

Cálculo del tamaño de muestra para el estudio del subregistro de enfermedades de transmisión sexual en el Valle del Cauca

Sample size calculation for the study of unregistered sexually transmitted diseases in Valle del Cauca

José Rafael Tovar C.

Estadístico, magíster en estadística, profesor de la Universidad del Valle y la Pontificia Universidad Javeriana, Cali, Colombia.
Cibercorreo: jrtovar@puj.edu.co

Carlos Alberto Rojas A.

Médico, epidemiólogo, Ph. D. en epidemiología, profesor de la Facultad Nacional de Salud Pública, Universidad de Antioquia, Medellín, Colombia.
Cibercorreo: crojas@guajiros.udea.edu.co

Héctor Jairo Martínez R.

Médico, matemático, Ph. D., profesor de la Universidad del Valle, Cali, Colombia.
Cibercorreo: hector@univalle.edu.co

Recibido: 30 Enero 2006 Aprobado: 7 Noviembre 2006

Tovar JR, Rojas CA, Martínez HJ. Cálculo del tamaño de muestra para el estudio del subregistro de enfermedades de transmisión sexual en el Valle del Cauca. Rev. Fac. Nac. Salud Pública. 2007; 25 (2): 110-116

Resumen

El cálculo del tamaño de muestra es un aspecto muy importante del diseño de estudios en muchas áreas de investigación. Generalmente, los investigadores saben que necesitan calcular el número de elementos a incluir en la muestra, pero desconocen la estrategia más apropiada para establecer dicho número, de manera que se cumplan los lineamientos teóricos desarrollados para obtener la mejor estimación posible del indicador que resume la información sobre la variable en la población de estudio (parámetro). Aquí se presenta la metodología utilizada para abordar un problema en el que se requería calcular el tamaño de muestra para estimar la proporción de casos de enfermedades de transmisión sexual no registrados en las bases de datos de la Secretaría de Salud Departamental del Valle del Cauca, tomando todos los municipios del departamento, excepto Cali por su

tamaño y heterogeneidad. En la metodología desarrollada, se estableció como diseño de muestreo apropiado aquel que incluye un procedimiento de estratificación con selección de conglomerados dentro de cada estrato en una primera etapa y posterior selección aleatoria de elementos en cada conglomerado en una segunda etapa. En la primera etapa, se seleccionaron aleatoriamente 30 municipios; en la segunda se tomó un total de 70 instituciones prestadoras de servicios de salud de carácter privado y 30 públicas (una por municipio). Para calcular el número de instituciones privadas se tomaron en cuenta características del municipio y el presupuesto disponible para realizar el estudio.

----- *Palabras clave:* muestreo, muestreo estratificado, muestreo por conglomerados, enfermedades de transmisión sexual

Summary

The sample size or number of elements to include in a study is one of the most important aspects in the design of a study. The researcher knows the need to calculate the number of subjects to include in an investigation but does not know what is the best plan or sampling design to use in order to get the best possible estimate of the parameter of interest. This work shows the methodology used to approach a problem in which it was necessary to calculate the sample size to estimate the proportion of unregistered sexually transmitted diseases cases in the data bases of the Public Health Department in Valle del Cauca, Colombia. In this analysis, data from all the cities in Valle del Cauca were included except the capital city,

Cali, because of its big size and heterogeneity. An appropriate sampling design including a stratification procedure with selection of clusters inside of each stratum and then a random selection of elements inside of each conglomerate were both defined. In the first step, 30 cities were selected and in the second step, a total of 70 private health care institutions were selected. The final sample was made up 100 institutions, 30 public (one per city) and 70 private. To calculate the number of private institutions, city conditions and the study budget were taken into account.

----- *Key words:* sampling studies, stratified sampling, cluster sampling, sexually transmitted diseases

Introducción

El cálculo del tamaño de muestra en cualquier clase de estudio implica, de parte del investigador, un conocimiento claro tanto del fenómeno estudiado como de la población donde realizará sus evaluaciones.¹ En este artículo se presenta la manera como se abordó una situación en la cual se necesitaba calcular un tamaño de muestra bajo condiciones complejas respecto de la composición de la población objeto de estudio y limitaciones en el presupuesto con el que se contaba para la realización del levantamiento de datos de campo. Se requería calcular la proporción de casos de enfermedades de transmisión sexual (ETS) que no aparecen registrados en las bases de datos de la Secretaría de Salud Departamental del Valle del Cauca (SSD), durante los años 1999 y 2000. Esta es una situación crítica para los programas de control de estas enfermedades en cualquier sistema de salud.²

Para tener la información real de la proporción de subregistro, habría que visitar los 41 municipios del departamento y encuestar a las personas encargadas del registro y manejo de los datos en cada una de las instituciones prestadoras de servicios de salud (IPS) públicas y privadas, lo que elevaría los costos desproporcionadamente. Por lo tanto, para abordar el problema, se consideró como población al conjunto de IPS de los municipios del departamento, excluyendo a Santiago de Cali, cuyo tamaño poblacional es mucho mayor que el del resto de municipios, razón por la que se diseñó un plan de muestreo específico para la ciudad de Cali. De acuerdo con las características del estudio, se diseñó un esquema de muestreo en el que se calculó la cantidad de municipios a visitar y, dentro de estos, se calculó el número de IPS que el presupuesto y las consideraciones técnicas y teóricas propias de la metodología de muestreo permitían incluir. En cada IPS se contaron los casos de ETS registrados durante el período de estudio para compararlos luego con los registrados en las bases de datos de la SSD.

Este artículo está compuesto de dos partes. La primera incluye el desarrollo teórico que permite obtener las ecuaciones necesarias para el cálculo del tamaño de muestra, conservando la consistencia con el diseño de muestreo. La segunda parte incluye el uso de las ecuaciones obtenidas en una situación real que requería el cálculo del tamaño de muestra para estudiar un problema de salud pública.

Tamaño de muestra

Antes de calcular el tamaño de la muestra se debe definir el diseño de muestreo más apropiado, teniendo en cuenta el contexto y estructura de la población estudiada y algunas condiciones de las unidades de estudio. El cálculo del tamaño de muestra para los diseños que incluyen estratos con conglomerados en su interior implica varios pasos: calcular el número de conglomerados de la muestra, su distribución por estrato y el número de unidades de estudio dentro de cada conglomerado.

Para el cálculo del número de conglomerados en la muestra, se deben considerar dos aspectos: la cantidad total de unidades primarias dentro de los estratos y su variabilidad dentro de ellos. El número de conglomerados en la muestra se calcula haciendo una modificación a la ecuación desarrollada por Cochran para estimar medias usando muestreo estratificado cuando se tiene una población finita y se quiere que el número de elementos por estrato en la muestra sea proporcional a la cantidad de elementos en la población. Dicha ecuación es:^{3, 4, 5, 6}

$$n = \frac{\sum_{h=1}^l W_h \left[\frac{N_h P_h Q_h}{N_h - 1} \right]}{V(\hat{p}) + \left((1/N) \left\{ \sum_{h=1}^l W_h \left[\frac{N_h P_h Q_h}{N_h - 1} \right] \right\} \right)}, \quad (1)$$

$$Q_h = 1 - P_h \quad V(\hat{p}) = \left(\frac{d}{Z_{\alpha/2}} \right)^2$$

$$W_h = \frac{N_h}{N}, \quad N = \sum_{h=1}^l N_h$$

donde:

l : número de estratos. Puede establecerse naturalmente por el comportamiento de la variable en el entorno o por el investigador de acuerdo con su conocimiento de la situación y el contexto;

d : diferencia máxima esperada entre el estimador y el parámetro investigado. Valor establecido por el investigador;

$Z_{\alpha/2}$: percentil $(1-\alpha/2)*100$ de la distribución normal estándar, siendo α la probabilidad de cometer error tipo I, establecida por el investigador;

$V(\hat{p})$: varianza esperada para el estimador del parámetro;

W_h : peso del estrato, calculado como la relación entre el número de conglomerados del estrato (N_h) sobre el total de conglomerados en la población ($N = \sum_{h=1}^l N_h$);

P_h : proporción de casos de interés en el estrato h , la cual debe ser aproximada bien sea por información obtenida a través de algún estudio previo o por datos de una prueba piloto. Cuando no se tiene ningún conocimiento previo sobre su valor y se espera que P no sea un valor cercano a cero o a uno (en el intervalo entre 0,2 y 0,8, por ejemplo), algunos autores recomiendan tomar $P_h = 0.5$ para maximizar la varianza y así obtener el mayor tamaño de muestra posible, mientras esté fijo el error de estimación establecido a priori.^{7, 8, 9, 10}

Una vez obtenido el número total de conglomerados, usando el mapa del departamento, se estableció la cantidad de municipios ubicados dentro del estrato y la

proporción que comprendía respecto al total para luego calcular la cantidad de unidades a evaluar en cada conglomerado. Esto se puede hacer considerando el presupuesto destinado para recolectar los datos de campo, el cual se distribuye entre los estratos y puede ser variable dependiendo de diferentes aspectos relacionados con el contexto físico y geográfico de la población estudiada. Dada la anterior restricción, la cantidad de unidades de estudio para un conglomerado cualquiera en un estrato específico, que permite tener una varianza mínima para el estimador del parámetro, se calcula ajustando las fórmulas dadas por Cochran cuando se hace un muestreo bietápico con conglomerados de diferente tamaño y costos fijos. La ecuación usada para calcular el número de unidades a evaluar es:^{3, 6, 10, 11}

$$m_{hi} = f_{hi} \bar{m}_h \quad i=1, 2, \dots, n_h \quad h=1, 2, \dots, l$$

$$f_{hi} = \frac{M_{hi} \sigma_{hi}}{\sum_{i=1}^{n_h} M_{hi} \sigma_{hi}} \quad \bar{m}_h = \frac{C_h - C_{2h} n_h}{C_{2h} n_h} \quad h=1, 2, \dots, l \quad (2)$$

$$\sigma_{hi} = \sqrt{\frac{M_{hi} P_{hi} (1 - P_{hi})}{M_{hi} - 1}} \quad i=1, 2, \dots, n_h \quad h=1, 2, \dots, l$$

donde:

m_{hi} : número de unidades de estudio en el conglomerado i del estrato h que deben incluirse en la muestra;

M_{hi} : total de unidades de estudio que conforman el conglomerado i en el estrato h ;

n_h : número de conglomerados del estrato h incluidos en la muestra;

c_h : cantidad de dinero presupuestada para encuestar en el estrato h ;

c_{1h} : costo que no depende de la aplicación de la encuesta propiamente, sino de factores como transporte y actividades relacionadas con la visita a los conglomerados del estrato h ;

c_{2h} : costo por unidad de aplicación de la encuesta en el estrato h , el cual puede estar asociado al precio del papel, costos de impresión, fotocopias, salario de encuestadores, etc.;

P_{hi} : proporción de subregistro esperada en el conglomerado i del estrato h ;

\bar{m}_h : número promedio de unidades que se puede muestrear en el estrato h . Se calcula teniendo en cuenta la cantidad de dinero presupuestado para el estrato y los costos del muestreo dentro del mismo.

Planteamiento del problema

En el año 2001, la SSD tenía un sistema de bases de datos en el que se registraban los casos de enfermedades de notificación obligatoria, incluidas las enfermedades

de transmisión sexual. Con base en la forma como venía operando el sistema de captura y registro de casos en las diferentes instituciones del departamento, y en concordancia con lo reportado en la literatura médica,¹² se tenía la hipótesis de que los archivos de la entidad no registraban la cantidad real de casos de ETS identificados en las IPS del departamento, razón por la que se debía estar presentando un índice de subregistro que era importante estimar.

Marco muestral

Se contó con el listado oficial de IPS registradas en la SSD en el año 2001, que debió ser depurado y actualizado, pues existían registros duplicados, registros de instituciones que ya no prestaban sus servicios como IPS, ausencia de IPS existentes y direcciones sin actualizar. Pese al tiempo y costo adicional —no presupuestados— que demandó la limpieza del marco muestral, es muy probable que el costo para el proyecto hubiese sido mayor si se hubieran seleccionado las IPS y programado las visitas sin haber realizado la depuración del marco muestral.⁵

Definición del parámetro

En la definición de subregistro, se consideró que a partir de la reforma al sistema de salud en Colombia, se produjo un aumento considerable del número de IPS de carácter privado, mas no así de las públicas. Por esta razón, en muchos municipios del departamento del Valle existe una sola IPS pública y un número variado de IPS privadas. La percepción de los funcionarios de la SSD es que a esta dependencia suelen reportarse regularmente los casos de ETS de la IPS públicas, pero solo unos pocos de los casos atendidos en las entidades privadas, lo cual es consistente con reportes de la literatura.^{12, 13, 14}

Se procedió bajo la hipótesis de que la IPS visitada durante el estudio entregaría información de una cantidad de casos mayor o igual a la registrada en la base de datos de la SSD. Cuando la IPS no figuraba en la base de datos de la SSD, se tomó como cero la cantidad de casos registrados. El parámetro de interés era la proporción de casos no reportados en el departamento, la cual se puede calcular con base en la proporción municipal. El índice municipal de subregistro (IMS) se construyó como el cociente entre el número de casos no reportados a la SSD sobre el total de casos identificados en el municipio durante el período de estudio, es decir:

$$IMS = \frac{u - v}{u} \quad (3)$$

donde:

u : número total de casos identificados en el municipio durante el periodo;

v : número de casos registrados en la base de datos de la SSD durante el mismo periodo.

Después de obtenerse las proporciones de subregistro municipal, el índice departamental de casos de ETS no registrados (IDS) se calculó como

$$IDS = \frac{\sum_{i=1}^N IMS_i \mu_i}{\sum_{i=1}^N u_i} \quad (4)$$

donde

IMS_i : índice de subregistro de casos de ETS en el municipio i ;

u_i : total de casos de ETS identificados en el municipio i ;

N : número de municipios del departamento ($N = 41$), excluida Cali.

Diseño del plan de muestreo

Como el número de IPS de un municipio es proporcional a su tamaño, se supuso que estaba correlacionado directamente con la cantidad de casos de ETS no registrados y se definieron cuatro estratos ($l = 4$) para clasificar los municipios de acuerdo con el número de IPS privadas registradas en el marco muestral.

En la primera etapa del muestreo, se seleccionaron los municipios (conglomerados) teniendo en cuenta su distribución dentro de los estratos antes mencionados. El número de municipios seleccionados constituyó a su vez la muestra de IPS públicas, ya que, como se mencionó antes, en estos municipios solo existe una IPS pública. En la segunda etapa se seleccionaron aleatoriamente IPS de carácter privado dentro de cada municipio seleccionado en la etapa anterior.

Primera etapa del muestreo

Para calcular el número de municipios en la muestra, se tomó un nivel de confianza del 95% ($Z_{\alpha/2} = 1.96$), un error de estimación (d) de 0,1 y una proporción esperada de subregistro del 50% ($P_h = 0.5$) en todos los

estratos, bajo el supuesto de que el subregistro de casos en el departamento no podía llegar a ser tan grande como para sobrepasar el 80% de los casos, pues según funcionarios de la SSD, en los municipios pequeños las personas no utilizan las IPS privadas o estas no existen. Tomar este valor como primera aproximación permitía también maximizar la varianza del estimador dejando el error de estimación fijo. El número de municipios por estrato, el tamaño de muestra calculado, el factor de ponderación para hacer una asignación proporcional al tamaño del estrato y el número de municipios a incluir en la muestra (n_h) aparecen en la tabla 1.

Segunda etapa de muestreo

Análisis de costos

La SSD divide el departamento del Valle en cuatro zonas geográficas que agrupan los municipios. Para facilitar el cálculo de los costos de transporte (c_1) se utilizaron tres de las zonas establecidas (norte, centro y sur) y el único municipio ubicado en la zona oeste se incluyó metodológicamente en la zona norte. El valor de c_2 se obtuvo tomando una misma cantidad de dinero como pago a la encuestadora, al que se le sumó una cantidad por concepto de materiales para la encuesta (papel, tinta) y los costos de las fotocopias. La ecuación para obtener el c_1 por estrato es:

$$c_{1h} = \sum_{z=1}^3 m_{zh} t_z, \quad n_h = \sum_{z=1}^3 m_{zh} \quad (5)$$

donde:

m_{zh} : número de municipios del estrato h en la zona geográfica z ;

t_z : cantidad promedio de dinero que cuesta la realización de encuestas en un municipio ubicado en la zona z (el desplazamiento desde Cali hasta el municipio, el costo de encuestar la IPS pública, etc).

Tabla 1. Distribución de la muestra de municipios entre estratos y factores de ponderación

Estrato h	Número total de municipios en el estrato N_h	Factor de ponderación $W_h = N_h / N$	Número de municipios a incluir en la muestra por estrato $n_h = W_h n$
1. 0-3 IPS privadas	27	0,660	20
2. 4-10 IPS privadas	9	0,220	7
3. 11-20 IPS privadas	2	0,049	1
4. 21 o más IPS privadas	3	0,073	2
Totales	$N = 41$	1,000	$n = 30^{**}$

** Valor calculado con la ecuación 1 usando un alfa de 0,05 y tomando una proporción de subregistro del 50% en todos los municipios.

Muestra de instituciones

En todos los municipios de la muestra seleccionada en la primera etapa, se visitó la IPS pública que centraliza el registro de casos de ETS. En aquellos municipios donde no hay IPS privadas, solamente se tomaron los datos de la IPS pública. Finalmente, el número de instituciones a incluir en el estudio fue de 70 IPS de tipo privado (número calculado utilizando la ecuación 2) y 30 de carácter público (tabla 2).

Estimador e intervalos de confianza para el parámetro

El estimador por estrato se calcula con la ecuación^{3-10, 11}

$$\hat{p}_h = \frac{\sum_{i=1}^{n_h} M_{hi} \hat{p}_{hi}}{\sum_{i=1}^{n_h} M_{hi}} \tag{6}$$

$$\hat{p}_{hi} = \frac{u_{hi} - v_{hi}}{u_{hi}} \quad u_{hi} = t_{hi} + v_{hi} + \sum_{j=1}^{m_{hi}} r_{hij}$$

donde:

t_{hi} : número de casos registrados en la institución pública que centraliza el registro en el municipio i del estrato h , que no aparecen en la base de datos de la SSD;

v_{hi} : número de casos del municipio i en el estrato h registrados en la base de datos de la SSD;

r_{hij} : número de casos encontrados al encuestar la j -ésima IPS privada del municipio i estrato h ;

M_{hi} : total de (IPS) dentro del municipio i del estrato h ;

n_h : número de municipios de la muestra pertenecientes al estrato h .

A partir de las fórmulas desarrolladas por Scheaffer y Mendenhall, se derivaron las ecuaciones del estimador de la proporción departamental de casos de ETS no registrados en la SSD y su varianza.^{3,4} La forma del estimador del parámetro y su varianza es:

$$\hat{p} = \sum_{h=1}^t W_h \hat{p}_h \quad \hat{V}(\hat{p}) = \sum_{h=1}^t W_h^2 \hat{V}(\hat{p}_h) \tag{7}$$

donde $\hat{V}(\hat{p}_h)$ es la varianza del estimador en cada estrato, de modo que

$$\hat{V}(\hat{p}_h) = a s_h^2 + b, \quad a = \left(\frac{N_h - n_h}{N_h} \right) \left(\frac{1}{n_h \bar{M}_h^2} \right),$$

$$b = \frac{1}{n_h N_h \bar{M}_h^2} \sum_{i=1}^{n_h} M_{hi}^2 \left(\frac{M_{hi} - m_{hi}}{M_{hi}} \right) \left(\frac{P_{hi}(1 - P_{hi})}{m_{hi} - 1} \right), \tag{8}$$

$$s_h^2 = \frac{\sum_{i=1}^{n_h} M_{hi} (\hat{p}_{hi} - \hat{p}_h)^2}{n_h - 1} \quad \bar{M}_h = \frac{\sum_{i=1}^{n_h} M_{hi}}{n_h}$$

Finalmente, el intervalo de confianza para el parámetro es $\hat{p} \pm Z_{\alpha/2} \sqrt{\hat{V}(\hat{p})}$ (si la confianza es del 95%, $Z_{\alpha/2} = 1,96$).

Conclusiones

En el diseño del plan de muestreo, las características del contexto y el entorno propios del problema estudiado definen la estrategia más adecuada para seleccionar las unidades de estudio y garantizar el cumplimiento de los requerimientos metodológicos del trabajo y, por ende, de los supuestos teóricos que respaldan las técnicas estadísticas a utilizar en el análisis de los datos. Del plan de muestreo depende también el tamaño de muestra, la forma de calcularla y la ecuación utilizada para calcular el estimador del parámetro de interés.

Existen muchas rutinas computacionales para calcular el tamaño de la muestra en un estudio, pero las mismas generalmente se ajustan al diseño de muestreo

Tabla 2. Número de instituciones privadas por estrato a incluir en la muestra

Estrato h	Número de municipios del estrato a incluir en la muestra n_h	Cantidad de instituciones en los municipios de la muestra M_{hi}	Número de instituciones a visitar m_h^*
1. 0-3 IPS privadas	20	10	10
2. 4-10 IPS privadas	7	39	19
3. 11-20 IPS privadas	1	14	7
4. 21 o más IPS privadas	2	65	32
Totales	30	128	70

* Valor calculado utilizando la ecuación 2. El valor de los costos no se reporta porque es información confidencial del estudio.

más sencillo, el muestreo aleatorio simple. Cuando se requieren diseños más sofisticados, las rutinas de computador ya no son tan útiles, puesto que se requiere mucho más conocimiento del problema específico para establecer la manera como se lo abordará desde la perspectiva del tamaño muestral.

En este estudio, el tamaño de la población de unidades primarias (conglomerados) es de 41 y el tamaño de muestra obtenido en esta etapa fue de 30, lo que representa tener que visitar cerca de 75% de los municipios. Esta cantidad tan alta de conglomerados se debió a la falta de información previa sobre la varianza del estimador, lo que obligó a utilizar el máximo valor que esta puede tomar cuando se trata de una proporción (0,25). Es decir, que para el cálculo se partió de que la proporción de casos no registrados, tanto en el departamento como en sus municipios, era del 50%, lo que maximiza el tamaño de muestra posible que se tendría manteniendo fijo el error de estimación. Sin embargo, cuando se calculó el número de instituciones, la cantidad fue de 70 entre 128 (55%), lo cual también representaría una cantidad considerable desde el punto de vista teórico, pero esta vez lo que realmente delimitó el número de IPS incluidas en la muestra fue la cantidad de dinero con la que se contaba para realizar el muestreo en cada uno de los estratos.

Si se tiene en cuenta que lo ideal para obtener el valor más aproximado a la cantidad real de casos de ETS no registrados en el departamento es realizar un censo entre las instituciones de todos sus municipios (lo cual no era posible por restricciones económicas), entonces la muestra de 70 IPS podría verse como una cantidad reducida.

Generalmente, el presupuesto hace que el investigador termine realizando su estudio con una cantidad menor de unidades de las que —por consideraciones teóricas e ideales— debería tener. Dentro de los fines prácticos establecidos, el desarrollo teórico realizado para obtener la ecuación para el cálculo del tamaño de muestra mostró que visitar 100 instituciones, entre públicas y privadas, era posible contando con las condiciones logísticas y el cronograma establecido para la realización del trabajo de campo.

En este trabajo se definió el indicador de subregistro sobre la base de que el número total de casos observado en la institución visitada sería la cantidad total atendida en el período de estudio y de que los casos reportados a la SSD serían siempre una cantidad menor o igual a lo encontrado en la IPS. Sin embargo, en este estudio no se realizó confrontación caso a caso de lo registrado en la IPS con lo registrado en la SSD,¹⁵ de haberlo hecho, posiblemente se tendrían sujetos que habrían figurado en la base de la SSD pero no en la IPS, lo cual es poco probable, pero no imposible. De haberse dado esta situación, la ecuación 3 sufriría algunos cambios, pero, en general,

el resto de las ecuaciones planteadas seguirán siendo las mismas, puesto que no dejaría de tratarse de una proporción de casos no registrados.

De acuerdo con la forma como se definió el indicador de subregistro, este se estimó a través de una proporción; sin embargo, es de anotar que el IMS pudo haber tomado otra forma, como por ejemplo, el total de casos no registrados, es decir, la diferencia entre lo observado en la IPS y el número de casos registrados en la SSD en el momento específico de la encuesta. También se pudieron haber revisado las diferencias en los registros durante varios años y haber calculado un promedio de ellas. En estos dos casos, la forma de construir el estimador habría sido diferente y muchos de los cálculos mostrados aquí habrían resultado más complejos, requiriendo cambios en las fórmulas del tamaño de muestra mas no en el diseño del plan de muestreo; este es independiente de la forma como se defina el parámetro a estimar en la población.

El ajuste del marco muestral debe ejecutarse mucho antes de realizar las encuestas, incluso antes de calcular el tamaño de muestra. Igualmente, el ejercicio de aplicación a un grupo piloto es bastante útil para evaluar tanto la calidad del marco como los instrumentos de recolección de datos.

Es importante mencionar que la experiencia descrita en este artículo puede tener aplicaciones en el estudio de otros problemas de salud en el contexto del actual Sistema General de Seguridad Social y Salud de Colombia que tengan como unidad muestral la IPS; por ejemplo, la identificación de otros eventos de salud de notificación obligatoria y eventos no infecciosos como trauma y violencia, pero también la evaluación de la calidad en la atención en las IPS públicas y privadas y su comparación.^{13, 14}

Finalmente, el estudio del subregistro en enfermedades de transmisión sexual es un tema complejo, pues existen varios niveles de subregistro.¹⁶ En la experiencia que se describe, se propuso estudiar la contribución al subregistro debida a la falla en la notificación de los casos identificados en las IPS de los diferentes municipios. La estimación de otros niveles de subregistro escapa a las posibilidades de este tipo de estudio, como es el caso de los pacientes con una ETS que consultan y son tratados en las farmacias locales y que nunca llegan a la IPS y que, por lo tanto, se quedan sin identificar.¹⁶

Agradecimientos

Los autores expresan sus agradecimientos al personal de digitadores de la SSD por su colaboración en el manejo de los archivos de datos de enfermedades de registro obligatorio, a la enfermera epidemióloga Patricia Bustamante por sus aportes al conocimiento de la población de estudio, al estadístico Helmer Zapata por su colaboración en la fase de diseño del plan de muestreo,

a la doctora María Consuelo Miranda y la enfermera Ana Marlen Obando del Centro Internacional de Entrenamiento e Investigaciones Médicas (CIDEIM) por el apoyo prestado en la depuración y limpieza del marco muestral y en el levantamiento de datos de campo.

Referencias

1. Silva LC. Diseño razonado de muestras y captaciones de datos para la investigación sanitaria. Madrid: Ediciones Diaz de Santos; 2000
2. Hassig S, Hoffman I, Hamilton H. Seguimiento y evaluación de programas de ETS. En: Dallabetta GA, Laga M, Lamptey PR, editores. El control de las enfermedades de transmisión sexual: un manual para el diseño y la administración de programas. Guatemala: AIDSCAP/Family Health Internacional; 1997. p. 297-312
3. Cochran WG. Técnicas de Muestreo. México: Compañía Editorial Continental ; 1980.
4. Seijaz ZF. Investigación por muestreo. Caracas: División de Publicaciones de la Universidad Central de Venezuela, Facultad de Ciencias Económicas y Sociales; 1981.
5. Pandurang VS. Teoría de Encuestas por muestreo con aplicaciones. México: Fondo de Cultura Económica; 1956.
6. Lemeshow S, Hosmer Jr DW, Lwanga SK. Sample size determination in health studies: a user's manual. Geneva: World Health Organization; 1989.
7. Hernández SR, Fernandez CC, Baptista LP. Metodología de la Investigación. 4ª ed. México: Mc Graw Hill; 2006.
8. Silva LC. Cultura Estadística e investigación científica en el campo de la salud: una mirada crítica. Madrid: Ediciones Diaz de Santos; 1997
9. Barlett JE, Kortlik J, Higgins C. Organizacional Research: Determining Appropriate Simple Size in Survey Research. *Information Technology, Learning and Performance Journal*. 2001; 19 (1)
10. Martínez C. Estadística y Muestreo. 12ª ed. Colombia: Ecoe Ediciones 2000
11. Scheaffer R.L, Mendenhall W, Ott L. Elementos de Muestreo. 3ª ed. México: Grupo Editorial Iberoamericana S.A. de C.V; 1987
12. Ward JW, Greenspan JR. Public health surveillance for HIV/AIDS and other STDs: guideposts for prevention and care. In: Holmes K, Sparling, F, Mardh, P, Lemon, S, Stamm, W, Piot, P., editor. Sexually Transmitted Diseases. New York: McGraw-Hill; 1999. p. 1337-52
13. Ross MW, Courtney P, Dennison J, Risser JM. Incomplete reporting of race and ethnicity in gonorrhoea cases and potential bias in disease reporting by private and public sector providers. *Int J STD AIDS*. 2004 Nov;15(11):778.
14. Tao G, Carr P, Stiffman M, DeFor TA. Incompleteness of reporting of laboratory-confirmed chlamydial infection by providers affiliated with a managed care organization, 1997-1999. *Sex Transm Dis*. 2004 Mar;31(3):139-42.
15. Niccolai LM, Kershaw TS, Lewis JB, Cicchetti DV, Ethier KA, Ickovics JR. Data collection for sexually transmitted disease diagnoses: a comparison of self-report, medical record reviews, and state health department reports. *Ann Epidemiol*. 2005 Mar; 15(3): 236-42.
16. Aral SO, Wasserheit JN. STD-related health care seeking and health service delivery. In: Holmes K, Sparling, F, Mardh, P, Lemon, S, Stamm, W, Piot, P., editor. Sexually Transmitted Diseases. New York: McGraw-Hill; 1999. p. 1295-305.