dinero y otros recursos. Es una cuestión de eficiencia. Si se ha calculado que son necesarios 300 individuos, y la población de referencia la forman 3.000 personas, estudiarlas a todas supone desperdiciar una buena cantidad de recursos, ya que, si la selección de la muestra y la medición de las variables se realizan correctamente, se obtendrá un resultado válido estudiando una muestra representativa de 300 personas.

Una segunda ventaja del muestreo, también relacionada con la validez del estudio, es que las mediciones tenderán a ser de mayor calidad, ya que se pueden dedicar más esfuerzos al estudiar un menor número de individuos. Paradójicamente, podría darse el caso de que se obtuvieran resultados

más fidedignos al estudiar una muestra que a toda la población.

También es un problema de cantidad de información. Al poder invertir más recursos en el estudio de cada individuo, pueden medirse otras variables, lo que no es posible hacer en toda la población. Si algunas de estas variables son potenciales factores de confusión, su medición posibilita su control en el análisis, lo que aumenta la validez de los resultados del estudio.

Ejemplo 16.1. En un estudio cuyo objetivo era conocer la prevalencia de factores de riesgo cardiovascular en jóvenes de 16 a 19 años, se seleccionó aleatoriamente una muestra de 1.176 estudiantes en los que se evaluaron unos pocos factores de riesgo (Andersen et al, 1989). A partir de esta muestra se seleccionó uno de cada cuatro estudiantes, en los que además se analizó su perfil lipídico, se midieron la presión arterial, el pliegue corporal, y el pulso en condiciones basales y después de una actividad intensa. Con la finalidad de establecer la representatividad de la segunda muestra se compararon la edad, el sexo, la talla y el peso con la primera, y no se encontraron diferencias significativas. El hecho de estudiar un conjunto de variables de forma más exhaustiva en una muestra reducida persigue el objetivo de alcanzar mayor exactitud y estandarización en las medidas llevadas a cabo.

PLANIFICACIÓN DEL MUESTREO

Las *unidades de muestreo* son los elementos sobre los que se aplicará la técnica de selección. Pueden ser personas, centros de salud o escuelas, por ejemplo.

Si se quiere estimar la prevalencia de hipertensión arterial (HTA) en una población, serán las personas; si se pretende conocer el número de escuelas de una provincia en las que se imparte educación sanitaria, serán las escuelas.

Las unidades de muestreo no tienen que coincidir necesariamente con las *unidades de análisis*. En un estudio para conocer la prevalencia de HTA en la población escolar, las unidades de muestreo pueden ser las escuelas, pero se analiza en cada una la presión arterial de sus alumnos (unidades de análisis).

El primer paso para planificar un muestreo suele ser la definición del *marco muestral*, que es esencialmente una lista, o forma de identificación, de las unidades de muestreo sobre las que se aplicará el proceso de selección. La definición de este marco debe ser cuidadosa, ya que condiciona la interpretación de los resultados.

Ejemplo 16.2. En un estudio sobre la asociación entre la adaptación social y la depresión, Luty et al (2002) comprobaron que esta asociación no variaba entre los diferentes países, pero sí dependía del marco muestral del estudio, ya que se observaron resultados diferentes en la comunidad y en los centros sanitarios.

Ejemplo 16.3. Fuchs et al (2001) compararon la prevalencia de HTA que se obtenía al utilizar dos marcos muestrales diferentes: los adultos residentes en una población y los que acuden a una unidad de prevención de la HTA en un centro sanitario. Las diferencias en la prevalencia observada (el 24 frente al 42%, respectivamente) ilustran la influencia del proceso de muestreo sobre la evaluación de los factores de riesgo en la población.

El siguiente paso es decidir cómo seleccionar la muestra. En líneas generales, las técnicas de muestreo pueden clasificarse en probabilísticas y no probabilísticas.

TÉCNICAS DE MUESTREO

Muestreo probabilístico

Se define como el proceso de selección en que todos los individuos candidatos tienen una probabilidad conocida, distinta de cero, de ser incluidos en la muestra. Se utiliza alguna forma de selección aleatoria para obtener las unidades que serán estudiadas.

Existe una tendencia a considerar sinónimas las expresiones *muestra probabilística* y *muestra representativa*. El uso de una técnica de muestreo probabilístico

tiende a asegurar que se obtendrá una muestra representativa, en especial si la población y la muestra son de gran tamaño. Sin embargo, puede que no sea así, ya que el propio azar puede conducir a una muestra que no tenga la misma distribución de las variables de interés que la población de referencia, sobre todo si su tamaño es reducido. Por otro lado, pueden obtenerse muestras representativas utilizando técnicas no probabilísticas.

En todas las técnicas probabilísticas la selección de las unidades se realiza al azar y se evita la posible parcialidad, consciente o inconsciente, de los investigadores. Por esta razón, es más probable que las muestras tiendan a ser representativas, aunque ello no significa que estén exentas de error. En realidad, la muestra seleccionada no es más que una de las muchas que podrían obtenerse de la población de referencia. Es probable que la estimación de la variable obtenida a partir de una muestra difiera por azar (error aleatorio o debido al muestreo), y lo es más cuanto menor es el tamaño de la muestra. El uso de muestras probabilísticas permite la aplicación de técnicas estadísticas capaces de cuantificar este error aleatorio.

Muestreo aleatorio simple

Es aquella técnica en la que cada unidad del marco muestral tiene la misma probabilidad de ser escogida y en la que cada una de las posibles muestras del mismo tamaño tienen la misma probabilidad de ser escogidas. El muestreo aleatorio simple es, pues, un procedimiento equiprobabilístico, pero no todo método equiprobabilístico es aleatorio simple.

Ejemplo 16.4. Silva (1993) ilustra este concepto con el siguiente ejemplo: supongamos un grupo de 10 niños ordenados alfabéticamente, de los que se quiere seleccionar una muestra de 5 niños. Consideremos los tres siguientes procedimientos de selección: *a)* tomar los 5 primeros de la lista; *b)* dividir el listado en dos grupos (los 5 primeros y los 5 últimos), lanzar una moneda al aire y tomar como muestra los elementos del primer grupo en caso de obtener cara, y los del segundo en caso de obtener cruz, y *c)* numerar los 10 elementos y obtener 5 números aleatorios. El método *a* no es probabilístico, ya que los últimos 5 niños del listado no podrán salir en la muestra. El método *b* es equiprobabilístico porque todos los niños tienen la misma probabilidad de salir en la muestra, pero no es aleatorio simple, ya que sólo dos muestras pueden ser elegidas (existen subconjuntos de 5 elementos que no podrían ser seleccionados). El método *c* es aleatorio simple, ya

que cualquiera de los 252 subconjuntos diferentes de 5 niños puede ser seleccionado y con la misma probabilidad.

Como se deduce del ejemplo anterior, el primer paso es preparar una lista de las unidades de muestreo, numerándolas, por ejemplo, de forma secuencial. Este proceso puede ser fácil y rápido si la población de referencia se encuentra en una base de datos informatizada, o muy laborioso si se realiza manualmente y la población es amplia. A continuación, se seleccionan tantos números aleatorios como elementos debe tener la muestra. La selección de números aleatorios suele realizarse con un ordenador o una calculadora de mano.

Esta técnica es sencilla y de fácil comprensión, y permite el cálculo rápido de medias y variancias. Requiere, sin embargo, que se conozca de antemano la lista completa de todas las unidades de muestreo. Si la lista contiene fallos u omite segmentos de la población, se perderán las ventajas propias del muestreo aleatorio, ya que todas las unidades no tendrán la misma probabilidad de aparecer en la muestra.

Muestreo aleatorio estratificado

Se trata de una modificación del método anterior que intenta asegurar que la muestra presente la misma distribución que la población en relación con determinadas variables.

La población se divide en estratos en función de las categorías de las variables por las que se desea estratificar, formando subgrupos con alguna característica en común y mutuamente excluyentes. A continuación, se escoge una muestra aleatoria de cada estrato, manteniendo las proporciones observadas en la población de referencia (*muestreo aleatorio estratificado proporcional*).

Los estratos deben definirse en función de variables que puedan influir sobre los resultados. Si se desea determinar la prevalencia de HTA en una comunidad, posibles factores de confusión serían, entre otros, la edad, el sexo o la presencia de obesidad, ya que son variables relacionadas con la presencia de HTA. Si se aplicara un muestreo aleatorio simple, podría obtenerse una muestra en la que la distribución de estas variables no fuera la misma que en la población, lo que conduciría a un sesgo en los resultados. Un muestreo estratificado por estos factores garantizaría una distribución similar a la de la población de referencia.

Ejemplo 16.5. Supongamos un marco de muestreo de 2.000 personas en las que se desea estimar la prevalencia de HTA, estudiando para ello una muestra de 250 sujetos. Se conoce que la prevalencia de HTA varía en función del sexo y la presencia de obesidad. Por tanto, los investigadores desean que la muestra en que se realizará el estudio refleje fielmente la distribución, según el sexo y la presencia de obesidad que tiene la población de referencia, que se muestra en la tabla 16.1. Así, por ejemplo, existe un 12% de hombres obesos, por lo que, de los 250 sujetos de la muestra, un 12% deberá tener estas características. Por tanto, de los 240 hombres obesos de la población, deben seleccionarse aleatoriamente 30 para formar parte de la muestra. Se procede de la misma forma con los otros estratos (tabla 16.1).

La correcta aplicación de esta técnica requiere que se conozca la distribución en la población de la(s) variable(s) empleada(s) para la estratificación y que la variabilidad entre los distintos estratos sea lo más amplia posible, mientras que dentro de cada estrato debe ser lo más pequeña posible. Así, se obtendrán subgrupos homogéneos respecto a la posible variable

Tabla 16.1 Muestreo estratificado (ejemplo 16.6)

Distribución de una población de 2.000 personas en función del sexo y la obesidad

	Varones	Mujeres	Total
Obesidad	240 (12%)	320 (16%)	560 (28%)
No obesidad	960 (48%)	480 (24%)	1.440 (72%)
Distribución de una muestra de 250 personas que mantiene la misma distribución en función del sexo y la obesidad que la población general			
	Varones	Mujeres	Total
Obesidad	30 (12%)	40 (16%)	70 (28%)
No obesidad	120 (48%)	60 (24%)	180 (72%)

de confusión (todos los miembros de un estrato la presentan o no) y heterogéneos entre sí (miembros de distintos estratos no comparten la variable).

En general, con el muestreo estratificado se obtienen estimaciones más precisas que con el muestreo aleatorio simple, porque la variancia total se basa en la de cada uno de los estratos, y éstos son más homogéneos que la población. El aumento de precisión que se consigue depende de las diferencias entre los estratos. Cuanto mayor sea esta diferencia, y menor la variabilidad dentro de cada estrato, mayor será la reducción de la variancia que se produce. Por tanto, siempre que se utiliza un muestreo estratificado, debe tenerse en cuenta en el análisis, al realizar las estimaciones.

Cuando alguno de los estratos tiene un tamaño reducido, puede desaparecer esta ventaja, ya que disminuiría la precisión de la estimación en dicho estrato y podría afectar a la de la estimación global. En esta situación, puede aplicarse un *muestreo no proporcional* que sobrerrepresente este estrato, aplicando sobre él una fracción de muestreo diferente. Esta alternativa dificulta el análisis, ya que la estimación del parámetro en el global de la muestra deberá hacerse a partir de cada estrato por separado, y ponderando la diferente variabilidad de cada uno de ellos. De hecho, el muestreo no proporcional se utiliza más cuando la finalidad es realizar comparaciones entre los estratos, con el fin de asegurar que cada uno de ellos tiene el tamaño suficiente para obtener estimaciones con la suficiente precisión.

Ejemplo 16.6. Supongamos que se lleva a cabo una encuesta entre los médicos de atención primaria para conocer su opinión sobre un determinado problema de salud. La distribución de médicos según su ámbito de trabajo es la siguiente: 230 (56%) trabajan en el medio urbano, 160 (39%) en el semiurbano y 20 (5%) en el rural. Si se seleccionara una muestra de 100 médicos y se mantuvieran las mismas proporciones, sólo se conseguirían las respuestas de 5 médicos rurales, lo cual sería insuficiente para conocer su opinión y compararla con la de otros grupos. Los investigadores deciden aplicar diferentes fracciones de muestreo en cada uno de los estratos (urbanos, 20%; semiurbanos, 30%; rurales, 100%), con lo que la distribución de encuestados es de 46, 48 y 20, respectivamente. Hay que tener en cuenta las diferentes fracciones si se quieren realizar estimaciones globales de toda la muestra, lo que añade cierta dificultad al análisis. Si se tiene la intención de realizar un análisis sólo de los subgrupos por separado, no es necesario tener en cuenta las diferentes fracciones de muestreo, porque cada estrato puede considerarse una población diferente.

Muestreo en múltiples etapas

Esta técnica consiste en seleccionar unidades de muestreo de una población (unidades primarias) y, en una segunda etapa, obtener una muestra de cada una de las unidades primarias (unidades secundarias). Por ejemplo, si se desea estudiar la prevalencia de escoliosis en niños de edad escolar, una manera de seleccionar la muestra sería elegir al azar escuelas y, dentro de cada una, una muestra de niños. Se pueden usar el número de etapas que sean necesarias y, en cada una de ellas, aplicar un método diferente de muestreo (simple, estratificado, sistemático). Cuando en el estudio se incluyen todas las unidades secundarias (es decir, la fracción de muestreo es del 100%), se denomina *muestreo en conglomerados*.

Esta técnica es útil cuando la población de referencia es muy grande y está dispersa, como en el caso de la población escolar de un país o una gran ciudad. El uso de una técnica aleatoria simple sería muy costoso, ya que la muestra incluiría a pocos estudiantes de cada institución y los investigadores tendrían que ir a casi todas ellas para entrevistarlos o examinarlos, con lo que se rentabilizaría muy poco cada viaje. Además, es probable que no se disponga de un listado de todos los niños en edad escolar de una ciudad, pero sí de una lista de escuelas, y que cada una de ellas tenga el listado de sus alumnos.

Su desventaja principal es que si las unidades primarias contienen personas similares en relación con el fenómeno de interés, la precisión de la estimación será menor que si se utiliza un muestreo aleatorio simple, por ejemplo. Si se considera un estudio sobre los hábitos alimentarios de una comunidad y se encuesta a 100 individuos obtenidos mediante un muestreo aleatorio simple, el grado de información y la precisión de las estimaciones no será igual que si se estudian 100 individuos (unidad secundaria) procedentes de 25 familias (unidad primaria). Los hábitos alimentarios son muy parecidos entre miembros de una misma familia, por lo que, en realidad, es casi como si sólo se estudiara a 25 personas. Además, si una o varias familias son muy numerosas, contribuirán de una forma desproporcionada al tamaño de la muestra.

Otro inconveniente es que la variabilidad es mayor que las técnicas anteriores, y su cálculo es complejo. Una solución parcial es extraer una muestra de mayor tamaño y considerarla como aleatoria simple, y procurar que las unidades de muestreo sean lo más pequeñas posible.

Ejemplo 16.7. Un estudio tenía por objetivo conocer las creencias y opiniones de los escolares respecto al tabaco, el alcohol y el ejercicio físico (Comín et al, 1989).

Se seleccionó una muestra representativa de escolares de la ciudad de Barcelona. La técnica de muestreo empleada fue la de conglomerados en varias etapas. Se tomó como unidad de muestreo el aula, estratificando por las siguientes variables: curso, tipo de escuela (pública, privada) y tamaño del curso (número de aulas). Es importante remarcar que todos estos factores están relacionados con las variables estudiadas; el curso se asocia a los hábitos de vida a través de la edad, y el tipo de escuela, a través de la clase social. La estratificación permite obtener estimaciones más precisas. El empleo del muestreo por conglomerados es prácticamente inevitable al estudiar la población escolar. El aumento de la variabilidad que acompaña a la utilización de conglomerados se minimiza al tomar el aula, en lugar del colegio, como unidad primaria de muestreo, y se compensa parcialmente con la estratificación.

Muestreo sistemático

Este procedimiento se basa en aplicar alguna regla sistemática simple, como elegir uno de cada n individuos. Los pasos a seguir son los siguientes: en primer lugar, se calcula la *constante de muestreo* (k), dividiendo el total de la población candidata por el tamaño de la muestra deseado. A continuación, se extrae la primera unidad al azar entre las k primeras unidades de muestreo, y se le suma la constante, sucesivamente, hasta completar el tamaño de la muestra.

Ejemplo 16.8. Supongamos que deseamos obtener una muestra de 200 historias clínicas de un archivo de 1.000 historias. La constante de muestreo es $1.000/200 = 5$, lo que significa que se escoge una de cada 5 historias. La primera se elige de forma aleatoria entre los números del 1 al 5. Si la elegida es la 2, la siguiente es la 7 ($2 + 5$), después la 12 ($7 + 5$), la 17 ($12 + 5$), y así sucesivamente, hasta conseguir las 200 historias necesarias.

La primera ventaja del muestreo sistemático es que es fácil de aplicar y, en ocasiones, es más cómodo y práctico que el muestreo aleatorio simple. Es más sencillo, por ejemplo, escoger todas las historias cuyo número termina en cero que a partir de un listado de números aleatorios. La segunda es que no siempre es necesario tener de antemano una lista completa y exhaustiva de toda la población. Los sujetos se pueden ir seleccionando a medida que el estudio avanza, siempre y cuando lleguen en algún orden preestablecido, como por ejemplo, los pacientes que acuden a una consulta. La tercera ventaja es que, cuando la población de referencia

está ordenada siguiendo una tendencia conocida (de mayor a menor, de más viejo a más joven, etc.), el muestreo sistemático asegura una cobertura de unidades de todos los tipos. Si interesa extraer una muestra de los médicos de una provincia y el colegio profesional tiene las listas ordenadas por el año de licenciatura, podría darse el caso de que con un muestreo aleatorio simple predominaran los médicos de una determinada edad. Sin embargo, con un muestreo sistemático, se aseguraría una muestra que representara todos los años de licenciatura.

La limitación más importante para su uso aparece cuando la constante que se aplica está relacionada con el fenómeno que se quiere medir. En este caso se pueden hallar estimaciones muy distorsionadas de los parámetros poblacionales.

Ejemplo 16.9. Supongamos que se quiere estudiar la actividad asistencial de un equipo de atención primaria y que la unidad de muestreo son los días de la semana. Si el primer día seleccionado fuera un lunes y se aplicara una constante ($k = 7$), se recogería únicamente la actividad realizada los lunes de cada semana. Con este procedimiento se sobrestimaría la media real de visitas, ya que tradicionalmente los lunes son los días de mayor actividad. Como puede deducirse, estas situaciones son muy poco frecuentes y, en todo caso, es relativamente fácil intuir que la constante de muestreo está relacionada con el fenómeno de estudio.

Las ventajas y desventajas de cada uno de los métodos de muestreo probabilístico se encuentran en la *tabla 16.2*.

Muestreo no probabilístico

En el muestreo no probabilístico se desconoce la probabilidad que posee cada unidad de ser incluida en la muestra, y la selección se realiza mediante métodos en los que no interviene el azar.

El uso de pruebas estadísticas requiere que la muestra analizada sea aleatoria. Su aplicación en muestras no obtenidas por un método aleatorio se hace asumiendo que están libres de sesgos y que son representativas de la población de referencia.

La técnica más frecuente es el *muestreo consecutivo*. Consiste en seleccionar a los pacientes que cumplen los criterios de selección especificados en el protocolo del estudio, a medida que acuden a la consulta durante un período determinado.

Ejemplo 16.10. Lee et al (2002) compararon las características de 1.546 pacientes que acudieron consecutivamente a un centro de salud con las de

Tabla 16.2 Ventajas y desventajas de las principales técnicas de muestreo probabilístico

Ventajas	Desventajas
Muestreo aleatorio simple	
<ul style="list-style-type: none"> – Sencillo y de fácil comprensión – Cálculo rápido de medias y variancias 	<ul style="list-style-type: none"> – Requiere que se posea de antemano un listado completo de toda la población – Cuando se trabaja con una muestra pequeña, es posible que no sea representativa de la población
Muestreo aleatorio estratificado	
<ul style="list-style-type: none"> – Tiende a asegurar que la muestra representa adecuadamente a la población en función de unas variables determinadas – Se obtienen estimaciones más precisas – Se pueden aplicar distintas fracciones de muestreo en cada estrato 	<ul style="list-style-type: none"> – Se ha de conocer la distribución en la población – Requiere que se posea de antemano un listado completo de toda la población – Cálculo de las estimaciones más complejo que en el muestreo aleatorio simple
Muestreo en etapas múltiples	
<ul style="list-style-type: none"> – Es muy eficiente cuando la población es muy grande y dispersa – La necesidad de listas de las unidades de una etapa se limita a aquellas unidades de muestreo seleccionadas en la etapa anterior 	<ul style="list-style-type: none"> – Las estimaciones son menos precisas que con el muestreo aleatorio simple o estratificado – El cálculo de las estimaciones es complejo
Muestreo sistemático	
<ul style="list-style-type: none"> – Fácil de aplicar – No siempre es necesario disponer de una lista de toda la población – Cuando la población está ordenada siguiendo una tendencia conocida, asegura una cobertura de unidades de todos los tipos 	<ul style="list-style-type: none"> – Si la constante de muestreo está asociada con el fenómeno de interés, se pueden obtener estimaciones sesgadas

la población de referencia. Observaron que los pacientes incluidos en la muestra consultaban con mayor frecuencia, eran mayores, con peor estado de salud (mayor prevalencia de HTA, accidente cerebrovascular y cáncer), menos fumadores y vivían solos, con mayor frecuencia.

El muestreo consecutivo es la técnica más utilizada en los ensayos clínicos. Si se evalúa la eficacia de una medida terapéutica contra la gastroenteritis, no se dispone al inicio del estudio de un listado de la población que desarrollará la enfermedad, por lo que no puede aplicarse un muestreo aleatorio simple. En esta situación, el equipo investigador se limita a reclutar a los individuos que cumplen los criterios de selección a medida que son diagnosticados.

La inclusión de *voluntarios* es otra técnica de muestreo no probabilístico. En general, es preferible rechazar su colaboración, ya que las verdaderas razones que llevan a una persona a prestarse para un estudio sin haber sido seleccionado para ello suelen estar relacionadas con otras características que afectan a la representatividad de la muestra.

Otra técnica utilizada en encuestas de opinión y de mercado, pero muy poco en la investigación epidemiológica, es el *muestreo por cuotas*, en el que la composición general de la población de referencia en términos de características como la edad, el sexo, o la clase de educación se decide, o se conoce, de antemano. Seguidamente, se determina el número de personas requerido, o las cuotas, para los hombres y las mujeres, según las diferentes edades y la clase de educación, y el único requerimiento es conseguir el número adecuado de personas para llenar cada una de las cuotas.

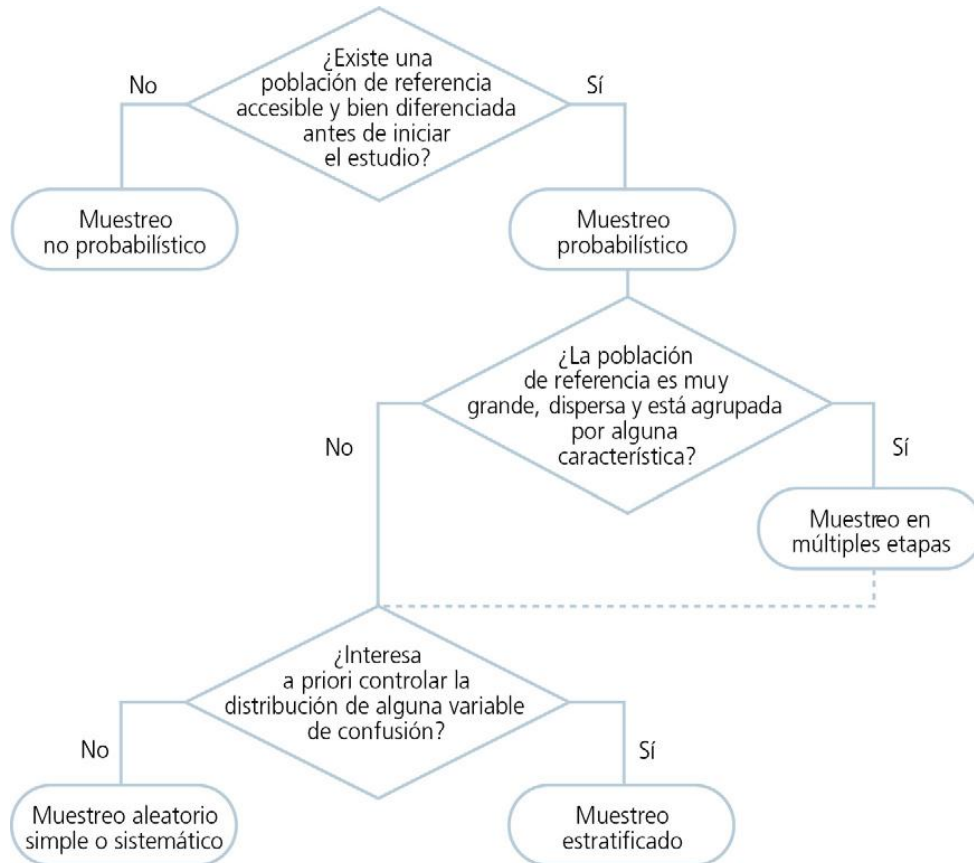
En algunas ocasiones, en especial cuando tanto el tamaño de la población de referencia como el de la muestra son muy pequeños, la selección puede hacerse por criterios razonados. Por ejemplo, si se desea conocer el funcionamiento de los centros de salud de una provincia a través de un análisis exhaustivo de dos de ellos, es preferible seleccionar los dos centros que puedan considerarse los más representativos, en vez de dejar la selección al azar.

En las *técnicas adaptativas* el diseño del muestreo se adapta en función de las observaciones que se

van realizando durante el estudio, por ejemplo, solicitando a los participantes en un estudio sobre consumo de drogas que informen al investigador sobre otros consumidores. Dependiendo del mecanismo de adaptación, reciben nombres como *bola de nieve*, *paseos aleatorios* o *muestreo en red*. Estas técnicas son especialmente útiles cuando la población de interés es difícil de alcanzar, ya que, además de la

ganancia que suponen en términos de eficiencia en relación con las técnicas convencionales, pueden ser la única forma operativa de obtener una muestra lo suficientemente amplia. Su principal inconveniente es que deben aplicarse procedimientos estadísticos que tengan en cuenta su utilización.

La figura 16.1 presenta un algoritmo para la selección de la técnica de muestreo.



Las preguntas son orientativas y han de realizarse en función de las ventajas y desventajas de cada técnica

Figura 16.1 Algoritmo para la selección de la técnica de muestreo.

BIBLIOGRAFÍA DE LOS EJEMPLOS

Andersen LB, Henckel P, Saltin B. Risk factors for cardiovascular disease in 16-19 year old teenagers. *Int J Med*. 1989;225:157-63.

Comín E, Nebot M, Villalbí JR. Exercici i consum de tabac i alcohol dels escolars de Barcelona. *Gac Sanit*. 1989;3:355-65.

Fuchs SC, Petter JG, Accordi MG, Zen VL, Pizzolo Jr AD, Moreira

LB, et al. Establishing the prevalence of hypertension. Influence of sampling criteria. *Arq Bras Cardiol*. 2001;76:445-52.

Lee ML, Yano EM, Wang M, Simon BF, Rubenstein LV. What patient population does visit-based sampling in primary care settings represent? *Med Care*. 2002;40:761-70.

Luty SE, Joyce RR, Mulder RT, McKenzie JM. Social adjustment in depression: the impact of depression severity, personality, and clinic versus community sampling. *J Affect Disord*. 2002;70:143-54.

Silva Ayçaquer LC. Muestreo para la investigación en ciencias de la salud. Madrid: Díaz de Santos; 1993.

BIBLIOGRAFÍA

Armitage P, Berry G, Matthews JNS. Statistical methods in medical research. 4th ed. Oxford: Blackwell Science; 2002.

Kelsey JL, Wittmore AS, Evans A, Thompson WD. Methods in observational epidemiology.

2nd ed. New York: Oxford University Press; 1996.

Phillips CV, LaPole LM. Quantifying errors without random sampling. BMC Med Res Methodol. 2003;3:9.

Rodríguez Osuna J, Métodos de muestreo. Colección «Cuadernos

metodológicos» núm. 1. Madrid: Centro de Investigaciones Sociológicas; 1991.

Thompson SK, Collins LM. Adaptive sampling in research on risk-related behaviors. Drug Alcohol Depend. 2002;68(Suppl 1):S57-67.