

2005

longitudinal surveys
methodology

luzetarako inkesten
metodologia

metodología de las
encuestas
longitudinales

PETER LYNN

45

Índice

TABLAS.....	2
FIGURAS.....	2
1. ¿QUÉ ES UNA ENCUESTA LONGITUDINAL?	3
1.1 Puntos fuertes de las encuestas longitudinales.....	5
Puntos fuertes de la recogida de datos	5
Puntos fuertes del análisis	7
1.2 Los puntos débiles de las encuestas longitudinales.....	10
Puntos débiles de la recogida de datos	10
Puntos débiles del análisis.....	11
1.3 Estructura de errores en las encuestas.....	12
2. DISEÑO DE ENCUESTAS Y DISEÑO MUESTRAL PARA ENCUESTAS LONGITUDINALES	15
2.1 Paneles fijos	15
2.2 Paneles fijos más “nacimientos”	15
2.3 Paneles Repetidos	18
2.4 Paneles rotativos.....	19
2.5 Paneles divididos (Split panel)	22
2.6 Diferencias entre los diseños	23
2.7 Representación de los “nacimientos”	24
Cobertura de las nuevas altas	25
Cobertura de las altas de corto plazo.....	25
3. FALTA DE RESPUESTA Y DESGASTE DE LAS ENCUESTAS LONGITUDINALES.....	27
3.1 Componentes de la falta de respuesta	27
Razones para una falta de contacto	28
Razones para negarse a cooperar con una encuesta.....	30
3.2 Características especiales de la falta de respuesta en las encuestas longitudinales.....	32
Falta de contacto.....	32
Negativas.....	33
3.3 Patrones de la falta de respuesta en las encuestas longitudinales	35
4. ERRORES DE MEDICIÓN EN LAS ENCUESTAS LONGITUDINALES.....	41
4.1 Introducción a los errores de medición.....	41
4.2 Errores de medición en las estimaciones de los cambios	43
4.3 El efecto junta (Seam effects).....	45
4.4 Métodos para la reducción los errores de medición en las encuestas longitudinales	47
Entrevistas dependientes	47
Calendarios y métodos relacionados.....	57
5. PESOS Y ATRIBUCIONES DE LAS ENCUESTAS LONGITUDINALES.....	62
5.1 Introducción a la ponderación.....	62
5.2 Pesos longitudinales	62

5.3 ¿Qué combinaciones de ciclos?	64
5.4 ¿Qué variables?	66
5.5 Introducción a la Imputación	67
5.6 Imputación longitudinal.....	68
5.7 Revisión de las imputaciones	69
REFERENCIAS	71

TABLAS

Tabla 1: Beneficios percibidos e inconvenientes de la participación en una encuesta.....	32
Tabla 2: Error de medición de una variable categórica	43
Tabla 3: Matriz de transición verdadera	43
Tabla 4: Matriz de transición observada, con medidas sujetas a error de medición.....	44
Tabla 5: Ingresos: informe de ciclo t condicional a informe de ciclo $t-1$	51
Tabla 6: Indicadores de cobro de ingresos provenientes de encuestas y de datos administrativos	52
Tabla 7: Características de los informantes independientes y reactivos en cuanto a fuentes de ingreso.....	53
Tabla 8: Porcentaje que declara cambios en las características de su trabajo	55
Tabla 9: Correlativos socio demográficos del efecto de las entrevistas dependientes sobre las estimaciones del cambio	56
Tabla 10: Poblaciones de estudio y muestras	63

FIGURAS

Figura 1: Marco de errores de las encuestas	13
Figura 2: Errores de Falta de Observación	14
Figura 3: Diseño de Panel Fijo.....	16
Figura 4: Diseño de Panel Fijo más “Nacimientos”	16
Figura 5: Diseño de Panel Repetido.....	18
Figura 6: Diseño de Panel rotativo: 1-1-1 Patrón de Rotación	20
Figura 7: Diseño de Panel rotativo: 1-1-0-0-1-1 Patrón de Rotación	22
Figura 8: Diseño de Panel dividido (Split panel).....	23
Figura 9: Muestreo de los nuevos nacimientos.....	26
Figura 10: Posibles patrones de respuesta por ciclo según la política 1: todas las unidades en todos los ciclos.....	37
Figura 11: Posibles patrones de respuesta por ciclo según la política 2: entrevistados del ciclo 1 en todos los ciclos	37
Figura 12: Posibles patrones de respuesta por ciclo según la política 3: sólo entrevistados del ciclo anterior	38
Figura 13: Patrones de respuesta por ciclos del Estudio de Cohortes 5 para Jóvenes de Inglaterra y Gales	38
Figura 14: Un calendario histórico de sucesos	60

1. ¿QUÉ ES UNA ENCUESTA LONGITUDINAL?

Una *encuesta longitudinal* es aquella que reúne datos, a partir de los mismos elementos muestrales, en múltiples ocasiones a través del tiempo. No se trata de un concepto contérmino con el de los *datos longitudinales*. Los datos longitudinales son datos que se *remiten* a los mismos elementos muestrales en múltiples ocasiones a través del tiempo, pero no es necesario que hayan sido *recogidos* en múltiples ocasiones. Pueden haber sido recogidos en una sola ocasión, mediante la memoria retrospectiva o recopilando información proveniente de registros. Por lo tanto, el rasgo distintivo de las encuestas longitudinales es el proceso de recogida de datos, aunque las encuestas longitudinales producen, por supuesto, datos longitudinales. Es por este motivo que, en lo que respecta a la metodología de las encuestas longitudinales, han de tomarse en consideración muchos aspectos de dichos datos.

Las encuestas longitudinales varían enormemente en términos de la naturaleza de la información recogida, la naturaleza de la población que se está estudiando, y los objetivos primarios (Binder, 1998). Algunos ejemplos que pueden ilustrar dichas variaciones son:

- ***Las encuestas sobre negocios realizadas por las oficinas de estadística nacionales o regionales.*** Estas encuestas tienden a reunir un abanico limitado de información, normalmente restringido a unos indicadores económicos clave. Los datos pueden ser recopilados a intervalos frecuentes, es decir, mensual o trimestralmente. Sus objetivos principales son por lo general la publicación de series regulares de estadísticas sobre los totales, las medias y los cambios netos entre los períodos, a menudo relativos a los campos que se interrelacionan, como son los de las regiones y las industrias;
- ***Las encuestas para jóvenes que acaban el colegio, los graduados y los que están en período de prácticas.*** Es frecuente que las instituciones que ofrecen educación o formación, como son las universidades u organismos gubernamentales responsables de la política relativa a este campo, quieran evaluar los resultados de dicha educación o formación a un micro nivel (estudiantes). Estos resultados son,

por lo general, de medio o largo plazo y por ello es necesario mantenerse en contacto con los estudiantes/practicantes durante algún tiempo después haber terminado sus estudios. A menudo se utilizan estudios longitudinales con este fin, reuniendo datos de los estudiantes/practicantes en diversas ocasiones, tal vez empezando en la época en que todavía son estudiantes/practicantes y durante los años posteriores a la finalización de sus cursos. La información que se recoge es a menudo bastante compleja, incluyendo tal vez los historiales completos de sus trabajos y actividades entre cada ciclo de encuesta y tal vez también las razones de los cambios y de las decisiones que se tomaron. A veces se incluye en la encuesta uno o más grupos de 'control', en un intento por evaluar el impacto de la educación/formación;

- **Las encuestas de los hogares panel.** En muchos países se llevan a cabo encuestas panel a largo plazo sobre la población general de los hogares. La más antigua, la Encuesta Panel sobre la Dinámica de la Renta (Panel Survey of Income Dynamics, PSID), de los EEUU, lleva entrevistando a las mismas personas desde 1968. Estas encuestas tienen un propósito general y muy diverso, y recoge datos de conducta, actitudes y circunstancias sobre un abanico de temas sociales y económicos. El objetivo principal es proporcionar un material rico en datos que puedan ser utilizados por una amplia variedad de usuarios, para una serie de propósitos. La estructura de los datos es compleja, e incluye entrevistas con cada persona del hogar de cada miembro de la muestra en cada ciclo, además de información del nivel del hogar y, a menudo, otros instrumentos para encuestas, como son los cuestionarios que cada persona debe cumplimentar o mediciones relacionadas con la salud.

También existe una variación considerable entre las encuestas longitudinales en cuanto a las limitaciones de índole práctico, tan diversas como el nivel de los recursos financieros disponibles, los medios a través de los cuales se puede acceder a la población del estudio, y las normas sobre la responsabilidad del entrevistado. Toda esta variación tiene como resultado una amplia variedad de diseños para encuestas, como se puede apreciar en la sección 2.

1.1 Puntos fuertes de las encuestas longitudinales

Los puntos fuertes de las encuestas longitudinales están asociados con la recogida de datos y las posibilidades de análisis que no se pueden conseguir con encuestas transversales o que no se pueden conseguir de una manera suficientemente fiable y precisa.

Puntos fuertes de la recogida de datos

- a) Es posible reunir historias continuadas mucho *más largas* sobre sucesos y transiciones que las que se podrían reunir retrospectivamente en una sola entrevista, solamente por el volumen de los datos involucrados (y de allí la longitud de la entrevista o cuestionario);
- b) Es posible reunir datos más *precisos* de los que sería posible en una sola entrevista con recuerdos retrospectivos, en los cuales los datos podrían estar sujetos a importantes errores de memoria.
- c) Es posible reunir información sobre expectativas y alternativas que no estén afectadas por sucesos y resultados posteriores, y también sobre los sucesos y resultados posteriores de las mismas unidades muestrales.

En particular, la duración de la historia que se está recopilando y la precisión de los datos están normalmente interrelacionados. Casi todas las preguntas de las encuestas requieren algún tipo de recuerdo por parte del entrevistado. Pero en la medida en que las circunstancias o sucesos que son el objeto de las preguntas sean actuales o recientes, menor será la confianza que se le atribuirá a los recuerdos del entrevistado. Si el propósito de la encuesta es reunir un registro de todo lo que sucedió en un tipo de suceso en particular, el margen de error de los informes del entrevistado dependerá del *período de recuerdo* y de la *relevancia* de los sucesos en cuestión. En el caso de un suceso de gran relevancia, como el de dar a luz o el del día de la boda, la mayoría de entrevistados podrá recordar los detalles esenciales de dicho suceso muchos años después. En esos casos, será fácil recoger el historial de fertilidad y matrimonio a lo largo de la vida en una sola entrevista y obtener datos con una precisión razonable. Pero cuando se trata de sucesos menos relevantes, como son la compra diaria de los alimentos, puede ser que los detalles básicos permanezcan en la memoria

del entrevistado sólo durante unos días. En una sola entrevista, sólo se podrán obtener datos de una precisión razonable sobre las compras de los últimos días. Si se formulan preguntas sobre las compras de rutina durante un período más largo, las respuestas (en caso de que los miembros de la muestra están dispuestos a darlas) estarán sujetas a un margen de error mayor. Por ello, si se han de reunir los historiales del comportamiento relacionado con la compra durante un período de semanas o meses, con una precisión razonable, es necesario recoger los datos de los entrevistados a intervalos regulares, probablemente al menos una vez por semana.

Por supuesto, muchos sucesos tienen una relevancia intermedia entre un nacimiento y la compra de rutina. Muchas personas pueden recordar con una precisión razonable los períodos de empleo, desempleo, educación y otras situaciones de sus actividades durante varios meses y posiblemente años. Pero, por supuesto, la *dificultad* de la tarea de recordar variará de acuerdo con los entrevistados, dependiendo del número y de la naturaleza de los sucesos que han experimentado. Esta variación puede ser verdaderamente considerable cuando se trata de historiales de actividades económicas, planteando un dilema para los diseñadores de las encuestas. Si una encuesta busca reunir historias completas sobre las actividades, a lo largo de varios años, de una muestra de personas que pueden tener unas experiencias que varíen notablemente, como una sección transversal de la población general, el intervalo ideal entre los ciclos de la encuesta será muy distinto para los distintos miembros de la muestra. Pero es difícil poder predecir esto con anterioridad, y tampoco es práctico, a veces, tener distintos intervalos entre ciclos para individuos diferentes. En vez de ello, se elige un intervalo estándar. Las entrevistas a intervalos anuales pueden no ser efectivas en el caso de las personas cuyas circunstancias no varían mucho (por ejemplo, para los jubilados o para quienes permanecen en el mismo trabajo durante muchos años). La cantidad marginal de información que se recoge en cada entrevista, en comparación con el coste, será pequeña. Pero las entrevistas anuales pueden suponer un esfuerzo de memoria considerable para las personas que experimentan muchos períodos cortos de empleo, quizás intercalados con períodos de desempleo u otras actividades. Así, para obtener el máximo beneficio de la capacidad que tienen las encuestas longitudinales de recopilar historias más largas y / o más precisas de los sucesos, el encargado de diseñar la encuesta necesita entender cuáles son las tareas de memoria y narración que se le pedirán a los entrevistados, además de la relación que tiene esta tarea con el intervalo entre ciclos y cómo puede variar ello de acuerdo con los miembros de la muestra.

Por muchas razones, un fechado preciso de los sucesos es al menos tan importante como un recuerdo preciso de los detalles del suceso. Pero es posible que los entrevistados no puedan recordar con precisión la fecha de un suceso específico, aunque sí puedan recordar el suceso en sí. Por ello, las preguntas de recuerdo retrospectivo formuladas en una sola entrevista pueden dar como resultado estimaciones de frecuencias y medidas asociadas que resulten sesgadas. Un fenómeno del cual se ha informado muchas veces es el del 'solapamiento', en el cual los entrevistados narran sucesos como si hubieran tenido lugar dentro de un período de referencia cuando, de hecho, sucedieron hace más tiempo. La encuesta panel ofrece una ventaja más, al recoger las fechas de los sucesos. Cada entrevista después de la primera queda 'delimitada' por la anterior, de manera que cada suceso previo puede ser descartado de los informes de la entrevista actual para evitar así el solapamiento. Con ello se supone, por supuesto, que se concluirá sin ambigüedades si los informes de dos entrevistas consecutivas narran el *mismo* suceso. A veces es difícil de lograr, especialmente cuando un entrevistado tiende a vivir sucesos de la misma naturaleza con frecuencia, pero en el caso de muchos tipos de datos de encuestas, sí se puede lograr.

A menudo, lo que se busca son datos sobre expectativas y alternativas (véase c) más arriba), con el fin de presentar un informe de evaluación de los resultados de varios tipos y para ayudar a entender los procesos que originan esos resultados. Para los entrevistados es muy difícil recordar cuáles eran sus expectativas en un momento determinado o las razones por las cuales tomaron algunas decisiones. En vez de ello, se da una tendencia, en muchas situaciones, que es la de "volver a crear" las razones a la luz de las experiencias posteriores. Si a raíz de alguna decisión que tomó el entrevistado en un momento determinado se obtuvo algún resultado positivo, el entrevistado puede racionalizar, *post hoc*, que la razón para tal o cual decisión debe haber sido la de conseguir dicho resultado. Por ello, las preguntas relativas a las expectativas y motivaciones de una persona deben hacerse en esa misma época. Es por ello que sólo una encuesta longitudinal puede vincular las respuestas a tales preguntas con los datos sobre los resultados.

Puntos fuertes del análisis

Está claro que es artificial separar las ventajas de los análisis de los datos de las encuestas longitudinales de las ventajas de los análisis, dado que la razón para la recogida de

ciertos tipos de datos es para poder realizar determinados tipos de análisis. Las ventajas primordiales de los datos longitudinales (que, en la mayoría de los casos, sólo pueden recogerse mediante encuestas longitudinales) son analíticas e incluyen los siguientes aspectos:

- a) El análisis del cambio bruto;
- b) El análisis del cambio medio a nivel de unidad;
- c) El análisis de la estabilidad o inestabilidad en cuanto a características;
- d) El análisis de las características en términos de tiempo de los sucesos o circunstancias, como son la frecuencia, el momento y la duración;
- e) Análisis de la naturaleza ordinal de los sucesos, lo cual sirve, a menudo, para clarificar problemas de causalidad.

El análisis del cambio bruto es quizás uno de los objetivos más comunes de las encuestas longitudinales. Las encuestas transversales repetidas pueden ser utilizadas para calcular el cambio neto, como por ejemplo, el cambio de la tasa de empleo dentro de una población en particular. Pero sólo una encuesta longitudinal puede identificar hasta qué punto, dicho cambio, está compuesto por diferentes elementos del cambio bruto. Por ejemplo, si se observa la misma tasa de paro en dos puntos en el tiempo, ¿se trata de las mismas personas que están en paro en ambas ocasiones? ¿O es que existen flujos opuestos y equivalentes con y sin paro (y, en ese caso, de cuánto tiempo y qué tipo de gente los componen, etc.)? Este es el tipo de preguntas que se pueden formular en las encuestas longitudinales.

También puede resultar interesante el cambio a nivel individual, independientemente del interés en el cambio neto a nivel de la población. Por ejemplo, entender la naturaleza y las características asociadas el cambio de estado civil a nivel individual, la composición del hogar, la renta, etc., es de gran interés para muchos analistas, y sólo se puede conseguir mediante una encuesta longitudinal. Sin embargo, a menudo, un cambio a nivel individual solo puede interpretarse adecuadamente dentro del contexto de los cambios que se producen dentro de un período de tiempo considerable. Por ejemplo, una encuesta longitudinal de 2 ciclos puede ser un buen vehículo para medir el cambio de la renta personal entre dos puntos en el tiempo. Ello permitirá al analista descomponer el cambio neto de la renta en sus componentes de cambio bruto, es decir, que podrá estimar la distribución de los cambios de ingresos a nivel individual. Pero mientras que la distribución de muestras de los cambios de ingresos a nivel individual puede estimar bien la distribución de la población, cada cambio individual observado puede no

representar bien el cambio en la renta 'media' para tal individuo durante un período de tiempo. Si quisiéramos estudiar las asociaciones entre las características personales y un cambio en la renta, una única medida de cambio entre los dos puntos en el tiempo puede no ser un buen indicador de uso. Con un panel multiciclo que recoge las medidas de la renta en cada ciclo, existe la posibilidad de elaborar una medida del cambio 'medio' de la renta durante un período relativamente largo para cada miembro de la muestra o una medida de la variación del cambio.

En efecto, el grado en que unas medidas tales como la renta sean estables o inestables durante un período de tiempo es de gran importancia para la política. Las encuestas panel con muchos ciclos pueden constituir unas buenas medidas de estabilidad de muchas características. Los analistas de la pobreza, por ejemplo, han utilizado datos panel de los hogares para demostrar que existe una inestabilidad considerable a lo largo del tiempo en cuanto a la situación de la pobreza de muchos individuos y hogares en los países occidentalizados. Mientras que la proporción de hogares pobres puede ser relativamente estable a lo largo del tiempo, puede haber muchas entradas y salidas de la pobreza. Una alta proporción de hogares puede experimentar al menos un período de pobreza a lo largo de mucho tiempo, mientras que muy pocos hogares pueden permanecer continuamente en la pobreza a lo largo de todo el período. Esta percepción proporcionada por las encuestas longitudinales puede haber variado el enfoque de la política desde unas características (estables) asociadas con la propensión a la pobreza en un punto del tiempo hasta un mejor entendimiento de las dinámicas y de los factores asociados con el empobrecimiento o con la imposibilidad de salir de la pobreza.

Entender la duración de los períodos de una situación en particular y los factores asociados con la salida de dicha situación (que a veces se denomina 'persistencia') es importante no sólo para la pobreza, sino también para muchos otros problemas como son el paro, el estado civil y de pareja, la participación en educación y formación, y la rentabilidad de una empresa. El modelo aleatorio y el análisis de supervivencia son las técnicas utilizadas para entender mejor la propensión del cambio (en cualquier situación de interés fundamental) y los factores asociados con dicho cambio. Estas técnicas requieren datos longitudinales, los cuales deben provenir normalmente de encuestas longitudinales. Los investigadores quisieran poder identificar no sólo los factores *asociados* con el cambio, sino también los factores que *causan* el cambio. La comprensión de los factores causales puede tener implicaciones directas en la política. Los datos longitudinales pueden ser de gran ayuda para establecer la causalidad, ya

que se puede entender el orden cronológico de los sucesos y de los cambios. Por ejemplo, una encuesta transversal puede establecer una asociación entre A y B. Pero una encuesta longitudinal puede establecer que para la mayoría de las unidades de la población que han experimentado tanto A como B, A sucedió antes que B, siendo más probable que A causara B, y no que B hubiera sido la causa de A (aunque, por supuesto, un tercer factor, C, puede haber sido la causa, tanto de A como de B y siempre se debe considerar esta posibilidad).

1.2 Los puntos débiles de las encuestas longitudinales

Las encuestas longitudinales también tienen algunas limitaciones en relación con otras encuestas. Se debe prestar mucha atención a éstas en la etapa de diseño de la encuesta y en la etapa del análisis.

Puntos débiles de la recogida de datos

Existen dos aspectos de la recogida de datos para encuestas que son únicos de las encuestas longitudinales y potencialmente perjudiciales:

- a) El condicionamiento de los paneles.
- b) El desgaste de los paneles.

El condicionamiento de los paneles se refiere a la posibilidad de que las respuestas proporcionadas por una persona que ya ha tomado parte en la encuesta con anterioridad puedan diferir de las respuestas que esa persona hubiera dado si estuviera participando por primera vez en la encuesta. Dicho de otro modo, la respuesta puede estar 'condicionada' por la experiencia previa de su participación en una encuesta. Por ello, esto tiene relación con todos los datos recogidos en las encuestas longitudinales, aparte de los recogidos en el primer ciclo. Existen dos maneras en las que puede ocurrir este condicionamiento: la forma en que los entrevistados *informan* sobre los sucesos, el comportamiento o las características pueden cambiar. O puede cambiar el propio *comportamiento*.

Por ejemplo, una encuesta para personas en el paro puede encontrar que más personas informan que han realizado algún tipo de actividad relacionada con la búsqueda de trabajo en el segundo ciclo más que en el primer ciclo. Ello puede reflejar un aumento genuino

de la medida en que esa actividad tiene lugar (independientemente de que tome o no parte en la encuesta). Pero también puede deberse a un condicionamiento de panel. Ello obedecer a que durante la primera entrevista algunos miembros de la muestra conocieron de la existencia de posibles actividades para la búsqueda de empleo que no estaban haciendo, de manera que después empezaron a hacer este tipo de actividades. Entonces sí era cierto que se daba un aumento en el alcance de la actividad, pero sólo entre algunos miembros de la muestra, no dentro de la población como un todo. El *comportamiento* de los miembros de la muestra se ha visto condicionado por la primera entrevista. Otra posibilidad es que la experiencia de la primera entrevista pueda haber afectado la manera en que algunos miembros de la muestra respondieran a las preguntas de la segunda entrevista, aunque su verdadero comportamiento en cuanto a la búsqueda de trabajo pueda no haber cambiado. Tal vez en la primera entrevista descubrieron que si informaban que no estaban realizando ningún tipo de actividad les preguntaban otra serie de preguntas sobre por qué no participaban en dicha actividad. De manera que, para que la segunda entrevista fuera más corta, o, para evitar preguntas incómodas, ahora informan que han participado en esta actividad en particular. En este caso, el *informe* de los miembros de la muestra se ha visto condicionado por la experiencia de la primera entrevista.

El desgaste de la muestra (también denominado “desgaste del panel”) se refiere a la continua pérdida de la muestra, debido a la falta de respuesta en cada ciclo de la encuesta longitudinal. La tasa de respuesta de cada ciclo de una encuesta longitudinal puede ser tan válida como la de cualquier otra encuesta pero después, de, digamos, cinco ciclos, la proporción de unidades de muestra que han respondido en cada ciclo puede ser bastante baja. Por este motivo, la tasa de respuesta efectiva de un análisis longitudinal (para el cual se requieren datos de cada ciclo) puede ser más baja que la tasa de respuesta que normalmente se obtiene en las encuestas transversales. Más adelante, en el capítulo 3, trataremos los temas de cómo y por qué se produce un desgaste en las muestras y qué se puede hacer para minimizar sus impactos negativos.

Puntos débiles del análisis

Muchas veces, las encuestas longitudinales no son tan adecuadas como las encuestas transversales para aportar estimaciones transversales. Esto puede parecer un punto débil, pero simplemente se trata de que las encuestas longitudinales no han sido diseñadas para ello. En

comparación con las estimaciones de una encuesta transversal, las estimaciones transversales de una encuesta longitudinal (desde el ciclo 2 en adelante) tienen más probabilidades de padecer un error de cobertura (porque la muestra se eligió hace mucho tiempo y puede no haber incluido las últimas modificaciones con respecto al interés de la población). Por otra parte, una muestra de una encuesta longitudinal puede adolecer de una baja tasa de respuesta en comparación con la de una encuesta transversal (aunque ello no necesariamente implique un error de falta de respuesta mayor). El diseño de una encuesta longitudinal puede ser a menudo ajustado para mejorar la calidad de las estimaciones transversales que se pueden realizar (véase capítulo 2) aunque con mucha probabilidad ello requiera de muchos recursos y pueda desviarse de los objetivos principales de la encuesta.

1.3 Estructura de errores en las encuestas

En este apartado, presentamos un marco general para los errores de las estimaciones que derivan de una encuesta, basándonos en Grooves (1989). Esto nos proporciona el contexto de las fuentes de error que han de ser tratadas en los cuatro capítulos restantes de este texto. A lo largo de dichos capítulos, haremos referencia al marco de errores. Una encuesta de errores es simplemente una diferencia entre una estimación basada en una encuesta, \hat{Y} , del parámetro de la población y el verdadero valor del parámetro, Y . Sin embargo, durante la etapa de diseño de la encuesta estamos normalmente muy preocupados por el error de *esperado* de nuestras estimaciones, así que nos interesan más las propiedades del estimador y que un valor específico realizado de la estimación, \hat{Y} . El *error cuadrático medio* del estimador es la cantidad que se usa normalmente para medir los errores de las encuestas:

$$MSE(y) = E(y - Y)^2$$

Esto se puede descomponer como sigue:

$$\begin{aligned} MSE(y) &= E(y - E[y])^2 + (E[y] - Y)^2 \\ &= Var(y) + Bias^2(y) \end{aligned}$$

De manera que, para cada estimador de interés, el error de encuesta tiene un componente tanto de varianza como de sesgo. En la Figura 1, resumimos los distintos orígenes potenciales de error. Cada uno de estos orígenes puede contribuir a la varianza del estimador o

al sesgo, o a ambos. Esto se indica con una “B” y con una V” en la última línea del diagrama. Los orígenes de los errores pueden ser útilmente clasificados en errores debidos a una falta de observación de cada unidad de la población de estudio (“errores de falta de observación”) y los que se deben a una observación defectuosa de las unidades que se estudian (“errores de observación”).

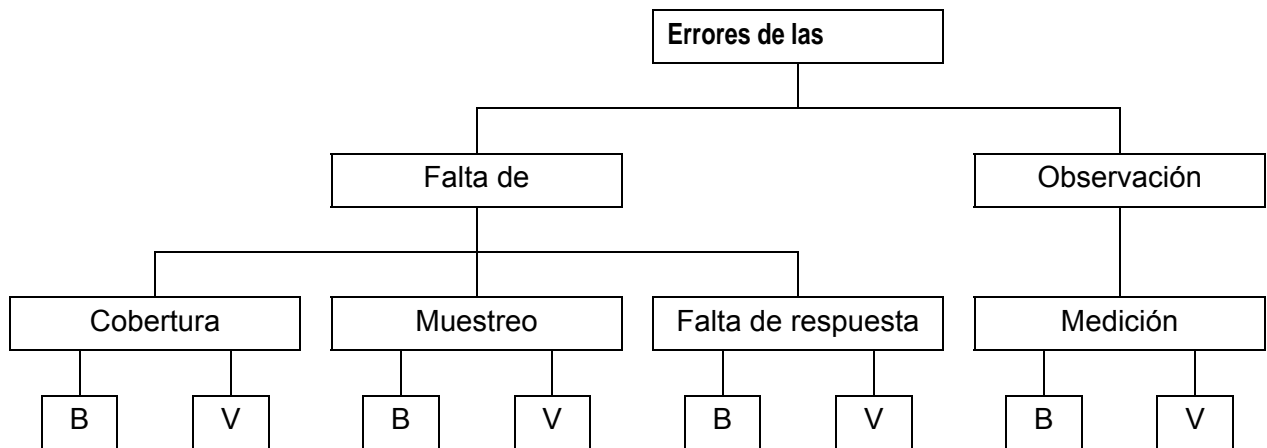


Figura 1: Marco de errores de las encuestas

Los errores de falta de observación pueden surgir en tres etapas del proceso de la encuesta. En primer lugar, el marco de muestreo o método de muestreo puede no proporcionar una cobertura completa de la población. Si algunas unidades de población tienen una posibilidad cero de ser seleccionadas, entonces, ello supondrá un error de cobertura. En segundo lugar, elegimos sólo una muestra de las unidades del marco y esto supone un error de muestreo, ya que las unidades incluidas en la muestra pueden no tener exactamente las mismas características que el conjunto completo de las unidades del marco. En tercer lugar, normalmente no logramos obtener observaciones de cada unidad incluida en la muestra: tenemos algunas faltas de respuesta, que pueden suponer un error de falta de respuesta. Estas tres etapas del proceso de observación y sus errores asociados han sido ilustrados en la Figura 2. Los errores de observación son básicamente colindantes con el error de medición, si consideramos todo el proceso de formulación preguntas, registro de las respuestas, codificación, ingreso y procesado de datos constituyen el proceso de medición (algunos autores sostienen que el error de medición proviene solamente del proceso de formulación de preguntas y registro de las respuestas y denominan los errores que surgen en etapas posteriores como *errores de procesado*. En todo caso, puede ser útil identificar y estudiar las etapas en las cuales surgen los errores).

Este es un marco general que sirve para cualquier encuesta. En el caso de las encuestas longitudinales, los orígenes de los errores son los mismos, pero la naturaleza de los errores y las técnicas que pueden utilizarse para reducirlos son totalmente distintas. Este texto se centrará en dichos errores.

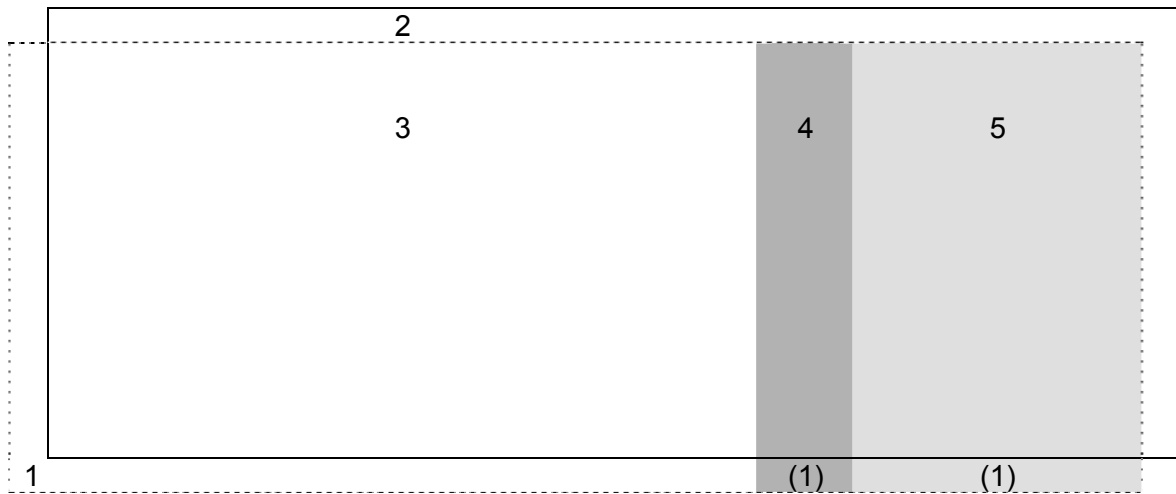
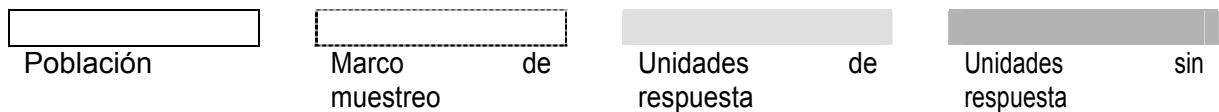


Figura 2: Errores de Falta de Observación



- 1: Sobrecobertura del marco (sin errores, si se han identificado)
- 2: Subcobertura del marco (error de cobertura)
- 3: Unidades no recogidas en la muestra (error de muestreo)
- 4: Unidades sin respuesta (error de falta de respuesta)
- 5: Unidades de respuesta (error de observación)

2. DISEÑO DE ENCUESTAS Y DISEÑO MUESTRAL PARA ENCUESTAS LONGITUDINALES

El capítulo 1 mostraba que las encuestas longitudinales pueden ser de muchos tipos. En este capítulo esbozamos los tipos más importantes de diseño y explicamos por qué se utilizan. El diseño de la encuesta longitudinal puede clasificarse en cinco amplias categorías: paneles fijos, paneles fijos con nacimientos, paneles repetidos, paneles rotativos y paneles de división.

2.1 Paneles fijos

Este panel intenta recoger los datos de las encuestas de las mismas unidades en múltiples ocasiones. Después de una selección inicial de la muestra, ya no se pueden agregar modificaciones. En principio, la única pérdida para la muestra elegible es a través de las “muertes” de la población (por ejemplo, la muerte de una persona, el cese de una empresa). Ilustramos esto en la Figura 3 con el ejemplo de un diseño de panel fijo de 5 ciclos, ignorando la falta de respuestas. Se elige una muestra de la población de interés para tomar parte en el ciclo 1. Poco después, tendrá lugar el ciclo 2. Para este momento, algunas unidades muestrales ya no están en la población de intereses y por ello ya no son elegibles para la encuesta. A menudo, nos referimos a ello como “muertes” muestrales. En el caso de una encuesta de personas, puede incluir la muerte propiamente dicha, así como la emigración, por ejemplo. En el caso de una encuesta de negocios, puede consistir en el cierre de ese negocio. Por ello, la muestra elegible para la recogida de datos del ciclo 2 es menor que la del ciclo 1. La medida en que sea menor depende del intervalo de tiempo entre los ciclos y la tasa de muertes de la población de estudio.

2.2 Paneles fijos más “nacimientos”

Es como un panel fijo, excepto porque se han añadido a la población muestras regulares de los “nacimientos” recientes. Por ejemplo, en cada ciclo de recogida de datos, se puede añadir una muestra de unidades de “nacidos” desde el ciclo anterior. Esto puede ser preferible antes que un panel fijo si hay números no triviales de nacimientos durante la vida de un panel y existe un deseo de representar la población transversal en el momento de cada ciclo, así como a la población longitudinal de las unidades que sobreviven desde el ciclo 1. La

mayoría de las encuestas panel para hogares tienen este diseño, dado que en cada ciclo se agrega una muestra de los “nacimientos” en el rango de edad elegible. La Figura 4 ilustra dicho diseño. Se puede observar que si la tasa de muertes es similar a la tasa de nacimientos (indicada en la escala vertical de la Figura 4) el tamaño muestral general elegible para la recogida de datos de cada ciclo permanece más o menos constante.

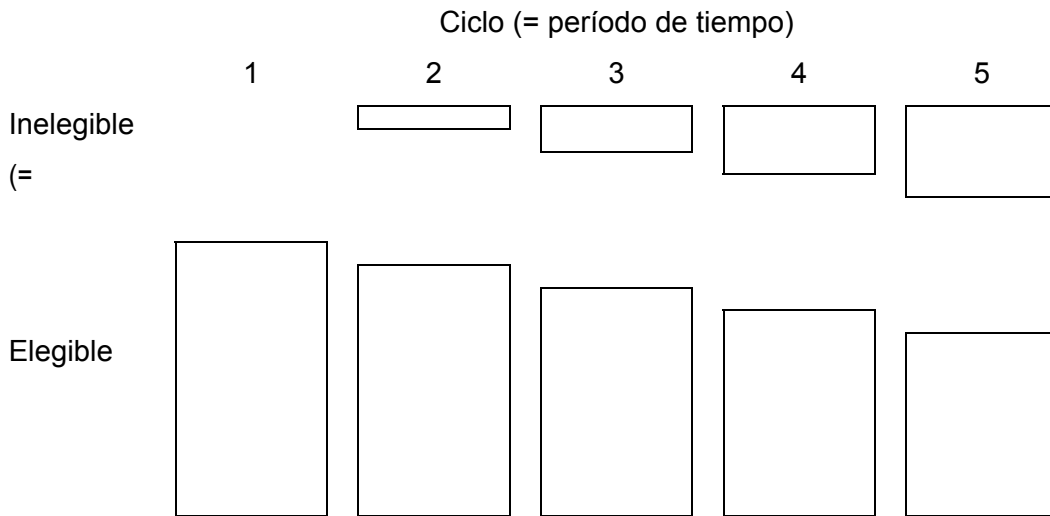


Figura 3: Diseño de Panel Fijo

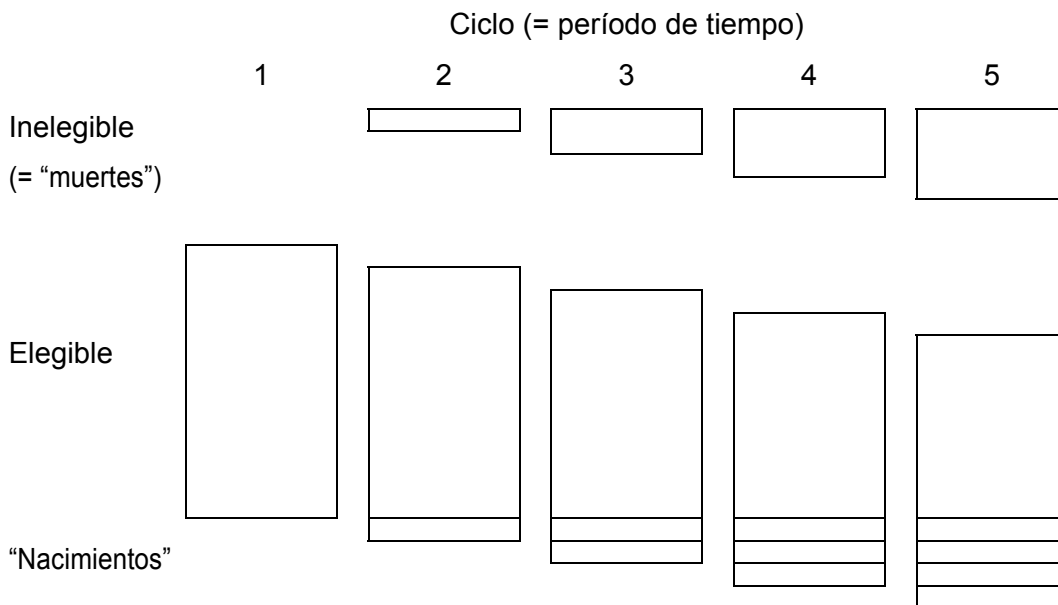


Figura 4: Diseño de Panel Fijo más "Nacimientos"

Una ventaja importante de un diseño de este tipo antes que el diseño de panel fijo es que la muestra de cada ciclo representa la población de transversal actual, facilitando que se realicen estimaciones transversales en paralelo con las estimaciones longitudinales.

En cuanto al análisis longitudinal, este diseño también presenta la ventaja de que permite una mejor representación de la población de sucesos tales como los períodos de transición que tienen lugar durante un período en particular. Por ejemplo, supongamos que la encuesta es sobre personas que viven en una ciudad en particular y que los ciclos de la encuesta tienen lugar con intervalos de un año. Supongamos también que la entrevista de la encuesta formula preguntas sobre los períodos de empleo durante el último año y que nos interesa estudiar la naturaleza de los cambios de trabajo en la ciudad durante un período de cinco años para poder comprender mejor las dinámicas del mercado de trabajo. Con un diseño de panel fijo (sin muestras adicionales de “nacimientos”), en el ciclo 2 los cambios de trabajo que observamos se limitan a personas que han vivido en la ciudad por lo menos un año. En el ciclo 3, se limitan a las personas que han vivido en la ciudad por lo menos dos años. Y así sucesivamente. Después de cinco años de la encuesta, la muestra acumulada de los cambios de trabajo será una subrepresentación de los cambios experimentados por otras personas que se acaban de trasladar a la ciudad. Los cambios de trabajo experimentados dentro del primer año del traslado a la ciudad sólo estarán representados en los datos del ciclo 1. Los cambios de trabajo vividos dentro del primer y segundo año después del traslado a la ciudad estarán representados sólo en los datos de los ciclos 1 y 2. Y así sucesivamente. Y es probable que los cambios de trabajo vividos justo después de trasladarse a la ciudad sean muy diferentes de los que han experimentado después de un tiempo viviendo en la ciudad. De manera que, la muestra observada de los cambios de trabajo estarán sesgadas con respecto a la población de todos los cambios de trabajo de la ciudad durante el período de cinco años.

Si en cada ciclo agregamos una muestra de personas que se han trasladado a la ciudad en el último año, solucionaremos este problema. La muestra de cambios de trabajo observados durante un período de años debería ser una representación de todos los cambios de trabajo (véase el apartado 2.7). Sin embargo, obsérvese que las muestras adicionales de los “nacimientos” no pueden ser utilizadas para todos los tipos de análisis longitudinales. Si deseamos estudiar el cambio a un micro nivel entre los años t y $t+4$ (por ejemplo, los ciclos 1 y 5 de la Figura 4), para lo cual necesitamos las observaciones de los dos años correspondientes a cada unidad muestral, entonces, las muestras de los nacimientos

agregados a la encuesta de cada uno de los años correspondientes no pueden contribuir al análisis.

2.3 Paneles Repetidos

Este diseño consta de una serie de encuestas panel, que pueden o no superponerse en el tiempo. Normalmente, se diseña cada panel para representar una población *equivalente*, es decir, la misma definición de población aplicada en un punto del tiempo diferente. Es frecuente encontrar este diseño en las encuestas para los jóvenes que acaban el colegio o los graduados universitarios. Cada panel consiste en una muestra de una cohorte en particular de un año de edad elegida en años diferentes, constando cada panel de por lo menos tres ciclos en los últimos tres años.

La Figura 5 ilustra un ejemplo simple de este tipo de diseño. En este ejemplo, cada panel tiene tres ciclos de recogida de datos y las fechas de estos tres ciclos corresponden con las fechas del principio de un nuevo panel (obsérvese que los nacimientos y las muertes que se muestran en la Figura 3 y la Figura 4 están omitidos en la Figura 5 por motivos de claridad aunque, en la práctica, pueden ser un rasgo de cada panel).

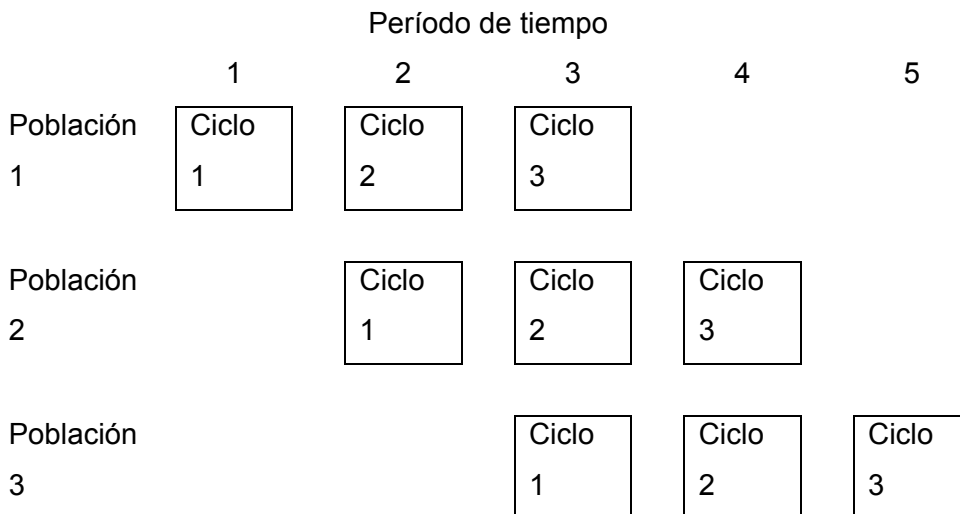


Figura 5: Diseño de Panel Repetido

La población j puede consistir, por ejemplo, en los alumnos que se gradúan en el año j . Les mandan un cuestionario en cada uno de los años j , $j+1$ y $j+2$. Pero en el año $j+1$ la nueva

población de alumno que se gradúan ese año recibe también un cuestionario por primera vez, y así sucesivamente.

A diferencia de los diseños de panel fijo y de panel fijo más los nacimientos, las muestras de las cuales se recogen los datos en un período de tiempo en particular no representan colectivamente una población relevante. En el caso del 3er año, en la Figura 5, los datos del ciclo 1 se recogen de la muestra de los graduados de 3er año, los datos del ciclo 2 se recogen de la muestra de los graduados del 2do año, y los datos del ciclo 3 se recogen de la muestra de los graduados de 1er año. Son tres muestras distintas que representan tres poblaciones deferentes. Y es muy posible que se formulen preguntas distintas en cada ciclo. Así que no se agregarían estos tres conjuntos de datos para formar las bases de ningún análisis. En vez de ello, el objetivo principal de las encuestas que tienen este diseño es normalmente entender las dinámicas del cambio por cada población (es decir, cuando se realizan análisis longitudinales de los tres ciclos de datos de una población en particular) y luego, comparar estas dinámicas de estimación entre las poblaciones.

Evidentemente, no es necesario que el intervalo de tiempo entre el inicio de cada nuevo panel sea el mismo que el intervalo de tiempo entre cada ciclo de recogida de datos de un panel. Por ejemplo, un nuevo panel de graduados puede empezar solo una vez cada dos años, mientras que la recogida de datos de cada panel puede tener lugar cada seis meses.

2.4 Paneles rotativos

Cuando se trata del diseño de un panel rotativo, se sustituyen proporciones predeterminadas de unidades muestrales en cada ocasión del trabajo de campo. Normalmente, cada unidad permanecerá en la muestra durante el mismo número de años. Se ha utilizado una gran variedad de patrones de rotación, pero daremos a conocer la idea con un sencillo patrón en el cual cada unidad permanece en la muestra durante tres ciclos, sustituyéndose un tercio de la muestra cada vez. Así lo ilustra la Figura 6. Aunque parece muy similar al diseño de panel repetido de la Figura 5, es fundamentalmente diferente, ya que cada muestra pretende representar la *misma* población. Por ello, las tres muestras entrevistadas en cualquier período pueden ser combinadas para hacer estimaciones de transversales. Las encuestas que utilizan diseños de panel rotativo recogen, por lo general, los mismos datos de cada unidad en cada ciclo, con el fin de permitir dicha combinación.

En algunas situaciones, dependiendo del intervalo entre los ciclos y la tasa de nacimientos de la población, puede ser necesario ponderar los nacimientos recientes de la muestra al realizar las estimaciones transversales. Por ejemplo, supongamos que en cada período de tiempo, el 5% de las unidades de la población “mueren” y son reemplazadas por un número similar de “nacimientos”. Solamente las unidades 2 y 3 contendrán las unidades nacidas desde el período 1 y sólo la muestra 3 contendrá las unidades nacidas desde el período 2. La muestra total incluida en el período 3 estará compuesta por 1.75% de unidades nacidas desde el período 2, 3.5% de nacidos entre los períodos 1 y 2 y 94.7% de nacidos antes del período 1; mientras que la población contendrá 5% de nacidos desde el período 2, 4.75% de los nacidos entre los períodos 1 y 2, y 90.25% de nacidos antes del período 1. Las estimaciones transversales podrán fácilmente resultar sesgadas si no se tomara en cuenta correctamente la composición de esta muestra.

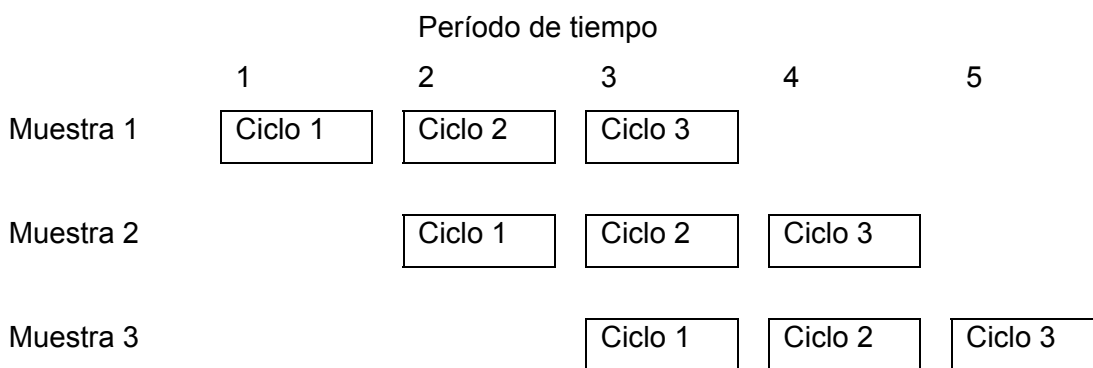


Figura 6: Diseño de Panel rotativo: 1-1-1 Patrón de Rotación

De acuerdo con las observaciones de Kalton y Citro (1993) un panel rotativo es sencillamente un caso especial de un panel repetido con superposiciones. Es especial dado que el patrón de superposición es fijo, y normalmente está equilibrado y ello es así precisamente porque cada panel que está ‘vivo’ en un momento dado está diseñado para representar a la *misma* población, permitiendo la combinación de los paneles para una estimación transversal. Los diseños de panel rotativo se usan a menudo cuando los objetivos principales son estimaciones transversales y estimaciones a corto plazo del cambio neto y bruto. En muchos países, las encuestas sobre la mano de obra tienen un diseño de panel rotativo.

Varios diseños de paneles rotativos comunes comprenden unidades que no se incluyen en la encuesta en todos los períodos, desde el momento en que entran por primera vez en el panel hasta el momento en que salen. Un ejemplo de ello se ilustra en la Figura 7. La razón por la que se adoptan dichos diseños está relacionada con el equilibrio entre, por una parte, la carga de respuesta y la representación de los nacimientos recientes, y, por otra parte, la varianza de las estimaciones. Supongamos que los períodos de tiempo de la Figura 7 son en realidad trimestres y que el objetivo principal de la encuesta es proporcionar estimaciones del cambio de un trimestre con el siguiente (“cambio trimestral”). Supongamos también que hay un componente estacional por cambiar, lo cual es algo frecuente en el caso de las estadísticas de los negocios. Entonces habrá interés por comparar el cambio trimestral en cualquier trimestre en particular con el cambio trimestral del *mismo* trimestre del año anterior. Por ejemplo, el cambio del período 7 (en relación con el período 6) será comparado con el cambio del período 3 (en relación con el período 2). La diferencia de niveles entre los períodos 3 y 7 también pueden ser de interés directo.

Podemos ver que con el patrón de rotación de la Figura 7, la mitad de las unidades muestrales del período 3 son también unidades muestrales del período 7. Y la mitad de las unidades muestrales que se pueden utilizar para estimar el cambio del período 2 al 3 también se pueden utilizar para estimar el cambio del período 6 al período 7. Este alto grado de *superposición muestral* reduce la varianza de las estimaciones del cambio. Un grado mayor de superposición reduciría aún más la varianza, pero esto también incrementaría la carga de respuesta. Situándonos en los extremos, una superposición de 100% solo podría conseguirse con el diseño de panel fijo presentado anteriormente. Pero ello significaría que las unidades muestrales tendrían que permanecer en la muestra indefinidamente.

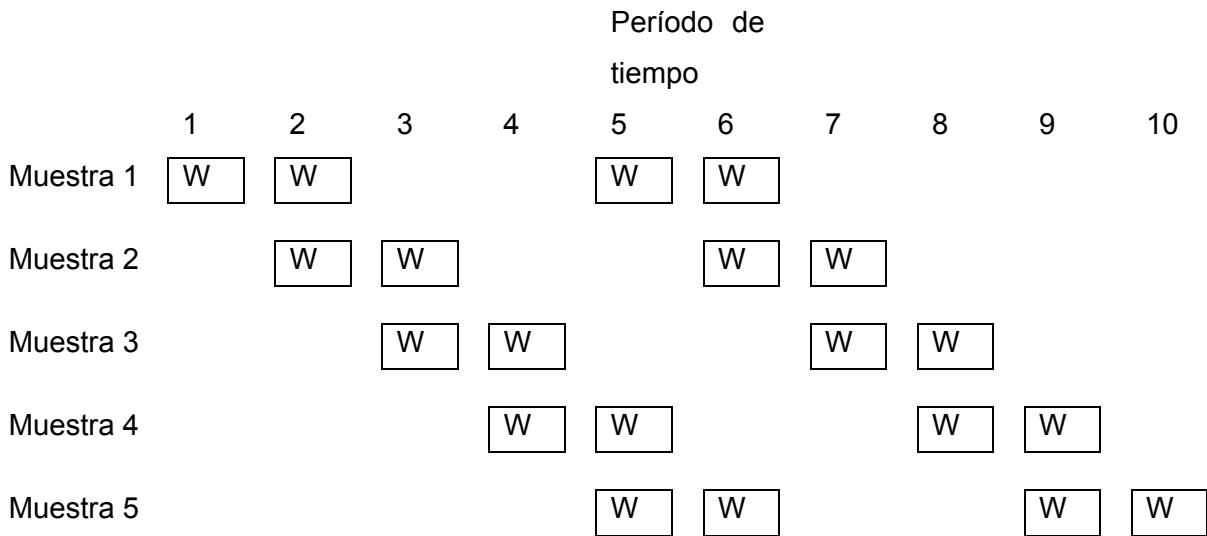


Figura 7: Diseño de Panel rotativo: 1-1-0-0-1-1 Patrón de Rotación

2.5 Paneles divididos (*Split panel*)

Este panel incluye una combinación de muestras transversales y panel en cada ocasión del trabajo de campo. Un diseño común descrito por Kish (1987, p.181-183) consiste en una muestra de panel fijo de donde se recogen datos de cada ocasión, más una muestra de transversal complementaria de cada ocasión (véase ilustración de la Figura 8). Una serie de encuestas transversales en las cuales una proporción de los elementos de muestra es deliberadamente retenida en la muestra para encuestas consecutivas, a la que Kalton y Citro (1993) denomina *encuesta de superposición* también se puede considerar como un tipo de panel dividido. Normalmente se toman en cuenta estos diseños cuando lo que se busca es realizar estimaciones, tanto de transversales como longitudinales, o también en caso de las dudas antes mencionadas sobre la capacidad de que las muestras longitudinales puedan proporcionar estimaciones transversales adecuadas.

Una variación es una serie de encuestas transversales repetidas, pero con un estrato de $P=1$ (las unidades del estrato están incluidas en la muestra de la encuesta con certeza). Con este tipo de diseño, se incluirá, en la práctica, un subconjunto de unidades muestrales en cada ocasión. En este caso, el objetivo es una serie de estimaciones transversales, pero es posible realizar estimaciones aleatorias con relación al estrato $P=1$. Muchas encuestas sobre negocios

tienen esta característica, conformando los negocios más grandes (en términos de, digamos, facturación o número de trabajadores) el estrato $P=1$.

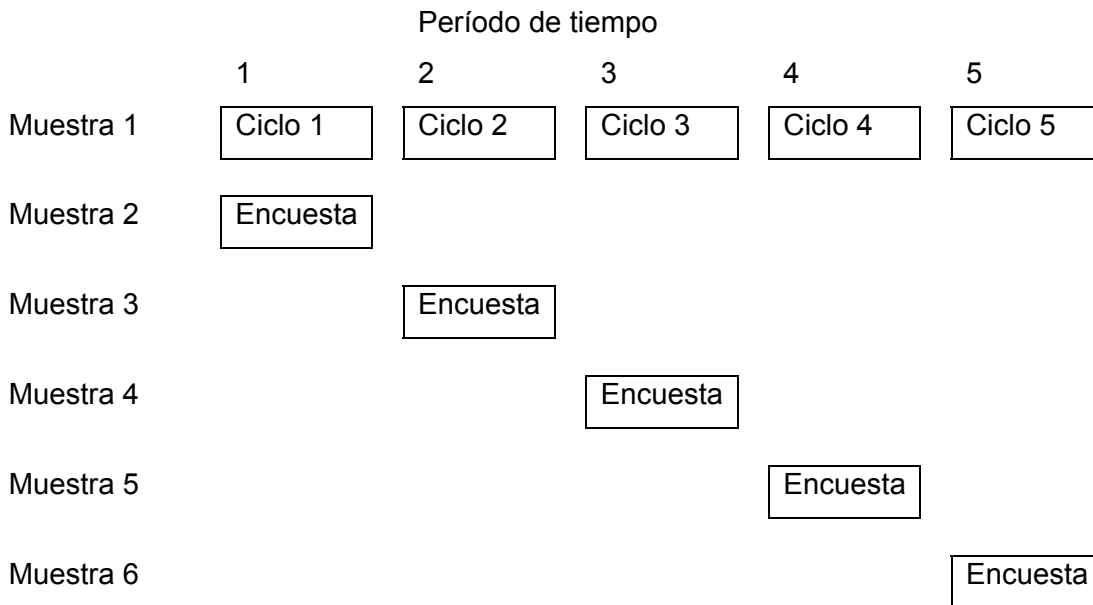


Figura 8: Diseño de Panel dividido (Split panel)

2.6 Diferencias entre los diseños

Los cinco tipos de diseños de encuestas longitudinales reseñados en los apartados 2.1 a 2.5 son, evidentemente, sólo una tipología muy diversa. Esta clasificación no describe toda la gama de diseños posibles. Por ejemplo, cada panel de un diseño de panel repetido puede o no incluir muestras regulares adicionales de nacimientos. Y en el caso de los paneles repetidos, los paneles rotativos y los paneles divididos, existe una amplia variedad de patrones posibles. La variante más apropiada de un diseño longitudinal dependerá de los objetivos de análisis y estimación, así como de la naturaleza de los datos que han de recogerse, y de las limitaciones prácticas.

Un aspecto importante es el de la posibilidad de que la ejecución de la encuesta ofrezca estimaciones transversales y longitudinales. En principio, un panel fijo más los nacimientos puede ofrecer esta posibilidad, pero en muchas situaciones, agregar muestras regulares de los nacimientos a un panel es una tarea muy compleja y costosa. Una manera más eficaz de conseguir una representatividad transversal puede ser elegir una muestra transversal nueva.

Se puede utilizar un diseño de panel dividido en este caso, o, si es necesario observar el cambio bruto de interés sólo durante períodos de tiempo relativamente cortos, un diseño de panel rotativo.

Un Segundo aspecto que se debe tomar en cuenta es la unidad principal del análisis. En buena parte del análisis longitudinal, la unidad de análisis es algún tipo de *suceso*, antes que una persona, un hogar o negocio que experimenta dicho suceso. Por, ejemplo, en el caso del análisis de las transiciones del mercado laboral, la unidad de análisis puede ser una transición entre dos estados de actividades (por ejemplo, de situación en paro o con trabajo). Por ello, la muestra debe ser representativa de todos esos sucesos durante un período específico. Esto sólo se puede lograr si la muestra es representativa de todas las *personas* que han vivido dichos sucesos, dado que los datos los proporcionan las personas. Con esto se subraya la importancia de incluir en el marco de la muestra a todas las personas que puedan haber experimentado dichos sucesos, incluso si la posibilidad de haber vivido dicho suceso hubiera sido de sólo un breve período de tiempo. Dicho de otro modo, es preferible incluir siempre en la muestra todos los nacimientos en la población de personas “en riesgo”. La conclusión será diferente si el objetivo principal del análisis fuera, digamos, estimar las características de las personas que experimentan patrones de sucesos o períodos en algunos estados durante un lapso de tiempo en particular. En este caso, la prioridad estaría en hacer las observaciones sobre una muestra de personas que están “en riesgo” durante todo el período.

2.7 Representación de los “nacimientos”

Un objetivo importante de muchos diseños para encuestas longitudinales es el de asegurar que los “nacimientos” de la población estén correctamente representados en la muestra. En este capítulo, subrayamos algunos aspectos importantes sobre el muestreo de los nacimientos.

La mejor manera de asegurar que todos los nacimientos de una población tengan la posibilidad de ser elegidos para la encuesta es tomar muestras del *flujo* de los nacimientos. Ello significaría, ya sea, tomar muestras constantemente en el tiempo u obtener muestras de un registro permanente de nuevas altas. Por ejemplo, en el caso de los inmigrantes (legales) de un país, tomaríamos muestras constantemente de las personas que entran en el país en los

aeropuertos, los puertos marítimos y otras fronteras, fijándonos en ellos para identificar a los inmigrantes, o podríamos tomar muestras de los registros oficiales de inmigrantes. Pero con frecuencia es imposible tomar muestras con asiduidad y no existe tal registro permanente. En ese caso, el procedimiento disponible más adecuado sería seleccionar una muestra transversal nueva a intervalos regulares, revisándola para identificar nuevas altas desde la última muestra. Con este enfoque, es importante que cada muestra proporcione una buena cobertura de las altas recientes. También es importante ser consciente de la posibilidad de que las nuevas altas han podido salir de la población de nuevo para el momento en que se ha elegido la siguiente muestra. En los dos apartados siguientes analizaremos estos problemas.

Cobertura de las nuevas altas

Es frecuente que un marco de muestreo, que sería una cobertura aceptable de la población total (stock), resulte inadecuado como marco para una nueva población entrante (flujo). Ello puede deberse a que el mecanismo por el cual se agregan las nuevas altas al marco tienen un desfase, así que no está al actualizado. O podría deberse a que simplemente las nuevas altas pueden tener el tipo de características que se asocian con la ausencia del marco. Así, simplemente elegir una nueva muestra del mismo marco que fue utilizado para elegir el ciclo 1 original puede no ser adecuado. En vez de ello, podría ser necesario ampliarlo mediante procedimientos especiales. Si la cobertura de las nuevas altas es deficiente, se podrían producir errores de cobertura (véase el apartado 1.3).

Cobertura de las altas de corto plazo

Algunas unidades pueden salir de la población de estudio en un período relativamente corto después de haber entrado. Los ejemplos incluyen nuevos negocios que cierran o se venden a una empresa más grande o personas que, en cuanto cumplen los 18 años, emigran (en el caso de una encuesta de residentes de 18 años o más). Cuando se toma una muestra de los nacimientos recientes identificándolos en una muestra transversal, sólo aquellos nacimientos que han permanecido en la población hasta el momento en que se elige la muestra tendrán posibilidades de ser elegidos. Otros serán omitidos, insistimos, pudiendo ocasionar un error de cobertura. La medida de esta subcobertura dependerá, por supuesto, de la prevalencia de estas “altas a corto plazo” y del intervalo entre las ocasiones en que se elabora una nueva muestra. Ello se ilustra en la Figura 9.

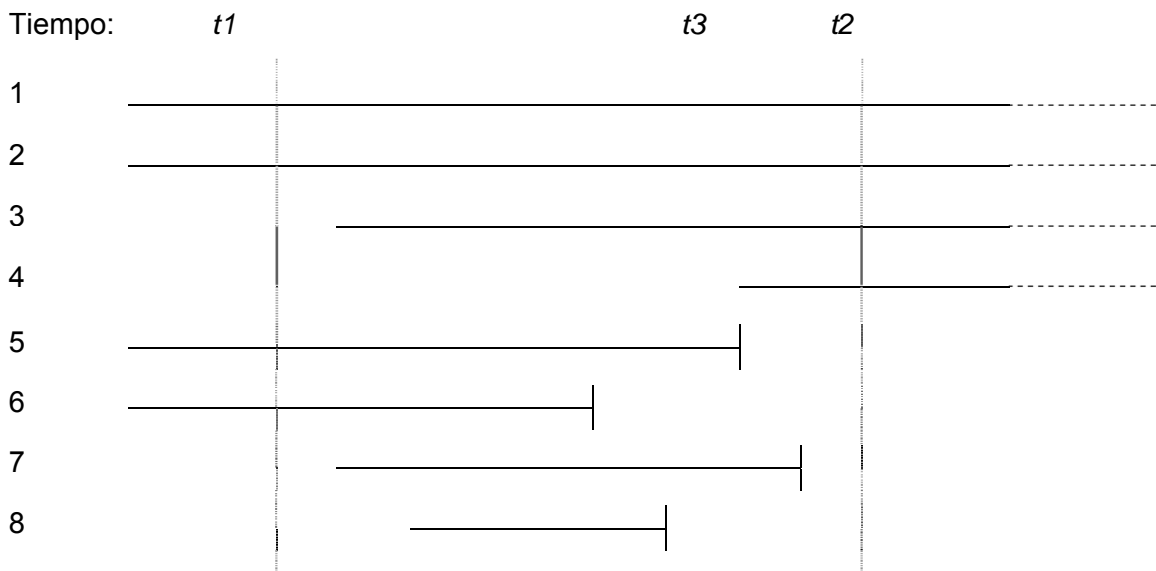


Figura 9: Muestreo de los nuevos nacimientos

Cada línea horizontal de la Figura 9 representa una unidad de la población. La longitud de la línea indica el período de tiempo durante el cual permanece la unidad en la población. Las líneas punteadas verticales t_1 y t_2 corresponden a los puntos en el tiempo en los cuales la muestra inicial (t_1) y la primera muestra extra de nuevos nacimientos (t_2) son elegidos. Una línea de rayas indica a los miembros de la población que avanzan en el futuro. Las unidades 1, 2, 5 y 6 están en el marco de la muestra t_1 . Las unidades 1, 2, 3 y 4 están en el marco de muestra t_2 , aunque, si se eligen las unidades 1 y 2, éstas aparecerían como no elegibles. Las unidades 3 y 4 han entrado en la población desde t_1 y son por ello, elegibles para poder ser incluidas en la muestra de los nacimientos recientes que se agreguen a t_2 . Sin embargo, las unidades 7 y 8 no están incluidas en el marco de muestreo ni en t_1 ni en t_2 , dado que entraron en la población después de t_1 y salieron antes de t_2 . Si el objeto del análisis es estudiar las características de algunas transiciones que tuvieron lugar durante el período, entonces, si las unidades 7 y 8 experimentaron alguna de estas transiciones, la subcobertura podría ocasionar errores. Una manera sencilla de minimizar el probable alcance de dicha subcobertura es incrementar la frecuencia con la cual se agregan las nuevas muestras a la encuesta. Por ejemplo, si la primera muestra extra de los nuevos nacimientos fuera elegida no en el tiempo t_2 , sino más bien en el tiempo t_3 , entonces, la unidad 7 estaría en el marco de la muestra y la unidad 8 sería la única unidad con cero probabilidades de selección.

3. FALTA DE RESPUESTA Y DESGASTE DE LAS ENCUESTAS LONGITUDINALES

La falta de respuesta es no poder obtener medidas completas y útiles sobre todas las unidades de medida de una encuesta (Groves et al, 2002). *La falta de respuesta de la unidad* quiere decir un fracaso total del intento de obtener cualquier medición de unidades de muestras en particular. Cuando hablamos de la *falta de respuesta de un elemento* nos referimos a los casos en que sí se han obtenido mediciones en relación con algún elemento, pero no todos los elementos de las unidades de muestra. En esta sección trataremos la falta de respuestas de una unidad. La falta de respuestas de un elemento se trata en la sección 5, pero Särndal y Lundström (2005) tratan este tema con más profundidad.

En el caso de una encuesta longitudinal, se puede producir una falta de respuesta en cada ciclo de la recogida de datos. Sus efectos tenderán a ser acumulativos, en el sentido de que el número de unidades de muestra que no proporcionan respuestas en cada ciclo normalmente excede el número (si lo hay) de los que vuelven a participar en la encuesta tras no haber proporcionado una respuesta anteriormente. Por ello, la falta de respuesta tiene el efecto de erosionar continuamente la muestra, haciendo que la muestra de respuesta sea cada vez más reducida con el transcurso del tiempo. Por esta razón, este proceso, que sólo se produce en el caso de las encuestas longitudinales, se denomina siempre como *desgaste de muestra*. En esta sección, indicaremos cuáles son las causas de la falta de respuesta y de desgaste de las encuestas longitudinales, así como los efectos que pueden tener sobre las estimaciones de las encuestas, y qué se puede hacer para reducir el alcance y el efecto de la falta de respuesta y el desgaste. El apartado 3.1 resume las razones por las que surgen las faltas de respuesta response en general. Los apartados 3.2 y 3.3 tratan los temas que son específicos de las encuestas longitudinales.

3.1 Componentes de la falta de respuesta

La falta de respuesta de una unidad puede ocurrir por varias razones. Pueden clasificarse de la siguiente manera, correspondiendo con los pasos del proceso de intento de conseguir una respuesta de una unidad muestral:

- Un fallo del encargado de recoger los datos en la localización / identificación de la unidad muestral;
- Un fallo producido al ponerse en contacto con la unidad muestral;
- La unidad muestral se niega a participar;
- La imposibilidad de la unidad muestral de participar (por ejemplo, enfermedad, ausencia, etc.);
- La imposibilidad entre el encargado de la recogida de datos y de la unidad muestral de comunicarse (por ejemplo, barreras lingüísticas);
- Una pérdida accidental de los datos / el cuestionario.

A menudo, se las clasifica en tres extensas categorías con el fin de que sirvan para debatir y presentar los resultados de las encuestas, a saber, “falta de contacto” (las primeras dos categorías), “negativa”, (la tercera) y “otras razones para la falta de respuesta” (la cuarta, quinta y sexta). Las razones por las que puede surgir cualquiera de estos tipos de resultados pueden ser muchas y variadas, dependiendo de las características particulares de una encuesta. Aquí resumimos brevemente los principales factores determinantes del nivel de cada tipo de falta de respuesta. Para un análisis más profundo, véase Lynn (2006).

Razones para una falta de contacto

El no lograr localizar o identificar una unidad muestral tiene normalmente que ver con la inexactitud de la información del marco de muestreo, como son el nombre, la dirección o los detalles del número de teléfono, que pueden no estar actualizados o incluso equivocados. La imposibilidad de ponerse en contacto con un miembro identificado de la muestra es un resultado que se aplica principalmente a las encuestas mediante entrevistas (ya sea en persona o por teléfono). En el caso de las encuestas para rellenar en casa, el no recibir un cuestionario se debe generalmente a un fallo en la localización de la unidad muestral, es decir, el cuestionario en papel ha sido enviado a una dirección equivocada, o una solicitud de correo electrónico para rellenar un cuestionario web ha sido enviada a una cuenta de correo equivocada.

En el caso de las encuestas con entrevistas, la imposibilidad de ponerse en contacto con una persona es el resultado de la interacción entre

- a) las fechas y el número de intentos que hace el entrevistador para ponerse en contacto, y
- b) las horas en que se puede localizar al miembro de la muestra

En el caso de las encuestas sobre la población general de los hogares, muchos estudios de investigación han examinado los patrones de visita de los entrevistadores y cómo se relacionan con los resultados de las encuestas o con los patrones domésticos de los hogares de la muestra. Estos estudios corresponden a las encuestas en persona (Swires-Hennessy y Drake, 1992; Campanelli *et al*, 1997; Phillipens y Billiet, 2004) y a las encuestas telefónicas (Kulka y Weeks, 1988; Bennett y Steel, 2000). Podemos resumir algunas pautas generales de la siguiente manera:

Encuestas para los hogares en persona:

- Los entrevistadores deberán hacer un mínimo de 7 visitas a cada dirección antes de aceptar un resultado de “falta de contacto” si la proporción general de la falta de contacto llega a bajar hasta el 4% (resultados correspondientes al Reino Unido, los Estados Unidos y los Países Bajos);
- A medida que el número de visitas aumenta, la probabilidad condicional de encontrar a alguien en casa en la siguiente visita se reduce;
- Las noches de los sábados y los domingos son los momentos con más probabilidades de encontrar a alguien en casa, seguidos de otras noches durante la semana;
- La probabilidad de localizar a alguien es mayor los fines de semana que durante el día en los días de semana;
- Localizar a alguien en un momento que no sea el adecuado para la encuesta sigue siendo valioso, ya que permite acordar una cita (por ejemplo, ponerse en contacto los domingos por la noche, en el caso del Reino Unido);
- Si se hacen las visitas cuando es menos probable que los miembros de la muestra estén en casa, los entrevistadores tendrán que hacer más visitas para lograr una tasa de respuesta en particular.

En el caso de las encuestas telefónicas:

- Se necesita un sistema para la planificación de las visitas para asegurar que, al final del trabajo de campo, se haya intentado localizar todos los casos de la muestra una suficiente cantidad de veces, que cubra distintos momentos del día y distintos días de la semana (y distintas semanas);
- Si un número está ocupado, el mejor momento para llamar de nuevo es de 10 a 30 minutos después. Si esto no es posible, otro buen momento es el día siguiente a la misma hora.
- Si nadie contesta (un número de una empresa), el mejor momento para intentarlo otra vez es el día siguiente;
- Si nadie contesta (número privado, de día) el mejor momento para intentarlo de nuevo es de 2 a 6 horas más tarde;
- Si nadie contesta (número privado, de noche) el mejor momento para intentarlo de nuevo es la noche siguiente.

Razones para negarse a cooperar con una encuesta

En la década pasada se ha realizado un gran volumen de investigación sobre las razones por las cuales las personas cooperan o no cooperan con las encuestas muestrales. Con ello se han conseguido avances considerables en la comprensión del proceso de participación en una encuesta. Es importante entender por qué se producen las negativas a realizar una encuesta, ya que a menudo constituyen una gran proporción de la falta de respuesta de una encuesta.

La decisión sobre si cooperar o no es el resultado de una interacción entre el entrevistador y el miembro de la muestra. El comportamiento del miembro de la muestra y del entrevistador durante la interacción se verá muy influenciada por dos conjuntos de factores. Se les puede denominar en sentido general como “el entorno social” y el diseño de la encuesta (por supuesto, tanto el entrevistador como el miembro de la muestra tendrán sus propias características y predisposiciones y dependerá de ello la forma en que actúen estos dos conjuntos de factores). El entorno social incluye el grado de cohesión social, la legitimidad de las instituciones, y demás. Ello tiene influencia sobre el grado de responsabilidad social que siente una persona de la muestra y las estrategias de persuasión y las estrategias de toma de

decisión que emplean los entrevistadores y los entrevistados respectivamente. Igualmente, es probable que el entorno inmediato en el cual tiene lugar la entrevista para la encuesta afecte la voluntad de participar en ella por parte del miembro de la muestra. Los factores relevantes incluyen la comodidad y la aparente seguridad.

Muchos aspectos del diseño de una encuesta afectan las tasas de respuesta. Los que tienen una importancia particular en relación con las encuestas longitudinales se tratan en el apartado 3.3. más adelante. Otros aspectos, amplios, del diseño para encuestas pueden considerarse como limitaciones en relación con la interacción entre el miembro de la muestra y el entrevistador. El modo en que se realiza la entrevista es muy importante. Los entrevistadores se sienten mucho más limitados en cuanto a la forma de comunicarse con un miembro de la muestra si ésta se realiza por teléfono o si están de pie uno frente al otro. No pueden mostrarle al miembro de la muestra ningún documento ni su carné de identidad, no pueden utilizar lenguaje corporal ni gestos, ni nada parecido. Estas limitaciones pueden contribuir con los bajos niveles de éxito que parece que tienen los entrevistadores en su intento por evitar las negativas de las encuestas telefónicas. El modo en que los entrevistadores presentan la encuesta también puede tener que ver con la longitud y el contenido de la entrevista. Por ejemplo, si el miembro de la muestra parece, en general, dispuesto pero parece que no tiene mucho tiempo disponible en esos momentos, entonces, ante la perspectiva de una larga encuesta, un entrevistador puede sugerir volver en un momento más conveniente (“retroceso y regreso”), antes que pedirle que empiecen con la entrevista inmediatamente. Pero si la entrevista es corta, al entrevistador le puede ser mejor sugerir que empiece la entrevista inmediatamente. Estas tácticas pueden tener diferentes consecuencias para el resultado de la encuesta.

Por último, ante una petición de tomar parte en una encuesta, un miembro de una muestra seguramente sopesará rápidamente (y en muchos casos, inconscientemente) los posibles beneficios (ventajas) e inconvenientes (desventajas) si accede a ella (Groves et al, 2000). El reto para el diseñador de la encuesta es encontrar formas de enfatizar las ventajas y atenuar las desventajas. Esto se complica con el hecho de que los distintos miembros de la muestra tendrán distintos puntos de vista sobre qué ventajas y desventajas potenciales son las que les interesan a ellos. Los principales beneficios e inconvenientes, que probablemente sean relevantes para muchos miembros de una muestra, han sido resumidos en la Tabla 1.

Beneficios	Inconvenientes
Contestar las preguntas será divertido (tema interesante)	Me tomará un tiempo que podría pasar de manera más útil haciendo otra cosa
La experiencia de la entrevista será agradable (interacción social)	Puede hacerme sentir incómodo o tenso (tema delicado)
Haré algo de provecho (estoy de acuerdo con los objetivos que me han indicado y creo que la encuesta puede alcanzar aquellos objetivos)	Puedo estar poniendo en riesgo mi seguridad (invitando a un extraño a mi casa, dándole información personal)
Recibiré algún beneficio inmediato (pago, vale de compra, regalo)	Podría haber otras consecuencias no deseadas (marketing directo, más entrevistas, obligaciones legales)

Tabla 1: Beneficios percibidos e inconvenientes de la participación en una encuesta

3.2 Características especiales de la falta de respuesta en las encuestas longitudinales

Los tipos principales de falta de respuesta presentados en el apartado 3.1 se aplican a todas las encuestas, incluyendo las encuestas longitudinales. Sin embargo, las razones por las que surgen, el impacto que puedan tener, y las técnicas que se pueden utilizar para hacerle frente pueden ser muy distintas en el caso de las encuestas longitudinales. En este apartado subrayamos aquellos aspectos que son únicos de la falta de respuesta en el contexto de las encuestas longitudinales.

Falta de contacto

Tras el primer ciclo de recogida de datos, un componente importante de la falta de contacto de las encuestas longitudinales se debe a la movilidad geográfica de los miembros de la muestra. Entre los ciclos, una parte de ellos se irá a casa, cambiará de trabajo, cambiará de número de teléfono o cambiará de dirección de correo electrónico. Si la organización de la encuesta depende de alguno o de todos estos detalles para ponerse en contacto con el miembro de la muestra, entonces necesitará tomar otras medidas para poder establecer contacto en un ciclo posterior. También deberá hacer que los entrevistadores estén disponibles para intentar ponerse en contacto con los miembros de la muestra en todos los lugares donde puedan haberse trasladado los miembros de la muestra (no hay problema en el caso de las encuestas telefónicas, postales o a través de la web, pero sí que es importante para las

encuestas cara a cara). Por ejemplo, las encuestas de los hogares del Reino Unido con recogida de datos con intervalos de un año revelan que cerca del 10% de las personas en promedio cambian de dirección entre cada par consecutivo de ciclos de encuestas (Laurie et al, 1999).

La organización puede tomar medidas como las que presentamos a continuación con el fin de hacer frente a la pérdida potencial de sumas debido a la falta de contacto.

- Recogida de información de contacto durante el primer ciclo y en cada ciclo posterior. Esta información incluye el número de teléfono (incluyendo el móvil); el nombre y el número de teléfono de amigos o familiares; detalles de contacto del lugar de trabajo; direcciones de correo electrónico. Esta información podrá ser utilizada en caso necesario para establecer contacto en los siguientes ciclos.
- Use distintos modos de contacto con los miembros de la muestra entre los ciclos. Incluya mecanismos que pueda utilizar el miembro de la muestra para informar a la organización de la encuesta sobre cualquier cambio de dirección. Por ejemplo, los mailings (como son los boletines y las tarjetas de cumpleaños) pueden incluir tarjetas postales prepagadas; los correos electrónicos pueden incluir un enlace hacia una página web de la encuesta con una sencilla forma de respuesta, etc.;
- Establezca métodos de seguimiento especiales para localizar los detalles de contacto de los miembros de la muestra que se han trasladado y cuya dirección actual se desconoce. La naturaleza de estos métodos dependerá de las circunstancias. Por ejemplo, puede existir algún tipo de dato oficial que se pueden consultar.

Negativas

Además de los rasgos universales de los rechazos que hemos descrito en el apartado 3.1, existen dos características especiales que son importantes para las encuestas longitudinales. La primera es que la participación requiere un compromiso considerable de parte de los miembros de la muestra. No se trata de una sola entrevista, sino de varias, a lo largo de un período de tiempo. A veces nos referimos a ello como una *gran responsabilidad para el entrevistado*. Por ello, puede ser necesario ofrecer incentivos o motivación especiales para compensar. Normalmente, (pero no siempre), las encuestas longitudinales ofrecen a los miembros de la muestra un pequeño pago por cada entrevista, u otra forma de regalo, además

de hacer un particular esfuerzo por hacer que el miembro de la muestra se sienta importante, insustituible, parte del estudio y persuadiéndolo de que el propio estudio es valioso.

La segunda característica especial de las encuestas longitudinales en lo que se refiere a las negativas es que, después del primer ciclo, los miembros de la muestra ya han pasado por la experiencia de una entrevista para una encuesta y, por eso, tienen una clara idea de lo que se trata, qué tipo de preguntas les harán, lo difícil que les puede resultar, etc. Esto es muy distinto de una situación normal en una entrevista, donde el miembro de la muestra solo tendrá una vaga idea y una impresión general de lo que le están pidiendo que haga en el momento en que le piden que coopere. Es por este motivo que, cuando se trata de una encuesta longitudinal, se debe tratar de que la entrevista sea una experiencia lo más agradable posible para el entrevistado. Si al entrevistado le parece que la entrevista es difícil, frustrante, desagradable, aburrida o simplemente demasiado larga, estará menos dispuesto a participar en el siguiente ciclo.

Un aspecto importante de las entrevistas para las encuestas longitudinales es si conviene o no, en el caso de que sea posible, asignar el mismo entrevistador al entrevistado cada ciclo. En términos de efecto sobre la tasa de respuesta, la evidencia es más bien ambigua. La mayor parte de los estudios que pretenden demostrar un efecto de este tipo son no experimentales (Waterton y Lievesley 1987, Rendtel 1990, Rope 1993, Taylor *et al* 1996) y, por ello, confunden la estabilidad del entrevistador con los efectos del área. Tres estudios relacionados con este tema constituyen una excepción. Todos utilizan datos de un diseño experimental que interrelacionaban entrevistadores y a áreas de la British Household Panel Survey (Encuesta Panel de Hogares Británicos). Campanelli y O'Muircheartaigh (1999) no encontraron ningún efecto de continuidad en el ciclo 2. Laurie *et al* (1999) extendió el análisis a los ciclos 3 y 4 y encontró diferencias significativas. Campanelli y O'Muircheartaigh (2002) analizaron nuevamente los mismos datos y concluyeron que las diferencias aparentes podrían deberse al desgaste no aleatorio del entrevistador. Por ello, en conclusión, no existe evidencia directa de que mantener la continuidad del entrevistador mejore las tasas de respuesta. Puede resultar de ayuda en el caso de algunos miembros y en el caso de otros, no.

Otra táctica que se puede utilizar para reducir las tasas de negativa es intentar "convertir" una negativa en un enfoque de seguimiento. Una manera de hacerlo es realizando el seguimiento por teléfono, directamente desde la oficina de la encuesta. Puede ser muy efectivo

(Burton et al, 2006), ya que la persona que hace el seguimiento puede comprobar por qué el miembro de la muestra la rechazó y puede intentar encaminar el asunto. A menudo, se dan negativas debidas a alguna circunstancia temporal (por ejemplo, un período muy ocupado, tenso o hasta traumático en la vida de un miembro de la muestra) que puede no ser el caso si se hace un intento por revertir esa negativa unas semanas más tarde. O puede ser que el miembro de la muestra sencillamente no se sintiera a gusto con el entrevistador que le visitó (por motivos de personalidad, apariencia, género, etc.), lo cual es una buena razón para que otra persona haga el intento de persuadir al entrevistado. Si el miembro de la muestra se deja convencer de que un entrevistador vaya a visitarle (si es necesario, otro entrevistador), se podrá acordar otra cita.

Una táctica general que puede usarse en las encuestas longitudinales para reducir el riesgo de negativas es “confeccionar” un enfoque, para cada miembro de la muestra, que encaje con sus circunstancias y preferencias particulares. A partir del ciclo 2, se conoce mucha información sobre cada miembro de la muestra antes de que un entrevistador se ponga en contacto con él. Esta información incluye tanto las respuestas de las encuestas y los datos del proceso de los ciclos (por ejemplo, a qué horas del día se consiguió o no encontrar a la persona; si hubo que intentar cambiar una negativa y cuáles fueron las razones para la negativa inicial). Esta información puede utilizarse para identificar un mejor enfoque el próximo ciclo. Algunos aspectos del enfoque que pueden depender de esta información previa incluyen las horas de los intentos de contacto, el contenido de cualquier carta de presentación, cualquier otra comunicación por escrito, la naturaleza de cualquier incentivo que se le ha ofrecido, los mensajes que ha subrayado el entrevistador al presentar la encuesta, e incluso el modo de la recogida de datos.

3.3 Patrones de la falta de respuesta en las encuestas longitudinales

En una encuesta longitudinal, una falta de respuesta puede surgir en cualquier ciclo de la encuesta. Cuando esto sucede en el caso de un miembro de la muestra que ya había respondido en el último ciclo, es frecuente denominar este hecho como una *falta de respuesta al ciclo*, para distinguirla de los casos en que no se obtiene ningún dato para un miembro de la muestra. Existen entonces muchos patrones de respuesta de ciclo que pueden surgir de un total de muchos ciclos de una encuesta. El número de patrones posibles depende de la política

de recogida de datos de la encuesta. A continuación, ilustramos los patrones que pueden presentarse según estas tres políticas comunes:

1. Intento de recoger datos de todas las unidades elegibles en cada ciclo;
2. Intento de recoger datos de cada ciclo de todas las unidades elegibles que responden en el ciclo 1. Esta política se adopta normalmente cuando el objetivo principal de la encuesta es entender el cambio de la situación en el momento del ciclo 1. Por ello, los datos de ciclos posteriores no son muy valiosos si faltan los datos del ciclo 1.
3. Intento de recoger datos de cada ciclo de todas las unidades elegibles que respondieron en el ciclo anterior. Se puede adoptar esta política cuando el objetivo principal de la encuesta requiere datos completos de cada ciclo.

A modo de ejemplo, supongamos que tenemos una encuesta de 4 ciclos. Entonces, según la política 1, existen 16 patrones de respuesta posibles, como se indica en la Figura 10. Los recuadros sombreados representan una respuesta y los recuadros en blanco, una falta de respuesta. Según la política 2, existen 9 posibles patrones de respuesta (Figura 11) y según la política 3, existen 5 patrones posibles (Figura 12). Estos patrones de respuesta tienen importancia en cuanto al modo en que se pueden utilizar en el análisis y a las técnicas de ajuste de la falta de respuestas, como es la ponderación, lo cual será tratado en el capítulo 5.

Ciclo:	1	2	3	4
1	■	■	■	■
2	■	■	■	□
3	■	■	□	■
4	■	□	■	■
5	□	■	■	■
6	■	■	□	□
7	■	□	■	□
8	□	■	■	□
9	■	□	□	■
10	□	■	□	■
11	□	□	■	■
12	■	□	□	□
13	□	■	□	□
14	□	□	■	□
15	□	□	□	■
16	□	□	□	□

Figura 10: Posibles patrones de respuesta por ciclo según la política 1: todas las unidades en todos los ciclos

Ciclo:	1	2	3	4
1	■	■	■	■
2	■	■	■	□
3	■	■	□	■
4	■	□	■	■
5	■	■	□	□
6	■	□	■	□
7	■	□	□	■
8	■	□	□	□
9	□	□	□	□

Figura 11: Posibles patrones de respuesta por ciclo según la política 2: entrevistados del ciclo 1 en todos los ciclos

Ciclo:	1	2	3	4
1				
2				
3				
4				
5				

Figura 12: Posibles patrones de respuesta por ciclo según la política 3: sólo entrevistados del ciclo anterior

En la Figura 13 presentamos un ejemplo de los patrones de respuesta observados a lo largo de los ciclos de una encuesta longitudinal. Este ejemplo utiliza datos de la cohorte 5 del England and Wales Youth Cohort Study (Estudio de Cohortes de Jóvenes de Inglaterra y Gales), una encuesta enviada por correo para rellenar en casa dirigida a personas de entre 16 y 19 años en Inglaterra y Gales (Lynn et al, 1994). Los miembros de la muestra recibieron cuestionarios con un intervalo de un año, con el primer ciclo en otoño del año en que habían acabado sus estudios obligatorios. La política de recogida de datos para esta encuesta era tratar de recoger los datos de cada ciclo de todos los miembros de la muestra (política 1) y la encuesta tenía tres ciclos de recogida de datos.

Ciclo:	1	2	3	No. de	% de
1				8,396	44.9
2				2,555	13.7
3				2,660	14.2
4				900	4.8
5				386	2.1
6				575	3.1
7				352	1.9
8				2,855	15.3

Figura 13: Patrones de respuesta por ciclos del Estudio de Cohortes 5 para Jóvenes de Inglaterra y Gales

La Figura 13 ilustra una serie de características comunes de patrones de respuesta de ciclos de las encuestas longitudinales. En primer lugar, el patrón que más prevalece es la

respuesta completa. En este caso, el 45% de los miembros de la muestra han respondido en los tres ciclos. Sin embargo, aunque este es el patrón que prevalece, no constituye una mayoría en la muestra. En otros contextos, una tasa de respuesta de 45% podría considerarse decepcionante, hasta preocupante. El siguiente patrón de respuesta es la falta de respuesta completa: 15% de los miembros no responden en ningún ciclo. Ello sugiere que hay algún tipo de consistencia en la conducta de respuesta de los miembros de la muestra: una mayoría de los miembros de la muestra responde en *todos* los ciclos o no responde en *ningún ciclo*.

Los siguientes patrones más predominantes son los patrones de desgaste: 14% de los miembros de la muestra responden en cada uno de los dos primeros ciclos pero no en el tercer ciclo y hasta un 14% responde sólo en el primer ciclo. Es bastante común que los miembros de la muestra respondan hasta cierto punto de la vida de una encuesta longitudinal y después dejen de responder.

Los patrones de respuesta en los que los miembros de la muestra responden en un ciclo habiendo dejado de responder en al menos una ocasión (patrones 4, 5, 6 y 7 de la Figura 13) son más bien poco comunes. Como ya se ha sugerido, el uso que se puede hacer de los datos de dichos entrevistados es normalmente limitado. Estos entrevistados no pueden contribuir a las estimaciones del cambio desde la edad de 16 años (ciclo 1). En el caso de los patrones 6 y 7, no pueden contribuir con ningún tipo de estimaciones de cambio, ya que sólo se les ha observado en una ocasión. Y es relativamente caro recoger datos de estos patrones de respuesta. Recuérdese que, en cada ciclo, se envían tres recordatorios a cada miembro de la muestra que todavía no ha respondido. Por definición, esto significa que los tres recordatorios se mandan a cada persona que, por finalmente, no responde. Considerando un promedio de 1.5 recordatorios por entrevistado, a lo largo de los tres ciclos, un total de 169,463 cuestionarios fueron enviados para obtener 36,457 cuestionarios respondidos, es decir, un promedio de 4.65 recordatorios por cuestionario respondido. Si, en vez de ello, se hubiera adoptado la política de enviar recordatorios sólo a los entrevistados del ciclo 1, el total de los cuestionarios enviados habría caído hasta 138,667, y el total de los cuestionarios respondidos habría sido de 34,758, un promedio de 3.99 recordatorios por cuestionario. Y si la encuesta hubiera adoptado la política de escribirle solo a los entrevistados en el ciclo anterior, se habrían recibido 33,858 cuestionarios respondidos a un coste de sólo 3.22 recordatorios cada uno. En cuanto a esta política, entonces, escribir a todos los miembros de la muestra en cada ciclo

puede considerarse como una manera cara de obtener una cantidad relativamente pequeña de datos adicionales que, en todo caso, tienen una utilidad limitada.

Tal vez también vale la pena tomar en cuenta las razones de los patrones de respuesta de un ciclo encontradas en esta encuesta. En el Reino Unido, una considerable minoría de jóvenes abandona el hogar poco después de acabar sus estudios obligatorios. Es posible que en el caso de muchos componentes de este subgrupo, la dirección que se obtuvo del marco de la muestra (dirección de casa durante el último año de colegio) ya estuviera desfasada para el momento en que se hizo el ciclo 1 de la encuesta. Por ello, el cuestionario puede no haber llegado a estos miembros de la muestra. En términos más generales, los jóvenes son más flexibles entre los 16 y 19 años, mudándose con frecuencia a distintos lugares para trabajar o estudiar. En particular, la mayoría de los que fueron a la universidad se habrán trasladado a algún lugar distinto (muy pocos estudiantes universitarios siguen viviendo con sus padres mientras están en la universidad en el Reino Unido) entre los ciclos 2 y 3. En el ciclo 1 (y en el ciclo 2), el cuestionario incluía un espacio para registrar los detalles de contacto (nombre, dirección y número de teléfono) de los hogares y otras personas que pudieran estar al corriente sobre dónde estarían en el futuro en caso de trasladarse. Se recurría a ellos en el siguiente ciclo en los casos en que les fuera devuelto el correo o cuando no recibían respuesta después del primer recordatorio.

Por consiguiente, es probable que haya habido miembros de la muestra que no recibieran el cuestionario del ciclo 1 ("falta de contacto"). Si esto sucedió, es probable que tampoco hayan recibido el cuestionario en los siguientes ciclos. Por otra parte, entre los miembros de la muestra que respondieron en el ciclo 1, había una alta probabilidad de contacto en ciclos posteriores, dados los detalles extra recogidos en el cuestionario del ciclo 1 (aunque no todos los entrevistados proporcionaron estos detalles), así que es bastante probable que buena parte de la falta de respuesta de los ciclos 2 y 3 de los entrevistados en el ciclo 1 (patrones de respuesta 2, 3 y 4) se debiera a negativas antes que a una falta de contacto. El cuestionario era bastante similar cada año, así que puede haber parecido repetitivo e innecesario para algunos miembros de la muestra. Es difícil persuadir a los miembros de la muestra de lo contrario sin la intervención de un entrevistador.

4. ERRORES DE MEDICIÓN EN LAS ENCUESTAS LONGITUDINALES

4.1 Introducción a los errores de medición

Un *error de medición* ocurre cuando el valor de un elemento disponible de análisis (es decir, del conjunto de los datos) correspondiente a un entrevistado no corresponde con el valor ideal que habría sido registrado si el concepto subyacente de interés hubiera estado perfectamente calculado. Nos referiremos a estos dos valores como valores de *observación* y de *verdadero valor* respectivamente (a la vez que reconocemos que es frecuente que nos sea difícil determinar con exactitud cuál es el verdadero valor). Existen muchas razones por las cuales la observación puede diferir del verdadero valor:

- El entrevistado puede no comprender la pregunta tal y como se desea (especialmente si la pregunta no estaba bien diseñada);
- El entrevistado puede no recordar de memoria (toda) la información necesaria para contestar la pregunta;
- El entrevistado puede cometer errores de juicio o estimación al convertir la información que recuerda en respuesta;
- El entrevistado puede dar una respuesta erróneamente, deliberadamente o por accidente;
- En una encuesta con entrevistas, el entrevistador puede registrar la respuesta erróneamente;
- En una encuesta con lápiz y papel (para que el propio entrevistado la rellene o mediante entrevista), la respuesta puede resultar ilegible;
- Si una respuesta requiere una codificación posterior a la encuesta, puede haberse aplicado un código equivocado;
- Si los datos requieren registrar los datos después de la entrevista, puede ocurrir un error de teclado;
- Otros errores durante la edición y el procesado de los datos pueden afectar los datos.

Lo que nos interesa en este caso no es analizar por qué ocurren estos distintos tipos de error y qué se puede hacer para reducirlos, en general. Para eso se pueden consultar otros textos (un buen comienzo sería Groves et al, 2004, Biemer et al, 1991 y Biemer y Lyberg, 2004). En vez de ello, nuestro interés se centra en cómo influyen estos errores en las estimaciones de las encuestas, especialmente dentro del contexto de una encuesta longitudinal

y en las técnicas específicas que se pueden utilizar en las encuestas longitudinales para minimizar los impactos negativos de los errores de medición.

El efecto neto de todos los errores individuales de los datos es que inducen a un error de medición en las estimaciones de la encuesta (véase el apartado 1.3). Pensemos en un simple ejemplo de estimación de la media de una variable continua, como es la renta. Si los errores de medición son aleatorios, es decir, que tienen una media de cero, entonces la media de las observaciones será equivalente a la media de los verdaderos valores. Dicho de otro modo, el error de medición no ha introducido ningún sesgo en la estimación de la encuesta, pero sí habrá introducido una varianza extra (véase la Figura 1). Por otra parte, si los errores de medición (o alguno de ellos) son sistemáticos, entonces, la media de la observación sí *diferirá* de la media de los verdaderos valores, así que se habrá introducido un sesgo. Esto puede suceder si, por ejemplo, unas personas que tienen unos ingresos muy bajos tienden a informar que ganan más de lo real, tal vez por razones de aspiraciones sociales, sin un informe de otras personas que tiendan a proporcionar una cifra inferior a la real, que los equilibre. En general, sí hay una diferencia si los errores de medición son *aleatorios* (media de cero) o *sistemáticos* (media por encima de cero).

Ahora supongamos que queremos estimar la asociación entre la renta y la ocupación. Incluso si los errores de medición en la renta observados son aleatorios, harán que la *asociación* parezca débil. De manera que los errores de medición en un objeto individual pueden causar un error sistemático (sesgo) en una estimación de una encuesta. Lo mismo se puede observar si nos fijamos en una estimación de la proporción de unidades en una categoría en particular de una variable categórica. Veamos el ejemplo de la Tabla 2, donde se ha dividido una muestra de 1000 personas en tres categorías de la variable “situación de la actividad económica”. Supongamos que la categoría ha sido observada correctamente el 90% de las veces, pero que otro 10% de las observaciones se han clasificado mal, pudiendo haber sido clasificadas en cualquiera de las dos categorías equivocadas. Nuevamente, el error de medición aparentemente aleatorio ha distorsionado las proporciones de cada categoría. Observamos que un 52.6% tiene trabajo, cuando la verdadera proporción dentro de la muestra es de 56.0%, etc. Una estimación basada en la encuesta sobre la proporción que trabaja presentaría un sesgo a la baja.

		Categorías observadas			Total
		Parados	Empleados	Otros	
Categorías verdaderas	Parados	144	8	8	160
	Empleados	28	504	28	560
	Otros	14	14	252	280
Total		186	526	288	1000

Tabla 2: Error de medición de una variable categórica

Ahora trasladaremos el análisis del impacto del error de medición sobre las estimaciones de una encuesta al tipo de estimaciones que nos interesan más en el caso de las encuestas longitudinales.

4.2 Errores de medición en las estimaciones de los cambios

Supongamos que queremos estimar la proporción de personas que cambian de estatus de actividad entre un ciclo de una encuesta y el siguiente, y que las observaciones de la Tabla 2 representan el primer ciclo. Supongamos que la Tabla 3 representa la relación entre los valores verdaderos de cada ciclo. Dicho de otro modo, esta la *verdadera* matriz de transición del estatus de actividad de un ciclo a otro.

Ahora supongamos además que las observaciones del ciclo 2 están sujetas exactamente al mismo tipo de error de medición que en las observaciones del ciclo 1. Es decir, 90% de los valores se observan correctamente y el otro 10% es clasificado aleatoria y erróneamente en una de las dos categorías. Entonces, la matriz de transición *observada* será como se indica en la Tabla 4.

		Ciclo 2			Total
		Parados	Empleados	Otros	
Ciclo 1	Parados	60	50	50	160
	Empleados	60	400	100	560
	Otros	40	40	200	280
Total		160	490	350	1000

Tabla 3: Matriz de transición verdadera

		Ciclo 2			
		Parados	Empleados	Otros	Total
Ciclo 1	Parados	59	67	60	186
	Empleados	77	336	113	526
	Otros	50	64	174	288
Total		186	467	347	1000

Tabla 4: Matriz de transición observada, con medidas sujetas a error de medición

Se puede observar que la matriz de transición observada es realmente muy diferente de la verdadera, pese al relativamente modesto nivel de error de medición aparentemente aleatorio. Por ejemplo, la proporción verdadera que no cambió de estatus de actividad entre los dos ciclos es de 66.0%, mientras que la proporción observada es de 56.9%. La verdadera proporción de personas desempleadas en el ciclo 1 que estaban trabajando en el ciclo 2 es de 31.2%, mientras que la proporción observada es de 36.0%. Y la verdadera proporción de personas clasificadas como “Otros” en el ciclo 1 que trabajaban en el ciclo 2 es de 14.3%, mientras que la proporción observada es de 22.2%. Un error de medición ha sesgado las estimaciones de cualquiera de estos parámetros de cambio.

Con frecuencia, nos interesa, no (solamente) la estimación de los niveles de cambio, sino también la estimación de la asociación de otras variables con el cambio (o la estabilidad). Por ejemplo, podríamos querer saber cuáles son las *características* de las personas que cambian de estar en el paro a tener un trabajo, y tal vez también compararlos con las características de las personas que permanecen en el paro. En nuestro ejemplo anterior, existen 50 miembros de la muestra que realmente cambian del paro al trabajo, pero observamos 67. De hecho, solo 41 de ellos han pasado del paro al trabajo ($50 * 0.9 * 0.9$), mientras que los otros 26 han sido clasificados erróneamente en la matriz de transición. De manera que basaríamos nuestras estimaciones de las características de las personas que han pasado por esta transición en 67 personas, de las cuales sólo 41 deberían haber sido incluidas en el análisis, y también estaremos omitiendo nueve personas que deberían haber sido incluidas. Es poco probable que el efecto neto de una clasificación errónea sería el de no tener efectos sobre nuestras estimaciones. Por ello, es muy importante tratar de clasificar las unidades muestrales correctamente tomando en cuenta si han experimentado o no cambios o han realizado transiciones.

4.3 El efecto junta (Seam effects)

A menudo, las encuestas longitudinales intentan recoger historias continuas, por ejemplo, ingresos, estatus de las actividades económicas, matrimonio y fertilidad, etc. Ello se hace recogiendo datos en cada ciclo sobre el período de tiempo desde el ciclo anterior y reuniendo después todas estas historias cortas y creando una larga historia. Dichos datos a menudo adolecen de inconsistencias en las “juntas” entre dos entrevistas. Por ejemplo, un entrevistado puede informar en un ciclo que no tiene trabajo en ese momento, pero en el siguiente ciclo, seis meses después, informa que lleva seis meses trabajando. Ello puede llevar al analista a pensar que debe haber empezado a trabajar inmediatamente después de la entrevista anterior. O bien, la descripción del entrevistado de su trabajo en cada ciclo es muy diferente, llevando tal vez al analista a deducir que debe haber habido un cambio de trabajo justo después de la entrevista anterior. Como consecuencia de ello, el número de cambios de estatus, es normalmente “mucho mayor entre los meses en los que se recogen los datos de ciclos diferentes que entre los meses en los cuales se recogen datos del mismo ciclo” (Kalton, Miller y Lepkowski 1992, p. 13). La concentración de transiciones entre los períodos de referencia, conocidos como el “efecto junta”, afecta a todas las encuestas panel y pueden ser fundamentales (véase, por ejemplo, Burkhead y Coder 1985; Hill 1987; Kalton y Miller 1991; Lemaitre 1992; Moore y Kasprzyk 1984).

Los efectos junta, entonces, surgen de la combinación de datos de múltiples ciclos cuando los datos están sujetos a un error de medición. Los errores no son específicos de las encuestas longitudinales (aunque sí lo sea su visibilidad como efecto junta). Veamos una encuesta que tenga un período de referencia de m meses, donde, en la primera entrevista, en el mes m , le preguntan al entrevistado sobre su situación en los meses 1, 2..., m . En la segunda entrevista, en el mes $2m$, le preguntan sobre su situación en los meses $m+1$, $m+2$, ..., $2m$. Cuando se reúne la información de ambas entrevistas para crear una historia continua, es normal atribuir un número desproporcionado de cambios de situación a los meses m and $m+1$, la ‘junta’ entre los períodos de referencia.

El analista puede *deducir* que ha habido un cambio durante la junta, si el informe de situación de la primera entrevista (en el mes m) correspondiente al mes m no encaja con el

informe retrospectivo de la segunda entrevista (en el mes $2m$) correspondiente al mes $m+1$. Estos desencajes pueden ocurrir por varias razones (véase Martini 1989; Young 1989):

1. *Omisión o error de informe.* Supongamos que en cada entrevista se pregunta sobre el cobro de una fuente de ingresos en particular y que el entrevistado ha recibido esta fuente de ingresos continuamente a lo largo del período de referencia. Puede ser que informe correctamente que ha recibido dicha fuente de ingresos en, digamos, todas las entrevistas menos una, con el resultado de una desviación aparente de la fuente en el mes $m+1$, seguido de un regreso a dicha fuente en el mes $2m+1$;
2. Clasificación errónea o re-definición de la información pasada. Por ejemplo, un entrevistado puede informar retrospectivamente que su actividad laboral del $m+1$ era 'cuidar de su hogar', aunque en el momento m afirmaba que no tenía trabajo;
3. Ubicación errónea de los sucesos en el tiempo. Si un entrevistado experimentó un cambio desde la entrevista anterior pero recordó erróneamente que había tenido lugar hace más tiempo, puede ser que indique correctamente su estatus actual en $2m$, pero que también indique la misma situación para los meses anteriores al período que está recordando hasta $m+1$, interpretándose que había habido un cambio durante la junta (y no más recientemente);
4. Codificación de errores, especialmente para los elementos codificados en marcos de codificación complejos, como son la labor y la ocupación. Los cambios dudosos pueden ocurrir porque los entrevistados usan palabras diferentes para describir la misma ocupación o labor, o por una descripción ambigua o por errores de codificación.

Un cambio en una junta puede ser *imputado* si falta la fecha del cambio de estatus. A menudo se imputan fechas que podían estar entre los sucesos anteriores o posteriores. Si la fecha atribuida cae en el período de referencia anterior, la fecha se ubica, por defecto, al principio del período de referencia (es decir, en la junta), en lugar de invalidar la información de la entrevista anterior. El razonamiento es que la información de la primera entrevista está más cercana en el tiempo a los propios sucesos y por ello tiene más probabilidades de ser más verosímil (Halpin 1998).

4.4 Métodos para la reducción los errores de medición en las encuestas longitudinales

Entrevistas dependientes

El término *entrevista dependiente* hace referencia a entrevistas estructuradas, en donde la *elección de las preguntas y/o el contenido de las preguntas* varía entre los miembros de la muestra, dependiendo de las respuestas dadas por el miembro de la muestra en una entrevista previa. Por ejemplo, una encuesta longitudinal puede intentar actualizar información recogida en un ciclo 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).

En algunos aspectos, esta técnica es similar a la que se usa normalmente en las encuestas transversales. Primero, el “dirigir” o “saltarse” (Oppenheim, 1992, ch.6), la *elección de la pregunta* depende de la(s) respuesta(s) dadas a una o más preguntas anteriores. En segundo lugar, la *formulación exacta de una pregunta* puede estar adaptada dependiendo de las respuestas a las preguntas anteriores (por ejemplo, “...su trabajo actual...” en el caso de un entrevistado que ha respondido que actualmente está trabajando, y “...su trabajo más reciente...” en el caso de un entrevistado que ha respondido que actualmente está en el paro pero que sí ha tenido un trabajo antes). La diferencia es simplemente que la información utilizada para determinar qué pregunta se debe formular, o el contenido de la pregunta, viene de la propia entrevista, mientras que, en el caso de la entrevista dependiente, la información ya se conoce antes de comenzar la entrevista. Ello es un reto extra cuando se trata de importar la información adecuadamente a la entrevista actual, pero también ofrece más oportunidades, ya que el entrevistador puede consultar la información antes de diseñar el instrumento de la encuesta. Igualmente, es posible “limpiar” los datos textuales que se van a utilizar para formular la pregunta, para que la estructura propuesta encaje mejor en la pregunta. El proceso de extraer los datos que se necesitarán durante la entrevista dependiente, de limpiarlos o corregirlos y ofrecerlos a los entrevistadores de una manera apropiada, es a menudo conocido como “alimentar” los datos de una encuesta (Corti y Campanelli, 1992; Jabine, 1990).

En el caso de las entrevistas con lápiz y papel, alimentar los datos de la encuesta es laborioso y susceptible de error. Por ello, pocas entrevistas utilizaban la entrevista dependiente

con anterioridad al uso de los métodos de entrevista asistidos por ordenador (CAI). La entrevista dependiente sólo se utilizaba antes cuando se tenía clara evidencia de que la calidad de los datos resultantes mejoraría notablemente (Neter and Waksberg, 1964) o si la naturaleza de los datos que se debían alimentar y la manera en que debían ser utilizados por los entrevistadores era simple. (Holt, 1979). El uso de CAIs (entrevistas personales asistidas por ordenador CAPI y entrevistas telefónicas asistidas por ordenador (CATI)) facilitó enormemente el uso de las entrevistas dependientes, al no existir necesidad de transcripción manual, ya que el peso de buscar la información relevante y la responsabilidad de corregir el contenido de la pregunta correctamente recaía sobre el entrevistador.

Hay muchas maneras posibles de formular y estructurar las preguntas dependientes, pero una diferencia clave son los métodos *proactivos* y *reactivos*. Una entrevista dependiente *proactiva* (PDI) se denomina así porque la información de la entrevista anterior se ofrece proactivamente como parte del proceso de interrogación (Brown *et al*, 1998). 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 (Bureau of Labor Statistics and US Census Bureau, 1997). Si la respuesta es afirmativa, se supone que la labor en el trabajo no ha cambiado y se le pregunta al entrevistado si sus actividades o deberes han cambiado desde la entrevista anterior. Si el entrevistado informa que no ha habido cambios en sus actividades o deberes, entonces se le 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. El uso de estas preguntas dependientes redujo enormemente el cambio aparente (que se pensaba había sido dudosa en gran medida) y también se ocupaba del tema de las quejas de los entrevistados en cuanto a la repetitividad (Cantor, 1991; Norwood y Tanur, 1994; Polivka y Rothgeb, 1993). Los estudios sobre la encuesta US Survey of Income and Program Participation (SIPP) (Hill, 1994) y la encuesta British Household Panel Survey (BHPS) (Sala and Lynn, 2004) llegaron a conclusiones similares. Ambas encuestas introdujeron posteriormente la PDI para las preguntas sobre la ocupación y la actividad económica. Además de los temas de la ocupación y actividades económicas, los detalles de la composición de los hogares están entre los tipos de preguntas para las que se utiliza normalmente la PDI (Mathiowetz and McGonagle, 2000).

En cuanto a las entrevistas dependientes *reactivas* (RDI), 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. Esta información se usa en la posterior edición de los datos (Hale and Michaud, 1995). Otros ejemplos de la RDI son los de las encuestas de los Estados Unidos, US Government Agricultural Surveys, en donde se pregunta a los agricultores sobre los cambios de las superficies de sus cultivos (Pafford, 1988) y a los ganaderos se les pregunta sobre los cambios en el número de sus cabezas de ganado (Stanley and Safer, 1997).

La razón principal por la que se prefiere la entrevista dependiente de la independiente es para reducir un error de medición, especialmente cuando se cree que se dan cambios dudosos. Existen pruebas (Hill, 1994; Lynn et al, 2004a; Rips, 2000; Webber, 1994) que avalan el punto de vista de que si se realiza un cuestionamiento independiente habrá una tendencia a sobreestimar el cambio, especialmente en el caso en que las categorías incluyen largas listas de elementos similares o cuando las respuestas con final abierto requieren codificación posterior para marcos complejos. Por otra parte, como sugieren Bates y Okon (2003) la PDI podría dar pie a algún sesgo, haciendo que un cambio dudoso sea simplemente reemplazado por una estabilidad dudosa, aunque no se tenga evidencia de ello (Lynn et al, 2004b). La RDI debería evitar ese posible sesgo, aunque puede no tener tanta efectividad como la PDI en la reducción del cambio dudoso (Lynn et al, 2004a; Sala y Lynn, 2004). Otras razones para preferir el modo de entrevistas dependientes incluyen algunos problemas sobre la responsabilidad del entrevistado y el entrevistador, además de los costes (Jäckle, 2005). Si existe una estabilidad considerable en los valores verdaderos, la PDI tiene el potencial de reducir (y significativamente, en algunos casos) el número de preguntas que es necesario formular y el número de respuestas con final abierto que los entrevistadores necesitan registrar y después codificar. Weinberg (2002) sostiene que la introducción de las entrevistas dependientes ha reducido la duración de la entrevista para el SIPP. Jäckle (2005) también ofrece pruebas de cómo las preguntas dependientes pueden reducir los tiempos de las entrevistas.

Ejemplo 1 de entrevista dependiente

Lynn et al (2004a) informa sobre un experimento en el cual una muestra de más de 1,000 personas, a quienes se había entrevistado previamente unos 16 meses antes, es distribuida aleatoriamente en tres grupos de tratamiento: el de las entrevistas independientes (INDI), las entrevistas dependientes proactivas (PDI) y las entrevistas dependientes reactivas (RDI).

Un conjunto de preguntas tiene que ver con las fuentes de ingresos. Los entrevistados bajo la modalidad de entrevistas INDI debían mirar cuatro tarjetas, cada una de las cuales mostraba una lista de posibles fuentes de ingresos. La primera tarjeta enumeraba 6 tipos de pensión, la segunda enumeraba 10 prestaciones relacionadas con incapacidades o perjuicios, la tercera enumeraba otras 9 prestaciones y la cuarta enumeraba otras 8 fuentes diversas de ingresos, más una categoría que podía reunir las todas, “cualquier otro pago regular”. El entrevistado debía decir si había recibido algún tipo de ingreso o pago desde la entrevista anterior. Primero se preguntó a los entrevistados de la PDI por cada fuente que había sido mencionada en la entrevista anterior como actualmente recibida, “De acuerdo con nuestros informes, cuando le entrevistamos la última vez, el día <fecha>, estaba usted percibiendo <fuente> usted sólo o conjuntamente. ¿Durante cuántos meses, desde entonces, ha recibido <fuente>? Luego les mostraron cuatro tarjetas y les preguntaron si había percibido cualquier otro tipo de ingresos enumerados. Primero se preguntó a los entrevistados de la RDI la pregunta estándar INDI, utilizando las cuatro tarjetas. Luego, por cada fuente que había sido mencionada en la entrevista anterior, pero no en la actual, se le preguntó al entrevistado, “¿Puedo verificar algo? De acuerdo con nuestra información, usted ha percibido en el pasado <fuente>. ¿Ha recibido <fuente> en algún momento desde <fecha>?”

Se halló que (Tabla 5) el nivel de información sobre una fuente de ingresos en particular, entre los miembros de la muestra que habían mencionado esa fuente en el ciclo anterior, era significativamente inferior en el caso de INDI que con al menos uno de los métodos dependientes en 7 de 8 fuentes de ingreso, con un tamaño de muestra suficientemente grande como para un análisis significativo. Ello sugiere que hay un fallo de información en el caso de INDI.

	INDI	RDI	PDI
Fuente de ingresos			
Pensión de jubilación NI	99	100	100
Pensión del ex empleador	91	100*	100*
Prestación por incapacidad	71	96*	85
Complemento de ingreso	82	83	98**
Subsidio por hijos	68	86**	93***
Crédito tributario para familias de clase obrera	57	68	87*
Subsidios para la vivienda	78	94**	94**
Subsidios sociales para impuestos	79	94**	95**

Nota: Los porcentajes correspondientes a RDI y PDI se comparan por separado con el porcentaje correspondiente a INDI utilizando un test de Pearson χ^2 en la tabla 2 x 2 correspondiente, con una corrección de correlación intrahogares, aplicada en *Stata* utilizando *svytab*. * indica $0.01 < P \leq 0.05$, ** $0.001 < P \leq 0.01$, *** $P < 0.001$.

Tabla 5: Ingresos: informe de ciclo t condicional a informe de ciclo $t-1$

En el caso de las fuentes de ingreso 6 y 8 (todas son prestaciones sociales), se obtuvieron los datos de validación de registros del gobierno y se utilizaron para estimar directamente los niveles de error de medición de los datos de las encuestas. En la Tabla 6 observamos que las entrevistas dependientes sí que parecen haber reducido los errores de medición, aunque los efectos no sean uniformes. La tasa negativa falsa (informe erróneo por omisión) es significativamente más baja para la DI, en comparación con la de INDI, en el caso de dos fuentes de ingresos, mientras que la tasa positiva falsa (informe erróneo por exceso de información...) no es significativamente diferente para ninguna fuente. El error de medición general (la diferencia entre el valor verdadero tal y como se mide con los datos de la administración y el valor observado en la encuesta) es menor que el de RDI en el caso de estas fuentes de ingresos y menor en el caso de PDI para una fuente.

Proporción de filas	de		Verdadero negativo	Verdadero positivo	Falso negativo	Falso positivo	Admin.	Encuesta	Dif.
			p_{00}	p_{11}	p_{10}	p_{01}	$p_{1\bullet}$	$p_{\bullet 1}$	$p_{01} - p_{10}$
Pensión jubilación	de	INDI	.698	.298	-	.004	.298	.302	.004
		PDI	.663	.330	.004	.004	.333	.333	.000
		RDI	.664	.321	.011	.004	.332	.325	-.007
Prestación incapacidad	por	INDI	.882	.057	.057	.004	.115	.061	-.054
		PDI	.897	.058	.042	.004	.100	.062	-.038
		RDI	.869	.073	.040	.018	.113	.091	-.022*
Complemento de ingreso	de	INDI	.790	.179	.023	.008	.202	.187	-.015
		PDI	.785	.180	.035	-	.215	.180	-.034
		RDI	.818	.168	.015	-	.183	.168	-.015
Subsidio por hijos		INDI	.767	.172	.050	.012	.221	.183	-.038
		PDI	.774	.192	.008**	.027	.199	.218	.019*
		RDI	.770	.208	.018*	.004	.226	.212	-.014*
Crédito familiar		INDI	.901	.057	.023	.019	.080	.076	-.004
		PDI	.877	.092	.004*	.027	.096	.119	.023
		RDI	.894	.077	.026	.004	.102	.080	-.022
Subsidios para la vivienda		INDI	.767	.179	.038	.015	.218	.195	-.023
		PDI	.644	.264	.061	.031	.326	.295	-.031
		RDI	.668	.274	.029	.029	.303	.303	.000**

Notas: p_{ab} es la proporción de entrevistados del grupo ab, donde = 1 si así lo indican los datos de la administración, 0 si no; b = 1 si así lo indica la respuesta de la encuesta, 0 si no. Las columnas con el título "Admin." y "Encuesta" indican tasas de prevalencia por cobro estimado de los datos de la administración y de la encuesta, respectivamente. Las bases son 262 casos INDI, 261 PDI y 274 RDI. Información más detallada disponible en Lynn et al (2004b).

Tabla 6: Indicadores de cobro de ingresos provenientes de encuestas y de datos administrativos

Este estudio también aporta indicios de que todos los miembros de la muestra no tenían las mismas probabilidades de proporcionar un informe erróneo por omisión sobre sus fuentes de ingresos con la modalidad INDI. Algunas personas parecían más capaces que otras de informar que estaban percibiendo ingresos sólo en respuesta a una pregunta con DI. Se trataba de personas relativamente jóvenes, personas que no vivían con un cónyuge o pareja y personas con discapacidad (Tabla 7). Ello sugiere que el error de medición puede estar

correlacionado con las características de los entrevistados, lo cual introduciría con seguridad un sesgo en algunos tipos de estimaciones de encuestas.

	Informantes independientes	Informantes reactivos
	%	%
Hombres	32	33
Nacidos antes de 1943	55	25**
Con trabajo remunerado	26	33
Jubilados	47	22**
Receptores de pensión NI (en ciclo t-1)	48	22**
Niños menores de 12 en el hogar	21	14
Vive con cónyuge o pareja	58	31**
Estudios universitarios	4	6
Salud general "excelente" o "buena"	52	50
Con incapacidad reconocida	23	39*
Ha vivido en el hogar más de 1 año	94	92
Usa coche regularmente	43	44
Tiene teléfono móvil	52	61
Le gusta su vecindario actual	90	83
<i>Base</i>	<i>198</i>	<i>36</i>

Nota: El análisis se basa en 234 entrevistados RDI que declaran percibir al menos una de las 8 fuentes de ingreso enumeradas en la Tabla 5. Los informantes independientes son los que siempre declararon dichas fuentes en respuesta a una pregunta independiente; los informantes reactivos son los que declararon al menos una de dichas fuentes sólo como respuesta a la pregunta de seguimiento reactiva, no habiendo identificado inicialmente la fuente ante una pregunta independiente. * indica $0.01 < P \leq 0.05$, ** $0.001 < P \leq 0.01$, *** $P < 0.001$

Tabla 7: Características de los informantes independientes y reactivos en cuanto a fuentes de ingreso

Ejemplo 2 de entrevista dependiente

En el mismo estudio experimental descrito en el ejemplo 1, también se hicieron preguntas sobre el empleo. Más específicamente, se preguntó a los entrevistados que trabajaban en esos momentos sobre tres características de su trabajo (ocupación, estatus como trabajador y si el entrevistado tenía responsabilidades directivas o de supervisión) y tres características del empleador (actividad económica, tipo de organización, número de trabajadores). En el caso del grupo INDI, las preguntas eran las mismas que en la entrevista anterior. Al grupo PDI se le proporcionó la respuesta que había dado en la entrevista anterior y

se le preguntó si la situación todavía era la misma. Si el entrevistado contestaba que no, se le hacía una pregunta estándar. Primero se preguntó al grupo RDI la pregunta estándar, pero después se le formuló una pregunta de verificación para que el entrevistado confirmara si se había producido algún cambio desde la última vez. En cuanto a la ocupación y al empleador, la pregunta de verificación fue formulada a *todos* los entrevistados de la RDI, aportando las respuestas recibidas la última vez. En lo que respecta al estatus del trabajador, el estatus de los directivos y el número de trabajadores, sólo se hacía la pregunta de verificación si la respuesta dada no correspondía con la respuesta recibida en el ciclo 8. En este caso nos centramos en la estimación del cambio (con relación a la entrevista anterior) de cada una de estas seis características, resumiendo algunos resultados de Sala y Lynn (2004).

Se descubrió que el método PDI produce niveles de cambio más bajos en cuanto a la ocupación, la actividad económica y el número de trabajadores observados (Tabla 8, primeras tres columnas de cifras). Por ejemplo, utilizando la UK Standard Occupational Classification (Clasificación Ocupacional Estándar del Reino Unido) (SOC), con las ocupaciones clasificadas en 371 grupos, se puede observar que con INDI, 53% de los entrevistados con trabajo parecían haber cambiado de ocupación entre las dos entrevistas, mientras que en el caso del grupo PDI, esta proporción cae hasta el 24%. 43% de los entrevistados INDI parecen haber cambiado de actividad económica en la que trabajan (utilizando 503 clases de actividades económicas) cayendo hasta el 20% con el método PDI.

Esta reducción en el cambio observado parece representar una reducción de error de mediciones, ya que el efecto de PDI es especialmente claro entre los entrevistados que no han informado de ningún cambio de trabajo entre los ciclos (Tabla 8, las tres últimas columnas de cifras). Los niveles de cambio en cuanto a las características del trabajo entre los entrevistados con INDI que no han declarado haber cambiado de trabajo siguen siendo demasiado altos.

Porcentaje que declara cambios	Todos los entrevistados que trabajan durante ambos ciclos			Todos los entrevistados con la misma ocupación en ambos ciclos		
	INDI	RDI	PDI	INDI	RDI	PDI
<u>Ocupación</u>						
Trabajos directivos (2)	14.7	9.2	8.9	16.1	1.3***	2.2***
Trabajador/ autónomo	2.7	2.9	1.3	0	2.2	0.9
Grupo principal SOC (9)	31.0	28.3	20.1*	23.7	15.2	0***
Grupo minoritario SOC (76)	46.5	35.5	20.8***	33.0	20.7+	0***
Grupo unitario SOC (371)	52.8	41.3+	23.5***	37.1	25.0	0***
<u>Organización contratante</u>						
Tipo de organización (8)	10.1	9.2	6.7	8.0	7.7	4.3
Número de trabajadores (9)	49.6	42.4	32.8**	38.4	37.7	18.9**
Secciones SIC (17)	26.6	22.0	16.4*	14.7	11.4	0***
Divisiones SIC (60)	30.9	28.0	17.9*	18.9	15.9	0***
Grupos SIC (222)	38.8	40.9	20.1***	25.3	29.5	0***
Clases SIC (503)	43.2	43.2	20.1***	30.5	33.0	0***
<i>Base</i>	<i>146</i>	<i>138</i>	<i>150</i>	<i>97</i>	<i>92</i>	<i>106</i>

Nota: las bases de algunas estimaciones son ligeramente más pequeñas de lo indicado debido a la falta de respuesta de algunos elementos. Los apartados de RDI y PDI se comparan por separado con INDI utilizando un test Pearson χ^2 unilateral en la tabla 2x2 correspondiente. El signo + indica $0.06 \geq P > 0.05$; * $0.05 \geq P > 0.01$; ** $0.01 \geq P > 0.001$; *** $0.001 \geq P$.

Tabla 8: Porcentaje que declara cambios en las características de su trabajo

Sala y Lynn (2004) también estudiaron las características demográficas de los entrevistados cuyas respuestas tenían que ver con el método de formulación de las preguntas. Descubrieron que el método PDI tiene menos probabilidades de marcar una diferencia en las estimaciones del cambio en cuanto a los entrevistados de menos de 36 años que en el caso de los que tenían 36 años o más. Para los tres niveles de la Clasificación Industrial Estándar (SIC) (grupos 17, 60 y 222) y para el nivel de SOC del grupo 76, se observa una reducción significativa del nivel de cambio sólo en el caso de los grupos de más edad. En cuanto al nivel SOC del grupo 9, en las responsabilidades, el tipo de organización y el número de trabajadores, se observa una reducción significativa sólo para uno de los dos grupos de más edad. Igualmente, el efecto del PDI parece más importante en el caso de los entrevistados con más preparación académica. En cuanto a dos variables, la que corresponde a las

responsabilidades directivas y al grupo 371 de SOC, lo mismo parece ser cierto cuando se trata de RDI. En cuanto al sexo, la DI parece tener más influencia en el caso de los hombres que en el de las mujeres, aunque esas diferencias son contradictorias. Existen tres medidas en las cuales las entrevistas RDI afectan sólo a los hombres. En el caso del PDI, hay dos medidas que influyen sólo en los hombres y una que sólo afecta a las mujeres.

				Edad			Notas		
		Hombres	Mujeres	18-35	36-50	51+	Baja	Media	Alta
Responsabilidades directivas	PDI				*				
	RDI	*							**
Trabajador/autónomo	PDI								
	RDI								
Grupo unidad SOC (371)	PDI	***	**	*	***	**	*	***	***
	RDI	*			*				**
Grupo minoritario SOC (76)	PDI	***	**		***	**		**	***
	RDI								
Grupo principal SOC (9)	PDI	**			*				*
	RDI								
Tipo de organización	PDI				*				
	RDI								
Número de trabajadores	PDI		**			*	*		*
	RDI								
Secciones de SIC (17)	PDI				*	**			*
	RDI	*							
Divisiones de SIC (60)	PDI	*			**	**	*		*
	RDI			*					
Grupos de SIC (222)	PDI	**	**		**	**	*		*
	RDI								
Clases de SIC (503)	PDI	***	**		***	***	*	*	**
	RDI								

Notas: para cada celda de la tabla, la proporción de entrevistados que muestran un cambio de característica de trabajo relevante es comparada con el grupo de INDI utilizando una prueba unilateral denominada Pearson χ^2 test. Las notas se definen de la siguiente manera: 'Alto' es un 'A level' con pass o equivalente; 'Medio' es al menos un GCSE, 'O level', CSE o pass equivalente; 'Bajo' es un no pass en GCSE o equivalente. * indica $0.05 \geq P > 0.01$; ** $0.01 \geq P > 0.001$; *** $0.001 \geq P$

Tabla 9: Correlativos socio demográficos del efecto de las entrevistas dependientes sobre las estimaciones del cambio

En resumen, estos dos ejemplos de los efectos de las entrevistas independientes parecen mostrar que:

- Los errores de medición (en el caso de las entrevistas independientes) no son aleatorios y por ello introducen sesgo en las estimaciones de los niveles y las estimaciones del cambio;
- Las entrevistas dependientes pueden reducir este sesgo;
- Sin embargo, las entrevistas PDI y RDI no efectivas por igual (en nuestros ejemplos, solo el método PDI parece efectivo para el cambio de las características de trabajo, pero ambos parecen igualmente efectivos en lo referente a las fuentes de ingresos);
- La posibilidad de que la entrevista PDI introduzca algunos errores (diferentes) no debería dejarse de lado, aunque no encontremos pruebas de que tales errores sean importantes.

Calendarios y métodos relacionados

Tal y como se sostenía en el apartado 1.1, uno de los puntos fuertes de las encuestas longitudinales es que los investigadores de encuestas necesitan confiar menos en la memoria y capacidades de recuerdo de los entrevistados. Los datos que cubren largos períodos de tiempo pueden obtenerse preguntado sólo sobre períodos de tiempo relativamente cortos en cada entrevista. Sin embargo, aún así es necesario algún tipo de recuerdo. El intervalo entre los ciclos de una encuesta puede durar de una semana a cinco años o más, dependiendo de la naturaleza de la encuesta. De manera que, incluso en el caso de las encuestas longitudinales, los errores de medición que pueden ocurrir debido a la incapacidad de los entrevistados de recordar los sucesos con suficiente precisión es un tema de interés. Los factores que pueden influir en la capacidad de los entrevistados de recordar sucesos incluyen no sólo la longitud del tiempo transcurrido desde el suceso, sino también la importancia del suceso y el que ocurrieran otros sucesos en ese lapso que puedan distorsionar el recuerdo del suceso en cuestión (Bound et al, 2001; Eisenhower et al, 1991; Mathiowetz y Duncan, 1988; Waksberg y Valliant, 1978).

Los investigadores han descubierto que la capacidad de los entrevistados de recordar sucesos como respuesta a las preguntas de una encuesta, pueden mejorar con la ayuda de “señales” o ayudas de memoria adecuadas. Una sencilla manera de ayuda para la memoria es simplemente conceder al entrevistado tiempo para pensar en las circunstancias que rodeaban

el suceso en cuestión y otros sucesos que puedan haber sucedido en esas fechas. Una manera de facilitar estos pensamientos es formular algunas preguntas “contexto”. Se trata de preguntas de encuesta que se formulan con el fin de intentar proporcionar algún contexto para el proceso de pensamiento del entrevistado, aunque las respuestas a esas preguntas puedan no ser necesarias para los fines del análisis. Por ejemplo, si una encuesta debe formular preguntas sobre los cambios de trabajo durante un período de tiempo en particular, estas preguntas podrían ir precedidas por algunas preguntas sobre algún traslado de casa o cambios de pareja durante el mismo período.

Otra manera de mejorar esta idea es el uso de Calendarios de la Historia de los Sucesos, (EHCs), antes denominados Life History Calendars (LHCs). Los EHCs han sido concebidos para facilitar un recuerdo retrospectivo de datos objetivos, partiendo del modelo estándar de formular preguntas fijas (sin variación en las palabras) en un orden fijo (sin variaciones en el orden de las preguntas) (Belli et al, 2001; Freedman et al, 1988). Los EHCs han sido diseñados para recoger datos cronológicos para varios dominios en paralelo, utilizando un enfoque de entrevistas flexible. Esto aprovecha las estructuras idiosincráticas de la memoria autobiográfica. Dicho de otro modo, cada persona guarda y recuerda la información de su memoria, de modos distintos y, por ello, un enfoque estándar para las entrevistas no será lo más óptimo para algunos entrevistados. El EHC es un cuadro que utilizan el entrevistador y el entrevistado para indicar cuándo tuvieron lugar los sucesos. Los EHCs también han sido utilizados para las entrevistas telefónicas.

Por ejemplo, Belli et al (2001) describen un EHC que recoge datos en siete dimensiones: los sucesos que marcan un hito para una persona, la residencia, la composición hogar, el trabajo y otras actividades, el tiempo que se pasa lejos del trabajo, y las prestaciones sociales. Se pueden utilizar los EHCs para recoger datos sobre una amplia gama de dimensiones, dependiendo del tema de la encuesta.

La Figura 14 presenta un ejemplo de un EHC de datos en cinco dimensiones: ubicación de la residencia, composición hogar, educación y prestaciones laborales y sociales. En este caso, los datos se registran en meses y el período de referencia son tres años. Los años y los meses han sido colocados previamente en el extremo superior de la página y los meses también se repiten dos tercios por debajo de la página para facilitar la referencia. Dentro de cada dimensión, los datos indican los períodos de cada situación, estando cada período

marcado por algún suceso. Por ejemplo, la dimensión 3 es la ubicación de la residencia. El entrevistado vivía en Londres SE3 al principio del período de referencia, hasta julio de 2003. Este período terminó con el suceso del traslado a Colchester, donde vive el entrevistado desde ese momento. Se han marcado los sucesos (y el principio / final del período de referencia) con una "X", y los períodos de una situación en particular se indican con líneas.

Podemos observar la utilidad del diseño del EHC como una ayuda a la memoria si observamos la dimensión 6, trabajo. La entrevistada tenía un trabajo a tiempo completo durante el tiempo que vivió en Londres. Esto terminó el mismo mes en que se trasladó de casa, en Julio de 2003, tras un mes sin trabajo. Luego, empezó un trabajo de media jornada y aumentó sus horas hasta que, en enero de 2004 volvía a trabajar a tiempo completo. Este trabajo de jornada completa terminó en septiembre de 2004, y el mes siguiente aparece un niño en el hogar. Lo que sucedió fue que la encuestada tuvo un hijo en octubre de 2004. Este suceso, y el traslado de casa en julio de 2003 son con toda probabilidad sucesos muy significativos que el entrevistado puede recordar, y (incluida la fecha) bien. Pero también tienen conexión con sus experiencias laborales (y con las prestaciones sociales). De manera que, indicar primero la ubicación y la composición del hogar en el EHC, sirve como un marco útil para recordar las fechas de los cambios de trabajo y de prestaciones sociales percibidas. Parece muy probable que las fechas relativas a los períodos de transición en el trabajo y a los subsidios sociales percibidos sean más precisas que si se hubiera pedido información de ellos utilizando un cuestionario estándar, sin el contexto inmediato (y visual) de un EHC. Es más, existe evidencia de que los EHCs mejoran la calidad de los informes retrospectivos, mediante el recuerdo de los sucesos y el posterior fechado de ellos (Axinn et al 1997; Belli et al, 2001; Belli et al, 2004; Caspi et al, 1996).

1 Year	2003												2004												2005											
2 Month	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D
3 Where	London SE3																																			
	Colchester, Essex																																			
4 Household																																				
Spouse	X												X	X											X											
Partner																																				
Children																						X														X
Mother																																				
Father																																				
Sibling(s)																																				
Other rel																																				
Non-rel																																				
Alone																																				
Non-HH																																				
5 Education																																				
Full-time																																				
Part-time																																				
None																																				
6 Employment	J <th>F</th> <th>M</th> <th>A</th> <th>M</th> <th>J</th> <th>J</th> <th>A</th> <th>S</th> <th>O</th> <th>N</th> <th>D</th> <th>J</th> <th>F</th> <th>M</th> <th>A</th> <th>M</th> <th>J</th> <th>J</th> <th>A</th> <th>S</th> <th>O</th> <th>N</th> <th>D</th> <th>J</th> <th>F</th> <th>M</th> <th>A</th> <th>M</th> <th>J</th> <th>J</th> <th>A</th> <th>S</th> <th>O</th> <th>N</th> <th>D</th>	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D
30 hrs+	X																																			
10-29 hrs																																				
1-9 hrs																																				
None																																				
7 Welfare																																				
Work credit																																				
Unemp																																				
Child credit																																				

Figura 14: Un calendario histórico de sucesos

Además de servir como ayuda a la memoria, mejorando así la calidad de la información aportada por los entrevistados, los EHCs también pueden servir para facilitar verificaciones para el momento de editar las entrevistas. Las presentaciones visuales facilitan que el entrevistador pueda identificar contradicciones en los registros de las respuestas, como un registro simultáneo de las situaciones que deberían ser mutuamente exclusivas (como son las dos categorías de “trabajo” en la Figura 14), o la falta de datos de una dimensión que debería ser representada en cada período de tiempo (como la dimensión 2, ubicación del lugar de residencia, en la Figura 14). 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.

5. PESOS Y ATRIBUCIONES DE LAS ENCUESTAS LONGITUDINALES

5.1 Introducción a la ponderación

Ponderar consiste en darle a cada unidad muestral de una encuesta un valor numérico (peso), representando la contribución que hará la unidad a las estimaciones basadas en los datos de la encuesta (Lynn, 2004). Los pesos han sido diseñados para que la muestra pueda ser representativa de la población de estudio. El peso de cada unidad de respuesta particular se puede interpretar como el número relativo de unidades de población que representa. El cálculo y la aplicación de los pesos es parte del proceso de inferencia estadística, por la cual se pueden sacar conclusiones sobre una población de interés sobre la base de las informaciones de una muestra obtenida de la población.

Existen cuatro razones principales para realizar la ponderación: corregir las diferencias de las tasas de cobertura, para elegir las probabilidades y probabilidades de respuesta (y falta de respuesta), y para corregir los efectos de las varianzas de muestreo aleatorio. Con respecto a las fuentes de los errores de encuesta presentados en la Figura 1, estas responden a los objetivos de minimizar un sesgo de cobertura, un sesgo de muestreo, un sesgo por falta de respuesta y una varianza de muestreo. Sea cual fuere la fuente del error, el objetivo básico de la ponderación es asegurar que los grupos de las unidades de muestreo estén representados en la misma proporción en la que aparecen dentro de la población. Estos factores sirven para cualquier encuesta. Los rasgos de la ponderación que son distintivos en el caso de las encuestas longitudinales tienen que ver principalmente con la definición de la población de estudio y con el tratamiento de los patrones de falta de respuesta complejos. En el caso de algunas encuestas longitudinales, pueden ser necesarios también algunos métodos especiales para estimar las probabilidades de selección. En esta sección, nos centraremos en estos rasgos distintivos de las encuestas longitudinales.

5.2 Pesos longitudinales

Muchas encuestas longitudinales proporcionan dos tipos de pesos a los usuarios de datos: “pesos transversales” y “pesos longitudinales”. Esta distinción refleja una diferencia en la población que se ha de representar, lo cual, a su vez, está relacionado con diferentes objetivos de estimación. Como ejemplo, veamos una encuesta simple de 2 ciclos con datos recogidos de

los puntos en el tiempo $t1$ y $t2$. Los investigadores pudieron utilizar los datos de una de las siguientes maneras:

1. Para realizar estimaciones longitudinales, utilizando los datos de $t1$ y $t2$;
2. Para realizar estimaciones transversales, utilizando sólo los datos de $t1$;
3. Para realizar estimaciones transversales, utilizando sólo los datos de $t2$.

Estos tres tipos de análisis son distintos *tanto* en términos de qué unidades de respuesta contribuyen al análisis *como* en términos de población a la cual se refieren las estimaciones, como se resume en la Tabla 10. Para cada tipo de análisis, se pueden elaborar las ponderaciones de cualquiera de las maneras habituales, lo cual normalmente implica dividir la población e grupos y calcular el peso de cada grupo como la proporción del número de unidades de población con relación al número de las unidades de muestra. Así que, por cada tipo de análisis, los numeradores (los tamaños o proporciones de la población) y los denominadores (tamaños o proporciones de muestra entrevistados) de los pesos serán diferentes, de manera que se necesitan tres grupos de pesos diferentes. Los pesos correspondientes al análisis del tipo 1 se denominan pesos longitudinales, mientras que los análisis del tipo 2 y tipo 3 son tipos de pesos transversales.

Tipos de análisis	Población de estudio	Unidades de respuesta
1	En una población de $t1$ y $t2$ (N_{12})	Respuestas en $t1$ y $t2$ (n_{12})
2	En una población de $t1$ (N_1)	Respuestas en $t1$ (n_1)
3	En una población de $t2$ (N_2)	Respuestas en $t2$ (n_2)

Tabla 10: Poblaciones de estudio y muestras

Obsérvese que el tamaño de la población del análisis del tipo 1 (N_{12}) no es necesariamente mayor que, ya sea, N_1 o N_2 . Si hay algún nacimiento o muerte en la población entre $t1$ y $t2$, que es normalmente el caso, entonces $N_{12} < N_1$ y $N_{12} < N_2$. Si los nacimientos sobrepasan las muertes, entonces $N_1 < N_2$ y *vice versa*. Las relaciones equivalentes tienen relación con los tamaños de la muestra. Un problema común de las encuestas longitudinales es encontrar una buena manera de estimar la distribución de la población, entre las variables relevantes, para cada una de las poblaciones del estudio. Las

estimaciones transversales de la población se pueden conseguir fácilmente de fuentes externas, pero ese casi nunca es el caso de las poblaciones longitudinales. Normalmente, un enfoque basado en un modelo se usa para estimar las características de las poblaciones longitudinales, tomando la población transversal del ciclo 1 como punto de partida y después utilizando estimaciones basadas en muestras de los nacimientos y muertes para dar forma a los cambios de la estructura de la población. Esto puede ser dudoso, dado que es probable que las estimaciones que se basan en muestras estén sujetas a errores de falta de respuesta, que es una de las fuentes de error cuyos errores que quisiéramos que corrigieran las ponderaciones.

5.3 ¿Qué combinaciones de ciclos?

En el caso del ejemplo de la Tabla 10, sería posible crear tres grupos de pesos y ponerlos a disposición de los usuarios de los datos. Pero en general existen posibles poblaciones $2^t - 1$ que pueden ser representadas por una encuesta longitudinal de ciclo t , de las cuales t son poblaciones transversales y $2^t - (t + 1)$ son poblaciones longitudinales. Por ello, existen, potencialmente, conjuntos t de pesos transversales y conjuntos $2^t - (t + 1)$ de pesos longitudinales que podrían crearse. En el caso de las encuestas con más de 2 o 3 ciclos, no sería factible crear todos estos conjuntos de pesos. Por ejemplo, si $t=10$, entonces $2^t - 1 = 1,023$. También podría ser confuso para los usuarios tener tantos conjuntos de pesos disponibles. Y probablemente no es necesario, de todos modos, ya que muchos de los conjuntos de pesos serían tan parecidos unos con otros que la elección entre ellos no tendría prácticamente ningún efecto sobre las estimaciones.

Una solución a este problema sería proporcionar a los usuarios de los datos todos los datos necesarios para calcular los pesos correspondientes a cualquier combinación de ciclos (y también algunas pautas o incluso un programa que calcule los pesos). Luego, cada usuario puede especificar el grupo de ciclos relevantes a su análisis y producir pesos según sus necesidades. Sin embargo, esto se hace pocas veces, ya sea porque no se pueden obtener algunos datos necesarios al nivel de la unidad o porque los usuarios prefieren claramente que se les proporcionen los pesos listos para utilizar.

Una alternativa práctica es que quien tiene los datos en su poder elabore las ponderaciones para un subconjunto limitado de posibles combinaciones de ciclos. Ello debería venir acompañado de una pauta para los usuarios sobre qué hacer si no se dispone de datos sobre la combinación que les interesa. La selección de las combinaciones de ciclos debería guiarse por los usos principales (más probables) de los datos. Por ejemplo, si el objetivo principal de la encuesta es permitir el análisis del cambio en relación con los datos de la línea de base que se recogieron en el ciclo 1, entonces, no tiene mucho sentido elaborar ponderaciones para combinaciones de ciclos que no incluyen el ciclo 1. Si se incluye un módulo de preguntas sobre un tema en particular sólo en los ciclos 1, 4, 7 y 10, entonces, esa combinación en particular debería ser una candidata fuerte para la ponderación. Para casi todas las encuestas longitudinales, el grupo completo de ciclos debería ser una de las combinaciones para las que se elaboran las ponderaciones. La única excepción debería darse si, por cuestiones de diseño, no existen unidades que puedan ser elegibles para la recogida de datos de cada ciclo.

Sin embargo, es importante saber que realizar análisis basados en los entrevistados de un conjunto en particular de ciclos que utilizan los pesos diseñados para un conjunto diferente de ciclos no es lo más indicado. Tomemos como ejemplo la encuesta de 3 ciclos de la Figura 13 (apartado 3.3) y supongamos que sólo se proporciona un conjunto de pesos longitudinales, diseñados para hacer que el conjunto de personas que respondieron a los tres ciclos sea representativo de la población longitudinal de los 3 ciclos. Supongamos que queremos estimar algún parámetro de cambio entre el ciclo 1 y el ciclo 3, para lo cual sólo necesitamos utilizar los datos recogidos en los ciclos 1 y 3. Podríamos utilizar todas las unidades que tienen patrones de respuesta 1 (XXX) o 4 (X0X). Pero los pesos longitudinales serán fijados en 0 para las unidades con patrones de respuesta 4. Ningún peso longitudinal se define para estas unidades. Por ello, el 10% de los casos disponibles (900 de 9,296) serán apartados del análisis por no disponer de un peso adecuado. Para esta estimación, habría sido mejor elaborar un conjunto de pesos para representar el X?X de la población (las unidades que están en la población en la época de los ciclos 1 y 3 sin importar si también están en la población de la época del ciclo 2). Estos serían distintos de cero para todas las unidades de las muestras XXX y X0X.

Otra consideración importante es que generalmente se elaboran conjuntos de pesos en varios puntos distintos del tiempo durante la vida de una encuesta longitudinal. Esto se hace a menudo después poder contar con los datos de cada nuevo ciclo, ya que los investigadores

querrán analizar los últimos datos sin tener que esperar a que el próximo ciclo acabe. Así que, como mínimo, en cada ciclo se elaborará un conjunto de pesos que represente a la población longitudinal en todos los ciclos hasta la fecha. Ello significa que será posible encontrar los pesos más recientes para cada “muestra de desgaste”. Por ejemplo, después de cinco ciclos habrá pesos para las muestras X0000, XX000, XXX00, XXXX0 y XXXXX. Si para las encuestas se lleva una política de intento de recoger datos solamente de los entrevistados de un ciclo anterior (política 3 en el apartado 3.3), no se necesitarán más pesos que estos. De otra manera, la tarea consiste en identificar qué otras combinaciones de ciclos son lo suficientemente importantes como para garantizar el cálculo de los pesos.

5.4 ¿Qué variables?

Una vez que el investigador ha identificado las poblaciones longitudinales relevantes para las cuales se elaboran los pesos, todavía queda por identificar un *método* de cálculo de los pesos y un conjunto de *variables auxiliares* que definirán las clases de ponderación y los pesos. Los criterios tanto para el método como para las variables no son distintos de los de cualquier tipo de encuesta. Básicamente, (Lynn, 2004), el objetivo es elegir un método y un conjunto de variables de manera que cuando se utilice el método para crear un conjunto de clases definido por las variables, las clases resultantes tengan las siguientes propiedades:

- Las propensiones de inclusión (tasas de cobertura, probabilidades de selección, probabilidades de respuesta) varían de acuerdo con las clases;
- Los valores de las estadísticas de las muestras más importantes (por ejemplo, las medias, las proporciones, los coeficientes de regresión, etc.) varían de acuerdo con las clases;
- Los valores de las estadísticas de las muestras más importantes para las unidades incluidas y excluidas (las que han sido objeto de muestra o no, las de los entrevistados y no entrevistados) dentro de cada clase.

En muchos casos es recomendable calcular los pesos de cada fuente de errores de encuesta por separado (por ejemplo, los pesos de diseño por probabilidades de selección, los pesos post-estratificación por varianza de muestreo y pesos de falta de respuesta por propensión de respuesta). Los criterios se aplican en cada etapa.

En el caso de una encuesta longitudinal, debemos recordar que las estadísticas de muestra más importantes tenderán a ser medidas de cambio y medidas de asociación de otras variables con las medidas del cambio. Es probable que esto tenga consecuencias importantes en la creación de las clases de ponderación. Las variables auxiliares que se correlacionan con más claridad con estas medidas de cambio pueden ser con toda probabilidad variables de encuestas de los ciclos anteriores, y, en particular, las medidas del cambio de los períodos previos. Por esta razón, la ponderación de la falta de respuesta de las encuestas longitudinales se hace, muchas veces, en orden secuencial. En cuanto a la falta de respuesta del ciclo 1, es necesario utilizar datos externos a la encuesta como datos auxiliares. Pero a partir de ese punto (a menos que la encuesta incluya respuestas de los últimos ciclos de las unidades que no respondían al ciclo 1) la propensión de respuesta en los ciclos siguientes puede ser estimada según la respuesta del ciclo 1 (u otros ciclos previos). En su forma más simple, los pesos de la falta de respuesta (NR) correspondientes a las muestras de desgaste podrían ser calculadas de la siguiente manera:

- Los pesos del ciclo 1 NR (w_1) utilizan datos auxiliares externos a la encuesta;
- Los pesos del ciclo 2 NR, sujetos a la respuesta del ciclo 1 ($w_{2|1}$) utilizan los datos del ciclo 1 como datos auxiliares. El peso del ciclo 2 NR es $w_2 = w_1 \times w_{2|1}$;
- Los pesos de los ciclos 3 NR, de acuerdo con las respuestas del ciclo 2 ($w_{3|2}$) utilizan los datos de los ciclos 1 y 2 como datos auxiliares (tal vez incluyendo las medidas del cambio entre los ciclos 1 y 2). El peso del ciclo 3 NR es $w_3 = w_2 \times w_{3|2}$;
- Etc.

Esta forma simple, evidentemente, debería ser corregida si las unidades que tienen patrones de falta de respuesta tuvieran que ser incluidos en el análisis.

5.5 Introducción a la Imputación

Los investigadores de cualquier tipo de datos se encuentran normalmente con faltas de respuesta a algún elemento y deben encontrar maneras de hacer frente a esta circunstancia.

Pueden elegir el análisis de todo el caso o, mejor dicho, el análisis de lo que disponen de dicho caso, pero esto no es efectivo y se pierden datos. Si las variables son categóricas, entonces la falta de respuesta a un elemento puede ser tratada como una categoría sustantiva separada en el análisis. La ponderación puede ser utilizada para hacer frente la falta de respuesta, tal y como se ha mencionado anteriormente, pero la muestra disponible de las unidades de respuesta sería diferente para cada estimación, siendo necesario calcular los pesos cada vez. Esto tomaría mucho tiempo (uno de los atractivos de la ponderación como método de ajuste es que una vez que los pesos son calculados, pueden ser utilizados para todos los análisis).

Un método popular para resolver el problema de los datos que faltan es la *Imputación*. Esto consiste en asignar (imputar) un valor cuando falta un elemento en el conjunto de datos. Lo que atrae de la Imputación es que tiene como resultado un conjunto de datos completos, de manera que se pueden utilizar métodos de análisis estándar y el mismo conjunto de unidades contribuirá con cualquier estimación correspondiente a la misma población de estudio (con el análisis de caso disponible, las bases de las muestras pueden cambiar de una estimación a la siguiente creando contradicciones entre las estimaciones). En la práctica, los métodos estándar deberían ajustarse para responder al hecho de que parte de los valores de los datos no son, de hecho, valores observados, sino más bien, valores imputados.

Existen muchas maneras de elegir el valor que se ha de imputar. Las clases de métodos de imputación incluyen métodos deductivos y basados en reglas, la imputación de la media o el modo, la imputación de la media o el modo dentro de una clase, la imputación aleatoria, la imputación Hot deck, el equivalente de la función a distancia y la imputación de regresión. Como sucede con la ponderación, la elección del método de Imputación es sólo una parte de la historia. La otra alternativa importante es la alternativa de los datos auxiliares y cómo deberían ser utilizados.

5.6 Imputación longitudinal

En el contexto de una encuesta longitudinal, una característica importante de la imputación es con frecuencia que los datos auxiliares pueden incluir valores del período previo de la variable para la cual se está buscando un valor imputado(o incluso se pueden restringir a ese período). En una encuesta transversal, es posible que la renta deba ser imputada sobre la base de variables tales como el estatus de la actividad económica, la ocupación (en caso de

tener un empleo), el sexo, la edad y la composición del hogar. Pero tratándose de una encuesta longitudinal, la renta puede ser imputada sobre la base de la renta de cada ciclo previo. A menudo, la renta del ciclo anterior será una variable explicativa de la renta del ciclo actual mucho más adecuada que cualquier conjunto de medidas del ciclo actual. Por ello, igual que con la ponderación, los datos auxiliares correspondientes a la imputación pueden incluir las medidas de sondeo de los ciclos previos, lo cual es, por lo general, una ventaja.

5.7 Revisión de las imputaciones

Habiendo acabado, digamos, el ciclo 2 de una encuesta longitudinal, la renta recogida en el ciclo 1 puede ser, con toda probabilidad, la mejor variable explicativa disponible de la renta para el ciclo 2 y, por ese motivo, la mejor elección de la variable auxiliar para realizar la imputación de la renta. Pero luego, cuando los datos del ciclo 3 ya estén disponibles, se puede dar el caso de que algunos entrevistados que no habían informado sobre su renta en el ciclo 2 (y por ello se les imputaba) hayan proporcionado su información sobre la renta en los ciclos 1 y 3. Para estos entrevistados, el uso de los ciclos 1 y 3 en combinación puede proporcionar una imputación mejor (y distinta) para el ciclo 2 que el uso únicamente los datos del ciclo 1. Pero para ese momento los investigadores ya habrán estado utilizando los datos del ciclo 2 con la imputación que se entregó durante el ciclo 2. El encargado de proporcionar los datos se enfrenta entonces a un dilema, y tiene las opciones las siguientes:

- Entregar imputaciones revisadas (con suerte, mejores) con el fin de sustituir las que se presentaron con anterioridad, sobre la base de que siempre se deberían proporcionar los mejores datos posibles, aunque ello signifique que habrá contradicciones entre los análisis llevados a cabo en distintos puntos en el tiempo;
- No revisar ninguna imputación que ya se haya proporcionado, con el fin de evitar contradicciones entre los análisis. Si se adopta esta estrategia, las ventajas del uso de datos de ciclos subsiguientes como variables auxiliares para la imputación puede llevar a la conclusión de que no se deberían haber hecho las imputaciones hasta el ciclo subsiguiente. Por ejemplo, no se pondrán a disposición las imputaciones de los valores que faltaban en ciclo 2 hasta que se publiquen los datos del ciclo 3. Sin embargo, ello todavía puede llevar a estimaciones contradictorias.

- Proporcionar imputaciones revisadas y también seguir proporcionando las imputaciones originales, con el fin de que los investigadores puedan verificar la fiabilidad de los resultados en relación con los procedimientos de imputación.

El debate sobre si se deben o no revisar las imputaciones es un asunto importante cuando se trata de las encuestas longitudinales. Se han adoptado distintas políticas según cada encuesta y es difícil ser categórico, ya que la mejor política dependerá de las características de cada encuesta en particular. Sin embargo, existe la idea general de que es importante que se tome en cuenta la opinión de los usuarios cuando haya que tomar decisiones sobre la política, y además, en la etapa más temprana posible.

REFERENCIAS

- Axinn, W.G., Barber, J.S. & Ghimire, D.J. (1997) The neighbourhood history calendar: a data collection method designed for dynamic multilevel modelling, *Sociological Methodology* 27, 355-392.
- Belli, R.F., Shay, W.L. and Stafford, F.P. (2001) Event History Calendars and Question List Surveys, *Public Opinion Quarterly*, Vol. 65, 45-74.
- Belli, R. F., Lee, E. H., Stafford, F. P. and Chou, C-H. (2004). Calendar and question-list survey methods: Association between interviewer behaviors and data quality. *Journal of Official Statistics*, 20, 185-218.
- Bennett, D.J. & Steel, D. (2000) An evaluation of a large-scale CATI household survey using random digit dialling, *Australian and New Zealand Journal of Statistics* 42, 255-270.
- Biemer, P.P. & Lyberg, L.E. (2003) *Introduction to Survey Quality*, New York: John Wiley & Sons.
- Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A. & Sudman, S. (ed.s) (1991) *Measurement Errors in Surveys*, New York: John Wiley & Sons.
- Binder, D. (1998) Longitudinal Surveys: why are these surveys different from all other surveys? *Survey Methodology*, 24:2, 101-108
- Bound, J., and Krueger, A.B., "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" in *Journal of Labor Economics*, Vol. 9(1), 1991, pp. 1-24.
- Brown, A., Hale, A. and Michaud, S. (1998) 'Use of Computer Assisted Interviewing in Longitudinal Surveys', in M. P. Couper, et al. (eds) *Computer Assisted Survey Information Collection*, New York: John Wiley and Sons.
- Bureau of Labor Statistics and Bureau of the Census (1997). CPS Questionnaire. Available at <http://www.bls.census.gov/cps/bqestair.htm>.
- Burkhead , D. and Coder, J. (1985) 'Gross Changes in Income Recipiency from the Survey of Income and Program Participation', *Proceedings of the Social Statistics Section*, Washington DC, American Statistical Association, 351-356.
- Burton, J., Laurie, H. & Lynn, P. (2006, forthcoming) The long-term effectiveness of refusal conversion procedures on longitudinal surveys, *Journal of the Royal Statistical Society (Series A): Statistics in Society*, 169:3
- Campanelli, P., Sturgis, P. and Purdon, S. (1997) *Can you hear me knocking: an investigation into the impact of interviewers on survey response rates*. London: National Centre for Social Research.

Campanelli, P. and O'Muircheartaigh, C. (1999) Interviewers, Interviewer Continuity, and Panel Survey Nonresponse, *Quality & Quantity*, 33:1, 59-76

Campanelli, P. and O'Muircheartaigh, C. (2002) The Importance of Experimental Control in Testing the Impact of Interviewer Continuity on Panel Survey Nonresponse, *Quality & Quantity*, 36:2, 129-144

Cantor, D. (1991). Draft recommendations on dependent interviewing, unpublished manuscript, Rockville MD, Westat.

Caspi, A., Moffitt, T.E., Thornton, A., Freedman, D., Amell, J.W., Harrington, H., Smeijers, J. & Silva, P.A. (1996) The life history calendar: a research and clinical assessment method for collecting retrospective event-history data, *International Journal of Methods in Psychiatric Research* 6, 101-114

Corti, L. and Campanelli, P. (1992). The utility of feeding forward earlier ciclo data for panel studies, pp. 109-118 in Westlake, A. et al (ed.s) *Survey and Statistical Computing*, North Holland, Elsevier.

Dillman, D.A. (2000) *Mail and internet surveys: the tailored design method* (2nd ed.). New York: Wiley.

Dillman, D.A., Eltinge, J.L., Groves, R.M. & Little, R.J.A. (2002) Survey nonresponse in design, data collection, and analysis. In: R.M. Groves, D.A. Dillman, J.L.Eltinge & R.J.A.Little (ed.s), *Survey nonresponse*. New York: Wiley.

Eisenhower, D., Mathiowetz, N.A., Morganstein, D. (1991) Recall Error: Sources and Bias Reduction Techniques, in Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A., Sudman, S. (eds.), *Measurement Errors in Surveys*, New York: Wiley.

Freedman, D., Thornton, A., Camburn, D., Alwin, D. & Young-DeMarco, L. (1988) The life history calendar: a technique for collecting retrospective data, *Sociological Methodology* 18, 37-68.

Groves, R. M. (1989) *Survey Errors and Survey Costs*, New York: John Wiley & Sons.

Groves, R.M. & Couper, M.P. (1998) *Nonresponse in Household Interview Surveys*. New York: John Wiley.

Groves, R.M., Dillman, D.A., Eltinge, J.L. & Little, R.J.A. (2002) *Survey Nonresponse*, New York: John Wiley & Sons.

Groves, R.M., Singer, E. and Corning, A. (2000) Leverage-saliency theory of survey participation: description and an illustration, *Public Opinion Quarterly* 64, 299-308

Groves, R.M., Fowler, F.J.Jr., Couper, M.P., Lepkowski, J.M., Singer, E. & Tourangeau, R. (2004) *Survey Methodology*, New York: John Wiley & Sons.

Halpin, B. (1998) 'Unified BHPS Work-Life Histories: Combining Multiple Sources Into a User-Friendly Format', *Bulletin de Méthodologie Sociologique*, 60(Oct.): 34-79.

Hill, D. H. (1987) 'Response Errors Around the Seam: Analysis of Change in a Panel with Overlapping Reference Periods', *Proceedings of the Survey Research Methods Section*, Washington DC, American Statistical Association.

Holt, M. (1979). The use of summaries of previously reported interview data in the National medical Care Expenditure Survey: a comparison of questionnaire and summary data for medical provider visits, in Sudman, S. (ed.) *Health Survey Research Methods*, Washington DC, US Department of Health and Human Services.

Jabine, T.B. (1990). *SIPP Quality Profile*, Washington DC, US Department of Commerce, US Bureau of the Census.

Jäckle, A. (2005) Does Dependent Interviewing Really Increase Efficiency and Reduce Respondent Burden? Working Paper 2005-11 of the Institute for Social and Economic Research. Colchester: University of Essex.

<http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2005-11.pdf>

Kalton, G. & Citro, C.F. (1993) Panel surveys: adding the fourth dimension. *Survey Methodology* 19 205-215.

Kalton, G. and Miller, M. E. (1991) 'The Seam Effect with Social Security Income in the Survey of Income and Program Participation', *Journal of Official Statistics*, 7(2): 235-245.

Kalton, G., Miller, D. P. and Lepkowski, J. (1992) 'Analyzing Spells of Program Participation in the SIPP' Technical Report, Survey Research Centre, Ann Arbor: University of Michigan.

Kasprzyk, D., Duncan, G., Kalton, G. & Singh, M.P. (ed.s) (1989) *Panel Surveys*, New York: John Wiley & Sons.

Kish, L. (1987) *Statistical Design for Research*, New York: John Wiley & Sons.

Kulka, R.A. & Weeks, M.F. (1988) Toward the development of optimal calling protocols for telephone surveys: a conditional probabilities approach, *Journal of Official Statistics* 4, 319-3

Laurie, H., Smith, R. and Scott, L. (1999) Strategies for reducing nonresponse in a longitudinal panel survey, *Journal of Official Statistics* 15:2, 269-282

Lemaitre, G. (1992) 'Dealing with the Seam Problem for the Survey of Labour and Income Dynamics', *SLID Research Paper Series*, No. 92-05, Ottawa: Statistics Canada.

Lynn, P. (2004) Weighting. In: Kempf-Leonard, K. (ed.), *Encyclopedia of Social Measurement*. London: Academic Press.

Lynn, P. (2005) Outcome Categories and definitions of response rates for panel surveys and other surveys involving multiple data collection events from the same units. Colchester: ISER, University of Essex.

Lynn, P. (2006, forthcoming) The problem of non-response. In: Hox, J., de Leeuw, E. and Dillman, D.A. (ed.s), *The International Handbook of Survey Methodology*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Lynn, P., Purdon, S., Hedges, B. & McAleese, I. (1994) *The Youth Cohort Study: An assessment of alternative weighting strategies and their effects*, Employment Department Research series YCS Report no.30

Lynn, P., Jäckle, A., Jenkins, S. and Sala, E. (2004a) The effects of dependent interviewing on responses to questions on income sources, Working Paper 2004-16 of the Institute for Social and Economic Research. Colchester: University of Essex.

<http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2004-16.pdf> (also forthcoming in *Journal of Official Statistics*)

Lynn, P., Jäckle, A., Jenkins, S. and Sala, E. (2004b) The impact of interviewing method on measurement error in panel survey measures of benefit receipt: evidence from a validation study, Working Paper 2004-28 of the Institute for Social and Economic Research. Colchester: University of Essex.

<http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2004-28.pdf>

Lynn, P. J, Nicolaas, G., Beerten, R., Laiho, J., and Martin, J. (2005) *Recommended Standard Final Outcome Categories and Standard Definitions of Response Rate for Social Surveys* (2nd edition). Colchester: ISER, University of Essex.

Mangione, T.W. (1995) *Mail surveys: improving the quality*. Thousand Oaks, California: Sage.

Martini, A. (1989) Seam effect, recall bias, and the estimation of labor force transition rates from SIPP, *Proceedings of the Section on Survey Research Methods 1989*, American Statistical Association, 387-392.

Mathiowetz, N., Duncan, G. (1988) Out of work, out of mind: response errors in retrospective reports of unemployment, *Journal of Business and Economic Statistics*, Vol. 6, 221-229.

Mathiowetz, N.A. and McGonagle, K.A. (2000) An assessment of the current state of dependent interviewing in household surveys, *Journal of Official Statistics*, 16:4, 401-418.

Moore, J. and Kasprzyk, D. (1984) 'Month-to-Month Reciprocity Turnover in the ISDP', *Proceedings of the Survey Research Methods Section*, Washington DC, American Statistical Association, 726-731.

Morton-Williams, J. (1993) *Interviewer approaches*. Aldershot: Dartmouth.

Neter, J. and Waksberg, J. (1964) A study of response error in expenditures data from household interviews, *Journal of the American Statistical Association*, 59, 18-55.

Norwood, J.L. and Tanur, J.M. (1994) Measuring unemployment in the nineties, *Public Opinion Quarterly*, 58:2, 277-294.

Oppenheim, A.N. (1992) *Questionnaire Design, Interviewing and Attitude Measurement*, New Edition, London, Pinter Publishers.

Pascale, J. and Mayer, T.S. (2004) Exploring confidentiality issues related to dependent interviewing: preliminary findings, *Journal of Official Statistics*, 20:2, 357-377.

Pafford, B. (1988) The influence of using previous survey data in the 1986 April ISO Grain Stock Survey, Washington DC, National Agricultural Statistical Service.

Philippens, M. & Billiet, J. (2004) Monitoring and evaluating nonresponse issues and fieldwork efforts in the European Social Survey. Proceedings of the European Conference on Quality and Methodology in Official Statistics (Q2004), Mainz, May 2004 (CD-ROM). Wiesbaden: Federal Statistical Office Germany.

Polivka, A. and Rothgeb, J. (1993) Redesigning the CPS questionnaire, *Monthly Labor Review*, 116:9, 10-28.

Rendtel, U. (1990) Teilnahmebereitschaft in Panelstudien: Zwischen Beeinflussung, Vertrauen und Sozialer Selektion, *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 42:2, 280-299

Rips, L. (2000) Unraveling the seam effect, Proceedings of the Section on Survey Research Methods 2000, American Statistical Association, 465-470.

Rope, D. (1993) Preliminary longitudinal nonresponse research with the CPS and CE. 4th international workshop on household survey nonresponse, Bath, UK

Sala, E. and Lynn, P. (2004) Measuring change in employment characteristics: the effects of dependent interviewing, Working Paper 2004-26 of the Institute for Social and Economic Research. Colchester: University of Essex.

<http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2004-26.pdf>

Särndal, C.-E. & Lundström, S. (2005) *Estimation in surveys with nonresponse*. Chichester: Wiley.

Singer, E., Van Hoewyk, J. & Gebler, N. (1999) The effects of incentives on response rates in interviewer-mediated surveys, *Journal of Official Statistics*, 15, 217-230.

Swires-Hennessy, E. & Drake, M. (1992) The optimum time at which to conduct interviews, *J. Market Research Society* 34, 61-72.

Taylor, B., Heath, A. and Lynn, P. (1996) The British Election Panel Study 1992-95: response characteristics and attrition, Centre for Research into Elections and Social Trends Working Paper no.40, University of Strathclyde

Trivellato, U. (1999) Issues in the design and analysis of panel studies: A cursory review, *Quality & Quantity*, 33:3, 339-352

Waksberg, J., Valliant, R. (1978) Final Report on the Evaluation and Calibration of NEISS (Westat, Inc. for Consumer Products Safety Commission).

Waterton, J. and Lievesley, D. (1987) Attrition in a panel survey of attitudes, *Journal of Official Statistics* 3:2, 267-282

Webber, M. (1994) The Survey of Labor and Income Dynamics: Lessons Learned in Testing, SLID Research Paper 94-07, Statistics Canada, Ottawa.

Weinberg, D.H. (2002) The Survey of Income and Program Participation – Recent History and Future Developments, SIPP Working Paper 232, US Department of Commerce, Bureau of the Census, Washington DC.

Young, N. (1989) 'Ciclo-Seam Effects in the SIPP', ASA Survey Research Methods Section, 393-398.