
estudios estadísticos y prospectivos

Hacia un sistema integrado de encuestas de hogares en los países de América Latina

Juan Carlos Feres

Fernando Medina



NACIONES UNIDAS



División de Estadística y Proyecciones Económicas

Santiago de Chile, enero de 2001

Este documento fue preparado por Juan Carlos Feres, Jefe de la Sección de Estadísticas Sociales de la División de Estadística y Proyecciones Económicas, y Fernando Medina, Asesor Regional en Estadísticas Sociales de la misma División. Una versión preliminar de este trabajo fue presentada al I Taller regional del “Programa para el Mejoramiento de las Encuestas y la Medición de las Condiciones de Vida en América Latina y el Caribe” (MECOVI), realizado en Aguascalientes, México, del 1 al 3 de abril de 1998.

Las opiniones expresadas en este documento, que no ha sido sometido a revisión editorial, son de exclusiva responsabilidad de los autores y pueden no coincidir con las de la Organización.

Publicación de las Naciones Unidas

LC/L.1476-P

ISBN: 92-1-321704-7

Copyright © Naciones Unidas, enero de 2001. Todos los derechos reservados

Nº de venta: S.01.II.G.7

Impreso en Naciones Unidas, Santiago de Chile

La autorización para reproducir total o parcialmente esta obra debe solicitarse al Secretario de la Junta de Publicaciones, Sede de las Naciones Unidas, Nueva York, N. Y. 10017, Estados Unidos. Los Estados miembros y sus instituciones gubernamentales pueden reproducir esta obra sin autorización previa. Sólo se les solicita que mencionen la fuente e informen a las Naciones Unidas de tal reproducción.

Índice

Resumen	5
Introducción	7
I. Bases de un sistema integrado de encuestas de hogares (SIEH)	9
1. Elementos centrales	9
2. Núcleo básico de encuestas a incorporar en un SIEH.....	17
3. Obstáculos y desafíos	20
4. Lineamientos de acción	22
II. Algunas consideraciones sobre el diseño estadístico de las encuestas de hogares	25
1. Marco de muestreo	25
2. Cartografía.....	30
3. Selección de muestras.....	33
III. Control de calidad en el diseño y ejecución de encuestas de hogares	35
1. Introducción.....	35
2. Ámbito metodológico	36
3. Levantamiento de la información	42
4. Tratamiento de la información	45

Índice de recuadros

Recuadro 1: Algunas ventajas de un sistema integrado de encuestas de hogares (SIEH).....	10
Recuadro 2: Diferencia entre un SIEH y una encuesta multitemática “única”	15
Recuadro 3: Núcleo básico de encuestas de un SIEH	17

Resumen

Este documento representa un primer esfuerzo orientado a sistematizar los elementos de carácter técnico que se consideran esenciales en el diseño y puesta en operación de un Sistema Integrado de Encuestas en Hogares (SIEH). En la primera parte del trabajo se incluyen algunas reflexiones acerca de determinados aspectos conceptuales y metodológicos relevantes para el diseño de un SIEH. Asimismo, se establecen las principales diferencias que existen entre un sistema de este tipo, conformado por diversas encuestas que captan información complementaria de los hogares en distintos momentos de tiempo, y una encuesta multitemática única, que persigue recabar información -en un sólo operativo- sobre una amplia gama de características de los hogares y de las personas. Al respecto, se advierte acerca de la importancia de atender a la confiabilidad de las estimaciones, debido, por ejemplo, a que los tamaños de muestra que se emplean no siempre resultan apropiados para muchas de las variables y/o subpoblaciones que se busca estudiar.

A su vez, el documento plantea la necesidad de reforzar los aspectos técnicos vinculados, entre otros, a la construcción del marco de muestreo, la determinación del tamaño de muestra y el efecto de diseño, así como al cálculo de errores de muestreo, con el fin de determinar la validez y confiabilidad estadística de la información que se genera, y que normalmente se utiliza para diseñar y evaluar políticas y medir el impacto de los programas públicos en los niveles de vida de la población.

Introducción

Los altos índices de pobreza y desigualdad que caracterizan la situación socioeconómica de la mayoría de los países de América Latina y el Caribe, y la diversidad de las carencias a que se ven enfrentados diferentes sectores de la población, plantean la necesidad de encarar el diseño y puesta en práctica de políticas y programas cada vez más eficaces para superar los rezagos existentes y mejorar las condiciones de bienestar y equidad social.

Lo anterior se traduce, naturalmente, en una *gran demanda de información* oportuna, adecuada y de buena calidad, con el objeto de dar seguimiento y evaluar el desarrollo social. Sin embargo, y pese a los apreciables avances logrados en los últimos años, subsisten aún en muchos países *déficit* significativos en este campo, especialmente en cuanto a la cobertura geográfica y temática de las encuestas, su periodicidad, la consistencia de los datos generados por diferentes fuentes y la aptitud de las mismas para responder a los requerimientos de información impuestos por las políticas y programas sociales. Esto ha llevado en algunos casos a la *proliferación de encuestas*, o bien a *ampliar el contenido de los cuestionarios* más allá de lo que razonablemente permite la definición de cada operativo y las características y limitaciones técnicas de los instrumentos utilizados.

Por tal motivo, el BID, el Banco Mundial y la CEPAL han formulado una iniciativa conjunta, expresada en el *Programa para el Mejoramiento de las Encuestas y la Medición de las Condiciones de Vida en América Latina y el Caribe (MECOVI)*, con el fin de apoyar los esfuerzos que sobre el particular vienen desplegando los gobiernos y los organismos estadísticos de los países de la región.

El propósito central del citado Programa, es contribuir a establecer y/o fortalecer los *Sistemas de Encuestas de Hogares*, constituidos por un conjunto de encuestas debidamente integradas, y que representan una de las principales fuentes de información para el análisis, evaluación y seguimiento de la realidad social y las condiciones de vida de la población. Con ello se busca, al mismo tiempo, dotar al sector público y privado de los elementos informativos necesarios que garanticen un adecuado monitoreo de las políticas en marcha y contribuyan al diseño de programas eficientes orientados a la superación de la pobreza y el mejoramiento de las condiciones de vida de la población.

En este contexto, la presente nota tiene por objeto, en primer lugar, formular de manera sintética y con *carácter* preliminar, algunos lineamientos generales acerca del denominado *Sistema Integrado de Encuestas de Hogares* (SIEH), a efectos de incentivar la discusión orientada a precisar sus bases conceptuales, describir sus bondades y limitaciones, e identificar los obstáculos que se enfrentan en los diferentes países para el diseño y puesta en práctica de un sistema de este tipo.

Por otra parte, y debido al auge que han tenido las encuestas de hogares en América Latina y el Caribe, para el diseño y seguimiento de políticas públicas, en la segunda sección del documento se abordan aspectos inherentes a los procedimientos técnicos asociados al diseño óptimo de encuestas, poniéndose énfasis en la necesidad de contar con marcos de muestreo y plataformas cartográficas apropiadas, sobre la base del apoyo que brindan para ello los modernos sistemas de digitalización y georeferenciación. A la vez, se describen los aspectos estadísticos vinculados con la selección de la muestra, en las diferentes etapas que conlleva este proceso, y se destacan algunos de los problemas que se pueden enfrentar en caso de que no se apliquen procedimientos adecuados.

En la última sección, en tanto, se argumenta acerca de la necesidad de instrumentar procedimientos de documentación, que permitan conocer los detalles asociados a todas las etapas involucradas en el diseño de una encuesta, desde la planeación hasta su ejecución en campo y el análisis de resultados, incluyendo aquellos aspectos de interés que pueden contribuir a generar un acervo documental que incida en el conocimiento y amplia difusión de las buenas prácticas relativas a las encuestas de hogares.

Con todo, este documento representa sólo un primer esfuerzo, orientado a sistematizar los elementos de carácter técnico que se consideran esenciales en el diseño y puesta en operación de un Sistema Integrado de Encuestas en Hogares (SIEH). Por ende, se espera que las ideas que contiene sean mejoradas y actualizadas, a la luz de las experiencias que se desarrollan en los países y los estudios que vienen realizando tanto organismos internacionales especializados, universidades y centros de investigación como también empresas privadas que participan en la ejecución de encuestas en hogares a gran escala.

I. Bases de un sistema integrado de encuestas de hogares (SIEH)

1. Elementos centrales

1.1 Definición

El Sistema Integrado de Encuestas de Hogares, es una estrategia de producción de información acerca de las características demográficas y socioeconómicas de los hogares y las personas, basada en el *diseño y ejecución coordinada* de un conjunto de encuestas que, siendo diferentes entre sí en cuanto a sus objetivos y alcances, comparten un cierto marco conceptual y metodológico, complementan sus temas de investigación, sincronizan su realización en el tiempo y optimizan el uso de los recursos humanos y financieros.

Dada las aptitudes y restricciones propias de cada tipo de encuesta, la variedad de las temáticas que abordan y las múltiples demandas que se proponen atender, el SIEH constituye un esquema de trabajo que persigue recopilar información mediante diferentes instrumentos y en distintos períodos del año, conforme a una estructura programática que refleje las prioridades nacionales, los temas de interés, la oportunidad en que se requieren los datos, los recursos disponibles, así como el uso de los métodos y procedimientos estadísticos más apropiados para generar la información, en el contexto de un marco conceptual que integre y complemente las distintas investigaciones.

1.2 Objetivos

La creación, perfeccionamiento y consolidación de un SIEH obedece, principalmente, a tres propósitos básicos:

- Mejorar la calidad, cobertura, oportunidad, consistencia y complementariedad de la información que se obtiene mediante un conjunto articulado de encuestas de hogares.
- Atender adecuadamente a los diversos requerimientos de información del área económica y social, tanto en lo relativo a la naturaleza y periodicidad de la misma, como a su pertinencia para la formulación, seguimiento y evaluación de las políticas y programas sociales.
- Aumentar la eficiencia en el uso de los recursos.

Recuadro 1

ALGUNAS VENTAJAS DE UN SISTEMA INTEGRADO DE ENCUESTAS DE HOGARES (SIEH)



1.3 Diseño conceptual

No cabe postular un *modelo único* para el diseño de un sistema integrado, ya que la composición, alcance y complejidad de un SIEH está estrechamente vinculada a la realidad específica de cada país, en términos de sus necesidades de información, la infraestructura existente en materia de encuestas, la capacidad institucional y los recursos disponibles. Sin embargo, es posible identificar ciertos *aspectos básicos comunes* vinculados, por ejemplo, a la complementariedad temática de las encuestas, la construcción y uso de marcos compartidos, el diseño de muestras maestras o la utilización de los insumos de trabajo, generando con ello economías de escala que redundan favorablemente en el costo de un sistema continuo de encuestas.

En líneas generales, las diversas encuestas de hogares incorporadas en un SIEH se conciben como un *sistema* integrado debido a que:

- Complementan información de la población y de los hogares referida a distintas áreas temáticas, y programan su investigación en el tiempo de acuerdo a la naturaleza de cada tema y a los intereses específicos de la unidad ejecutora y de los usuarios de la información.
- Cada una de ellas responde a un diseño conceptual y estadístico particular y utiliza instrumentos de captura de los datos definidos en función de los temas específicos que investiga, pero preserva la homogeneidad conceptual en los temas comunes.
- Las encuestas cuentan con un marco muestral común y las respectivas muestras se pueden extraer, de estimarse conveniente, a partir de un marco maestro que evita las duplicidades y reduce al máximo las visitas a los mismos hogares, incrementando las tasas de respuesta.
- El diseño de algunas de las encuestas permite el análisis "longitudinal", en virtud del esquema de rotación de la muestra o porque ésta contempla expresamente una componente de "panel", con traslape parcial de los hogares que se investigan.
- Comparten recursos y permiten dar respuesta a los requerimientos de los usuarios de la información de manera eficiente y oportuna, reduciendo costos y optimizando la infraestructura de campo.

A su vez, la integración también hace referencia a la consistencia conceptual y metodológica de las encuestas de hogares con otras fuentes de información del sistema estadístico nacional, particularmente con los censos de población y vivienda, los registros administrativos y la contabilidad macroeconómica, generando con ello un sistema coherente que complementa información y preserva la comparabilidad básica entre las distintas fuentes.

Cada encuesta que forma parte del SIEH posee características particulares, en atención a sus objetivos específicos, diseño muestral (tamaño de muestra, precisión estadística, variables de interés), contenido temático, periodicidad, cobertura geográfica y costo. En el punto 2 se presentan los lineamientos de un modelo básico simplificado que considera tres clases de relevamientos, a saber: *encuesta de propósitos múltiples, con un módulo central orientado a dar seguimiento a la evolución del empleo y el mercado de trabajo; encuesta de condiciones de vida; y encuesta de ingresos y gastos familiares.*

No existe una regla universal para definir las encuestas que debieran ser integradas en un SIEH. Sin embargo, es frecuente encontrar que los siguientes temas suelen habitualmente ser objeto de investigación por medio de encuestas en hogares:

- Características demográficas de la población
- Migración interna e internacional
- Mortalidad y fecundidad
- Características físicas de la vivienda y disponibilidad de servicios básicos
- Tenencia de bienes electrodomésticos y de vehículos de transporte
- Niveles de escolaridad de los miembros del hogar
- Empleo, desempleo e inserción laboral
- Funcionamiento y estructura del mercado de trabajo
- Nivel, origen y composición del ingreso

- Estructura del gasto y del consumo de los hogares
- Salud y nutrición
- Acceso a los programas públicos y privados de asistencia social
- Existencia de personas discapacitadas en el hogar
- Violencia intrafamiliar y maltrato infantil
- Consumo de drogas y sustancias prohibidas
- Uso de tiempo

Normalmente los estudios se realizan por medio de *investigaciones específicas* para cada tema, o bien -teniendo como base un tema central- se incorporan al cuestionario de la encuesta las preguntas, o en ocasiones los *módulos*, relativos al conjunto de fenómenos que se desea investigar. En todos los casos, el diseño estadístico de la encuesta debiera prever esta situación, a fin de que no se produzcan sesgos en los datos y se afecte severamente la precisión de los resultados obtenidos, ya que el error de muestreo está altamente correlacionado con el tamaño de muestra y la varianza de los estimadores.

A su vez, debe considerarse que las *unidades de análisis* para los distintos temas son diferentes. En efecto, a pesar que, por ejemplo, tanto en una encuesta de empleo como en una de presupuestos familiares las unidades de selección de última etapa son las viviendas, en las investigaciones sobre fuerza de trabajo los esfuerzos se centran en conocer las características de ocupación de la Población Económica Activa (PEA), mientras que en un estudio sobre ingresos y gastos las unidades de análisis son las familias. Asimismo, es necesario considerar que el *tiempo de entrevista* y la duración del trabajo de campo se incrementan de manera considerable en las encuestas que tienen por objeto analizar las fuentes de ingreso y los hábitos de consumo de la población. Esta situación se refleja en el diseño de un *cuestionario* amplio y detallado, que permita registrar las distintas modalidades de ingreso y el destino del gasto de los hogares. No abundan los estudios analíticos que demuestren que a partir de un número reducido de preguntas sea posible estimar con precisión aceptable la estructura del gasto de los hogares o la distribución personal del ingreso¹.

Por otra parte, y como se señala en el capítulo II en que se presentan algunas consideraciones sobre el diseño estadístico de las encuestas de hogares, es importante identificar y seleccionar con claridad las *variables de estratificación* que resulten adecuadas para la conformación del marco de muestreo. Sobre el particular, cabe señalar las diferencias que existen en la variable que se utiliza para la estratificación del marco de muestreo cuando se trata, por ejemplo, de una encuesta de empleo o de una orientada a investigar el presupuesto familiar. Así, mientras en el primer caso la variable adecuada puede ser la PEA, la categoría ocupacional o la rama de actividad económica de los ocupados, en el caso de las encuestas de ingresos y gastos es claro que la variable de estratificación debe estar altamente correlacionada con el nivel socioeconómico de las unidades de observación y análisis, el cual se puede aproximar a través del ingreso o el gasto de las unidades de consumo. Asimismo, en cada investigación se debe identificar la *variable de diseño* que resulta más apropiada, y calcular el tamaño de muestra óptimo, considerando el efecto de conglomeración (efd) de la variable de interés y la tasa de respuesta esperada para cada una de las variables sobre

¹ Por el contrario, investigaciones recientes realizadas por el Banco Mundial, han corroborado que la precisión de los estimadores disminuye en la medida que se emplea un menor número de preguntas para indagar acerca del destino del gasto de los hogares o el origen de sus ingresos (véase CEPAL, "Medición del gasto en las encuestas de hogares". Tercer taller regional del Mecovi, Resumen y Conclusiones. CEPAL, LC/R.1914, septiembre de 1999).

las cuales se desea obtener estimación². Por último, hay que determinar la correlación que existe entre la variable de diseño y las características de la población que se desean estudiar con fines de inferencia estadística. Por ejemplo, las encuestas de empleo se diseñan para estimar con una determinada precisión la tasa de desocupación abierta (tda), pero no necesariamente para realizar estimaciones de pobreza a partir del nivel de ingreso de los hogares, ni tampoco para estudiar el total de empleo generado, el número de subempleados invisibles o el volumen de personas que participan en el servicio doméstico, entre otros estimadores que frecuentemente se construyen a pesar de que no hayan sido considerados en el diseño estadístico de la muestra.

1.4 Plan operativo

El diseño y puesta en práctica del SIEH implica definir, entre otras cosas, los temas a investigar, el marco conceptual general y de cada investigación, el tipo de encuestas que se levantarán, la periodicidad y cobertura geográfica y socioeconómica de cada una de ellas, y los recursos destinados a la realización de las mismas. A su vez, se deben abordar los aspectos metodológicos vinculados a la construcción del marco maestro de hogares que se utilizará para las distintas rondas, definiendo criterios eficientes para su administración, así como procedimientos técnicos que garanticen la actualización permanente de la cartografía. Asimismo, deben considerarse los aspectos estadísticos para el diseño y selección de las muestras, los métodos de estimación y el procedimiento que se aplicará para el cálculo de los errores de muestreo y los procedimientos de campo que se pondrán en práctica para la ejecución de las diversas encuestas.

En la etapa de diseño del SIEH, es necesario establecer con claridad aspectos tales como los objetivos de cada investigación, las unidades de análisis a investigar, los dominios de estudio, las subpoblaciones de interés, los productos, las fechas para las que se espera disponer de la información y el plan de explotación y análisis de los resultados. A su vez, se debe garantizar un planteamiento conceptual y metodológico que integre a las distintas investigaciones, y construir un marco de muestreo maestro que permita la selección de las diferentes muestras, ya sea para completar las rondas de una misma encuesta o para permitir la selección de muestras independientes. Por ende, se requiere además tener en cuenta la relación que existe entre las variables de estratificación -utilizadas para mejorar la eficiencia del diseño-, los parámetros de la población que se desea estimar y la(s) variable(s) considerada(s) para la determinación del tamaño de muestra.

Para la puesta en operación del SIEH, en tanto, es menester hacer hincapié en la necesidad que el diseño del programa considere la *coordinación operativa* y la *integración sustantiva* de las diferentes encuestas a realizar. La coordinación operativa del sistema, se relaciona con la utilización de los insumos comunes para la realización del trabajo de campo, en el sentido de garantizar el uso eficiente de los recursos disponibles. A su vez, la integración sustantiva se refiere a la homogeneidad conceptual y metodológica que debiera existir entre las distintas encuestas que forman parte del sistema, la cual debiera basarse en las recomendaciones internacionales vigentes y en un marco coherente con el sistema nacional de información estadística y geográfica, de manera que la información generada revele una adecuada consistencia y permita lograr la complementariedad temática del programa. De modo que el diseño del SIEH debiera tener como objetivo orientarse a maximizar las ventajas de coordinación e integración en el desarrollo de las diferentes etapas.

² Para la descripción de los procedimientos estadísticos que se aplican para este fin, véase Medina, F., "Tamaño óptimo de muestra en encuestas de propósitos múltiples", en CEPAL, Planificación y desarrollo de encuestas de hogares para la medición de las condiciones de vida, LC/R. 1826, julio de 1998.

1.5 Diferencia entre un SIEH y una encuesta multitemática "única"

La formación de un SIEH implica armonizar y complementar las distintas encuestas que se desean realizar, cada una de las cuales presenta, en función de sus objetivos y diseño, determinadas aptitudes y limitaciones para captar información socioeconómica y demográfica de los hogares. En ese sentido, el sistema integrado difiere del concepto tradicional de encuesta única, politemática, que persigue investigar con un sólo instrumento, en un sólo ejercicio de campo y bajo el mismo diseño estadístico, una variedad extremadamente amplia de fenómenos, pretendiendo atender simultáneamente, y con la misma precisión estadística, diversos objetivos analíticos -para subpoblaciones muy disímiles y heterogéneas-, lo que por lo general dificulta la determinación de un tamaño óptimo de muestra e incide en la confiabilidad de los resultados para las distintas variables de interés.

Como se sabe, en la práctica todas las encuestas de hogares que se llevan a cabo se pueden considerar de propósitos múltiples. Sin embargo, así como rara vez se diseña una encuesta para obtener estimaciones acerca de una sola variable, tampoco es usual -ni lógico- pretender diseñarla de manera que cubra, con precisión adecuada, todos los requerimientos de información existentes³. Este punto de vista goza de general aceptación, al extremo que en la práctica las oficinas de estadística realizan normalmente distintos tipos de encuestas, con carácter complementario, para atender a las variadas demandas de información que les plantean los usuarios. Tales encuestas tienen objetivos diferentes y, por ende, los diseños de muestra se deben elaborar considerando distintas unidades de selección, observación y análisis, con dominios de estudio específicos, asumiendo los aspectos estadísticos vinculados a la determinación de los tamaños de muestra (variabilidad del estimador, efecto de diseño y tasa de respuesta estimada), así como aquellos relativos a la cobertura geográfica y temporal, y la calendarización de los trabajos de campo y entrega de resultados.

Así, mientras en el esquema conceptual de un SIEH subyace la premisa de que los requerimientos informativos de un cierto ámbito socioeconómico se cubren mediante un conjunto de investigaciones diferentes, pero vinculadas entre sí, en las encuestas multitemáticas del tipo de las que se comentan -y que por lo general se llevan a cabo de manera esporádica- se intenta recabar en un mismo levantamiento información sobre una amplia gama de temas, normalmente muy distintos entre ellos, que difícilmente pueden ser atendidos con la misma precisión estadística (véase el recuadro 2).

³ Esta situación es la que frecuentemente se enfrenta cuando se pretende diseñar una investigación por muestreo que permita el estudio de los niveles de vida de las familias. En la mayoría de las aplicaciones prácticas se pretende obtener, en un sólo ejercicio estadístico, información sobre distintas variables relacionadas con las condiciones de pobreza de los hogares, así como con los factores de carácter estructural que inciden en la distribución del ingreso. Además, es frecuente que se incorporen múltiples temas adicionales, lo cual necesariamente incide en la precisión estadística de los resultados, generando sesgos de estimación que por lo general se desconocen y no se avalúan.

Recuadro 2

DIFERENCIA ENTRE UN SIEH Y UNA ENCUESTA MULTITEMÁTICA "ÚNICA"

SIEH	
<p style="text-align: center;">Conjunto de encuestas, donde cada una tiene:</p> <ul style="list-style-type: none"> - Diseño estadístico particular - Metodologías de investigación apropiadas a cada tema - Distintos instrumentos de captación de los datos - Temas relativamente acotados - Periodicidad ajustada a los requerimientos de cada área 	
ENCUESTA POLITEMÁTICA "AMPLIA"	
<p style="text-align: center;">Por lo general utiliza:</p> <ul style="list-style-type: none"> - Un Diseño para toda la investigación - Metodologías e instrumentos únicos para casi todos los Temas - Temática extremadamente amplia y diversa - Periodicidad más bien esporádica 	
<p>A mayor amplitud temática</p> <p>Mayor tamaño de muestra requerido → precisión de los estimadores → costo de la investigación</p> <p>Mayor tiempo de entrevista → calidad de la información → costo de la entrevista</p> <p>A menor periodicidad</p> <p>Menor oportunidad del dato → relevancia para la toma de decisiones → capacidad de acción en el corto plazo</p> <p>Encuestas independientes</p> <p>Registro con información limitada → plantea desafío de la integración → acota las posibilidades analíticas</p>	
<p>PRECISION ↔ COSTO ↔ OPORTUNIDAD ↔ ALCANCE ANALITICO</p>	

Identificar unidades de muestreo comunes y recabar información sobre múltiples temas para la misma unidad de observación es una ventaja importante, ya que permite identificar las relaciones de causalidad existentes entre diversas variables que inciden en el comportamiento de un determinado fenómeno. Asimismo, el hecho que el levantamiento se concentre en un momento específico, puede llevar a reducir los costos de operación -por unidad de observación-, sin desconocer que esta situación obliga a disponer de la totalidad de los recursos para cuando se fija la fecha en que se efectuará el operativo de campo, a partir del cual se generará la información de interés. Sin embargo, con respecto al costo se debe tener presente, además, que si en una encuesta de este tipo el propósito fuera investigar en un sólo ejercicio todos los temas con el rigor requerido, es muy probable que tanto la amplitud del cuestionario, el tiempo de entrevista y el tamaño de muestra, se incrementarían en forma considerable, al punto de elevar su costo más allá que el de cualquier encuesta continua que recabe datos sobre temas específicos.

De allí que por el volumen de información que normalmente recopilan estas "encuestas integradas" y los costos asociados a su diseño y operación, es muy difícil darles un carácter

permanente o aplicarlas en forma continua, dado que la mayoría de los países no disponen de recursos suficientes para financiarlas con el presupuesto regular con que cuentan las oficinas de estadística.

A su vez, desde un punto de vista técnico, la generación de información por esta vía también puede presentar algunos problemas, que no son menores. Debido a su carácter multitemático, estas encuestas enfrentan habitualmente un conflicto entre la precisión de los resultados y el tamaño de muestra utilizado. Como ha sido documentado por diferentes investigadores⁴, en una encuesta de propósitos múltiples muy amplios, es difícil lograr una adecuada compatibilidad entre los objetivos de la investigación, el tamaño de muestra óptimo y los recursos disponibles. Lo anterior significa que para lograr similar precisión para todas las variables de interés, se tendría que incrementar de manera considerable el tamaño de muestra y, por ende, el presupuesto de la encuesta. En ese sentido, la práctica común indica que necesariamente se debe sacrificar la precisión estadística de muchas de las estimaciones y, por ende, reconsiderar los objetivos de la investigación, poniendo énfasis en los verdaderos alcances del estudio, las restricciones del análisis y la confiabilidad de la información generada para dominios de estudio específicos.

Para ilustrar esta situación, baste considerar el caso en que a partir de una encuesta de empleo se pretenda estimar el ingreso y el consumo promedio de los hogares, con la misma precisión estadística con que se obtiene la tasa de desocupación abierta (TDA). Una primera consideración que se debiera realizar durante la etapa de análisis de los datos, es preguntarse cuál fue la variable utilizada para la determinación del tamaño de muestra: la TDA o el coeficiente de variación del ingreso (gasto) promedio del hogar; asimismo, también es importante conocer cuál fue el efecto de diseño y la tasa de respuesta que se utilizó para ajustar el tamaño de muestra y finalmente determinar el error de muestreo de los estimadores.

Al respecto, es conocido el hecho que el número de selecciones que se requieren para estimar con una precisión determinada una proporción o tasa, es menor que el requerido para calcular los promedios o totales poblacionales⁵. De modo que si el tamaño de muestra de una encuesta se calcula a partir de la proporción de población desocupada, con un cierto nivel de confianza, es muy probable que las estimaciones para los promedios de ingreso y gasto tengan una mayor variabilidad y su error relativo sea superior al requerido por el investigador para realizar inferencias y adoptar decisiones de política. Problema que se agudiza aún más cuando a partir de la muestra se desea realizar inferencias para dominios de estudio con características específicas, y no definidos entre los objetivos de la encuesta. Por ejemplo, es muy frecuente que con datos de una encuesta de hogares que se registró el ingreso familiar, se determine el número de indigentes que están desempleados y que provienen, por ejemplo, del sector primario de la economía. En ese caso, es muy probable que este segmento de población no tenga una correlación significativa con la variable de diseño y, por lo tanto, la baja frecuencia de observaciones con esta característica se refleje en el elevado error de muestreo de los estimadores. Asimismo, es común que los analistas procedan a ajustar modelos econométricos para subpoblaciones con pocas observaciones, confiando en las propiedades de normalidad de los estimadores para muestras pequeñas. Sin embargo, en esas circunstancias, además de considerar el número de observaciones involucradas, es necesario evaluar la consistencia del método de estimación utilizado, a efectos de garantizar que se cumplan los supuestos del modelo y se logre la precisión estadística deseada.

Por su parte, bajo el esquema del SIEH las mediciones que se efectúan sobre temas específicos tienen, por lo general, mayor precisión estadística y menor costo de operación, y por tanto, generan información más confiable y pueden efectuarse con mayor regularidad. Al mismo tiempo, las encuestas permanentes tienen la capacidad de captar información sobre temas

⁴ Véase por ejemplo, Kish, L., "Survey sampling", New York, Wiley & Sons, 1965, y Medina, F., op. cit.

⁵ Cochran, W., "Sampling techniques", second edition, New York, Wiley, 1963.

adicionales al objetivo central para el que fueron diseñadas, en la medida que se hagan las modificaciones pertinentes en el tamaño de muestra y la conformación del marco de muestreo guarde relación con la temática que se desea medir. Esto sin duda incrementa los costos de operación, pero de manera marginal, ya que existe una infraestructura para el trabajo de terreno que se aprovecha con mayor intensidad. A su vez, mediante este tipo de estrategias es posible generar información para estudios longitudinales, que permitan analizar la dinámica de determinados fenómenos económicos y sociales.

Sin embargo, a pesar de las bondades señaladas, las encuestas que se realizan en el marco de un SIEH difícilmente se pueden integrar desde un punto de vista temático, por lo que el principal desafío -como se menciona más adelante- consiste en diseñar métodos adecuados para complementar los datos obtenidos mediante las diferentes encuestas.

2. Núcleo básico de encuestas a incorporar en un SIEH

En una versión inicial y simplificada, el SIEH se puede concebir a partir de la integración de por lo menos *tres tipos de encuestas de hogares*:

- Encuestas de Empleo (de propósitos múltiples)
- Encuestas de Condiciones de Vida
- Encuestas de Ingresos y Gastos

Cada una de estas encuestas posee *características particulares*, en atención a sus objetivos específicos, diseño muestral, contenido temático, periodicidad, cobertura geográfica y costo.

A continuación se describen esquemáticamente algunos de estos aspectos, con la finalidad de contribuir a avanzar en la discusión orientada a precisar los elementos e interrelaciones que pueden conformar un sistema integrado de encuestas de hogares en los países de América Latina (véase el recuadro 3).

Recuadro 3
NÚCLEO BÁSICO DE ENCUESTAS DE UN SIEH

TIPO DE ENCUESTA	PERIODICIDAD	COBERTURA GEOGRÁFICA
DE EMPLEO	Continua	Grandes ciudades o áreas metropolitanas
	Trimestral	Resto del área urban
	Semestral	Área rural
		↓ NACIONAL
MEDICIÓN DE LAS CONDICIONES DE VIDA	Cada 2 años	NACIONAL (con representatividad a nivel de comunas o distritos)
INGRESOS Y GASTOS	Cada 10 años	NACIONAL
	(Anual)	
	Cada 5 años	
	(A mitad del decenio) (Trimestral)	URBANA

2.1 Encuestas de empleo (de propósitos múltiples)

2.1.1 Objetivos específicos

Medición y seguimiento del empleo, desempleo y caracterización del mercado de trabajo; e investigación de temas asociados.

2.1.2 Contenido temático

Módulo central: Características generales de la población, educación, medición del empleo, desempleo, caracterización del mercado de trabajo e ingresos.

Suplementos: Ampliación y profundización de los temas anteriores y complementación de nuevos temas. Los módulos especiales debieran, además:

- Formar parte de un programa con frecuencias predeterminadas.
- Aplicarse a la totalidad de la muestra o a una submuestra específica, según la naturaleza del tema.
- Incluir un número de preguntas que no represente una carga excesiva para los hogares de la muestra, al punto de afectar la confiabilidad de la información recogida en el módulo central e incrementar innecesariamente la tasa de rechazo.

2.1.3 Periodicidad (mínima)

Grandes ciudades o áreas metropolitanas: **Continua**, con estimaciones mensuales (mediante la técnica de promedios móviles) para las principales variables del mercado de trabajo.

Resto del área urbana : **Trimestral**, con integración de la medición del último mes del trimestre en las grandes ciudades, para generar estimaciones del área urbana nacional.

Área rural: **Semestral**, con integración de la medición urbana del segundo y cuarto trimestre, para generar estimaciones para el contexto nacional.

2.1.4 Cobertura geográfica

Nacional, sujeta a las frecuencias señaladas en el punto anterior.

2.2 Encuestas de medición de las condiciones de vida

2.2.1 Objetivos específicos

Evaluación y seguimiento de las condiciones sociales y de las situaciones de pobreza. Persiguen investigar con mayor extensión o profundidad ciertos aspectos de las condiciones de vida de las personas, como el acceso a determinados satisfactores y, muy especialmente, el impacto de políticas y programas sociales en el bienestar de diferentes sectores de la población.

2.2.2 Contenido temático

Además de las características generales y ocupacionales de la población, investigan con bastante detalle no sólo el origen de los recursos (ingresos) globales del hogar, sino también el acceso efectivo a determinados programas públicos y privados (alimentación escolar, subsidios alimentarios o de vivienda, transferencias monetarias o en especie, financiamiento en efectivo,

capacitación, etc.) y a los servicios públicos (salud, educación, infraestructura básica, saneamiento ambiental, etc.). Ello permite medir el impacto de las políticas y evaluar el grado de focalización de esos programas, conjuntamente con el análisis de importantes atributos de los grupos beneficiarios, a partir de información del hogar y de sus miembros recogida por la encuesta y no así por los registros administrativos propios de cada programa.

2.2.3 Periodicidad

Cada dos años, con una duración del trabajo de campo de entre uno y tres meses.

2.2.4 Cobertura geográfica

Nacional, con representatividad a nivel de unidades geográficas pequeñas (comunas, localidades o distritos).

2.3 Encuestas de ingresos y gastos

2.3.1 Objetivos específicos

Análisis de la estructura del ingreso y el consumo de los hogares; cálculo de las ponderaciones del Índice de Precios al Consumidor; determinación de las elasticidades de los diferentes bienes y servicios; formación del ingreso y; estudio del nivel y distribución del ingreso y el gasto familiar.

2.3.2 Contenido temático

Características generales de las personas, hogares y viviendas y registro detallado de las fuentes de ingreso y el destino del gasto (monetario y no monetario) de las familias. Cabe señalar que a partir de esto último es posible definir -entre otros- las canastas básicas de alimentos cuyo valor se utiliza en la estimación de las líneas de pobreza, así como de aquellas que se requieren para construir y actualizar el índice de precios al consumidor. Asimismo, estas encuestas efectúan un control (estadístico) de la estacionalidad del ingreso y de los gastos, en la medida que se entrevista a distintos hogares habitualmente a lo largo de todo un año, lo que redundará en una calidad relativamente buena de la información. A ello contribuyen también los instrumentos y procedimientos de captación, los métodos de evaluación (balance ingreso-gasto, verificación de precios y presentaciones de los distintos productos en los lugares de compra) y la selección adecuada del entrevistado que posea la mayor información acerca de los hábitos de gasto de la familia (en muchos casos informante directo). Además, suelen integrarse en el marco conceptual de las Cuentas Nacionales, especialmente en lo que respecta a la estructura de consumo de los hogares, y su temática hace posible -en algún grado- el estudio del comportamiento de los gastos de consumo en respuesta al impacto de determinadas políticas económicas de coyuntura.

2.3.3 Periodicidad

Cada cinco años⁶. Eventualmente, la encuesta de mitad del decenio podría realizarse con una duración del trabajo de campo inferior a doce meses (por ejemplo un trimestre⁷), seleccionándose ese período de acuerdo a criterios de representatividad de determinados promedios anuales.

⁶ Actualmente, en los países de América Latina y el Caribe estas encuestas se realizan, en general, cada 10 años. Una excepción importante es México, donde se ha conformado un programa de carácter permanente, que contempla el levantamiento de una encuesta de este tipo cada dos años.

2.3.4 Cobertura geográfica

Nacional. De seguirse la modalidad señalada en el punto anterior, podría ser una encuesta nacional (la de duración anual) y otra de cobertura sólo para las zonas urbanas.

3. Obstáculos y desafíos

3.1 Mejorar los marcos de muestreo

Diseñar y construir *marcos completos o exhaustivos* es una tarea de la mayor importancia, a fin de mejorar la cobertura de las encuestas. Esta completitud no obsta para que en ocasiones puedan existir regiones del país o segmentos de la población que sean excluidos deliberadamente del marco de muestreo, por razones ligadas a los objetivos de las distintas investigaciones, falta de recursos, limitaciones de acceso, baja incidencia del fenómeno a estudiar o escasa población registrada⁸. En el capítulo II se examinan en detalle algunos problemas y desafíos que supone avanzar en esta etapa crucial de la instalación y desarrollo de un SIEH.

Asimismo, como se ha insistido, *mejorar y actualizar el marco* que se utiliza para la selección de la muestra es un requisito indispensable para incrementar la calidad de la información que se genera por medio de las encuestas de hogares. Sin embargo, se constata que generalmente los países no dedican recursos suficientes a esta actividad, lo que puede llevar a que en los períodos intercensales se trabaje sin considerar que las ciudades se transforman, las viviendas aumentan, las edificaciones individuales derivan en estructuras de departamentos, la población crece y se mueve libremente en el territorio o que las manzanas –o bloques de viviendas- aumentan de tamaño y desbordan los límites establecidos para las unidades primarias y secundarias de muestreo. Cuando esto sucede, se alteran los pesos relativos de las distintas unidades de selección, modificando las probabilidades de selección de las viviendas, sin que la estructura operativa de la encuesta esté en capacidad de incorporar estos cambios en la base de datos y en la cartografía en uso, lo que limita las posibilidades de que la encuesta se convierta en una buena alternativa para captar la dinámica demográfica, con relación al crecimiento, localización y composición de la población.

Habitualmente la expansión de los resultados de una encuesta se hace con base a una proyección de población que no siempre es apta para incorporar todos los señalamientos anteriores, y de manera especial la distribución de la población en el territorio, lo cual puede distorsionar las tasas de crecimiento poblacional de las distintas agrupaciones geográficas o administrativas del país. Además, en la medida que una encuesta se realiza en una fecha muy distanciada del último evento censal, las diferencias entre las proyecciones de población y los recuentos de personas obtenidos por medio del censo se pueden agudizar, incidiendo en la confiabilidad de los resultados de la encuesta.

Por otra parte, una vez que se dispone de los resultados definitivos del censo, las estimaciones pasadas de las encuestas normalmente no se corrigen, lo que afecta la *evolución de las series históricas* e introduce sesgos de estimación en la medida que los datos se utilicen para efectuar proyecciones sobre el comportamiento de distintas variables de interés.

⁷ En el caso de México, este es el período de estudio definido y que coincide con el del trabajo de campo. Al respecto, debiera profundizarse en el análisis de las implicaciones técnicas de esta opción, especialmente en lo que se refiere a la calidad y consistencia de la información generada, vis a vis aquella que se obtiene mediante la utilización de un período anual.

⁸ En las situaciones en que esto ocurre, se debiera tener la previsión técnica necesaria para considerarlo en el cálculo de los factores de expansión y registrarlo en los documentos metodológicos de la encuesta, con el fin de que los usuarios conozcan las restricciones en el uso de la información.

De allí que sea muy importante generar consenso respecto de la necesidad de *actualizar el marco de muestreo por lo menos una vez al año*, aprovechando para ello la infraestructura existente para la realización del trabajo de campo de las encuestas periódicas. Por ejemplo, durante el operativo de campo se puede adoptar la estrategia de asignar un día más en el trabajo de los enumeradores, a fin de que recorran su zona de trabajo y registren las diferencias que existan entre el plano de línea utilizado para identificar su área de trabajo y la situación observada en el momento de realizar las visitas a los hogares. Del mismo modo, se debe actualizar el listado de viviendas, como una actividad rutinaria previa a la selección de las viviendas que formarán parte de la muestra.

Otra actividad a la cual se deben dedicar recursos suficientes, es la *sistematización del marco de muestreo y de la información que sirve de base para su conformación y actualización*. En ese sentido, la *digitalización de la cartografía* se presenta como una tarea necesaria que permite disponer de planos de línea de alta calidad, que apoyan de manera eficaz las labores de identificación de los segmentos de viviendas y de las unidades de observación, así como la actualización inmediata de bases de datos georeferenciadas a partir de un Sistema de Información Geográfica (SIG), representando un insumo fundamental para el desarrollo de cualquier investigación por muestreo. A su vez, la sistematización de los procedimientos de estratificación, selección de la muestra, determinación de las probabilidades de selección, y cálculo y corrección de los factores de expansión, permiten llevar un mejor control de las actividades de campo y reducir los errores humanos en los procesos de trabajo, evitando manipulaciones innecesarias que pueden conducir a sesgos de estimación en los resultados.

3.2 Vincular los diseños de las encuestas

En el contexto de un SIEH es evidente que los diseños de muestra se diferencian según las características propias de cada encuesta (objetivos, contenido temático, periodicidad, unidad de análisis, etc.). De allí, que sean explícitas también las diferencias relativas a las variables utilizadas para la estratificación del marco de muestreo y la determinación del tamaño de muestra, las unidades de selección, observación y análisis y los dominios de estudio sobre los que generan información, así como la confianza y precisión de los indicadores.

La construcción de marcos maestros de muestreo contribuye a reducir el costo de diseño y operación de las encuestas, en la medida que el presupuesto requerido para su conformación se distribuye entre las distintas investigaciones que se desean realizar. A su vez, bajo una estrategia integrada, los diseños de muestra permiten definir el tamaño apropiado de las unidades de primera y segunda etapa de selección, así como el número de las mismas que debe seleccionarse, a fin de satisfacer las necesidades planteadas en un programa de mediano plazo definido para recabar información sobre diferentes temas.

Mediante esta práctica, también es posible diseñar procedimientos de administración eficientes, de tal suerte que se eviten al máximo las duplicidades de las unidades de observación y los hogares no sean visitados en forma repetida, buscando minimizar el rechazo por parte de los informantes.

Sin embargo, uno de los grandes retos estadísticos que se deben afrontar en el futuro mediano es el desarrollo de procedimientos teóricos y metodologías de carácter práctico que permitan *interrelacionar los resultados* obtenidos por medio de las diversas encuestas que conforman un sistema integrado. Ello es así, dado que a pesar que en un SIEH se logre la complementariedad temática y se sincronice la ejecución de las diferentes encuestas, no es clara la manera de integrar

los resultados para los efectos de la caracterización de fenómenos multivariados, como es el caso del nivel de vida de los hogares⁹.

En efecto, sabemos de la conveniencia de interrelacionar las bases de datos de diferentes encuestas de hogares que estudian distintos temas. No obstante, en el contexto de un SIEH normalmente las encuestas que se realizan son diseñadas con base en muestras independientes, donde los hogares visitados difieren entre las distintas investigaciones. Más aún, incluso en el caso en que se tratara de los mismos hogares, pero bajo esquemas de selección distintos, subsisten las diferencias anotadas con respecto a la estratificación del marco de muestreo, las variables de diseño, los pesos relativos de las unidades de selección, los factores de expansión y la no respuesta, que constituyen -entre otras limitantes- algunos de los problemas que impiden sobreponer dos bases de datos para complementar el estudio de un determinado fenómeno.

De modo que, aun cuando el SIEH calendarize en un primer momento, por ejemplo, una encuesta de fuerza de trabajo y luego una encuesta de ingresos y gastos, no es posible utilizar, al menos en forma directa, la información de empleo de la primera para complementar la segunda, y evitar así que nuevamente se tenga que captar información sobre las características ocupacionales de la población. Pese a que en los dos casos se trata de una encuesta de hogares, cuyas muestras pueden haber sido seleccionadas a partir de un mismo marco de muestreo, las distintas probabilidades de selección de las observaciones limitan la operación deseada.

3.3 Optimizar el uso de los recursos

Mejorar la eficiencia en el uso de los recursos técnicos, humanos y materiales, generar sinergías entre los distintos operativos que se llevan a cabo, aprovechar intensamente la infraestructura disponible e incentivar la producción de economías de escala, son algunos de los desafíos a los que se ven permanentemente enfrentadas las oficinas nacionales de estadística, y que cobran particular importancia en el contexto de instalación y desarrollo de un SIEH.

De allí que, ante la crónica insuficiencia de recursos que suele afectar a estas instituciones, se debe reforzar el empeño por desarrollar procedimientos que contribuyan a optimizar las diversas actividades, no sólo en cuanto a su calidad técnica sino también en lo que se refiere al uso de los recursos disponibles. Baste señalar, tan solo a modo de ejemplo, los avances posibles de lograr en este sentido en las tareas propias de la fase de diseño y selección de la muestra; o bien, en las de actualización del marco de muestreo, que sigue siendo una de las principales limitaciones para reducir el margen de error de las estimaciones generadas a partir de las encuestas. Sobre esto último, y dado que en muchos casos es difícil que en el corto plazo se disponga de los recursos necesarios para mantener al día la información cartográfica, es conveniente aprovechar al máximo las visitas al terreno para obtener información sobre los cambios que se observan en la distribución y destino de las viviendas, la creación de áreas de nuevas construcciones, la desaparición de unidades habitacionales, etc.

4. Lineamientos de acción

En los últimos tiempos, y como fruto de una creciente demanda de información del área social, en la región se han evidenciado con mayor fuerza ciertos vacíos con relación especialmente a la cobertura temática y la periodicidad de las encuestas, como también respecto a la consistencia de los datos generados por diferentes fuentes y a la aptitud de las mismas para responder a las

⁹ La razón fundamental es que en las diferentes encuestas los hogares en muestra son distintos, lo cual no permite integrar la información. Por otra parte, y dado que los tamaños de muestra difieren según el tipo de investigación, se presenta un problema adicional que se refiere a la definición de las probabilidades de selección, y por ende al cálculo de los factores de expansión que se utilizan para ponderar las observaciones en muestra.

necesidades impuestas por las políticas y programas sociales. Esto ha llevado en algunos casos a la proliferación de encuestas, o bien a ampliar el contenido de los cuestionarios más allá de lo que razonablemente permite la definición de cada operativo y las características y limitaciones de los instrumentos utilizados. Un ejemplo de ello es el intento de abordar la investigación de los temas propios de las encuestas de ingresos y gastos mediante procedimientos más ágiles y baratos, pero menos eficaces, a partir de las encuestas de propósitos múltiples.

De allí que, teniendo presente que la situación en que se encuentran los distintos países de la región en materia de infraestructura de encuestas de hogares es muy diversa, para los próximos años se hace imprescindible impulsar acciones diferenciadas, acordes a la experiencia y grado de avance en que se encuentra cada país. Estas acciones, sin embargo, debieran apuntar en último término a establecer en el mediano y largo plazo sistemas integrados de encuestas de hogares, que permitan aprovechar de manera simultánea y coherente las ventajas propias de cada relevamiento.

Desde el punto de vista regional, existen dos grandes áreas de tareas técnicas asociadas a este esfuerzo. La primera comprende todas aquellas materias referidas al muestreo, en la perspectiva de la construcción de un sistema integrado de encuestas (desarrollo y perfeccionamiento del marco muestral, definición de muestras maestras, diseño de muestras panel o semipanel, cálculo apropiado de tamaños de muestra para encuestas multitemáticas, etc.). La segunda, en tanto, abarca la recopilación y sistematización de experiencias acerca del diseño y contenido de los cuestionarios, técnicas de investigación, organización del trabajo de campo, métodos de evaluación de la calidad y confiabilidad de los resultados, integración de productores y usuarios y pertinencia de las diferentes encuestas en función de los objetivos perseguidos.

A su vez, el establecimiento y/o desarrollo de un sistema integrado de encuestas de hogares supone introducir mejoramientos institucionales y técnicos en los países, que en gran medida son funcionales al mejoramiento de las distintas investigaciones. Entre las tareas a asumir pueden destacarse:

- i) el fortalecimiento del nivel técnico de los funcionarios de las oficinas de estadística, tanto a nivel central como regional, incluyendo la conformación de un equipo permanente o parcialmente permanente de encuestadores;
- ii) la actualización o establecimiento de una muestra maestra, así como la definición de los procedimientos técnicos apropiados para la selección de submuestras;
- iii) la incorporación de una componente de panel en todos los relevamientos¹⁰;
- iv) la conformación de equipos de trabajo que incluyan la entrada de datos en campo y una alta capacidad de reentrevista, acompañado de un intenso programa de supervisión;
- v) la incorporación de programas de suplementos (módulos) preestablecidos y flexibles;
- vi) en suma, el fortalecimiento institucional de las oficinas de estadística, con la radicación en ellas de todos los operativos.

El estado general del sistema de encuestas de cada país debiera servir de base para definir las acciones conducentes a su mejoramiento. Con el fin de ilustrar algunos cursos de acción en términos de productos que pueden constituir los primeros resultados de un programa integrado, se presentan a continuación tres ejemplos.

¹⁰ Esto se aplica actualmente en la mayoría de las encuestas de fuerza de trabajo que se realizan en los países de la región. Sin embargo, es especialmente importante que también las encuestas orientadas a evaluar los niveles de vida, la evolución de la pobreza, la desigualdad y el desarrollo humano consideren un panel de rotación apropiado, con el fin de darle seguimiento a la acción pública y su efecto en el bienestar de la población.

En un primer caso, se debieran sentar las bases para el mejoramiento del sistema integrado concentrando los esfuerzos en el fortalecimiento de la encuesta de hogares de propósitos múltiples. Esto supone, en primer lugar, aumentar su frecuencia, llegando incluso a atribuirle un carácter continuo, superando de esta manera la situación que presentan algunos de los países en cuanto a levantar la encuesta sólo una o dos veces al año. En segundo lugar, ampliar el ámbito geográfico de la encuesta, con el fin que del área metropolitana o de algunas zonas urbanas se pase a cubrir a la población de la totalidad del territorio nacional. En tercer lugar, ampliar también la cobertura temática de la encuesta, a efectos de complementar la medición del empleo y del desempleo con información relevante para un abanico de temas sociales, que permitan establecer relaciones de causalidad y medir el efecto de las políticas. En cuarto lugar, en el caso de los países más avanzados de este grupo, se sugiere la incorporación de módulos especiales, con el objeto de permitir explorar en mayor profundidad determinadas áreas de la realidad social. Por último, progresar en los aspectos metodológicos, operativos y de procesamiento de la información, con el fin de ganar en calidad y oportunidad.

En un segundo caso, los esfuerzos de nuevos levantamientos debieran concentrarse en el relevamiento de una encuesta del tipo caracterización socioeconómica, que apunte a complementar la información de la encuesta continua con un análisis más pormenorizado de las familias afectadas por situaciones de pobreza. Especial importancia tendría investigar el alcance geográfico y por estratos sociales de las políticas de gasto público, incluyendo el examen de los diversos programas que los gobiernos vienen promoviendo en los últimos años en su lucha contra los rezagos sociales. Se trataría de establecer un programa que contemple la realización de este tipo de encuestas cada dos o tres años. En este sentido, y recordando que se persigue mejorar la calidad de los datos, es imprescindible aprovechar la experiencia que ya ha acumulado la región en esta clase de encuestas.

Por último, en un tercer caso, los esfuerzos de nuevos relevamientos debieran centrarse en establecer un programa de encuestas de ingresos y gastos, en un modelo simplificado pero de mayor periodicidad, con el fin de conseguir cifras de gran calidad en torno a la evolución de los ingresos y de los gastos y de la estructura del consumo de las familias. Se trata de un objetivo ambicioso, pues requiere de mayores recursos y de una estructura institucional más consolidada. A la vez, permitiría tener un conocimiento más cabal de los efectos que los cambios estructurales en la producción, en la ocupación y en la distribución del ingreso vienen teniendo sobre la trama social, en términos de la distribución de los frutos del crecimiento económico en una nueva modalidad de desarrollo.

II. Algunas consideraciones sobre el diseño estadístico de las encuestas de hogares

1. Marco de muestreo

1.1 Generalidades

Un marco de muestreo se define como una lista exhaustiva - organizada en forma de base de datos- que contiene todos y cada uno de los elementos de una población de interés que participarán en las distintas fases de selección de la muestra. A su vez, el marco está formado por un conjunto de mapas y planos a diferentes escalas, que permiten la delimitación física de las diversas unidades de selección. Asimismo, se considera parte fundamental de éste los registros físicos de las unidades de viviendas, así como los listados en que se detallan las referencias que permiten a los encuestadores la plena identificación de las viviendas seleccionadas, así como de aquellas que sin ser parte de la muestra pertenecen a la población objeto de estudio.

El marco debiera contener información sobre la división administrativa y geográfica del país -subdivisiones políticas o zonificación estadística definida para efectuar la enumeración del censo de población-, como también sobre los volúmenes de viviendas y de población total, por grupos de edad y sexo, entre otras variables necesarias para clasificar a los hogares de acuerdo a determinadas características según los objetivos específicos de la encuesta.

En un buen marco de muestreo, todos los miembros del universo de estudio deben tener una probabilidad conocida, y distinta de cero, de formar parte de alguna muestra. Sin embargo, se debe señalar que existen estudios en que deliberadamente se excluyen algunas zonas o regiones del país, o determinadas subpoblaciones, para fines de muestreo; estos casos no violan el principio señalado, pero ello debe aclararse en los documentos que hacen referencia a la población encuestada a fin de que no se generalicen los resultados a sectores no considerados en el universo de estudio.

Por último, el marco debe estar exento de duplicaciones y omisiones, y actualizarse de manera permanente, con el objeto de reflejar los cambios estructurales que se van produciendo en la geografía del país, la conformación de las unidades de selección, su distribución física, el surgimiento de nuevas edificaciones y las construcciones que se generan debido al crecimiento de la población, así como la bajas registradas debido a demoliciones, fenómenos naturales o por el hecho de que algunas viviendas se hayan transformado en negocios comerciales.

1.2 Información del marco de muestreo

La información contenida en un marco de muestreo puede referirse a áreas geográficas o a una lista de unidades de viviendas o personas. Los estadísticos de encuestas se refieren a estas alternativas como marco de áreas y de lista, respectivamente. En la mayoría de los países de la región se utiliza habitualmente una combinación de ambos para organizar la información necesaria para elaborar los diseños de muestra de las encuestas, así como la planeación del operativo de campo.

Es común que en las encuestas donde las unidades de observación se seleccionan en varias etapas -esquemas polietápicos-, se parta de la elección de un conjunto de áreas geográficas perfectamente delimitadas, y posteriormente se elabore un marco de lista que incluye únicamente a las viviendas de las unidades que pertenecen a las áreas previamente seleccionadas. Es decir, se elabora un marco de muestreo para cada una de las etapas involucradas en la selección de la muestra. Mediante este procedimiento se reducen los costos de construcción y actualización del marco, ya que para las fases subsecuentes éstos sólo incluyen aquellas unidades que fueron seleccionadas en las etapas anteriores.

Por otra parte, los marcos suelen incorporar información sobre las diferentes unidades de muestreo, las que son agrupadas conforme a criterios que permitan mejorar la eficiencia del diseño. En aquellos casos en que se trata de un esquema polietápico, a las unidades utilizadas en la primera etapa se les denomina unidades primarias de muestreo (UPM's), y a las que se seleccionan al final, unidades últimas de muestreo (UUM's). Entre esos dos extremos existen múltiples niveles, pero lo habitual es que en las aplicaciones de carácter práctico los esquemas de selección comprendan, como máximo, dos o tres etapas.

En la realidad, es usual que las organizaciones nacionales de estadística definan la construcción de un único marco con el propósito de realizar diferentes rondas de una misma encuesta o bien investigaciones puntuales de objetivos diversos. En este caso, el marco que permite la elaboración de los distintos diseños, así como la elección de las diferentes muestras, se denomina *Marco Maestro de Muestreo* (MMM), que representa un insumo fundamental en el diseño de un SIEH. Asimismo, es en este contexto en el que se generan las *muestras maestras*, definidas como toda muestra a partir de la cual pueden seleccionarse submuestras independientes -denominadas réplicas- o grupos de unidades con algún nivel de traslape que se utilizan para diferentes rondas de una encuesta o para proyectos específicos.

1.3 Construcción del marco de muestreo

Al finalizar la enumeración censal, es práctica común que las áreas encargadas de las encuestas en las ONE's dediquen esfuerzos y recursos a la construcción del marco maestro de muestreo. Esta etapa se asocia a la conformación de una base de datos que permite el diseño y selección de diferentes tipos de muestras, vale decir que posibilita la realización de distintas encuestas o la formación de réplicas -o submuestras interpenetrantes- necesarias para la ejecución de las encuestas periódicas o continuas que forman parte de los programas permanentes que llevan a cabo los países.

En la mayoría de los países en que se realizan encuestas de hogares, el diseño de las mismas y la selección de las unidades de observación se efectúa a partir de información obtenida de los censos de población y vivienda. Sin embargo, en ocasiones, y ante la falta de información censal actualizada, se recurre al uso de registros administrativos o mapas del catastro elaborados por otras dependencias públicas para identificar a las unidades de vivienda. De modo que a pesar de la existencia en esos países de diversos marcos de muestreo, es necesario reconocer que deben modernizarse los métodos de trabajo utilizados para su administración y actualización, a fin de incorporar las nuevas tecnologías y las metodologías de punta que vienen siendo desarrolladas, lo que además permitiría contar al mediano plazo con cartografía automatizada y una base de datos adecuada para el manejo de la información.

La preparación y ejecución de los censos de población demanda la inversión de un monto considerable de recursos, dentro de lo cual -como se verá más adelante en el punto 2- la elaboración y actualización de la cartografía es una tarea fundamental. Es así que el marco de muestreo para las encuestas se va formando de hecho en las etapas de campo previas a la ejecución del censo, debido a que los trabajos censales requieren disponer de una base cartográfica actualizada y de buena calidad.

Una vez que concluyen los trabajos de enumeración censal, surge la oportunidad de diseñar los planes y programas de las encuestas con las que se produce información para atender las demandas de los usuarios en los períodos intercensales. En términos generales, la conformación de los marcos de muestreo sigue los pasos que se detallan a continuación.

En primer lugar, la base informativa generada a partir del trabajo de campo de los censos posibilita la construcción de un marco de áreas, integrado por mapas de línea a diferentes escalas y una base de datos con la división geográfica, administrativa y estadística del país, así como con alguna información de base que permita conocer el total de viviendas en los distintos contextos geográficos, además de una primera aproximación al número de personas que habitan en cada uno de ellos. Esta información debiera ser respaldada en un sistema automatizado, conformado por una base de datos relacional, que permita a los usuarios la consulta directa, y por la cartografía digitalizada, orientada a la elaboración de mapas y a su actualización en forma automatizada.

Durante la elaboración del marco, cada unidad estadística debiera identificarse en las diferentes fases y acompañarse de información complementaria que haga posible definir su importancia relativa respecto a las demás (número de viviendas, población total, población por edad y sexo, etc.). También es importante incorporar datos sobre algunas otras características de las unidades de selección, con el objeto de contar con elementos que faciliten efectuar agrupaciones -estratificaciones- que mejoren la eficiencia del diseño y disminuyan la variabilidad de los estimadores (en caso que la variable de estratificación esté altamente correlacionada con los parámetros que se desean estimar).

En segundo lugar, las unidades censales se utilizan por lo general para formar agrupaciones (que se conocen con el nombre de conglomerados y que también forman parte del marco de muestreo por áreas) definidas comúnmente como unidades de primera etapa o unidades primarias

de muestreo (UPMs). Según las características de las unidades censales, es posible que éstas puedan utilizarse como conglomerados (teniendo presente que en cualquier encuesta la primera etapa es la que genera la mayor contribución a la varianza total del diseño).

Como ha sido ampliamente documentado en textos especializados y en diversos informes técnicos elaborados por la División de Estadística de las Naciones Unidas (véase United Nations), las unidades de observación agrupadas en conglomerados son muy distintas entre sí (heterogéneas) y tienden a incrementar la varianza de los estimadores, ya que estos conglomerados son generalmente pequeños en comparación con el universo de estudio. Sin embargo, en la práctica del muestreo es común la formación de conglomerados, ya que representan espacios geográficos que se constituyen como unidades de primera etapa para la selección de la muestra (en los documentos técnicos de Naciones Unidas a estas unidades se les denomina áreas de enumeración).

Al interior de las unidades de primera etapa existe una cierta afinidad entre los elementos que las conforman, por lo que su contribución a la varianza total (intraconglomerados) es menor que aquella que aporta las diferencias entre las agrupaciones (variación entre conglomerados). Luego, al interior de los conglomerados se definen agrupaciones más pequeñas constituidas por grupos de viviendas, que usualmente se conocen como segmentos compactos o áreas de listado y que se denominan unidades secundarias de muestreo (USMs). En esta etapa se requiere contar con un listado exhaustivo de las agrupaciones, en el que se detallen las características del sector y permitan a los entrevistadores la identificación de las viviendas seleccionadas. Finalmente, al interior de las USMs se listan todas las viviendas, las que son consideradas como unidades últimas o finales de muestreo (UUMs).

1.4 Estratos y conglomerados

La estratificación se refiere a la subdivisión de una población determinada en subconjuntos con características propias. Esta acción se lleva a cabo como una etapa previa a la selección de la muestra y la(s) variable(s) que se utiliza(n) para ello debe(n) contener información acerca de todas las unidades de la población.

El objetivo de este procedimiento es reducir la varianza del parámetro de interés, por lo que se sugiere que las variables de estratificación deben estar altamente correlacionadas con aquella(s) utilizada(s) para la determinación del tamaño de muestra. De modo que los estratos son, por definición, homogéneos en su interior, lo que, de paso, establece una diferencia fundamental respecto a las características de los conglomerados.

Asimismo, los estratos no suelen utilizarse como unidades de selección debido a que es normal que todos ellos estén representados en la muestra, mientras que la naturaleza de los conglomerados los define como las típicas unidades de selección en las primeras etapas, ya que su heterogeneidad garantiza que en su interior queden incluidos diferentes elementos de la población.

Como ya se señaló, en general la formación de conglomerados tiende a incrementar la varianza muestral. Sin embargo, su utilización en el diseño de la muestra permite flexibilidad en el número de etapas de selección y reduce substancialmente los costos asociados a la construcción del marco y a la realización del trabajo de campo.

Es común que en la mayoría de las encuestas de hogares que se llevan a cabo en los países de la región se combinen las técnicas de estratificación y conglomeración, como condición necesaria e indispensable para garantizar un diseño muestral óptimo. Sin embargo, cabe precisar que la estratificación se realiza con un objetivo específico, que está inducido por los fines de la encuesta, y guarda estrecha relación con las variables de interés. En otras palabras, es muy difícil que exista una estratificación óptima que satisfaga, con la misma precisión estadística, las necesidades de

todos los temas incorporados en una encuesta y, por ende, a los distintos usuarios de la información¹¹.

Lo anterior es relevante en la medida que existen iniciativas que promueven la realización de encuestas multitemáticas que no siempre evalúan, con el rigor técnico del caso, la correlación que existe entre la estratificación del marco muestral, la variable de diseño para la determinación del tamaño de muestra y el error relativo de los diferentes indicadores que se construyen con base a las preguntas contenidas en el cuestionario y asociadas a los diferentes temas que se investigan.

Es probable que estos estudios partan de la hipótesis que existe una posible relación entre algunos de los indicadores obtenidos, la estratificación del marco y la variable de diseño. Sin embargo, también es probable que esta correlación no sea significativa desde el punto de vista estadístico, en cuyo caso se comprometen o invalidan las inferencias que se puedan hacer sobre el comportamiento de un fenómeno perteneciente a un determinado dominio de estudio. De modo que resulta indispensable estimar los errores de muestreo y calcular el efecto de diseño de las variables, con el propósito de evaluar si el tamaño de muestra utilizado resultó apropiado para los fines de la encuesta.

Por último, es importante señalar que, a diferencia de la etapa de estratificación que tiene un fin específico para cada encuesta, los trabajos que se realizan para la formación de conglomerados son de beneficio común y se pueden compartir entre las diferentes encuestas de un programa permanente o en el contexto de un SIEH. Esto significa que la definición y formación de los conglomerados, la sistematización de la información de base, la elaboración del material cartográfico y su actualización, son insumos que se pueden aprovechar para desarrollar varias rondas de una misma encuesta o para realizar investigaciones independientes sobre distintos temas, utilizando mecanismos de estratificación apropiados que mantengan una correlación significativa con la variable de diseño. Así, el costo que implica la administración de los conglomerados, con objeto de mantenerlos actualizados, se amortiza entre las diferentes rondas o encuestas, generando economías de escala que permiten la puesta en marcha de un programa integrado de encuestas de hogares a costos razonables, que informe de manera periódica sobre una amplia gama de temas de interés.

1.5 Sistematización de los marcos de muestreo

Conforme a lo expuesto, es necesario que las ONEs elaboren marcos de muestreo ad-hoc para cada una de las etapas incluidas en el diseño. Estas actividades forman parte de las labores rutinarias de las divisiones de encuestas; sin embargo, lo que aún no está muy extendido en los países de la región, es que se incorporen en los trabajos de sistematización las metodologías y facilidades que brindan los sistemas modernos de manejo de bases de datos, que constituyen una herramienta sumamente útil para apoyar la construcción y administración de los marcos de muestreo.

En este sentido, se debieran definir y asignar prioridad al logro de algunos objetivos básicos a mediano plazo, tales como: i) establecer un vínculo estrecho entre el material cartográfico elaborado para los censos de población y vivienda y las necesidades informativas para la formación de los marcos de muestreo requeridos para el diseño de las encuestas; y ii) construir marcos de muestreo sistematizados que permitan, entre otras cosas, la actualización automática de la información, el manejo de los procedimientos estadísticos para la selección de las unidades de observación, el cálculo de los factores de expansión, la generación de cartografía automatizada y la administración y actualización de las bases de datos.

¹¹ Esta consideración es de especial interés sobretodo en las encuestas de propósitos múltiples, ya que es muy frecuente que ello se pase por alto, sin que se midan sus consecuencias en la posible pérdida de precisión de algunos estimadores.

2. Cartografía

2.1 Cartografía automatizada y los Sistema de Información Geográfica (SIGs)

La identificación de las unidades de vivienda que forman parte de la muestra, requiere privilegiar su posición física antes que su precisión geográfica. Esto significa que se deben establecer claramente las diferencias entre cartografía automatizada y las facilidades y alcances de un Sistema de Información Geográfica (SIG), ya que excederse en las pretensiones de sistematización implica un gasto considerable de recursos y tiempo, que podría no justificarse ante la ausencia de una estrategia institucional para desarrollar un proyecto que vincule la ejecución de encuestas con el uso de la información estadística y geográfica en un mismo plano.

Debido a los avances tecnológicos ocurridos en los últimos años, resulta sumamente tentador el modernizar los procedimientos actuales de manejo de información. De allí, la notable relevancia que han cobrado los SIG's para la gestión de proyectos estadísticos, y el interés que ha surgido por incorporarlos en diversos campos de trabajo. La naturaleza de esta herramienta la hace muy apropiada para ubicar en un mismo plano información estadística y geográfica, de diferentes temas, en el medio físico. De hecho, la posibilidad de sobreponer cartografía básica y temática representa una de las enormes ventajas de esta tecnología, y en la actualidad los SIG's se utilizan en áreas tan diversas como la gestión ambiental, el desarrollo urbano, la mercadotecnia, los servicios públicos y privados, el desarrollo de infraestructura básica, entre otras.

Sin embargo, en el ámbito del muestreo las capacidades y usos de los SIG's se dan en un contexto más amplio que la generación de cartografía automatizada para el trabajo de campo de las encuestas. De hecho, un SIG se puede definir como un sistema de información que permite el registro de datos, su almacenamiento, la verificación de su consistencia, la integración, la manipulación y análisis, así como el despliegue visual de la información en un contexto espacial (Chorley, 1987). En términos generales, éste consta de una plataforma mínima de información, la cual debe ser resguardada en una base de datos relacional, con atributos específicos para las observaciones, a fin de que permita el manejo de las variables y la superposición espacial de diferentes temas. Además, el hecho de disponer de cartografía georeferenciada en medios magnéticos brinda la posibilidad de ubicar físicamente la información con precisión métrica.

La idea de incorporar la herramienta de los SIG's en la gestión de proyectos estadísticos fue manifestada por J. Durbin en la reunión de la Royal Statistical Society en 1987. Durbin manifestó que “si a un mapa se le puede agregar información que permita ubicar a las oficinas de correo, los registros de electores, o cualquier otro tipo de información, se podría formular una estrategia estadística para relacionar los datos económicos y sociales provenientes de los censos y las encuestas con otros datos de utilidad para la sociedad”. De modo que contar con cartografía automatizada representa una gran ventaja para los administradores de los programas de encuestas. Sin embargo, se pueden elaborar cartas y planos de línea en medios magnéticos, sin que esto signifique la disponibilidad de un SIG. Igualmente, se puede disponer de un sistema de base de datos que resguarde la información del marco de muestreo y sobreponerla en los planos de línea o cartas básicas o temáticas, sin que ello represente trabajar con un SIG.

Un aspecto que distingue a los SIG's de los actuales sistemas de manejo de información estadística y geográfica con fines de muestreo, es el hecho que para garantizar la operación eficiente de un SIG es condición indispensable que la cartografía básica se encuentre georeferenciada y se disponga de una base de datos con estructura relacional que permita el manejo de información de varios temas.

En la actualidad, la mayoría de las oficinas de estadística de la región no tienen como responsabilidad la generación de cartografía o de información sobre temas relativos al medio físico. De allí que los alcances de un SIG rebasan las atribuciones y necesidades de las divisiones de encuestas, por lo que se debe ser prudente para enfrentar el desarrollo de actividades en este campo, sin contar con un proyecto que contemple la vinculación de los trabajos de las ONEs con las tareas que realizan los Institutos Geográficos Nacionales.

Lo anterior no significa que se deba abandonar la idea de diseñar un sistema eficiente que, además de servir para los fines de ejecución de las encuestas, contribuya al estudio de diversos temas de interés en donde el posicionamiento espacial de las unidades de vivienda resulte fundamental. Por el contrario, se debiera identificar las áreas de trabajo en donde los SIG's podrían apoyar la gestión de proyectos estadísticos. En primer lugar, el uso de los SIG's puede asociarse con las etapas de diseño, recolección de información y análisis de resultados. En segundo lugar, es evidente que en relación con el diseño de encuestas, los SIG's pueden contribuir a la formación, actualización y administración de los marcos de muestreo y, en la medida que se relacionen bases de datos sobre diferentes temas, apoyar la construcción de marcos maestros. Asimismo, en el diseño estadístico de la encuesta es posible utilizar variables de la base de datos para estratificar el universo de estudio y, además, apoyar la formación de las unidades de selección, la delimitación de segmentos en las zonas urbanas y las áreas de conteo en las zonas rurales. Por su parte, la etapa de recolección de datos en campo puede ser aprovechada para captar información sobre los servicios públicos y privados disponibles en las áreas visitadas, así como su ubicación geográfica. Esta información es de gran utilidad y puede utilizarse para estudios en el ámbito del desarrollo urbano y la generación de servicios en zonas carenciadas. Por ejemplo, establecer la distancia de las viviendas a las escuelas y centros de salud y abastecimiento, o bien, determinar la distancia promedio entre las viviendas y las tomas públicas en aquellas áreas rurales que carecen de suministro de agua por tubería.

En la etapa de análisis se puede distinguir al menos dos fases: el análisis de consistencia de la información de la encuesta y la vinculación de ésta con datos disponibles sobre el medio físico (variables geográficas). En cada una de ellas los programas de análisis son, por lo general, diferentes; sin embargo, ya existen paquetes estadísticos que contemplan rutinas para el manejo de bases de datos geográficas bajo un enfoque SIG (por ejemplo, el módulo especial del Statistical Analysis System - SAS).

En síntesis, la automatización de las actividades relativas al diseño de encuestas es posible enfrentarlas por etapas. En primera instancia, y en ausencia de un proyecto nacional de SIG, las oficinas de estadística debieran orientar sus esfuerzos a sistematizar la información disponible sobre el marco de muestreo y digitalizar la cartografía en uso, sin prestar gran atención a la precisión geográfica de la ubicación física de las viviendas. No obstante, en contacto con los institutos geográficos nacionales, se debiera cautelar que en el diseño y desarrollo del sistema se observen las normas técnicas que permitan posteriormente relacionar la información de las viviendas con otros temas.

Luego, en el contexto de un proyecto más amplio, se puede asumir la tarea de georeferenciar la cartografía. En el caso que se disponga de una red geodésica, ésta se podrá utilizar con fines de actualización. Así, las brigadas de campo, mediante el uso del GPS (global position system), pueden contribuir a mantener vigente la cartografía, sobreponiendo información sobre los rasgos físicos del terreno, la infraestructura, los servicios, las vías de comunicación, etc. Cabe señalar que las metodologías modernas sobre desarrollo regional utilizan información del medio físico combinada con datos socioeconómicos y demográficos, a fin de lograr una mejor aproximación a las interacciones entre los fenómenos geográficos y sociales. Sin embargo, se debe estar consciente

que las actividades en el ámbito geográfico suelen tener fines más amplios que rebasan por mucho los requerimientos de la elaboración y administración de un marco de muestreo.

2.2 Requerimientos de la digitalización

En términos generales, para disponer de un mapa digitalizado es necesario cumplir un conjunto de requisitos básicos, entre los que cabe destacar:

- Contar con los mapas a digitalizar. Estos debieran incluir, como mínimo, la división político-administrativa del país (estados, provincias, municipios, etc.). No obstante, también es posible elaborar productos cartográficos a partir de listas de coordenadas con información en cualquier sistema de referencia (cartesiano o euclidiano), ya sea que provengan de planos elaborados directamente en campo en forma manual o con ayuda de un GPS. A su vez, otra alternativa consiste en identificar determinadas zonas sobre la base de fotografías aéreas o imágenes de satélite de alta resolución.

- Definir el nivel para el cual se requiere interpolar los datos para efectos de su representación espacial y precisión estadística, dado que las encuestas normalmente recaban información sólo de una parte del universo de estudio.

- Disponer de programas para la digitalización y captura de los datos numéricos referidos a la información del marco de muestreo o de las variables de la encuesta. Para la digitalización existen distintas alternativas como: AUTOCAD, ARC/INFO, GRASS e IDRISI. Por su parte, para la información numérica es posible utilizar hojas de cálculo como EXCEL, o manejadores de bases de datos como DBASE, FOXPRO o ACCES, entre otras.

- En un mapa digitalizado cada entidad geométrica que lo compone tiene asociado un atributo, con valor numérico, que se utiliza como identificador y permite relacionar el mapa con la base de datos tabular.

- En un contexto ideal, la actualización cartográfica se puede realizar a partir de dotar a las brigadas de campo de un GPS portátil, con el fin de referir la posición geográfica del lugar en donde se efectúa la encuesta, con una exactitud tal que permita su representación espacial y su manejo cartográfico.

- Para cumplir con eficiencia estas tareas se requiere, naturalmente, capacitar al personal en el manejo de información estadística y geográfica, así como en las funciones de digitalización, manejo de bases de datos y análisis espacial.

Los equipos y sistemas recomendables dependen de la magnitud y temporalidad del proyecto. Así, si el objetivo es diseñar un sistema permanente de información, con cobertura nacional, se sugiere utilizar estaciones de trabajo con plataforma UNIX y la infraestructura necesaria para la edición e impresión de mapas. A su vez, si el propósito es generar herramientas para la elaboración de diagnósticos y trabajos de investigación que apoyen la toma de decisiones a nivel local, entonces existe la posibilidad de operar con computadoras personales con una configuración mínima que garantice capacidad de almacenamiento y velocidad de proceso (Pentium 150 mhz., 32 mb de memoria ram, lectora de CD-Rom, monitor SVGA con 2 mb de memoria y tarjeta de gráficos para resolución de 1024 columnas por 768 renglones).

La elección de la mesa digitalizadora es independiente del tipo de plataforma que se disponga. Obviamente se debe velar porque tenga un tamaño adecuado. Existen en el mercado varias opciones (CALCOMP, ALTEK, GTCO, KURTA) y el proveedor deberá entregar los programas de enlace entre la tableta digitalizadora y el sistema de cómputo que se esté utilizando.

Finalmente, los dispositivos para la salida de los mapas pueden ser “plotters” de plumas para hojas de papel de 36 pulgadas e impresoras a color de inyección de tinta.

3. Selección de las muestras

Uno de los objetivos fundamentales del diseño muestral es definir el procedimiento que se aplicará para seleccionar las unidades de observación que formarán parte de la muestra. En términos generales, lo más apropiado sería seleccionar directamente las unidades de observación a partir de un listado que considere a todas y cada una de las unidades objeto de estudio. Sin embargo, proceder de esta manera incrementa apreciablemente el costo del trabajo de campo ya que, en caso que existiera un listado que incorporara todos los elementos del universo de estudio, la dispersión de la muestra implicaría grandes desplazamientos de los entrevistadores para visitar las distintas unidades seleccionadas, lo que incrementaría los costos a un nivel que haría casi inviable cualquier proyecto de encuesta.

Debido a esos altos costos asociados a la selección de una muestra aleatoria (en diseños monoetápicos con selección por muestreo aleatorio simple), es una práctica habitual el que los países utilicen diseños polietápicos; vale decir, métodos de selección en que las unidades de observación son elegidas luego de diversas fases que combinan procedimientos de conglomeración y estratificación. A su vez, en estos casos es necesario construir un marco para cada una de las etapas definidas para la selección de la muestra.

De esta manera, se busca formar unidades de marco de tamaños similares, a fin de reducir la variabilidad de las probabilidades de selección. En el caso en que las unidades formadas no difieren considerablemente en relación al número de unidades que las forman, es posible aplicar un procedimiento de selección que asigne igual probabilidad a todas las unidades. Sin embargo, es frecuente que existan diferencias considerables en relación al tamaño, por lo que en estas situaciones se sugiere que la asignación de las unidades en muestra se efectúe asignándoles una probabilidad proporcional a su tamaño, expresado por ejemplo en el total de viviendas que incluyen o de alguna otra variable que esté altamente correlacionada con las características de la población que se desea estimar.

Es necesario controlar cuidadosamente la aplicación del procedimiento de selección expresado en el esquema de muestreo adoptado, y llevar un registro detallado de las unidades incluidas en la muestra y de sus probabilidades de selección, a efectos de que sea posible construir en forma clara el factor de expansión que permitirá extrapolar los resultados observados en la muestra al total del universo de estudio.

Previo a la selección de la unidades de vivienda en los listados correspondientes, se debiera depurar la lista disponible a partir de un recorrido de campo que permita conocer la situación de las viviendas que formarán parte de la investigación. Así, se debiera actualizar el listado identificando bajas y altas de unidades de vivienda, poniendo especial énfasis en las de nueva creación y en aquellas que se dedican a una actividad económica pero que a la vez operan como lugar de residencia de las familias.

En la fórmula aplicada para el cálculo del factor de expansión, se debieran precisar las diferencias entre los pesos asignados en gabinete a las unidades seleccionadas y aquellos determinados como fruto de los recorridos de actualización en campo. Es habitual que existan

diferencias entre ambas mediciones, por lo que en el hecho resulta muy difícil preservar la característica de autoponderación en el diseño de una encuesta¹².

Cuando no se sigue este procedimiento, suele adoptarse el supuesto de que no han ocurrido cambios significativos entre la fecha de elaboración del marco de muestro y la de realización del trabajo de campo, lo cual obviamente es incorrecto. A fin de intentar superar esta situación, la práctica cotidiana supone que el problema se resuelve, al menos en parte, expandiendo los resultados de la encuesta a una proyección de población que tome en cuenta la dinámica demográfica del país, pero que pocas veces considera la distribución espacial de la población. Ello permite reducir los errores en los estimadores nacionales, lo que no impide que en muchas ocasiones se agudicen las diferencias regionales, generando tasas de crecimiento inconsistentes entre dos o más encuestas de un mismo programa¹³.

Otro elemento que también es importante señalar en este apartado, es el que se refiere a la práctica, muy difundida en algunas oficinas de estadística de la región, de efectuar reemplazos de viviendas en los casos de pérdida de muestra por motivos de no respuesta. De hecho, es común que las altas tasas de respuesta que reportan algunas investigaciones por muestreo escondan prácticas deliberadas de reemplazos efectuados a criterio del entrevistador y sin ningún control por parte de los encargados del diseño estadístico de la encuesta. Como se sabe, estas sustituciones alteran las probabilidades de selección de las viviendas e introducen sesgos en la información que son difíciles de evaluar. De modo que para lograr la precisión deseada en las estimaciones, se sugiere no fomentar estas iniciativas, sino más bien, alternativamente, ajustar el tamaño de muestra por motivos de no respuesta, con el objeto de obtener durante la ejecución del trabajo de campo el número mínimo de selecciones que garantice la confianza requerida.

Aún en la situación en que los reemplazos se efectúan previo a la ejecución del trabajo de campo, se deben extremar precauciones. Es bien sabido que en ocasiones se hacen selecciones adicionales, con objeto de prever las pérdidas de observaciones en terreno; sin embargo, cuando se emplea esta práctica, se debe considerar que las probabilidades de selección de las viviendas que funjirán como reemplazos son diferentes al resto, generando con ello sesgos de estimación, además de no reflejar con fidelidad el comportamiento del informante ante la encuesta, así como los errores inherentes a la posible falta de actualización del marco de muestreo.

¹² Hay personas que creen que la autoponderación es una característica deseable en los diseños de muestra y que incluso esta cualidad facilita el manejo de la base de datos y la expansión de los resultados. Sin embargo, debido a los cambios físicos ocurridos en el terreno, esta condición difícilmente se puede preservar y la práctica común indica que la mayor parte de las muestras no son autoponderadas. Además, el uso intensivo de microcomputadoras facilita enormemente el manejo de la información, por lo que la existencia de múltiples factores de expansión no representa ningún problema.

¹³ La utilización de proyecciones de población para la expansión de los resultados es común en muchas de las encuestas de hogares que se efectúan en los países de la región. Sin embargo, a pesar de ser la solución más barata e idónea ante la ausencia de un marco de muestreo previamente actualizado, a la larga puede generar graves problemas a los usuarios de la información y a las propias oficinas de estadística. Se debe considerar el hecho de que las proyecciones de población en el mejor de los casos se efectúan para el contexto nacional, estatal, provincial o departamental, pero en muy pocas ocasiones para divisiones geográficas menores. Esto significa que a partir de ellas no es posible conocer los cambios de población que se registran en el ámbito de áreas de muestreo, lo cual puede traducirse en inconsistencias en las estimaciones generadas.

III. Control de calidad en el diseño y ejecución de encuestas de hogares

1. Introducción

El diseño y ejecución de encuestas de hogares por muestreo exige un sólido sustento técnico, así como la aplicación de sistemas modernos de gestión y administración que aseguren la producción de información estadística de calidad. El cumplimiento de los objetivos de la investigación se traduce en el logro de una plena satisfacción de los requerimientos de los usuarios, donde se minimicen los errores y sesgos de la información y se alcance eficiencia en el uso de los recursos disponibles mediante la generación de resultados confiables, oportunos y a bajo costo.

En ese sentido, todas las etapas involucradas en el diseño y ejecución de una encuesta son esenciales para garantizar la calidad de la información. De modo que en cada una de ellas se debieran contemplar y poner en práctica mecanismos de control que permitan identificar y corregir las distorsiones y posibles errores en los datos recabados.

Bajo esta perspectiva, de orden más bien sistémico, a continuación se analizan sucintamente las diversas etapas que caracterizan la ejecución de una encuesta de hogares, a la vez que se sugieren algunos criterios, procedimientos y medidas de control que se consideran de utilidad para reducir los errores y sesgos de los datos y, en consecuencia, elevar la calidad de la información.

2. Ambito metodológico

2.1 Diseño conceptual

Bajo la óptica de la calidad total, los objetivos de las encuestas se deben orientar preferentemente a satisfacer, en tiempo y forma, las necesidades de los usuarios de la información: analistas, diseñadores de políticas y tomadores de decisiones. De modo que en la etapa de definición de los métodos y procedimientos para llevar a cabo la investigación, se debe dar respuesta a preguntas tales como qué tipo de información se desea obtener; de qué manera se alcanzarán los objetivos de la encuesta y cumplirán los plazos establecidos para la entrega de resultados; cómo se relacionarán los datos generados con aquellos disponibles de otras fuentes y cuál es la precisión deseada para las estimaciones de interés.

a) Requerimientos de información

Un primer requisito indispensable es definir con claridad el fenómeno a estudiar, con el fin de identificar los objetivos generales y específicos de la encuesta. Son pocas las ocasiones en que no existe ninguna información acerca del tema que se desea analizar, por lo que es aconsejable hacer primeramente una revisión bibliográfica y de los antecedentes disponibles de estudios similares realizados en el pasado, en el país y en el extranjero, así como efectuar consultas con especialistas y usuarios de la información a efectos de precisar los alcances de la investigación y los indicadores que se desean obtener.

Asimismo, es necesario tener claro los usos que se harán de la información recopilada, lo que permitirá identificar los dominios de estudio, establecer las hipótesis de comportamiento, formular posibles relaciones de causalidad entre las variables estudiadas y definir los niveles de desagregación y precisión requeridos para los datos. Esta última información constituye, a su vez, un insumo fundamental para determinar el tamaño de la muestra y definir los procedimientos para la selección de las observaciones que formarán parte de la investigación.

b) Marco conceptual

Una vez identificados los requerimientos de información, la elaboración del marco conceptual y metodológico debiera orientarse a cumplir con los objetivos de la encuesta, considerando las recomendaciones internacionales vigentes. A su vez, si entre las distintas encuestas que forman parte del sistema integrado existiese traslape de algunos temas, las definiciones y conceptos debieran homologarse a fin de evitar inconsistencias en los resultados proporcionados por las diferentes fuentes de datos.

Es habitual que algunas encuestas investiguen determinados temas con mayor o menor profundidad que otras, situación que también puede redundar en que los resultados obtenidos no sean estrictamente comparables. Las encuestas de empleo, por ejemplo, normalmente profundizan en la identificación de las personas que manifiestan estar desocupadas (cesantes o buscando empleo por primera vez), o bien que están afectadas por distintas formas de subempleo. Para ello, y en atención a las recomendaciones internacionales sobre la materia, se formulan varias preguntas, algunas de las cuales actúan como filtro, con el objeto de distinguir a estas subpoblaciones de interés, pero que a la vez determinan la medición de la fuerza de trabajo. Asimismo, existen también algunas encuestas orientadas a evaluar los niveles de vida de las familias que estudian el empleo como una variable importante para explicar la formación del ingreso de los hogares; sin embargo, es frecuente que las preguntas que se incluyen para captar la condición de ocupación de la población en edad de trabajar adolezcan de cierta superficialidad, lo que genera discrepancias en los resultados respecto a otras estadísticas oficiales sobre empleo y desempleo.

Por su parte, la sistematización del marco conceptual y metodológico de la encuesta debe traducirse en la elaboración de un conjunto de manuales que sirvan para apoyar las actividades de capacitación de los supervisores, entrevistadores, críticos y codificadores, y del personal de todo nivel involucrado en la ejecución del proyecto. En estos manuales se debieran incorporar los conceptos y definiciones de las preguntas contenidas en el cuestionario, ejemplificando con gran detalle las variantes que pueden enfrentarse durante la realización del trabajo de campo, especificando lo que se considera y lo que se excluye en cada caso, y la manera de actuar y decidir ante distintas situaciones prácticas. Asimismo, dichos manuales debieran adoptarse como documentos de consulta que ayuden a aclarar dudas y resolver situaciones complejas que puedan surgir durante las visitas a los hogares¹⁴.

c) Cuestionario y plan de tabulados

El logro de los objetivos de una encuesta de propósitos múltiples se basa en gran medida en la disposición de los hogares a colaborar, así como en la calidad de los datos que proporcionan los informantes. De manera que la etapa del diseño del cuestionario se puede considerar como una de las más críticas en la planeación de una encuesta por muestreo, ya que su aceptación por parte de los informantes está condicionada a disponer de un instrumento de fácil entendimiento y estructurado en términos lógicos. Un cuestionario bien redactado y ordenado atrae la atención del informante y facilita su relación con el entrevistador.

El cuestionario es el instrumento que permite conducir la entrevista y evita que el empadronador elija el orden y redacción de las preguntas, o las categorías de respuesta. A su vez, establece el flujo adecuado de las preguntas e identifica a las subpoblaciones de interés en cada tema. Por otra parte, para facilitar su manejo en campo el cuestionario debe tener un formato adecuado y ser resistente a las inclemencias climáticas y al constante manoseo de que son objeto.

Dado que el cuestionario representa la expresión operativa de los objetivos de la investigación, su diseño debe guardar estrecha relación con los demás componentes de la encuesta. Por ende, es muy recomendable que su confección esté a cargo de un equipo interdisciplinario integrado por especialistas en los temas de la encuesta, diseño estadístico, enfoque conceptual, personas con experiencia en el trabajo de campo y los encargados del procesamiento de datos, entre otros. En la actualidad es común encontrar que tiendan a repetirse los estudios sobre diversos temas en un mismo país e incluso entre países, por lo que la experiencia en el diseño de cuestionarios debe entenderse como un proceso acumulativo. Sin embargo, esto no significa que un formato elaborado con objetivos específicos pueda aplicarse en cualquier lugar o circunstancia sin que se efectúen las modificaciones pertinentes en atención a las características del entorno cultural, el nivel de instrucción de la población objetivo, la calificación del personal de campo o el grado de desarrollo del sistema estadístico nacional.

Como se ha señalado en diversos documentos técnicos, es importante tener consciencia que las personas que redactan el cuestionario no son las mismas que van a responder la entrevista, por lo que se debe pensar siempre en los informantes al momento de formular las preguntas. En este sentido, deben omitirse al máximo los tecnicismos y las palabras de uso poco frecuente ya que su inclusión seguramente contribuirá a generar confusión y a desestimular la participación de los entrevistados.

El buen conocimiento del tema a investigar, así como las consultas a los usuarios, facilitan una apropiada redacción de las preguntas. Asimismo, las pruebas de campo en zonas geográficas diversas que presenten situaciones poco comunes y para diferentes estratos socioeconómicos,

¹⁴ Apoyándose en los métodos modernos de comunicación, es aconsejable que esta información se encuentre en línea en una página WEB, con el fin de que todo el personal de las oficinas regionales lo pueda consultar en cualquier momento.

permiten afinar la redacción y ayudan a percibir la aceptación de los informantes y su disposición a colaborar. No siempre una redacción detallada genera mejor información que las preguntas breves y directas. Sin embargo, por lo general un desglose amplio de las opciones de respuesta constituye un mecanismo adecuado para ganar en precisión y reducir las omisiones, al tiempo que en muchas variables económicas permite la comparabilidad con datos a nivel macro y con registros administrativos.

El diseño físico del formulario puede basarse en un formato que registre en una primera parte información de carácter general, como el nombre de la encuesta, el organismo que la ejecuta, las disposiciones legales que amparan su aplicación, el número de folio e información de control que permita determinar con facilidad la ubicación geográfica de la unidad de observación, el tiempo de duración de la visita al hogar y el resultado final de la entrevista. A continuación, es aconsejable incorporar preguntas que informen sobre las características básicas de las unidades de observación. Así, en el caso de las encuestas de hogares se puede registrar, entre otras, las características de la vivienda y el número de hogares que la componen, sus materiales de construcción, la disponibilidad de servicios básicos y el equipamiento doméstico.

La segunda parte del formato se puede destinar a consignar información general sobre los miembros del hogar, que permita una primera clasificación de la población objeto de estudio para iniciar el registro de datos sobre los temas centrales de la investigación. También, es común que en esta sección se indague acerca del uso del tiempo de los miembros del hogar

Finalmente, las preguntas relacionadas con el o los temas específicos de la encuesta deberán estructurarse conforme a una lógica simple, donde los aspectos de carácter general antecedan a las preguntas más puntuales sobre los tópicos de interés. Los primeros cuestionamientos deben ser breves y fáciles de responder, a fin de que el entrevistado adquiera confianza e incremente su interés por participar. Asimismo, las preguntas comprometedoras o íntimas se debieran ubicar de la mitad del cuestionario hacia adelante, para aplicarlas una vez que el entrevistador haya establecido una relación de confianza con el informante. En todos los casos el cuestionario debiera disponer de suficiente espacio en blanco, para que el entrevistador pueda hacer las aclaraciones pertinentes que ayuden a los responsables de la crítica y codificación a entender el porqué se registró esa información.

En relación al plan de tabulados, éste suele verse simplemente como la acción que permite decidir a los responsables de la difusión sobre el esquema básico de publicación de resultados. Sin embargo, hay que enfatizar que esta fase constituye una oportunidad para identificar, junto con los usuarios, los dominios analíticos en los que se desea desagregar la información, así como para conocer cuáles serán los indicadores que se desean calcular y las respectivas subpoblaciones de interés. Asimismo, el análisis preliminar de los cruces que se quieren realizar contribuye a estimar el número mínimo de observaciones necesarias para garantizar la confiabilidad estadística de las relaciones que se establezcan entre categorías de las variables estudiadas, lo cual ayudará a la determinación del tamaño de muestra necesario.

Por otra parte, el plan básico de tabulados debe ser interpretado como un primer intento por sintetizar los resultados que se generarán, y en ningún caso debiera considerarse como un procedimiento orientado a la explotación total de la información. Por mas tabulados y gráficas que se incluyan en una publicación, nunca se podrá llegar a satisfacer todas las demandas de los usuarios o lograr una completa y exhaustiva explotación de la base de datos.

El verdadero valor de los resultados generados por medio de una encuesta por muestreo está determinado, en última instancia, por el uso analítico que se haga de la información, con el fin de establecer un sistema de datos que describa y/o contribuya a explicar la ocurrencia de un fenómeno

de interés, a evaluar el impacto de las políticas públicas o a determinar cuáles son las causas asociadas a la evolución de una variable determinada.

Una práctica que se viene desarrollando frecuentemente, es poner a disposición de los usuarios discos magnéticos o compactos con los tabulados de la encuesta. Esto constituye una interesante y eficaz forma de incorporar el desarrollo tecnológico a los procedimientos de difusión de información estadística, que facilita al usuario la consulta y almacenamiento de los datos. No obstante, en ningún caso releva de la necesidad de agregar mayor valor a la información de la encuesta, mediante los estudios analíticos que normalmente se elaboran en los centros de investigación y de diseño de políticas, tanto públicos como privados.

En ese sentido, cabe enfatizar que los usuarios especializados abocados a evaluar políticas o a determinar relaciones de causalidad, requieren disponer esencialmente de las bases de microdatos, con el fin de ampliar y profundizar sus posibilidades de análisis. De modo que las unidades ejecutoras debieran asumir que las actividades de la encuesta continúan más allá de la generación de tabulados y la publicación de resultados, y deben estar dispuestas a proporcionar a los usuarios los archivos con los registros individuales, así como asesoría estadística para garantizar el buen uso de los datos y preservar la confidencialidad de la información.

d) Estructura funcional

Se requiere establecer con claridad los diferentes niveles de mando que intervienen en el desarrollo de la encuesta, así como las funciones y responsabilidades asociadas a cada uno de ellos. Para esto, es recomendable elaborar un organigrama que represente la estructura jerárquica y las interrelaciones de las distintas áreas que intervienen en el diseño y ejecución de la encuesta. Asimismo, debiera elaborarse un manual de operación que defina los procedimientos administrativos que corresponde cumplir para la solicitud y comprobación de recursos, el envío y recepción de materiales, el uso de vehículos y combustible, los diversos gastos que se requieren para la elaboración y actualización del marco de muestreo, la impresión de materiales, la ejecución del trabajo de campo, etc.

De paso, esta información resulta esencial para futuras investigaciones, en la medida que los diseños óptimos de encuestas requieren disponer de datos sobre los costos de ejecución en las diferentes fases, para determinar con exactitud el número óptimo de unidades de selección de primera y segunda etapa y, en otro orden, permite también elaborar controles para la distribución de los materiales de trabajo (útiles de oficina, cuestionarios, manuales, etc.) y diseñar diversos formatos para el manejo de las actividades de la encuesta¹⁵.

2.2 Diseño estadístico

El diseño estadístico de una encuesta comprende al menos el desarrollo de las siguientes actividades: i) Determinación del tamaño de muestra; ii) Elaboración del marco de muestreo; iii) Selección de las unidades que formarán parte de la muestra; y iv) Definición de los procedimientos de estimación y cálculo de los errores de muestreo. Cada una de estas actividades debe ser planeada y ejecutada con rigor y eficiencia, con el fin de contribuir a minimizar los errores de muestreo y los sesgos de estimación.

¹⁵ Tal vez una falencia importante de las oficinas nacionales de estadística de los países de la región, es la escasa información que registran con relación a las etapas de planeación y operación en campo del proyecto. Son pocas las ocasiones en que se informa acerca de los costos unitarios asociados a la encuesta, se hace una evaluación pormenorizada de la calidad del marco de muestreo y de las causas más frecuentes de rechazo o se consignan las mejores prácticas observadas durante la entrevista, la supervisión, crítica o registro de datos.

a) Tamaño de la muestra

Uno de los primeros requerimientos del diseño estadístico de una encuesta es identificar, dentro del conjunto de variables para las que se propone generar información, cuál(es) de ella(s) se utilizará(n) para determinar el tamaño de la muestra. Al mismo tiempo, se debe decidir si se obtendrán estimaciones para proporciones, promedios o totales, así como la precisión y confianza deseada para los resultados esperados.

Conocer la varianza de las variables de diseño es fundamental para calcular el número óptimo de selecciones que permitan lograr estimaciones con la precisión establecida. No obstante, es común que en muchas encuestas (sobre todo las que estudian el mercado laboral y la fuerza de trabajo y también las que estudian las condiciones de vida) el tamaño de muestra se calcule como si el objetivo de la encuesta se centrara exclusivamente en estimar proporciones, aunque posteriormente la información sea utilizada para calcular promedios y totales. En este sentido, se debe enfatizar que esta práctica tiene consecuencias en el error de muestreo, ya que éste se incrementa necesariamente por el simple hecho que la muestra calculada no resulta suficiente para garantizar la precisión esperada en estimaciones de promedios y totales. A su vez, también es habitual que no se disponga de estimaciones acerca del error de muestreo de la variable de interés para los diseños estratificados y de conglomerados. En este caso, comúnmente el tamaño de muestra se determina como si se fuera a efectuar una selección con igual probabilidad, y posteriormente el valor obtenido se incrementa por un factor de ajuste denominado efecto de diseño (efd), así como por las posibles pérdidas de observaciones que se puedan originar durante la ejecución del trabajo de campo, o tasa de no respuesta esperada (tnr).

Esta fase suele desarrollarse en forma rutinaria y la gran mayoría de las veces no se le presta la debida atención. Como se sabe, un menor número de observaciones que las necesarias, incrementa los errores de muestreo e introduce sesgos importantes en los datos.

Cuando no se dispone de valores actualizados sobre el efecto de conglomeración del marco (efd) y se utiliza alguno obtenido de una investigación "similar", es muy probable que se subestime el tamaño de muestra requerido para obtener resultados con la precisión deseada. Asimismo, si se sobreestima el valor de este factor de ajuste, el número de selecciones se incrementa innecesariamente lo cual redundará en un desperdicio de recursos. Es importante recordar que cada encuesta tiene características propias y que éstas varían conforme a las circunstancias de la investigación, la participación de los informantes, la calidad de los encuestadores, etc. En tal sentido, se señala que deben actualizarse los coeficientes de variación y efectos de diseño de las variables, ya que el no hacerlo puede significar pérdida de precisión por malos cálculos en el tamaño de muestra¹⁶.

Por último, para prever la pérdida de observaciones en campo, por motivos de rechazo de los informantes y defectos del marco, y preservar el grado de precisión, es recomendable introducir ajustes en el tamaño de la muestra. De modo que se considera una mala práctica incorporar reemplazos de manera deliberada, ya que además de alterar las probabilidades de selección de las unidades de observación, induce a un mal uso por parte de los entrevistadores e introduce sesgos de estimación difíciles de evaluar.

b) Marco de muestreo

Para seleccionar las unidades que formarán parte de la encuesta, se requiere disponer de una relación que identifique de manera inequívoca a cada una de ellas. Sin embargo, por lo general no se dispone de un listado que permita la ubicación y selección directa de las unidades de

¹⁶ De hecho, es preciso anotar que cada variable tiene asociado un efecto de diseño, por lo que el utilizar un solo valor en la determinación del tamaño de muestra puede conducir a errores de cálculo.

observación, lo que hace necesario elaborar diseños polietápicos que suponen la confección de marcos de muestreo para cada fase de selección.

A su vez, para minimizar los errores de selección e identificación de las unidades se requiere que el marco de muestreo esté actualizado y no adolezca de duplicaciones u omisiones. En tal sentido, y a pesar que el marco puede haberse elaborado con información reciente, es necesario efectuar conteos que permitan actualizar el número de viviendas existentes en las unidades seleccionadas, y luego listarlas, a fin de identificar aquellas que son susceptibles de ser entrevistadas. En el caso en que se trate de una encuesta de hogares, se eliminan de este proceso los establecimientos económicos y las viviendas de uso colectivo, por no formar parte de la población objeto de estudio.

El listado de viviendas debiera contar con un croquis que permita ubicar físicamente la manzana o área geográfica seleccionada, así como el segmento donde se encuentran las viviendas que se desea visitar. A fin de que los entrevistadores puedan identificarlas con precisión, se deben describir los rasgos físicos perdurables de la zona y de la vivienda, de tal suerte que sean fácilmente identificables en el terreno.

Asimismo, para no incorporar distorsiones en la selección de la muestra, durante la etapa de elaboración del marco se deben asignar a las unidades definidas medidas de tamaño (pesos) actualizadas, con el objeto de diferenciar su importancia relativa tanto en el proceso de selección como en la fase de generación de resultados.

En todos los casos es deseable que los procedimientos utilizados tengan la capacidad de ser replicables, en el sentido que si otra persona siguiera las instrucciones definidas necesariamente debiera aproximarse al mismo resultado.

A fin de apoyar la consolidación y resguardo de la información, es importante recurrir a métodos de automatización que permitan clasificar e indexar la base de datos en función a los atributos de interés definidos por los usuarios, a fin de facilitar los procedimientos de estratificación y conglomeración en el caso de las encuestas polietápicas.

c) Selección de la muestra

La etapa de selección de la muestra es fundamental para elegir de manera inequívoca a las unidades que formarán parte de la investigación. Así, a partir del esquema de selección definido se deberán seleccionar las unidades en las diferentes etapas, identificando el estrato de procedencia, así como el área geográfica a la que pertenece el conglomerado seleccionado. En todos los casos, se deberá registrar el "peso" de la unidad seleccionada con el fin de poder determinar la probabilidad de selección de las viviendas y los factores de expansión asociados.

La incorporación de los Sistemas de Información Geográfica (SIGs) a la ejecución de proyectos de generación de información estadística es una realidad en los países de la región. En este sentido, es cada vez más común que los marcos de muestreo estén resguardados en medios magnéticos y los procedimientos de conglomeración y selección de la muestra se hagan en forma automatizada. Sin embargo, se deberá tener cuidado en la ejecución mecánica de los procedimientos ya que para efectos del diseño y la estrategia de control y supervisión del trabajo de campo, no siempre las soluciones matemáticas obtenidas mediante algún algoritmo de optimización resultan ser las más eficientes desde el punto de vista práctico. Asimismo, se debe recordar que para efectos de muestreo se debe dar prioridad a la identificación física de la vivienda y no a la precisión geográfica de su ubicación, por lo que la falta de referencia métrica de la cartografía no debe ser impedimento para automatizar estas fases del proceso y disponer de planos de línea que se puedan actualizar e imprimir en forma automática.

d) Procedimiento de estimación y cálculo de los errores de muestreo

La manera de evaluar si una encuesta cumplió con las especificaciones previstas en el diseño, en relación a la confiabilidad de los estimadores, es mediante el cálculo de los errores de muestreo para las variables de interés. De modo que una vez definidas las expresiones que se utilizarán para construir los estimadores, se debe especificar claramente en el diseño de la muestra el procedimiento que se aplicará para el cálculo de los errores de muestreo.

Pese a que en los documentos elaborados por los países muchas veces se indica que para la estimación de la varianza se pueden utilizar algunos de los distintos métodos conocidos (Muestras Replicadas, Replicaciones Repetidas y Balanceadas, Replicaciones Repetidas por el método Jackknife y el método de Series de Taylor), en la mayoría de los casos la información contenida en la ficha técnica de la encuesta no permite evaluar la factibilidad de aplicar estas alternativas. Por ejemplo, el método de muestreo replicado requiere que se hayan realizado al menos dos selecciones (un número par) al interior de cada unidad primaria de muestreo (UPM). Así, si esta información no se consigna en el reporte de la encuesta y la base de datos no permite identificar las selecciones por UPM, el analista no estará en condiciones de calcular el error de muestreo.

Para efectuar las estimaciones es necesario asociar a cada una de las observaciones el inverso de su probabilidad de selección, llamado también factor de expansión. En esta tarea se deben tener en cuenta las diferentes etapas de selección de la muestra, así como los pesos relativos de las unidades de selección. Asimismo, en el cálculo del factor de expansión se requiere considerar los ajustes necesarios por no respuesta y por la post-estratificación de los datos, toda vez que los valores muestrales de las estructuras por edad y sexo se contrasten con los resultados obtenidos a partir de un censo reciente o de una proyección de población confiable.

3. Levantamiento de la información

3.1 Organización del trabajo de campo

Una de las etapas más importante en el desarrollo de una encuesta es la ejecución del trabajo de campo. Es posible elaborar un buen cuestionario, o efectuar un óptimo diseño de muestra e impartir una excelente capacitación a los encuestadores; sin embargo, la falta de una adecuada organización y supervisión del trabajo de campo puede llegar a invalidar los resultados de la investigación.

Por ende, es imprescindible que la unidad ejecutora confeccione un manual en el que se especifiquen detalladamente los aspectos organizativos del levantamiento y las facultades, atribuciones y responsabilidades del personal, así como las relaciones entre las diferentes instancias que participan en la planeación y ejecución del trabajo de campo.

A su vez, se deben definir los controles necesarios para garantizar que se minimicen los errores de muestreo y ajenos al muestreo, así como los sesgos involuntarios que se pueden introducir producto de decisiones inadecuadas adoptadas por supervisores y/o entrevistadores. Al respecto, es recomendable que el equipo de campo evite adoptar iniciativas sustantivas sin el previo consentimiento de la coordinación central del proyecto, y al mismo tiempo asegure que los criterios aplicados para la clasificación de las respuestas estén de acuerdo con los lineamientos entregados en la etapa de capacitación y consignados en el manual correspondiente.

Por otra parte, es conveniente elaborar bitácoras de problemas surgidos en las diferentes zonas donde se esté llevando a cabo la encuesta y de sus respectivas soluciones, a fin de dar a conocer al personal –especialmente al que está en campo– las decisiones adoptadas ante cada situación. Asimismo, y valiéndose de las facilidades que ofrecen los medios de comunicación

modernos, como el correo electrónico, las oficinas regionales debieran estar permanentemente informadas sobre los criterios asumidos por la coordinación del proyecto, respecto a las dudas surgidas durante el trabajo de campo en las distintas zonas geográficas en donde se realiza la investigación.

3.2 Capacitación del personal de campo y de oficina

La puesta en marcha de un adecuado programa de capacitación del personal involucrado en el levantamiento de la información es otra tarea fundamental. Este programa debiera abarcar todos los niveles de la unidad ejecutora, a la vez que es necesario que contenga prácticas en campo que permitan conocer el desempeño y rendimiento de la estructura operativa que participará en la fase de ejecución: supervisores, entrevistadores, codificadores, choferes, etc. Es aconsejable que tales actividades comprendan la capacitación en sala, con ejercicios prácticos, a partir de juego de roles, en donde se ponga a prueba la claridad conceptual y de redacción del cuestionario, así como la capacidad de los manuales para responder a dudas concretas planteadas en situaciones reales. Asimismo, se sugiere realizar evaluaciones permanentes que fomenten el intercambio de puntos de vista entre los responsables de la elaboración del marco conceptual y metodológico, por una parte, y los entrevistadores y el personal encargado de hacer la supervisión y la crítica-codificación de la información, por la otra. Al mismo tiempo, se debieran planear sesiones de prácticas de campo en distintos ámbitos geográficos, seguidas del correspondiente análisis de evaluación que permita perfeccionar los procedimientos, así como precisar conceptos o incorporar eventuales correcciones que afecten incluso la redacción de algunas preguntas. Respecto a esto último, es necesario recalcar que la correcta comprensión por parte de los empadronadores, tanto de los conceptos a investigar como de las preguntas que serán formuladas a los entrevistados, son una condición esencial para disminuir los errores de no muestreo y los probables sesgos de la información.

3.3 Distribución de la muestra y de las cargas de trabajo

Existen otras actividades relativas al levantamiento de los datos a las que también se les debe prestar especial atención. Una de ellas se refiere a lograr una adecuada distribución de la muestra de conformidad a los plazos estimados para la realización de la encuesta. Esta debe ser congruente con el calendario en términos del tiempo asignado para el levantamiento y la entrega de resultados, buscando el equilibrio entre la disponibilidad de recursos y las necesidades del trabajo de campo. A su vez, se debiera apuntar a que los diferentes períodos proyectados para la recopilación de la información (semana, mes, trimestre, etc.) mantengan un balance adecuado, con el objeto de no saturar las actividades en alguna fecha determinada que pudiera incrementar desproporcionadamente la carga de trabajo de los encuestadores y afectar la confiabilidad de la información. Asimismo, es importante distribuir la muestra en función de las condiciones climáticas, privilegiando aquellas zonas o regiones de difícil acceso o en las que sólo se puede transitar en ciertas épocas del año, a fin de garantizar que las entrevistas se realicen en los plazos señalados.

Por otra parte, la asignación de las cargas de trabajo a los encuestadores y supervisores debiera estar acorde con la capacidad física del personal, las condiciones del terreno y el tiempo estimado para llevar a cabo la entrevista. En este sentido, el número de visitas diarias que razonablemente puede realizar un entrevistador dependerá de la extensión del cuestionario y la complejidad de las preguntas, la distancia entre las viviendas (tanto en las zonas urbanas como en las rurales), la duración promedio de la entrevista y el número de visitas necesarias para lograr su cooperación. Se debe velar por no sobrecargar al personal de campo, ya que ello podría llevar a que las entrevistas se realicen sin el debido cuidado o, peor aún, se estimule el autollenado de cuestionarios por parte de los enumeradores. En ambos casos, se incrementan los sesgos de

estimación, con el agravante que, como se ha dicho, es muy difícil evaluar su efecto en la calidad de los resultados¹⁷.

3.4 Supervisión del trabajo de campo

La supervisión del trabajo de campo es otra de las tareas fundamentales para garantizar buenos resultados. En primera instancia, se debiera conformar un equipo encargado de la supervisión de los entrevistadores sin que sus integrantes tengan la responsabilidad directa de realizar entrevistas. Ello implica destinar recursos para que este grupo de personas, con mayor nivel de capacitación y entrenamiento, evalúen la labor de los enumeradores y además dediquen tiempo para realizar reentrevistas que permitan corroborar la veracidad de la información proporcionada por los miembros del hogar. De este modo, la supervisión permite aclarar dudas que se presentan en los datos, verificar si el encuestador está cumpliendo con su trabajo y, en general, evaluar la calidad de la información recopilada.

Actualmente, las oficinas de estadística están preocupadas por reducir el tiempo que media entre la ejecución del trabajo de campo y la entrega de resultados, lo cual sin duda es muy importante para mejorar la oportunidad de la información de cara al diseño y evaluación de políticas y la toma de decisiones. En esta dirección se han realizado ejercicios que incorporan el uso de computadoras portátiles para ingresar en campo la información proporcionada por los hogares, desarrollando en forma paralela la fase de crítica y análisis de consistencia.

A primera vista esta situación podría considerarse ideal, ya que la información se ingresa y valida en el momento mismo en que se captura, a partir del diseño de sistemas en línea que permiten definir criterios de consistencia de los datos. Sin embargo, se debe tener presente que es muy difícil que los diseños de gabinete prevean todas las contingencias que surgen en la realidad, y por lo tanto cubran suficientemente los diversos aspectos relativos a la congruencia de la información. De allí que, especialmente en encuestas complejas, la experiencia hace aconsejable mantener y reforzar la etapa de crítica y codificación, como una tarea fundamental para detectar errores e inconsistencias no identificados en la fase de supervisión, lo que contribuye a mejorar la calidad de la información.

Por otra parte, en aquellas encuestas que contienen preguntas cuyas respuestas deben codificarse después de realizar el análisis de un conjunto de variables y la consulta de nomencladores específicos, es difícil realizar esta actividad en forma automática. Un ejemplo de ello es la codificación del tipo de ocupación o la rama de actividad económica en que se desempeña la fuerza de trabajo, que en la medida que se registre con un amplio nivel de desagregación requiere de gran experiencia por parte de los codificadores para lograr una adecuada clasificación, así como el uso de catálogos especializados, no siempre de fácil interpretación y entendimiento.

La captura y consistencia de la información en campo no implica que la congruencia interna de los datos sea suficiente para minimizar los errores de muestreo y garantizar la validez de los resultados. Como se sabe, el error cuadrático medio de los estimadores ($ECM = V(x) + \text{Sesgo}^2$) se compone del error de muestreo -varianza del estimador- más el sesgo al cuadrado¹⁸, acotando que este último componente no puede evaluarse estadísticamente. A su vez, se debe recordar que los errores ajenos al muestreo generan los mayores sesgos y distorsiones respecto de la realidad, y es difícil -por no decir imposible- que la incorporación de procedimientos automáticos puedan

¹⁷ Se debe insistir en la necesidad de efectuar periódicamente evaluaciones del trabajo de los encuestadores con el fin de conocer su rendimiento y confiabilidad. Por ejemplo, es aconsejable definir indicadores de eficiencia en el llenado de las secciones del cuestionario, en los tiempos definidos para tal fin, así como mediciones de su eficacia a partir de los reportes elaborados por los supervisores en las viviendas que se decida volver a visitar como parte del programa regular de supervisión en terreno.

¹⁸ Recordemos que el sesgo de un estimador se define como la diferencia entre el valor generado del estimador y su valor esperado ($\theta - E(\theta)$).

disminuirlos de manera significativa. En caso que existan errores en los datos, tales procedimientos automatizados tenderán más bien a estandarizarlos pero no a eliminarlos. De modo que el propósito de reducir el tiempo empleado en el procesamiento de la información no debe llevar a descuidar el resto de la etapas. Así, no sirve de mucho obtener resultados rápidos si los tamaños de muestra utilizados no son apropiados para cumplir con los objetivos planteados y no se conocen los errores de muestreo asociados a las estimaciones, ni la consistencia de la información con otras fuentes de datos disponibles. Más aún, la pronta disponibilidad de resultados no elimina de por sí las malas prácticas en el uso de la información en el caso en que las investigaciones que se efectúen apliquen, por ejemplo, métodos de análisis que no consideren como es debido la estructura del diseño de la muestra.

En la práctica, la mayoría de las oficinas de estadística de la región asignan importantes recursos a las tareas de supervisión y crítica-codificación. Muchas de ellas incluso han diseñado sistemas de captura y consistencia de la información que les permiten depurar las bases de datos en tiempo real, a medida que los datos se ingresan, con lo cual han reducido de manera apreciable el tiempo que transcurre entre la realización del trabajo de campo y la entrega de resultados. Sin embargo, corresponde insistir en que debe prestarse una mayor atención a la congruencia interna de los datos, la consistencia de la información generada por la encuesta con aquella otra disponible en el sistema nacional de información, así como a su comparación con los resultados de encuestas similares realizadas en el pasado.

Esto no desestima, sino muy por el contrario refuerza, los intentos por entregar resultados oportunos pero a la vez con mayor confiabilidad. El éxito de una investigación por muestreo no se basa en disponer de computadores en el terreno, si acaso se olvida que los errores de los datos se generan de hecho en diferentes instancias, y que un proceso automatizado no es suficiente por sí sólo para eliminar los sesgos. De cualquier manera, se debe seguir propugnando que se reduzcan los plazos para la entrega de resultados, ya que en ocasiones la información es conocida varios años después de concluido el trabajo de campo, con lo que ésta se incorpora al acervo histórico del país pero sin representar una contribución efectiva al estudio de la coyuntura y, por ende, a la toma de decisiones.

Por último, es preciso señalar que la fase de supervisión y reentrevista a los hogares no debe entenderse como una tarea que signifique volver a estudiar toda la muestra o seguir a los entrevistadores por toda su área de trabajo. Para cubrir satisfactoriamente esta etapa, basta con el uso del muestreo de lotes, lo cual permite llevar un registro y evaluar el trabajo efectuado por los entrevistadores, generar una calificación de su desempeño en el terreno y controlar la calidad de la información. De hecho, esta práctica se sugiere para las encuestas de panel donde los hogares son visitados en más de una ocasión. Además, es recomendable rotar a los entrevistadores a fin de minimizar las posibilidades de repetición de información en forma deliberada o el autollenado de cuestionarios. Esta práctica es frecuente, sobretodo cuando los entrevistadores visitan las mismas viviendas durante diversas rondas de una misma encuesta (por ejemplo, en las encuestas continuas de hogares que investigan el tema del empleo).

4. Tratamiento de la información

4.1 Crítica y codificación

En esta etapa se definen los códigos válidos para cada categoría de respuesta y sus correspondientes rangos, las subpoblaciones de interés para cada tema y las relaciones permitidas entre las variables.

Las oficinas encargadas de administrar las encuestas debieran disponer de un grupo de técnicos con gran experiencia y profundo conocimiento de los temas investigados y de los alcances del cuestionario, encargados exclusivamente de evaluar la calidad de la información recolectada y la asignación de códigos a las preguntas abiertas que requieren clasificarse.

A pesar de que resulta aconsejable que la mayor parte de las preguntas estén precodificadas en el formato del cuestionario, existen ciertos temas en los cuales -como se indicó más arriba- no es conveniente que se efectúe una asignación automática ya que habitualmente se cometen errores de interpretación y clasificación. Por ejemplo, y como ya se mencionó, cuando se investiga el tema del empleo y se pretende codificar el tipo de ocupación (principal y secundaria) que realizan los miembros del hogar que manifestaron estar ocupados durante el período de referencia de la encuesta, es habitual que se presenten errores en la codificación efectuada en campo. De allí que para lograr información de calidad, se solicita a los entrevistadores que obtengan de los informantes ocupados una amplia descripción de sus actividades, a los efectos que en la oficina -personal con amplia experiencia en esta tarea- pueda asignar posteriormente un código apropiado a partir de las alternativas consignadas en la clasificación de ocupaciones elaborada por los organismos internacionales especializados en el tema del empleo. Asimismo, hay situaciones en las que tampoco se puede inferir en forma directa la codificación de la rama de actividad de la empresa, por lo que es necesario revisar catálogos específicos elaborados para este fin.

Se debe estar conscientes, sin embargo, que es normal que se presenten problemas de campo no previstos en los manuales y formatos de llenado. De hecho es prácticamente imposible prever todas las situaciones y reflejarlas en la modalidad de captura, crítica y validación de los datos en campo, sin quedar expuesto a que se generen una cantidad considerable de errores de clasificación y sin que los programas de verificación sean capaces de detectar tales inconsistencias. Del mismo modo, es preciso reiterar una vez más que los sesgos en los datos son difíciles de evaluar y en la mayor parte de los casos son más dañinos que los errores de muestreo.

Por otra parte, la etapa de crítica y codificación también permite evaluar el trabajo de los supervisores y representa otra instancia de control de la actuación de los encuestadores. Llevar registros apropiados y desarrollar actividades de actualización del personal son mecanismos que ayudan a reducir los errores ajenos al muestreo, sobre todo en las encuestas permanentes y/o continuas.

4.2 Falta de respuesta

La pérdida de información es una situación que comúnmente se presenta durante la ejecución del trabajo de campo. De allí que llevar un control adecuado de sus causas y su distribución entre grupos de hogares y regiones del país, sea visto como un elemento fundamental para reducir sesgos de estimación y preservar al máximo la calidad de la información. La no respuesta siempre va a existir, pero lo importante es tener una idea precisa de su magnitud y localización geográfica, a fin de realizar los ajustes pertinentes en el tamaño de muestra, los factores de expansión y reducir las causas que la originan.

La falta de actualización del marco de muestreo es una de las situaciones que provocan que no se identifiquen los segmentos seleccionados o se subregistre el número de viviendas existentes en alguna zona determinada. Esto puede preverse y controlarse, efectuando recorridos en las zonas seleccionadas y elaborando conteos y listados de viviendas en las fechas cercanas a la realización del trabajo de campo. En muchas ocasiones la desactualización del marco ocurre a gran velocidad, por lo que es recomendable que no exista mucha diferencia entre la fecha de actualización y la correspondiente a la realización del trabajo de campo.

A su vez, la no respuesta por rechazo del informante se puede minimizar mediante revisitas a los hogares por parte de los supervisores, o informando previamente sobre la realización de la encuesta por medio de cartas dirigidas a los jefes de los hogares donde se expongan sus objetivos y la fecha en que serán visitados por personal de la oficina de estadística.

A partir de experiencias previas, es posible estimar la tasa de respuesta esperada e incrementar en la fase de diseño estadístico el tamaño de muestra, a fin de lograr el número mínimo de entrevistas para obtener la precisión requerida en las estimaciones. Como se indicó, no se considera recomendable realizar sustituciones de las viviendas no entrevistadas por otras que hayan sido preseleccionadas y definidas como reemplazos. Este procedimiento introduce sesgos y altera las probabilidades de selección de las observaciones. En todo caso, como una actividad posterior a la conclusión del trabajo de campo y antes de ajustar los factores de expansión, se sugiere efectuar un análisis de la no respuesta total, por causas, por entrevistador, por grupos de hogares y por zonas geográficas, a fin de evaluar posibles errores de omisión o subespecificación. Mediante este procedimiento se pueden identificar posibles sesgos u omisiones, determinar la calidad del marco y evaluar la eficiencia de la labor realizada por los enumeradores. La concentración de la no respuesta en alguna zona o grupo social específico puede distorsionar la calidad de la información, y en ocasiones incluso invalidar los resultados de la investigación.

Cuando el número de entrevistas no logradas es poco significativo o se encuentra dentro del rango estimado en el cálculo del tamaño óptimo de muestra, la eliminación de éstas observaciones no afecta la precisión estadística de los datos. Sin embargo, será necesario ajustar los factores de expansión con el objeto que los resultados generados por medio de la muestra se aproximen a los valores verdaderos del universo de estudio (post-estratificación). Dependiendo de la magnitud y distribución de la no respuesta, los factores de expansión originales tenderán a modificarse, por lo que es casi un hecho que los diseños autoponderados perderán esa característica al finalizar la etapa del trabajo de campo. Así, un buen registro y control de la falta de respuesta es un requisito para garantizar la confiabilidad de los datos reportados.

Por otra parte, es también necesario prestar atención a la no respuesta parcial, la cual ocurre cuando algunas preguntas o secciones del cuestionario no son respondidas por el hogar o alguno de sus miembros. Cuando la frecuencia de ocurrencia es baja, las observaciones que presentan esta condición pueden ser eliminadas sin que ello afecte la precisión de los resultados. De lo contrario, es posible aplicar procedimientos automáticos para imputar valores faltantes, lo que sin embargo no necesariamente contribuye a mejorar la calidad de las estimaciones y puede llevar a incrementar el error del estimador¹⁹.

En efecto, esta práctica permite reducir el número de preguntas sin respuesta, pero la imputación no siempre es el método apropiado para reducir los errores de estimación. A pesar de que las observaciones ubicadas en una misma unidad muestral tienen una alta correlación en relación con la variable de estratificación, esto no necesariamente significa que ésta se preservará para cualquier característica de la población que se desee estimar.

4.3 Captura y congruencia de los datos

La fase de captura de datos representa el primer paso para la sistematización de la información. El desarrollo tecnológico y los programas disponibles permiten que esta etapa sea ahora más simple que las prácticas adoptadas en el pasado. Actualmente, la capacidad de las computadoras personales y las facilidades para compartir recursos que brindan las redes locales han desplazado a los antiguos "main frames", que requerían de amplios espacios, ambientación especial

¹⁹ La mayoría de los métodos estadísticos para la imputación de datos parten del supuesto de normalidad de la variable de interés, o de la premisa de que el esquema de muestreo que generó la muestra fue con la misma probabilidad para todas las observaciones, lo cual no necesariamente se cumple para el caso de una encuesta compleja.

y complicados lenguajes de programación. A su vez, los modernos sistemas de digitación permiten el diseño de pantallas a imagen del cuestionario, la definición de criterios de validación, la verificación de información durante la captura y la grabación en línea de los datos. Los tiempos de respuesta se han acortado y la mayor velocidad y capacidad de memoria permiten reducir los plazos para la depuración y procesamiento de la información. Por último, también se sugiere aquí aplicar métodos de muestreo por lotes para verificar la calidad de la información capturada, así como para evaluar el desempeño del personal encargado de la digitación.

4.4 Consistencia de la información

Esta etapa se realiza toda vez que se ha concluido con las tareas de digitación, y comprende al menos las dos etapas siguientes: i) análisis de consistencia interna y ii) validación de los resultados, con encuestas sobre el mismo tema realizadas en el pasado, con otras encuestas que generen datos sobre temas comunes, datos macroeconómicos, estudios internacionales y registros administrativos. En este contexto, esta actividad es fundamental ya que la calidad de la información no se puede calificar de manera exclusiva por la tasa de respuesta, por la baja incidencia de preguntas sin respuesta o la rapidez para la obtención de resultados.

Los datos generados se considerarán de buena calidad y útiles para la toma de decisiones en la medida que sus órdenes de magnitud guarden relación con el resto de cifras micro y macro del sistema nacional de información. Asimismo, se podrán considerar adecuados desde el punto de vista estadístico en la medida que el orden de su error absoluto y relativo esté en un rango aceptable, el cual fue definido al momento de establecer los objetivos de la investigación.

4.5 Precisión estadística de los estimadores

La evaluación de la calidad de la información desde la perspectiva estadística se efectúa a partir del cálculo de los errores de muestreo y del efecto de diseño. En primer lugar, los responsables de la administración del proyecto deben verificar si el margen de error utilizado en la determinación del tamaño de muestra está acorde con los resultados observados en la encuesta para la(s) variable(s) de diseño y algunas otras que estén altamente correlacionadas con ésta(s). Asimismo, se debe calcular el efecto de diseño y verificar si su valor es similar al que se utilizó para ajustar el tamaño de muestra debido a la conglomeración del marco.

Estas comparaciones mínimas, permitirán determinar si el esquema de muestreo aplicado fue adecuado, y si el número de observaciones seleccionadas fueron suficientes para generar resultados con la precisión requerida. En caso que el error estimado se encuentre en un rango aceptable, será posible afirmar que los resultados obtenidos se pueden utilizar para establecer relaciones de causalidad y realizar inferencia al universo de estudio con la confianza asumida en los objetivos de la investigación. En ese sentido, es preciso recordar que el error de estimación y el efecto de diseño son diferentes para cada variable y que su valor cambia entre las diferentes encuestas que se realizan, por lo que en la medida de lo posible se debe evitar utilizar resultados de otras investigaciones para estimar tamaños de muestra, calcular coeficientes de variación, límites de intervalos de confianza o probar la significancia estadística de algún parámetro.

4.6 Documentación de los procesos de la encuesta

Las actividades posteriores a la generación de resultados también deben considerarse como fases importantes en el control de calidad de la información. La documentación de las diferentes fases del proceso de la encuesta, así como de las decisiones que se asumieron ante problemas concretos, resultan fundamentales para evaluar la consistencia de los resultados y explicar distorsiones o sesgos en los datos en caso de que estos existan. Esta afirmación es válida para todas

las etapas de la encuesta, pero sobre todo en la que se vincula en forma directa con el tratamiento de la información y los procedimientos de ajuste e imputación que se aplican para suplir los datos faltantes.

En la medida que no se documenten las decisiones adoptadas sobre, por ejemplo, la concentración de la falta de respuesta total y parcial, los motivos y criterios para el ajuste de los factores de expansión, los procedimientos utilizados para la corrección de los ingresos y la imputación de la condición de actividad de la población ocupada, los usuarios de la información carecerán de elementos para determinar si sus análisis y las hipótesis en las que se basan los modelos de comportamiento se cumplen en función de la naturaleza de los datos.

Adicionalmente, se debe enfatizar la conveniencia de proveer a los usuarios de todos los elementos necesarios para que juzguen la calidad de la información y evalúen la posibilidad de utilizarla para la toma de decisiones. Así, se debe mejorar el contenido de la ficha técnica que acompaña la documentación de la encuestas. En el mejor de los casos, los documentos que se ponen a disposición de los usuarios contienen un resumen muy apretado del diseño estadístico de la encuesta y de algunos de los parámetros que se utilizaron para determinar el tamaño de la muestra. Sin embargo, son muy pocas las ocasiones en que se divulga el esquema de selección de la muestra y se identifican las unidades primarias y secundarias de selección, la distribución geográfica y por estrato social de la no respuesta, el miembro del hogar que proporcionó la información, el número de preguntas que se tuvieron que imputar, etc.

Se debe recordar que una buena manera de sensibilizar al usuario sobre los malos usos de la información es indicarle con claridad las limitaciones de los datos y las restricciones impuestas por el tamaño de muestra y los objetivos de la investigación, en aquellos casos en que la información no puede desagregarse para ciertos dominios de estudio. Además, deben proporcionársele los insumos informativos que le permitan estimar el error de muestreo de las variables que le interesen.

4.7 Difusión de resultados y relación con los usuarios

Las responsabilidades de las oficinas de estadística no se acaban con la divulgación del informe que contiene los resultados básicos de la encuesta. La relación con los usuarios debe ser permanente y cubrir un amplio espectro de actividades, que van desde la definición conjunta de los objetivos de la investigación a la retroalimentación mutua a partir de la entrega de las bases de datos con los registros individuales de la encuesta.

Es recomendable llevar un control de las personas e instituciones a las que se les facilita la información, y establecer algún tipo de convenio en que la oficina de estadística se comprometa a brindar asesoría sobre aspectos específicos de la encuesta y el archivo de datos, y los usuarios asuman el compromiso de consultar sobre la confiabilidad de la información antes de su publicación y la responsabilidad por los errores y/o imprecisiones que se cometan en la interpretación de los resultados.



NACIONES UNIDAS

CEPAL

ECLAC

Serie

estudios estadísticos y prospectivos

Números publicados

- 1 Hacia un sistema integrado de encuestas de hogares en los países de América Latina, Juan Carlos Feres y Fernando Medina (LC/L.1476-P), N° de venta: S.01.II.G.7, (US\$ 10.00), 2001.
[www](#)

- El lector interesado en números anteriores de esta serie puede solicitarlos dirigiendo su correspondencia a la Unidad de Distribución, CEPAL, Casilla 179-D, Santiago, Chile. No todos los títulos están disponibles.
- Los títulos a la venta deben ser solicitados a la Unidad de Distribución, CEPAL, Casilla 179-D, Santiago, Chile, Fax (562) 210 2069,

Nombre:

Actividad:.....

Dirección:.....

Código postal, ciudad, país:

Tel.:..... Fax:..... E.mail: