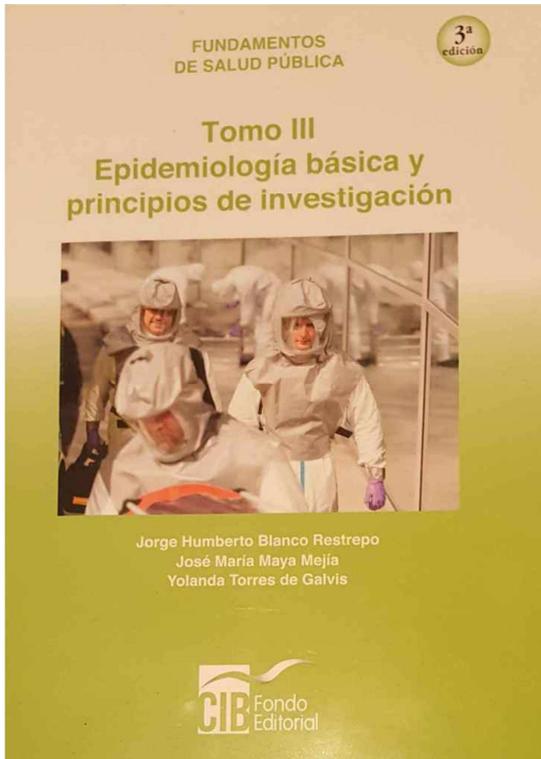


MUESTREO Y SU EMPLEO EN LA INVESTIGACIÓN EPIDEMIOLÓGICA

Dr. Luis Carlos Silva Ayçaguer, Escuela Nacional de Salud Pública de Cuba



Capítulo 6 del libro "Epidemiología básica y principios de investigación" Tomo III, Blanco, Maya y Torres Eds., 3ª edición, Medellín: CIB Fondo Editorial (COLOMBIA); 2016.

INTRODUCCIÓN

El muestreo es una disciplina estadística orientada a dar solución a un problema virtualmente ineludible, cualquiera sea el problema de investigación epidemiológica: establecer qué parte de una realidad ha de examinarse para poder realizar inferencias confiables.

Salvo que se trate de un estudio teórico, de una revisión bibliográfica sistemática o de un estudio que maneje información secundaria (ya procesada por servicios estadísticos centrales, por ejemplo), en los estudios epidemiológicos será siempre necesario recoger información

empírica, realizar mediciones o interrogatorios. En la mayoría de las situaciones prácticas, razones de factibilidad o de financiamiento imposibilitan la recogida de información sobre todos los miembros de la población acerca de los cuales nos estamos formulando determinadas preguntas.

La necesidad de encarar un problema epidemiológico valiéndose de la observación de una parte del entorno en que se produce, entonces, se presenta con extrema frecuencia. Dicho en términos más simples: para responder a las preguntas que usualmente se formula la epidemiología, con pocas excepciones, hay que acudir a la selección de muestras.

Las preguntas de investigación conciernen, ya sea de manera explícita o implícita, a lo que regularmente llamamos *universo* o *población*: un conjunto de unidades de análisis que por lo general, aunque no necesariamente, son personas.

La definición del término *muestra* que figura en el Diccionario de la Real Academia Española reza así: "Parte o porción extraída de un conjunto por métodos que permiten considerarla como representativa de él".¹ Este es una de los tantos ejemplos que prueban que la docta institución espera por una mejoría importante en materia científica. Una muestra es, por definición, un subconjunto cualquiera de dicho universo. Los métodos pueden ser "malos" o "buenos" a los efectos de conseguir representaciones adecuadas del conjunto inicial, pero esa capacidad no debe integrar la definición de *muestra*, del mismo modo que no exigimos a una mujer que quiera mucho a su hijo para considerar que es una "madre". Precisamente, el papel del muestreo como disciplina es proveer a los investigadores de recursos para que las muestras que obtengan sean "buenas" muestras. De eso trata el presente capítulo.

La necesidad de operar con ellas se deriva claramente de que resulta literalmente imposible estudiar a todos los individuos que integran una población, como se ha dicho. Es importante, sin embargo, comprender que, incluso cuando la realización de un censo fuera factible, el empleo de muestras suele ser conveniente. Dos razones fundamentan tal afirmación: que a los efectos de la mayoría de las tareas investigativas, la calidad del proceso no se incrementa una vez alcanzado un tamaño mínimo necesario y, por otra parte, que cuanto mayor sea la muestra, mayor será el riesgo de que la calidad de los datos obtenidos sea deficiente.

De lo que se trata en síntesis, es de garantizar que el proceso *inductivo* que supone obtener conclusiones sobre una realidad sobre la base de una porción de dicha realidad, pueda llevarse delante de manera fructífera.

Como es bien conocido, un eje taxonómico esencial en el marco epidemiológico es el que clasifica los estudios según su propósito: en *descriptivos* (los que pretenden caracterizar cómo es una realidad) y *analíticos* o *explicativos* (los que procuran determinar por qué esa realidad es como es). Los estudios de este último tipo, sean experimentales o de intervención, u observacionales (de cohorte por una parte, y de casos y controles por otra) tienen demandas muy específicas en relación con las muestras, las cuales son objeto de examen en otros capítulos del presente libro. Consecuentemente, las técnicas de muestreo en las que se pone el énfasis en este capítulo son aquellas que procede aplicar específicamente ante preguntas de investigación que apuntan en dirección *descriptiva*.

El diseño de una muestra adecuada con este fin, supone el cumplimiento de tres tareas fundamentales, a saber:

- Determinar el método de selección de la muestra
- Fijar el tamaño adecuado que dicha muestra debe alcanzar
- Establecer los procedimientos que habrán de aplicarse a los datos así obtenidos

Esas son en esencia las encomiendas básicas que el muestreo, como parte del método científico, ha de resolver o ayudar a resolver. La solución correcta de dichas tareas exige poseer conocimientos y estar en condiciones de emplear recursos bastante especializados, tales como manejar la teoría de probabilidades, dominar aplicaciones informáticas relativamente complejas y desplegar delicadas técnicas de encuestas, entre otros. En el cuerpo de este capítulo, iremos considerándolas, por ese orden. Pero el abordaje de dichos temas no puede sino ser epidérmico debido a la naturaleza especializada del muestreo. Se intenta, sin embargo, recorrer los conceptos fundamentales que necesita el lector para tener un intercambio inteligible con los especialistas y facilitar la lectura de materiales más avanzados, en los cuales se basa la presente síntesis. De ellos, el que directamente se sugiere como posible fuente de profundización y que ha sido más empleado para la presente exposición es el libro titulado "Diseño razonado de muestras y captación de datos para la investigación sanitaria".²

1.1 Determinar el método de selección de la muestra. Población y parámetros.

Supongamos que estamos ante una *población* finita cualquiera; o sea, un conjunto de entes susceptibles, al menos teóricamente, de ser listados.

Las preguntas que intentan responder los epidemiólogos conciernen por lo general a poblaciones bien definidas, tales como los sujetos que han padecido un infarto en los últimos dos años, los escolares que residen en una comunidad, los perros callejeros en una ciudad, las consultas estomatológicas ofrecidas durante un año a pacientes en los servicios de una clínica dental.

Nótese que el término “población” se emplea en un sentido amplio, no se trata necesariamente de grupos humanos, como ocurre con los dos últimos ejemplos mencionados. Se denomina *unidades de análisis* a los elementos o sujetos que componen la población y que habrían de ser objeto de interrogatorio o medición.

La fuente informativa de la cual se obtendrá la muestra se conoce en el marco del muestreo como *marco muestral*. Para fijar ideas, supongamos que se estudia una población de adultos de un barrio con vistas a explorar sus conocimientos sobre hipertensión arterial. Cada adulto es una unidad de análisis, el ente que se examina. Si se tiene una lista de todos los adultos residentes en el barrio, este propio listado es el marco muestral. Muchas veces, tal lista no existe ni se puede conseguir. Pero, en lugar de contar con un listado de adultos, puede tenerse uno de las manzanas del barrio, algunas de las cuales se elegirán para el estudio. En ese caso, las unidades primarias de muestreo serán las manzanas. Si, una vez hecha esta selección, se toman sólo algunos adultos de las manzanas elegidas para integrar la muestra final, entonces habría dos procesos sucesivos de selección. Existen en este caso las llamadas unidades de muestreo de primera, y de segunda etapa (manzanas, y adultos respectivamente). El conjunto de listas necesarias para completar este esquema de selección (primero de manzanas, y luego de adultos dentro de cada una de las manzanas seleccionadas) constituyen el marco muestral del estudio.

Supongamos también que se quiere conocer una característica general o *parámetro*, definido por el resultado numérico de "combinar" de alguna manera prefijada los datos que se obtendrían en el supuesto de que se midieran todos los elementos de la población. Tal sería el caso, por ejemplo, de un porcentaje, una media poblacional o del número total de individuos que comparten cierto rasgo. Por ejemplo, pudiera ser el porcentaje de adultos de la comunidad que poseen suficiente información acerca de la hipertensión, o el gasto promedio de los hipertensos en medicamentos para mantener bajo control sus cifras tensionales, o de la edad mediana de las embarazadas que padecen hipertensión en la población.

Cabe aclarar que, aunque aquí, con el fin de simplificar la exposición, nos estamos refiriendo a un solo parámetro, en la práctica regular lo típico es no haya un único parámetro que se quiere estimar, sino varias decenas de ellos. Por ejemplo, en un estudio sobre riesgos laborales, el objetivo central puede que sea estimar el porcentaje de sujetos que se han accidentado en el último quinquenio, pero difícilmente el interés se ciña al conocimiento de ese número; típicamente, se deseará obtener tablas que contemplen aspectos tales como ese mismo dato por grupos de edad, el número de trabajadores que laboran sin los debidos equipos de protección, las profesiones más expuestas a daños laborales, etc.

Si en lugar de proceder a obtener los datos necesarios para todas y cada una de las unidades de la población de interés (todos los adolescentes de la comunidad, la totalidad de los enfermeros del país o todas las niñas y adolescentes de la provincia), se opta por examinar solo una parte de tales unidades, y estimar el número desconocido a partir de dicha información, entonces decimos que se ha planteado un *problema de muestreo* y, junto con él, un *problema de estimación*.

1.2 Representatividad y azar

La noción de *representatividad* sólo tiene un alcance intuitivo. No existe una definición formal que permita establecer si una muestra concreta es o no representativa, ni estimar siquiera su grado de representatividad. Esta noción, sin embargo, integra el repertorio de nuestras intuiciones desde muy temprano, y es tan familiar que resulta muy fácil hallar ejemplos de la vida cotidiana que ilustren su empleo.

En cualquier caso, la idea que mejor sintetiza el concepto de representatividad muestral es la siguiente: para conseguirla, lo que debe procurarse es que la muestra exhiba internamente el mismo grado de diversidad o variabilidad que la población. De hecho, cuando “sentimos” que una muestra no es representativa de la realidad es precisamente porque sabemos que dicha realidad es más variable que la muestra. Por ejemplo, los médicos colombianos constituyen una muestra de individuos de esta nacionalidad, pero no representa adecuadamente a todos ellos porque dicha muestra es mucho más uniforme (al menos respecto de muchísimas variables, tales como salario, hábitos alimentarios, zona de residencia, estrato al que pertenecen, etc.) que la población general.

Es natural, en síntesis, el afán por contar con muestras representativas. La pregunta medular es cómo actuar para conseguirlo. Una primera respuesta que viene a la mente es que para ello sería conveniente o necesario emplear el azar en el proceso de selección. Y en efecto, esta es una demanda bien establecida. Sin embargo, estrictamente considerado, no es difícil convencerse de que el azar, por sí mismo, no garantiza la representatividad (bien pudiéramos tener la “mala suerte” de que en una muestra de colombianos elegida aleatoriamente, la presencia de médicos sea muchísimo más alta que en la población).

¿Cuáles son entonces los méritos del azar? Se reducen a dos: en primer lugar, por su conducto se satisface la exigencia intuitiva y natural de eliminar -o mitigar al menos- la carga subjetiva que podría influir en la elección de los elementos que se van a examinar y, por ende, que pudiera gravitar sobre las conclusiones. El azar juega el papel de seguro contra distorsiones sistemáticas, sean o no deliberadas, y ese papel es por lo general insustituible, aunque no se garantice por su conducto la representatividad. En segundo lugar, y esto es realmente esencial, sólo cuando se hace uso del azar se podrá adicionar a las estimaciones una medida del grado de precisión con que se han realizado. De este aspecto nos ocuparemos más adelante.

En este contexto, procede enfatizar que, en general, no se emplea un método porque *siempre* funcione eficientemente, ni se descarta porque *siempre* fracase. La elección racional de un procedimiento -independientemente de lo que se procure resolver con él- poco tiene que ver con la infalibilidad; con lo que tiene que ver es con la evaluación del riesgo de que nos conduzca al fracaso: si éste es muy pequeño, se adopta; si es muy grande, se desdeña. Así nos conducimos cotidianamente cada vez que elegimos racionalmente una manera dada para resolver algo (modificar las opiniones de quienes nos rodean, comprar un producto, trasladarnos a algún sitio, etc.) En el caso de la selección de muestras, casi siempre es *posible* que nos toque en suerte una muestra “deficiente” aunque usemos el azar; pero si los tamaños de la población y de la muestra no son muy reducidos, entonces *muy probablemente* la muestra resultante será representativa a cualquier efecto de interés. Dicho de otro modo, las muestras “deficientes” serán relativamente pocas. La que resulte seleccionada cuando se aplique un procedimiento aleatorio no tiene una confiabilidad intrínseca; una vez elegida la muestra, incluso, no sabremos si ella es “buena” o no. Pero podremos estar razonablemente confiados en que lo será, especialmente cuando, además, la muestra tenga un tamaño suficientemente grande.

1.3 Los métodos de selección

Las formas en que puede intervenir el azar en el proceso de selección de la muestra son muy diversas. La elección de aquella que se empleará en un caso concreto es materia de especialistas y existen programas informáticos eficientes para ponerla en práctica (uno especialmente amigable es el módulo destinado al muestreo en EPIDAT, aplicación que se puede descargar gratuitamente

de <http://dxsp.sergas.es/>). En la presente sección se enumerarán, explicarán e ilustrarán muy brevemente los procedimientos más empleados.

Todos ellos comparten un rasgo que resulta cardinal a la hora de seleccionar una muestra: el procedimiento ha de otorgar a cada elemento de la población una probabilidad mayor que cero y conocida de antemano. Los diseños que cumplen esta condición son calificados como *probabilísticos*. La importancia de que así sea será explicada más adelante.

1.3.1 Muestreo simple aleatorio (MSA)

Este es el más sencillo, importante y conocido procedimiento de selección. En primer lugar, es un método básico como componente de casi todos los diseños que se explican más adelante. En segundo lugar, el muestreo simple aleatorio se utiliza ampliamente en el diseño experimental (en particular, en la fase de asignación de casos a los tratamientos en el marco de los ensayos clínicos controlados) y constituye el más notable punto de contacto de la teoría de muestreo en poblaciones finitas con toda la estadística inferencial clásica. Es un recurso que otorga igual probabilidad de selección a todos los sujetos de una lista. Sin embargo, esa no es su definición, como erróneamente suele creerse. La definición formal tiene cierta sutileza. Reza así: un método de selección de n elementos de un conjunto de tamaño N (donde n es menor o igual que N) es un MSA si todos los subconjuntos posibles de tamaño n de dicho conjunto tienen la misma probabilidad de ser la muestra resultante. Hay muchos otros métodos equiprobabilísticos que no cumplen esta condición.

La aplicación práctica de este procedimiento supone, en principio, el empleo de algún dispositivo físico que produzca desenlaces aleatoriamente (tal como extraer bolas numeradas de una urna) o de una tabla de números aleatorios donde se hayan registrado los desenlaces una vez empleado reiteradamente tal dispositivo. Sin embargo, en la actualidad se usan programas muy simples de computación que operan con los llamados números pseudoaleatorios, los cuales han hecho caer en desuso aquellas tablas, muy utilizadas hace solo unos años.

1.3.2 Muestreo sistemático (MS)

Este procedimiento de selección es esencialmente equivalente al **MSA** desde el punto de vista teórico. Sin embargo, es mucho más sencillo de aplicar. En esencia, consiste en ir tomando sujetos a intervalos regulares de un listado. Por ejemplo, tomar uno de cada $K=11$ de la lista de todos los alumnos de una escuela a partir de un número elegido al azar entre 1 y 11. Si por ejemplo, tuviéramos $N=1214$ alumnos en dicha lista, y si el número inicial que se ha elegido al azar hubiera sido el $r=7$, entonces quedarían en la muestra los sujetos que ocupan los lugares 7, 18, 29, 40, 51 ... y así sucesivamente hasta que se haya "recorrido" el listado completo.

En general, los sujetos así elegidos serán: $r, r+K, r+2K, r+3K...$ hasta $r+(n-1)K$ donde n es tal que $r+nK$ es el menor número de esa secuencia que supere a N .

La sencillez del método es especialmente acusada en determinadas situaciones, para las que no es preciso siquiera tener una lista física de los N sujetos, ni asignar números previamente a los elementos de la población, sino que basta con disponer de un marco muestral con cierta ordenación explícita o implícita (visitas sucesivas de pacientes a un centro, renglones en un libro de registro, o historias clínicas en un estante).

Por ejemplo, si se desea una muestra de altas hospitalarias correspondientes a cierto año en determinado hospital y el estudio se está planificando en algún momento del año anterior al de su desarrollo, entonces obviamente se desconocen los individuos específicos que integrarán la

población de interés (o sea, la población de sujetos que habrán de recibir el alta en el hospital); no se sabe siquiera el número de elementos que tendrá dicha población. Es imposible, por tanto, delimitar los integrantes de la muestra a través de un **MSA**. Sin embargo, como es obvio, puede establecerse algún sistema que permita dar una ordenación consecutiva a los pacientes que van a ir recibiendo el alta, y preestablecer con toda precisión los elementos que integrarán una muestra elegida mediante **MS** de esa futura población de altas.

1.3.3 Muestreo aleatorio estratificado (MAE)

Como se explicó en la sección 1.3.1, el **MSA** es una técnica de selección que depende exclusivamente del azar. Y como ya se discutió, sabemos que el azar no es una garantía de representatividad, aunque también conocemos que para muestras grandes (por ejemplo, de 100 o más elementos) probablemente la representatividad sea conseguida. Ahora bien, si suponemos que en la población bajo estudio pueden identificarse diferentes grupos de elementos cuya representación en la muestra quisiera asegurarse (por ejemplo, diferentes grupos ocupacionales en un estudio sobre riesgos de intoxicación o diferentes zonas de una ciudad), y si se cuenta con listas separadas para tales grupos, entonces puede aplicarse un *muestreo aleatorio estratificado*. Ello consiste en la conformación de listas separadas para cada uno de tales grupos y seleccionar submuestras (mediante alguna técnica de selección, que puede ser **MSA** o **MS**) dentro de cada uno de ellos. Actuando de esa manera, no sólo se asegura la representación de cada grupo en la muestra, sino que se puede controlar de antemano qué parte de la muestra habrá de proceder de cada uno de ellos.

Con el fin de obtener una muestra final que tenga en sí una variabilidad similar a la de la población (que era la idea central asociada a la representatividad), lo ideal sería conseguir que los subconjuntos conformados (los estratos) en que ella se divide fuesen internamente homogéneos y -consecuentemente- diferentes o heterogéneos entre sí.

En principio, la estratificación se aplica directamente a las unidades de análisis. Por ejemplo, se pueden dividir los pacientes hospitalizados según los servicios donde se hallan y luego tomar una muestra simple aleatoria de pacientes independiente para cada uno de los servicios (estratos) considerados.

Al aplicar un **MAE**, una vez que se ha determinado el tamaño de muestra, es necesario determinar cómo va a distribuirse ese tamaño entre los estratos que se han definido; o sea, debe decidirse el modo en que se dividirá el tamaño general en los tamaños correspondientes a los estratos respectivos. A tal proceso se le llama *asignación de tamaños muestrales a los estratos*. La forma más común de llevar adelante ese proceso es la llamada *asignación proporcional*, que sencillamente consiste, como el nombre indica, en repartir el tamaño global proporcionalmente a los tamaños de los estratos.

Cabe subrayar que el MAE también se puede aplicar a unidades conglomeradas. Para ilustrar esta modalidad muestral supongamos que se quiere caracterizar una población de trabajadores censados en una provincia de acuerdo al grado de contacto que han tenido con pesticidas, y que los investigadores saben que los patrones de tales contactos dependen en buena medida del tipo de actividad laboral (servicios, industrial o agrícola). Los municipios en que se divide la provincia pueden dividirse en aquellos que tienen cada una de las actividades mencionadas como la predominante. Es muy intuitivo que para tener un cuadro real de la magnitud y la estructura del problema, sería conveniente examinar personas procedentes de cada uno de los tres "estratos" posibles: residentes en municipios con actividad predominante de servicios, industrial y agrícola. En lugar de obtener un listado global de todos los municipios, el muestrista podría procurarse

listas separadas de tales zonas pero divididas según actividad predominante. Se eligen algunos municipios de cada tipo, de manera que todos ellos queden representados en la muestra. La del ejemplo precedente es una estratificación de municipios; la muestra se completará luego seleccionando individuos en casa municipio elegido. La estratificación no se realiza directamente sobre las unidades de análisis, pero estamos de todos modos ante un MAE, solo que concierne a una población de municipios.

1.3.4 Muestreo por conglomerados (MPC)

El ejemplo con que concluye la sección precedente, además de concernir al muestreo estatificado, también ilustra el llamado *muestreo por conglomerados*. Con esta denominación se alude a los diseños que no se basan en elegir directamente las unidades de análisis sino en seleccionar primero, siempre mediante procedimientos que apelan al azar, tales como el MSA o el MS, grupos de tales unidades (los conglomerados) y luego las unidades propiamente dichas dentro de los grupos o conglomerados.

Bajo este modelo entonces, la población se divide en cierto número de partes o conglomerados a las que se llama *unidades de primera etapa* (UPE). Cuando de cada UPE se seleccionan unidades de análisis, estamos ante un muestreo "bietápico" (salvo que se tomen todas las unidades contenidas en las UPE elegidas, en cuyo caso el muestreo se denomina *monoetápico*). Sin embargo, las etapas pueden ser más de dos; es decir, cada UPE a su vez atomizarse en subconjuntos llamados unidades de segunda etapa (USE), y así sucesivamente, hasta llegar a cierta etapa cuyas unidades estén directamente integradas por los sujetos que serán objeto de interrogatorio o medición. La idea estriba, sintetizando, en seleccionar un número de UPE, luego tomar un conjunto de USE dentro de las que integran las UPE seleccionadas, y continuar "bajando" hasta llegar a seleccionar las unidades de análisis que conformen la muestra final.

Consideremos como ejemplo una ciudad de 800 mil habitantes que se divide en sectores censales (UPE), que dichos sectores se consideran divididos en manzanas (USE) y éstas en viviendas, las cuales están finalmente integradas por individuos, que son en este caso las unidades de análisis. Por ejemplo, supongamos que se seleccionan 40 sectores entre todos los que existen en la ciudad, y que dentro de cada uno de ellos se toman 5 manzanas; supongamos que se continúa con la selección de 10 viviendas en cada una de las 200 manzanas ya obtenidas y que, finalmente, se escoge un individuo dentro de cada vivienda elegida, para determinar así 2000 elementos que integran la muestra de la población de 800 mil. El proceso ha transitado en este caso por cuatro etapas de selección: sectores, manzanas, viviendas e individuos.

Un distintivo clave del proceso de selección en este tipo de muestreo consiste en que la necesidad de listas de las unidades de muestreo en una etapa se limita a aquellas unidades seleccionadas en la etapa anterior. Así, en nuestro ejemplo, se necesita un listado de sectores; luego uno de manzanas, pero solo para los 40 sectores escogidos; en su momento se necesitará un listado de viviendas, pero solamente para las 200 manzanas elegidas en dichos sectores; y finalmente una lista de sujetos para cada una de las 2000 viviendas seleccionadas.

Como es obvio, la conformación de estas listas no puede producirse al unísono, sino que tiene que ser escalonada, ya que en una etapa aún no se conoce qué listas han de ser necesarias en la siguiente. Esto economiza recursos y favorece la agilidad del proceso; por ejemplo, la lista de residentes en una vivienda puede conformarse in situ, cuando ésta sea por primera vez visitada para realizar la encuesta correspondiente. Tal característica hace del MPC el más empleado en la práctica epidemiológica, ya que sería impensable, por ejemplo, hallar una lista de los 800 mil habitantes, y mucho menos aplicar un método directo como el MSA a dicha lista para obtener

2000 individuos, pues ello supondría tener que recorrer la ciudad palmo a palmo en lugar de ceñirse a la visita de solo 200 manzanas.

1.3.5 Muestras complejas (MC)

Cuando se conforman estratos de UPE (los conglomerados se agrupan previamente) y en cada uno de los estratos se aplica un MPC, se suele decir que se ha aplicado un procedimiento que da lugar a un “muestra compleja”. Siendo algo más específico, el contexto de un diseño muestral complejo es, en general, el siguiente: la población inicial se divide en estratos; dentro de cada estrato se eligen al menos dos conglomerados o unidades de primera etapa (UPE). Dentro de cada una de las UPE así elegidas en la población se selecciona cierto número de unidades de análisis (las que serán medidas o interrogadas) usando una o más etapas de selección. Por otra parte, además de la estratificación y las etapas del muestreo, la selección en cada una estas etapas puede llevarse a cabo de distintos modos: por ejemplo, mediante MSA o con probabilidades proporcionales al tamaño de las unidades que contiene. Obviamente, en la práctica se utilizan distintas combinaciones de los elementos que intervienen en el diseño muestral: puede ocurrir que la selección se realice en varias etapas pero sin estratificación, o que se aplique un muestreo estratificado de conglomerados en una única etapa.

Este es el más complicado y versátil de los métodos usuales, y se emplea cuando el universo es muy grande (por ejemplo, los ancianos de una gran ciudad o el conjunto de los estudiantes de la enseñanza primaria de un país). Evidentemente, el diseño de muestras complejas es materia altamente técnica y es una tarea que toca inequívocamente a los estadísticos especializados en muestreo.

2. Determinación del tamaño de muestra adecuado

2.1 Importancia del tamaño muestral

Determinar con qué tamaño de muestra se habrá de operar es algo crucial. Si bien es fuente de no pocos tics y frecuentes malentendidos, lo cierto es que se trata de algo simplemente ineludible, pues todos los estudios necesitan una muestra, independientemente de que su magnitud se decida por una vía u otra.

La relevancia del tema tiene dos vertientes: por una parte, de ello va a depender el costo del estudio y, por otro, la precisión de las estimaciones y el grado de fiabilidad que puedan tener en general los resultados que se obtengan, también dependen vitalmente del tamaño empleado. En principio, cuanto mayor sea la muestra, mayor será su representatividad (aunque ha de quedar claro, una vez más, que un tamaño “grande” no es condición suficiente para conseguirla); y, al mismo tiempo, cuanto mayor sea, más caro será llevarlo adelante.

Siendo así, es obvio que hay que hallar una especie de compromiso: que la muestra no sea tan pequeña como para invalidar las inferencias que se basarán en ella, y que no sea tan grande que no se disponga de recursos para estudiarla. Dicho de otro modo, el problema consiste en hallar el número mínimo de unidades con el cual puedan contestarse las preguntas de investigación.

En el caso de los estudios que aquí nos ocupan, los estudios descriptivos (diagnósticos de salud, caracterizaciones epidemiológicas, o estudios de prevalencia), dar respuestas a tales preguntas supone realizar estimaciones paramétricas, fundamentalmente de porcentajes, razones, medias, varianzas y totales poblacionales pero, ocasionalmente, también la estimación de otros parámetros, tales como coeficientes de correlación o de regresión.

2.2 Procedimientos para determinar el tamaño de muestra

Para introducir este tema, consideremos el caso en que se quiere estimar un porcentaje P mediante **MSA**. A continuación se esboza la solución convencional que se hallará en cualquier libro de texto para este problema.

La fórmula universalmente aceptada es la siguiente: $n = \frac{n_0}{1 + \frac{n_0}{N}}$ donde $n_0 = \frac{P(1-P)Z_\alpha^2}{e^2}$ y donde, a su vez, N es el tamaño poblacional, e es el máximo error absoluto que el investigador está dispuesto a admitir al estimar P y Z_α representa al percentil $\left(1 - \frac{\alpha}{2}\right) 100$ de la distribución normal estandarizada (por ejemplo, si $\alpha = 0,05$, entonces $Z_\alpha = Z_{0,975} = 1,96$).

Se plantea que usualmente "se puede tener una idea aproximada" del valor de P a partir de un estudio piloto, o de una investigación similar realizada anteriormente.

Imaginemos que se trata de una población de $N = 2800$ sujetos hospitalizados sobre la que se realiza un estudio de satisfacción. Admitamos, como es usual, que se trabajará con una confiabilidad del 95 % ($\alpha = 0,05$) y que un valor de $e = 3\%$ se considera aceptable. Finalmente, consideraremos $P = 86\%$, "idea aproximada" que se ha obtenido, por ejemplo, a partir de un estudio anterior. Así las cosas, la aplicación de las fórmulas expuestas arroja que $n = 434$.

Ahora bien, ésta es la solución que se da al caso en que se usa el muestreo simple aleatorio. Si el diseño muestral es otro, en particular si se trata de un muestreo en etapas (que es el que se utiliza por lo menos 90 de cada 100 veces en la práctica), el error que regularmente se comete al estimar P es mayor que el corresponde al **MSA** (supuesto, desde luego, que se está usando el mismo tamaño de muestra). Dicho de otro modo: con muestreo por conglomerados, para alcanzar el grado de precisión prefijado, es necesario seleccionar una muestra de tamaño mayor que el que demanda el **MSA**.

La sugerencia que se da entonces es, simplemente, multiplicar el tamaño anterior por un número mayor que la unidad: el llamado *efecto de diseño*, usualmente denotado como *deff*. Es decir, el tamaño corregido sería: $n_c = (n)(deff)$.

Para conocer más sobre el efecto de diseño y cómo se calcula, se sugiere al lector acudir a la literatura recomendada, pero a los efectos de este recorrido general, basta con que comprenda intuitivamente que es un número que intenta aprehender el grado en que se pierde representatividad debido a que se ha hecho un diseño estructuralmente más complejo (pero más simple desde el punto de vista operativo) que el MSA. Su nombre proviene del entendido de que un MSA es un método donde "no hay diseño" ya que depende exclusivamente del azar sin intervención alguna del muestrista una vez decidido el tamaño de muestra.

¿Qué otros elementos o alternativas pueden contemplarse para fijar el tamaño muestral? En lugar de extendernos en la respuesta a esta pregunta, invitamos al lector a que la halle a lo largo de la lectura del siguiente acápite, destinado a reflexionar críticamente sobre el empleo de fórmulas como las que se han incluido arriba.

2.3 Observaciones críticas relacionadas con las técnicas para determinar el tamaño muestral

Curiosamente, a pesar de ser uno de los temas más borrosamente solucionados por la estadística, la uniformidad del tratamiento que le dan los textos a la determinación del tamaño de muestra es casi total; ello hace pensar a muchos que la armonía entre la teoría y la práctica correspondiente, es completa. Sin embargo, como se expone a continuación, ese no es ciertamente el caso.

Este polémico tema ha sido tratado con bastante detalle en un capítulo de un libro destinado a incrementar la cultura estadística de los investigadores.³ Dicho capítulo contiene, sin embargo, un extenso material relativamente complejo para no estadísticos, de manera que en lo que sigue se abordará este asunto de una manera más ligera y general.

El sensible y crucial asunto de determinar el tamaño muestral adecuado constituye una tarea muy singular porque en ella confluyen dos circunstancias especiales y en cierto sentido contradictorias: en primer lugar, que este es un asunto al que hay que atender insoslayablemente: como se señalaba al comienzo de esta sección, por una u otra vía, no hay más remedio que decidir un tamaño muestral y, en segundo lugar, que no existe una solución formal que sea universalmente satisfactoria.

No se trata de que no existan fórmulas que supuestamente aportan la solución, como vimos en el acápite precedente para un ejemplo concreto, sino que ellas plantean tantos o más problemas de los que resuelven por la simple razón de que, salvo situaciones excepcionales, para aplicarlas hay que fijar de antemano valores desconocidos que solo pueden ser determinados mediante una gran dosis de imaginación. Salvo alguna pincelada aquí o allá, el conflicto es predominantemente pasado por alto. Ante tal contingencia, aparecen de inmediato las "soluciones" rituales.

Se ha consolidado entre estudiantes e investigadores la convicción de que para cada problema existe un número que puede ser "descubierto" por los especialistas, independientemente del enfoque personal; un número que puede determinarse técnicamente por aquellos "elegidos", capaces de desentrañar complejas formulaciones. Muchos metodólogos profesionales suelen poner en un serio aprieto a modestos investigadores exigiéndoles que justifiquen formalmente el tamaño muestral que han elegido por analogía con lo que han visto en la literatura, o porque es el que permiten sus recursos. Es una situación irónica, porque con alta probabilidad la inmensa mayoría de los propios inquisidores se verían en similar dificultad si, en lugar de dedicarse a pedir respuestas, tuvieran que producirlas.

En el epígrafe precedente dimos un ejemplo de la "solución oficial" para la situación más simple dentro de los estudios descriptivos. Como veremos de inmediato, este planteamiento contiene no pocas píldoras difíciles de tragar. Varios de estos problemas son usualmente omitidos cuando se aborda el tema. A continuación se enumeran y comentan cinco de ellos tomando como paradigma, nuevamente, el caso más simple: la estimación de P en el contexto descriptivo. Para casos más complejos, la situación no hace sino agravarse.

2.3.1 La demanda inicial es directamente cuestionable

Retomemos el ejemplo de la sección precedente donde se quería conocer el porcentaje de pacientes satisfechos en la población de sujetos hospitalizados: se había fijado $e = 3\%$ y una confiabilidad del 95 %, y se sabía que $N = 2800$; pero para hacer uso de la fórmula fue necesario conocer nada menos que el propio valor de P , el porcentaje de pacientes satisfechos. En el ejemplo habíamos admitido que $P = 86\%$, número procedente de un estudio anterior. Pero si miramos el asunto con un espíritu menos concesivo, tendremos que admitir que estamos en un

círculo vicioso: todo el proceso tiene como finalidad conocer el valor de P pero, para empezar a desarrollarlo, es necesario conocer precisamente P . Nótese que no se trata de un detalle lateral: un enfoque riguroso aconsejaría detener el proceso en este punto por elementales razones de coherencia lógica. El planteamiento es análogo, por poner un ejemplo, a que un viajero que se halle en Medellín y quiera trasladarse a Bogotá, solo pueda adquirir el billete de avión que lo conduzca a la capital acudiendo personalmente a una oficina radicada en ... Bogotá. Para apreciar mejor esta inconsistencia, obsérvese que, puesto que vamos a trabajar con "una idea aproximada" del valor de P , estamos condenados a conocer solo "una idea aproximada" del valor de n necesario. Si aquella aproximación fuera mala (muy errónea), así será el tamaño muestral, algo obviamente inconveniente. Pero si, en cambio, estuviéramos verdaderamente persuadidos de que la pre-estimación de P fuera muy cercana a su verdadero valor, entonces la situación sería directamente incoherente, porque en tal caso no haría falta hacer la estimación y con más razón sería innecesario el cómputo de n .

2.3.2 Las formulas exigen el empleo de datos sometidos a un alto grado de subjetividad

La "pre-estimación" de P , además de ser conceptualmente conflictiva, demanda de un acto intrínsecamente cargado de subjetividad, pues puede tomarse como referencia un trabajo previo u otro, o puede no existir tal trabajo y tener entonces que inventarlo, o acudir al recurso manido (y equivocado, como se explica detalladamente en un artículo aparecido en 2000 en la revista Gaceta Sanitaria⁴) recurso de partir de que $P=50\%$.

Examinemos ahora los otros elementos de la fórmula. Se supone que hay un "error máximo" que se puede aceptar. Es obvio que, ineludiblemente, se trata de una decisión esencialmente subjetiva. Volvamos al ejemplo: ¿qué error ha de considerarse suficientemente pequeño como para que resulte admisible? No parece existir árbitro alguno que determine sin margen para el debate si un error de un 3% -cómo habíamos establecido en el ejemplo- es admisible; o si, para que lo sea, éste no debe exceder, por ejemplo, el 2% de pacientes satisfechos. Algo muy similar ocurre con la confiabilidad. ¿Tomar 95%? ¿o 99%, como sugieren otros con no menos argumentos?

La realidad es tal que dos investigadores independientes, siguiendo la misma estrategia general (usando las mismas fórmulas), pueden obtener tamaños abismalmente diferentes, solo por el hecho de que han fijado -legítimamente ambos- de manera diferente, los valores que tienen que incluir dentro de las fórmulas para llegar al tamaño muestral. Adviértase que si, por ejemplo, tomáramos $e = 2\%$, confiabilidad del 99% ($Z_\alpha = Z_{0,95} = 2,57$) y un valor de $P = 74\%$, acaso obtenido de un estudio piloto - todos valores tan legítimos como los anteriores- entonces obtendríamos un tamaño casi cuatro veces mayor: $n = 1505$.

2.3.3 Casi nunca se quiere estimar un único parámetro

Una encuesta, en la realidad, nunca se proyecta para estimar uno o dos parámetros, sino que normalmente exige realizar decenas o incluso cientos de estimaciones. Debe notarse que una tabla demanda una estimación por cada celda y, en ocasiones, suelen construirse muchas tablas con varias celdas en cada cual. Es evidente, entonces, que se podrían aplicar las fórmulas como las arriba expuestas para cada uno de estos parámetros, lo cual produciría un amplísimo abanico de posibles tamaños de muestra, todos candidatos a ser el tamaño muestral con el que a la postre

se trabajaría (este sí, necesariamente único). Esta polivalencia suele ocultarse bajo la premisa de que se puede elegir el parámetro "más importante" (en el mejor de los casos, "los dos o tres más importantes"). Obviamente, incluso suponiendo que siempre tuviera sentido hacer tal discernimiento, el margen de subjetividad para decidir cuáles son tales "parámetros más importantes" es enorme. Y aun así, esos "más importantes" pueden dar lugar a números muy dispares entre sí.

2.3.4 La necesidad de realizar estimaciones dentro de subconjuntos poblacionales no es tomada en cuenta

Lo usual es que se seleccione una muestra general de unidades y luego se hagan, además de análisis globales, otros circunscritos a subconjuntos de la población, basados -naturalmente- en los correspondientes subconjuntos muestrales, cuyos tamaños son necesariamente menores (a veces mucho menores) que el de la muestra original. Tal contingencia no suele ser tomada en consideración cuando se aplican las fórmulas. Por ejemplo, en el estudio de los pacientes hospitalizados que se ha bosquejado, es muy probable que se quieran hacer estimaciones del porcentaje de sujetos satisfechos, no solo para toda la población de hospitalizados, sino también dentro de cada uno de los sexos, o según servicios del hospital, o con acuerdo a lapsos de estancia hospitalaria.

En síntesis, lo que se quiere subrayar en este punto es que, al computar un tamaño muestral empleando fórmulas, normalmente se actúa pensando en estimaciones para la población completa; pero el error "máximo" que se está dispuesto a cometer y que se usa con el fin de realizar aquella determinación, resulta mucho menor que el que verdaderamente se comete en el marco de la tabulación verdadera.

2.3.5 El costo de los procesos es casi siempre mucho más determinante que cualquier consideración teórica

Realizar el análisis del tamaño muestral pensando solo en términos de *eficacia*, sin tener en cuenta la *eficiencia* es poco realista: si no hubiese limitaciones de recursos, en muchos casos no se plantearía siquiera el uso de muestras, sino que se investigaría íntegramente la población. Las disponibilidades de tiempo, personal y presupuesto deben ocupar (y de hecho ocupan) un lugar determinante en las decisiones, aunque muchas veces la influencia de esta limitante se ejerza de manera implícita o solapada.

2.4 Consideraciones finales sobre el tamaño muestral

La mayoría de los textos y de los profesores pasan por alto las realidades que se acaban de enumerar. Afortunadamente, hay algunas excepciones; una de ellas es, en mi opinión, notable, pues se encuentra en el famoso libro debido a Kenneth Rothman del año 1986, en el cual se reconoce con crudeza la inviabilidad de una solución teórica cuando escribe⁵:

En resumidas cuentas, el problema de determinar el tamaño de muestra más adecuado no es de naturaleza técnica; no es susceptible de ser resuelto por vía de los cómputos, sino que ha de encararse mediante el juicio, la experiencia y la intuición.

Otra publicación más reciente pero tanto o más relevante es debida a Schulz y Grimes y apareció en 2005 en Lancet.⁶ Allí se hace notar que ante la naturaleza difusa de la teoría para determinar tamaños muestrales, se suele apelar a la aplicación de un ajuste retrospectivo (el llamado *retrofitting*), consistente en: fijar primero el tamaño muestral que, sin uso de fórmula alguna, se

considera factible y adecuado, y luego usar las fórmulas para justificarlo eligiendo los valores que han de ponerse en la fórmula para llegar a dicho tamaño.

Lo más importante es comprender que, cualquiera que sea el tamaño de muestra, los errores de muestreo pueden ser calculados a posteriori. O sea, las fórmulas para el cómputo de errores contemplan explícitamente los tamaños muestrales empleados; la estructura de esas fórmulas es tal que el investigador se verá imposibilitado de sacar partido concluyente a la información si el tamaño de la muestra fuera insuficiente, pero por lo menos siempre podrá conocer el grado de confianza que pueda depositar en sus resultados.

A pesar de todas estas críticas, las fórmulas pueden tener utilidad como recurso orientativo, siempre que se usen de modo flexible, se reconozcan sus enormes limitaciones y se complementen con una reflexión crítica y abarcadora. En cualquier caso, es ilustrativo reparar en que no casualmente una gran cantidad de los trabajos serios y trascendentes (por ejemplo, los que aparecen en revistas de alto impacto, tal y como se demuestra en el trabajo de Silva y Alonso, 2013,⁷ que se cita en las lecturas recomendadas), no se detienen a explicar con taparrabos técnicos sus tamaños muestrales. Usualmente se circunscriben a comunicar los que fueron usados. Los juicios que nos puedan merecer sus conclusiones no se modificarían en absoluto por el hecho de que consiguieran o no persuadirnos de la legitimidad técnica de aquella decisión, sino, en todo caso, por la longitud concreta de los intervalos de confianza.

3. Establecer los procedimientos que habrán de aplicarse a los datos obtenidos

3.1 Probabilidad de inclusión y ponderaciones

Un paso de crucial relevancia en el proceso histórico de consolidación de la teoría del muestreo fue dado con la introducción del azar en el acto de selección muestral. Sin él, no tendríamos muestras probabilísticas. Se recordará que se cumple esta condición cuando cada sujeto de la población tiene una probabilidad no nula y conocida de antemano de integrar la muestra. Dicho número es conocido como la “probabilidad de inclusión” del elemento. Una vez seleccionada la muestra, conoceremos la probabilidad de inclusión que tuvo cada una de las unidades que la integran. Esta probabilidad depende, naturalmente, de la forma en que se haya llevado adelante la selección.

La importancia de este conocimiento reside en que es imprescindible para poder hacer adecuadamente tanto las estimaciones puntuales como los errores asociados a ellas, un elemento metodológico, este último, del que nos ocupamos en la siguiente sección.

¿Cómo intervienen estas probabilidades en ambos procesos? Desborda el alcance de este capítulo dar una respuesta detallada a esta pregunta. El lector, sin embargo, ha de tener presente que, cuando las muestras no son equiprobabilísticas (no todos los sujetos de la muestra tienen igual probabilidad de selección), a la hora de procesar los datos será necesario incorporar ponderaciones. La ponderación que corresponde a un sujeto de la muestra –también conocida como *factor de expansión*– se calcula, generalmente, como el inverso de su probabilidad de inclusión.

La suma de las ponderaciones correspondientes a cada uno de los miembros de la muestra es igual al tamaño poblacional. Coherentemente con esta propiedad, la ponderación que

corresponde a un sujeto de la muestra puede interpretarse como el número de unidades de la población a quienes dicho sujeto representa.

3.2 Error de muestreo

Las inferencias que se realizan a partir de una muestra necesariamente conciernen a una población. Cabe recalcar que el investigador nunca se hace preguntas sobre la muestra, la cual no es más que un intermediario metodológico para resolver su tarea. Ahora bien, entre el verdadero valor del parámetro poblacional y su estimación, siempre habrá una diferencia. Nuestra esperanza es que ella sea pequeña, de modo que podamos considerar a la segunda como si fuera el valor real del parámetro. Esta diferencia es lo que se conoce como *error asociado a la estimación*, algo que se puede considerar connatural al proceso inductivo y al hecho de que operamos con una muestra. Más específicamente, dicha diferencia se conoce como *error de muestreo*. Tal error, puede a su vez ser estimado: es la distancia máxima estimada entre el valor estimado y el parámetro desconocido. Una de las tareas del muestrista consiste en identificar el procedimiento de cálculo de dicha estimación.

Supongamos que cierto estudio tiene entre sus propósitos conocer el gasto promedio en medicamentos en que incurren mensualmente los núcleos familiares de cierta comunidad francesa. Los investigadores conocen por estudios similares que dicho valor debe estar entre €2.00 y €6.00, pero necesitan un dato más preciso.

Supongamos que tras una encuesta por muestreo se obtuvo, en calidad de estimación, que dicho gasto promedio asciende a €3.25. ¿Qué significa ese dato si no se tiene una idea del posible error cometido? Virtualmente nada. Sin embargo, si además se informase que el error máximo es $e=0,15$ €, lo que se está afirmando es que el verdadero gasto promedio se encuentra casi con seguridad entre €3.10 y €3.40. Ahora, si en lugar de €0.15, el error asociado a la estimación ascendiera a €3,20, entonces lo único que se conocería, en definitiva, sería que el verdadero gasto medio está entre €0.05 y €6.45. O sea, no se agregaría absolutamente nada a lo que ya se conocía de antemano. Es preciso comprender, en síntesis, que la estimación aislada, sin una medida del error que pueda estar afectándola, carece virtualmente de interés.

No explicaremos aquí cómo se calculan los errores de muestreo, pero es importante que el lector tenga presente cuatro aspectos concernientes a este tema:

1. el error de muestreo solo se puede calcular si la muestra fue probabilística
2. las fórmulas para computarlo dependen del diseño muestral empleado
3. dichas fórmulas son diferentes para cada tipo de parámetro
4. la magnitud del error depende vitalmente del tamaño de la muestra (a mayor tamaño, menor error, y viceversa)
5. es necesario disponer de algún software especializado para resolver el problema, ya que los programas convencionales no contemplan el diseño muestral empleando al calcular el error

3.3 Estimadores puntuales y por intervalo

Como ya se ha insistido, en el contexto de los estudios descriptivos la caracterización que se procura realizar casi siempre supone la estimación de parámetros. Concentrémonos en un parámetro genérico al que llamaremos P (que puede representar, por ejemplo, la prevalencia de asmáticos o el número total de cardiopatas en una comunidad). Supongamos que estamos en el primer caso y que un técnico de salud ha estimado ese número mediante un diseño probabilístico

y que comunica que la estimación de P es igual a 11 %. Esto, obviamente, no quiere decir que la prevalencia sea exactamente igual a 11 %. Cuando se nos dice que la estimación es 11 %, sabemos que esto significa que el verdadero valor de P es "más o menos" igual a 11. La demanda, natural, sin embargo es conocer cuánto "más o menos".

Cuando un titular de prensa afirma algo como lo siguiente: "*A partir de una rigurosa muestra aleatoria de 300 hombres de la capital, los expertos han estimado que el porcentaje de hombres que usan preservativos regularmente se incrementó en el curso del último año de 31,4 % a 35,7 %*", pero escamotea (quizás porque no considera necesario comunicarlo o porque la fuente informativa no lo consigna) que el error máximo asociado a esta última estimación es de 4,6 %, probablemente el lector común quede automáticamente persuadido de que ese consumo ha aumentado. Sin embargo, en tal caso, lo que realmente se sabe es que el porcentaje actual está, muy probablemente, entre 31,1 % y 40,3 %, de modo que pudiera ser incluso inferior al 31,4 % obtenido antes.

Es decir, resulta lógica la aspiración de contar no solo con una estimación puntual de P , sino también con una idea del error máximo que pudiera estar afectando a dicha estimación. Esto equivale a disponer de un intervalo en cuyo seno quepa esperar que se halle el parámetro.

Un *intervalo de confianza* para un parámetro P es, informalmente definido, una pareja de números P_1 y P_2 entre los cuales podamos "estar confiados" que se halla el parámetro en cuestión. Es obvio que no es lo mismo creer que P es 11 % más o menos 0.5 % (el intervalo sería 10,5%-11,5%), que creer que es 11 % más o menos 4 % (el intervalo sería 7%-15%), o que solo saber que es "aproximadamente" igual a 11 %, pero sin tener la menor idea acerca de cuál es la magnitud del error posible.

Cabe apuntar, de paso, que esa es, en esencia, la razón por la cual las "votaciones" que suelen incluirse en los sitios WEB y periódicos digitales, carecen de todo fundamento científico como instrumento para conocer «la opinión pública».

Asociadas a un método de muestreo, hay potencialmente muchas muestras posibles. Sin embargo, una vez aplicado, se obtiene solo una de ellas. La virtud del intervalo que se obtiene para la única muestra que se selecciona en la práctica estriba en que se produjo por un método confiable; o sea, se podrá estar razonablemente confiado en que ese único intervalo que tocó en suerte no estará entre los pocos teóricamente obtenibles que no contienen al parámetro. Consecuentemente, si el intervalo se construye por un método tal que acierta a "atrapar" a P , por ejemplo, el 95 % de las veces, se dice que " P se halla entre P_1 y P_2 con una confiabilidad del 95 %". El intervalo de confianza incorpora la estimación del error introducida en la sección anterior a la estimación puntual.

En síntesis, es muy cuestionable que se prescindiera de intervalos de confianza puesto que su función es la de permitir juicios sobre el grado en que podemos considerar que realmente conocemos el parámetro que quiere estimarse. Dicho de otro modo: resulta crucial contar no solo con una estimación puntual, sino también con una idea del error máximo que pudiera estar asociado a dicha estimación, lo cual equivale a disponer de un intervalo en cuyo seno quepa esperar que se halle el parámetro. Un trabajo que prescindiera de esta información tiene un valor casi tan reducido como uno en que los resultados hayan sido inventados.

El *International Committee of Medical Journal Editors* en sus casi universalmente aceptados *Requisitos Uniformes para los Manuscritos Enviados a Revistas Biomédicas*,⁸ desde 1998 y hasta la última versión, de 2013 (en la sección donde se pauta cómo debe manejarse la información

estadística) es categórico al respecto cuando consigna textualmente: “Cuantifique los hallazgos y preséntelos con los indicadores adecuados de medición de error o inseguridad, tal como los intervalos de confianza”

Bibliografía

¹ Diccionario de la Real Academia Española, 21ª Edición, Madrid: Espasa Calpe; 2011.

² Silva LC. Diseño razonado de muestras y captación de datos para la investigación sanitaria. Madrid: Díaz de Santos; 2000.

³ Silva LC. Cultura estadística e investigación en el campo de la salud: Una mirada crítica Madrid: Díaz de Santos; 1997.

⁴ Silva LC. Nueva visita al supuesto de máxima indeterminación y al empleo de errores absolutos y relativos. Gaceta Sanitaria 2000; 14(3):254-257.

⁵ Rothman JK. Modern epidemiology. Little, Brown and Col., Boston; 1986.

⁶ Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. The Lancet 2005; 365:1348-1353.

⁷ Silva LC, Alonso P. Explicación del tamaño muestral empleado: una exigencia irracional de las revistas biomédicas. Gaceta Sanitaria 2013; 27(1): 53-57.

⁸ International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. <http://www.icmje.org>. Updated octubre 2013. Accesible en febrero 2, 2014.