

Población de estudio y muestreo en la investigación epidemiológica

A lo largo de los números anteriores hemos ido abordando las diferentes etapas de un proceso de investigación para acabar centrándonos en los diferentes tipos de estudios epidemiológicos, en los próximos capítulos abordaremos los aspectos que debemos tener en cuenta a la hora de abordar un diseño de muestreo.

La pregunta de investigación que nos hemos planteado es el núcleo central de la investigación indicándonos los posibles diseños de investigación que podemos seguir y delimitándonos cual es nuestra población de estudio.

Niveles de población

Podemos definir población como un agregado total de casos que cumple con una serie predeterminada de criterios, por lo que cuando hablamos de población no nos referimos de forma exclusiva a seres humanos sino que también podemos referirnos a historias de un hospital determinado, escuelas de enfermería, etc,

A la población de estudio se le conoce como **población diana** o **población blanco** y viene delimitada por características demográficas, sociales, hábitos de vida, problemas de salud, etc. De esta manera nuestra población de estudio puede ser mujeres en edad fértil, cuidadores de ancianos, usuarios de drogas por vía parenteral, mujeres con cáncer de mama o institutos públicos de secundaria.

En toda investigación la población diana es el universo al que queremos generalizar nuestros resultados. Dentro de esta población diana se encuentra la **población accesible** que consta del conjunto de casos que satisfacen los criterios predeterminados y que al mismo tiempo son accesibles para el investigador; Por lo que la población accesible viene determinada por consideraciones prácticas en función de la accesibilidad que tengamos a los sujetos (existencia de registros, circunstancias que faciliten la colaboración, etc.). En una investigación en la que nuestra población diana sean mujeres embarazadas nuestra población accesible puede ser mujeres embarazadas que acudan a los controles prenatales en un área determinada.

Al identificar una población el investigador debe especificar los criterios que definen quienes deben incluirse, estos criterios deben estar bien especificados ya que mediante ellos se decidirá si un individuo se incluirá o no en la población de interés. Estos criterios de selección nos van a delimitar a la **población elegible**.

Los criterios que especifican las características que la población debe tener se denominan criterios de elegibilidad o criterios de inclusión (por ejemplo tener más de 18 años, tener personas mayores de 65 años a su cargo, etc.).

A veces, una población se define en función de las características que sus miembros no deben tener, para ello se establecen criterios de exclusión (por ejemplo pueden excluirse personas que no hablen español)

Los criterios de selección suelen reflejar alguno de los siguientes aspectos:

- Costo: algunos criterios reflejan restricciones de costo. Por ejemplo si excluimos a personas que no hablan español no significa que no nos interese ese colectivo sino que no tenemos posibilidades de contratar traductores.
- Problemas de orden práctico: problemas para incluir a personas de zonas rurales.
- Posibilidad de participar en un estudio: puedes ser necesario excluir de un estudio a personas con una neoplasia en fase terminal.
- Consideraciones de diseño ya que a veces es adecuado definir una muestra más homogénea para controlar variables extrínsecas.

En la mayor parte de las investigaciones no estudiamos a toda la población elegible sino que tomamos una muestra de esa población. Debemos por tanto considerar que el investigador suele obtener una muestra de la población elegible con miras a generalizar los resultados a la población diana, por lo que en el paso de un nivel a otro de población (que hemos esquematizado en la tabla 1) debemos siempre preguntarnos si existen características que puedan hacer diferente a una población de otra, y si es así, ver si esas características pueden influir en los resultados de la investigación.

El último nivel de población serían los participantes en el estudio, ya que en la mayor parte de las ocasiones los sujetos que participan en el estudio no suelen coincidir exactamente con la muestra diseñada debido a pérdidas en el reclutamiento, rechazo a colaborar, etc.



Tabla 1. Niveles de población

Fuente: Rebagliato M, Ruiz I, Arranz M, *Metodología de investigación en epidemiología*. Edic. Díaz de Santos, Madrid, 1996.

Muestreo

El término muestreo se refiere al proceso de selección de una parte o subconjunto de la población para que represente al conjunto, por tanto una muestra es una parte o subconjunto de la población que mediante las técnicas estadísticas nos permiten a partir de los datos muestrales inferir resultados a la población. Las unidades que constituyen las muestras y las poblaciones se conocen como elementos o unidades de análisis. El elemento es pues la unidad básica acerca de la cual se recaba información.

Al no estudiar a la población en su conjunto, la muestra siempre nos va a dar menos información que si tuviésemos a la población total. La razón para elegir una muestra y no toda la población es porque somos incapaces de estudiar a toda la población ya sea por cuestión de tiempo, dinero u otros recursos.

El principal criterio para evaluar la muestra de un estudio epidemiológico es "la representatividad", el objetivo del diseño de muestreo es conseguir muestras **representativas** es decir muestras cuyas características claves se aproximen estrechamente a las de la población o dicho de otra manera que la muestra que hemos elegido sea una replica en miniatura de la población de la que la hemos obtenido y que por tanto tenga el mismo grado de diversidad que la población.

Por desgracia no hay un método que garantice que la muestra sea representativa sin obtener información de toda la población, por lo que el concepto de representatividad muestral es un concepto intuitivo, no sabemos si la muestra es representativa, puesto que la representatividad esta vinculada con el grado de variabilidad de la población, lo que sabemos es que los diseños de muestreo en los que interviene el azar produce en la mayor parte de las veces muestras representativas aunque no garantizan la representatividad.

Las técnicas estadísticas derivadas de la teoría de muestreo nos permiten, a partir de los datos obtenidos en la muestra, estimar los valores en la población de la que hemos obtenido la muestra, para ello debemos hacer un plan de muestreo que nos indique el cómo y el cuánto, es decir, cómo vamos a seleccionar a los individuos y qué tamaño tiene que tener nuestra muestra, las diferentes formas de seleccionar las unidades que forman nuestra muestra es lo que se conoce como diseños de muestreo.

Los diseños de muestreo se clasifican en probabilísticos y no probabilísticos

- Diseños probabilísticos son aquellos en los que se utiliza algún sistema de selección aleatoria para garantizar que cada unidad de la población tenga una probabilidad específica de ser seleccionada, por lo que toda unidad tiene una probabilidad de ser elegida y esa probabilidad es conocida de antemano.

Las principales técnicas probabilísticas son:

- Muestreo aleatorio simple
 - Muestreo sistemático
 - Muestreo estratificado
 - Muestreo por conglomerados
- Diseño no probabilístico donde las unidades se escogen utilizando métodos en los que no interviene el azar, de modo que no es posible estimar la probabilidad de que cada elemento haya sido incluido en la muestra y no todos los elementos tiene la posibilidad de ser incluidos.

Las técnicas más utilizadas son:

- Muestreo consecutivo
- Muestreo de conveniencia
- Muestreo a criterio

Los investigadores trabajan con muestras y no con poblaciones por cuestiones económicas y de eficiencia, sin embargo debemos de ser conscientes que la muestra nos da menos información que la población total y que los datos provenientes de una muestra pueden conducir a conclusiones erróneas, uno de los motivos por los que puede ocurrir esto es porque exista un sesgo de muestreo.

El sesgo de muestreo consiste en la representación sistemáticamente excesiva o deficiente de un sector de la población en la muestra según un atributo importante para el problema que se investiga. En la mayor parte de las ocasiones el investigador no es consciente de la existencia del sesgo. Si en una investigación con los estudiantes de una escuela de enfermería se entrevista sistemáticamente a uno de cada diez que entra a la biblioteca de la escuela la muestra estaría sesgada a favor de los que acuden regularmente a la biblioteca.

El grado en que el sesgo de muestreo puede ocasionar dificultades posteriores depende de la homogeneidad de la población respecto de las variables que se investigan, si la variable fuese totalmente homogénea bastaría con una unidad de análisis para llegar a conclusiones validas acerca de toda la población, sin embargo esto no es lo que suele ocurrir en los variables que estudiamos donde la heterogeneidad es un factor prácticamente constante en todas las variables (edad, estado civil, ingresos, nivel de dependencia, etc.). Para evitar la presencia de sesgos es importante reflexionar sobre el diseño de muestreo que vamos a utilizar y contestarnos a la pregunta ¿tenemos indicios que nos indiquen que la muestra que vamos a elegir tenga un menor grado de diversidad que la población de la que la hemos obtenido? o dicho de otra manera ¿tenemos indicios para sospechar que la muestra no sea representativa de la población de la que la hemos obtenido? si es así deberemos modificar nuestro diseño de muestreo para evitar generalizaciones erróneas a la población.

En el próximo capítulo abordaremos los diferentes diseños de muestreo.

Bibliografía consultada:

1. Polit, Hungler, *Investigación científica en ciencias de la salud*, 5ª edición, Edit. McGraw-Hill Interamericana, México 2002.
2. Rebagliato M, Ruiz I, Arranz M, *Metodología de investigación en epidemiología*. Edit. Díaz de Santos, Madrid, 1996.
3. Silva LC, *Muestreo para la investigación en ciencias de la salud*.
4. Edit. Díaz de Santos, Madrid, 1993.