

## CAPÍTULO 7

# PROCESAMIENTO DE DATOS

*Este capítulo ha sido escrito para coordinadores de encuesta, expertos en procesamiento de datos y personas de recurso técnico. Provee información sobre cómo:*

- Prepararse para procesar los datos
- Montar un sistema para manejar el procesamiento de datos
- Ingresar los datos
- Editar los datos y crear un archivo de datos ‘depurado’ para análisis
- Producir tabulaciones con los indicadores
- Archivar y distribuir los datos

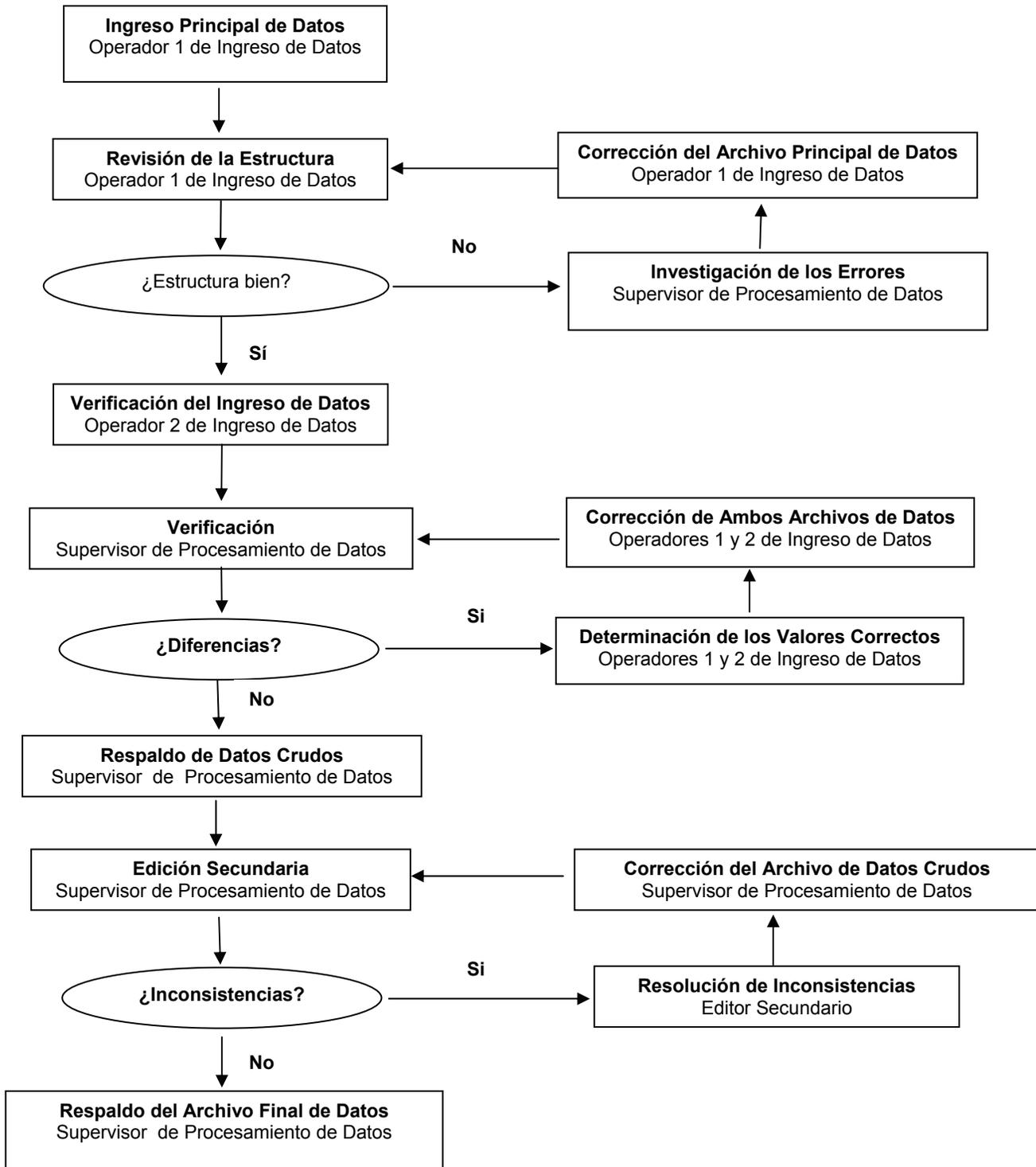
El sistema de procesamiento de datos MICS3 está diseñado para producir los primeros resultados de una encuesta unas pocas semanas después de terminado el trabajo de campo. Este capítulo contiene información que le ayudará a llevar a cabo la planeación y el trabajo preliminar de preparación para hacer realidad esta meta. El capítulo comienza con información general sobre el sistema de procesamiento de datos MICS3. A continuación, presenta en detalle cada uno de sus componentes, proporcionando referencias sobre fuentes de información adicionales, cuando corresponde. Finaliza con un juego de tres listas de control que le ayudarán a realizar el procesamiento de datos de su encuesta exitosamente.

### INFORMACIÓN GENERAL

La razón por la cual el sistema de procesamiento de datos MICS3 puede lograr una reducción de tiempo tan considerable se debe a que los datos se procesan paralelamente con el trabajo de campo. Los datos de cada conglomerado se almacenan en archivos de datos separados y son procesados tan pronto los cuestionarios regresan del campo. Este enfoque divide el procesamiento de datos en segmentos discretos y permite que el procesamiento de datos progrese mientras se lleva a cabo el trabajo de campo. En consecuencia, cuando se terminan los últimos cuestionarios y se devuelven a la sede, la mayor parte de los datos ya han sido procesados.

El procesamiento de datos por conglomerado no es difícil pero sí requiere una organización meticulosa. El sistema de procesamiento de datos se puede dividir en tres fases: preparación, procesamiento primario de datos y procesamiento secundario de datos. Cada una de estas fases se resume en una de las tres sub-secciones a continuación y cada una tiene asociada una lista de control que se encuentra al final de este capítulo.

**Tabla 7.1**  
**El Sistema de Procesamiento de Datos MICS3**



### **PREPARACIÓN PARA EL INGRESO DE DATOS**

La meta de prepararse para la fase de ingreso de datos es el estar listos para comenzar poco después de iniciar el trabajo de campo. La fase de preparación comprende los siguientes pasos:

- La obtención del equipo de computación y la organización de un recinto para el procesamiento de datos
- La identificación y contratación del personal adecuado
- La adaptación de los programas de computación al cuestionario específico de país
- La creación de un sistema para el manejo de cuestionarios y archivos de datos

### **PROCESAMIENTO PRIMARIO DE DATOS**

La meta del procesamiento primario de datos es producir archivos de datos depurados y editados. El procesamiento primario de datos comprende los siguientes pasos:

- Ingresar en un archivo de datos todos los cuestionarios de un conglomerado
- Revisar la estructura del archivo de datos
- Ingresar los datos por segunda vez y luego verificar el archivo de datos
- Respalidar el archivo de datos revisado y verificado
- Llevar a cabo la edición secundaria del archivo de datos
- Respalidar el archivo de datos editado o final

El flujo del procesamiento primario de datos se resume en el diagrama de flujo de la página anterior. Tome cuidadosa nota sobre el hecho que la revisión de la estructura, la verificación del ingreso de datos y la edición secundaria son procedimientos iterativos que se repiten hasta que los problemas se hayan resuelto o se haya establecido que son aceptables.

### **PROCESAMIENTO SECUNDARIO DE DATOS**

La meta del procesamiento secundario de datos es producir archivos de análisis de datos y crear las tablas estándar de MICS3. El procesamiento secundario de datos comprende los siguientes pasos:

- Concatenar todos los archivos de datos de los conglomerados en un solo archivo de datos
- Exportar los datos a SPSS
- Calcular los factores de ponderación
- Computar el índice de riqueza
- Recodificar las variables para simplificar el análisis
- Crear las tablas necesarias para analizar los datos
- Archivar y distribuir los archivos de datos

## PERSONAL E INFRAESTRUCTURA

### PERSONAL

El equipo de procesamiento de datos para una encuesta MICS3 incluye cuatro tipos de personal: un administrador de cuestionarios, los operadores de ingreso de datos, los editores secundarios y un supervisor de procesamiento de datos. Cada puesto de trabajo tiene responsabilidades específicas y fusionar las mismas podría afectar la calidad de sus datos.

El **administrador de cuestionarios** (o **editor de planta**) revisa y organiza los cuestionarios cuando llegan del campo. Cuando llega un conglomerado a la oficina de procesamiento de datos, él o ella revisa que todos los cuestionarios estén presentes y listos para ser ingresados. Si faltan cuestionarios, él o ella debe resolver el problema con la ayuda del equipo de trabajo de campo (los pasos exactos que debe dar el administrador de cuestionarios se detallan más adelante).

Los **operadores de ingreso de datos** ingresan los datos. Deben tener experiencia previa en el ingreso de datos y estar familiarizados con los cuestionarios. Una forma de lograr esto es hacer que los operadores de ingreso de datos asistan a la capacitación de entrevistadoras. Antes de comenzar el ingreso de datos, se debe llevar a cabo una capacitación independiente de dos o tres días para que los operadores se familiaricen con el programa de ingreso de datos y con el ritmo del sistema de procesamiento de los mismos. Al término de la capacitación, los operadores de ingreso de datos deben sentirse cómodos con el programa de ingreso de datos y estar conscientes de sus responsabilidades diarias. El número requerido de operadores de ingreso de datos depende del número de computadores disponibles, como se verá en detalle más adelante.

Los **editores secundarios** investigan y resuelven inconsistencias complejas descubiertas por el programa de edición secundaria. Deben tener una excelente comprensión de los cuestionarios y de las metas de la encuesta. Las directrices de edición se presentan en el Anexo Siete para ayudarles durante el proceso de edición secundaria. Una encuesta típica requerirá de uno o dos editores secundarios.

El **supervisor de procesamiento de datos** es un miembro fundamental del equipo de procesamiento de datos. Él o ella adapta los modelos de los programas para ajustarlos a los cuestionarios de su país y supervisa todas las labores de procesamiento de datos. El supervisor de procesamiento de datos debe tener experiencia en la administración del procesamiento de datos de encuestas o censos a gran escala, una excelente comprensión del cuestionario y conocimiento en programación de los paquetes de software de CPro y SPSS. El supervisor de procesamiento de datos debe estar disponible a tiempo completo durante el período de ingreso, edición y tabulación de datos.

El supervisor de procesamiento de datos debe haber sido identificado en las etapas iniciales de la planeación de la encuesta para que él o ella se pueda involucrar en la revisión del cuestionario MICS3. Esta persona debe ser consultada para asegurar que los esquemas de codificación que se

usen en el cuestionario sean consistentes y no tengan ambigüedades, y para que se incluya toda la información de identificación necesaria. El supervisor de procesamiento de datos también debe estar en capacidad de ayudar en las revisiones finales del cuestionario con base en la experiencia adquirida durante el ingreso de los cuestionarios de la prueba preliminar.

### COMPUTADORES Y OTRO EQUIPO

La lista a continuación muestra el equipo necesario para el procesamiento de datos:

- Computadores para el ingreso de datos
- El computador para el supervisor de procesamiento de datos
- Un dispositivo de archivo secundario (p. ej., dispositivo portátil de puerto USB o una unidad de CD-RW)
- Disquetes (o un medio para que los operadores puedan transferir archivos al supervisor de procesamiento de datos, p. ej., por intermedio de una red)
- Una impresora
- Papel
- Cartuchos de tinta / cintas de impresora
- Estabilizadores de corriente
- Proveedores de energía sin interrupciones (UPS)
- Bolígrafos verdes

Los computadores para el ingreso de datos deben tener procesadores Pentium, Windows 95 en adelante, 32 megabytes de memoria RAM o más, 1 gigabyte de espacio libre en el disco duro o más y/o unidades de disquete de 3.5" (o estar conectados en red). El número de computadores para el ingreso de datos depende del tamaño de la muestra de la encuesta, del número de horas que trabajará un operador de ingreso de datos por semana, del espacio disponible y del calendario de la encuesta. Para obtener un estimado del número de computadores necesarios para el ingreso de datos, usted deberá estimar el tiempo que le tomará a un operador ingresar los cuestionarios de un hogar típico y multiplicarlo por el número de hogares esperados de acuerdo con el diseño de la muestra. Si no logra estimar el tiempo que tomará ingresar los cuestionarios, use 20-30 minutos por hogar como guía aproximada, dependiendo del número esperado de mujeres, niños y niñas de cada hogar. Multiplique este estimado de tiempo por el número de hogares para obtener el total de horas requeridas para ingresar los datos. Divida este estimado por el número de horas que trabajará cada operador por semana y luego por el número de semanas que tendrá para completar el ingreso de datos (si espera completar la labor una semana después que los últimos cuestionarios hubiesen sido devueltos del campo).

Por ejemplo, si el tamaño de la muestra de la encuesta es de 6.000 hogares y cada hogar toma 20 minutos para ingresar, el tiempo total que se necesita para ingresar todos los hogares es de 2.000 horas. Si dispone de 8 semanas para completar el ingreso de datos, entonces necesitará 250 horas por semana. Si cada operador de ingreso de datos trabaja 40 horas por semana, necesitará 7 computadores y 7 operadores. A veces es posible organizar turnos dobles de ingreso de datos,

siendo así usado cada computador por dos operadores de ingreso de datos cada día. Cada operador trabajaría digamos que 6 horas, permaneciendo así el computador en uso 12 horas por día.

El computador del supervisor deberá tener un procesador más rápido, Windows 98 en adelante, 64 megabytes o más en RAM, 1 gigabyte de espacio libre o más en el disco duro, una unidad de disquete de 3.5" (o estar conectado a la red de computadores de ingreso de datos) y un dispositivo secundario de almacenamiento.

Los proveedores de energía sin interrupción y los estabilizadores de corriente son indispensables si el país en el cual trabaja sufre cortes de energía. Los bolígrafos verdes se deberán utilizar cada vez que un miembro del equipo de procesamiento de datos modifique los datos de un cuestionario. Los bolígrafos verdes le dan visibilidad a estos cambios frente a los registros originales del entrevistador (tinta azul) y a cualquier cambio efectuado por el equipo de trabajo de campo (tinta roja).

#### **PAQUETES DE SOFTWARE**

Los programas estándar para el procesamiento de las encuestas MICS3 se desarrollaron en CPro 2.6 y SPSS. CPro, que ha sido usado para procesar tanto encuestas como censos, se desarrolló en cooperación con la Oficina de Censo de Estados Unidos (*US Bureau of the Census*), ORC Macro International y SerPro Ltda. Se puede bajar gratuitamente de la página web de la Oficina de Censo de Estados Unidos<sup>1</sup>. SPSS es un paquete de software comercial que se encuentra disponible por intermedio de UNICEF y muchos otros proveedores de software.

#### **ESPACIO DE OFICINA**

Se requieren oficinas separadas para el ingreso de datos y la edición de los mismos. La oficina para el ingreso de datos debe ser lo suficientemente amplia como para que cada operador de ingreso de datos tenga espacio para su computador y el cuestionario en el cual está trabajando. Deben existir escritorios o mesas de trabajo y suficientes tomacorrientes. El recinto debe ser fresco, bien iluminado y libre de polvo y humedad hasta donde sea posible. En los países de clima caliente, es necesario que el lugar tenga aire acondicionado. Cada computador debe estar conectado a un proveedor de energía sin interrupciones (UPS). Si los cortes de luz son frecuentes o prolongados, se necesitará otro tipo de suministro de energía de emergencia, tal como un generador.

La oficina de edición de datos es para el administrador de cuestionarios y los editores secundarios. También deberá ser fresca y bien iluminada y tener espacio suficiente para que los editores puedan revisar los cuestionarios. Idealmente, la oficina de edición deberá contener suficiente estantería o gabinetes para archivar los cuestionarios en forma organizada. Si los

---

<sup>1</sup> La dirección de la página web es: <http://www.census.gov/ipc/www/cspro/>

cuestionarios no se pueden guardar en la oficina de edición, deben guardarse cerca, de tal forma que sean fácilmente accesibles puesto que serán usados durante diferentes etapas del procesamiento. Tenga la precaución de no subestimar la cantidad de espacio que se vaya a necesitar para guardar los miles de cuestionarios que tendrá en la oficina cuando finalice el trabajo de campo.

### **ADAPTAR LOS PROGRAMAS ESTÁNDAR**

Como se describe en el Capítulo 3, el modelo de cuestionario MICS3 se debe adaptar a la situación de cada país. Esto significa que los programas modelo de ingreso de datos, edición y tabulación también requieren modificaciones para que guarden consistencia con los cambios efectuados en el cuestionario. Cuantos más cambios tenga el cuestionario modelo, tanto más tiempo se requerirá para la adaptación y prueba de los programas. Por ejemplo, si se agregan nuevas preguntas al cuestionario, se deberán efectuar las correspondientes adiciones a los programas de ingreso de datos, de edición y de tabulación.

Este proceso será considerablemente más fácil si se mantiene la numeración de las preguntas del cuestionario modelo. Si se agregan preguntas, habrá que agregar una letra a la numeración existente (p. ej., una pregunta que se inserte entre WS4 y WS5 deberá llevar el número WS4A). Igualmente, si se eliminan preguntas, no se deberán reenumerar las preguntas restantes. Adicionalmente, cuando se agregan categorías de codificación a las que aparecen en el cuestionario modelo, deben agregarse al final de la lista existente, dejando los demás códigos intactos. La adaptación de programas de ingreso de datos y edición debe estar terminada antes de la prueba preliminar. Los cuestionarios de prueba pueden ser ingresados y editados usando los programas. Este procedimiento cumplirá con dos propósitos. Revelará los problemas de codificación y los patrones de pases en los cuestionarios y también mostrará cualquier error de los programas. Cuando se haya completado la prueba y el cuestionario esté finalizado, se podrán efectuar los cambios finales de los programas. Las secciones a continuación ofrecen una guía básica para modificar los diccionarios de datos modelo y las aplicaciones modelo de CSPro. En documentos independientes encontrará un resumen más detallado de los contenidos de las aplicaciones de CSPro.

Incluso si no agrega preguntas a los cuestionarios modelo, los diccionarios de datos modelo y las aplicaciones contienen algunos ítems que deben ser actualizados (p. ej., el rango aceptable de la fecha de la entrevista, el rango aceptable para el número de conglomerado, etc.). Estos ítems son necesariamente específicos del país y usted debe completarlos. Así, aunque su país use el cuestionario modelo, usted tendrá que adaptar los programas estándar.

## LOS DICCIONARIOS DE DATOS

En la Encuesta de Indicadores Múltiples por Conglomerados, los conjuntos de preguntas relacionadas (p. ej., sobre mortalidad materna, uso de anticonceptivos e inmunización) se encuentran agrupados en módulos que a su vez se agrupan en cuestionarios (en particular, para hogares, mujeres y niños/niñas menores de cinco años). En CSPro, los diccionarios se usan para describir esta estructura de datos: un grupo de variables relacionadas (preguntas) conforma un registro (módulo), y un grupo de registros conforma un nivel (cuestionario). Éstos se guardan en un archivo de diccionario (con extensión *dcf*). Adicionalmente al diccionario de datos, existen formas asociadas al diccionario que se usan para el ingreso de datos. Normalmente, existe una forma para cada registro. Los formatos se guardan en un archivo de formatos (con extensión *fmf*). Los archivos *dcf* y *fmf* se pueden modificar directamente. La mejor forma de hacerlo es abriendo el archivo de formas de CSPro. Eso le dará acceso al diccionario de datos junto con las formas, y de esa manera se garantizará que continúen sincronizados. Se recomienda mantener un respaldo de los modelos del diccionario de datos y del archivo de formas para que sirva de referencia.

Existen tres tipos de cuestionarios MICS. El Cuestionario de Mujeres Individuales y el Cuestionario de Niños/as Menores de Cinco corresponden a una sola unidad de análisis: una mujer y un niño/niña, respectivamente. El Cuestionario del Hogar contiene dos unidades de análisis: el hogar y los miembros del hogar. Todos los tipos de cuestionario están archivados en *mics3.dcf* y *entry.fmf*.

### VARIABLES Y NIVELES DE IDENTIFICACIÓN

En CSPro todo cuestionario debe tener una serie de variables que lo identifiquen exclusivamente. Por ejemplo, un hogar se identifica por sus números de conglomerado y de hogar. Las variables que identifican un cuestionario se conocen como variables de identificación. En la tabla a continuación verá la lista de tipos de cuestionario y sus variables de identificación.

**Cuadro 7.2**  
**Tipos de Cuestionario y sus Variables de Identificación**

Cuestionario	Número de Conglomerado	Número de Hogar	Número de Línea
Hogar	HH1	HH2	
Mujer	HH1	HH2	LN
Niños/Niñas	HH1	HH2	LN

Como puede ver en la tabla anterior, las mujeres y los niños/niñas tienen las mismas variables de identificación. Dado que cada miembro de hogar aparece listado en una línea separada de la lista de hogares, no habrá dos mujeres o niños/niñas con el mismo número de línea, incluso si se encuentran en el mismo hogar. Por lo tanto, el número de línea identifica exclusivamente a una

sola mujer o niño/niña cuando se combina con el número de conglomerado y el número de hogar.

En un diccionario de CSPro, el nivel está definido por un conjunto de variables de identificación. En el diccionario MICS3 existen dos niveles: hogares e individuos (específicamente, mujeres elegibles y niños/niñas elegibles). Los hogares corresponden al primer nivel, en tanto que mujeres y niños/niñas pertenecen al segundo nivel. Esta estructura jerárquica es natural puesto que en el cuestionario de MICS3 toda mujer o niño/niña pertenece a un hogar mientras que un hogar puede tener muchas mujeres o niños/niñas.

Los cuestionarios de mujer y los cuestionarios de los niños/niñas se guardan en el mismo nivel, porque cada uno corresponde a un cuestionario que se aplica a un solo miembro del hogar. La aplicación de ingreso de datos contiene lógica que omita las formas pertenecientes a los niños/niñas cuando se ingresa un cuestionario de mujer y omita las formas pertenecientes a las mujeres cuando se ingresa un cuestionario de niño/niña. Así, aunque los cuestionarios de mujer y los cuestionarios de niño/niña se guardan ambos como casos del nivel dos, no tienen variables en común exceptuando las variables de identificación.

## MÓDULOS

El diccionario de datos se diseñó para reflejar la estructura modular de los cuestionarios de MICS3. Cada módulo se guarda en su propio registro (excepción: los módulos de Listado de Hogar y de Niños/niñas Huérfanos y Vulnerables a causa de VIH/SIDA, donde cada uno tiene dos registros debido a su estructura poco usual) en *mics3.dcf* y cada registro tiene una forma (o dos, como en el caso del módulo de Características del Hogar) asociado con él en *entry.fmf*. Así, si su país no usa un módulo en particular, lo puede eliminar borrando su registro y su forma (y eliminando cualquier lógica adicional que haga referencia a él en la aplicación de ingreso de datos).

Los módulos disponibles para el cuestionario de hogar (con el(los) código(s) del módulo entre paréntesis) son: Información del Hogar (HH), Listado del Hogar (HL y TO), Educación (ED), Agua y Saneamiento (WS), Características del Hogar (HC), Mosquiteros Tratados con Insecticida (TN), Niños/as Huérfanos y Vulnerables por VIH/SIDA (OV y OR), Trabajo Infantil (CL), Disciplina del Niño/a (CD), Discapacidad (DA), Mortalidad Materna (MM) y Yodación de Sal (SI).

Los módulos disponibles para el cuestionario de mujeres son: Panel de Información de la Mujer (WM), Mortalidad en la Niñez (CM), Toxoide Tetánico (TT), Salud Materna y del Recién Nacido (MN), Matrimonio/Unión (MA), Seguridad de la Tenencia y Durabilidad de la Vivienda (ST), Anticoncepción (CP), Mutilación Genital Femenina (FG), Actitud hacia la Violencia Doméstica (DV), Comportamiento Sexual (SB) y VIH/SIDA (HA).

Los módulos disponibles para el cuestionario de niños/niñas son: Panel de Información de Menores de Cinco Años (UF), Registro de Nacimiento y Aprendizaje Temprano (BR), Desarrollo del Niño/a (CE), Vitamina A (VA), Lactancia (BF), Atención a enfermedades (CA), Malaria (ML), Inmunización (IM) y Antropometría (AN).

### CONVENCIONES PARA LA DENOMINACIÓN DE LAS VARIABLES

Las variables se denominan de acuerdo con el módulo del cuestionario en el cual se encuentran y con el número de pregunta cuya respuesta contienen. Por ejemplo, la pregunta 9 en el Listado del Hogar se guarda en una variable denominada HL9. Algunas preguntas se dividen en dos o más partes, con las diferentes partes identificadas con una letra única. Cada parte de dichas preguntas se guarda en una variable independiente. Los nombres de estas variables independientes incluyen letras que identifican las partes de la pregunta. Por ejemplo, la pregunta 11 del módulo de Salud Materna y del Recién Nacido consta de dos partes. La primera parte de esta pregunta se guarda en la variable *MN11A* y la segunda parte se guarda en *MN11*.

Algunas preguntas constan de dos o más partes de categorías de respuesta. Estas preguntas se guardan en una variable única y las categorías de respuesta se definen como sub-ítems. Cuando estas preguntas involucran fechas, las letras *d* (para día), *m* (para mes) y *y* (para año) se anexan al nombre de la variable básica para crear el nombre de los sub-ítems. Por ejemplo, en la pregunta 6 del Panel de Información de la Mujer, se requiere el día, mes y año de nacimiento de la mujer. Su respuesta se guarda en *WM8*, que consta de tres sub-ítems: *wm8d*, *wm8m* y *wm8y*.

Algunas preguntas tienen una estructura en la cual la primera parte de la respuesta es la forma de la respuesta y la segunda parte es la respuesta. Estas preguntas se guardan en una variable única, y la forma y la respuesta se definen como sub-ítems. El nombre del sub-ítem que guarda la forma de la respuesta es el nombre de la variable con la letra *u* (para unidades) anexa a él, en tanto que el nombre del sub-ítem que guarda la respuesta corresponde al nombre de la variable con la letra *n* (para número) anexa a él. Por ejemplo, la pregunta 13 del módulo de Salud Materna y del Recién Nacido registra cuánto tiempo después de su nacimiento el niño/niña recibió leche materna por primera vez. La encuestada podría responder en horas o días. La respuesta se guarda en la variable *MN13* con los sub-ítems *mn13u* y *mn13n*.

### PREGUNTAS CON RESPUESTAS MÚLTIPLES Y VARIABLES ALFANUMÉRICAS

En el cuestionario existe un número de preguntas que permiten respuestas múltiples. Estas preguntas se identifican en el cuestionario por medio de códigos de respuesta alfanuméricos (es decir, letras de la A hasta la Z). En el diccionario de datos, la respuesta a una pregunta de múltiples respuestas se guarda en una variable alfanumérica cuya longitud es igual al número máximo de respuestas potenciales. Éstas son las únicas variables alfanuméricas del diccionario. Cada variable alfanumérica consta de un sub-ítem para cada código de respuesta del cuestionario. El nombre de uno de estos sub-ítems está conformado por el nombre de la variable más el código de la respuesta que representa ese sub-ítem.

Por ejemplo, la segunda pregunta en el módulo de Salud Materna y del Recién Nacido registra a todas las personas que le brindaron cuidado prenatal a una mujer antes de su último parto. Los códigos de respuesta potencial son A, B, C, F, G, H, X y Y. En consecuencia, la variable *mn2* tiene una longitud de 8 caracteres con ocho sub-ítems: *mn2a*, *mn2b*, *mn2c*, *mn2f*, *mn2g*, *mn2h*, *mn2x* y *mn2y*.

### CONVENCIONES DE CODIFICACIÓN

Los modelos de diccionario usan una codificación estándar para determinadas respuestas. Primero hablaremos sobre las convenciones de codificación para las variables numéricas. La respuesta “Otro” siempre se codifica con un 6 antecedido por nueves. Las respuestas inconsistentes se codifican siempre con un 7 antecedido por nueves. La respuesta “No sabe” se codifica siempre con un 8 antecedido por nueves. Las preguntas con una respuesta faltante (es decir, la entrevistadora no registró la respuesta de una pregunta aplicable) se codifica siempre con un 9 antecedido por nueves. Las preguntas que no aplican a un encuestado se codifican siempre en blanco. El Cuadro 7.3 abajo resume las convenciones de la codificación estándar.

**Cuadro 7.3**  
**Resumen de la Convenciones de Codificación Estándar**

Respuesta	Longitud de la Variable				
	Alfabética	Un carácter	Dos caracteres	Tres caracteres	Cuatro caracteres
Otro	X	6	96	996	9996
No/Ninguna	Y	na	na	na	na
Inconsistente	na	7	97	997	9997
No sabe	Z	8	98	998	9998
Faltante	?	9	99	999	9999
No aplica	En blanco	En blanco	En blanco	En blanco	En blanco

Puesto que los códigos del 6 al 9 están reservados para uso especial, cualquier pregunta que requiera más de 6 categorías de respuesta deberá tener categorías de respuesta de 2 dígitos antecidas por ceros (p. ej.: 01, 02, 03, 04, 05, 06, 07, 96, 97, 98 y 99).

Para las variables alfanuméricas, la respuesta “Otro” se codifica siempre con una X, la respuesta “No sabe” con una Z, un valor faltante se codifica siempre usando el carácter de pregunta (?), y lo que no es aplicable se codifica en blanco.

### RANGOS

La mayoría de las preguntas de los cuestionarios MICS3 tienen rangos de respuesta definidos. Los rangos para las variables se definen en el diccionario *mics3.dcf*. CSPro verifica durante el ingreso de datos que todo valor que se ingrese en una variable se encuentre dentro de los rangos

definidos para esa variable. CSPro permite un amplio número de rangos para cada variable, así las preguntas con rangos no-consecutivos de respuesta (p. ej., 1-8, 96, 98 y 99) se deben definir usando varios rangos (p. ej., 1-8, 96, 98, 99 en vez de 1-99). Aunque los rangos del diccionario son útiles para la verificación de rangos sencillos, los rangos más complicados o condicionales (p. ej., consistencia entre el día y el mes en una variable de fecha) se deben verificar en las aplicaciones de ingreso o edición de datos.

### APLICACIÓN DE INGRESO DE DATOS

La aplicación de ingreso de datos es un programa extenso y complejo. Las limitaciones de espacio no permiten que la describamos en detalle en este capítulo. En vez de ello, esta subsección se enfocará en algunos temas generales importantes relacionados con la aplicación de ingreso de datos.

#### PASES

Los cuestionarios MICS usan los pases con mucha frecuencia. Los pases son instrucciones del cuestionario que le indican al entrevistador que pase por alto todas las preguntas entre la pregunta actual y una pregunta posterior del cuestionario. Los pases en un cuestionario deben coincidir con los pases del programa de ingreso de datos correspondiente. Los pases en un programa de ingreso de datos definen la ruta de ingreso de datos. CSPro implementa estrictamente la ruta de ingreso de datos cuando se usan los comandos *pase a (skip to)* o *pase a siguiente (skip to next)*.

#### MENSAJES DE ERROR

Es útil que aparezca un mensaje de error cuando un operador de ingreso de datos ingresa el valor de una variable que es inconsistente con la información ingresada previamente. Ese mensaje de error debe explicar la naturaleza del problema y proporcionar cualquier información que pudiera ayudar a resolver la inconsistencia. En CSPro, la función *errmsg* muestra un mensaje de error con texto definido por el usuario cuando se activa. Los mensajes de error del programa de ingreso de datos se numeran y guardan en el archivo *entry.mgf*. El texto, número y las inconsistencias que llevan a cada uno de los mensajes que aparecen, se encuentran detallados en la lista del Anexo Siete, al igual que las directrices para resolverlos.

Usted deberá revisar su cuestionario para determinar si alguna pregunta que ha sido agregada requiere verificaciones de consistencia. En caso afirmativo, usted deberá agregarle la lógica al programa de ingreso de datos, al programa de edición o a ambos para verificar la consistencia. Cuando usted agrega una verificación de consistencia, asegúrese de agregar el correspondiente mensaje de error al archivo de mensajes de ingreso de datos o de edición. Además, si usted agrega mensajes de error, asegúrese de no usar un número existente de mensaje de error.

Algunos mensajes de error traen a continuación el comando *reenter* (reingresar) que regresa al campo que se está ingresando. Esto obliga al operador a resolver el error antes de continuar. Debido a que el operador de ingreso de datos en algunas ocasiones deberá ingresar correcciones, se requiere una cuidadosa supervisión. Si usted agrega sus propios mensajes de error, considere cuidadosamente si desea obligar al operador de ingreso de datos a resolver el problema antes de seguir adelante. En ese caso, a continuación de su mensaje de error incluya el comando *reenter*.

### VARIABLES ALFANUMÉRICAS

La aplicación de ingreso de datos verifica que las variables alfanuméricas se ingresen correctamente. Ejecuta cuatro revisiones para cada variable alfanumérica. Primero, revisa que el valor ingresado contenga únicamente los códigos listados en el cuestionario (es decir, ejecuta una verificación de rango). Segundo, revisa que las respuestas se ingresen en orden alfabético (p. ej., *ACG* y no *GAC*). Tercero, revisa que, si los códigos de “No sabe” o “Ninguno” (generalmente la letra ‘Y’) se incluyen en la respuesta, entonces no existe ninguna otra respuesta (p. ej., no permitiría la respuesta *ACY*). Cuarto, revisa que si el código de faltante (‘?’) aparece incluido en la respuesta, entonces no existe ninguna otra respuesta (p. ej., no permitiría la respuesta *AC?*).

La aplicación de ingreso de datos también reorganiza los valores ingresados en las variables alfanuméricas de tal forma que cada respuesta se guarda en el sitio que establece su sub-ítem. Por ejemplo, para la variable MN2, la respuesta *ACG* se reorganizará para quedar *A C G* , donde aparece un espacio en blanco entre la *A* y la *C* y la *C* y la *G* y tres espacios en blanco a continuación de *G*.

### FUNCIONES DEFINIDAS POR EL USUARIO

Una buena característica de CSPro es que permite a los programas definir sus propias funciones. Dichas funciones se conocen como funciones definidas por el usuario y pueden ser de utilidad. Especialmente, permiten que uno evite rescribir el código usado frecuentemente. Las funciones definidas por el usuario siempre aparecen en la parte superior de una aplicación de CSPro. La aplicación de ingreso de datos *entry.app* contiene 14 funciones definidas por el usuario. Usted no necesita modificar estas funciones, pero debe comprender lo que hacen si desea entender la aplicación de ingreso de datos.

La función *valid* revisa si el valor de una variable corresponde a uno de los valores especiales: inconsistente, no sabe, faltante o no aplica. Si el valor de una variable es no aplica, la función *natozero* lo cambia a ‘0’, permitiendo que se pueda sumar a otra variable (para conocer un ejemplo de cómo usarla, refiérase al procedimiento *cm9*).

Las siguientes 7 funciones definidas por el usuario (*zscoef*, *dabs*, *zspct*, *zseval*, *zscr*, *zsanth*, y *agemth*) se emplean en el Cuestionario de Niños/Niñas para calcular las medidas antropométricas que se encuentran al final del módulo de antropometría. Se usa la función *agemth* para calcular la edad del niño/niña en meses. A continuación se usa la función *zsanth*. Esta función activa *zseval*,

*zscr* y *zspct*. La función *zseval* llama a *zscoef*, y *zspct* llama a *dabs*. Estas funciones las encontrará únicamente en las variables de antropometría, y si las llega a encontrar, sabrá que están calculando y luego verificando las medidas antropométricas.

El código de la función *agemth* calcula la edad del niño/niña en meses. Teniendo en cuenta que la antropometría es altamente sensible a la edad, la edad del niño/niña debe basarse en la edad del menor en días. El código calcula primero el número de días que han transcurrido entre el comienzo del año y el nacimiento del niño/niña. A continuación calcula el número de días transcurridos entre el comienzo del año y la fecha de la entrevista. Finalmente agrega el número de días de los años que van entre el año de nacimiento y el año de la entrevista al número de días que van desde el comienzo del año hasta la fecha de la entrevista. La diferencia entre estos dos números de días corresponde a la edad del niño/niña en días. Entonces, éstos se convierten en la edad del niño/niña en meses al dividir por 30,4373 (el número promedio de días de un mes durante cuatro años). Debido a la necesidad de exactitud, la edad del niño/niña en meses se calcula con dos decimales.

Las funciones *vdvalid*, *vdoi* y *vdo* verifican que las fechas de vacunación ingresadas en el módulo de inmunización sean consistentes, que no tengan una fecha posterior a la de la entrevista y una fecha anterior a la del nacimiento, respectivamente. La función de *endmess* (abreviatura de ‘end message’ o ‘terminar mensaje’) muestra un mensaje al final de un cuestionario que le pregunta al operador de ingreso de datos si desea revisar el cuestionario actual o continuar al siguiente. Finalmente, la función *alphack* ejecuta las verificaciones de las variables alfanuméricas detalladas en la sub-sección anterior.

### **FECHAS Y CÓDIGO DE MES DEL SIGLO**

Los programas modelo (incluyendo la aplicación de ingreso de datos) usan los códigos de mes-siglo (CMC) para la mayoría de las fechas. El CMC para una fecha es el número de meses a partir de diciembre de 1899. Por ejemplo, el CMC para enero de 1900 es 1; el CMC para marzo de 2000 es 1203. El CMC para una fecha se calcula en la siguiente forma: reste 1900 al año de la fecha, multiplique el número resultante de años por 12, y luego súmele al resultado el número de mes de la fecha. Por ejemplo, el CMC de marzo de 2000 se calcula como  $(2000-1900) \times 12 + 3$ .

La aplicación de ingreso de datos usa cuatro funciones para simplificar el trabajo con los códigos de mes-siglo. Dos de estas funciones, *setlb* y *setub*, calculan los límites inferior y superior respectivamente, para el CMC de la fecha de un evento. Las otras dos funciones, *adjlba* y *adjuba*, ajustan los límites inferior y superior, respectivamente, del CMC de la fecha de un evento (p. ej., el cumpleaños de un menor) donde también se especifica una edad. El Cuadro 7.4 a continuación resume estas funciones.

**Cuadro 7.4**  
**Funciones de CS Pro para Simplificar el Trabajo con Códigos de Siglo-Mes**

<pre>lcmc = setlb (month, year, minimum);</pre>	<p>Los argumentos de la función son un <i>mes (month)</i>, <i>año (year)</i> y un CMC <i>mínimo (minimum)</i>. Si el <i>año</i> y el <i>mes</i> son válidos, el CMC se calcula y es presentado. Si el <i>año</i> no es válido, se presenta el <i>mínimo</i>. Si el <i>mes</i> no es válido, presenta el CMC de enero del <i>año</i>.</p>
<pre>ucmc = setub (month, year, maximum);</pre>	<p>Los argumentos de la función son un <i>mes (month)</i>, <i>año (year)</i> y un CMC <i>máximo (maximum)</i>. Si <i>año</i> y <i>mes</i> son válidos, el CMC se calcula y es presentado. Si el <i>año</i> no es válido, se presenta el <i>máximo</i>. Si el <i>mes</i> no es válido, se presenta el CMC de diciembre del <i>año</i>.</p>
<pre>t = adjlba (lcmc, ucmc, di, di, age); if t &lt; 0 then   e = errmsg (2171); else   lcmc = t; endif;</pre>	<p>Los argumentos de la función son el CMC mínimo de una fecha (<i>lcmc</i>), el máximo CMC de una fecha (<i>ucmc</i>), el CMC para la fecha de la entrevista (<i>di</i>), y una <i>edad (age)</i>. La función incrementa <i>lcmc</i> usando <i>age</i> y <i>di</i>. La función presenta la fecha resultante si ésta es mayor o igual a <i>lcmc</i> y menor o igual que <i>ucmc</i>. Si la fecha resultante es menor que <i>lcmc</i>, la función presenta <i>lcmc</i>. Si la fecha resultante excede <i>ucmc</i>, la función presenta -1.</p>
<pre>t = adjuba (lcmc, ucmc, di, di, age); if t &lt; 0 then   e = errmsg (2171); else   ucmc = t; endif;</pre>	<p>Los argumentos de la función son el CMC mínimo de una fecha (<i>lcmc</i>), el CMC máximo de una fecha (<i>ucmc</i>), el CMC para la fecha de la entrevista (<i>di</i>), y una <i>edad (age)</i>. La función reduce <i>ucmc</i> usando <i>edad</i> y <i>di</i>. La función presenta la fecha resultante si es mayor o igual a <i>lcmc</i> y menor o igual a <i>ucmc</i>. Si la fecha resultante es mayor que <i>ucmc</i>, la función presenta <i>ucmc</i>. Si la fecha resultante es menor que <i>lcmc</i>, la función presenta -1.</p>

**RECEPCIÓN DE LOS CUESTIONARIOS PROCEDENTES DEL CAMPO**

Cuando los cuestionarios de un conglomerado llegan del campo, el administrador de cuestionarios debe verificar el número de cuestionarios con la hoja de control de conglomerados. Si falta algún cuestionario, el administrador de cuestionarios debe contactar al equipo de campo para ver si es posible encontrarlo. En caso contrario, el equipo de campo debe rehacer el cuestionario faltante en caso de que ello sea posible (es decir, si el equipo de trabajo de campo se encuentra cerca del conglomerado). Si esto no es posible y el cuestionario faltante corresponde a un Cuestionario de Hogar, el administrador de cuestionarios debe agregar a los cuestionarios del conglomerado un cuestionario en blanco (excepto por las variables de identificación) con un código de resultado de '6'. Si el cuestionario faltante corresponde a un cuestionario de mujer o de niño/niña, el administrador de cuestionarios debe cambiar los totales de los cuestionarios completados en la portada del hogar al cual pertenece la mujer o niño/niña.

El administrador de cuestionarios debe registrar la llegada de los cuestionarios del campo en el formulario de seguimiento de conglomerados. El formulario de seguimiento de conglomerados se encuentra disponible en ambos formatos, el electrónico y el impreso. Al final del presente capítulo, encontrará un ejemplo del formulario en formato impreso, y la versión electrónica se encuentra en el menú del supervisor. Antes de recibir algún cuestionario procedente del campo, es necesario hacer una lista de los conglomerados en orden ascendente en el formulario de seguimiento de conglomerados. A medida que se van recibiendo los cuestionarios de los conglomerados, la información de cada uno se puede ir registrando en el espacio correspondiente. Adicionalmente, para cada conglomerado, el administrador de cuestionarios se debe asegurar que:

- Se encuentren todos los cuestionarios listados en la hoja de control de conglomerados
- Los cuestionarios del hogar, de mujeres y de niños/niñas correspondientes a cada hogar estén juntos
- Los cuestionarios estén ordenados por número de hogar en orden ascendente
- Todos los cuestionarios de mujeres estén juntos, clasificados en orden ascendente por el número de línea de la mujer, seguidos por todos los cuestionarios de niños/niñas clasificados también en orden ascendente según el número de línea
- Todos los códigos geográficos y la información de la entrevista estén completados en el Panel de Información del Hogar al comienzo de cada Cuestionario del Hogar

Una vez finalizada la verificación y resuelto cualquier problema, el administrador de cuestionarios debe registrar los números de cuestionarios de hogares, mujeres y niños/niñas en el formulario de seguimiento de conglomerados. Él debe crear también una hoja de resumen de conglomerados. La hoja de resumen de conglomerados debe ser una hoja de papel grueso que se dobla sobre un extremo del paquete de cuestionarios y muestra el número del conglomerado en negrita. La hoja de resumen de conglomerados también debe presentar el número de cuestionarios en el conglomerado, como se muestra en el Cuadro 7.5 adelante.

Una vez creada la hoja de resumen de conglomerados, el supervisor debe asignarle el conglomerado a un operador de ingreso de datos (tenga en cuenta que el operador asignado debe ingresar todos los cuestionarios del conglomerado). El administrador de cuestionarios deberá registrar el nombre del operador de ingreso de datos en el formato de seguimiento de conglomerados.

**Cuadro 7.5**  
**Muestra de Hoja de Resumen de Conglomerados**

101	
Total de Cuestionarios de Hogar	_____
Total de Hogares Completados	_____
Total de Cuestionarios de Mujeres	_____
Total de Mujeres Completados	_____
Total de Cuestionarios de Niños/Niñas	_____
Total de Niños/niñas Completados	_____

Cuando se retiren los cuestionarios de su lugar de almacenaje para ser ingresados o editados, se deberán volver a empaquetar siempre en el mismo orden y devolver al mismo lugar. El manejo estricto y sistemático de los cuestionarios hará que los diferentes pasos del procesamiento de datos se lleven a cabo en forma más rápida y eficiente.

Cuando un conglomerado ha sido ingresado en el formulario de seguimiento de conglomerados, el supervisor debe revisar los cuestionarios y verificar que estén listos para ser ingresados. Las verificaciones específicas que se deben llevar a cabo se encuentran detalladas en El Cuadro 7.6 adelante. Una vez finalizada la verificación y resuelto cualquier problema que se hubiese detectado, el supervisor le asigna el conglomerado al un operador de ingreso de datos.

### INGRESO PRINCIPAL DE DATOS

Es mejor que se comiencen a ingresar los datos inmediatamente después que las entrevistadoras hayan iniciado el trabajo de campo. Esto le permitirá ubicar y corregir errores que determinadas entrevistadoras o equipos pudieran estar cometiendo. Problemas serios que pudiesen escapar a la atención del supervisor de campo, se pueden detectar a tiempo para re-entrenar al personal de campo y corregir los errores graves. Los supervisores de campo son responsables de revisar los cuestionarios para verificar que estén completos y sean consistentes, y clasificar cualquier respuesta que el entrevistador no hubiese podido codificar. Cuando los cuestionarios se devuelven a la oficina, no deberían requerir sino verificaciones mínimas como éstas, como se describe arriba. El ingreso de datos debe comenzar tan pronto los datos de un conglomerado son devueltos a la sede.

**Cuadro 7.6**  
**Verificaciones en la Oficina**

1. Asegúrese que todos los Cuestionarios del Hogar del conglomerado estén organizados en orden ascendente por número de hogar.
2. Asegúrese que estén completados todos los códigos geográficos y de información de la entrevista del Panel de Información del Hogar al comienzo de cada Cuestionario del Hogar.
3. La elegibilidad de cada miembro de hogar debe verificarse en el Formulario de Listado de Hogares. Para ser elegible para los módulos de mujer, la persona debe ser mujer con edades de 15 a 49 años. Para ser elegible para los módulos de trabajo infantil, la persona debe tener de 5 a 14 años (o el rango de edad específico del país). Para ser elegible para los módulos de salud infantil, la persona debe ser menor de 5 años. Para cada persona elegible, el código de elegibilidad debe aparecer marcado con un círculo en el número de línea correspondiente o en el número de línea de la madre o persona encargada (HL6-HL8 del Formulario de Listado del Hogar). Para otros miembros del hogar, las columnas de elegibilidad deben aparecer en blanco en el Listado del Hogar, y se ingresarán como 00. Para resolver cualquier problema, siga las instrucciones de las directrices de edición (Anexo 6) para los mensajes 0101-0132.
4. El número total de mujeres y niños/niñas elegibles registrado en el Panel de Información del Hogar debe coincidir con el número de mujeres y niños/as elegibles registrado en el Formulario de Listado de Hogares.
5. El número total de mujeres y niños/niñas elegibles registrado en el Panel de Información del Hogar debe coincidir con el número de cuestionarios de mujeres y niños/niñas del hogar.
6. El número de entrevistas de mujeres y niños/niñas completadas registradas en el Panel de Información del Hogar debe coincidir con el número de cuestionarios de mujeres y niños/niñas que tengan el código de resultado '1' ('Completo') para ese hogar.
7. Asegúrese que todos los Cuestionarios de Mujeres Individuales estén juntos, organizados en orden ascendente según el número de línea de mujer (WM4), seguidos por todos los Cuestionarios de Niños/niñas organizados en orden ascendente según el número de línea de niño/niña (UF4).

El proceso de ingreso de datos está regido por una aplicación de CSPro llamada *entry.ent*. Esta aplicación ejecuta dos labores principales. Primero, para un hogar determinado, asegura que todos los cuestionarios (y solamente esos cuestionarios) sean ingresados y que toda la información contenida en ellos (y solamente esa información) sea ingresada. Específicamente, la

aplicación de ingreso de datos comienza ingresando todos los datos de un Cuestionario del Hogar. Una vez ingresados estos datos, revisa el Listado del Hogar y determina cuáles miembros del hogar son elegibles para los cuestionarios de mujeres. Habiendo identificado estos miembros, la aplicación de ingreso de datos pide los datos de cada individuo desde el cuestionario de mujeres, en orden ascendente según el número de línea. A continuación, la aplicación de ingreso de datos ejecuta el mismo procedimiento con los miembros del hogar que son elegibles para el cuestionario de niños/niñas. Cuando se ha completado el ingreso de todos los cuestionarios del hogar, se guardan los datos de todo el hogar y la aplicación sigue adelante con el siguiente hogar del conglomerado, en caso de haberlo.

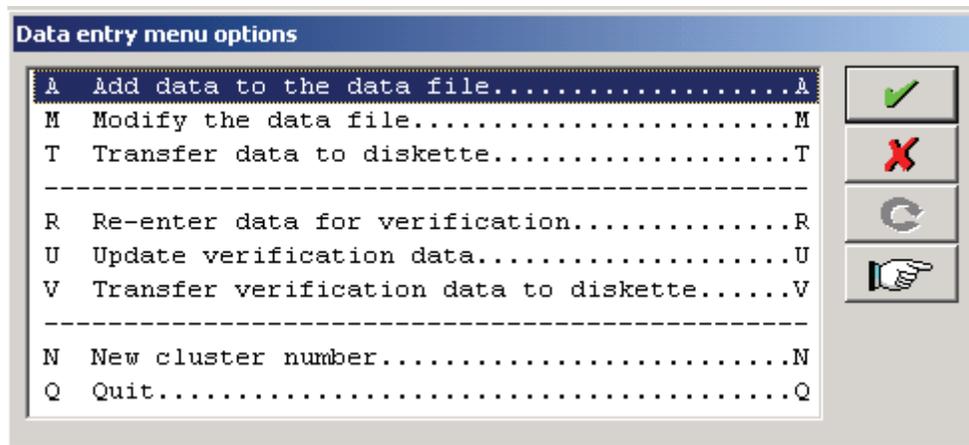
Teniendo en cuenta que los datos se guardan únicamente cuando todas las mujeres y niños/niñas elegibles pertenecientes al hogar han sido ingresados, los operadores de ingreso de datos no deben abandonar sus computadores en medio del ingreso de los datos de un hogar. Antes de tomar un receso o de terminar el trabajo del día, se deben ingresar todos los cuestionarios pertenecientes a un hogar. Además, se recomienda copiar los datos al computador del supervisor o a un disquete como medida de precaución. Adicionalmente, todas las tardes, el supervisor debe copiar el contenido de *c:\mics* y de todos los sub-directorios a la unidad de almacenamiento secundario. Esta medida de seguridad le permitirá al supervisor recuperar la información en caso de que su computador presente alguna falla.

Además de controlar los cuestionarios que se ingresan, la aplicación de ingreso de datos controla rigurosamente el patrón de pases de un cuestionario. Es decir, que solamente pedirá respuestas a las preguntas que se debieron hacer, teniendo en cuenta las respuestas anteriores. Por ejemplo, si se ingresa el valor 2 en una variable *cm1* (es decir, la mujer jamás ha dado a luz), la aplicación de ingreso de datos le pedirá a continuación un valor para la variable *ma1*, pasando por alto todas las variables pertenecientes a las mujeres que han dado a luz.

La segunda labor de la aplicación de ingreso de datos es minimizar los errores de ingreso de datos. La aplicación de ingreso de datos ejecuta esta labor llevando a cabo verificaciones a medida que se ingresan los datos. Si un valor ingresado para determinada pregunta se encuentra fuera del rango de valores del cuestionario o si se detecta alguna otra inconsistencia básica, la aplicación mostrará un mensaje de error y le pedirá al operador de ingreso de datos que resuelva la inconsistencia antes de seguir adelante. Las inconsistencias más complejas, cuya resolución reduciría considerablemente el ritmo del ingreso de datos, no se revisan durante el proceso de ingreso de datos y en cambio, se someten a verificación durante la edición secundaria.

## EL MENÚ DE INGRESO DE DATOS

El menú de ingreso de datos simplifica la labor del ingreso de datos. El menú de ingreso de datos es creado por la aplicación de CSPro *entry\_menu.bch* (no es necesario que usted modifique esta aplicación). El texto a continuación muestra el menú de ingreso de datos.



Las opciones *A*, *M* y *T* se usan únicamente si el operador de ingreso de datos está ingresando el archivo de datos principales. Las opciones *R*, *U* y *V* se usan únicamente si el operador de ingreso de datos está ingresando el archivo de verificación de datos. Las opciones *N* y *Q* se pueden usar en cualquier circunstancia.

La opción *A* permite que el operador de ingreso de datos agregue casos al archivo de datos principales, la opción *M* le permite modificar los casos existentes en el archivo principal de datos, y la opción *T* le permite copiar el archivo de datos principales que esté completo a un disquete para que pueda ser transferido al computador del supervisor. La opción *R* permite que el operador de ingreso de datos agregue casos en el archivo de verificación de datos, la opción *U* le permite modificar los casos existentes en el archivo de verificación de datos, y la opción *V* le permite copiar el archivo de verificación de datos que esté completo a un disquete para que pueda ser transferido al computador del supervisor.

La opción *N* le permite al operador de ingreso de datos cambiar el número de conglomerado para poder ingresar datos de otro conglomerado. La opción *Q* sale del menú de ingreso de datos.

## ESTRUCTURA DEL DIRECTORIO DE UN COMPUTADOR DE INGRESO DE DATOS

En un computador de ingreso de datos, todos los archivos y programas aparecen listados en el directorio *c:\mics\CSPro* o en uno de sus sub-directorios. Los sub-directorios se denominan *data*, *dicts*, *entry*, y *veri*. El directorio *data* contiene todos los archivos de datos principales que se han ingresado en el computador. El directorio *dicts* contiene todos los diccionarios de datos. El directorio *entry* contiene la aplicación de ingreso de datos y la aplicación que crea el menú de

ingreso de datos. El directorio *veri* contiene todos los archivos de verificación de datos que han sido ingresados en el computador.

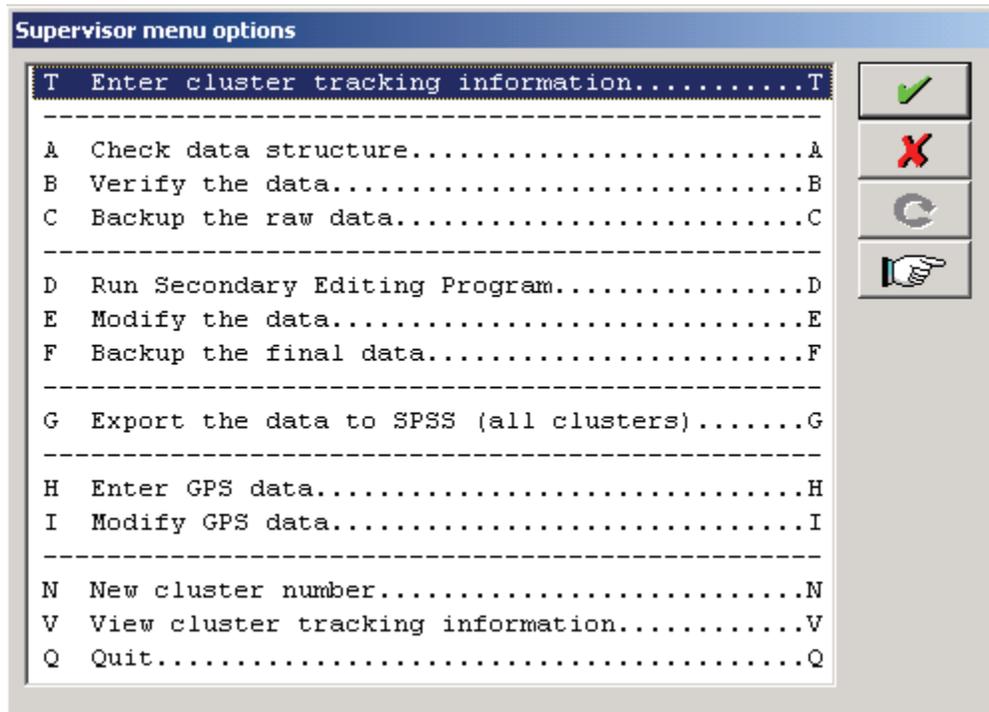
Estructura del directorio del computador de un operador de ingreso de datos:

```

MICS
  CSPRO
    DATA      Directorio de datos principales
    DICTS      Directorio de diccionarios de datos
    ENTRY      Directorio de aplicación de ingreso de datos
    VERI       Directorio de datos de verificación
    
```

### MENÚ DEL SUPERVISOR

El menú del supervisor le ayuda al supervisor de procesamiento de datos a manejar el sistema de procesamiento de datos MICS3. Se abre ejecutando la aplicación de CSPro *super\_menu.pff*. Primero, el menú pide el número de conglomerado a procesar. Tan pronto se ingresa el número de conglomerado, aparece el menú del supervisor mostrado abajo. Esta ilustración está seguida por un resumen de cada opción.



**OPCIÓN T. INGRESO DE INFORMACIÓN DE SEGUIMIENTO DE CONGLOMERADO**

Esta opción muestra el formulario electrónico de seguimiento de conglomerados para que el supervisor de procesamiento de datos pueda ingresar la información de seguimiento de conglomerados. La información se debe ingresar cuando los cuestionarios de un conglomerado apenas se reciben del campo, cuando el conglomerado se asigna a un operador para el ingreso de datos principales, y cuando se asigna el conglomerado a un operador para el ingreso de datos de verificación. La otra información del formulario de seguimiento electrónico de conglomerados se actualiza automáticamente a medida que el supervisor avanza a través del menú del supervisor.

**OPCIÓN A: VERIFICACIÓN DE LA ESTRUCTURA DE DATOS**

Esta opción verifica la estructura del archivo de datos del conglomerado al ejecutar la aplicación *check.bch*. Después de revisar la estructura del archivo de datos, *check.bch* produce un archivo de salida (que aparece automáticamente en la pantalla) que resume las cantidades de cada tipo de cuestionario y muestra el número de cuestionarios que fueron completados. El supervisor de procesamiento de datos debe comparar esta información con el formulario de seguimiento de conglomerados y asegurarse que ambas fuentes coinciden. En caso contrario, el supervisor de procesamiento de datos debe identificar el problema (p. ej., el operador de ingreso de datos olvidó ingresar un hogar) y resolverlo revisando cuidadosamente los cuestionarios del conglomerado. Tan pronto se resuelve el problema (sea que se actualice el archivo de datos o se actualice el formulario de seguimiento de conglomerados), el supervisor de procesamiento de datos debe volver a correr *check.bch*. Únicamente cuando *check.bch* muestre el mismo número de cuestionarios que el formulario de seguimiento de conglomerados, podrá el supervisor de procesamiento de datos asignar los cuestionarios a un segundo operador de ingreso de datos para el ingreso de los datos de verificación.

**OPCIÓN B: VERIFICAR LOS DATOS**

Esta opción compara el archivo de datos principales de un conglomerado con su archivo de verificación de datos usando la herramienta de comparación de CSPro. Si se presentan diferencias entre los archivos de datos, éstas aparecerán en la pantalla. Esta salida se debe imprimir y entregar a los operadores de ingreso de datos responsables del ingreso del conglomerado. Trabajando conjuntamente, los operadores de ingreso de datos deben consultar los cuestionarios y determinar el valor correcto de todo caso en el cual sus archivos de datos no coincidan. Una vez establecidos los valores correctos, cada operador deberá actualizar su archivo de datos. En ese punto, se deberán comparar nuevamente los archivos. Cuando ya no queden diferencias entre ambos archivos, se podrá seguir adelante con el procesamiento del conglomerado.

**C: RESPALDO DE LOS DATOS CRUDOS**

Esta opción hace un respaldo de los datos crudos copiando el archivo de datos principales verificados al subdirectorio *backup* en el computador del supervisor. Se debe correr después de completar las revisiones de la estructura y la verificación y antes de efectuar la edición secundaria. Los datos crudos se guardan para documentar la condición de los datos antes de ser editados.

**OPCIÓN D: CORRER EL PROGRAMA SECUNDARIO DE EDICIÓN**

Esta opción efectúa la revisión de inconsistencias complejas ejecutando la aplicación *editing.bch*. La salida de este programa aparece en pantalla y se debe imprimir si incluye algún mensaje de error. De ser así, la lista de inconsistencias y los cuestionarios del conglomerado se deben entregar a un editor secundario. El editor secundario, basado en su conocimiento del cuestionario y en el manual de edición en el Anexo 7, resolverá cada una de las inconsistencias de la lista. Una vez concluida la labor de los editores secundarios, éstos devuelven la lista de inconsistencias y de las acciones requeridas al supervisor de procesamiento de datos, quien implementará los cambios requeridos (ver opción *E* a continuación). El supervisor de procesamiento de datos vuelve a correr *editing.bch*. Si no aparecen más mensajes de error, se puede continuar con el procesamiento del conglomerado; si se presentan mensajes de error, se debe entregar la lista de los mismos y los cuestionarios al editor secundario para una edición adicional. En algunos casos, el editor secundario podría considerar aceptables los mensajes de error y ya no habrá necesidad de correcciones adicionales de los datos. Si los únicos mensajes que quedan son aquellos que fueron considerados aceptables por el editor, entonces se habrá completado el proceso de edición secundaria.

**OPCIÓN E. MODIFICAR LOS DATOS**

Esta opción ejecuta la aplicación *entry.ent* para que el supervisor de procesamiento de datos pueda implementar los cambios solicitados por los editores secundarios. Después de usar esta opción, el supervisor de procesamiento de datos debe regresar a la opción *D* para asegurarse que los cambios de los datos han corregido las inconsistencias como se esperaba y que no se han creado nuevas inconsistencias.

**OPCIÓN F. RESPALDO DE LOS DATOS FINALES**

Esta opción copia los archivos de los datos finales al subdirectorio *final* del computador del supervisor de procesamiento de datos. Los datos almacenados en este directorio se concatenarán más adelante y posteriormente se exportarán a SPSS.

**OPCIÓN G. EXPORTAR LOS DATOS A SPSS**

Esta opción concatena todos los archivos de datos del subdirectorio *final* en un solo archivo y luego lo exporta ejecutando *export.bch*. La aplicación produce cuatro archivos de texto ASCII y su correspondiente archivo de descripción SPSS para cada archivo de texto. Un archivo de texto contiene los hogares, otro contiene los miembros de hogares, otro contiene mujeres y otro contiene niños/niñas.

**OPCIÓN H. INGRESAR LOS DATOS DEL GPS**

Esta opción permite que el supervisor de procesamiento de datos ingrese los datos de ubicación geográfica ejecutando la aplicación *gpsentry.ent*. A diferencia del programa de ingreso de datos principales, esta aplicación permite que el supervisor de procesamiento de datos ingrese al mismo tiempo la cantidad de conglomerados que desee. La aplicación solicita que el supervisor de procesamiento de datos ingrese dos veces los datos del GPS para prevenir errores de digitación.

**OPCIÓN I. MODIFICAR LOS DATOS DEL GPS**

Esta opción permite que el supervisor de procesamiento de datos modifique la información de ubicación del GPS ejecutando la aplicación *gpsentry.ent*.

**OPCIÓN V. VER LA INFORMACIÓN DE SEGUIMIENTO DE CONGLOMERADO**

Esta opción muestra la información de todos los conglomerados almacenada en el formulario de seguimiento de conglomerados.

**OPCIÓN N. SELECCIONE UN NUEVO CONGLOMERADO**

Esta opción cambia el número de conglomerado de tal forma que el supervisor de procesamiento de datos pueda comenzar a procesar un nuevo conglomerado.

**ESTRUCTURA DEL DIRECTORIO DEL COMPUTADOR DE PROCESAMIENTO DE DATOS DEL SUPERVISOR**

El sistema de procesamiento de datos de MICS3 usa una estructura específica tanto en los computadores de ingreso de datos como en el computador del supervisor de procesamiento de datos. La estructura de los computadores de ingreso de datos se verá a continuación. En el computador del supervisor, todos los archivos y programas relacionados con SPSS están guardados en el directorio *c:\mics\spss* o en uno de sus subdirectorios. En el computador del supervisor, todos los archivos y programas relacionados con CSPro están guardados en el directorio *c:\mics\CSPro* o en uno de sus subdirectorios. Los subdirectorios se denominan *backup*, *dicts*, *entry*, *export*, *gps*, *final*, *raw*, *super* y *weights*.

Estructura del directorio del computador del supervisor:

MICS		
	CSPRO	
	BACKUP	Directorio de respaldo
	DICTS	Directorio de diccionarios de datos
	ENTRY	Directorio de la aplicación de ingreso de datos
	EXPORT	Directorio de aplicación de exportación
	FINAL	Directorio de datos editados finales
	GPS	Directorio de aplicación de ingreso de datos de GPS
	RAW	Directorio de datos crudos
	SUPER	Directorio de aplicación de edición del supervisor
	WEIGHTS	Directorio de aplicación de factor de ponderación de la muestra

El directorio *backup* contiene un respaldo de los archivos de datos que han sido revisados y verificados estructuralmente, pero no han sido editados. El directorio *dicts* contiene todos los diccionarios de datos. El directorio *entry* contiene la aplicación de ingreso de datos y la aplicación que crea el menú de ingreso de datos. El directorio *export* contiene los programas usados para exportar datos de CSPro a SPSS. El directorio *final* contiene el respaldo de los archivos de datos que han sido revisados, verificados y editados estructuralmente. El directorio *raw* contiene los archivos de datos que han sido transferidos de los equipos de ingreso de datos. El directorio *super* contiene las aplicaciones que ejecutan las revisiones estructurales, las verificaciones y la edición secundaria y la aplicación que crea el menú del supervisor. El directorio *weights* contiene la hoja de cálculo que calcula los factores de ponderación.

### REVISIÓN DE LA ESTRUCTURA

Es esencial que los datos tengan solidez estructural. El programa de ingreso de datos implementa la mayor parte de las verificaciones de consistencia estructural, pero no puede revisarlo todo sin verse limitado seriamente. En consecuencia, es necesario ejecutar un programa de revisión estructural después de completar el ingreso de datos principales. El programa de revisión de la estructura asegura que el número de cuestionarios del archivo de datos coincida con el número de cuestionarios que llegaron del campo y ejecuta algunas revisiones adicionales en la estructura de un cuestionario individual.

El programa de revisión de la estructura se denomina *check.bch*. Esta sección se enfocará en lo que hace el programa en vez de cómo lo hace. El programa es complejo y no permite explicaciones fáciles. La mejor forma de comprender la lógica en el programa es estudiarlo cuidadosamente después que haya entendido lo que hace el programa. A continuación se muestra un ejemplo de una salida crucial de *check.bch*.

```
MICS Data Structure Check
Cluster: 3
```

Households			Women				Children			
Total	Comp	Incomp	Eligible		Interviewed		Eligible		Interviewed	
			HH12	FOUND	HH13	FOUND	HH14	FOUND	HH15	FOUND
2	1	1	5	5	4	4	4	4	3	3

El primer bloque de salida muestra un resumen del número total de hogares y sus códigos de respuesta. El segundo bloque de salida muestra el resultado de cuatro conteos del número de mujeres elegibles. Las cifras de las columnas hh12 y hh13 corresponden a la cantidad de mujeres elegibles y entrevistadas de acuerdo con el Panel de Información del Hogar. Las cifras de las columnas *found* representan el número de cuestionarios de mujeres y de cuestionarios de mujeres completados, respectivamente, en el archivo de datos.

En el tercer bloque de salida muestra el resultado de cuatro conteos del número de niños/niñas menores de 5 años elegibles. Los conteos en las columnas hh14 y hh15 corresponden a la cantidad de niños/niñas elegibles y entrevistados según el Panel de Información del Hogar. Los conteos en las columnas *found* del representan el número de cuestionarios de niños/niñas y de cuestionarios de niños/niñas completados, respectivamente, en el archivo de datos.

El supervisor debe imprimir la salida de *check.bch* y comparar la información que contiene con el formulario de seguimiento de conglomerados. Si se presenta una diferencia entre las dos cifras de los cuestionarios, el supervisor y el operador de ingreso de datos deben usar el listado de errores y los cuestionarios del conglomerado para determinar la causa del problema estructural. Cuando esas causas se hayan identificado, deberán ser corregidas por el operador de ingreso de datos. A continuación se deberá volver a correr la aplicación de revisión de la estructura para verificar que el problema ha sido corregido sin crear un nuevo problema. Únicamente cuando las cifras generadas por *check.bch* coincidan con aquellas que se encuentran en el formulario de seguimiento de conglomerados se podrá comenzar el ingreso de los datos de verificación.

La aplicación *check.bch* produce también una lista de todos los hogares del conglomerado. Aparece en pantalla cada número de hogar y el código de resultado, junto con un recuento de los cuestionarios de mujeres y niños/niñas elegibles, si se completó la entrevista del hogar (es decir, el código de resultado de hogares es igual a 1). A continuación se muestra un ejemplo de salida de un hogar.

```
MICS Data Structure Check
Household: 1
Result: 1
```

Women				Children			
Eligible		Interviewed		Eligible		Interviewed	
HH12	FOUND	HH13	FOUND	HH14	FOUND	HH15	FOUND
4	4	3	3	2	2	1	1

El listado de hogares puede ser útil para la identificación de la fuente de un problema a nivel de conglomerado. Por ejemplo, suponga que la hoja de seguimiento de conglomerados muestra una

lista de 20 hogares en el conglomerado, pero solamente aparecen registrados 19 en el archivo de datos. Al comparar el listado de hogares con los cuestionarios del conglomerado, usted podrá identificar el hogar que no fue ingresado.

## VERIFICACIÓN

La verificación por doble digitación, se efectúa por medio de una aplicación de comparación de CSPro. La aplicación de comparación se denomina *compare.cmp* y contiene una lista de todas las variables (ítems). En la forma en que se encuentra configurado el programa actualmente, los ítems marcados se compararán durante la verificación y los ítems no marcados no se compararán durante ese proceso. Únicamente existe una variable que no está marcada (el código del operador de ingreso de datos, *hh16*) y se recomienda que *ninguna variable adicional* se mantenga sin marcar, ya que las diferencias que se pudieran presentar en otras variables podrían afectar la calidad de los datos.

La aplicación de comparación compara el archivo de ingreso de datos principales (que ha sido copiado al equipo del supervisor) con el archivo de ingreso de datos de verificación (que ha sido copiado a un disquete (o unidad de red, si se emplea una red)) y produce una lista de diferencias, en caso de que existan. Si no hay diferencias, el supervisor deberá hacer un respaldo de los datos sin procesar y continuar con la edición secundaria.

En caso de existir diferencias, la lista de las mismas se debe imprimir y entregar a los operadores de ingreso de datos. A continuación, los operadores de ingreso de datos usarán la lista de diferencias y los cuestionarios del conglomerado para revisar cada diferencia y registrar en la lista el archivo que debe ser corregido. Cuando se hayan investigado todas las diferencias, los operadores de ingreso de datos corregirán cualquier error en sus archivos, volverán a copiar los archivos de datos a sus disquetes (o a la red) y se compararán nuevamente los archivos. Este proceso deberá continuar hasta que los archivos sean idénticos.

## EDICIÓN SECUNDARIA

La experiencia ha demostrado que las inconsistencias simples se pueden identificar y corregir de manera práctica durante el ingreso de datos. Sin embargo, los errores más complejos de consistencia se deben resolver examinando cuidadosamente el cuestionario. Este tipo de revisión de consistencia se implementa mejor en forma separada, imprimiendo la lista de errores que a su vez, puede servir para marcar las correcciones. Este paso se conoce como edición secundaria.

El programa de edición secundaria de datos se denomina *editing.bch*. Ejecuta una larga lista de revisiones de consistencia (p. ej., si las edades y las fechas de nacimiento son coherentes) y produce una lista de inconsistencias encontradas en el archivo de datos crudos. El supervisor de procesamiento de datos debe imprimir esta lista y entregarla junto con los cuestionarios del conglomerado a uno de los editores secundarios. El editor secundario revisa entonces la lista de errores y las respuestas del cuestionario. Usando las directrices de edición (ubicadas en el Anexo

7) y su conocimiento del cuestionario, el editor secundario anota entonces una corrección en el listado de errores o anota que no es necesario tomar acción. Cuando el editor secundario ha revisado todos y cada uno de los mensajes de error, devuelve el listado con las anotaciones al supervisor de procesamiento de datos. A continuación, el supervisor de procesamiento de datos efectúa entonces los cambios sugeridos en los datos crudos. Una vez hecho esto, el supervisor de procesamiento de datos vuelve a correr el programa de edición. Únicamente cuando el programa de edición deje de producir mensajes de error se podrá continuar con el procesamiento del conglomerado.

Tres aspectos del proceso de edición son fundamentalmente importantes. Primero, en cada error, el editor secundario debe examinar cuidadosamente el cuestionario correspondiente. Segundo, el editor secundario siempre debe referirse a las directrices de edición antes de desarrollar una solución al problema. Tercero, el proceso de edición se debe repetir hasta que no queden más mensajes de error.

Cuando se haya concluido esta tercera labor, el supervisor de procesamiento de datos podrá hacer el respaldo de los datos editados. Ahora, se consideran limpios y se pueden usar para construir los archivos de análisis.

### EXPORTAR LOS DATOS A SPSS

Una vez concluido el procesamiento de datos primarios, usted dispondrá de un archivo de datos limpios para cada conglomerado de su muestra. En tanto que el procesamiento primario de datos se efectúa empleando CSPro, el procesamiento secundario de datos se hace primordialmente en SPSS. Por lo tanto, el primer paso en el procesamiento secundario de datos es la conversión de los datos del formato de datos CSPro al formato de datos SPSS. Esto se lleva a cabo usando la opción de ‘*Export the data to SPSS*’ (‘Exportar datos a SPSS’) del menú del supervisor.

Cuando usted selecciona esta opción, todos los archivos de datos del subdirectorio *final* (es decir, todos los archivos de datos que han sido verificados, revisados y editados) se concatenan en un solo archivo de datos denominado *all.dat*. A continuación, este archivo de datos se exporta a SPSS por medio de la aplicación *export.bch*. Esta aplicación crea cuatro archivos de datos ASCII (*mych.dat*, *myhh.dat*, *myhl.dat* y *mywm.dat*) y cuatro archivos de descripción de SPSS (*mych.sps*, *myhh.sps*, *myhl.sps* y *mywm.sps*) en el directorio *c:\mics\spss*. Aunque los archivos de descripción de datos de SPSS podrán leer los archivos de datos ASCII, no los guardarán. Para lograr que los archivos de descripción de datos guarden la información en el formato SPSS, el comando

```
save outfile = `filename.sav`.
```

se debe agregar al final de cada archivo de descripción de datos. La palabra *filename* se debe reemplazar por *hh*, *hl*, *wm* o *ch*, dependiendo del tipo de archivo de datos. Una vez que este comando ha sido modificado adecuadamente y se ha agregado a cada archivo de descripción de

datos, al ejecutar los archivos de descripción de datos de SPSS se crearán los archivos de datos *hh.sav*, *hl.sav*, *wm.sav* y *ch.sav* de SPSS.

### CREAR UN ARCHIVO DE ANÁLISIS

La estructura del archivo de datos simplifica el proceso de ingreso de datos durante el procesamiento primario de datos. Sin embargo, esta estructura no es la mejor para el análisis de los datos recolectados, y por eso la primera labor después que los datos se han transferido a SPSS, es recodificar variables para facilitar el análisis y hacerlo más eficiente. Esta labor se conoce como la creación de un archivo de análisis. Esta sección presentará un detalle de los pasos necesarios para la creación de archivos de análisis MICS3. Los archivos de análisis que resultarán al seguir estos pasos pueden ser usados por los modelos de planes de tabulación y son apropiados para ser distribuidos a investigadores.

#### RECODIFICAR LAS VARIABLES

Los programas *makehl.sps*, *makewm.sps* y *makech.sps* de SPSS recodifican las variables existentes para crear nuevas. Variables que se emplean en varias tabulaciones se recodifican en estos programas y luego se guardan. La codificación restante se efectúa en los programas de tabulación y es temporal.

La recodificación de la mayoría de las variables usa comandos estándar de SPSS y no se abordará aquí. Sin embargo, existe un enfoque usado frecuentemente que requiere explicación: la recodificación de las variables a 0 o 100. Esta recodificación inusual se hace únicamente con fines de presentación. Cuando SPSS muestra porcentajes en una tabla, presenta todas las categorías de la variable. Para muchas de las tablas del plan de tabulación, solamente estamos interesados en una sola categoría. Si le asignamos un valor de 100 a esa categoría y un valor de 0 a todas las demás categorías, la media de la variable será el porcentaje de encuestados en esa categoría. En consecuencia, pedirle a SPSS que muestre el promedio de la nueva variable nos dará únicamente el porcentaje que nos interesa ver.

Por ejemplo, la variable *hal* registra si una mujer ha escuchado sobre el SIDA. Asume un valor de 1 si la mujer ha escuchado sobre el SIDA y un valor de 2 si no ha sido así. Estamos interesados en mostrar el porcentaje de mujeres que ha escuchado sobre el SIDA. En el programa *makewm.sps*, la variable *hal* se recodifica en la variable *aids*. La variable *aids* asume el valor de 100 si la mujer ha escuchado hablar del SIDA y un valor de 0 si no ha sido así. La media de la variable *aids* es el porcentaje de mujeres que han escuchado sobre el SIDA. Para comprender por qué es así, tenga en cuenta el ejemplo a continuación:

Mujeres que han escuchado sobre el SIDA	10
Número total de mujeres	20
Porcentaje de mujeres que han escuchado sobre el SIDA	$(10 / 20) * 100 = 50$
Media de la variable SIDA	$(10*100 + 10*0)/20 = 10*100/20 = (10/20)*100 = 50$

### CALCULAR Y AGREGAR LOS FACTORES DE PONDERACIÓN

Si se usaron marcos muestrales independientes para las distintas regiones (o dominios) en la primera etapa del muestreo, la muestra nacional no se seleccionó con probabilidad proporcional al tamaño. Esto podría suceder también si usted estratificó de acuerdo con algún otro factor (p. ej., urbano/rural o tugurio/no tugurio) y si tomó diferentes fracciones (proporciones) de la muestra en cada estrato. Estas muestras no son auto-ponderadas, y usted debe ponderar su muestra cuando reporta estimaciones nacionales. Es decir, usted debe asegurarse que cada sub-muestra – por ejemplo, cada región (o dominio) – contribuya solamente lo que contribuiría en el caso de que la muestra de la encuesta a nivel nacional se hubiese seleccionado con probabilidad proporcional a tamaño.

Si su muestra no es auto-ponderada, usted debe calcular los factores de ponderación y agregarlos a sus archivos de análisis. Esta tarea se logra usando la hoja de cálculo *weights.xls* y los programas *weights\_table.sps*, *weights.sps* y *weights\_merge.sps* de SPSS. La hoja de cálculo se usa para calcular los factores de ponderación. Tiene dos hojas de cálculo: *calculations* y *output*. La hoja de cálculo *calculations* lleva a cabo los cálculos. La hoja de cálculo *output* contiene solamente los factores de ponderación y una lista con los números de conglomerados; este formato es útil para leer los datos contenidos a SPSS. El programa *weights\_table.sps* produce los datos necesarios para calcular los factores de ponderación. El programa *weights\_merge.sps* agrega los factores de ponderación apropiados a los archivos de análisis. El programa *weights.sps*, el cual usted jamás ejecutará directamente, describe la estructura de los datos en la hoja de cálculo de salida.

El proceso para calcular los factores de ponderación y agregarlos a sus archivos de análisis se puede dividir en siete pasos:

**Paso 1:** Ajuste el número de filas en las hojas de cálculo de *calculations* y *output* de tal forma que tengan en una fila por conglomerado en su encuesta. Después de agregar o borrar filas, asegúrese que al hacerlo no afectó la fila de totales de la hoja de cálculo de *calculations*.

**Paso 2:** Ingrese en *weights.xls* las ponderaciones del diseño de la muestra. Si sus ponderaciones varían entre conglomerados dentro de un estrato o dominio específico, debe completar ambas, la columna de fracción de la muestra del conglomerado y la columna de fracción de la muestra del estrato (o dominio) con la información provista por el experto en muestreo de la encuesta. Si las ponderaciones varían entre estratos (o dominios) pero entre conglomerados dentro de los estratos (es decir, la muestra es auto-ponderada dentro de los estratos o dominios), ingrese el valor 1 en la

columna de fracción de la muestra de conglomerado y complete la columna de fracción de la muestra del estrato (o dominio) usando la información provista por el experto en muestreo de la encuesta.

**Paso 3:** Actualice la definición de estratos (o dominios) en las líneas que van del 3 al 10 en el programa *weights\_table.sps*. Los programas estándar asumen que los estratos están compuestos por todas las combinaciones de área (es decir, urbana y rural) y región y que existen cuatro regiones (deberá modificar el programa para reflejar los estratos o dominios que use en su muestra).

**Paso 4:** Ejecute el programa *weights\_table.sps*.

**Paso 5:** Copie la información de la tabla y péguela en la hoja de cálculo *calculations* de *weights.xls*. Cuando haya completado este paso, *weights.xls* calculará automáticamente los factores de ponderación.

**Paso 6:** Guarde la hoja de cálculo *output* como un archivo de valor separado por comas (*comma-separated value*) (\*.csv) bajo el nombre de *weights.csv* en el directorio *c:\mics\weights*.

**Paso 7:** Ejecute el programa *weights\_merge.sps*. Cuando haya concluido el séptimo paso, asegúrese de revisar la lista de salida para ver si hay mensajes de error, y abra los archivos de análisis y confirme que las ponderaciones se han fusionado correctamente.

## CALCULAR Y AGREGAR EL ÍNDICE DE RIQUEZA

El plan de tabulaciones de MICS incluye como variable de fondo un índice de riqueza de hogares. Este índice de riqueza es calculado por el programa *wealth.sps*, que crea el archivo de datos *wealth.sav* que contiene variables de identificación, una variable con el nivel de riqueza de cada hogar y una variable con el índice de riqueza de cada hogar. El programa *wealth.sps* produce primero las frecuencias de todas las variables de los hogares relacionadas con riqueza o bienes. A continuación, recodifica las variables que describen los bienes del hogar y de individuos en variables dicotómicas. Luego, el programa usa análisis de factores (específicamente análisis de componentes principales) para calcular el grado de riqueza para cada hogar. Finalmente, usa el grado de riqueza para crear los quintiles de riqueza (es decir, el índice de riqueza) de los hogares, y luego los guarda en un archivo de datos de SPSS. La selección de las variables que se incluirán en el análisis de factores es fundamental, y no se debería hacer sin consultar cuidadosamente sus frecuencias. En [www.childinfo.org](http://www.childinfo.org) encontrará información sobre variables incluidas en el análisis. Una vez calculado el índice de riqueza, deberá ejecutar el programa *wealth\_merge.sps* que lo agregará a sus archivos de análisis. Asegúrese de revisar la lista de salida para verificar si hay mensajes de error y de abrir los archivos de análisis y confirmar que el índice de riqueza se ha fusionado correctamente.

### **AGREGAR LAS LECTURAS DEL *GPS***

Para el trabajo de campo, algunas encuestas de MICS3 tomarán las lecturas de los sistemas globales de posicionamiento (*GPS – global positioning systems*) para sus conglomerados. Las lecturas de GPS, que localizan los conglomerados con exactitud, se pueden usar para agregar otras bases de datos geográficos (p. ej., datos pluviométricos) a la base de datos de MICS3. Si su encuesta está usando las lecturas del GPS, usted deseará fusionarlas con sus archivos de análisis. Esta tarea se lleva a cabo por medio de dos programas de SPSS: *gps.sps* y *gps\_merge.sps*. El programa *gps.sps*, que usted jamás ejecutará directamente, describe la estructura del archivo de datos *gps.dat* (que se crea por medio de la aplicación de ingreso de datos *gpsentry.bch* de CSPro). Si usted ha cambiado el diccionario *gps.dic* de CSPro, deberá actualizar *gps.sps* para reflejar sus cambios. El programa *gps\_merge.sps* fusiona las lecturas de GPS con los archivos de análisis. Usted no debería verse en la necesidad de modificar este programa. Para fusionar las lecturas de GPS con sus archivos de análisis, ejecute *gps\_merge.sps*. Asegúrese de revisar la lista de salida para ver si hay mensajes de error y de abrir los archivos de análisis y confirmar que las lecturas del GPS se han fusionado correctamente.

### **TABULACIÓN**

El Capítulo 8 describe en detalle el proceso de análisis de datos y la elaboración de informes. Correr las tabulaciones es uno de los componentes más importantes de esta actividad. Un plan modelo de tabulación (Capítulo 8 y Anexo 6) y los programas de tabulación para SPSS acompañan este manual. Existe un programa de tabulación SPSS para cada tabla en el plan de tabulación de MICS3. El nombre de cada programa consta de la letra ‘t’ seguida por el número de la tabla correspondiente al plan de tabulación. Por ejemplo, el programa *t1.sps* crea la Tabla 1 del plan de tabulación.

Antes de correr las tabulaciones de un informe, es fundamental producir un conjunto de distribuciones de frecuencia (sin ponderar) para cada variable del archivo de datos. Las frecuencias se deben revisar para buscar valores inusuales, aquellos que se encuentran fuera del rango de la mayoría de las respuestas, y aquellos que pudieran ser respuestas improbables a la pregunta relevante. Por ejemplo, una respuesta de “53” a la pregunta sobre el número de horas que un niño trabajó en los quehaceres del hogar durante la semana anterior parece altamente improbable y demasiado precisa. La información de identificación en este tipo de casos se debe registrar por escrito y los valores del archivo de datos se deben comparar con los cuestionarios originales.

### **REVISAR LOS MODELOS DE PROGRAMAS**

Todo programa de tabulación debe ser revisado cuidadosamente. Es importante verificar si las variables que se usan en el programa de tabulación existen en su archivo de datos. En caso contrario, revise si la variable tiene una importancia primaria o secundaria. Si una variable de importancia primaria no existe en su archivo de datos, usted deberá quitar la tabla o pedirle a un

analista que la rediseñe. Si falta una variable de importancia secundaria, quite del programa de tabulación todas las referencias relacionadas con la variable y efectúe cualquier otro ajuste que sea necesario debido a su ausencia.

Toda actividad de recodificación también se debe revisar cuidadosamente. Si en su cuestionario existen variables que no tienen categorías estándar, cualquier actividad de recodificación que involucre a dichas variables debe ser examinada. Si su cuestionario contiene variables no estándar, éstas se deben recodificar para que pueda usarlas en una tabulación.

Usted también debe revisar cualquier operación de fusión si su cuestionario emplea identificadores de caso que no se encuentran en el cuestionario estándar. Existe un número de fusiones en los programas de tabulación que solamente funcionarán si se utilizan identificadores únicos.

### **APLICAR LOS FACTORES DE PONDERACIÓN DE LA MUESTRA**

La ponderación en los programas de tabulación es directa, excepto cuando involucra el comando *aggregate*. Si la meta del comando *aggregate* es acumular casos para calcular el numerador y el denominador, las ponderaciones se deben aplicar antes del comando *aggregate*. No se deben usar cuando esté trabajando con el archivo resultante; ya ha sido ponderado.

Por ejemplo, la tabla HH.1 en el Anexo 7 contiene la tasa de respuesta de hogares. La tasa de respuesta de hogares es difícil de calcular porque requiere la división de una variable por otra en la tabla. Una solución para este problema es crear un archivo agrupado que contiene el conteo de los hogares de la muestra, de los hogares ocupados y de los hogares entrevistados. El archivo agrupado va a contener un caso por cada categoría de la variable de quiebre especificada (p. ej., urbana/rural).

Las ponderaciones se deben aplicar cuando se crea el archivo agrupado para generar el numerador ponderado (el conteo de los hogares entrevistados) y el denominador ponderado (el conteo de hogares ocupados). Una vez creado el archivo agrupado, la tasa de respuesta de hogares por cada categoría de la variable de quiebre es el numerador dividido por el denominador.

Si la meta del comando *aggregate* es crear una estadística de resumen para casos individuales, las ponderaciones se deben aplicar después del comando *aggregate*. Por ejemplo, la Tabla HH.3 en el Anexo 7 contiene información sobre el porcentaje de hogares que tienen por lo menos un niño/niña menor de 15 años.

Esta información no está presente en el archivo de datos de hogares, pero se puede crear agrupando el archivo del listado de hogares. Las variables de quiebre comprenden el número de conglomerado y el número de hogar. Se aplican las ponderaciones después de la agrupación, porque estamos interesados en el porcentaje ponderado de los hogares con por lo menos un

niño/niña menor de 15 años, y no en el número ponderado de niños/niñas menores de 15 de cada hogar.

### EL COMANDO *INCLUDE*

El programa *tables.sps* de SPSS correrá todos los programas de tabulación juntos. Consta de una serie de comandos *include* de SPSS que ejecutan los programas de tabulación en forma individualizada. Si SPSS encuentra un error en el programa que se está incluyendo (es decir, se está ejecutando por un comando *include*), inmediatamente detendrá el programa de ejecución y regresará al programa que incluyó el programa (es decir, el programa que contenía el comando *include*).

Debido a esto, usted debería usar *tables.sps* solamente después de haber revisado, modificado y probado todos los programas individuales de tabulación. Asegúrese de eliminar cualquier comando *include* que ejecute un programa de tabulación que no esté usando.

El comando *include* le impone cuatro restricciones a los programas que ejecuta. La primera restricción es que cada comando debe comenzar en la primera columna del programa. Esta restricción parece limitar los espacios al inicio de las líneas del programa, pero a una línea se le puede incluir espacio si ésta comienza con el carácter '+'. Los comandos a continuación ilustran el uso del carácter '+':

```
do if (cage >= 6 and cage <= 9).
+ compute solids = 0.
+ if (BF3G = 1) solids = 100.
end if.
variable labels solids "Solid foods".
```

La segunda restricción impuesta por el comando *include* es que si el comando continúa a lo largo de líneas múltiples, la columna 1 de las líneas de continuación debe aparecer en blanco. El ejemplo a continuación ilustra un comando de líneas múltiples que respeta esta restricción.

```
add files
  /file=*
  /file='tmp6.sav'.
```

Observe que los sub-comandos de las líneas segunda y tercera aparecen con espacio en dos columnas (Aunque solamente es necesario que aparezcan con espacio en una columna para satisfacer la restricción, aparecen con espacio en las dos columnas para conservar la consistencia con el estilo de programación de MICS3).

Las restricciones tercera y cuarta que impone el comando *include* son que los cierres de comandos son opcionales y que un asterisco (\*) ubicado en la primera columna de una línea indica que corresponde a una línea de comentario. Ninguna de estas restricciones afecta nuestros programas de tabulación.

Adicionalmente a *tables.sps*, existe un programa de SPSS que automatiza la creación de archivos de análisis. Ese programa se denomina *CSPro.sps*. Ese programa se puede usar únicamente cuando todos los programas componentes han sido ejecutados y se ha comprobado que funcionan. Es útil para volver a crear archivos de análisis cuando se ha efectuado un cambio en uno de los programas de creación de archivos. Asegura que todos los programas de creación de archivos de análisis se ejecutarán en el orden correcto.

### ARCHIVO Y DISTRIBUCIÓN DE DATOS

Un componente importante – y con frecuencia olvidado – del procesamiento de datos es el de archivo y documentación de los archivos de datos. Adicionalmente, sea que los archivos de datos estén disponibles ampliamente o únicamente al interior de una sola institución, es imperativo establecer algunas directrices de distribución con bastante anticipación. Esos pasos – archivar, documentar y distribuir – requieren una inversión en tiempo y esfuerzo. Sin embargo, la inversión queda bien justificada debido a un número de razones:

- **Incremento de la relación costo-efectividad en la recolección de datos.** Recolectar datos en una encuesta es una actividad costosa y de trabajo exhaustivo. Para justificar esa inversión, los datos recolectados se deben aprovechar al máximo. Poner los archivos de datos a la disposición de otros investigadores incrementa la relación costo-efectividad de las actividades de la encuesta.
- **Incremento en la apropiación de los datos y la aceptación de los resultados por parte del país.** Cuando el archivo de datos se encuentra a la disposición de otros para su uso, el proceso de recolección de datos gana credibilidad. Así se demuestra que quienes hacen la labor de recolección de datos tienen confianza en sus hallazgos y la accesibilidad de otros investigadores a los archivos de datos significa que los resultados pueden ser replicados y verificados por otros.
- **La capacidad de analizar tendencias.** Con frecuencia los resultados publicados, correspondientes a encuestas diferentes, no permiten la comparación directa. Por ejemplo, un informe de encuesta podría definir a los informantes adultos como aquellos cuya edad es de 15 años o más mientras que otra podría definir a los adultos como aquellos de 18 años o más. Sin los archivos de datos, lo mejor que se podría hacer es una comparación imprecisa de los dos conjuntos de resultados. Sin embargo, cuando los archivos de datos de las dos encuestas se encuentran disponibles, con frecuencia los resultados se pueden volver a tabular de tal forma que se puedan comparar directamente, permitiendo que se saquen conclusiones respecto a las tendencias.
- **La capacidad de hacer comparaciones entre países o al interior de los mismos.** Con frecuencia es aleccionador comparar resultados entre países, sea dentro de una sub-región o entre regiones. Estas comparaciones facilitan la identificación de áreas donde es necesario hacer énfasis en un programa específico o donde los programas han sido

particularmente exitosos. Además, podría ser de utilidad comparar los resultados de diferentes encuestas en un mismo país. En ocasiones esto se hace para validar resultados inesperados (p. ej., cuando la mortalidad infantil es más baja de lo esperado) o para evaluar los efectos de una metodología específica de recolección de datos (p. ej., usar tarjetas de vacunación en lugar de información sobre vacunaciones provista por las madres). Para poder llevar a cabo estos tipos de análisis, los investigadores necesitan acceder a los archivos de datos con el fin de calcular cifras que sean directamente comparables.

- **Permite el análisis a profundidad de áreas temáticas importantes por parte de especialistas.** Debido a la presión por informar rápidamente los hallazgos, la información que se presenta en un informe de encuesta normalmente incluye únicamente los hallazgos básicos de la misma. Un archivo de datos bien documentado y disponible permitiría la conducción de análisis exhaustivos en áreas temáticas específicas y esos análisis podrían ser implementados por especialistas en el tema que pudieran no pertenecer al personal de la institución que recolectó los datos.

El archivo de análisis de MICS3 se debe archivar, documentar y distribuir. Por lo menos la documentación que acompaña los archivos de análisis debe incluir una copia del informe completo, una copia del cuestionario y una descripción del diseño de la muestra. La documentación deberá incluir también un libro de códigos que contenga la localización y descripción de cada variable del archivo de análisis (éste se puede crear fácilmente en SPSS). También se deben archivar y hacer disponibles por solicitud, copias de todos los programas y archivos usados durante el procesamiento de la encuesta. Una copia de los archivos de análisis y su documentación deben ser enviadas a la Oficina Regional de UNICEF y a UNICEF Nueva York (Sección de Estadísticas y Seguimiento). Finalmente, es necesario establecer una política y los procedimientos para la distribución del archivo de datos a otros interesados.

**Cuadro 7.7**  
**Listas de Control**

Antes del Trabajo de Campo:

- Obtener los computadores y otros equipos de procesamiento.
- Organizar la oficina o espacio para el procesamiento de datos
- Contratar al supervisor de procesamiento de datos y demás personal.
- Crear un sistema para organizar las actividades de procesamiento.
- Adaptar los programas para que sean coherentes con el cuestionario de la prueba preliminar.
- Ingresar y editar los cuestionarios de la prueba previa.
- Finalizar los programas basados en la experiencia de los cuestionarios de prueba y los finales.

Durante el Trabajo de Campo:

- Recibir los cuestionarios del campo.
- Asignar el ingreso principal de datos.
- Revisar la estructura del archivo de ingreso principal de datos.
- Asignar el ingreso de datos de verificación.
- Verificar que los archivos de datos principales y de verificación sean idénticos.
- Respaldo del archivo de datos crudos.
- Llevar a cabo la edición secundaria.
- Respaldo del archivo de datos finales.

Después del Trabajo de Campo:

- Exportar los datos a SPSS.
- Recodificar las variables.
- Calcular e incluir los factores de ponderación, el índice de riqueza y los datos del GPS.
- Correr los programas de tabulación.
- Archivar los datos y desarrollar una política y un sistema de distribución (p. ej., un sitio web).
- Enviar los archivos de análisis, su documentación y todos los programas a UNICEF.

**Cuadro 7.8**  
**Formulario de Seguimiento de Conglomerados**

Número de Conglomerado	Fecha de Recibido	Hogares Seleccionados	Número de cuestionarios			Ingreso de datos			Fecha Terminación de Revisión	Fecha de Terminación de Verificación	Fecha de Respaldo de Datos Crudos	Fecha de Edición	Fecha de Respaldo Final
			Hogares	Mujeres	Niños y niñas	Nombre del Operador	Número del Operador	Fecha de Creación					
						M:							
						V:							
						M:							
						V:							
						M:							
						V:							
						M:							
						V:							
						M:							
						V:							
						M:							
						V:							
						M:							
						V:							
						M:							
						V:							