

Recomendaciones Metodológicas para las Rondas de Seguimiento de la Encuesta Longitudinal de Protección Social (ELPS)

Joaquín Prieto, Lucía Madrigal, Francisca Gallegos y Emilio Bravo 2015



Recomendaciones Metodológicas para las Rondas de Seguimiento de la Encuesta Longitudinal de Protección Social (ELPS)

Joaquín Prieto¹²

17 de junio de 2015

Resumen: La preparación y levantamiento de las rondas de seguimientos de la ELPS en varios países de América Latina presenta varios desafíos para los responsables de su diseño e implementación. Este documento sugiere un set de productos y actividades para realizar entre cada levantamiento de datos. Al mismo tiempo entrega un conjunto de recomendaciones para evaluar la calidad de los datos recogidos en la primera ronda de la ELPS. Se hace especial énfasis en las recomendaciones para reducir el error de la no respuesta en las siguientes rondas, y minimizar los errores de medición propios de una encuesta panel. Además se hacen sugerencias para reducir el error total de la encuesta cuando la ELPS se aplica en cuestionarios impresos y no aparatos electrónicos. Por último, se entregan sugerencias para la diseminación y promoción de las bases de datos junto a su documentación. Las recomendaciones se basan en la experiencia de encuestas paneles en otros países e investigaciones que han estudiado los tipos de errores en encuestas longitudinales y apuntan a mejorar la calidad de datos en la ELPS.

¹ Ingeniero Civil Industrial y Magíster en Economía Ambiental de la Universidad Católica de Chile. Candidato a Doctor en Políticas Sociales en la London School of Economics UK y profesor asociado al Departamento de Sociología de la universidad Alberto Hurtado. Actualmente es profesor visitante en el Programa Internacional de Derechos Humanos, Boston College US. jjprieto@uahurtado.cl; +1 617 775 7690 ; 205 Walden Street 4R, Cambridge, MA, 2140, USA.

² Participaron en la elaboración del presente documento Lucía Madrigal, Francisca Gallegos y Emilio Bravo.

Contenidos

1. PRÓLOGO.....	3
2. INTRODUCCIÓN	4
3. RECOMENDACIONES PARA EVALUAR LA CALIDAD DE DATOS EN LA PRIMERA RONDA ...	5
3.1 EVALUACIÓN DEL ERROR TOTAL DE LA ENCUESTA	5
3.1.1 <i>Conceptos y definiciones</i>	5
3.1.2 <i>Lecturas sugeridas</i>	7
3.1.3 <i>Recomendaciones específicas</i>	7
3.2 USO DE LA PARADATA PARA MEJORAR CALIDAD DE DATOS DE LA ELPS	8
3.2.1 <i>Conceptos y definiciones</i>	8
3.2.2 <i>Lectura sugerida</i>	9
3.2.3 <i>Recomendaciones específicas</i>	9
3.3 PROPUESTA DE EVALUACIÓN GENERAL DE CALIDAD DE DATOS DE LA PRIMERA RONDA DE LA ELPS.....	10
3.3.1 <i>Etapas críticas del trabajo de campo de la primera ronda de la ELPS</i>	10
3.3.2 <i>Procedimiento de evaluación de la calidad de datos para las etapas críticas</i>	22
4. RECOMENDACIONES PARA REDUCIR EL ERROR DE NO RESPUESTA EN LAS RONDAS DE SEGUIMIENTO	32
4.1 LA ATRICIÓN DE LA MUESTRA	32
4.1.1 <i>Conceptos y definiciones</i>	32
4.1.2 <i>Determinantes de la no respuesta y la atrición</i>	32
4.1.3 <i>Estrategias de seguimiento de las personas que se mueven en una encuesta panel</i>	34
4.1.4 <i>Lecturas sugeridas</i>	34
4.1.5 <i>Recomendaciones específicas</i>	35
4.2 PROPUESTA PARA EL SEGUIMIENTO DE LA MUESTRA EN LAS PRÓXIMAS RONDAS	36
5. RECOMENDACIONES PARA REDUCIR EL ERROR DE LA MEDICIÓN EN LAS RONDAS DE SEGUIMIENTO	39
5.1 EL ERROR DE MEDICIÓN EN ENCUESTAS PANEL	39
5.2 RECOMENDACIÓN PARA REDUCIR EL EFECTO MEMORIA EN ENCUESTAS PANEL.....	39
5.2.1 <i>Uso de entrevistas dependientes</i>	39
5.2.2 <i>Uso de calendario de la historia de los eventos</i>	41
5.3 PROPUESTA DE CAMBIOS PARA EL CUESTIONARIO DE SEGUIMIENTO	43
6. RECOMENDACIONES PARA REDUCIR EL ERROR TOTAL DE LA ELPS EN LAS RONDAS DE SEGUIMIENTO EN MODO PAPI.....	51
6.1 FLUJO DE INFORMACIÓN PARA LA IMPLEMENTACIÓN DE LAS CINCO SUGERENCIAS EN LA ELPS	51
6.2 SUGERENCIAS PARA REDUCIR EL ERROR DE LA ENCUESTA REALIZADA EN MODO PAPI	56
6.2.1 <i>Supervisión de las capacitación de los encuestadores.</i>	56
6.2.2 <i>Revisión y validación/codificación de las encuestas realizadas</i>	59
6.2.3 <i>Elaboración y análisis de la paradata</i>	63
6.2.4 <i>Digitación de las encuestas con primer nivel de revisión consistencia</i>	63
6.2.5 <i>Análisis de consistencia de todas las variables en las bases de datos</i>	64
7. RECOMENDACIONES PARA LA DISEMINACIÓN Y PROMOCIÓN DE LAS BASES DE DATOS Y DOCUMENTACIÓN DE LA ELPS	65
7.1 DOCUMENTACIÓN REQUERIDA EN UN ESTUDIO PANEL.....	66
7.2 SITIO WEB DE UN ESTUDIO PANEL.....	67
7.3 SEMINARIO Y TALLERES	68
7.4 EXPERIENCIA EXITORA DE DISEMINACIÓN Y PROMOCIÓN DE DATOS LONGITUDINALES	68
7.5 HOMOGENIZACIÓN Y HOMOLOGACIÓN DE LAS BASES DE DATOS DE LA ELPS ENTRE PAÍSES.....	71
8. REFERENCIAS.....	72

1. Prólogo

Las encuestas longitudinales son instrumentos dinámicos que buscan capturar la realidad actual de las personas. Para mantener la relevancia de éstos es necesario tener en cuenta los cambios en el contexto, las realidades sociales, el avance en la teoría, o lo último en metodología de la investigación. Dado esto, actualizar los cuestionarios es esencial, sin embargo, adaptar las preguntas sin perder la capacidad de hacer comparaciones significativas entre ellas, a través de los años es un gran desafío.

No existe un único documento en la literatura que abarque todos los distintos aspectos a considerar cuando se preparan los cuestionarios de seguimiento (segundas rondas y en adelante). Sin embargo, existe mucha y distinta literatura que discute los distintos temas importantes a tomar en cuenta en este proceso. El desgaste de la muestra a través de los años, los motivos aceptables para hacer cambios a las preguntas y para añadir preguntas adicionales, el tiempo de aplicación de la encuesta, la identificación de preguntas complicadas para los entrevistados, el periodo óptimo para preguntar por la historia de las personas, entre otros temas a considerar.

Los temas anteriormente mencionados son importantes a considerar cuando se prepara un cuestionario de seguimiento. Sin embargo, evitar lo más posible el desgaste de la muestra entre rondas es el tema al que se le debe de prestar mayor importancia. La mayoría de las encuestas longitudinales pierden muchos individuos entre olas y esto tiene efectos en el poder de predicción de las variables y provoca que la encuesta sea poco útil para sacar conclusiones válidas. Para evitar el desgaste es crucial contar con buenas estrategias de seguimiento a los individuos. En los países en desarrollo un problema grave es la movilidad de las personas por trabajo o en busca de oportunidades de éste.

Este documento plantea algunos de estos temas y presenta algunos lineamientos generales que permiten guiar a los equipos de los países ELPS en la preparación de sus siguientes rondas. Este documento es una guía general que no es exhaustiva y la idea es que sirva de base para la discusión de los temas que surjan por lo que es un documento de trabajo.

2. Introducción

La preparación y levantamiento de las rondas de seguimientos de la ELPS presenta varios desafíos para los responsables de su diseño e implementación. Este documento sugiere una serie de recomendaciones en términos de productos y actividades para aquellos países que ya han realizado la primera ronda de la ELPS y están planificando las próximas rondas de seguimiento. Las recomendaciones se basan en la experiencia de encuestas paneles en otros países e investigaciones que han estudiado los tipos de errores en encuestas longitudinales y apuntan a mejorar la calidad de datos en la ELPS.

De esta manera, las recomendaciones se han agrupado en cinco grandes temas. Cada tópico corresponde a un capítulo del documento. En el primer capítulo se hace una serie de recomendaciones para evaluar la calidad de los datos recogidos en la primera ronda de la ELPS. La propuesta se basa en la perspectiva de error total de la encuesta y el uso de la paradata recogida durante el trabajo de campo de la línea de base. El segundo y tercer capítulo se focalizan en entregar sugerencias para reducir el error de la no respuesta en las siguientes rondas de la ELPS, y minimizar los errores de medición propios de diseñar el cuestionario de seguimiento en las encuesta panel. Se incluye un capítulo de sugerencias para reducir el error total de la encuesta cuando la ELPS se aplica en cuestionarios impresos y no en aparatos electrónicos como notebooks o tablets. El último capítulo, entrega recomendaciones para que los países logren un mayor impacto en la diseminación y promoción de las bases de datos junto a su documentación.

En cada capítulo se repite el mismo formato. En primer lugar se entregan antecedentes generales y se exponen conceptos y definiciones que buscan justificar de algún modo o hacer sentido a las sugerencias. Posteriormente, se entregan recomendaciones bibliográficas que profundizan en los temas abarcados. Después se listan recomendaciones específicas o de desarrolla en detalle una recomendación que contiene varias sugerencias integradas entre sí.

Es posible que varias recomendaciones se hayan desarrollado o se estén aplicando en algunos países. En otras ocasiones pueden resultar novedosas para los equipos técnicos responsables de la ELPS. La relevancia de este documento radica en poder visualizar ciertos aspectos técnicos del desarrollo de encuestas longitudinales con foco en la seguridad social. El objetivo final es poder traducir estas sugerencias en propuestas acotadas y concretas que permitan mejorar la calidad de datos de la ELPS y facilitar al usuario y analista de las bases de datos un trabajo más eficiente.

El prólogo del documento junto a las secciones 4.2 "Propuesta para el seguimiento de la muestra en las próximas rondas" y 5.3 "Propuesta de cambios para el cuestionario de seguimientos" fueron escritas por Lucía Madrigal. Las secciones 3.3 "Propuesta de evaluación general de calidad de datos en la primera ronda de la ELPS" y el capítulo 6 "Recomendaciones para reducir el error total de la ELPS en las rondas de seguimiento en modo PAPI" fueron desarrolladas por Joaquín Prieto, Francisca Gallegos y Emilio Bravo. Las otras secciones fueron escritas por el autor principal de este documento.

3. Recomendaciones para evaluar la calidad de datos en la primera ronda

Las sugerencias presentadas en este capítulo se basan en dos temas desarrollados en la metodología de encuestas durante los últimos años: la evaluación del error total de la encuesta, y el uso de la paradata para la calidad de datos. A partir de estas dos perspectivas se proponen una serie de recomendaciones específicas para aplicar en los países donde ya se ha terminado de aplicar la primera ronda. Se incluye además una propuesta de análisis de la calidad de datos de la línea base a partir de la identificación de trece etapas críticas del levantamiento de los datos en términos de la contribución al error total de la encuesta.

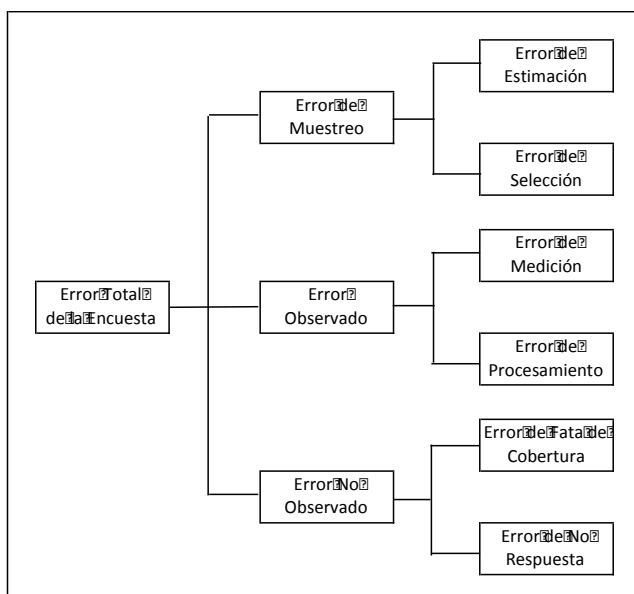
3.1 EVALUACIÓN DEL ERROR TOTAL DE LA ENCUESTA

3.1.1 Conceptos y definiciones

El error de cualquier encuesta es la diferencia entre el resultado obtenido en una pregunta determinada y su verdadero valor. La calidad de una encuesta tiene relación con los errores de los datos. En términos formales esto significa que la calidad del estimador de un parámetro de la población en estudio está en función de esos errores. Mientras más grande sea el cúmulo de errores de los datos existirá una mayor probabilidad que los parámetros estimados de la población estén distorsionados. Todos los fenómenos que causan estas distorsiones se denominan fuentes de error. Las principales fuentes de error en una encuesta son la no respuesta, la cobertura del marco muestral, el diseño del cuestionario, el encuestador, el modo de recolección de los datos, el entrevistado, y el procesamiento de datos, entre otras cosas (Biemer & Lyberg, 2003).

El impacto que tienen todas las fuentes de error en las estimaciones es conocido como el error total de la encuesta (Groves, 1989). El desafío que enfrenta cualquier agencia para levantar información de una población determinada es minimizar el error total de la encuesta dentro del presupuesto disponible para el proyecto. De esta manera la tensión en cualquier diseño e implementación de una encuesta es encontrar el equilibrio entre el presupuesto y un error total que permita obtener resultados precisos y confiables.

Figura 1: Componentes del Error Total de la Encuesta basados en Bethlehem (1999).



Lamentablemente, no hay ninguna encuesta que no esté sujeta a errores, y estos provienen de distintas causas. El primer paso para reducir esos errores es comprender su origen para cada uno de ellos. Basados en el trabajo de Bethlehem (1999) se identifican en la Figura 1 seis tipos de errores en una encuesta. Estos errores pueden agruparse en: i) *errores de muestreo*, ii) *errores observados*, iii) y *errores no observados*.

Los *errores de muestreo* son introducidos por el diseño de muestreo. Estos se deben al hecho que las estimaciones son basadas en una muestra de la población total. Los errores de muestreo desaparecerían si toda la muestra fuera observada (e.g. es el caso de los censos). Los errores de muestreo pueden dividirse en error de selección y error de estimación. El *error de estimación* da cuenta del efecto causado por usar una muestra probabilística. Cada nueva selección aleatoria de la muestra resultará en un elemento diferente y así en un diferente valor del estimador. Es posible controlar el error de estimación a través del diseño de muestreo. Por ejemplo, el error de estimación puede ser reducido aumentando el tamaño de la muestra o tomando una selección de probabilidad proporcional a los valores de una variable auxiliar conocida. El *error de selección* ocurre cuando existe un error en la probabilidad de selección que es usada para calcular el estimador. Por ejemplo, la verdadera probabilidad de selección de un miembro del hogar puede diferir si el encuestador aplica de forma equivocada el procedimiento de selección de la persona a encuestar (uso de Tabla de Kish en el caso de la ELPS en algunos países).

Los *errores observados* refieren a los errores cometidos durante el proceso de obtención y captura de las respuestas del entrevistado. Los errores observados pueden dividirse en dos. El *error de medición* y el *error de procesamiento*. Un error de medición ocurre cuando los entrevistados no entienden una pregunta o no quieren entregar la respuesta verdadera, o si el encuestador comete un error al registrar la respuesta. Además, se consideran como partes de este tipo de error el efecto del entrevistador, el efecto del fraseo de las preguntas y el efecto de la memoria. En otras palabras, el error de medición causa una diferencia entre el verdadero valor y el valor procesado en la encuesta. Un error de procesamiento es un error que aparece durante el procesamiento de datos. Por ejemplo, en el momento que un digitador digita un valor erróneo en un computador del valor que está registrado en una encuesta realizada en papel.

Los *errores no observados* aparecen porque las mediciones no pueden realizarse. Los errores no observados se dividen en dos. El *error por falta de cobertura* y el *error de no respuesta*. El error por falta de cobertura ocurre cuando elementos de la población objetivo no están presente en el marco muestral. Por ejemplo, si el marco muestral es el censo de un país, y éste no ha sido actualizado en varios años, las viviendas que fueron construidas durante el periodo posterior al censo no tienen probabilidad de ser seleccionadas. El error de no respuesta surge cuando los elementos seleccionados en la muestra no entregan la información requerida. Por ejemplo, los casos donde no se logra contacto con la persona a entrevistar o la persona se niega a responder la encuesta. También se considera como error de no respuesta las preguntas o módulos que no fueron contestados. Por ejemplo, en los casos que la entrevista fue interrumpida y no se logra terminarla.

Aunque se tomen medidas para reducir los errores en cada etapa del diseño e implementación de la encuesta siempre ocurrirán errores. Por lo tanto, es importante revisar los errores en los datos recolectados, y cuando sea posible, corregir dichos errores. Una recomendable manera de realizar este trabajo es a través de una evaluación del error de la encuesta (Biemer & Lyberg, 2003). El objetivo de una evaluación del error de la encuesta es evaluar, comprender, y arreglar - de ser posible - los errores provenientes de las fuentes de error. Los errores de medición y procesamiento, y en algunos casos específicos el error de selección son posible detectar y corregir. Este proceso de corrección de errores se conoce con el nombre de *edición de datos*. Los fenómenos como falta de cobertura y no respuesta requieren de una solución diferente relacionada con el uso de pesos ajustados en el cálculo de las estimaciones.

Se desprenden entonces de esta breve presentación de conceptos y definiciones, que es importante para los diseñadores de la ELPS y los potenciales usuarios de las bases de datos conocer la contribución de cada fuente de error en el error total de la encuesta.

3.1.2 Lecturas sugeridas

- Si se desea profundizar un poco más en la relación entre los errores y costos de una encuesta se sugiere leer *Survey Errors and Survey Costs* de Robert Groves (1989). En este libro se entregan estrategias para buscar el óptimo de un problema dual: ajustar el presupuesto a los objetivos de la investigación o ajustar la precisión de las mediciones a un presupuesto fijo.
- Biemer y Lyberg (2003) en su libro *Introduction to Survey Quality* proponen varios métodos de evaluación para los distintos tipos de errores. Además los autores realizan una pormenorizada discusión sobre medidas de calidad de una encuesta.
- Una clara y precisa sistematización de la detección y corrección de errores de una encuesta se encuentra en el libro *Applied Survey Methods: A Statistical Perspective* de Bethlehem (2009). Se recomienda revisar el subcapítulo 8.4 *Data Editing Strategies*, específicamente las estrategia de Macroedición de los datos.

3.1.3 Recomendaciones específicas

Recomendación 1: Elaborar tablas que entreguen los *errores de estimación* del diseño muestral utilizado las variables más relevantes en poblaciones de interés de la ELPS.

Recomendación 2: Generar un do-file en Stata que chequee si la selección de la persona en cada vivienda está correcta. El objetivo sería confirmar que el *error de selección* en esa tercera etapa³ del diseño muestral es cercano a cero.

Recomendación 3: Realizar un informe que dé cuenta de los *errores de medición* detectados durante el desarrollo del trabajo de campo y el procesamiento de datos. Se recomienda elaborar dicho informe a partir de entrevistas o grupos focales a los encuestadores, jefes de terreno, revisores de encuestas, digitadores y programadores de las mallas de validación - rutinas que detectan los errores de preguntas no contestadas, los errores de dominio, los errores de consistencia y los errores de ruta (*skip pattern errors*) -. Además se sugiere incluir un análisis del comportamiento de cada una de las observaciones que contienen las bases de datos en la distribución de todas las observaciones.

Recomendación 4: Si las encuestas realizadas en papel fueron digitadas solo una sola vez, se sugiere realizar una muestra aleatoria de 400 encuestas y volver a digitalarlas. La base de datos debiera compararse con la base de datos de la primera digitación. Las diferencias debieran revisarse en la encuesta en papel e informar el porcentaje de *errores de procesamiento* encontrados por encuesta.

Recomendación 5: Se sugiere elaborar un informe que acompañe el informe del diseño muestral que dé cuenta de las características del marco muestral utilizado en el diseño muestral. Se debieran incluir una descripción detallada - con la información disponible - de los crecimientos urbanos que no han sido actualizado en el marco muestral que fue utilizado con el fin de dimensionar el *error de falta de cobertura*.

³ La primera y segunda etapa corresponde a la selección de los sectores censales y la selección de las viviendas.

Recomendación 6: Realizar un análisis del *error de la no respuesta* de los casos donde después de aplicar la Tabla de Kish para seleccionar a la persona de la vivienda, no se pudo realizar la encuesta en la primera ronda. Se sugiere utilizar la información registrada en la matriz de selección aleatoria de la persona (edad, sexo y status laboral) y las razones que se recogieron en el Formulario de Contacto para explicar la no respuesta.

3.2 USO DE LA PARADATA PARA MEJORAR CALIDAD DE DATOS DE LA ELPS.

3.2.1 Conceptos y definiciones

La paradata es información sobre el proceso de levantamiento de datos. Se considera parte de la paradata las observaciones del encuestador, el día de la semana y horario de las visitas para encontrar al entrevistado, las distancias recorridas por el encuestador y sus gastos asociado a la movilización, o cualquier información registrada durante el trabajo de campo que pueda procesarse en una base de datos para ser analizada. La recolección de información auxiliar durante el proceso de recolección de los datos de una encuesta puede ayudar a los investigadores a entender y reducir el error total de la encuesta. La paradata puede informar a los investigadores sobre los distintos tipos de errores durante distintas etapas de la recolección de datos. En algunos casos, esta información pueden indicar problemas que pueden resolverse durante el trabajo de campo, otras veces solo es posible detectar los errores en el procesamiento de los datos, y en otro casos, la paradata contiene la información necesaria para modelar los errores junto a las bases de datos de la respuesta de la encuesta (Kreuter & Casas-Cordero, 2010)

Algunos ejemplos de la paradata que se generan durante diferentes etapas de un trabajo de campo son el movimiento de un empadronador - registrado a través de un GPS - durante el listamiento de viviendas de un sector censal y selección de la vivienda o las hojas de papel donde el encuestador listó a los miembros del hogar junto a la Tabla de Kish que utilizó para seleccionar a quien entrevistar (Eckman, 2013). Otros ejemplos de la paradata son los kilómetros recorridos de un encuestador durante un día de trabajo obtenidos en el momento de reembolsar los gastos de movilización (Wagner & Olson, 2001), o el Formulario de Contacto con la información de los intentos de contacto y las observaciones del barrio y de la vivienda de quienes responden o no la encuesta (Kreuter & Kohler, 2009). En el caso de encuestas realizadas en computador o *tablet* es posible capturar de manera automática y sin desfase otros datos auxiliares como el tiempo de aplicación entre preguntas, información geoespacial, correcciones realizadas por el encuestador, observaciones indicadas por el entrevistador, entre otra información recogida por el sistema de captura durante el proceso de encuestaje (Kreuter et al., 2010). En las encuestas realizadas en papel y lápiz esta información debe ser procesada después de ser aplicada la encuesta o en algunas situaciones no puede obtenerse.

En general, la paradata y la metadata se tratan conjuntamente. Por esta razón es importante comprender la diferencia entre ambos tipos de datos⁴. Kreuter (2013) utiliza una analogía con la

⁴ Un elemento que colabora para la confusión de ambas definiciones en la metodología de encuestas es que la misma fuente puede generar la paradata y la metadata. Por ejemplo, los archivos que almacenan la historia de contacto de un miembro de la muestra es la paradata. El día, la hora y el resultado - razones no logro de la encuesta - son paradata del proceso de contacto. Pero las mismas bases de datos también son usadas para calcular las tasas de respuesta de las encuestas realizadas. Esta información que tradicionalmente se entrega en la documentación de las bases de datos se denomina la metadata. Otros ejemplos de la metadata son el Libro de Código que detalla las variables creadas, editadas, y codificadas de la base de datos final. En general, la documentación de la calidad de datos de la encuesta debiera incluir la metadata.

fotografía digital para explicar la diferencia entre la metadata y la paradata. Al sacar una foto digital, la fecha y hora en que fue tomada es guardada en algún archivo de la cámara. De la misma manera es posible almacenar el tipo de lente que se usó, la apertura del diafragma y el tiempo de exposición de la fotografía tomada. Este tipo de información sería la metadata o los datos sobre los datos. En cambio, a la paradata corresponden a los datos sobre el proceso de generación del producto final. En el ejemplo de la fotografía digital dato por Kreuter (2013), la paradata serían los tipos de lentes y diferentes ángulos que el fotógrafo probó antes de sacar la fotografía. Incluso, se podría incluir en la paradata las palabras que usó el fotógrafo para conseguir que la persona aparezca sonriendo en la foto.

En general, las agencias que realizan el levantamiento de encuestas generan la paradata para monitorear el trabajo de campo. Pero no siempre la paradata es procesada de manera tal que permite informar estadísticas de evaluación, seguimiento y gestión de la recolección. Esta información es bienvenida con el fin de mejorar el desarrollo de las siguientes rondas del panel. Incluso puede llegar a ayudar a que la logística del terreno sea más eficiente, provocando por cierto un importante ahorro en los costos. Por último, la paradata es usada para mejorar las estimaciones de la encuesta, particularmente en los casos de no respuesta (Kreuter, 2013).

3.2.2 Lectura sugerida

Kreuter (2013) en su libro *Improving surveys with paradata: Analytic uses of process information*. recoge una serie de estudios que ha utilizado la paradata en cada uno de sus tres beneficios: i) mejorar la calidad de datos a través del monitoreo del trabajo de campo, ii) diseñar un trabajo de campo más eficiente reduciendo costos, y iii) mejorar las estimaciones de la encuesta.

3.2.3 Recomendaciones específicas

Recomendación 7: Identificar y listar todas las fuentes de la paradata del trabajo de campo de cada ronda. Algunos ejemplos de la paradata de la ELPS podrían ser: i) la papelería donde se registra los usos de viviendas de reemplazo y sus razones de hacerlo; ii) las hojas utilizadas para listar los miembros del hogar y hacer la selección utilizando la Tabla de Kish; iii) el día y hora de los intentos de contacto de la muestra y de la aplicación de la encuesta, iv) las razones de no contacto, la edad, sexo, nivel educacional y nivel de experiencia de los encuestadores, supervisores, y digitadores. v) Tiempos de llenado de la encuesta y de digitar en el computador las encuesta en papel. vii) Observaciones del encuestador, supervisores, y digitadores durante el trabajo de campo. Por ejemplo, las características de la vivienda y del barrio que los encuestadores registran al visitar las viviendas seleccionadas en la muestra panel.

Recomendación 8: En los casos donde la agencia responsable del trabajo de campo ya cuenta con la paradata procesada revisar que las bases de datos contengan una identificador o "llave" que permita unir las distintas bases de datos entre sí. Por ejemplo, una "llave" que permita unir las bases de datos de las encuestas hechas con las bases de datos de las características de los encuestadores, supervisores y digitadores. Otro ejemplo, es el caso de las encuestas no realizadas. Se debiera chequear que exista un identificador que permita unir la base de datos que contiene la historia de contacto con las bases de datos que guarda las observaciones que realizó el encuestador sobre la vivienda y el barrio. En los casos que la paradata no esté procesada, evaluar digitar la información o parte de ella. Por ejemplo, se podría crear una base de datos de todas las hojas donde el encuestador modificó - ya sea borrando o rayando sobre el papel - el orden de listamiento de los miembros del hogar o el número total de integrantes en el momento de la selección del entrevistado.

Recomendación 9: De disponer de la base de datos que tiene la historia de contacto de las encuestas realizadas en la primera ronda, identificar el día de la semana y hora de la entrevista junto

con el nombre del encuestador. Esta información debiera ser un insumo importante para el próximo trabajo de campo. El supervisor debiera asignarle el mismo encuestador a la persona ya entrevistada. La idea es que el encuestador vaya el mismo día y hora - por ejemplo, el viernes a las 6:00 PM - con el fin de aumentar la probabilidad de realizar la encuesta en el primer contacto. Este procedimiento tiene un impacto importante dentro de las estrategias para disminuir la no respuesta como se verá en la siguiente sección. La recomendación se entrega en este apartado para mostrar la importancia de la paradata.

Recomendación 10: De contar con la paradata y el identificador correspondiente para unir bases de datos, realizar una evaluación de los errores medición y de procesamiento de la encuesta identificando a los encuestadores y digitadores que presentan mayores tasas de errores. El total de error por encuesta podría calcularse de la información que arroja el proceso de validación automatizado - detección de errores por la malla de validación - o del número de datos faltantes en las 100 variables más sensibles de la ELPS. Esta información podría ser útil para el proceso de reclutamiento de la segunda ronda o para poner una mayor atención del trabajo que realizan los encuestadores o digitadores que presentaron mayores errores en sus encuestas.

3.3 PROPUESTA DE EVALUACIÓN GENERAL DE CALIDAD DE DATOS DE LA PRIMERA RONDA DE LA ELPS

A continuación se presenta una propuesta de control de calidad de datos de la primera medición de la ELPS considerando como elemento articulador la información de la paradata y el concepto de error total de la encuesta. La recomendación está diseñada para las ELPS que fueron realizadas en cuestionarios impresos o modo PAPI (*Paper-and-Pencil Interviewing*). Esta recomendación general se ha dividido en dos partes. En la primera parte se definen las etapas críticas del trabajo de campo desde su inicio hasta la producción final de la base de datos. Específicamente, se hace referencia a la identificación de los errores que se producen en las diferentes fases de implementación del levantamiento de información de la primera ronda de la EPLS. En la segunda sección, se presenta un procedimiento de evaluación de la calidad de datos para las etapas críticas. Este procedimiento describe los insumos y bases de datos necesarios para construir indicadores de calidad de datos. Además se definen y especifican algunos indicadores de calidad del trabajo de campo y del cuestionario base. Por ejemplo, los indicadores buscan evaluar el contacto con la vivienda seleccionada, la aplicación de la tabla de Kish, la aceptación de la entrevista completa por parte del entrevistado y la disminución de rechazos en entrevistas parciales y los errores de flujo, rango y consistencia de entrevistas realizadas. Por último, se entregan algunos comentarios a considerar para la programación y análisis de la información base para la evaluación de calidad de los datos. Específicamente, se expone la estructura esperada de los archivos ejecutables en Stata (do-file) para implementar en cada etapa de evaluación y la definición de cada uno de dichos archivos.

3.3.1 Etapas críticas del trabajo de campo de la primera ronda de la ELPS

Las etapas críticas del trabajo de campo de la ELPS pueden agruparse según el tipo de error que pueden provocar en la producción de los diferentes estimadores estadísticos. Identificar los errores de la encuesta tal como se discutió en la sección 2.1 permite identificar de una manera comprensible las etapas críticas de una encuesta compleja como es la ELPS. Cabe recordar que en la mayoría de los países la primera ronda se ha realizado o se va a realizar en modo PAPI.

El objetivo de relacionar los errores de una encuesta con las etapas críticas o situaciones que tienen la probabilidad de agregar algún error al Error Total de la Encuesta es evaluar con mayor atención los datos que se van generando en esas instancias. Las fuentes de errores que se sugiere analizar son los errores de muestra en la tercera etapa de selección, los errores de no respuesta en la primera ronda, los errores de medición, y los errores de procesamiento de la encuesta.

A continuación se describen 13 etapas críticas del levantamiento de los datos en términos de la contribución al error total de la encuesta en la primera ronda de la ELPS. Cada etapa crítica está asociada a un tipo de error y se describe detalladamente los errores que podrían detectarse.

Etapas Críticas 1 (error de muestreo): Contactar a la vivienda que fue seleccionada en la muestra

La selección de la vivienda (USM Unidad Secundaria Muestral) con muestreo probabilístico tri-etápico sin reemplazo, exige acceder a la vivienda seleccionada en forma inequívoca. Esta tarea se puede ver anulada por causas provocadas desde el diseño de muestra hasta el Entrevistador en el momento que intenta tomar contacto con la vivienda seleccionada. Por ello es una etapa que no debe descuidarse, ya que no solo está inserta al inicio del proceso de levantamiento de datos sino que es parte de la estrategia del diseño muestral.

Inicialmente debe existir una debida actualización de las unidades secundarias de muestreo en el marco muestral. Esto conlleva tener información actualizada de las reales direcciones en terreno de viviendas a seleccionar, o en su defecto referencias de entorno y características que la identifiquen inequívocamente para la ubicación por parte del Entrevistador. De no ser posible esta tarea deberá concebirse dentro del proyecto una etapa de empadronamiento previo a fin de obtener dicha actualización (Actualización de la muestra).

Posteriormente cuando ya se dispone inicialmente de las direcciones actualizadas, los registros que las contienen deben ser transmitidos a las hojas de ruta sin falta o errores de transcripción. En este punto es vital acompañar dicha información con la correspondiente cartografía actualizada, lo que mejorará notablemente una rápida y correcta ubicación en terreno.

En zonas rurales es habitual la identificación de las viviendas por referencias y características tanto de entorno como de la vivienda misma. En consecuencia la muestra en áreas rurales siempre deberá ir acompañada a terreno de una cartografía correspondiente con esas especificaciones ubicadas en ella. Esto se logra realizando en último término con croquis describiendo accesos, caminos, senderos, hitos o referencias, etc. levantados previamente en la etapa de empadronamiento de la muestra.

El acceso y ubicación de la vivienda preestablecida debe ser controlada en terreno. Esto debe cumplirse porque puede existir un error involuntario o intencional en entrevistar un hogar que no corresponda a la vivienda seleccionada y debe ser corregido para recuperar la no respuesta producida.

Dado el carácter panel del estudio es de muy alto valor actualizar la precisión de ubicación de la vivienda tanto en los registros de la Hojas de ruta de la vivienda como en la cartografía asociada a ella. Esta tarea de mejoramiento puede obtenerse desde los entrevistadores y controladores en su visita a terreno durante el proyecto y deben establecerse procedimientos ulteriores que rescaten estas mejoras para los controles y principalmente las re-visitas en olas posteriores.

Etapas Críticas 2 (error de muestreo): Aplicación de la Tabla de Kish para seleccionar a la persona

La aplicación de la Tabla Kish constituye una etapa crucial en la selección de los miembros de la muestra panel ya que un error en su aplicación puede invalidar un cuestionario completado aunque esté perfectamente respondido, es decir un cuestionario en que el entrevistador aplicó correctamente las preguntas de él y la persona entrevistada (que no correspondía), le responde correctamente y en forma completa.

La experiencia indica que existen múltiples factores que inciden en un error en la aplicación de una Tabla Kish, a continuación se enuncian los principales que debieran considerarse en cualquier evaluación de calidad.

Error de dirección de entrevista / Hogar no corresponde al direccionado: Esto ocurre cuando se listan las personas de un hogar que no corresponde a la dirección que se debe entrevistar. Esto puede ser provocado por la inexactitud en la descripción del hogar (Ej. Múltiples viviendas en una misma dirección de calle), errores en el empadronamiento, registro o transcripción de la dirección o un fraude del entrevistador.

Mal registro del total de personas del hogar mayores de 15 años: Esto ocurre en muchos casos por una mala información hacia el entrevistado para que informe sobre la totalidad de personas mayores de 15 años que efectivamente forman parte del hogar y en consecuencia entrega un número mayor (incluye personas que no corresponden) o uno menor (omite personas). Este punto es importante de resaltar ya que es muy probable que si la persona que sobra o falta no es la sorteada el entrevistador sólo se daría cuenta (probablemente) de ello cuando esté avanzado en el cuestionario. Aquí la experiencia también indica que en algunos casos el entrevistador hace caso omiso de este error, y continúa completando la entrevista considerando el costo que tiene para él reformular el cuestionario a otra persona, de la cual probablemente reciba un rechazo dado lo extenso del cuestionario que ya fue aplicado parcialmente a la persona incorrecta del hogar.

Registro errado de la edad de los integrantes del hogar: Ocurre cuando el registro de edades no corresponde al real y el mal registro puede provenir de un informante no calificado para haber dado la información de los componentes del hogar como también de un manejo fraudulento de parte del entrevistador para entrevistar a una determinada persona (generalmente presente) y no a otra (generalmente ausente). Se sugiere siempre marcar en una variable del cuestionario diagramado la persona del hogar que entrega esta información, de tal forma de resolver dudas respecto de donde se obtuvo esa información (listado de personas, total de mayores de 15 años, nombres, edades, sexo).

Registro errado del sexo de la persona: Este error también se presenta con cierta frecuencia y generalmente viene dado por descuido en el registro de la persona ya sea en la captura del nombre que al entrevistador lo hace registrar el sexo de la persona sin preguntar e inferirlo del nombre; o bien al descuido en registrar el código de sexo.

Mala interpretación de la Tabla de Kish: Esto ocurre cuando el entrevistador teniendo bien registrados los integrantes del hogar y los dígitos del folio realiza un cruce incorrecto de las variables (Ej. Permuta la aplicación entre filas y columnas de la Tabla de Kish). Este error también puede provocarse por una mala lectura del dígito de cruce de variables. Otro error puede estar en el mal registro de las sumas acumuladas de personas seleccionables que inducen a elegir una persona incorrecta

Mala transcripción del seleccionado desde la Tabla de Kish hacia la Hoja Composición de Personas en la Vivienda: Este error se manifiesta en ciertos casos cuando el entrevistador en forma descuidada cambia el dígito correcto desde la Tabla al registrarlo en la Hoja Composición de Personas en la Vivienda.

Mal uso de la calificación de elegibilidad de la persona: En algunas ocasiones algunos entrevistadores no tienen clara las razones o causas que eximen a una persona de ser elegible, estas razones o causas debieron haber sido claramente expuestas en las capacitaciones y debieran estar registradas en el Manual del entrevistador. También cabe la posibilidad de un mal uso la clasificación de elegibilidad por parte de malos entrevistadores, que mediante una clasificación

incorrecta provocan un cambio de la persona seleccionada obteniendo una que sí se encuentra presente y dispuesta a responderle la entrevista.

Etapa Crítica 3 (error de no respuesta): No hacer contacto con alguien de la vivienda

Una adecuada y constante aplicación de visitas programadas a la vivienda seleccionada puede subsanar problemas de contacto con los residentes de la vivienda seleccionada. En este sentido la supervisión de las fechas y horas de visita que realiza el entrevistador es una tarea relevante que debe cubrir su supervisor, de tal forma de realizar una distribución temporal de ellas cubriendo días hábiles y de fines de semana como también festivos para aumentar las probabilidades de contacto.

Sin duda los contactos son más difíciles cuando los hogares son de menos personas y más aún si ellas desarrollan actividades laborales fuera del hogar, sin embargo debe instruirse a los entrevistadores de ciertas técnicas para mejorar sus contactos, como lo son: 1) obtener información con vecinos confiables acerca de la ubicación horaria o lugar de las personas residentes en la vivienda seleccionada; 2) Obtención de registro de números telefónicos de la vivienda en fuentes públicas de información (Guías telefónicas; internet, etc.) para solicitar una hora de visita en la vivienda y exponer toda la información del estudio. La información previa dada a estos residentes debe estar estructurada y detallada de forma tal que no coartar la realización de la entrevista por una exposición deficiente o inadecuada del objetivo del contacto.

Una especial atención deben recibir los hogares unipersonales en que el residente no se logra contactar ya que la frecuencia de estos casos (sin logro) finalmente impactará en los estimadores al quedar esta tipología sub-representada en la muestra lograda.

La distribución de muestra que se realiza con los entrevistadores debe cuidar mantener un pronto y eficiente contacto. Por lo tanto la muestra asignada por unidad de tiempo debe asegurar una visita regulada (en fechas y horas diferentes) en a lo menos 4 oportunidades a la dirección de la vivienda, considerando además implícitamente los tiempos de traslado o transporte del entrevistador en su jornada diaria de trabajo desde su propia residencia a las áreas de la muestra.

En áreas rurales las visitas a terreno deben cuidar el buen uso de los recursos presupuestarios. Así en zonas de difícil o distante acceso por geografía, es necesario considerar siempre la posibilidad de lograr contacto con algún residente del hogar seleccionado en algún lugar distinto (cercano a la vivienda) y que haya sido referido a través de un vecino de su vivienda. Esta posibilidad también debe poder ser aplicada para entrevistar inmediatamente a la persona seleccionada si es habida en un lugar cercano. En este caso deben cuidarse las condiciones de entorno (ambiente reservado y tranquilo) del lugar de la entrevista para que no procuren omisiones o respuestas irreales provocadas por la presencia de otras personas en el lugar en que se realiza la entrevista.

En algunas zonas geográficas existen viviendas que mantienen un control previo de acceso a través de personal de seguridad o mantenimiento de áreas que restringe el acceso a las viviendas seleccionadas. Ejemplo de ello son los conserjes o guardias en los accesos a edificios o condominios habitacionales. Para estos casos se recomienda desarrollar Cartas a la Administración de condominios residenciales dirigidas a dichas agrupaciones a fin de obtener una respuesta positiva proveniente desde los mismos residentes de las viviendas ya informados y no una respuesta negativa de la Administración de tipo general.

La disposición de las personas a revelar información al entrevistador se ve fuertemente mejorada cuando se respalda a través de información actualizada y respaldada en páginas oficiales de instituciones del Estado o del equipo de investigación del estudio.

Etapa Crítica 4 (error de no respuesta): No lograr contacto con la persona seleccionada por la Tabla de Kish

El contacto del entrevistador con los residentes de la vivienda seleccionada debe revelar un alto grado de profesionalismo y corrección en su trabajo, de tal forma que genere un nivel de confianza suficiente en el entrevistado para acceder a responder todas las preguntas de la ELPS.

Entre las condiciones para generar confianza en el entrevistado están primeramente una presentación personal limpia y ordenada del entrevistador; buen conocimiento del estudio y proyecciones; correcto uso del lenguaje o dialéctico de comunicación con el entrevistado; transparencia en su trabajo entregando información útil al entrevistado para su disposición a cooperar en el estudio; uso permanente de credencial y Cédula de Identificación Civil (C.I) o Documento Nacional de Identidad (D.N.I).

Para la correcta aplicación de la Tabla de Kish es requisito indispensable obtener información fidedigna de los integrantes del hogar en cuanto su cantidad, nombres y edades. Esta es sin duda una de las etapas de mayor riesgo en No respuesta por la poca o nula disposición a colaborar por parte del o los integrantes del hogar. Este desinterés puede ser revertido habilitando el entrevistador de capacidades disuasivas de ese desinterés basadas en un conocimiento real y oficial entregado al residente, que sea fácilmente verificable por él. Complementariamente esta etapa crítica es mejorada con información preliminar escrita dirigida a las direcciones de ciertos estratos socioeconómicos o condominios habitacionales en que se prevea esta situación dada en estudios anteriores realizados en esos lugares.

Para los estratos socioeconómicos altos es siempre necesario disponer de información condensada en cartas; dípticos o trípticos; folletos que informen los contenidos y objetivos del estudio con referencias claras para que puedan contactar requiriendo más información.

Siempre el entrevistador debe adaptar sus horarios de trabajo a los de disponibilidad de los residentes del hogar seleccionado, mejorando las probabilidades de contacto con visitas distribuidas en diferentes días y horarios.

Paralelamente es muy importante mantener para este punto y para el proyecto en general una página o link en las redes sociales de comunicación (Internet u otra) de la población objetivo del estudio que informe del estudio ELPS durante su levantamiento indicando las probabilidades de solicitar respuesta a la población seleccionada.

Es de sumo cuidado que los números telefónicos centrales indicados en estos informativos tengan números válidos y de atención continua a los entrevistados consultantes. Este personal que responde las consultas telefónicas debe ser capacitado, para ser competente en su conocimiento del estudio y no causar un rechazo por respuestas insuficientes al entrevistado que realiza consultas previas a responder la ELPS. Debe considerarse que luego de cada llamada el entrevistado debe quedar realmente convencido de la veracidad y seriedad de este proyecto oficial. Esto también es válido para las consultas que puedan plantearse vía internet al sitio o email del proyecto ELPS en donde deberá haber un monitoreo y seguimiento de respuestas a las diferente consultas planteadas por los potenciales entrevistados y a la población en general. Las respuestas a consultas vía web pueden mejorar si se establecen preguntas frecuentes tipo que se presentan en la página web del proyecto con la debida y profunda respuesta oficial. En estos contactos preliminares de consultas del entrevistado, es muy útil explorar la posibilidad de concertar una visita programada de un entrevistador a su domicilio (vivienda seleccionada).

Etapa Crítica 5 (error de no respuesta): No lograr firmar el consentimiento informado del encuestado

La obtención de consentimiento informado por parte del entrevistado debe siempre estar fundada en un pleno conocimiento del estudio por parte del entrevistado ya que será parte fundamental en futuros levantamientos de este estudio panel. Por lo tanto desde este primer contacto es imprescindible otorgarle todas las respuestas e información en forma muy profesional a sus inquietudes, teniendo presente no inhibir su participación en el proyecto.

La información anexa complementaria es la oficial manejada desde la dirección central del proyecto y no debe usarse otra información sino sólo la oficial. La importancia de seguir esta norma radica en que toda la población debe ser sometida al mismo diagnóstico del estudio en igualdad de condiciones evitando inducciones de cualquier índole que puedan alterar los resultados de un real estado de las variables en la población.

Otro aspecto que se debe supervisar es la correcta aplicación de los incentivos (si los hubiera) al entrevistado para minimizar la no respuesta. Estos incentivos debieran ser solo los otorgados oficialmente por la dirección central del proyecto. Se supervisará la promoción de cualquier otro incentivo o promesas no autorizadas que realice el entrevistador pretendiendo obtener la entrevista con incentivos ajenos a los proyectados.

Para lograr resultados positivos en forma estandarizada que avalen una adecuada y oportuna información a los entrevistados se sugiere establecer protocolos documentados (folletos, dípticos, etc.) para aplicar esta etapa previa por parte del entrevistador y que posteriormente serán controlados.

En estudios panel realizados en otros países (UK ; Chile) un efecto disuasivo de rechazo en esta etapa crítica lo constituyen los regalos que se realizan a los entrevistados seleccionados (en dinero efectivo; tarjetas con carga monetaria para ser cobradas en artículos de su interés; souvenirs con logos del proyecto, etc.). Debe quedar claro que éstos deben ser informados y entregados sólo al concluir una entrevista completa y bajo ninguna forma condicionar su entrega a la firma del consentimiento y respuesta del Cuestionario de la ELPS. De esta forma queda sentado un buen precedente para obtener respuestas futuras positivas en las olas posteriores. Del mismo modo el nivel de costo de estos regalos de agradecimiento debe ser moderado para no influir negativamente en las futuras atenciones de este tipo que pudieran recibir los entrevistados.

Los documentos de consentimiento informado (hogar y personas) mantienen información relevante en su contenido, es de vital importancia su resguardo físico y además el respaldo actualizado de la información que contienen. Es importante saber en caso de pérdida de uno de estos documentos, cuál fue la secuencia que tenía registrada en su aprobación o desaprobación. No sería profesional volver a solicitar un nuevamente el consentimiento informado a un entrevistado que respondió la entrevista, sin tener esta información disponible registrada. Ej. Fechas de las visitas y su respuesta de autorización que buscamos nuevamente tener impresa.

En esta etapa nuevamente el profesionalismo de los entrevistadores es muy importante y se debe mantener cuidado en no incurrir en irresponsabilidades, descortesías, malas conductas, mala presentación personal, mal vocabulario, etc.

En todo instante se deberá procurar entregar garantías hacia el entrevistado respecto del resguardo de sus datos, es decir asegura la confidencialidad de sus respuestas.

Etapa Crítica 6 (error de no respuesta): No lograr terminar la encuesta.

La duración del cuestionario hace necesario un refresco motivacional consciente por parte de los entrevistadores que no debiera ser posterior a las preguntas relacionadas con la salud del entrevistado, de tal forma de distender la entrevista, haciéndola más amena y menos monótona para el entrevistado, sin desmedro en la aplicación correcta de cada pregunta.

En ciertas circunstancias especiales, ante imprevistos, emergencias, o cansancio extremo del entrevistado, se toma la decisión de interrumpir la entrevista para continuarla de común acuerdo en una fecha posterior. Estas situaciones deben ser monitoreadas por el supervisor de tal forma de no incurrir en irresponsabilidades en el cumplimiento de las citas acordadas entre entrevistador y entrevistado. Es relevante mantener suma cautela en un trato cordial, amable y responsable con cada uno de los entrevistados ya que son piezas fundamentales para la continuidad del proyecto en las olas posteriores.

Es importante entregar al inicio toda la información necesaria que está estipulada en los protocolos para evitar dudas en el entrevistado que provoquen una detención en el desarrollo de la entrevista. Un tema sensible que provoca reacciones adversas es la consulta de ingresos monetarios por lo que su planteamiento debe estar claro al inicio de la entrevista, asegurando siempre la confidencialidad de los datos

La lectura de las preguntas debe realizarse con claridad y fluidez dando sentido correcto a la pregunta.

El supervisor deberá tener siempre un seguimiento actualizado de las entrevistas en estado de resultado parcial y deberá procurar su cambio a estado de logro completo. Para ello puede en determinados casos realizar reasignaciones de la muestra con el fin de completar la entrevista aún con otro entrevistador si fuese necesario para así satisfacer la disponibilidad horaria del entrevistado.

Etapa Crítica 7 (error de medición): Saltos incorrectos en flujo de las preguntas

La capacitación del entrevistador debe considerar siempre a lo menos un par de ejercicios de desarrollo de aplicación de cuestionarios ELPS completos, es decir llenar dos entrevistas reales de prueba que sean acuciosamente revisadas por los supervisores para no sólo detectar errores en saltos o flujo; sino también en la revisión de universos de aplicación de las preguntas específicas que corresponda responder y las inconsistencias que pudiera generarse a partir las respuestas obtenidas. Estas pruebas deben realizarse con hogares tipo y no solo ensayar con hogares unipersonales.

Sin duda una herramienta efectiva de usar, para cubrir estandarizadamente los chequeos básicos lo constituye enlistar estos chequeos que deben revisarse para hacer un *check list* de cada uno de ellos.

Estos errores demandan una gran atención y concentración del entrevistador en la modalidad PAPI y que en contraste no aparecen en la modalidad CAPI. Por ello el entrevistador deberá estar siempre atento a la correcta lectura del Cuestionario ELPS.

El diseño del Cuestionario (y toda la documentación asociada) debe ser de una tipografía y diagramación formal que permita leer cada parte de su contenido sin inconvenientes, siguiendo el sentido lógico del desarrollo de las respuestas esperadas, en condiciones de luz natural (diaria).

El Cuestionario ELPS debe ser sometido a variadas pruebas preliminares que eliminen cualquier error en la edición impresa del mismo ya sea en enunciados; saltos; universos; nombres de variables; alternativas de respuesta; falta de espacios predefinidos de respuesta, etc.

El incurrir en un salto incorrecto puede llevar al rechazo de la entrevista por parte del entrevistado al observar desorden y falta de lógica en la formulación de las preguntas. Adicionalmente el incurrir en estos errores provoca en forma ineludible una pérdida de tiempo importante que va en desmedro del tiempo previsto para las preguntas siguientes, y causando además en el entrevistado un cierto nivel de fatiga y falta de concentración para emitir sus respuestas con información realmente válidas y significativas para el proyecto (evita profundizar para dar una respuesta más real y así poder terminar pronto con la entrevista).

Etapas Críticas 8 (error de medición): Missing y outliers en los datos

La primera responsabilidad de revisión del cuestionario reside en el mismo entrevistador. Se debe dejar constancia de su revisión escrita por parte de él en un recuadro dispuesto para ello (puede ser al final del cuestionario). Es común que los entrevistadores omitan registrar información para disminuir la duración de la entrevista. Ej. No marcar las respuestas No Sabe; No responde; sexo (que lo infieren del nombre), etc. Y dejan esta labor para realizarla una vez terminada la entrevista y en otro lugar. Esta estrategia es abiertamente deficiente para el proyecto ya que pueden registrarse u omitirse respuestas que no corresponden. Esta mala estrategia debe ser mostrada como tal en las capacitaciones ya que además obliga al entrevistador a “recrear” en su mente el hogar visitado y recorrer todo el cuestionario en la búsqueda de respuestas que no registró debidamente y claramente le demandará un tiempo no menor. Como resumen la entrevista debe ser completada con la última respuesta obtenida del entrevistado y sólo debiera completarse posteriormente tras una revisión con el entrevistado presente, chequeando que no falte o sobre información.

Debe concebirse un listado de todas las variables que deben traer respuesta a todo evento, como también realizar una tabulación de seguimiento a los niveles de *missing* que traen las entrevistas de tal forma de instruir y/o recapacitar a los entrevistadores para evitar sus ocurrencias. La presencia de valores *missing* siempre está latente de presentarse en el cuestionario, más aún cuando las variables que los contienen no forman partes de condiciones de universo o saltos que el entrevistador debe observar a continuación en el desarrollo del cuestionario.

Es primordial evitar la aprobación de entrevistas con *missing* en etapas previas a la digitación, por cuanto el programa de digitación puede exigir como requisito un valor válido en la variable con respuesta *missing*.

Para todas las variables que requieren información de ingresos monetarios o tenencia de activos o bienes de valor del entrevistado deben existir alternativas de respuesta que permitan una expresión clara de negativa a responder esas preguntas y bajo ninguna circunstancia deben quedar sin respuesta si correspondiera aplicarlas según el flujo de la entrevista.

Los valores *outliers* requieren ser detectados ya que representan, de no ser verdaderos, un claro sesgo en el estadístico que participe esa variable. Un área importante donde pueden generarse estos valores *outliers* es en el registro de ingreso donde existe la tendencia a anotar solo números significativos de la cifra monetaria. Ej. El entrevistado dice \$832.000 y el entrevistador registra \$832 para ahorrar tiempo o encapsular la respuesta en un cuadro de registro diseñado en forma deficiente para esa respuesta de \$832.000. Otros casos los constituyen las amplificaciones de los valores reales al escribir con más dígitos que los que corresponde, generalmente por mala audición de la cifra (en cuyo caso el entrevistador debe estar atento una pronunciación incorrecta del

entrevistado y repreguntar para precisar). Estas ampliaciones son también por descuido de los entrevistadores al realizar el registro (Ej. Le informan \$100.000 y anota \$1.000.000).

Sin duda estos errores de *missing* y *outliers* pueden ser producidos en la etapa posterior de digitación de los datos y también deben ser considerados en esa tarea de procesamiento (Etapa Crítica 10 vista más adelante). En estas etapas posteriores también surgen estos errores de *missing* y/o *outliers* por una mala caligrafía y/u ortografía del entrevistador que induce a una interpretación incorrecta del dato.

Los entrevistadores, supervisores y especialmente los Críticos validadores-codificadores deben tener establecidos los rangos posibles de respuesta que pueden obtener por parte de los entrevistados, de tal forma de corregir estos valores *outliers* incorrectos que provienen del mismo entrevistado. Estos rangos de referencia pueden provenir para ciertas variables de otros estudios oficiales realizados en la población objetivo. En el aporte a la prevención de valores *outliers* también debe entregarse una capacitación de rangos válidos de montos monetarios otorgados por instituciones estatales y que están normadas y son de público conocimiento y son consultados en algunas preguntas del cuestionario. Estos rangos de montos para estas respuestas (al igual que los requisitos para recibirlos) deben ser explicitados claramente en los Manuales del entrevistador; Supervisor y Crítico validador-codificador y por lo tanto estarán accesibles a cada uno de ellos para solicitar la corrección de valores *outliers* en estas variables.

Etapa Crítica 9 (error de medición): Inconsistencias de respuestas entre variables al interior del cuestionario

Esta etapa crítica no debe restringirse sólo a ser abordada en la sección correspondiente al Crítico validador-codificador o peor aún, postergarla para que las inconsistencias sean detectadas por la Malla de validación y consistencia final que contendría la completitud de todas las revisiones posibles, sino por el contrario deben distribuirse y aplicarse en todo el proceso previo a la formación de la base de datos

Las razones para anticipar la detección de los errores de inconsistencia en cada etapa de revisión; desde el Entrevistador que revisa su trabajo antes de finalizar su entrevista, y luego antes de entregarlo al Supervisor; y este último nuevamente revisa el cuestionario para que el Crítico validador-codificador realice una tarea de revisión más exhaustiva ; obedecen a una mejora sustantiva en la calidad de los datos, con correcciones efectivas y oportunas, evitando costos muy superiores si se postergan para ser tratados luego de obtener la base de datos.

En las inconsistencias deben estar consideradas las fallas de aplicación correcta de los universos pertinentes a cada pregunta, como asimismo las exigencias previas que deben cumplirse para registrar ciertas categorías de respuestas.

La detección de inconsistencias requiere un grado mayor concentración en la revisión que los errores de flujos; rangos y universos, ya que tiene implícita la observación de varias respuestas para diferentes variables en forma simultánea, y que generalmente están distribuidas en diferentes módulos del cuestionario.

Debe capacitarse a los entrevistadores en algunos tipos clásicos de inconsistencia como lo son la relación curso y nivel educacional vinculados con la edad de la persona (Ej. Adulto asistiendo a educación Inicial; nivel de competencias adquirido en la educación formal y capacitaciones vinculadas con el oficio que desempeña la persona (Ej. Paramédico con sólo educación inferior a al nivel superior). De esta forma se estarán evitando respuestas deficientes desde los informantes y que pueden obedecer por ejemplo a que son emitidas por un informante proxy como es el caso del

llenado del módulo I Composición y Características del hogar en que una persona Entrevistado (con Jefe de hogar) responde información de otros integrantes del hogar.

Etapa Crítica 10 (error de medición): Respuestas No sabe y No responde con alta frecuencia en múltiples variables

Esta etapa debe estar regulada en forma precisa antes de las capacitaciones de los entrevistadores de tal forma de prevenir su uso desmedido y procurar obtener de parte del entrevistado una respuesta con valor significativo para los estimadores estadísticos que se desean concluir.

Debe existir una tabulación permanente de las tasas de uso de estas alternativas de respuesta que son utilizadas como verdaderas “puertas escape” para terminar prontamente la entrevista y que son recursos utilizados tanto por algunos entrevistados como por algunos entrevistadores incorrectos o capacitados en forma deficiente.

En este grupo de tabulaciones también deben estar contenidas y en forma separadas las respuestas No responde.

Los resultados de estas tabulaciones deben ser analizadas con prontitud realizando en ellas una comparación entre entrevistadores, con el objetivo de recapacitar o reforzar instrucciones que disminuyan la frecuencia de estas alternativas de respuesta.

El entrevistador debe tener presente en todo instante que una respuesta No sabe o No responde puede obedecer a que el informante puede ser proxy en relación a la pregunta, y no es calificado para responderla en forma adecuada y debe en consecuencia procurar que el informante sea la persona referida en la pregunta, si estuviera presente en el hogar mientras realiza la entrevista.

Las preguntas consultantes de ingresos monetarios generalmente tienen una mayor tasa de respuestas del tipo No sabe/ No responde que deben ser disminuidas a partir de una capacitación adecuada de los entrevistadores para evitar recibir este tipo de respuestas. Igualmente estas tasas deben ser monitoreadas para detectar aquellas malas prácticas de algunos entrevistadores que ofrecen la alternativa No sabe/No responde antes de obtener las respuestas espontáneas de ingresos por parte de los entrevistados con el objetivo de atenuar un posible rechazo en la continuidad de la entrevista.

Etapa Crítica 11 (error de procesamiento): Digitación de las hojas de rutas y de las encuestas

La digitación es una etapa que en muchos estudios de modalidad PAPI suele postergarse más tiempo del programado inicialmente. Estos retrasos de su puesta en marcha producen una detención masiva del conjunto de datos que requiere prontamente ser analizados vía aplicación de una malla de validación inicial que acuse las fallas de revisión básicas en las etapas previas. Una de los objetivos que debe tener este proyecto es decrecer al mínimo la puesta en marcha de esta etapa de digitación y que permite obtener pronto una base de datos que sea analizada luego con la máxima rigurosidad.

La digitación de los cuestionarios y hojas de ruta (o formulario de Contacto) debe ser realizada en un ambiente protegido, con condiciones de luminosidad y espacios adecuados para mantener el orden imprescindible en el movimiento físico de la documentación que está siendo digitalizada.

El equipamiento computacional y programa de digitación deben ser aptos para estar configurados en la plataforma general que se exigirá a este proyecto para mantener en todo instante un alto nivel de seguridad en el manejo de datos.

Todo el proceso de digitación estará normado en documentación de respaldo, que entregue información de la ubicación exacta de cada cuestionario y la persona que está siendo responsable de éste en esa etapa de ubicación.

Esta etapa es delicada porque en ella se realiza una transcripción de la información desde la fuente manuscrita hacia una digitalizada en una operación realizada por personas que deben saber interpretar correctamente las respuestas.

La interpretación correcta exige en primera instancia que la fuente tenga una descripción clara y legible de la información. Es por tanto necesario mantener la exigencia de una escritura de buena caligrafía y ortografía a los entrevistadores, como también el registro de las respuestas en los lugares establecidos para ello. También en este punto incide nuevamente el diagramado y diseño del cuestionario en disponer de aquellos espacios adecuados para registrar las respuestas.

Los digitadores deben formar parte de una capacitación mínima acerca de los contenidos del Cuestionario ELPS, Hojas de ruta y Consentimientos informados antes de la digitalización de la información. Esto permitirá obtener solución por parte de ellos a problemas de interpretación de datos que vienen defectuosos en su registro.

Sin duda la interpretación en la lectura es una fuente de potencial error del digitador pero a ello se suma la correcta escritura de lo que interpretó o debiera interpretar, es decir lo que finalmente digita. En esta escritura pueden incorporarse errores involuntarios al digitar mal en el teclado del dispositivo de entrada de datos.

Visto lo anterior es imprescindible someter esta etapa a una estrategia de control de la digitación que reduzca los errores en esta etapa a una mínima, idealmente nula existencia.

La estrategia de control debe establecer rápidas correcciones en el proceso de digitación para eliminar la ocurrencia repetitiva de esos errores de digitación detectados.

La labor de control debe estar aplicada a todo el trabajo de digitación, en este aspecto la doble digitación al 100% de las entrevistas es sin duda una de los métodos más seguros para obtener un alto estándar de confiabilidad y calidad en una transcripción correcta, sin embargo sus costos limitan su aplicación en forma completa optándose muchas veces por disminuir ese porcentaje con selección aleatoria de porcentajes inferiores al 100%, bajo criterios de cotas mínimas de diferencia aceptables entre digitaciones de un mismo instrumento realizada entre dos digitadores.

La importancia adicional de tener confirmada una correcta digitación radica en la posibilidad que sean digitados valores válidos para las mallas de validación pero que no representan la respuesta escrita dada por el entrevistado. Estos errores si no se detectan en esta etapa pasan inadvertidos y pueden provocar sesgos no cuantificados en los estadísticos que se generen con esos datos erróneos.

Es importante establecer fórmulas adecuadas de cancelación de honorarios a los digitadores que no sólo incorporen la cantidad de trabajo realizado (digitado) sino también la calidad de dicho trabajo y que sea ponderado por sus resultados según los controles calidad.

Etapa Crítica 12 (error de procesamiento): Codificación

Las preguntas abiertas sujetas a una posterior digitación deben ser revisadas previamente por el entrevistador, y supervisor para que el Crítico validador-codificador realice su trabajo en forma lo más dinámica y correcta posible.

Es importante que las personas que codifiquen tengan una plataforma de comunicación oficial en que compartan sus criterios y que sea monitoreada por la dirección del proyecto para los casos frecuentes que ocasionan discusión o controversia en la aplicación de un determinado código

Esta etapa al igual que la digitación exige un control riguroso por cuanto también puede ser génesis de errores inadvertidos por las mallas de validación posterior. Los porcentajes de revisión de estas codificaciones (% de control) debe realizarla en primera instancia el Supervisor al chequear el trabajo realizado por el Crítico validador-codificador.

Las capacitaciones de los codificadores no sólo se deben limitar a informar sobre los listados o diccionarios de códigos que debe aplicar, sino también a la estructura orgánica que tienen dichos diccionarios y que en el caso de las clasificaciones de oficios y actividades económicas son muy relevantes, y su conocimiento les entregará a los codificadores una mayor efectividad en aplicar los códigos correctos.

La codificación debe ser realizada en un lugar que reúna las condiciones necesarias para un trabajo de alta concentración.

La codificación de buena calidad exige atender durante su ejecución la observación de otras variables que le dan consistencia válida al código. Un ejemplo de ello lo constituye la aplicación de códigos de oficio donde deben observarse la edad; sexo; educación formal; capacitaciones; lugar de trabajo; ingresos; tipo de contrato y otras variables complementarias; antes de aplicar el código de oficio considerado como correcto.

Deberá realizarse una capacitación con exigentes ejercicios en la aplicación de códigos.

Etapa Crítica 13 (error de procesamiento): Generación de bases de datos

Las bases de datos deben en primer lugar ser accedidas solo por personal autorizado y por lo tanto todo usuario de ellas ya sea total o parcialmente debe estar identificado y formar parte del personal autorizado del proyecto. Las bases de datos deben cuidar en todo momento el manejo indebido de la información nominada.

El manejo de la información de bases datos debe ser regulado y realizado en una plataforma informática que asegure su resguardo y evite fugas o puertas de acceso a personas no autorizadas. Por lo tanto no debiera haber manipulación en dispositivos portátiles de esta información (Discos externos de memoria; pendrive; accesos inalámbricos; etc).

Las bases de datos deben cuidar una estructura que optimice su tratamiento de validación y contenga todas las variables del estudio incluyendo las variables con respuestas codificadas.

Las bases de datos deben contener llaves de conexión (Ej. Folio+N° Vivienda+N° de orden de la persona) y que mantenga la innominación de la muestra para aquellas sub bases en que no sea necesaria la observación de la identificación de las personas o variables que precisen su dirección de vivienda.

Las bases de datos deben ser identificadas con versiones que aseguren un tránsito de fácil seguimiento de ellas durante el proceso de validación. Esto resulta importante de destacar ya que las bases van creciendo en registros conforme se levantan los datos. Y es importante identificar en un instante determinado si se está trabajando con bases actualizadas y cuáles son sus fechas de exportación o generación inicial.

Cada base de datos exportada deberá ser sometida a un análisis descriptivo acompañado en paralelo con un seguimiento en el cuestionario impreso para detectar pérdida de respuestas o variables de salida y que obedezcan a un manejo interno en la programación computacional que opera en la generación, transmisión, almacenamiento, selección y exportación de datos que configuran las bases de datos de salida.

Las estructuras de las bases de datos deben estar finalmente orientadas al usuario o investigador que hace uso de la información para representar los resultados muestrales.

El software de análisis de las bases de datos debe ser aceptado como válido por todas las partes involucradas en el proyecto. En general se utiliza el paquete de software estadístico Stata que es internacionalmente reconocido por su gran versatilidad y eficiencia en el análisis estadístico tanto en el mundo académico como de ciencias aplicadas.

3.3.2 Procedimiento de evaluación de la calidad de datos para las etapas críticas

Una vez identificadas las etapas críticas de la primera ronda de la ELPS resulta fundamental establecer un conjunto de indicadores que permitan medir concretamente la calidad de los datos. Para esto es fundamental contar con información del diseño de la muestra, participación de capacitaciones, el perfil de los encuestadores, los resultados de las visitas e información del proceso de levantamiento de información cuantitativa.

En consideración a ello, esta sección se conforma de dos apartados. El primero de ellos busca listar los insumos requeridos como fuentes de información y, el segundo, presenta los indicadores que se sugiere implementar como parte de un reporte de calidad de datos para la primera medición de la ELPS.

Insumos para la evaluación de la calidad de datos de la ELPS

El proceso de evaluación de la calidad de datos de la ELPS requiere de la identificación y preparación de un conjunto de instrumentos y bases de datos que se recomienda utilizar para conocer el estado de los indicadores de calidad del levantamiento de información.

Entre los instrumentos se distinguen el Cuestionario Base; las Hojas de Ruta y Seguimiento de la Muestra; archivos con información curricular del personal de trabajo de campo, cronograma de trabajo de campo, documento de distribución de la muestra en el trabajo de campo y archivos de variables críticas del cuestionario base.

Entre las bases de datos se considera el archivo electrónico con la muestra (viviendas seleccionadas, hogares y entrevistados); bases del cuestionario base proveniente de las exportaciones de la información digitalizada del levantamiento PAPI; bases de paradata con información del proceso de levantamiento; sus etapas de control, evaluación y ajustes.

Para el procesamiento de estos insumos se definen indicadores de calidad de datos que describen el conjunto de resultados y tasas generadas para evaluar y proyectar el comportamiento y calidad de atributos y variables relevantes del estudio. Dado que se considera que el estudio tiene una gran complejidad, sumado al hecho que el levantamiento se ha realizado en modo PAPI, se requiere realizar un seguimiento a los indicadores de calidad a través de archivos de programación ejecutables de un software estadístico como Stata (do-files). Estos archivos contienen la descripción de los insumos que se requieren para su procesamiento y las instrucciones estadístico-computacionales para generar los resultados. Los do-files generados para evaluar la calidad de datos

de la primera ronda de la ELPS pueden implementarse para monitorear la calidad de datos durante las siguientes mediciones de la ELPS.

Bases de datos requeridas para la evaluación de la calidad de datos

El producto final de todo levantamiento de información cuantitativo mediante encuestas corresponde a la base de datos. La generación de bases de datos, en cuanto es una etapa crítica, tiene la particularidad de contener información que no debe ser accedida ni intervenida por personal no autorizado, por tanto su manejo deberá cautelar las condiciones de confidencialidad y seguridad de los datos establecidas en el proyecto.

Un proceso que resguarde las condiciones de confidencialidad y seguridad de los datos implica tener un programa de digitación que produzca los datos en forma segura, que puede enriquecerse con encriptación de datos; transmisión en protocolos seguros (ftps); acopio centralizado de datos con respaldo periódico (backup); y accesos regulados y administrados con registro histórico de accesos a la base, entre otras acciones.

Para identificarlas las bases de datos, se sugiere que lleve el sufijo *_aammddvx* donde *aa mm dd* representan respectivamente el año, mes y día en que fue creada. El complemento *vx* representa la versión diaria creación. Ej de generarse tres bases con la misma información de variables durante una misma fecha *_aammdd*, la primera generada tendrá la identificación de sufijo *_aammddv1* ; la segunda *_aammddv2* y la última *_aammddv3*.

Las bases de información para la evaluación de calidad se dividen en 3 grandes grupos:

G bd 1.- Grupo de Bases de Muestrales: En este grupo están contenidas todas las bases que contienen información nominada de la muestra.

- Base de Datos Direcciones de la Muestra: *bddd_muestra_aammddvx.dta*

Esta base estrictamente reservada contendrá un registro para cada folio asociado a una dirección de la muestra ELPS en que se precisará la Departamento; Municipio; Distrito; y dirección precisa de la vivienda seleccionada. El formato estará sujeto a la entrega del organismo responsable de su desarrollo.

G bd 2.- Grupo de Bases Cuestionario base: Contiene todas las bases con información total o parcial del Cuestionario base ELPS manteniendo en forma separada la nominación de datos.

- Base Entrevistado *aammddvx.dta*: base de datos con la totalidad de variables cuyo universo sean los entrevistados.

Esta base tiene información correspondiente a los diferentes módulos del cuestionario aplicadas al entrevistado. El formato de este archivo se propone que sea matricial por folio y se estructura mediante registros lineales por folio.

- Base MiembrosHogar *aammddvx.dta*: base de datos con la totalidad de variables cuyo universo sean los integrantes del hogar del entrevistado.

Esta base tiene información correspondiente al folio; número de integrantes del hogar; listado de personas del hogar; sexo; edades. El formato de este archivo se propone que sea matricial por folio. Esta base es fundamental para el análisis de un correcto sorteo aleatorio para la elección de la persona a entrevistar. Se estructura mediante Estructura de registros matriciales por folio-persona.

- Base Hijos *aammddvx.dta*: base de datos con la totalidad de variables cuyo universo sean los hijos entrevistados.

Esta base contiene la información de los hijos del entrevistado consultados según el cuestionario. Se estructura mediante registros matriciales por folio e hijos.

- Base Carreras aammddvx.dta: base de datos con la totalidad de variables cuyo universo sean las carreras de los entrevistados.

Esta base contiene el módulo de información de carreras de educación terciaria, técnica y/o universitaria que ha recibido o está recibiendo el entrevistado. Se estructura mediante registros matriciales por folio, carreras.

- Base Capacitacion aammddvx.dta: base de datos con la totalidad de variables cuyo universo sean las actividades de capacitación de los entrevistados.

Esta base contiene el módulo de información de capacitación para el trabajo que ha recibido o está recibiendo el entrevistado. Se estructura mediante registros matriciales por folio, capacitaciones.

- Base Hlaboral aammddvx.dta: base de datos con la totalidad de variables cuyo universo sean las historias laborales del entrevistado.

Esta base contiene el módulo de historias laborales correspondientes al entrevistado en el período de análisis. Se estructura mediante matriciales por folio, período de historia laboral.

- Base CaracterHogar aammddvx.dta: base de datos con la totalidad de variables cuyo universo sean los integrantes del hogar del entrevistado.

Esta base contiene la información de la composición y características del hogar. Se estructura mediante registros matriciales por folio e integrante del hogar.

G bd 3.- Paradas: Este grupo estará formado por todas las bases que sean complementarias al proceso y son desarrolladas con objetivos específicos de evaluación y control de procesos antes, durante y posterior al proceso de levantamiento.

- Base Paradata_ aammddvx.dta: base de datos del proceso de desarrollo de la encuesta

Contiene la información correspondiente a las variables complementarias que acompañan el cuestionario base entre ellas: folio; Identificación y resultados de visitas; fechas y horas; identificación de entrevistador; supervisor; revisor crítico codificador; digitador; fechas de procesos; dirección de muestra. Se estructura mediante registros lineales por folio.

- Base de datos de la Hoja de RRNH: Nombre: bbdd_hrrnh_aammddvx.dta

Esta base contendrá un registro para cada folio de la muestra en que se digitalizará toda la información que se recopila en la Hoja de RRNH.doc. Esta disposición lineal (no matricial) mejora el manejo de análisis compuesto con otras bases de datos. Se creará en formato Excel para su formación y en formato Stata para su análisis.

- Base de datos de encuestadores: Nombre bbdd_iinfenc_aammddvx.dta

Esta base contiene un resumen estructurado de la totalidad de la información captada en el instrumento Información del encuestador (IINFENC.doc) ordenada en un registro por encuestador. Se creará una base de entrada en formato Excel y una de salida en formato Stata. De usarse la plataforma computacional para el ingreso de datos, solo se exportaran salidas en formato STATA. Las bases de salida (exportadas) serán protegidas con claves de acceso.

- Base de datos de supervisores: Nombre bbdd_iinfsup_aammddvx.dta

Esta base contiene un resumen estructurado de la totalidad de la información captada en el instrumento Información del supervisor (IINFSUP.doc) ordenada en un registro por supervisor. Se creará una base de entrada en formato Excel y una de salida en formato Stata. De usarse la

plataforma computacional para el ingreso de datos, solo se exportaran salidas en formato Stata. Las bases de salida (exportadas) serán protegidas con claves de acceso.

- Base de datos de Crítico Codificador : Nombre `bbdd_iinfcri_aammddvx.dta`

Esta base contiene un resumen estructurado de la totalidad de la información captada en el instrumento Información del crítico codificador (IINFCRI.xls) ordenada en un registro por crítico codificador. Se creará una base de entrada en formato Excel y una de salida en formato Stata. De usarse la plataforma computacional para el ingreso de datos, solo se exportaran salidas en formato Stata. Las bases de salida (exportadas) serán protegidas con claves de acceso.

- Base de datos de digitadores: Nombre `bbdd_iinfdig_aammddvx.dta`

Esta base contiene un resumen estructurado de la totalidad de la información captada en el instrumento Información del digitador (IINFDIG.doc) ordenada en un registro por digitador. Se creará una base de entrada en formato excel y una de salida en formato Stata. De usarse la plataforma computacional para el ingreso de datos, solo se exportaran salidas en formato Stata. Las bases de salida (exportadas) serán protegidas con claves de acceso.

- Base de datos de las 100 variables claves: Nombre `bbdd_100var_aammddvx`.

Esta base se forma con la estructura del archivo `Lista_100_var.xls` que va en crecimiento con los datos recopilados por los entrevistadores y servirá de base para los análisis de estas variables vinculadas a otras bases, en particular la base Nombre `bbdd_iinfenc_aammddvx.dta`.

Construcción de los indicadores de calidad de datos

El análisis de la calidad de datos en estudios socio-métricos esta crecientemente siendo objeto de desarrollo de nuevas metodologías de control de calidad tanto a nivel académico como de gerencia de proyectos. Estos trabajos se instalan con el espíritu de avanzar en disminuir los diferentes tipos de errores de una encuesta.

En la estandarización de estos análisis un papel importante lo representan los indicadores de calidad que evidencian y permiten monitorear el comportamiento de las variables observadas tanto del cuestionario mismo como aquellas del proceso de encuestaje.

Los indicadores que se sugieren para la evaluación de la calidad de datos de la primera ronda de la ELPS tienen relación con las tasas de respuesta y los datos del cuestionario base en las etapas de levantamiento y procesamiento de datos.

Uno de los objetivos que busca cumplirse en el desarrollo y resultados de estos indicadores es dejar un registro escrito de la evolución de estos elementos que pueden ser mejorados en las siguientes rondas de la ELPS.

A continuación se describen algunos indicadores de calidad para el proceso de levantamiento y del cuestionario base.

Indicadores de calidad de proceso de levantamiento:

a) Tasa de asignación de muestra de terreno por unidad área según cobertura:

Para la ejecución del levantamiento se realizan distribuciones de la muestra total N_t en X zonas o sedes.

En una zona i se tendrá una determinada asignación muestral $Nm(i)$, y a su vez se tendrá su correspondiente muestra asignada a personal de terreno $Nma(i)$.

La tasa de asignación de muestra en la sede (i) estará dada por $Nma(i)/Nm(i)$ y consecuentemente

la tasa de asignación de muestra total será:

$$\sum_{i=1}^x Nma(i)/Nt \text{ o } \sum_{i=1}^x Nma(i)/\sum_{i=1}^x Nm(i).$$

Estas tasas es necesario monitorearlas durante el levantamiento para que no se produzcan rezagos en la entrega que ocasionen un retraso o falta de entrevistas logradas finales.

b) Tasa de respuestas de visitas en muestra asignada por entrevistador:

En el trabajo de terreno un entrevistador tendrá asignada una muestra asignada Na para visitar y luego de los resultados de su trabajo tendrá una muestra asignada visitada Nav .

La tasa de muestra visitada por ese entrevistador estará dada por:

$$Nav/Na$$

Los valores Nav y Na serán acumulativos y la tasa estará sujeta a una fecha en que se evalúa su valor.

Esta tasa es conveniente observarla ya que refleja en el entrevistador, la postergación o abandono de visitas (al menos una vez) a la muestra asignada.

c) Tasas de Respuesta de visitas:

Las tasas de respuesta que se monitorean periódicamente son las tasas de respuesta definidas por la AAPOR en que previamente se definen:

EC = entrevista completa
EP = entrevista parcial
R = Rechazo y break-off
NC = Sin contacto
O = Otros
ED = Elegibilidad desconocida
NE = No elegible.

En nuestro caso se observarán las evoluciones de tasas en dos últimos etapas de la muestra tri-
etápica, es decir:

A nivel Hogar (USM Unidad Secundaria Muestral).

A nivel de persona seleccionada (UTM Unidad Terciaria Muestral).

La cuantificación de cada uno de estos elementos componentes de las tasas se obtendrá por la agregación de las frecuencias de los correspondientes códigos estipulados en las respectivas Hojas de Ruta del Hogar de la Persona (HRRNP).

Luego al interior de cada etapa se monitoreará por entrevistador; supervisores; zonas muestrales y global.

El detalle de cada tasa se describe a continuación.

Tasa de respuesta: Considera el porcentaje de encuestas completas en relación a la muestra de elegibilidad conocida más la desconocida.

$$\text{Tasa de respuesta: TTR} = \frac{(EC)}{(EC+EP)+(R+NC+O)+ED}$$

Tasa de Contacto: Considera el porcentaje de la muestra en que el entrevistador toma contacto o logra captar la atención para obtener un logro o un rechazo.

$$\text{Tasa de Contacto: TC} = \frac{(EC+EP)+(R+O)}{(EC+EP)+(R+NC+O)+ED}$$

Tasa de Rechazo: Considera la proporción de elementos de la muestra que rechazan completamente participar el estudio, medido sobre el total de casos potencialmente elegibles.

$$\text{Tasa de Rechazo: TR} = \frac{R}{(EC+EP)+(R+NC+O)+ED}$$

Tasa de Cooperación: Es la proporción de unidades muestrales entrevistadas sobre todas las unidades elegibles que fueron contactadas, tanto a nivel de hogar como de individuos.

$$\text{Tasa de Cooperación: TTC} = \frac{EC}{(EC+EP)+(R+NC+O)}$$

Tasa de No elegibilidad: Corresponde a la proporción de unidades no elegibles de la muestra, sobre la muestra elegible y elegibilidad desconocida y no elegible.

$$\text{Tasa de No Elegibilidad: TNE} = \frac{NE}{(EC+EP)+(R+NC+O)+(ED+NE)}$$

Donde, las elegibles son $(EC + EP) + (R + NC + O)$.

Indicadores de calidad de datos Cuestionario base

Estos indicadores representan resultados con los contenidos de información al interior del Cuestionario base de la ELPS. Los indicadores aquí obtenidos se sugiere acompañarlos en los análisis por otros indicadores equivalentes de estudios oficiales con resultados aprobados. Esto entrega información comparativa para alertar sobre la existencia de determinados sesgos que están ocurriendo en la formación de los estadísticos. (Ej. Una tasa conocida del porcentaje de población adscrita a algún sistema previsional nos entrega rangos en los cuales debieran comportarse las tasas obtenidas en este estudio. También en otras variables relevante como distribución por sexo y edades, etc.).

Los indicadores de calidad del cuestionario base se observarán en función del rendimiento del personal de trabajo, haciendo necesario establecer un conjunto de escalas: (1) Escala de entrevistador/crítico/digitador como unidad; (2) Escala de entrevistadores/crítico/digitador como promedio; (3) Escala de supervisor individual y (4) Escala de supervisor como promedio (Promedios generales).

Con estas escalas de medición será posible tener indicadores de desviaciones estándar y varianzas respecto de los promedios alcanzados entrevistadores/críticos/digitadores y/o supervisores. Estos

indicadores se obtienen luego de las etapas de crítica codificación y digitación por lo que pueden contener desviaciones con algunos sesgos de dichas etapas si estos últimos no fueron corregidos.

a) Tasas de entrevistados por rangos etáreos:

Esta corresponde a la proporción de la frecuencia para un rango etéreo de los entrevistados logrados respecto del total de entrevistados logrados. Una estratificación etérea adecuada es:

- edad_rango=1 Edad en variable a7 está en el intervalo [15 , 25)
- edad_rango=2 Edad en variable a7 está en el intervalo [25 , 35)
- edad_rango=3 Edad en variable a7 está en el intervalo [35 , 45)
- edad_rango=4 Edad en variable a7 está en el intervalo [45 , 55)
- edad_rango=5 Edad en variable a7 está en el intervalo [55 , 65)
- edad_rango=6 Edad en variable a7 está en el intervalo ≥ 65

Así se tendrá para el rango j que la tasa será:

$$\text{count}(\text{edad_rango} = j) / \sum_{x=1}^6 \text{count}(\text{edad_rango}_x)$$

b) Tasas de errores en distribución de variables relevantes:

En relación a los resultados dentro del estudio (con entrevistadores) se considera evaluar el promedio, para las 100 variables más relevantes de casos realizados con un 20% o más de valores perdidos y con un promedio de las variables numéricas que sea mayor o igual a dos desviaciones estándar. Así es posible distinguir entre tasa de valores perdidos y tasa de valores fuera del promedio.

Para la *tasa de valores perdidos* se calcula como la proporción de casos con esa advertencia sobre el total de casos aplicados por el encuestador.

$$\frac{C_{vm}}{C_{total}}$$

Donde,

C_{vm}= casos con más de un 20% de variables con problemas de valores perdidos

C_{total}= casos totales aplicados por el encuestador

Para la *tasa de valores fuera del promedio* se considera la proporción de casos con esa advertencia sobre el total de casos aplicados por el encuestador.

$$\frac{C_{fp}}{C_{total}}$$

Donde,

C_{fr}= casos con más de un 20% de variables fuera del promedio

C_{total}= casos totales aplicados por el encuestador

c) Tasas desviación según información de referencia:

En relación a resultados referidos desde otros estudios o informes oficiales, estos nos entregarán comparaciones preliminares con otras mediciones oficiales. En base a ello, es posible realizar la selección de un subconjunto de variables entre las 100 más relevantes que se vayan monitoreando según entrevistador y supervisor. Esta tasa se construye como la proporción de casos con diferencias significativas para el promedio o moda de alguna de las variables. Se pesquisará como error cuando un entrevistador o supervisor concentre más del 20% de sus variables por caso con diferencias de dos o más desviaciones estándares del promedio de referencia de otros estudios.

$$\frac{Cfr}{Ctotal}$$

Donde,

Cfp= casos con más de un 20% de variables fuera del promedio de referencia de otros estudios.

Ctotal= casos totales aplicados por el encuestado

Algunos comentarios sobre programación y análisis de la información base para la evaluación de la calidad de datos

El análisis de la información para la evaluación de la calidad de datos requiere establecer patrones de revisión estandarizados y homogéneos en todo su contenido, de tal forma de asegurar el mismo nivel de calidad en cada uno de sus registros.

Este análisis se ve fuertemente mejorado si se ejecuta en forma automatizada a cada una de las sub-bases primigenias u originales. Esta automatización permite en todo instante recrear o regenerar diferentes análisis con sus respectivos reportes, tomando como insumos las sub-bases originales y los correspondientes programas de análisis que a ellas les sean aplicados en forma pre-definida.

Sin duda el paquete estadístico de análisis Stata provee de potentes recursos para establecer esta programación a través de los llamados do-files. Los do-files permiten tomar como insumos las sub-bases originales y entregar resultados de análisis específicos según los requerimientos del estudio. Esto permite entender los análisis por terceros, replicarlos y mejorarlos para las siguientes rondas de la ELPS.

Junto a los registros originales obtenidos en la base de datos ELPS se debiera disponer de información administrativa estructurada en bases de datos y la base muestral que también serán inter-relacionadas mediante la programación de los do files pre-definidos.

A continuación se presenta una serie de recomendaciones para trabajar con los archivos de Stata en términos de formato y de estructura de trabajo identificando los insumos empleados para cada uno de ellos.

Nomenclatura de identificación de archivos do file

Al igual que las bases de datos los archivos do-files deben llevar el sufijo *_aammddvx* donde *aa mm dd* que representan respectivamente el año, mes y día en que fue modificado por última vez. El complemento *vx* representa la versión diaria de creación. Por ejemplo, de generarse tres do-file con modificaciones diferentes durante una misma fecha *_aammdd*, el primer do file generado tendrá la identificación de sufijo *_aammdd v1* ; el segundo *_aammddv2* y el último *_aammdd v3*.

Descripción resumida de contenidos de los archivos do file

Cada archivo do-file deberá llevar como nombre una descripción compacta de su contenido, a su vez en su interior en el encabezado o inicio deberá llevar información resumida de su contenido, responsables de su autoría y de los cambios realizados.

Estructura propuesta de programación de do-file

La estructura propuesta está orientada a tener una programación modular de los do-file asociados a la evaluación de calidad de datos en diferentes niveles, acompañados con una descripción objetiva de los insumos, procesos y resultados de cada una de sus partes y en su totalidad.

Ejemplo 1

Archivos de programación:

```
chkBaseEntrevistado_aammddvx.do
chkBaseMiembrosHogar_aammddvx.do
chkBaseHijos_aammddvx.do
chkBaseCarreras_aammddvx.do
chkBaseCapacitacion_aammddvx.do
chkBaseHlaboral_aammddvx.do
chkBaseCaracterHogar_aammddvx.do
```

Proceso para cada do-file :

Adiciona a la base insumo (una de las 8 sub bases del Cuestionario Base) variables auxiliares y de control de información para un análisis de consistencia inicial interna que implica en sí misma tener todas las variables iniciales con una distribución adecuada; saltos de flujo internos correctos; condiciones de universo correctamente cumplidas e inconsistencias internas entre variables completamente revisadas.

Nombre de producto de salida:

```
chkBaseEntrevistado_aammddvx.dta
chkBaseMiembrosHogar_aammddvx.dta
chkBaseHijos_aammddvx.dta
chkBaseCarreras_aammddvx.dta
chkBaseCapacitacion_aammddvx.dta
chkBaseHlaboral_aammddvx.dta
chkBaseCaracterHogar_aammddvx.dta
```

Ejemplo 2

Archivo de programación: chk_iinfenc_100var_aammddvx.do

Proceso: Une la base de datos de iinfencaammddvx.dta con la base de datos bbdd_100var_aammddvx.dta y analiza las 100 variables críticas de la ELPS a nivel de encuestador entregando variables auxiliares que dan cuenta de la calidad de los datos asociado a las personas que participaron en el trabajo de campo de la primera ronda de la ELPS

Nombre de producto de salida: chkBaseParadata100var_aammddvx.dat

Subproductos de salida: Calidad de datos por encuestador aammddvx.xls

Ejemplo 3

Archivo de programación: chk_BaseParadata_iinfenc_aammddvx.do

Insumos: BasesParadata_aammddvx.dta ; iinfencaammddvx.dta

Proceso(s): Genera base información para análisis cruzado de resultados de visita con información administrativa de los entrevistadores para su vinculación con tasas de respuesta.

Nombre de producto de salida: merge_paradata_iinfenc_aammddvx.dta

Subproductos de salida: Rendimientos y logros por encuestador aammddvx.xls

Ejemplo 4

Archivo de programación: do_chk_seleccion_kish_aammddvx.do

Insumo: MiembrosHogar_aammddvx.dta

Proceso: Adiciona a base MiembrosHogar_aammddvx.dta variables dummy que indican la ocurrencia errores en registro del número de personas seleccionables del hogar; ordenamiento de hombres y mujeres; ordenamiento en edades decreciente y selección de orden según Tabla Kish.

Nombre de producto de salida: chk_seleccion_kish_aammddvx.dta

Subproductos de salida: Lisado de errores de selección Kish por encuestador aammddvx.xls

Ejemplo 5

Archivo de programación: do_tasas_AAPOR_aammddvx.do

Insumo: merge_paradata_iinfenc_aammddvx.dta

Proceso: Adiciona a base merge_paradata_iinfenc_aammddvx.dta variables de tasas de respuesta; de cooperación; de rechazo y contacto; variables dummy que indican la ocurrencia errores en registro del número de personas seleccionables del hogar; ordenamiento de hombres y mujeres; ordenamiento en edades decreciente y selección de orden según Tabla Kish.

Nombre de producto: tasas_AAPOR_aammddvx.dta

Subproductos de salida: Tasas AAPOR por encuestador aammddvx.xls

4. Recomendaciones para reducir el error de no respuesta en las rondas de seguimiento

4.1 LA ATRICIÓN DE LA MUESTRA

4.1.1 Conceptos y definiciones

La no respuesta ocurre cuando una unidad de la muestra no responde a la solicitud de ser encuestada o no responde preguntas específicas del cuestionario. En esta sección solo se hará referencia al caso de no lograr encuestar a la persona seleccionada para un estudio determinado. Al igual que las encuestas de corte transversal, las encuestas longitudinales se ven afectadas por el error de la no respuesta de los miembros de la muestra. Sin embargo, una característica distintiva de las encuestas longitudinales es que la no respuesta - también llamada atrición o desgaste de la muestra - no es un evento único y tiende a acumularse en el tiempo con cada ronda de entrevistas.

El problema de una alta atrición es doble. Por un lado, el menor número de observaciones disminuye la eficiencia de los estimadores, amenazando la continuidad de un panel, especialmente si la muestra inicial fue relativamente pequeña. Por otra parte, si las características de los individuos que no vuelven a ser encuestados difieren sistemáticamente de las características de los que sí vuelven a serlo, entonces la atrición es selectiva. Al ser la atrición selectiva las estimaciones pueden resultar sesgadas y no se estaría representando a la población inicial (Watson & Wooden, 2009).

De esta manera los responsables del levantamiento de la encuesta debieran tener la presión de diseñar e implementar procedimientos en el trabajo de campo que busquen maximizar la retención de la muestra (Laurie et al., 1999), y si por algún motivo una encuesta no es realizable, entregar la mayor información posible sobre el individuo que no respondió la encuesta para ayudar a los analistas de los datos a realizar inferencias al enfrentarse a la información faltante (LaRoche, 2003). Lograr estos dos objetivos requiere de un buen conocimiento de los procesos que conllevan una respuesta (o encuesta realizada) y los factores que provocan un aumento de la atrición de la muestra (Falaris, 2003; Fitzgerald et al., 1998; Paredes et al., 2006; Uhrig, 2008; Watson, 2003).

4.1.2 Determinantes de la no respuesta y la atrición

La no respuesta se explica por varios motivos. La muerte de una persona que era parte de la muestra es un tipo de no respuesta, y sería una causa exógena a la agencia que levanta los datos. Hay situaciones donde las tasas de no respuesta si depende de los protocolos de seguimiento y de contacto definidos por la entidad que realiza el trabajo de campo. Por ejemplo, seguir o no seguir a un individuo que cambió de residencia (Hill, 2004)⁵, el número y horario de los intentos para contactar a la persona una vez localizada (Groves & Couper, 1998), y la duración del trabajo de campo (Nicoletti & Peracchi, 2005) son decisiones que tienen incidencia directa en los niveles de respuesta de una encuesta. Otros motivos que explican la no respuesta es la disposición que tienen las personas a cooperar una vez localizadas y contactadas. Esta cooperación está sujeta a factores como el uso de incentivos en la entrevista (Laurie & Lynn, 2009), la experiencia del encuestador y

⁵ Hill (2004) al comparar el nivel de no respuesta en 13 encuestas paneles de países en desarrollo estima que la atrición anual se encontró en un rango de 9 y 21% cuando no hubo seguimiento a las personas que se cambiaron de residencia. Cuando se realizó el esfuerzo de localizar a quienes se habían movido después de ser entrevistados la atrición anual bajó entre 1 y 8%.

si fue el mismo encuestador quien hizo la entrevista en la ronda anterior (Nicoletti & Buck, 2004), las características del entrevistado (Watson & Wooden, 2009), y el cansancio que pueden tener los entrevistados por participar en una encuesta que se repite cada cierto tiempo (Laurie et al., 1999). Cabe mencionar un análisis del contacto de la Encuesta Panel de Hogar de la Comunidad Europea realizado por Nicoletti y Peracchi (2005). Estos autores encuentran que la probabilidad de no contactar a una persona en la ronda siguiente es directamente proporcional al número de intentos que tuvo en la ronda anterior. Por ejemplo, la probabilidad de contactar a una persona que se visitó tres veces antes de hacer la entrevista es menor que una persona con similares características que se le aplicó la encuesta en la primera visita.

Hay instrumentos similares que se aplican en distintos países como es el caso de la EPLS. Esta situación permite analizar la no respuesta controlando por el cuestionario. Un estudio realizado por De Heer (1999) encuentra tres factores que explican los diferentes niveles de respuesta en la Encuesta de Fuerza Laboral (*Labor Force Survey*) que se realiza en 16 países. En primer lugar, factores del diseño. Por ejemplo, si el marco muestral fue obtenido de datos administrativos como en la ELPS Chile o de las unidades geográficas del último censo como en la ELPS de los otros países. Otro ejemplo de factor de diseño es si se recogen los datos en computador o en cuestionarios en papel y lápiz. En segundo lugar, las estrategias de trabajo de campo. Por ejemplo, el número de intentos de contacto, el protocolo de conversión de rechazos, el modo de agendar citas para la entrevista, el incentivos para el entrevistado y para el encuestador. Por último, los factores relacionados a la organización de la encuesta. Por ejemplo, si la encuesta es voluntaria u obligatoria y el tipo de contrato de los encuestadores.

En el caso de las encuestas longitudinales, Lepkowski y Couper (2002) sistematizan la comprensión de la no respuesta identificando tres procesos que están condicionados entre sí. De esta manera, el proceso de la no respuesta en las rondas de seguimiento se divide en: i) localización del entrevistado, ii) contacto del entrevistado una vez localizado, iii) cooperación del entrevistado una vez contactado. Hay consenso tanto en los países desarrollados como en los países en vías de desarrollo que el primer proceso en el esquema de Lepkowski y Couper (2002) es el que presenta un mayor desafío a las agencias que levantan los datos en las rondas de seguimiento. Por ejemplo, Laurie et al. (1999) encuentra en las primeras rondas del panel británico que uno de sus principales problemas a enfrentar está en la localización de los miembros de la muestra que se mueven. Para las encuestas longitudinales en países en desarrollo con procesos de migración rural-urbano aún vigentes la localización de las miembros del hogar que se van a vivir a otros lugares es el componente mayor de no respuesta cuando no se toman las medidas apropiadas para localizar al entrevistado (Alderman et al., 2001; Thomas et al., 2012).

Dado que la localización del entrevistado que cambia de residencia⁶ entre una ronda y otra impone una dificultad adicional a las agencias que levantan encuestas paneles es importante conocer los niveles de movilidad residencial en Latinoamérica. La tasa de movilidad residencial o migración interna en el continente ha mostrado una disminución entre los censos de los años 1990 y 2000. Para el periodo 1990 y 1995 la movilidad promedio fue de 5.1 por ciento y para el periodo 1995 y 2000 fue de un 4 por ciento (UN, 2011). Para el periodo 1995 y 2000 los países que presentaron una menor movilidad (entre el 3 y 4 por ciento) fueron Brasil, Guatemala y Nicaragua. Los que tuvieron una migración interna entre el 4 y 6 por ciento fueron Chile, Colombia, Ecuador y México. Los países que mostraron una mayor movilidad para el periodo estudiado fueron Paraguay y Uruguay

⁶ En general, las personas se mueven por razones relacionadas con la casa, los familiares, y el trabajo o educación. La distribución de estas tres razones varía a nivel de país y al interior de las regiones de cada país.

con una tasa cercana al 8 y 7 por ciento respectivamente. Al comparar la migración interna con países de otros continentes encontramos niveles similares o mayores a los de Paraguay y Uruguay. En el caso de Estados Unidos para el mismo periodo analizado es aproximadamente 9 por ciento (UN, 2011). Países europeos como Reino Unido, Francia y Holanda tuvieron una movilidad anual entre los años 1996 y 1997 de 8,4, 8,7 y 8,1 por ciento (Gregg et al., 2004). En el continente asiático, el país de Indonesia tuvo una movilidad residencial de un 12 por ciento entre 1993 y 1997 (Thomas et al., 2001).

4.1.3 Estrategias de seguimiento de las personas que se mueven en una encuesta panel

En los países desarrollados una de las estrategias de seguimiento de los miembros de la muestra que cambian de residencia es buscar las nuevas direcciones en directorios de teléfonos, datos administrativos obtenidos de instituciones privadas y públicas (Couper & Ofstedal, 2009). En los países en vías de desarrollo, en general, esta información no está disponible, por lo tanto, localizar a los entrevistados que se mueven presenta un gran desafío metodológico y económico para las agencias que realizan las entrevistas en esos países (Falaris, 2003, Hill, 2004). Sin embargo, hay experiencias que han sido exitosas en reducir la no respuesta en el proceso de localizar a los que se han movido entre una ronda y otra, a partir de la información de contacto recogida en la primera ronda (Thomas et al., 2001)⁷. En esta misma línea, Laurie et al. (1999) muestra en su estudio que los que entregaron una información de contacto al entrevistador en la primera ronda tuvieron una probabilidad más alta de continuar en la muestra a lo largo de las rondas. Esta estrategia de seguimiento es una técnica proactiva que implica un diseño y planificación anterior al trabajo de campo. Por ejemplo, obtener información de una o dos personas de contacto en el momento de la entrevista, registrar información que de indicios que el entrevistado se moverá a otro lugar (preguntarle directamente si espera hacerlo, información de otras residencias relacionadas con el entrevistado, experiencia de un seguimiento exitoso anterior) o mantener contacto con el encuestado entre rondas a través de una visita o llamado de teléfono. Una tercera estrategia de seguimiento se puede realizar durante la ejecución del trabajo de campo (Couper & Ofstedal, 2009). Por ejemplo, obtener la nueva dirección de los nuevos residentes de la vivienda o de los vecinos. En el caso que los nuevos residentes o vecinos saben dónde vive el encuestado pero no quieren dar la dirección deja una carta para que se la entreguen al entrevistado.

4.1.4 Lecturas sugeridas

- Groves, Dillman, Eltinge y Little, editores del libro *Survey Nonresponse* (2002), agruparon 29 artículos en cuatro temas: i) perspectivas de la no respuesta, que entrega varios marcos conceptuales de gran utilidad para entender como los investigadores en metodología de encuestas pueden aproximarse al problema de la no respuesta, ii) impacto del diseño de la encuesta en la no respuesta, incluye algunas reflexiones y buenas prácticas en los protocolos de encuestaje y población objetivo que están relacionados con las tasas de respuesta iii) la no respuesta en diversos tipos de encuestas, donde hay un trabajo que da cuenta de la no respuesta en la segunda ronda de una encuesta panel, e iv) inferencia estadística en el tratamiento de la no respuesta, que consiste básicamente en una descripción de los pesos, imputaciones, sustituciones y técnicas de estimación que intentan reducir los efectos de la no respuesta en la inferencia de los parámetros de la población estudiada. Para efectos de esta sección se sugiere

⁷ Indonesia cuenta con una encuesta panel (*Indonesia Family Life Survey* (IFLS)) que ha registrado altos niveles de participación pese a no ser un país desarrollado, tener una geografía territorial compleja y una movilidad residencial por sobre el 10 por ciento en cinco años. Por ejemplo, la tasa de no respuesta de la IFLS en la segunda ronda fue de solo un 6 por ciento.

la lectura de los trabajos que están en los dos primeros temas, y particularmente el trabajo de Lepkowski y Couper (2002) en el tercer tema.

- El libro *Nonresponse in Household Interview Surveys* (1998) escrito por Groves y Couper analiza el cruce de datos entre quienes respondieron y no respondieron las encuestas de hogares más importante de Estados Unidos y la información del censo de 1990. De esta manera, la riqueza de los datos que les permitió a Groves y Couper construir un modelo teórico y práctico para la comprensión de la no respuesta. De este modelo se desprende un conjunto de estrategias para reducir la no respuesta y ajustarla en los cálculos de los factores de expansión. Los autores también profundizan en los aspectos que determinan que una persona decida participar en una encuesta. Por ejemplo, estudian las influencias de las características del hogar, los factores sociales ambientales que se dan entre la interacción de los entrevistadores y los miembros del hogar como también aspectos del diseño de la encuesta.
- El reporte de Witoelar (2011) llamado *Tracking in Longitudinal Household Surveys* entrega una serie de recomendaciones para el seguimiento de la muestra en encuestas longitudinales de países en vías de desarrollo. Este documento preparado para el *World Bank Development Research Group* es un manual de buenas prácticas para minimizar atrición de la muestra. La mayoría de las sugerencias que se proponen en la sección 4.2 "Propuesta para el seguimiento de la muestra en las próximas rondas" provienen de este reporte.

4.1.5 Recomendaciones específicas

Recomendación 11: Crear una base de datos antes de comenzar el trabajo de campo que contenga toda la información que podría ser importante para localizar a un miembro de la muestra que se cambió de residencia. Esa información corresponde a los contactos que entregó el entrevistado en la primera ronda o la que se fue recibiendo entre las rondas producto de algún contacto telefónico u otro mecanismo. Además se sugiere analizar esa base de datos para comprender por qué no se dispone de esa información. Los casos sin información de contacto serán difícil de localizar si la persona ha cambiado de residencia entre una ronda y otra.

Recomendación 12: De contar para la segunda ronda con los mismos encuestadores que participaron en la primera ronda asignar a cada encuestador las viviendas que realizó en la ronda anterior. Además, utilizando el insumo generado en la recomendación 9, entregar a cada encuestador el día de la semana y horario de cuando la entrevista fue realizada para que considere en la planificación de sus visitas esa información. Como se señaló más arriba, existe evidencia empírica que tomando estas dos simples medidas se aumenta la probabilidad de contactar y lograr entrevistar a la persona en la siguiente medición.

Recomendación 13: Analizar la Historia de Contacto⁸ marcando aquellos casos que se logró la entrevista después de varios intentos. Hay estudios que demuestran que las encuestas realizadas después de varios intentos tienen una mayor probabilidad de no responder la encuesta en la próxima ronda. Por esta razón es importante saber con anticipación cuáles personas podrían presentar una mayor probabilidad de no contactarla o rechazar participar nuevamente en el estudio para así tomar las medidas necesarias antes del primer contacto.

⁸ Se entiende como Historia de Contacto, la fecha y horario de cada uno de los intentos que realiza el encuestador, primero para aplicar la tabla de Kish en la vivienda de la muestra y luego para contactar a la persona que fue seleccionada para participar en la ELPS.

4.2 PROPUESTA PARA EL SEGUIMIENTO DE LA MUESTRA EN LAS PRÓXIMAS RONDAS

Recolectar información para el re-contacto.

Ponerse en contacto con o en re-contacto con un miembro de la muestra en la segunda y en las olas posteriores de una encuesta de panel es ligeramente diferente a contactar con ellos en la primera etapa. Es importante considerar que el entrevistador está buscando una persona específica y no sólo a la persona que está viviendo en ese momento en la vivienda por lo que si el miembro de la muestra se ha mudado de casa se debe tratar de localizarlo y entrevistarlo en su nueva dirección.

- Entre las rondas de la encuesta, se sugiere mantener contacto con el entrevistado mediante cartas de agradecimiento, tarjetas de felicitación de cumpleaños, o tarjetas de cambio de dirección.
- Se recomienda hacer un primer contacto vía telefónica antes del levantamiento de la encuesta para saber cuándo son los mejores momentos o días para ir a la visita.
- Se sugiere recolectar datos de contacto del entrevistado pero también de alguien fuera del hogar al final de cada entrevista.
- Se recomienda pedir el teléfono de donde trabaja la persona, o donde pasa la mayor parte del tiempo en el día.

Se sugiere añadir algunas preguntas como en la tabla 1. Las variables de la tabla se llaman RC01 y RC02 (Re-Contacto 01 y 02).

Tabla 1. Información para el re-contacto.

<p>RC01. Si usted o algún miembro de su familia se llegará a cambiar de casa, ¿a quién le avisaría o quién nos podría dar información acerca de su paradero (algún pariente, amigo o familiar muy cercano)?</p> <p>1. Especifique 3. Misma Loc. / Col. / Mpio. / Del. / Edo. / País del encuestado 8. NS</p>	<p>REFERENCIA 1 (Alguien ajeno al hogar)</p> <p>1. Nombre y apellido: _____</p> <p>1. Parentesco: _____</p> <p>1. Calle: _____ 3. Mismo 8.NS</p> <p>1. Localidad/Colonia: _____ 3. Mismo 8.NS</p> <p>1. Municipio/ Delegación: _____ 3. Mismo 8.NS</p> <p>1. Referencia: _____</p> <p>1. Estado: _____ 3. Mismo 8.NS</p> <p>1. País: _____ 3. Mismo 8.NS</p> <p>1. Teléfono: [] [] [] [] [] - [] [] [] [] [] [] [] [] [] [] número ext. [] [] [] [] 2. Propio 3. Comunitario</p>
<p>RC02. ¿Me podría dar el nombre de algún otro pariente, amigo o familiar muy cercano que sepa de usted si se llegará a cambiar de casa?</p> <p>1. Especifique 2. Misma Loc. / Col. / Mpio. / Del. / Edo. / País del encuestado 8. NS</p>	<p>REFERENCIA 2 (Alguien ajeno al hogar)</p> <p>1. Nombre y apellido: _____</p> <p>1. Parentesco: _____</p> <p>1. Calle: _____ 3. Mismo 8.NS</p> <p>1. Localidad/Colonia: _____ 3. Mismo 8.NS</p> <p>1. Municipio/ Delegación: _____ 3. Mismo 8.NS</p> <p>1. Referencia: _____</p> <p>1. Estado: _____ 3. Mismo 8.NS</p> <p>1. País: _____ 3. Mismo 8.NS</p> <p>1. Teléfono: [] [] [] [] [] - [] [] [] [] [] [] [] [] [] [] número ext. [] [] [] [] 2. Propio 3. Comunitario</p>

Fuente: ENNVIIH, México 2005

Los resultados de las investigaciones recientes sugieren que las encuestas de hogares longitudinales deben considerar seriamente el seguimiento de los encuestados que se mueven, dentro de la comunidad y fuera de la localidad de referencia, lo que puede acarrear costos significativos. Como los encuestados se mueven y dejan su hogar línea de base, la muestra restante será menos

representativa de la muestra original, en particular ya que la migración es probablemente no-aleatoria. Para esto se deben establecer reglas de los individuos a seguir dependiendo de los objetivos de la encuesta: por ejemplo, variables como edad y género, pueden tener prioridad.

También se debe decidir qué tan lejos seguir a los entrevistados que se han mudado para poder tener en control los costos esto. Se pueden establecer fronteras administrativas, por ejemplo Witoelar (2011) recomienda el seguimiento de los individuos en tres fases para organizarlo mejor:

Pre-levantamiento: Esta fase implica estimar la potencial migración de la muestra con la información recopilada en la línea de base (si los entrevistados declararon que se iban a mudar, por ejemplo). Esto permite incluso crear algunos mapas o direcciones alternativas para que los encuestadores vayan preparados en el levantamiento.

- Se recomienda también que durante las capacitaciones de los encuestadores se enseñe cuál será el protocolo de seguimiento de los individuos.
- Se recomienda que los encuestadores y supervisores estén bien entrenados sobre el seguimiento y sepan cual son las implicaciones y beneficios de encontrar a las personas, y sobre todo el por qué es importante encontrar a las personas que se mudan.

Levantamiento: Si el entrevistado se ha movido a una zona diferente de dónde es la ruta que corresponde al equipo de encuestadores es importante que haya comunicación para que este entrevistado sea entrevistado por el equipo correspondiente a su nueva localización. En caso de que los entrevistados que se han movido no están dentro de la zona de la encuesta, se puede estimar un periodo especial para el seguimiento de estos individuos una vez haya terminado el levantamiento.

- Se recomienda tener a un **encargado de seguimiento** durante el levantamiento que pueda ayudar a coordinar estos casos.
- Se recomienda considerar tener un periodo especial para el seguimiento de algunos individuos después del levantamiento.

La experiencia en diversas encuestas longitudinales ha demostrado que es fundamental contar con una sola persona o un pequeño equipo asignado a la tarea de monitoreo de desgaste de la muestra durante el trabajo de campo principal (levantamiento) para la organización de las actividades de seguimiento. Este encargado deberá:

1. Seguir el protocolo de seguimiento y decidir quiénes de los que se han mudado tienen prioridad de ser encontrado, considerando costos y objetivos de la encuesta;
2. Resolver información contradictoria que pueda surgir (por ejemplo, información de la misma persona de distintos hogares);
3. Evaluar la calidad de la información para el seguimiento y así saber la probabilidad de encontrar a esa persona.
4. Organizar y priorizar los casos por región.
5. Asignar los casos de seguimiento a los equipos
6. Mantener un registro de los seguimientos
7. Monitorear la cantidad de mudanzas y migraciones

Post-levantamiento: Se recomienda generar estadísticas de los resultados obtenidos para poder medir la potencial migración para las siguientes rondas y respecto a esto, ajustar las medidas de re-contacto entre olas.

Protocolo de seguimiento

Durante el levantamiento, cada vez que los entrevistadores no pueden localizar a un entrevistado de seguimiento en la vivienda de referencia, se debe seguir un protocolo de seguimiento (Witoelar 2011). El protocolo de seguimiento se inicia con el entrevistador completando un “Formulario de seguimiento” con la información necesaria para encontrar a las personas que se mudan. Sin embargo, antes de concluir que el individuo se ha movido, el equipo debe asegurarse de que están en la vivienda correcta. Dirección de la calle, notas, bocetos acerca de la ubicación recogidos en rondas anteriores deben ser consultadas.

Habiendo asegurado que están en la vivienda correcta, los entrevistadores tendrán que identificar si toda la familia (todos los miembros de la línea de base) o cualquier miembro del hogar se ha movido. Los entrevistadores deben ir a la casa preparados con lista con la información de los hogares. Las listas precargadas / preimpresos deben enumerar todos los que estaban previamente en la lista del hogar. La información básica, como el sexo, edad, fecha de nacimiento y relación con el jefe debe ser incluido. Ver secciones 1 a 4 de este documento.

Formas para los individuos de seguimiento: El entrevistador debe tratar de obtener múltiples informantes confiables. Por lo tanto, es muy útil contar con una lista de posibles informantes en la encuesta de referencia que podrían proporcionar información sobre el paradero del entrevistado. Esta información de los informantes potenciales también debe estar en posesión del entrevistador durante la segunda encuesta. Esto debe ser una práctica estándar en cada ronda.

Las formas para los individuos de seguimiento deben incluir:

1. Nueva dirección
2. Teléfonos de casa, trabajo, familiares
3. Ruta y forma de llegar a la nueva dirección
4. Nuevo y viejo trabajo o lugar donde pasa la mayor parte del tiempo
5. Información sobre alguien cerca de la nueva dirección o en su nuevo o viejo trabajo
6. Información sobre alguien en la dirección vieja que pueda tener más información sobre su paradero.
7. La fecha en la cual se mudó
8. La razón por la que se fue
9. Su grado de educación más alto
10. Su estado marital
11. Su estado civil

Esta información deberá ser evaluada por el encargado del seguimiento para ver si es suficiente para poder encontrar a la persona.

5. Recomendaciones para reducir el error de la medición en las rondas de seguimiento

5.1 EL ERROR DE MEDICIÓN EN ENCUESTAS PANEL

El error de medición hace referencia a la posibilidad que el verdadero valor de una respuesta pueda ser distinto al valor recogido en la encuesta. Dado que en las encuestas longitudinales es de central interés estimar cambios en el tiempo cualquier inconsistencias de las mediciones puede contribuir a obtener resultados sesgados (Lynn, 2009). En las encuestas panel no solo el efecto del entrevistador, del fraseo de las preguntas o del modo de recolección de los datos contribuyen al error de medición (Biemer & Lyberg, 2003). El efecto memoria del entrevistado para reportar eventos pasados también aporta al error de medición. Los problemas más conocidos en este ámbito son el efecto telescopio y el efecto juntura.

El efecto telescopio o *telescoping* refiere a la dificultad que tienen las personas para precisar las fechas de ocurrencia de los eventos que recuerda. Neter y Waksberg (1964) acuñaron el concepto cuando se dieron cuenta en su investigación que los que responden tienden a reportar eventos como si hubiesen ocurrido hace menos tiempo. El uso del concepto *telescoping* sugiere que los eventos tienden a percibirse más cerca de lo que realmente están.

El efecto juntura o *seam effect* aparece cuando los entrevistados ubican eventos dentro del período de referencia de la ronda siendo que realmente sucedieron en otro periodo tiempo (Jackle & Lynn, 2004). Por ejemplo, un entrevistado puede informar en una ronda que no tiene trabajo en ese momento, pero doce meses después, en la siguiente medición, informa que lleva doce meses trabajando, y que comenzó a trabajar coincidentemente en el inicio del periodo de referencia. Otro ejemplo ocurre cuando el entrevistado reporta una fuente de ingreso permanente durante el periodo de referencia de la entrevista. Sin embargo, en una medición no entregó esa información. Al observar la información de todas las rondas se observará una disminución de la fuente de ingreso que calza con el periodo de la ronda donde omitió la información.

5.2 RECOMENDACIÓN PARA REDUCIR EL EFECTO MEMORIA EN ENCUESTAS PANEL

5.2.1 Uso de entrevistas dependientes

Las entrevistas dependientes buscan obtener el máximo beneficio de la capacidad que tienen las encuestas longitudinales, y a su vez disminuir los errores de medición. En las entrevistas dependientes el contenido de las preguntas varía entre los miembros de la muestra, dependiendo de las respuestas que la persona haya entregado en la ronda previa. Por ejemplo, se puede intentar actualizar la información recogida en una ronda anterior presentándole al miembro de la muestra esa información y pidiéndole que confirme si las circunstancias han cambiado o no (entrevista dependiente) en vez de simplemente pedirle que indique sus circunstancias actuales (entrevista independiente). Es decir, en la entrevista dependiente se utiliza información de la entrevista anterior para determinar qué pregunta se debe formular y para orientar su contenido. Mientras que la entrevista independiente tradicional se caracteriza por preguntar a las personas por su situación en diferentes puntos del tiempo, sin hacer referencia a las respuestas anteriores.

Algunos beneficios que se obtienen al utilizar las entrevistas dependientes son:

- Reduce el tiempo de entrevista y mejora del proceso de ésta (las personas sólo indican si algo ha cambiado con respecto al año anterior en vez de reconstruir respuestas completas), lo que reduce costos.

- Disminuye la atrición, ya que las personas no sienten que les están haciendo las mismas preguntas que la ronda pasada, con lo que se pueden sentir aburridos o irritados. Cuando hay poca motivación puede llevar a una reducción en la calidad de la respuesta lo cual a su vez puede llevar a que los encuestados no quieran ser entrevistados de nuevo.
- Reduce las omisiones y clasificaciones incorrectas mediante la entrega de una ayuda de memoria que sirve para refresca las respuestas anteriores.
- Puede reducir errores en el reporte de ingresos o información relacionada al empleo, ya que la entrevista dependiente al otorgar información de la entrevista anterior reduce el subreporte de fuentes de ingresos, sobre todo las que son más fáciles de olvidar.
- Reduce los cambios espurios en las características del empleo a través de las olas. Estos cambios espurios se dan aún más en los trabajos complejos, cuyas características son difíciles de explicar y por lo mismo es más probable que se reporte de manera inconsistente a través de las entrevistas.

Sin embargo, a pesar de que la entrevista dependiente reduce los errores que conducen al efecto juntura, no los elimina completamente:

- A pesar que la entrevista dependiente mejora el reporte de una fuente de ingresos durante el período de referencia, no mejora el reporte de las fechas de los cambios en la recepción de ingresos. Los errores de medición pueden ser causa de intentar reconstruir un pasado complejo utilizando reglas simples. Si eso es cierto, las medidas que se usan para reducir las dificultades del recuerdo no prevendrán el efecto juntura.
- Se puede dar que los individuos no hagan el esfuerzo cognitivo requerido para responder bien las respuestas, y simplemente confirmen la información previa, pero tampoco está comprobado que el control que ejerce este tipo de entrevista se deba a ese comportamiento conformista.
- Para entrevistas con lápiz y papel es trabajoso y susceptible de error alimentar los datos de la encuesta con información anterior, debido a que el peso de buscar la información relevante y la responsabilidad de corregir el contenido de la pregunta correctamente recae sobre el entrevistador. Por esta razón, este proceso es utilizado solo cuando existen métodos computacionales para realizar la encuesta.

Existen dos importantes métodos para aplicar entrevistas dependientes, a continuación una breve descripción de cada una.

Entrevista Dependiente Proactiva (PDI):

En ella, la información de la entrevista anterior se ofrece proactivamente como parte del proceso de interrogación. Por ejemplo, durante la encuesta sobre población *US Current Population Survey*, (CPS), a los entrevistados se les recuerda la empresa en la cual habían informado que trabajaban en la entrevista del trimestre anterior y se les pregunta si siguen trabajando en la misma empresa. Si el entrevistado informa que no ha habido cambios en sus actividades o deberes, entonces se lee una descripción de sus actividades y deberes y se le pide que confirme si esa información todavía es válida. Si es así, se asume que su ocupación sigue siendo la misma.

Se ha comprobado que el uso de estas preguntas dependientes reduce enormemente el cambio aparente (que se pensaba como dudoso) y también se hace cargo de las quejas de los entrevistados en cuanto a la repetitividad de las preguntas.

Ejemplo de pregunta proactivas:

“De acuerdo con nuestros informes, cuando le entrevistamos la última vez, el día (fecha), usted estaba recibiendo (fuente) usted sólo o conjuntamente. ¿Durante cuántos meses, desde entonces, ha recibido (fuente)? Luego se mostraba cuatro tarjetas con posibles fuentes de ingresos y se preguntaba si había recibido cualquier otro tipo de ingresos enumerados.

Entrevista Dependiente Reactiva (RDI):

En ella se ofrece la información sobre la entrevista anterior sólo como reacción ante ciertas respuestas. Por ejemplo, en la encuesta *Canadian Survey of Labour and Income Dynamics* (SLID) se usa la RDI para los datos sobre los sueldos: Si el entrevistado informa sobre una cantidad que es, ya sea inferior a la cantidad de la entrevista del año anterior, o superior en un 10%, entonces aparece un recuadro en la pantalla CAPI mostrando ambas cantidades e informando al entrevistador que debe averiguar y anotar las razones de esa diferencia.

Ejemplo de pregunta reactivas:

Primero se hace la pregunta estándar de entrevista independiente mostrando las 4 tarjetas con las listas de posibles fuentes de ingresos: “¿Usted ha recibido algún tipo de ingreso o pago desde la entrevista anterior?”. Luego, para cada fuente que había sido mencionada en la entrevista anterior, pero no en la actual se pregunta: “¿Puedo verificar algo?, De acuerdo con nuestra información, usted ha recibido en el pasado (fuente). ¿Ha recibido (fuente) en algún momento desde (fecha)?

5.2.2 Uso de calendario de la historia de los eventos

Una manera complementaria de reducir error de medición debido a la memoria es el uso de Calendario de la Historia de los Eventos o *Event History Calendar* (EHC). La idea es usarlos para que el entrevistado tenga tiempo de pensar en las circunstancias que rodeaban el evento en cuestión, además de otros hitos que puedan haber acontecido en esas fechas. Estos calendarios ayudan a formular algunas preguntas de “contexto”, para orientar el proceso de pensamiento del entrevistado, aunque las respuestas a esas preguntas puedan no ser necesarias para los fines del análisis. Concretamente este método se sugiere aplicar para el módulo de Historia Laboral del cuestionario, en que la ubicación de ciertos hitos en el tiempo se utilizan como guía para recodar la situación ocupacional del individuo durante los últimos meses a la fecha de la aplicación de la encuesta.

Este calendario además de servir como ayuda a la memoria puede facilitar verificaciones en el momento de editar las entrevistas. Las prestaciones visuales permiten que el entrevistador pueda identificar contradicciones en los registros de las respuestas. Por ejemplo, un registro simultáneo de situaciones que deberían ser mutuamente exclusivas. El entrevistador puede entonces investigar y resolver estas contradicciones con el entrevistado antes de pasar a la siguiente sección de la entrevista. Ello refuerza más la calidad de los datos resultantes.

A continuación se muestra el calendario utilizado para complementar el registro de la historia laboral de la tercera ronda de la Panel CASEN 1996-2001-2006. En esta encuesta la extensión de tiempo entre una ronda y otra obligaba a diseñar mecanismos que ayudarán a la memoria de los entrevistados. Los hitos que se preguntaban en primer lugar era cambio en la familia (por ejemplo si había nacido un niño o niña), problemas de salud (por ejemplo alguna enfermedad o ida al hospital), crisis económica y cambio de trabajo.

MATRIZ Historial Laboral. Jefe de Hogar

MATRIZ 2006-2001		Año	dic	nov	oct	sep	ago	jul	jun	may	abr	mar	feb	ene	Año	dic	nov	oct	sep	ago	jul	jun	may	abr	mar	feb	ene
H I T O S	Cambio en familia																										
	Problema de Salud																										
	Crisis Económica	2													2												
	Cambio de Trabajo	0													0												
	Trabajo Principal	0													0												
S T A T U S	Trabajo Secundario o pololo	6													5												
	Estaba Cesante																										
	Buscaba Trabajo por 1ª vez																										
	Era Inactivo																										
H I T O S	Cambio en familia																										
	Problema de Salud																										
	Crisis Económica	2													2												
	Cambio de Trabajo	0													0												
	Trabajo Principal	0													0												
S T A T U S	Trabajo Secundario o pololo	4													3												
	Estaba Cesante																										
	Buscaba Trabajo por 1ª vez																										
	Era Inactivo																										
H I T O S	Cambio en familia																										
	Problema de Salud																										
	Crisis Económica	2													2												
	Cambio de Trabajo	0													0												
	Trabajo Principal	0													0												
S T A T U S	Trabajo Secundario o pololo	2													1												
	Estaba Cesante																										
	Buscaba Trabajo por 1ª vez																										
	Era Inactivo																										

(Encuestador) Una vez terminado el llenado de la MATRIZ, vaya a la Hoja de Ruta y revise cuál era la situación laboral del Jefe de hogar en el año 2001. Con esta información, haga la pregunta 6b.

5.3 PROPUESTA DE CAMBIOS PARA EL CUESTIONARIO DE SEGUIMIENTO

1. Verificación de los datos

Es importante llevar impresa o precargada en las *tablets* o *notebooks* la información completa del individuo y de todos los integrantes del hogar para hacer una verificación de nombres, añadir personas nuevas y así verificar la lista completa. En caso de que el cuestionario sea en papel deberá llevarse una lista impresa con la información del hogar. Además es importante verificar que la dirección sea la correcta, anotar observaciones en caso algo haya sido modificado en la vivienda.

2. Identificador único de las personas del hogar

Para identificar de una forma eficiente al entrevistado y a los miembros del hogar a través de las rondas de la encuesta se sugiere usar una numeración única que no se modifique en las rondas subsiguientes. Un ejemplo para realizar esto de forma eficiente es:

Numerar a los miembros de la primera ronda (o línea de base) 1, 2, 3,...n. Donde el entrevistado deberá ser identificado además con el número 1. De esta forma si hay 4 integrantes del hogar 1 jefe, 2 esposa, 3 hijo, y 4 hija y el entrevistado es la esposa, su identificador único será 21.

En la segunda ronda, se recomienda que a los miembros nuevos del hogar se les identifique como 101, 102,... n.

En la tercera ronda, como 201, 202, 203...n.

Se recomienda que esta variable se nombre PID, por sus siglas en inglés *personal identification*, esto ya que es una práctica común en las encuestas de todos los países.

Para crear un identificador único de las personas del hogar y del entrevistado al PID se le agrega el identificador único del hogar que puede ser el folio de la entrevista, siempre y cuando este se mantenga siempre.

También es posible crear un identificador único agregando las variables geográficas del hogar más el folio.

Se recomienda nombrar esta variable HHID por sus siglas en inglés *household identification*, es una práctica común que permite que los usuarios puedan crear variables a nivel hogar.

3. Características del hogar: nuevos integrantes y personas que se van.

Se recomienda registrar a los miembros nuevos (nacimientos, gente que se muda) o si éste ya estaba en la ola anterior. Al igual se recomienda registrar a los que se fueron del hogar y conviene saber cuál es la razón de esto.

Esto es importante ya que es importante saber si el hogar cambió mucho de ronda a ronda ya que esto por ejemplo, puede cambiar fuertemente los gastos del hogar en salud, en alimentación, y cambio en las actitudes hacia las pensiones.

Además este ejercicio permite estimar la movilidad de los miembros del hogar (y del entrevistado) esperada en las siguientes rondas.

Se recomienda usar las siguientes preguntas en la lista de los integrantes del hogar (o roster).

¿Sigue [NOMBRE] viviendo en este hogar?

1. Sí
2. No

Esto se puede modificar para saber la ola en la que corresponde. En el caso de una segunda ola sería

1. 2012
2. No

¿Es [NOMBRE] miembro original de [...]?

1. 2012
2. 2014
3. No

Si la persona no sigue viviendo en el hogar, se pregunta:

¿A dónde se mudó [NOMBRE]?

1. Falleció
2. A otro lugar dentro de este municipio
3. A otro municipio dentro del departamento
4. A la capital
5. A otro departamento
6. A otro país
7. Otro: especifique.

Esto se modifica dependiendo el país y la forma como se divide geográficamente.

Si la respuesta es distinta a 1 o 7, se pregunta:

¿Por qué se mudó [NOMBRE]?

1. Encontró trabajo con mayor salario
2. Más trabajo disponible
3. Encontró trabajo y se fue
4. Para estudiar
5. Se casó
6. Asunto familiar (acompañar a algún familiar, cuidar a algún abuelo, divorcio)
7. Otro: especifique

A los miembros nuevos en el hogar se les pregunta (es decir, responde no en la pregunta de miembro original):

¿De dónde se mudó [NOMBRE]?

1. Nació
2. De otro hogar dentro de este municipio
3. De otro municipio dentro del mismo departamento
4. De la capital
5. De otro departamento
6. De otro país: especifique
7. No declarado en la primera ronda

Esto se modifica dependiendo el país y la forma como se divide geográficamente.

Además permite incluir miembros que por alguna razón no hayan sido declarados en la primera ronda.

Si la respuesta es distinta de 1, se pregunta:

¿Por qué se mudó [NOMBRE]?

1. Encontró trabajo con mayor salario
2. Más trabajo disponible

3. Para estudiar
4. Se casó
5. Asunto familiar (acompañar a algún familiar, cuidar a algún abuelo, divorcio)
6. Otro: especifique

4. Identificación de que es la misma vivienda con los mismos individuos

Una vez realizada la verificación de la dirección y de los miembros del hogar, se recomienda agregar una variable que identifique el hogar entrevistado está en la misma vivienda y tiene a los mismos individuos, esto se refiere a que el hogar pudo haber crecido pero el entrevistado y la mayoría de los miembros originales están todavía ahí.

Esta variable se recomienda nombrar “hogar panel”.

Si el entrevistado fue del hogar original para crear su nuevo hogar, la variable tomaría el valor de 0

Si se encontró al entrevistado con su familia original, la variable tomaría el valor de 1

Si el entrevistado cambio de vivienda pero sigue estando con la familia original, la variable tomaría el valor de 2

Si el entrevistado sigue en la misma vivienda pero cambiaron todos los miembros del hogar, la variable tomaría el valor de 3

Este es un filtro muy útil ya que permite saber fácilmente si las variables de hogar son comparables a través de las olas. Además de que las preferencias por la seguridad social pueden haber cambiado si el individuo entrevistado se mudó por haberse casado, o si se mudó a algún otro lugar. Este filtro se debe poner al principio de las secciones que puedan tener variables que cambian con el lugar de residencia y el hogar.

5. Cambios a las preguntas

Como se mencionó antes las encuestas longitudinales deben ser adaptadas al contexto por lo que hay algunas razones por las que es posible hacer cambios a la redacción de la pregunta. Esto, tratando de mantenerla lo más parecida a la pregunta original.

- Cuando existan cambios de tecnología o instrumentos nuevos relevantes a la pregunta
- Cuando las palabras no reflejan el uso del lenguaje actual
- Cuando las preguntas dejan de ser relevantes en el contexto social
- Cuando la pregunta no es utilizada mucho
- Cuando se quieren incorporar mejores prácticas de en la elaboración de preguntas, por ejemplo incorporar términos que puedan capturar mejor las situación actual
- Cuando hay cambios en el diseño del cuestionario

En algunos casos estas razones también llevan a la eliminación o incorporación de preguntas.

Se recomienda ampliamente que los cambios de una ronda a otra sean documentados exhaustivamente explicando razones para los cambios y posibles sesgos derivados del cambio en el fraseo de las preguntas.

6. Registro de los cambios

Siguiendo con lo mencionado anteriormente, es muy importante registrar los cambios que hay de una ola a la otra explicando cual es la justificación para ello. Esto es un ejercicio muy valioso que permite que los usuarios de la encuesta sepan que los cambios de una ola a otra. Se recomienda seguir el siguiente modelo de tabla 2 (abajo) cada fila corresponde a un módulo. La segunda columna especifica el módulo en la ola 2 y la tercera columna la ola 3, la cuarta columna explica las diferencias entre una y la otra. Esta tabla corresponde a la Encuesta de Nivel de vida de los Hogares de México, tercera ronda.

Tabla 2. Diferencias entre preguntas y secciones ENNVIIH 2 y ENNVIIH 3

Libro	ENNVIIH-2 Secciones	ENNVIIH-3 Secciones	Diferencias ENNVIIH-2 y ENNVIIH-3
C	DF (Definición)	DF (Definición)	1) Se modifica el pase de la pregunta 1- respuesta Si- DF02.
	LS (Lista de los miembros del hogar)	LS (Lista de los miembros del hogar)	<ol style="list-style-type: none"> 1) Se adiciona pregunta LS005 para saber si nació algún hijo después del 2002 y no ha sido enlistado. 2) Se adicionan Código de respuesta para pregunta LS05 (respuestas 16, 17 y 18). Se amplían las opciones para determinar la relación social entre personas. 3) Se adiciona Código de respuesta para la pregunta LS15 (respuesta 00). Se amplía opción para puntualizar grado escolar. 4) Se adiciona código de respuesta para pregunta LS19b (respuesta 10). Se amplía opción para señalar la causa de movilidad de una persona. 5) Se adiciona la pregunta LS00i. Ahora se pide registrar el pid_link (indicador único individual - código para proteger la confidencialidad de los datos personales) completo.
	CV (Características de la Vivienda)	CV (Características de la Vivienda)	<ol style="list-style-type: none"> 1) Se adiciona pregunta CV02a en caso de que el participante se encuentre pagando su vivienda, se pregunta a quién está pagando. 2) Se adiciona pregunta CV02b para saber un porcentaje aproximado del adeudo del valor de la vivienda. 3) Se adiciona pregunta CV15a para saber el número de días que le faltó el agua a la vivienda. 4) Se adiciona pregunta CV21 para saber si en los últimos 3 meses, el hogar se quedó sin comer por falta de dinero o recursos.
	SP (Salud y Planificación)	SP (Salud y Planificación)	Ninguna
	RC (Información para el Recontacto)	RC (Información para el Recontacto)	1) Se adiciona la opción de anotar número telefónico del lugar de trabajo.
	EH (Escolaridad de los Miembros del Hogar)	EH (Escolaridad de los Miembros del Hogar)	Ninguna
	CVO (Observación Directa de las Características de la Vivienda)	CVO (Observación Directa de las Características de la Vivienda)	Ninguna.
	NE (Notas de las sesiones de la entrevista)	NE (Notas de las sesiones de la entrevista)	Ninguna

7. ¿Cuándo modificar o dejar las preguntas del cuestionario?

Se identifican tres categorías de preguntas en los cuestionarios de seguimiento y a continuación se describe cuándo se recomienda modificarlas y por qué razones se prefiere que se mantengan igual.

- Preguntas control
- Preguntas que cambian en el tiempo
- Preguntas que no cambian en el tiempo

a) Preguntas control

Las preguntas control son las preguntas de las características básicas del individuo. Para lograr tener estimaciones acertadas es muy importante lograr que tengan el menor nivel de error en la recolección de los datos.

Para lograr esto, se recomienda determinar un conjunto de preguntas control que se repetirán en todas las olas. Por ejemplo:

- sexo,
- edad,
- fecha de nacimiento,
- relación de parentesco,
- nacional o extranjero,
- educación,
- situación de pareja actual,
- estado civil legal,
- número de hijos,
- sabe leer y escribir,
- ascendencia o raza o etnia,
- discapacidad.
- años trabajados
- años aportados

En este conjunto de preguntas se deberán considerar, además, aquellas que tengan una relación directa con el objetivo de estudio y con el contexto en cada país.

Éstas preguntas control aunque pueden ser repetitivas en la encuesta y parecer obvias, en general son rápidas de contestar y pueden tener error de captura o digitación.

Existen varias razones por las que las encuestas longitudinales como la Encuesta de Nivel de vida de los Hogares de México las incluyen en cada ola.

- Por simplicidad y no modificar el cuestionario todas las veces;
- Verificar los errores que hayan surgido tanto de digitación como de memoria;
- Introducir el tema y hacer que las preguntas fluyan más ya que si por ejemplo eliminamos las preguntas de educación y después de cualquier otro tema se pregunta algo de educación esto puede sesgar las respuestas;
- Ayuda a verificar la consistencia del cuestionario en general;
- Ayuda a verificar la eficiencia del entrevistador.

b) Preguntas que cambian en el tiempo

Se pueden identificar las preguntas derivadas de un evento específico y las periódicas.

En el caso de las preguntas derivadas de un evento específico: enfermarse, salir de la escuela, divorcio, etc. se sugiere seguir la siguiente estrategia: Cambiar el fraseo para tratar de capturar lo ocurrido en el periodo transcurrido entre la ola anterior y la actual y en los casos que aplique hacer una pregunta control que permita verificar la información de la ola anterior (este paso es opcional, pero podría ser útil).

Esto aplica a enfermedades, fallecimientos, divorcio, entre otros eventos.

Ejemplo en ola 1: ¿Cuántos hijos nacidos vivos ha tenido?

En ola 2: ¿Cuántos hijos nacidos vivos ha tenido en los últimos dos años? (desde la fecha de la última ola)

Pregunta opcional de control ¿En total cuantos hijos nacidos vivos ha tenido?

Cuando las preguntas derivadas de un evento específico son parte de un set de preguntas largo, se recomienda repetir las preguntas como estaban en la primera ola. Esto permite capturar cambios a la vez que verificar la información capturada en la primera ronda.

En el caso de discapacidad es útil volver a preguntar todo ya que de esta forma se evita cambiar mucho el cuestionario y se pueden capturar cambios. Esto también es una pregunta control.

c) Preguntas que no cambian en el tiempo

En esta categoría se puede incluir la educación y ocupación de los padres, el primer empleo, la fecha que por primera vez se afilió a la seguridad social, entre otras. Para esta categoría de preguntas estas deben eliminarse del cuestionario de seguimiento. Sin embargo debe de registrarse en un documento donde se justifica la razón de su eliminación.

8. Eliminación de preguntas

En caso de que se decida eliminar alguna pregunta dado que deja de ser relevante o que la gente no responde a ella, la numeración de preguntas subsecuentes no debe ser modificada.

Es importante considerar que el cambio en el orden de las preguntas puede afectar el flujo del cuestionario y posiblemente modificar las respuestas que los entrevistados proporcionan. Por ejemplo, si se pregunta sobre enfermedades y luego se pregunta sobre expectativa de vida, las personas pueden estar influidas por sus respuestas anteriores. Si se elimina la pregunta de enfermedades la pregunta sobre la expectativa de vida reflejara algo distinto. Esto es importante tomarlo en cuenta al cambiar el orden de las preguntas.

9. Introducción de nuevas preguntas

Se recomienda introducir nuevas preguntas al final de los módulos o de secciones (series de preguntas) para alterar el cuestionario lo menos posible. Se recomienda justificar y registrar todos los cambios.

10. Período de historia laboral

Se recomienda utilizar un mes de referencia ya que eso permite uniformar el criterio para todos y recibir información comparable. Esto es, se recomienda usar “Desde enero de 20XX a la fecha” que captura los últimos dos años entre olas.

11. Fuera de la vivienda

Se recomienda que el entrevistador recoja ciertas impresiones después de terminar la entrevista.

Observaciones. Solo para el encuestador una vez fuera de la vivienda

<p>O1. Vía de acceso principal a la vivienda (por observación del encuestador)</p> <p>1. Carretera/calle pavimentada o adoquinada 2. Empedrado 3. Lastrado/calle de tierra 4. Sendero 5. Río/mar 6. Otro. O1e. Especifique _____</p>	<p>O2. ¿El estado del material predominante en las paredes externas es?:</p> <p>Señale:</p> <p>1. Bueno 2. Regular 3. Malo</p>	<p>O3. ¿El estado del material predominante en el techo es?:</p> <p>Señale:</p> <p>1. Bueno 2. Malo 3. Regular</p>
---	---	---

<p>O4. ¿El estado del material predominante en los pisos es?:</p> <p>Señale:</p> <p>1. Bueno 2. Malo 3. Regular</p>	<p>O5. ¿La cuadra tiene alumbrado público?</p> <p>1. Sí 2. No</p>	<p>O6. ¿Cuál es el estado de ánimo con el que lo recibió la persona encuestada?</p> <p>1. Muy Bueno 2. Bueno 3. Regular 4. Malo 5. Muy Malo</p>
--	---	--

<p>Impresiones generales</p>	<p>Comentarios</p>
-------------------------------------	---------------------------

12. Medición del efecto de los cambios en los cuestionarios

Cambiar preguntas o secciones en un cuestionario puede tener efectos en las respuestas de las personas y crear sesgos. Por eso es importante justificar y dejar por escrito todos los cambios realizados de una ronda a otra.

Sin embargo, no es posible darse cuenta a simple vista de los cambios en las respuestas producidos por los cambios en los cuestionarios. Existen técnicas en la literatura de las encuestas que recomiendan que antes de hacer cambios sustanciales a los cuestionarios es con una vía experimental. Esto significa implementar cambios en preguntas en un grupo aleatorio de la población y otro grupo control reciba las preguntas anteriores. De esta forma, se podrá saber si hay diferencias sustanciales y estimar el efecto del cambio.

Esta técnica experimental se puede usar también cuando un cuestionario cambia de modo de aplicación PAPI a CAPI, cuando se introducen nuevos módulos, entre otros.

6. Recomendaciones para reducir el error total de la ELPS en las rondas de seguimiento en modo PAPI

A continuación se entregan recomendaciones que tienen como objetivo mejorar la calidad durante el trabajo de campo de las siguientes rondas de la ELPS en los países donde toman la decisión de completar los cuestionarios impresos en papel con un lápiz o modo PAPI (*Paper-and-Pencil Interviewing*). Antes de presentar las cinco sugerencias se entrega el flujo de información para la implementación de cada una de ellas y se detalla el proceso desde la toma de la entrevista hasta la generación de la base de datos depurada y dispuesta para la aplicación de factores de expansión y los análisis de usuarios finales.

6.1 FLUJO DE INFORMACIÓN PARA LA IMPLEMENTACIÓN DE LAS CINCO SUGERENCIAS EN LA ELPS

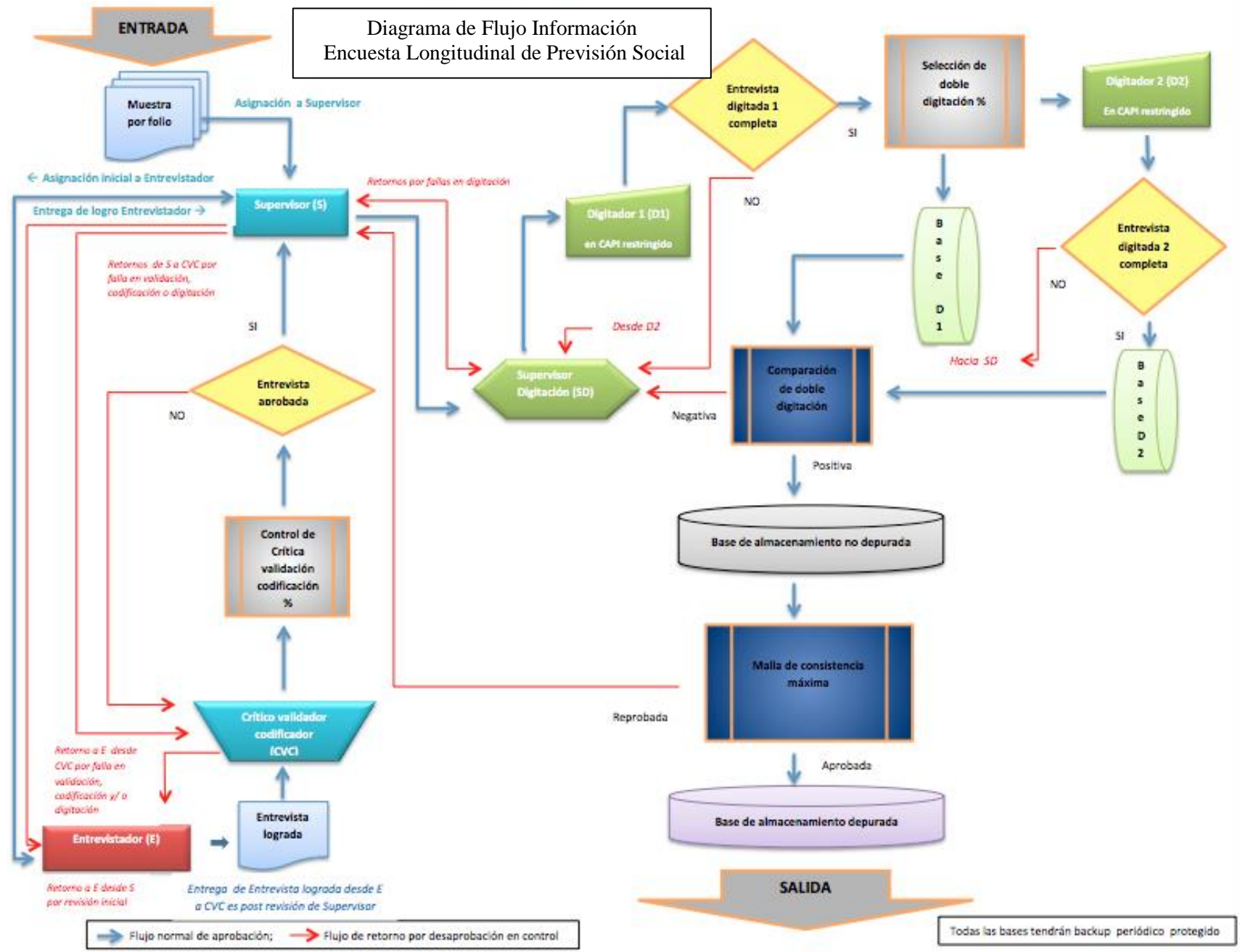
Las sugerencias que se listarán en este capítulo precisan de un proceso dinámico y evolutivo que va conformando la base de datos final de la ELPS junto con las bases anexas que registran información del proceso en sí. Para una visualización gráfica de las interacciones de las recomendaciones se presenta en primer lugar un diagrama de flujo de información de la ELPS aplicada en modo PAPI.

El diagrama describe el circuito que siguen los cuestionarios de la ELPS y las Hojas de rutas o Formulario de Contacto en el transcurso de validación de sus informaciones para ser incorporadas en una base de datos depurada que será la fuente digital para generar los análisis de resultados de la ELPS.

Para una mejor descripción que diferencie el proceso normal (con flujo positivo indicado ilustrado con conectores en color azul →) del proceso con procedimientos rectificadores (con conectores en color rojo →) se describirán a continuación en forma separada ambos circuitos.

El primer circuito de flujo de proceso normal se describirá según el recorrido que realice la información correctamente recolectada y también correctamente procesada.

El segundo circuito correspondiente a los flujos rectificatorios se describirá desde el punto de los diferentes entes o procesos que intervienen desde la entrada (muestra inicial) hasta la salida (base de datos final depurada).



Descripción del flujo normal.

Este flujo se describe por los conectores en color azul → y corresponde al recorrido de tránsito de la información desde su selección a nivel de USM (Unidad Secundaria Muestral) hasta formar parte de los registros de la base final depurada. Este recorrido sigue el siguiente curso según el sistema propuesto:

1. Desde la ENTRADA se tiene información detallada de la muestra asignada al Supervisor (S)
2. El supervisor (S) distribuye la muestra en su equipo de entrevistadores (asignación inicial a Entrevistador (E))
3. El Entrevistador (E) realiza la aplicación de selección de la persona en el hogar señalado y logra la entrevista, devolviendo: cuestionarios, hojas de ruta y consentimientos (SET) completados a su Supervisor (S).
4. El Supervisor (S) autoriza la entrega de la Entrevista lograda al Crítico validador-codificador, luego de haber realizado y aprobado su propia revisión inicial del SET.
5. El Crítico validador-codificador (CVC) procede a revisar procedimientos y contenidos de la información del SET
6. El Crítico validador-codificador (CVC) aprueba el Control Crítica validación-codificación, entregando SET al Supervisor(S).
7. El Set ingresa al Control de Crítica validación codificación a cargo del Supervisor (S) y tras su aprobación ella está a disposición para la etapa de digitación.
8. El Supervisor (S) recibe la documentación validada del SET y la entrega al Supervisor de digitación (SD).
9. El Supervisor de digitación (SD) entrega el SET al primer Digitador (D1) (de la doble digitación).
10. El primer Digitador (D1) digita y aprueba la digitación en Entrevista digitada 1 completa del SET.
11. El primer Digitador (D1) da origen a una base de datos Base D1 y el SET es ingresado al proceso de Selección de doble digitación que está a cargo del Supervisor de digitación (SD).
12. Habiendo sido seleccionado el SET para una doble digitación pasa a ser digitado nuevamente por el Digitador (D2) quien aprueba en Entrevista digitada 2 completa.
13. El segundo Digitador (D2) produce una segunda base de digitación Base D2 y devuelve el SET a su Supervisor de digitación (SD).
14. El programa informático realiza la Comparación de doble digitación , aprobando su validez en forma positiva (digitaciones de SET es idéntica en Base D1 y en Base D2)
15. El programa informático genera la Base de almacenamiento no depurada.
16. El programa informático ingresa el registro digital del SET a la Malla de consistencia máxima.
17. La Malla de consistencia máxima aprueba el registro digital del SET tras no detectar la existencia de ningún error o inconsistencia en la información del SET.
18. El registro digital del SET está dispuesto en la SALIDA de datos para su análisis estadístico de usuarios finales (aplicación de factores de expansión y análisis de representaciones estadísticas)

Descripción de flujos de retorno

Estos flujos se describen por los conectores en color rojo → y corresponde al recorrido de tránsito de la información entre etapas internas del proceso de digitalización validada de la información.

La descripción se hace por las llegadas y salidas de estos flujos de retorno (color rojo) en cada ente o etapa de composición de este proceso completo.

Supervisor

Entradas:

- 1) Retorno desde Supervisor de digitación (SD). Se produce por errores e inconsistencias que no permiten completar su digitación en Digitador D1 y/o Digitador D2.
- 2) Retorno desde Malla de consistencia máxima. Este retorno está acompañado de informes o listados de que explicitan detalladamente el conflicto que impide la validación. Corresponde a SETs que no aprobaron satisfactoriamente la aplicación de dicha malla. Estos pueden tener su origen en el proceso llevado a cabo previamente de levantamiento y/o digitación.

Salidas:

- 1) Retorno hacia el Entrevistador (E). Se produce por errores de levantamiento que comete el Entrevistador y le son indicados en forma personal por el supervisor en una revisión de Hojas de ruta; consentimientos y cuestionario ELPS (los tres instrumentos constituyen un SET). Este retorno debe contener la mayor y mejor instrucción posible del Supervisor hacia el Entrevistador para evitar los mismos errores o inconsistencias a futuro.
- 2) Retorno hacia el Crítico validador codificador (CVC). Se produce por errores de levantamiento que fueron detectados en la etapa de digitación y/o validación del SET.

Entrevistador (E)

Entradas:

- 1) Retorno desde el Supervisor (S). Se produce por errores de levantamiento que le fueron advertidos al Entrevistador por el Supervisor (S) en la entrega inicial de logro para obtener la aprobación de su revisión y entregarla al Crítico validador-codificador (CVC). Este retorno debe estar completamente corregido y aprobado por el Supervisor antes de ingresar la Encuesta lograda (SET) a la revisión que realiza el Crítico validador-codificador (CVC).
- 2) Retorno desde el Crítico validador-codificador (CVC). Se produce por errores de levantamiento que fueron detectados en la etapa crítica validación-codificación; digitación y/o validación del SET.

Crítico validador-codificador (CVC)

Entradas:

- 1) Retorno desde Supervisor (S). Se produce por errores e inconsistencias detectadas en las etapas superiores de digitación y aplicación de mallas de validación. Este retorno está acompañado de informes o listados de que explicitan detalladamente el conflicto que impide la digitación y/o validación.

2) Retorno desde Control de Crítica validación codificación. Este retorno procede del control que se realiza al su trabajo ya aprobado anteriormente por el mismo y en el cual se encuentran errores u omisiones que debe resolver a través una nueva crítica de validación y codificación..

Salidas:

1) Retorno hacia el Entrevistador (S). Se produce por errores de levantamiento que fueron detectados en la etapa crítica validación-codificación inicial o de un retorno superior proveniente de digitación y/o validación.

Entrevista aprobada

Salidas:

1) La salida negativa (No) hacia el Crítico validador-codificador (CVC) luego de Control de Crítica validación codificación, está representando las entrevistas que aún mantienen errores u omisiones y que no fueron detectados por el Crítico validador-codificador (CVC) en su revisión previa esta labor de control la realiza el Supervisor (S).

Supervisor de Digitación (SD)

Entradas:

1) Retorno desde Digitador 1 (D1). Se produce en las entrevistas que adolecen de fallas u errores que son detectados por el programa de digitación que contiene una programación con mínima validación de consistencia para generar la base de registros en Base D1.Estas fallas u errores no permiten completar la primera digitación.

2) Retorno desde Digitador 2 (D2). Se produce en las entrevistas que adolecen de fallas u errores que son detectados por el programa de digitación que contiene una programación con mínima validación de consistencia para generar la base de registros en Base D2.Estas fallas u errores no permiten completar la segunda digitación.

3) Retorno desde Comparación de doble digitación. Este retorno está acompañado de informes o listados de que explicitan detalladamente el conflicto que impide un resultado positivo. Este resultado le indica al Supervisor de Digitación (SD) transmitir su solución al Digitador D1; Digitador D2 o Supervisor (S) quién a su vez lo remitirá al Crítico validador-codificador si fuese de su responsabilidad.

4) Retorno desde Supervisor (S). Corresponde al flujo de entrevistas retornadas tras haber sido enviadas previamente según el punto de Salida 1 indicado a continuación.

Salidas:

1) Retorno hacia el Supervisor (S). Se produce por errores de las entrevistas que impiden la digitación de la información levantamiento y que proceden de su propia observación (Supervisor de digitación (SD)) o de sus respectivos digitadores.

Entrevista digitada 1 completa

Salidas:

1) La salida negativa (No) hacia el Supervisor de Digitación (SD) la ejecuta el Digitador 1 (D1), y está representando las entrevistas que no pueden ser digitadas completamente al no supera las exigencias mínimas establecidas para la formación de una base de datos comparable Base D1.

Entrevista digitada 2 completa

Salidas:

1) La salida negativa (No) hacia el Supervisor de Digitación (SD) la ejecuta el Digitador 2 (D2), y está representando las entrevistas que no pueden ser digitadas completamente al no supera las exigencias mínimas establecidas para la formación de una base de datos comparable Base D2.

Comparación de doble digitación

Salidas:

1) La salida negativa (No) hacia el Supervisor de Digitación (SD) la realiza el programa informático de comparación desarrollado e indica documentadamente los detalles de no concordancia de la Base D1 y la Base D2.

Malla de consistencia máxima

Salidas:

1) La salida Reprobada hacia el Supervisor (S), la realiza el programa informático de comparación desarrollado e indica en sendos informes el detalle de los errores en rangos y consistencias que afectan al SET de información

Sin duda esta es la salida más importante que debe atenderse sobre el proceso ya que su falta de resolución puede ocasionar deterioro en la representación de los resultados, aumentando la no respuesta.

La información contenida en este informe de retorno es dirigida al Supervisor (S) quien deberá resolver caso a caso su resolución dirigiéndola hacia sus responsables iniciales, siguiendo un ordenamiento jerárquico decreciente, para dejar constancia del retorno en todos los niveles de la cadena productiva y permitir un mejoramiento global en todo el equipo al compartir situaciones o resultados que deben evitarse.

6.2 SUGERENCIAS PARA REDUCIR EL ERROR DE LA ENCUESTA REALIZADA EN MODO PAPI

6.2.1 Supervisión de las capacitación de los encuestadores.

En el organigrama del trabajo de campo de la ELPS en modo PAPI se distinguen tres actores que interactuaran en el objetivo de completar correctamente el Cuestionario en modo. Ellos son los Entrevistadores; Supervisores y Críticos validadores-codificadores.

Los Entrevistadores y Críticos validadores-codificadores debieran ser entrenados en la misma capacitación. En el caso de los Supervisores, debieran asistir a las capacitaciones en calidad de monitores y asistentes para profundizar y aplicar conocimientos adquiridos en sus capacitaciones previas. Es decir, los Supervisores debieran ser capacitados previo a las capacitaciones de los Entrevistadores y Críticos validadores-codificadores.

Las capacitaciones de los Supervisores deben ser realizadas en forma centralizada por la dirección del proyecto en jornadas diarias de no menos de 3 días consecutivos. En ellas debieran ser instruidos por profesionales sectorialistas de las áreas de estudio consultadas en la ELPS y que se desempeñen en los organismos gubernamentales que solicitan el estudio, y que serán los usuarios principales de la información.

Se debiera impartir una capacitación a los supervisores en dos áreas:

- 1.-Contenidos administrativos (Control y gestión de recursos humanos y físicos en la sede)
- 2.-Contenidos del cuestionario ELPS e instrumentos asociados (Hojas de ruta, Consentimientos, Informes de evaluación de Críticos Validadores-codificadores; Digitadores)

Estas capacitaciones debieran ser programadas por módulos con espacios destinados al ejercicio y evaluación de los contenidos. Las evaluaciones debieran ser inmediatas, resolviendo dudas y errores en forma previa al inicio de la sección de instrucción siguiente.

En las capacitaciones de supervisores deberán quedar estipulados todos sus derechos y deberes dentro del proyecto como así mismo abordar en forma realista la disponibilidad de mantener esas funciones en períodos de alta demanda laboral que exijan trabajar en jornadas extra (Ej. Fines de semana; horas extra sobre la jornada diaria habitual). Una vez concluida esta etapa previa de capacitación de Supervisores se inicia la capacitación de los Entrevistadores.

La calidad del trabajo obtenida por cada Entrevistador no necesariamente tiene una correlación directa del conocimiento previo de conceptos insertos en la ELPS. Sino más bien por los niveles de conocimiento adquiridos durante las capacitaciones, su concentración, y responsabilidad en la ejecución de su trabajo asignado. Sin embargo el correcto conocimiento previo de estos conceptos en algunos entrevistadores, les otorga una ventaja inicial para lograr un buen desempeño en terreno. De esta forma entrevistadores que tengan conocimientos superiores en las áreas sociales y económicas, a priori podrían realizar su trabajo en forma un poco más precisa y correcta. Este antecedente no debe ir en desmedro de postulantes a entrevistadores con estudios en otras áreas que sólo tienen estudios formales de nivel medio y que a través de una buena aplicación en las capacitaciones y responsabilidad pueden obtener rendimientos efectivos sobresalientes.

La convocatoria de entrevistadores a la reuniones de capacitación se debiera realizar con anticipación (no inferior a una semana). La convocatoria debe priorizar la disposición horaria de trabajo y las direcciones de domicilio más cercanas a la zona de muestra. Una exigencia obligatoria que deben cumplir todas las personas que participen en este proyecto como Entrevistador, Supervisor y Crítico validador-codificador, es tener una escritura clara y legible en su trabajo.

Las capacitaciones a los entrevistadores tienen por objetivo principal entregarles la información necesaria para que cumplan su trabajo de aplicación del cuestionario ELPS en forma eficiente. Las capacitaciones se deben planificarán en forma modular para atender los aspectos administrativos y de aplicación técnica del cuestionario ELPS. El lugar de capacitación deberá ser del tipo aula o sala de clases que disponga de conectividad de dispositivos de presentaciones con multimedios (dispositivos audio-visuales).

Las capacitaciones deberán tener una estructura predefinida que aborde en forma ordenada todas las instancias que provoquen errores atribuibles tanto a su propio desempeño, como a los provenientes del informante o persona entrevistada. Se sugiere elaborar un "Listado de fuentes de error en etapas críticas", con cada uno de los aspectos mencionados en la sección de de las etapas críticas que sean

aplicables al entrevistador y se debieran explicitar en forma contextualizada para su mejor comprensión..

Al término de la capacitación conceptual y del total de preguntas del cuestionario ELPS se sugiere realizar ejercicios auto evaluados de *role playing* en que participen todos los entrevistadores. Estos ejercicios tiene por objetivo detectar y subsanar deficiencias para abordar entrevistados difíciles de convencer a participar en el estudio. En estos ejercicios se exponen diferentes técnicas para desbloquear esas dificultades. También se debe especial atención al uso del lenguaje correcto para ser aplicado y su presentación personal.

La comprobación de conocimientos y aptitudes adquiridas por los entrevistadores se sugiere basarla en una serie de instrumentos de evaluación. Los instrumentos debieran estar desarrollado específicamente para estos efectos y recoger tanto los resultados de ejercicios prácticos como de conocimiento teórico. Los ítems que se evalúan deberán tanto en la forma como en los contenidos, ser aprobados por la dirección central del proyecto antes de ser aplicados a los entrevistadores.

Al concluir cada capacitación en Sedes se les entregará un "Cuestionario de evaluación de la capacitación" a cada entrevistador (y/o crítico validador codificador) para que emita sus calificaciones y opiniones respecto de la capacitación recibida. Se sugiere realizarla en etapas según jornadas (módulos o diaria) para corregir eventuales deficiencias mayores y obtener al mismo tiempo un diagnóstico más reciente de parte del entrevistador (y/o crítico validador codificador).

Las evaluaciones serán estandarizadas e iguales para todos los entrevistadores, salvo aquellas sedes en que por razones particularidad en la muestra requieran un ítem modificado o nuevo. Este sería el caso de las sedes que contienen sólo áreas urbanas versus otras que sean mayoritariamente rurales.

La corrección de las evaluaciones de contenidos técnicos de la ELPS debiera estar a cargo del supervisor de sede de modo tal de conocer tanto las deficiencias de sus entrevistadores y postulantes a Críticos validadores-codificadores como las que evidencien un bajo o deficiente nivel de las capacitaciones. Esta corrección deberá ser inmediata para hacer uso de sus resultados antes del término de la capacitación.

Posteriormente se debe evaluar los niveles de asimilación de conocimiento que arrojen los instrumentos de evaluación y se deben focalizar esfuerzos en aquellos módulos o aspectos que resulten deficitarios precisando y corrigiendo las causas que provocaron resultados insuficientes.

Las escalas de evaluación y su graduación como también los criterios asociados a ellas debiera ser dados por la dirección central del proyecto, y así realizar análisis comparativo entre sedes a fin de detectar falencias en cada una de ellas y que deben ser corregidas antes del inicio del terreno.

Los resultados de estas evaluaciones debieran ser comunicadas a la dirección central en documentos de transmisión dispuestos para ello. Así la dirección central podrá contar con los resultados de las evaluaciones de cada postulante a entrevistador (y/o crítico validador codificador) y también la correspondiente a cada capacitación.

Todos los instrumentos de evaluación debieran ser identificados con la sede; nombre del evaluado; fecha de la evaluación, nombre del revisor. Estos documentos escritos se archivarán en forma numerada, siendo resguardados por el supervisor quien deberá entregarlos en un reporte a la dirección central. Bajo ninguna circunstancia pueden extraviarse estos instrumentos de evaluación.

Al comenzar su participación en el estudio los entrevistadores completarán una "Ficha de inscripción" que entregará información respecto de edad; dirección de domicilio; sus competencias

educacionales y especialmente laborales que serán observadas para reforzar conocimientos específicos para este estudio diferenciándolo de otros similares en que pudiera haber participado el entrevistador.

Se sugiere exigir a lo menos dos cuestionarios realizados en direcciones de prueba sobrante del estudio piloto en estratos socioeconómicos bajo y medio. Esto tiene por objetivo detectar capacidades insuficientes del entrevistador para abordar a los entrevistados. Estas dos entrevistas reales serán también parte de la evaluación del entrevistador y las pautas de evaluación debieran ser dadas en forma centralizada para realizar análisis comparativo. Al igual que los otros documentos de evaluación éstas deberán archivar.

6.2.2 Revisión y validación/codificación de las encuestas realizadas

El cumplimiento de cada una de las pautas de proceso entregadas en esta recomendación se sugiere ser atendido en forma permanente para lograr la calidad necesaria de la información levantada, ya que sin duda cubre la etapa del filtro más crítico para evitar malos resultados.

Es importante comenzar con una revisión previa de la muestra para cotejar direcciones muestrales con la dirección y cartografía que se enviará a terreno en las Hojas de ruta; realizar agrupaciones adecuadas de asignación.

Se sugiere elaborar n pautas escritas de revisión de Hojas de ruta; consentimientos informados y cuestionarios ELPS.

Se sugiere diseñar en la web un registro centralizado de "Disposición Final de Casos" (DFC) para indicar la evolución de cada folio de la muestra: Asignación; des-asignación; Estado de resultado de terreno; Estado/resultados de revisión de Supervisor; Estado/resultados de revisión de Crítico validador-codificador; Resultado de Control de Crítica validación codificación; Estado/resultados de revisión de primer digitador; Estado/resultados de revisión de segundo digitador; Resultados de Control de doble digitación; Estado de resultado pre-validación final; Estado de resultado post validación final. Este registro deberá estar actualizado permanentemente con cada revisión que se ejecute.

La información de la DFC debiera producir una base de datos Base DFC que la dirección central contrastará con la base generada a partir de la digitación de los cuestionarios ELPS y sus respectivas Hojas de ruta y consentimientos. Esta comparación tendrá por objetivo evaluar la corrección de ambas informaciones y a partir de sus diferencias corregir errores (Ej. Cuestionarios no procesados, Cuestionarios con identificación de folio incorrecto, mal ingreso en los reportes de avance y estados de resultados y controles).

Se sugiere que el diagramado del cuestionario ELPS tenga una página destinada a "Resultados de revisiones, controles y digitaciones". Una alternativa de ubicación es la contratapa (o última página del Cuestionario). En esta página de "Resultados de revisiones, controles y digitaciones" se debiera establecer espacios para el registro de confirmación de estas revisiones del Entrevistador; Supervisor, Crítico validador-codificador y digitadores. En estos espacios se debiera registrar el nombre y firma del responsable y fecha de revisión. La página de "Resultados de revisiones, controles y digitaciones" debiera contener variables dicotómicas que indicarán que un proceso determinado fue aprobado (Ej. Revisión del Crítico validador-codificador aprobada, con firma). La página de "Resultados de revisiones, controles y digitaciones" debiera ser digitada y formar parte de la base paradata.dta. Y sus variables serán controladas al ser comparadas con la de la base de Disposición Final de Casos (DFC).

El Entrevistador tiene la obligación de revisar cada entrevista una vez completada. Se sugiere que deje constancia de esta acción en la página destinada al Resultados de revisiones, controles y digitaciones.

Las Hojas de Ruta ; Cuestionarios ELPS y Consentimientos se completarán sólo con lápiz de tinta indeleble color azul o negro. Así si una entrevista se realiza en color azul, entonces su revisión se hará en color negro y viceversa.

La escritura de color rojo estará destinada a codificaciones y escritura del registro de sumas de control (Ej. Suma de ingresos por página).

Se establecerá un registro centralizado de fechas de entregas o asignaciones; recepciones; revisiones; controles y digitaciones de muestra. Ej. Esto puede ser realizado a través de cuentas en Dropbox para cada supervisor en que se reportaran los estados de cada entrevista y la etapa del proceso en que se encuentran.

Las Hojas de rutas y consentimientos informados serán los primeros documentos que deberá revisarse. Se condicionará la revisión de los Cuestionarios ELPS a una aprobación de estos dos instrumentos previos.

Para las Hojas de ruta se establecerán pautas que deberán confirmarse en el registro de revisión.

Se sugiere que el Supervisor revise completa y profundamente las primeras 5 encuestas de cada entrevistador en forma presencial para detectar todos los errores u omisiones que pudiera presentar. No debieran aceptarse revisiones no presenciales entre Entrevistador y Supervisor. Las correcciones se pueden realizar tajando lo incorrecto y colocando lo correcto con un color diferente al traído inicialmente desde terreno.

La revisión debiera seguir un orden según la lógica de llenado del cuestionario de acuerdo a las respuestas que sean requeridas. Esto no resta corregir anticipadamente otros errores que sean detectados previamente (Ej. Falta respuesta en variables que no aceptan *missing*).

La revisión del cuestionario ELPS se puede ver facilitado por el "Pauta de revisión del supervisor" (PRS) que deberá seguir en cada revisión de cuestionario o registro de persona.

Se sugiere seguir el orden de revisión de respuestas comenzando por universos; flujos y posteriormente las consistencias que indique el Listado de revisiones. Esto evita omitir algunas posibles revisiones que pudieran contener errores.

El supervisor debiera devolver el cuestionario sin su aprobación en caso que detecte errores no solucionados en una conciliación con el entrevistador y registrará esta situación en el "Registro de estados por folios". En este caso el entrevistador contactará al entrevistado nuevamente para completar correctamente el cuestionario y regresar hacia el Supervisor para su aprobación.

Los cuestionarios aprobados por el supervisor pueden ser entregados al entrevistador para que los entregue personalmente al crítico validador-codificador.

El Supervisor deberá copiar los resultados de su "Registro de Revisiones del Supervisor" (RRS) hacia el sistema central de Disposición Final de Casos (DFC).

El Crítico validador-codificador podrá recibir sólo cuestionarios que cuenten con la aprobación firmada previamente del supervisor (En la Hoja de RRS).

Se sugiere que el Crítico validador-codificador trabaje en un ambiente y lugar que reúna condiciones de espacio; silencio; ventilación y temperatura y seguridad que aseguren una concentración permanente en su trabajo.

El Crítico validador-codificador debiera revisar presencialmente y de manera profunda y exhaustiva las primeras 5 entrevistas realizadas por el entrevistador.

La revisión realizada por el crítico validador debiera tener la secuencia indicada en la "Pauta de revisión de Críticos validadores-digitadores" (PRCVC).

Se sugiere que la PRCVC sea elaborada a nivel central por la dirección del proyecto y constituirse así en las revisiones mínimas que deberán realizar los críticos validadores-codificadores en cada una de las entrevistas logradas incluyendo Hojas de ruta y consentimientos informados junto al Cuestionario ELPS. La PRCVC debiera llegar a ser uno de los instrumentos más importante y conocido para los Críticos validadores-codificadores en su trabajo.

Los resultados de la evaluación realizada por el crítico validador-codificador podrán ser indicados por él en el "Registro de Revisiones del Crítico validador-codificador" (RRCVC). En este registro se indicaran los resultados obtenidos en la aplicación de la PRCVC y los adicionales que encuentre en su revisión.

Se sugiere que el RRCVC tenga variables para registrar la situación y ubicación de los cuestionarios vistos por el Crítico validador-codificador sólo podrá establecer dos direcciones tras su trabajo (devuelto al entrevistador o aprobado y enviado a supervisor)

El RRCVC debiera tener además las tasas de resultados *missing* y respuestas No sabe /No responde encontradas en el cuestionario.

Los cuestionarios que sobrepasen límites preestablecidos para sus tasas de *missing* y respuestas No sabe /No responde, podrían ser devueltos a los entrevistadores para reducir esos valores críticos.

Para las codificaciones y registro de sumatorias de ingresos y montos declarados se sugiere diagramar los espacios para su registro en el cuestionario ELPS en forma anticipada.

Esto impacta positivamente en un mejor control y digitación de los datos. De no ser posible esta diagramación previa de estos espacios, el registro se puede realizar sobre el texto que representa el código y en la celda correspondiente a la variable y/o línea de registro.

Las sumatorias de ingresos o montos deben realizarse por páginas o a lo más módulos, para una detección más eficiente de los errores de lectura/interpretación y digitación de los valores.

Las codificaciones y sumatorias se sugiere realizarlas con lápiz de tinta indeleble color rojo.

Las codificaciones debieran ser requisito para la aprobación de un cuestionario ELPS por parte del crítico validador-codificador, es decir no se podrán aprobar cuestionarios sin tener todas las codificaciones y sumatorias de variables de ingresos y montos realizadas. Sin embargo se podrá pre-aprobar el cuestionario (para informar al supervisor y entrevistador), si el crítico-validador codificador considera suficiente las glosas de texto requeridas para aplicar luego un código preciso.

El Crítico validador-codificador debiera copiar los resultados de su RRCVC hacia el sistema central de Disposición Final de Casos (DFC).

El supervisor debiera tener un RCCVC. En este podría anotar los folios sujetos a revisión; fecha de solicitud del cuestionario; fecha de entrega de la evaluación al crítico validador-codificador; resultado de evaluación.

Los resultados del registro RCCCVC debieran ser incorporados a un registro de la DFC

La dirección central debiera fijar el % mínimo de revisión del supervisor al crítico validador-codificador.

Los controles realizados por el supervisor al crítico validador-codificador debieran ser aleatorios cuidando sesgos como supervisar solo cuestionarios hogares de pocas personas.

Se sugiere que los controles realizados por el supervisor al crítico validador-codificador cubran el trabajo de todos los entrevistadores que participen en esa sede.

El supervisor debiera tener un "Registro de Cuestionarios en Digitación" (RCD) en este documento se debiera registrar la fecha de ingreso del cuestionario a digitación; fecha de devolución de cuestionario digitado.

La entrega de cuestionarios aprobados para su digitación debiera ser desde el Supervisor hacia el Supervisor de Digitación.

Se sugiere que el Supervisor de Digitación reciba los cuestionarios para digitación desde el Supervisor de la sede. En la recepción debiera tener el documento de "Registro de Revisiones del Supervisor de Digitación (RRSD)" para anotar los cuestionarios recibidos. Estos cuestionarios se debieran disponer ordenadamente para ser puestos en la cadena de digitación que considera el primer y segundo digitador en forma secuencial.

Es importante que el RRSD tenga variables para registrar: folio; fecha de recepción de cuestionario de digitación; Fecha de entrega al primer digitador; Estado de primera digitación; Fecha de resolución de primera digitación; Fecha de aprobación o rechazo de primera digitación; Fecha de control de primera digitación; Fecha de entrega al segundo digitador; Estado de segunda digitación; Fecha de resolución de segunda digitación; Fecha de aprobación o rechazo de segunda digitación; Resultado de Control de doble digitación; Fecha de devolución cuestionario digitado al supervisor.

El Supervisor de Digitación debiera recibir desde el supervisor (y/o la dirección central a través de la observación de la DFC) los folios que deberá someter a doble digitación.

El Supervisor de Digitación debiera entrega el Cuestionario al primer digitador y éste hacer recepción ingresándolo en su hoja personal de "Registro de Digitaciones Digitador 1" (RD1) en este registrará; folio; fecha de ingreso a digitación; Observaciones de digitación; fecha de salida de digitación. En las Observaciones registrará siempre las causas de rechazo en digitación (Digitación 1 incompleta)

Se sugiere que el Supervisor de Digitación entregue el Cuestionario seleccionado para doble digitación al segundo digitador y éste hace recepción ingresándolo en su hoja personal de Registro de Digitaciones Digitador 2 (RD2) en este registrará; folio; fecha de ingreso a digitación; Observaciones de digitación; fecha de salida de digitación. En las Observaciones registrará siempre las causas de rechazo en esta segunda digitación (Digitación 2 incompleta)

El Supervisor de Digitación debiera revisar los resultados de las digitaciones mediante tres flujos: Devolución de cuestionarios de digitador 1; Devolución de cuestionarios de digitador 2; 3 Reporte de comparación de digitaciones.

Se sugiere que el Supervisor de Digitación redirige al Supervisor todos los cuestionarios separados en Digitación Aprobada sin doble digitación; Digitación Rechazada y Digitación Aprobada con doble digitación. A su vez el supervisor debiera devolver los cuestionarios con Digitación Rechazada a su Crítico validador-codificador para conciliar su solución con el entrevistador.

6.2.3 Elaboración y análisis de la paradata

La base de Disposición Final de Casos (DFC) se debiera combinar con la base de inscripción de entrevistadores y la base *paradata* proveniente de la base no depurada (sin aplicación de malla final de validaciones).

El cruce de bases se realizará en forma periódica. Se sugiere realizar cruces semanales que generarán reportes de errores de medición, procesamiento, cobertura y no respuesta que deberán resolver en la semana que le son informados (antes de la generación del siguiente reporte semanal). Se sugiere realizar cruces los días viernes en la tarde para disponer de los reportes a los supervisores el día lunes a primera hora.

Toda la generación de reportes; tasas de indicadores y errores se debieran generar a partir de la aplicación de programas desarrollados en Stata (do-file).

Se sugiere que los análisis de *paradata* reporten a cada supervisor el detalle de fuentes de no respuesta crítico (Ej. Indicadores que indican visitas sólo en días hábiles)

6.2.4 Digitación de las encuestas con primer nivel de revisión consistencia

La digitación se debiera realizar con personal calificado con experiencia en digitaciones masivas.

La selección aleatoria de los folios de re-digitación debiera venir desde la dirección central que tomara como fuente la base de digitación de primer digitador. Para la selección se recomienda una de las tres siguientes alternativas:

1. Redigitación del 100% del trabajo (Doble digitación completa). Este procedimiento es de costo más elevado, pero otorga mayor confiabilidad al considerar todas las digitaciones.
2. Redigitación decreciente desde 100% de los cuestionarios y documentos. Este procedimiento requiere establecer un valor máximo aceptable de discrepancias (*errormaxdig%*) entre las digitaciones originales y de control. En este máximo no debieran incluirse las variables críticas del cuestionario, es decir, aquellas que bajo todo evento deben estar correctamente digitadas. Luego las redigitaciones disminuyen desde un 100% paulatinamente hasta que no excedan el valor máximo aceptable de discrepancias (*errormaxdig%*), el porcentaje de cuestionarios sometido a doble digitación dentro de los márgenes aceptables (% de error menor a *errormaxdig%*) no debiera ser inferior a un 20% ó 10% en los casos de un buen desempeño de los digitadores. La redigitación decreciente requiere un equipo independiente de digitación.
3. Redigitación aleatoria de un porcentaje mínimo de cuestionarios (20%) que detecta discrepancias de digitación que tras un % límite anula el cuestionario original y obliga a su

redigitación por el digitador. Este procedimiento requiere menor cantidad de digitadores de control que los mencionados previamente.

El Programa de Digitación debiera ser un equivalente pseudo CAPI diseñado con los mismos nombres de variables que el Cuestionario impreso PAPI pero que no llevará los enunciados de sus preguntas sino sólo el respectivo *label* de ella. Se debiera implementar en él todas las validaciones de flujo y universos. Y sólo se habilitarán en él las validaciones de consistencia básica para las relaciones Edad-relación de parentesco; edad – educación; edad/sexo- recepción de beneficios de programas estatales. Se sugiere dejar para la malla de validación final todas las restantes validaciones de consistencia posibles de aplicar.

6.2.5 Análisis de consistencia de todas las variables en las bases de datos

Se sugiere que la aplicación de la Malla Final de Validación a la base no depurada genere reportes a los supervisores en que se indicarán los folios que deberán ser revisados nuevamente con el detalle adjunto de cada error detectado.

El retorno de un cuestionario por la aplicación de la Malla Final de Validación debiera quedar registrado en la base de Disposición Final de Casos (DFC) y su actualización debiera ser responsabilidad del Jefe de terreno (o la persona responsable de generar el reporte de la Malla Final de validación).

El reporte periódico de la Malla Final de Validación debiera contener: Identificación del supervisor; Número de errores detectados; Desglose de identificación de errores por folio y orden de persona indicando variables involucradas.

Estos cuestionarios reportados con errores desde la Malla Final de Validación debieran ingresar al circuito Entrevistador; Supervisor; Crítico validador-codificador; Supervisor de digitación (digitadores 1 y 2) según los errores o correcciones realizadas. Es decir recorrer todo el proceso desde el personal que realiza la corrección.

7. Recomendaciones para la diseminación y promoción de las bases de datos y documentación de la ELPS

Un acceso deficiente a las base de datos y documentación de encuestas sociales tiene efectos negativos en la investigación académica y en el diseño de políticas sociales, haciendo que éstas se orienten solo a la información disponible y que la calidad de los datos no mejore ni se exploten a cabalidad. A su vez, estos efectos negativos se traducen en que sean menos las personas pensando los problemas de la sociedad, y que éstas muchas veces repliquen sus esfuerzos al no estar claros los procedimientos que se siguieron para obtener y configurar la información. De esta forma se hace esencial transparentar y dar cuenta no solo los datos sino también las decisiones metodológicas detrás de éstos para promover la investigación social de calidad.

Para el caso de la ELPS, lograr que la información sea de fácil acceso y debidamente documentada, permitirá avanzar hacia los estándares internacionales de disponibilidad y educación en el manejo de bases de datos longitudinales. De esta manera, se va generando en Latinoamérica un precedente para futuros emprendimientos en la temática de previsión social o la evidencia necesaria para justificar el financiamiento de nuevas rondas de la ELPS.

Un punto importante que no puede ser obviado, es mencionar que no se debe pensar este paso como el solo disponer en un sitio web las bases de datos de la ELPS sino que debe existir una capacitación en el uso de las bases de datos. Es por esta razón que las instituciones responsables de la ELPS debieran situar el proceso del manejo, organización y diseminación de las bases de datos al mismo nivel que las etapas de diseño, metodología y levantamiento de datos. Considerar este punto en la etapa de gestación de la ELPS ayuda en asegurar que la información contenida en las bases de datos pueda ser utilizada correctamente para análisis posteriores.

Recogiendo la experiencia internacional Taylor (2000) entrega tres razones que justifican considerar la diseminación y promoción de las bases de datos como una etapa fundamental en una encuesta social compleja. En primer lugar hay una razón epistemológica: para que el conocimiento se reconozca como tal debe ser público. Desde una perspectiva económica, los altos costos de un proyecto como la ELPS solo se justifican si se genera un mayor dinamismo de conocimiento. La tercera razón tiene un fundamento ético. Aquellos que usan los recursos limitados disponibles para investigar tienen la responsabilidad de hacer que una amplia comunidad conozca los resultados.

La ansiedad por publicar de los investigadores, sumado a la falta de fiscalización acerca de los procedimientos utilizados durante el proceso y la falta de reconocimiento de poseer una base de datos bien documentada, hace que nuevos análisis por investigadores que no participaron directamente del estudio sean muy difíciles de lograr. Conocer y ser conscientes de las dificultades y virtudes de la base de datos es fundamental si se quieren realizar un estudio de alto nivel.

Una buena información y documentación de las bases de datos es importante para los llamados usuarios secundarios o quienes las utilizarán para análisis posteriores. En esta etapa la documentación e información debe responder a muchas preguntas, pero por sobre todo debe conservar el conocimiento adquirido por el equipo investigador que diseñó el estudio. Interrogantes del tipo: ¿cómo fue recogida la base de datos?, ¿cómo fue validada y por qué fue limpiada?, ¿cómo se utilizan la base de datos? deben ser aclaradas.

Estimular diferentes miradas y gatillar nuevas posibilidades de análisis son elementos que no pueden estar ausentes de una correcta documentación de las bases de datos. La información de las bases de datos debe cumplir a su vez un rol educador para los usuarios secundarios de tal forma

éstos las utilicen sin posibilidad de mal interpretaciones o producir resultados con errores por no usar bien las bases de datos.

La particularidad de las encuestas panel es que las bases de datos crecen, se complejizan y actualizan muy dinámicamente toda vez que una nueva ronda es lanzada. Es por eso que quedan obsoletos soportes de distribución como CD o impresiones. La irrupción de Internet como una herramienta de manejo flexible, rápida, dinámica y de alto alcance abre una gran gama de posibilidades y aplicaciones para una mantención y difusión de datos para este tipo de estudios.

Con el fin de estructurar algunas recomendaciones en este tema, se sugiere dividir la etapa de diseminación y promoción de las bases de datos en cuatro sugerencias que de alguna manera están relacionadas entre sí: i) documentación requerida en un estudio panel; ii) sitio web de un estudio panel; iii) seminario y talleres , y iv) ejemplo de un caso exitoso de promoción de los datos y documentación de una encuesta longitudinal.

7.1 DOCUMENTACIÓN REQUERIDA EN UN ESTUDIO PANEL

Una completa documentación es necesaria para el análisis de todo tipo de encuestas. Ahora bien, en encuestas tipo panel esta documentación es todavía más necesaria por las complejidades adicionales que presentan por sobre sus homólogas de corte transversal. De no proveerse un adecuado contexto de los datos e indicaciones adicionales, es muy fácil que el análisis termine en resultados erróneos.

Por esta razón, en la ELPS es importante liberar las bases de datos con una documentación de los más altos estándares. Taylor (2000) sugiere cuatro tipos de información en la documentación de los datos de una encuesta panel. Primero, todo lo relacionado con el diseño de la encuesta. Por ejemplo, la muestra, cuestionario, trabajo en terreno, codificación, edición, *linking*, tratamiento de los datos faltantes. Segundo, la información relacionado con el diseño del panel. Por ejemplo, las reglas de seguimiento, verificación del *linking*, la periodicidad de las mediciones. Tercero, los datos de la encuesta. Por ejemplo, las consistencia de los datos, los ajustes de la muestra, los controles de tabulación. Cuarto, los datos de la panel. Por ejemplo, las consistencias intertemporales, el ajuste de la muestra en el tiempo y también los controles de las tabulaciones desde una mirada dinámica.

En términos prácticos la misión de la documentación es entender mejor los datos y minimizar los errores en los análisis por parte del usuario. A continuación se exponen los diferentes documentos que se consideran imprescindibles para esta etapa. La mayoría de ellos se basan en las recomendaciones que realiza Lynn (2001) en el texto *Standard Quality Profiles for Longitudinal Studies* y que luego Peter Lynn aplica para la encuesta panel de hogares del Reino Unido (2006).

Documento 1: Introducción y descripción del estudio y de los objetivos de la investigación

Incluye los objetivos y propósitos de la investigación. La justificación del estudio. Información relacionada con otros sistemas bibliográficos. Corresponde a la información esencial que acompaña a la base de datos.

Documento 2: Información del contexto de la encuesta

Incluye detalles del equipo responsable del estudio, y la racionalidad detrás del diseño de la investigación.

Documento 3: Metodología y diseño de la encuesta

Describe la metodología utilizada. Incluye descripción de instrumentos utilizados: cuestionarios, información de los cuestionarios, temas de la encuesta. Descripción del diseño muestral (definición de población objetivo, mecanismo de selección de la muestra, detallar procesos, etc). Detalla el

trabajo en terreno y proceso de recolección de datos (Cronograma, trabajo en terreno, procedimientos de los encuestadores, reglas de seguimiento, atrición, etc.)

Documento 4: Recolección y Procesamiento de los datos

Descripción de procedimientos, técnicas y codificación usada.

Documento 5: Error de la muestra, información acerca de peso e imputación

Procedimiento y algoritmos usado, codificación de la imputación a través de marcadores (*flag*). Descripción de errores de medida, error de no respuesta, errores muestrales.

Documento 6: Información descriptiva

Notas de la estructura de los datos, notas acerca de los términos y conceptos usados, información de contexto, información técnica para la construcción de la base de datos. Notas acerca de la construcción de variables derivadas, etc.

Documento 7: Publicación y detalles del análisis

Referencias a todas las publicaciones relevantes para el análisis de datos

Documento 8: Consejos para usar datos linkeados

Recolecta información no solo a través del instrumento mismo sino que se vincula los resultados con otras plataformas de datos.

Documento 9: Declaración de confidencialidad y ética

Documento 10: Manual del Usuario

Explica el uso de las variables, su codificación y facilita el acceso a la principal información. Permite a los investigadores o analistas hacer uso de las bases de datos de manera sencilla.

7.2 SITIO WEB DE UN ESTUDIO PANEL

Para que el sitio web cumpla con los estándares internacionales debiese mezclar funcionalidad con estilo. Esto se logra pensando en este sitio web como la cara visible de la información recolectada, que refleje el espíritu de la investigación.

La información que debe estar presente está dirigida a esclarecer temas como:

- Descripción
- Objetivos
- Metodología
- Cuestionarios
- Estadísticas
- Definiciones
- Documentos
- Base de datos
- Documentación para el correcto uso de las bases de datos.

En resumen y basándose en los documentos propuestos anteriormente se enumeran los temas que el sitio web debiese contener:

1. Introducción y descripción del estudio y los objetivos de la investigación
2. Información del contexto de la encuesta
3. Metodología y diseño de la encuesta
4. Recolección y procesamiento de los datos
5. Error de la muestra, información acerca de peso e imputación
6. Información descriptiva

7. Publicación y detalles del análisis
8. Consejos para usar datos linkeados
9. Declaración de confidencialidad y ética
10. Manual del Usuario

Se sugiere que el sitio web esté compuesto por el siguiente esquema:

- Páginas con una interfaz que permita hacer visible todos los elementos de la información en la documentación de ELPS expuesto anteriormente. Esto se considera material esencial de este portal.

Además, se sugiere una serie de mecanismo enfocados al buen uso de los datos;

- Página que permita registrar un conjunto de atributos o características de los potenciales usuarios que soliciten disponer de la información recolectada.
- Página que permita descargar la información, a un usuario autorizado, utilizando para ello el sistema tradicional (username y password).
- Página con todos los antecedentes que permitan al usuario hacer un buen uso de la información, por ejemplo, el cuestionario, el manual del encuestador, los documentos desarrollados, etc.

Es recomendable tener un registro sobre la cantidad y nivel de publicaciones que se efectuarán a partir de las bases de datos, es por eso que con el registro de los potenciales usuarios se debiera tener la herramienta necesaria para intentar crear un mecanismo de seguimiento y conocer cuál fue el resultado final del uso de estas bases de datos.

7.3 SEMINARIO Y TALLERES

El paso de la diseminación debe favorecer diferentes miradas de las bases de datos y en general de la información que aporta esta encuesta. Esta visión desde distintas perspectivas no debe ser casual sino que intencionada y prevista desde los orígenes de la investigación. Por tal motivo la documentación de los datos debe motivar esta mirada distinta y debe gatillar nuevos horizontes investigativos. En complemento a lo anterior se hace necesario socializar resultados más concretos frente a la opinión pública. La manera de entregar los resultados de la ELPS en formato de seminarios y talleres favorece la multiplicidad de interpretaciones acerca de los fenómenos rescatados por el estudio. Además permite corregir y enseñar de mejor manera la forma en que se deban utilizar los datos.

El formato de seminario incrementa el impacto que tendrán los resultados a nivel país. Es atractivo concebir estas instancias como tribuna para que actores relevantes de la sociedad generen un discurso a partir de los datos levantados. Los medios de comunicación a su vez juegan un rol fundamental en este proceso de difusión de la información, que se verá potenciado por instancias como seminarios, que otorgan una oportunidad concreta de cobertura periodística.

El formato de talleres debieran estar más enfocado a nichos específicos de usuarios de la información centrándose en perfeccionar el uso de ésta. Los talleres si bien están pensados para un público amplio, pretenden ahondar en temas más técnicos sobre todo pensando en enseñar el buen uso y manejo de las bases de datos para análisis y validaciones posteriores.

7.4 EXPERIENCIA EXITORA DE DISEMINACIÓN Y PROMOCIÓN DE DATOS LONGITUDINALES

La estrategia utilizada en la primera encuesta longitudinal realizada en Latinoamérica para diseminar los resultados principales y las bases de datos entrega varias líneas de acción que pueden ser recomendaciones prácticas para la promoción de la ELPS en cada país y a nivel regional.

Los resultados de la tercera ronda de la Encuesta Panel de Caracterización Socioeconómica de Chile (EPCASEN) fueron presentados ante una audiencia experta en políticas sociales que bordeaba las 200 personas. Ese mismo día, a través de un portal web, se pudo acceder a las bases de datos (en formatos STATA y SPSS), los cuestionarios 2001 y 2006, un manual de uso de las bases y un análisis acerca de la atrición de la muestra. Esta completa gama de información que se puso a disposición de quien estuviese interesado. Todo esto se complementó con una minuta de resultados y una presentación de estos, ambos archivos dirigidos a entregar un acercamiento real de este tema a personas no expertas en esta área del conocimiento.

La EPCASEN 1996, 2001, 2006 era algo inédito en Chile y cuyos resultados despertaron gran atención en diversas esferas del saber.

La presentación de los resultados de la primera fase de análisis, dejó al descubierto que los estudios de este tipo son una fuente de información que puede ser utilizada por diversos profesionales y con distintos fines enfocados a mejorar la calidad de la investigación de la realidad socioeconómica del país.

Para desarrollar de mejor manera las razones que llevan a decir que esta fue una experiencia exitosa es necesario dividir arbitrariamente el análisis en dos: el Mundo Académico y el Mundo Social.

Mundo Académico

Para hablar de este punto hace falta remitirse a algunas cifras. Después de un año que las bases de datos se hicieron públicas alrededor de 160 personas habían accedido a ellas y las habían usado para análisis en sus diferentes campos. Se implementó un sistema en que los interesados(as) debieron completar una ficha de solicitud como la que se observa a continuación:

http://www.osuah.cl - Ficha de inscripción OSUAH - Mozilla Firefox

Solicitud Encuesta Panel CASEN 1996,2001,2006

Nombre :

Apellidos :

Correo Electrónico :
(se usará para acceder)

Teléfono:

Dirección :

Ciudad :

País :

Nombre de la institución donde trabaja:

Clasificación de la institución donde trabaja : Universidad

¿Cuál es su cargo o función en esta institución?: profesor

Indique su área principal de investigación: antropología

Indique el grado académico más alto alcanzado: Pregrado

Describe el objetivo de la investigación que realizará:

Terminado

El objetivo de este procedimiento era poder llevar un registro que permita monitorear quiénes y en qué áreas se está trabajando con los datos. Esto representa una serie de ventajas, dentro de las que destaca la capacidad de generar un diálogo y un intercambio de los estudios e investigaciones que

puedan surgir a partir de esta información. Además permite a la institución que proporciona las bases de datos entregar un servicio personalizado en caso de dudas, preguntas y sugerencias.

Del total de las personas que completaron la solicitud y accedieron a las bases de datos, cerca de un 93% estaba en Chile. Un 37% eran personas que cursan programas de magíster, en tanto que un 45% son alumnos de pregrado y alrededor de 10% de doctorado, el resto corresponden a otro tipo de estudios. Respecto de las instituciones a las que pertenecen las personas que requirieron las bases de datos, un 50% corresponde a universidades a lo largo de todo Chile. Cerca de un 22% son del sector público, 8% del sector privado y un 15% a centros de estudios. El resto no registra institución. Dentro de las áreas de investigación en que se desempeñaban quienes tienen las bases de datos, cerca de un 30% son del mundo de la economía seguido por un 23% del área sociología.

Mundo Social

La presentación de los resultados de la EPCASEN, estuvo precedida por una masiva conferencia de prensa. Cerca de 15 medios de comunicación hicieron eco de la información proporcionada Y cubrieron en sus secciones de economía y crónica los datos entregados por este estudio. Las apariciones en prensa escrita, radio y televisión estuvieron marcadas por el énfasis en el núcleo de “pobreza dura” (4,4% al cabo de las tres mediciones) y por las cifras referidas a vulnerabilidad (por ejemplo el 34,1% de las personas fueron pobres, alguna vez, en la última década).

La señal que se recoge, después de la presentación de los resultados de la primera fase de análisis, es que la información que entrega este tipo de investigaciones representa un tema de interés para ser comunicado, lo que refuerza la idea de que en la agenda pública los temas relacionados con políticas sociales tienen más cabida de lo que se cree, siempre y cuando se hable de datos frescos, inéditos y de calidad.

Es por esto que las encuestas longitudinales abren toda una nueva veta del conocimiento, no sólo para el mundo académico, sino que para el común de la gente, que poco a poco se debe ir a acostumbrando a reconocer las ventajas comparativas de este modo de acercarse a la realidad social.

Ejemplos de algunos registros de prensa aparecidos al día siguiente de la presentación de los resultados de la EPCASEN junto a la liberación de las bases de datos y documentación.



7.5 HOMOGENIZACIÓN Y HOMOLOGACIÓN DE LAS BASES DE DATOS DE LA ELPS ENTRE PAÍSES

Una estrategia de largo aliento para aumentar el número de estudios que utilizan la ELPS en cada país es armonizar las bases de datos que se han levantado en una sola base de datos de medidas de equivalencia. Las medidas de equivalencia permitirían que una variable de la ELPS de un país sea comparable con la variable de otra ELPS. De esta manera, el análisis comparado de los países en temas de seguridad social enriquecería profundamente la discusión para el diseño de políticas sociales en el ámbito de la protección social. La tarea que involucra la homogenización y homologación de las variables de las distintas rondas entre países es ardua y compleja. En general, este trabajo se realiza ex-post, y las personas responsables de esta delicada tarea invierten considerable tiempo y esfuerzo en definir las medidas de equivalencias de los conceptos y definiciones que ha utilizado cada país para adaptar su realidad a la ELPS.

Una de las experiencias internacionales más conocidas en este ámbito es el *Luxembourg Income Study* (LIS). Este proyecto comenzó el año 1983 y ha logrado armonizar datos de encuestas representativas de más de 30 países (Smeeding et al., 2002). Esta base de datos con información socioeconómica de los países participantes es utilizada por una amplia gama de organizaciones e investigadores porque reduce considerablemente el presupuesto de cualquier estudio. La razón está dada porque un investigador secundario no tiene que invertir recursos propios en armonizar las bases de datos de los países que le interesa comparar. Cabe destacar que el LIS ha armonizado solo bases de datos de encuestas de corte transversal y no encuestas panel.

Actualmente ocho países que están desarrollando encuestas longitudinales de hogares han logrado armonizar sus bases de datos para hacer los paneles comparables entre sí. El proyecto se llama *Cross-National Equivalent File* (CNEF) y es una gran base de datos que agrupa armonizadamente la información de las principales encuestas longitudinales de Alemania, Australia, Canadá, Corea, Estados Unidos, Gran Bretaña, Suiza y Rusia (Frick et al., 2007). Es una iniciativa que el Observatorio Regional de Protección Social debiera obtener innumerables sugerencias de querer desarrollar una base de datos que reúna las distintas ELPS.

8. Referencias

- Alderman, H., Behrman, J. R., Kohler, H.-P., Maluccio, J. A., & Watkins, S. C. (2001). Attrition in longitudinal household survey data. *Demographic Research*, 5(4), 79-124.
- Bethlehem, J. (2009). *Applied survey methods: A statistical perspective*. John Wiley & Sons, Inc., Hoboken, New Jersey. Disponible en <http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470373083.html>
- Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to survey quality*. John Wiley & Sons, Inc., Hoboken, New Jersey. Disponible en <http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471193755.html>
- Couper, M. P., & Ofstedal, M. B. (2009). Keeping in contact with mobile sample members. *Methodology of longitudinal surveys*, 183-203.
- De Heer, W. (1999). International response trends: results of an international survey. *JOURNAL OF OFFICIAL STATISTICS-STOCKHOLM-*, 15, 129-142.
- Eckman, S. (2013). Paradata for coverage research. In F. Kreuter (Ed.), *Improving surveys with paradata. Making use of process information* (pp. 97– 116). John Wiley & Sons, Inc., Hoboken, New Jersey. Disponible en <http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470905417.html>
- Falaris, E. M. (2003). The effect of survey attrition in longitudinal surveys: evidence from Peru, Côte d'Ivoire and Vietnam. *Journal of Development Economics*, 70(1), 133-157. Disponible en <http://www.sciencedirect.com/science/article/pii/S0304387802000792>
- Fitzgerald, J., Gottschalk, P., & Moffitt, R. A. (1998). An analysis of sample attrition in panel data: The Michigan Panel Study of Income Dynamics. National Bureau of Economic Research Cambridge, MA, USA. Disponible en <http://www.nber.org/papers/t0220>
- Frick, J. R., Jenkins, S. P., Lillard, D. R., Lipps, O., & Wooden, M. (2007). The Cross-National Equivalent File (CNEF) and its member country household panel studies. *Schmollers Jahrbuch*, 127(4), 627-654.
- Gregg, P., Machin, S., & Manning, A. (2004). Mobility and joblessness *Seeking a premier economy: The economic effects of British economic reforms, 1980-2000* (pp. 371-410): University of Chicago Press.
- Groves, R. M. (1989). *Survey errors and survey costs*. John Wiley & Sons, Inc., Hoboken, New Jersey. Disponible en <http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471678511.html>
- Groves, R. M., & Couper, M. P. (1998). *Nonresponse in Household Interview Surveys*. John Wiley & Sons, Inc., New York. Disponible en <http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471182451.html>

- Groves, R. M., Dillman, D. A., Eltinge, J. L., & Little, R. J. A., (eds). (2002). *Survey Nonresponse*. John Wiley & Sons, Inc., New York. Disponible en <http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471396273.html>
- Hill, Z. (2004). Reducing attrition in panel studies in developing countries. *International journal of epidemiology*, 33(3), 493-498. Disponible en <http://ije.oxfordjournals.org/content/33/3/493.full>
- Jackle, A., & Lynn, P. (2004). *Dependent interviewing and seam effects in work history data*.
- Kreuter, F. (2013). *Improving surveys with paradata: Analytic uses of process information*. John Wiley & Sons, Inc., Hoboken, New Jersey. Disponible en <http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470905417.html>
- Kreuter, F., & Casas-Cordero, C. (2010). Paradata. Working paper series of the German Council for Social and Economic Data 136, German Council for Social and Economic Data (RatSWD). Disponible en http://www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_136.pdf
- Kreuter, F., Couper, M., & Lyberg, L. (2010). *The use of paradata to monitor and manage survey data collection*. Paper presented at the Proceedings of the Joint Statistical Meetings, American Statistical Association. Disponible en http://www.amstat.org/Sections/Srms/Proceedings/y2010/Files/306107_55863.pdf
- Kreuter, F., & Kohler, U. (2009). Analyzing contact sequences in call record data. Potential and limitations of sequence indicators for nonresponse adjustments in the European social survey. *Journal of Official Statistics*, 25(2), 203. Disponible en <http://www.jos.nu/Articles/abstract.asp?article=252203>
- LaRoche, S. (2003). Longitudinal and Cross-Sectional Weighting of the Survey of Labour and Income Dynamics. Cat.No. 75F0002MIE, No. 007, Statistics Canada. Disponible en <http://www.statcan.gc.ca/pub/75f0002m/75f0002m2007007-eng.htm>
- Laurie, H., & Lynn, P. (2009). The use of respondent incentives on longitudinal surveys. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 205-233).
- Laurie, H., Smith, R., & Scott, L. (1999). Strategies for reducing nonresponse in a longitudinal panel survey. *Journal of Official Statistics*, 15(2), 269-282. Disponible en <http://repository.essex.ac.uk/8747/1/document.pdf>
- Lepkowski, J. M., & Couper, M. P. (2002). Nonresponse in the second wave of longitudinal household surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 259-272). John Wiley & Sons, Inc., New York. Disponible en <http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471396273.html>
- Lynn, P. (2001). Standard quality profiles for longitudinal surveys. Institute for Social and Economic Research, University of Essex, Colchester.
- Lynn, P. (2009). Methods for Longitudinal Surveys. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 1-19). Chichester, UK: John Wiley & Sons.

- Lynn, P., Buck, N., Burton, J., Laurie, H., & Uhrig, S. C. N. (2006). Quality Profile: British Household Panel Survey Version 2.0: Waves 1 to 13: 1991-2003. Institute for Social and Economic Research University of Essex.
- Neter, J., & Waksberg, J. (1964). A study of response errors in expenditures data from household interviews. *Journal of the American Statistical Association*, 59(305), 18-55.
- Nicoletti, C., & Buck, N. (2004). Explaining interviewee contact and co-operation in the British and German household panels. ISER Working Paper Series.
- Nicoletti, C., & Peracchi, F. (2005). Survey response and survey characteristics: microlevel evidence from the European Community Household Panel. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(4), 763-781.
- Paredes, R., Prieto, J., & Zubizarreta, J. R. (2006). Attrition in Longitudinal Data and Income Mobility in Chile. Mimeo, Observatorio Social, Universidad Alberto Hurtado. Disponible en https://www.dropbox.com/s/vezpy7te49rj4zy/06-Paredes_et_al.pdf?dl=0
- Smeeding, T. M., Jesuit, D. K., & Alkemade, P. (2002). The LIS/LES Project Databank: Introduction and Overview. *Schmollers Jahrbuch*, 122(3), 497-517.
- Taylor, M. F. (2000). Dissemination issues for panel studies. Metadata and documentation. In D. Rose (Ed.), *Researching Social and Economic Change. The use of Household Panel Studies* (pp. 146-162). London and New York: Routledge.
- Thomas, D., Frankenberg, E., & Smith, J. P. (2001). Lost but not forgotten: Attrition and follow-up in the Indonesia Family Life Survey. *Journal of Human Resources*, 556-592.
- Thomas, D., Witoelar, F., Frankenberg, E., Sikoki, B., Strauss, J., Sumantri, C., & Suriastini, W. (2012). Cutting the costs of attrition: Results from the Indonesia Family Life Survey. *Journal of Development Economics*, 98(1), 108-123.
- Uhrig, S. N. (2008). The nature and causes of attrition in the British Household Panel Survey. (No. 2008-05). ISER Working Paper Series. Disponible en <http://hdl.handle.net/10419/92025>
- UN, (United Nations). (2011). Population Distribution, Urbanization, Internal Migration and Development: An International Perspective. Department of Economic and Social Affairs. Population Division. New York. Disponible en <http://www.un.org/esa/population/publications/PopDistribUrbanization/PopulationDistributionUrbanization.pdf>
- Wagner, J., & Olson, K. (2001). *Where do interviewers go when they do what they do? An analysis of interviewer travel in two field surveys*. Paper presented at the Proceedings of the American Statistical Association, Section on Survey Research Methods, Miami, FL. https://www.amstat.org/sections/srms/proceedings/y2011/Files/301565_65170.pdf
- Watson, D. (2003). Sample attrition between waves 1 and 5 in the European Community Household Panel. *European Sociological Review*, 19(4), 361-378. Disponible en <http://esr.oxfordjournals.org/content/19/4/361.short>

- Watson, N., & Wooden, M. (2009). Identifying factors affecting longitudinal survey response. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 157-182). John Wiley & Sons, Chichester, UK. Disponible en <http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470018712.html>
- Witoelar, F. (2011). Tracking in Longitudinal Household Surveys. *The World Bank, Development Research Group, LSMS-ISA Integrated Surveys on Agriculture. Washington, DC*. Disponible en http://siteresources.worldbank.org/INTSURAGRI/Resources/7420178-1294259038276/Tracking_in_Longitudinal_Household_Surveys.pdf