



Documento de soporte del PaD FUNDACIÓ JAUME BOFILL

Panel de Desigualtats Socials a Catalunya

Manual de uso de las matrices

Olas de la A (2001/02) a la K (2012)

JUNIO 2014



Índex

Presentación.....	3
1 La encuesta PaD.....	4
1.1. Datos sobre estructura y desigualdades sociales.....	4
1.2. Aspectos objetivos y subjetivos de la vida de los individuos.....	4
1.3. Una encuesta longitudinal.....	4
1.4. Dos unidades de estudio: el hogar y sus individuos.....	6
1.5. El diseño de la muestra (año 2001).....	7
1.5.1. Criterios de seguimiento de la muestra	
1.5.2. Evolución de la muestra	
1.5.3. Remuestreos	
1.5.4. Ponderación	
1.5.4.1. Hogares	
1.5.4.2. Individuos	
1.6. Los cuestionarios.....	14
2 Las matrices de datos.....	15
2.1. Tipos de matrices.....	15
2.2. Estructura de las matrices.....	16
2.3. Tipos de variables i nomenclatura.....	16
2.4. Valores perdidos.....	18
2.5. Los identificadores individuales y de hogar.....	18
3 Los materiales de soporte.....	19
3.1. El informe longitudinal de contenidos.....	19
3.2. Los libros de códigos.....	20
3.3. Documentos metodológicos.....	20
Anexos	
Anexo 1. Tablas de referencia de los gráficos sobre muestra.....	22
Anexo 2. Construcción de los identificadores numéricos.....	24
Anexo 3: Cuestionario.....	25
Anexo 4: Muestra.....	28
Anexo 5: Trabajo de campo.....	30
Anexo 6: Datos.....	33
Anexo 7: Ficha técnica del análisis.....	35
Anexo 8: Ficha técnica de difusión.....	36

Presentación

El *Manual de uso* pretende ser una herramienta de apoyo para todos los investigadores e investigadoras que trabajen con la base de datos del Panel de Desigualdades Sociales en Cataluña (en lo sucesivo, PaD). Las particularidades metodológicas, técnicas y de contenido de una encuesta como el PaD hacen del todo recomendable la lectura detallada de este documento antes de profundizar en la explotación de los datos. Se ha procurado que fuera lo más breve posible, seleccionando cuidadosamente aquella información que parecía indispensable para poder entender adecuadamente el tipo de datos y la estructura de los ficheros del PaD.

En esta ocasión el *Manual de uso* se ofrece también en castellano, para permitir a los investigadores e investigadoras hispanohablantes la posibilidad de trabajar con los datos del PaD. Hay un apartado de la web en castellano [enlace](#) en el que se encuentra toda la documentación en esta lengua. No ha sido posible traducir la totalidad, ni de la documentación, ni de las matrices de datos disponibles. Sin embargo, y dada la proximidad del castellano con el catalán, se quiere animar al uso de la documentación y los datos que sean de interés, en su lengua original. En este sentido, se puede contar con la disponibilidad por parte del equipo para cualquier consulta o aclaración.

Puede dirigirse cualquier consulta o información sobre el PaD a:

Laia Pineda, coordinación general del PaD lpineda@fbofill.cat

Laura Morató, muestra, datos y análisis del PaD lmorato@fbofill.cat

La encuesta PaD

Datos sobre estructura y desigualdades sociales

El PaD tiene como objetivo principal aportar nueva información, en clave longitudinal, sobre las desigualdades sociales en Cataluña. La encuesta está pensada para obtener información sobre las oportunidades y limitaciones de las personas que viven en Cataluña en relación con diferentes aspectos. Los ámbitos principales en los que se centra el PaD, y respecto a los cuales se recoge información en cada ola, son:

- Situación personal y familiar
- Educación reglada y no reglada
- Trabajo productivo, reproductivo y conciliación
- Usos del tiempo
- Vivienda
- Posición económica
- Salud y discapacidades
- Posición y relaciones sociales
- Valores y actitudes

Aspectos objetivos y subjetivos de la vida de los individuos

Una de las particularidades de la encuesta PaD tiene que ver con la recogida de aspectos subjetivos de la vida de los encuestados. Las estadísticas oficiales disponibles (tanto respecto a Cataluña como a España y Europa) marcan la tendencia a una expansión de las estadísticas sobre los hechos (datos objetivables) y prestan poca atención a la valoración que hacen los individuos de esos hechos, es decir, a la vivencia personal que los hechos generan. Por ello, los cuestionarios del PaD intentan encontrar un equilibrio entre las preguntas que buscan recoger datos de carácter objetivo y las que se centran en los aspectos más subjetivos de la vida de los encuestados.

Una encuesta longitudinal

El PaD es la primera encuesta longitudinal que se ha impulsado desde Cataluña con una muestra estadística y suficientemente representativa de hogares catalanes¹. Optar por una encuesta longitudinal, dada la escasa experiencia que nuestro país ha acumulado en este tipo de encuestas, recomienda basarse en las encuestas europeas. Paneles como el *British Household Panel Survey* (BHPS) actualmente denominado *Understanding Society*,² el *German Socio-Economic Panel* (GSOEP)³ o el

¹ En Cataluña ha habido hogares que han formado parte de muestras de encuestas longitudinales impulsadas por el Instituto Nacional de Estadística (INE) como, por ejemplo, la Encuesta Continua de Presupuestos Familiares (ECPF), que suministra información sobre el origen y el volumen de las rentas de los hogares en España y cómo se utilizan en los diversos gastos de consumo, pero nunca han sido muestras suficientemente representativas desde el punto de vista estadístico como para realizar estudios específicos de Cataluña.

² <http://www.iser.essex.ac.uk/bhps>

³ <http://www.diw.de/en/soep>

*Swiss Household Panel (SHP)*⁴ son encuestas longitudinales de solvencia contrastada. En el momento inicial del PaD, la experiencia británica fue la principal referencia, sin que ello significara realizar una mera exportación de la herramienta. Sin embargo, en los años de desarrollo del PaD ha sido necesario también explorar las experiencias de otros países y otros paneles, adecuando el contenido de la encuesta a las últimas innovaciones que estos han implementado.

Desde un punto de vista estrictamente técnico, la característica básica de una encuesta de panel puro es que se efectúa siempre sobre la misma muestra de individuos a lo largo de las diferentes olas en que se realiza, al contrario que las encuestas de panel rotativo que renuevan parte de la muestra en cada edición. La periodicidad de un panel no está predeterminada, sino que se ajusta a la voluntad y/o posibilidades económicas de cada proyecto. En este sentido, el PaD es una encuesta de panel puro y su periodicidad es anual.

Sin embargo, cabe mencionar que, si bien la muestra inicial se mantiene en las olas sucesivas y el interés específico radica precisamente en su continuidad, el hecho de estar configurada por hogares e individuos confiere también cierta "vida" a la muestra. Por ese motivo, a lo largo de las olas, la muestra experimenta cambios, como la incorporación de nuevos individuos (nacimientos, formación de nuevas parejas, acogida de ancianos en el hogar, reagrupamientos familiares, etc.), la creación de nuevos hogares (con la salida del hogar paterno por parte de los jóvenes que crean un nuevo hogar, por divorcios o separaciones, etc.). En este sentido, la muestra de una encuesta panel combina la dimensión dinámica (de cambio) con la dimensión de estabilidad y continuidad.

Por otra parte, en una muestra panel también hay pérdidas muestrales tanto de hogares como de individuos, que se denominan *attrition*, y se deben a distintos motivos: muerte de los miembros, institucionalización (ingreso en un centro del tipo residencia de ancianos, internado de estudios, prisión, etc.), traslado de todo el hogar fuera del territorio catalán, negativa a seguir participando en la encuesta y otros. Este efecto de *attrition* obliga a realizar nuevas captaciones de hogares durante la vida de la encuesta, a fin de que la muestra no pierda representatividad. Por este motivo, el PaD ha recurrido a la captación de nuevos hogares en diferentes momentos de la encuesta para suplir así la erosión de la muestra. (Véase el apartado 2.5.1.2 Re muestreos)

Las ventajas de realizar el seguimiento de una muestra panel pueden resumirse en cuatro aspectos fundamentales:

- **Calidad:** Una muestra panel asegura poder recoger mucho mejor los datos: el regreso a los mismos hogares en cada ola permite depurar errores y validar los datos año tras año.
- **Cantidad:** La muestra panel permite desarrollar, con diferentes niveles de profundidad, distintos bloques de preguntas en cada una de las olas (un año se profundiza en el tema de la educación, otro en el trabajo doméstico, otro en temas de vivienda, etc.), de modo que puede acumularse mucha más información de cada ámbito a medida que se van sucediendo las olas.
- **Perspectiva del cambio:** La muestra panel permite recoger datos muy valiosos sobre todo tipo de cambios que experimentan tanto los hogares como los individuos (cambios de vivienda, cambios en la estructura de los hogares, cambios en el trabajo, etc.). El hecho de que se pregunte a los mismos individuos año tras año permite conocer no solo qué cambios

⁴ <http://www.swisspanel.ch/?lang=en>

se producen en sus vidas sino que también puede localizarse especialmente en qué grupos sociales se producen, con qué intensidad, sus consecuencias sobre otros aspectos de la vida personal y familiar, etc., sin los sesgos que se introducirían por el hecho de estar preguntando en cada nueva edición de la encuesta a una muestra nueva de individuos. Es la única herramienta que permite explicar la trayectoria de los individuos y de los hogares, es decir, de los procesos que viven en cada momento de sus vidas. De este modo, permite disponer de información privilegiada sobre los acontecimientos más relevantes del ciclo vital.

- **Significación:** La muestra panel implica que el análisis de los procesos es estadísticamente más significativo que el de un análisis transversal. De este modo, pueden realizarse afirmaciones más sólidas sobre los cambios que se producen en la realidad social, aunque estos sean, en términos porcentuales, más pequeños, ya que se reducen los márgenes de error que se derivan de cualquier muestra.

Dos unidades de estudio: el hogar y sus individuos

El PaD tiene la particularidad de tener una doble unidad de estudio: los hogares y sus individuos. Esta característica tiene que ver con la voluntad de ampliar al máximo el potencial explicativo de los datos recogidos. El análisis de los datos del PaD puede ser tanto de los hogares en sí mismos (movilidad geográfica, formación y disolución de hogares, estrategias en la educación de los menores de 16 años, etc.) como de los individuos que los conforman. A eso debe añadirse el interés analítico de disponer, respecto a los individuos, de datos sobre sus hogares, es decir, sobre su contexto inmediato.

Metodológicamente, las implicaciones de este hecho son:

- En cuanto a la muestra: debe efectuarse un muestreo en el que tanto el número de hogares como el de individuos responda a los márgenes de error máximos estipulados.
- En cuanto al cuestionario: ha sido necesario diseñar dos cuestionarios diferentes, uno para los hogares y otro para los individuos, para poder recoger la información en función de la unidad de análisis.
- En cuanto al trabajo de campo:
 - Proceso de entrevistas: ha sido necesario desarrollar un programa que permita codificar los cambios que se produzcan en la composición de los hogares y que ello se refleje inmediatamente durante el proceso de la entrevistas.
 - Captación de hogares: para formar parte de la muestra, los hogares deben acceder a participar en su conjunto. En los años de captación⁵ (1ª Ola y Olas de la 6ª a la 9ª) no se aceptan hogares con participación parcial, además de que tienen que estar de acuerdo en participar en las sucesivas olas de la encuesta.
- En cuanto a los ficheros de datos: para poder realizar un seguimiento de los individuos y los

⁵ La ola de la captación inicial de los hogares del PaD es la primera (2001-2002), pero a partir de la 6ª Ola (2007) efectuamos un remuestreo para iniciar una nueva captación parcial que compense las pérdidas muestrales y refuerce la muestra. Esta captación parcial se prolonga durante la 7ª, 8ª y 9ª Olas del PaD.

hogares a lo largo de las olas ha sido necesario construir identificadores de hogar y de individuo, invariables en el tiempo, que garanticen poder unificar ficheros de olas distintas, así como añadir información de los hogares a los individuos o a la inversa.

El diseño de la muestra (año 2001)

La muestra inicial del PaD consta de 2.000 hogares de Cataluña seleccionados aleatoriamente siguiendo un método estratificado de muestreo en etapas múltiples, con una selección aleatoria sistemática de las primeras unidades muestrales (secciones censales) y una selección aleatoria simple de las unidades finales (direcciones).

Se decide utilizar las secciones censales como primeras unidades muestrales, puesto que son las unidades territoriales estadísticas de uso público más pequeñas. Dado que se trata de agrupaciones territoriales con promedios de entre 1.000 y 1.200 personas, o aproximadamente 400 hogares, poseen una gran coherencia interna y la información disponible permite clasificaciones de estructura social y territorial diferentes. El número idóneo de hogares por encuestar en cada sección censal se fija en 10. Se cuenta con que, de media, los hogares de cada sección censal derivarán en 270 individuos para la muestra final. En relación con la muestra territorial fijada, deben seleccionarse 200 secciones censales de las 5.186 existentes en el año 1996 (actualización del censo de 1991).

A fin de garantizar una buena representatividad de la muestra, se combinan cuatro variables de estratificación, dos de tipo territorial y dos de tipo social:

- provincia (incluyendo la subdivisión en 3 subregiones en el caso de Barcelona);
- ámbito urbano-rural (en función del tamaño de municipio);
- variable construida de posición social (clasificación según \pm clase trabajadora);
- situación ocupacional (ordenación según número de activos/inactivos)⁶.

El resultado de la selección de secciones censales se concreta en 203 unidades ubicadas en 115 municipios de Cataluña.

La muestra de individuos se deriva de contabilizar 2,7 individuos de media en cada hogar seleccionado para el muestreo, resultando en 5.445 individuos para el conjunto de Cataluña.

Para poder garantizar una buena representatividad territorial en la muestra del PaD y compensar la gran concentración poblacional en la región metropolitana de Barcelona (donde habitan más de tres cuartas partes de los hogares catalanes) se opta por una muestra proporcional en lo que respecta a las provincias.

Por ese motivo, y partiendo del ± 5 % como margen máximo de error de la muestra asumible por resultados globales (con un intervalo de confianza del 95,5 %, con $k=2$, y en el supuesto de máxima indeterminación $p=q=50$), es necesario observar un mínimo de 400 hogares en cada una de las cuatro provincias catalanas. En lo tocante a la provincia de Barcelona, su diversidad y complejidad exigen una muestra específica de mayor tamaño, que se fija finalmente en 800 hogares, repartidos proporcionalmente, según la distribución de la población, entre Barcelona ciudad (261), la Región Metropolitana (470) y las Comarcas Centrales (69). Es así como se llega al tamaño final de la muestra, que estará formada por 2.000 hogares distribuidos proporcionalmente a escala provincial. En

⁶La variable de posición social se construye a partir de las preguntas 12 y 13 del Padrón de 1996.

la siguiente tabla, puede observarse el detalle de su distribución y los errores muestrales asociados.

Tabla 1. DISTRIBUCIÓN DE HOGARES E INDIVIDUOS DE LA MUESTRA ORIGINAL⁷ Y ERROR DE LA MUESTRA (1ª ola del PaD: 2001- 2)

ÁREAS	NÚMERO DE HOGARES	e* (%)	NÚMERO DE INDIVIDUOS	e* (%)
Barcelona	791	±3,54	2.213	±2,13
Ciudad BCN	252	±6,12		
Región Metropolitana	470	±4,61		
Comarcas centrales	69	±12,04		
Girona	400	±5,00	1.193	±2,89
Lleida	400	±5,00	1.175	±2,91
Tarragona	400	±5,00	1.176	±2,91
CATALUNYA	1991	±2,78	5.757	±1,32

*e = error de la muestra (intervalo de confianza=95,5%, k=2 | p=q=50)
Fuente: elaboración propia

Durante el trabajo de campo se realizan tres comprobaciones sobre la bondad de la muestra, tanto de los hogares como de los individuos captados (con el 11 %, el 50 % y el 100 % del trabajo de campo realizado), asegurándose de que no existan sesgos importantes que requieran rectificaciones de la muestra diseñada.

1.5.1. Criterios de seguimiento de la muestra

Cualquier encuesta debe marcar unos criterios sobre cual es la muestra que necesita para dar respuesta a sus objetivos. En el caso del PaD, una encuesta longitudinal pura, esto es capital. No sólo para la captación sino para saber en cada nueva ola a qué hogares o individuos se tiene que seguir encuestando y a cuáles no.

Poder formar parte de la muestra quiere decir ser considerado "elegible", y al contrario, no poder formar parte de la muestra, ser "inelegible". En cada ola, entonces, tiene que intentarse encuestar a todos los hogares, y los individuos, elegibles, que vengan de la ola anterior.

En el caso de los hogares, los criterios de elegibilidad son que el hogar esté domiciliado en Cataluña y que sea un hogar privado, es decir, no un centro o institución. En el caso de los individuos el criterio es que sean mayores de 15 años y que vivan en un hogar elegible.

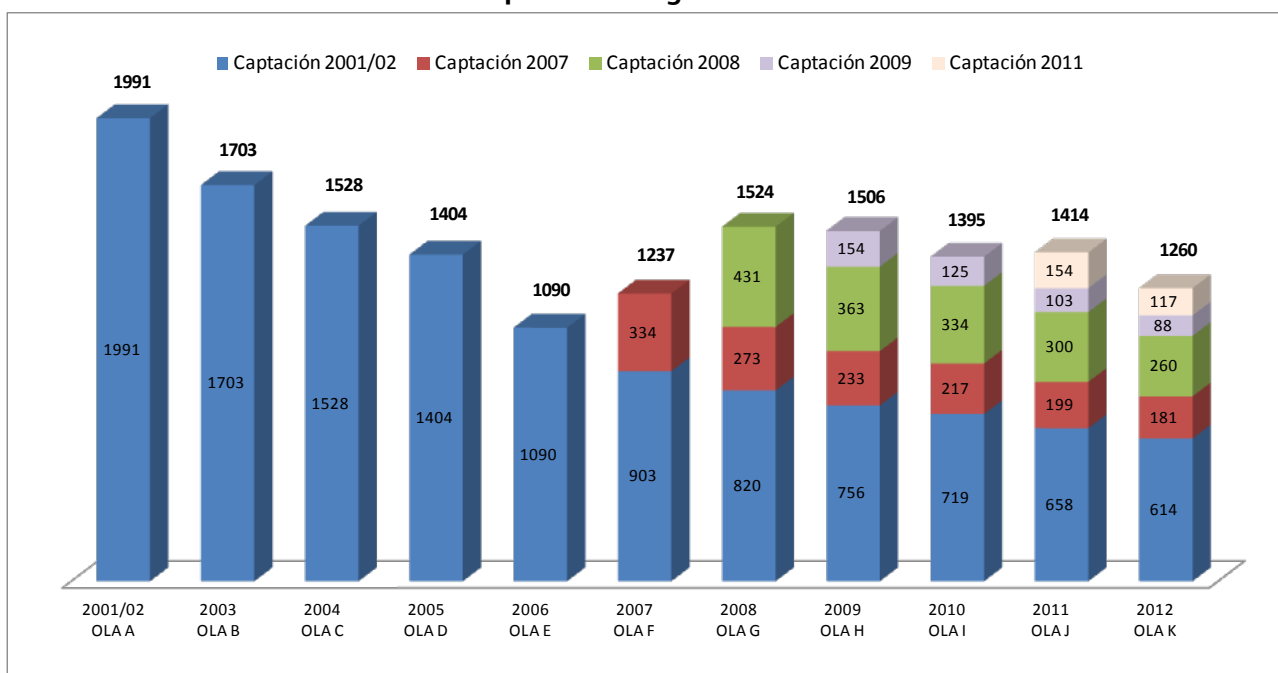
1.5.2. Evolución de la muestra

⁷ Se pueden consultar los errores muestrales de la muestra teórica en el anexo 4. Muestra.

Como ya se ha mencionado más arriba, las encuestas longitudinales, debido a su naturaleza periódica, van perdiendo casos de la muestra año tras año. Por un lado, los hogares y los individuos dejan de ser elegibles (por institucionalización, muerte o traslado fuera de Cataluña). Por el otro, dado que la respuesta a la encuesta PaD es voluntaria, pueden rechazar participar, o ser ilocalizables, entre otras causas.

Des de un punto de vista del análisis longitudinal, interesa conocer cual es la evolución de la muestra, para saber a cuantos hogares i/o individuos se puede seguir a lo largo de las diferentes olas del PaD. Con la voluntad de sintetizar esta evolución, los dos gráficos siguientes muestran el comportamiento de los hogares (gráfico 1) y los individuos (gráfico 2) captados, en las olas en las cuales ha habido captación⁸. Se tiene que tener en cuenta que este no es el total de hogares, ni de individuos que se incluyen en las matrices de datos ya que no se han incluido los nuevos hogares (creados por independización del hogar paterno, separación o divorcio...) ni los nuevos individuos que se han encuestado (porque han ido a vivir a un hogar de la muestra del PaD).

Gráfico 1. Evolución de la muestra captada de hogares

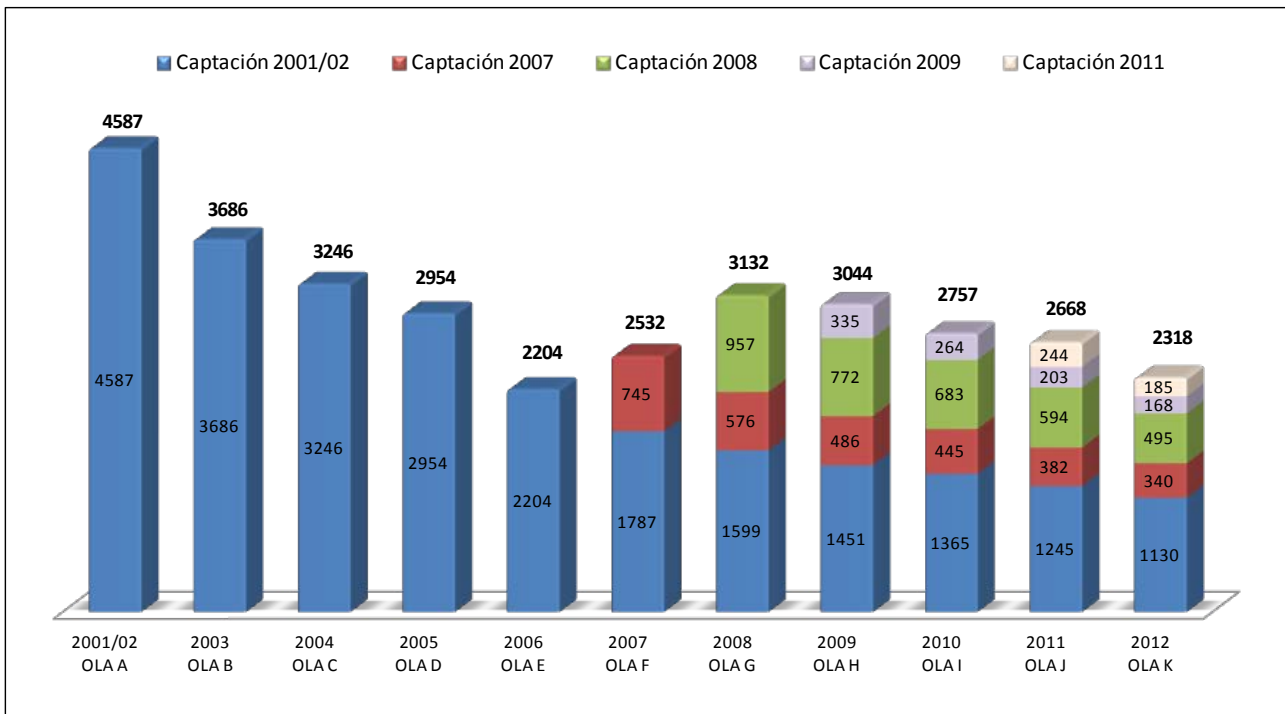


Font: elaboración propia con datos del PaD (1ª a 11ª olas)

Este gráfico incluye únicamente las incorporaciones de hogares a la muestra por captación, no las producidas por las dinámicas propias de los hogares.

⁸ Se pueden consultar las tablas, con los porcentajes de pérdida correspondientes, en el anexo 1.

Gráfico 2. Evolución de la muestra captada de individuos

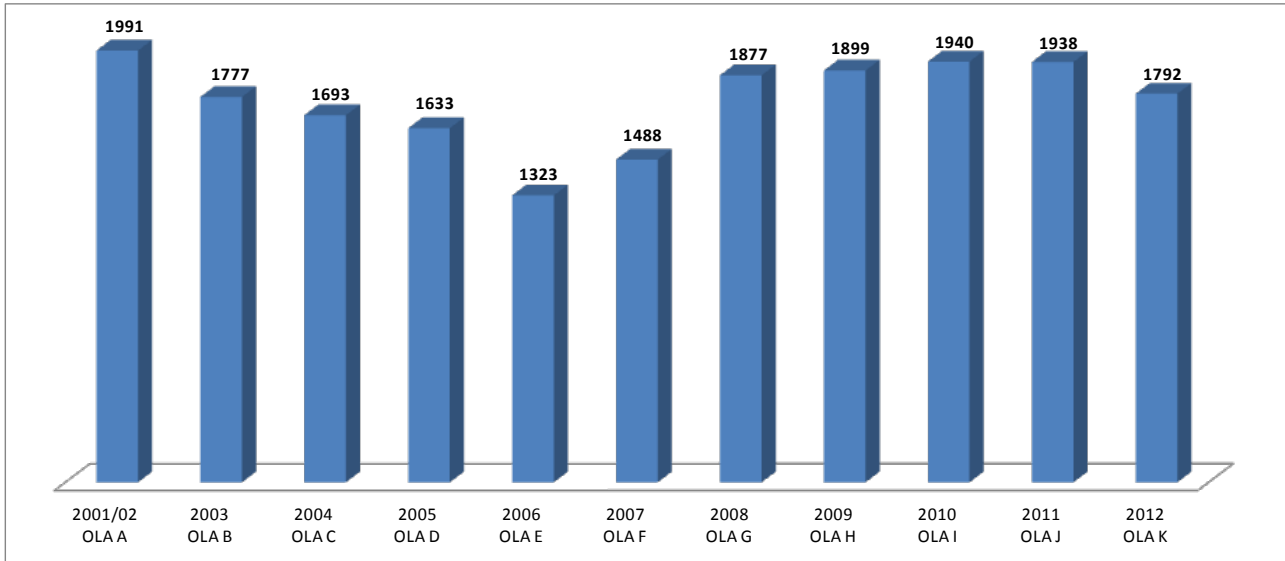


Font: elaboración propia con datos del PaD (1ª a 11ª olas)

Este gráfico incluye únicamente las incorporaciones de hogares a la muestra por captación, no las producidas por las dinámicas propias de los hogares.

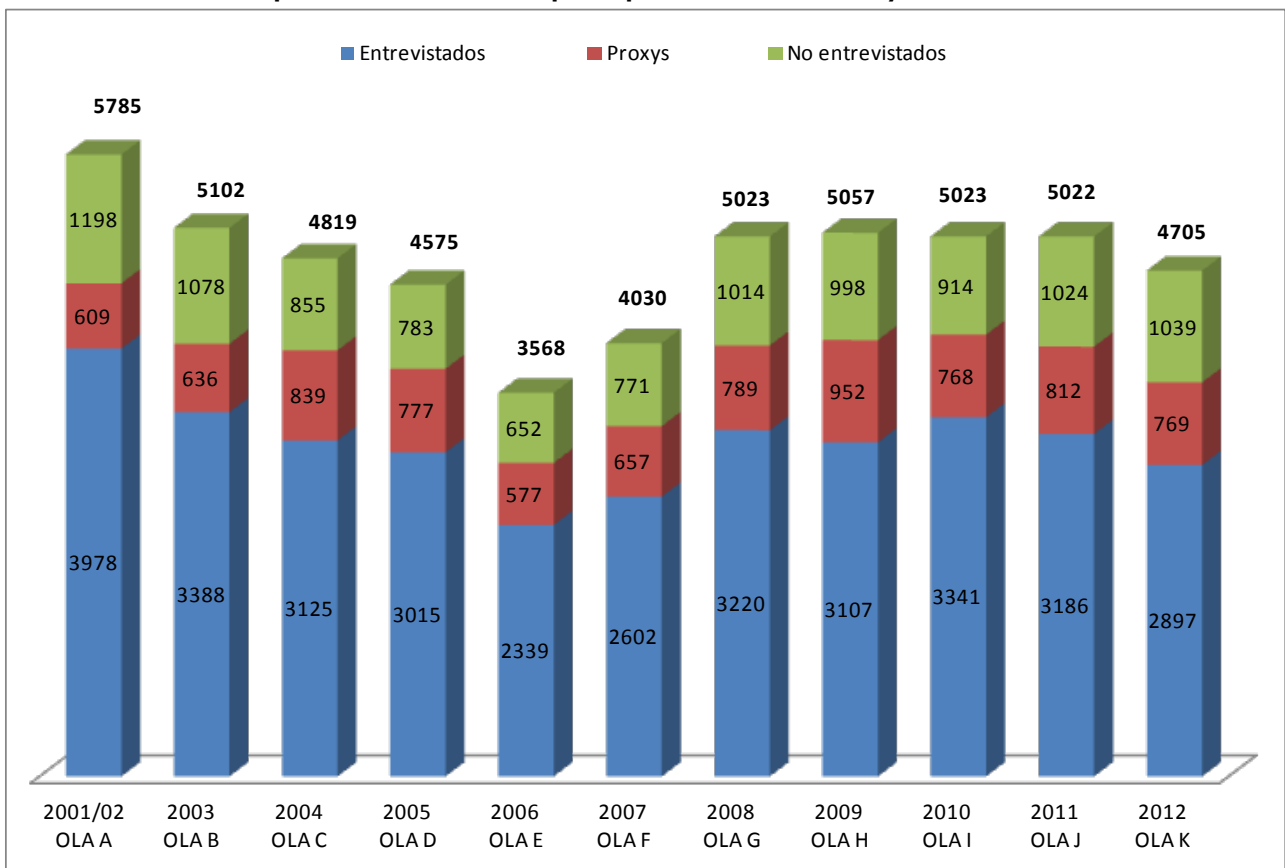
Desde un punto de vista transversal, interesa el total de observaciones que hay en cada ola. Los dos gráficos siguientes quieren presentar esta información de manera clara. En el caso de los individuos (gráfico 4), se ha especificado el tipo de cuestionario (entrevistados, proxys o no entrevistados) ya que cada uno de ellos provee de una cantidad de información diferente. En el apartado correspondiente a cuestionario, encontraréis información más detallada sobre los tipos de cuestionario.

Gráfico 3. Hogares entrevistados en el PaD por ola



Font: elaboración propia con datos del PaD (1ª a 11ª olas)

Gráfico 4. Individuos presentes en el PaD por tipo de cuestionario y ola



1.5.3. Remuestreos

Como se puede ver en los gráficos anteriores, la erosión muestral dejó los 1.991 hogares captados en la 1ª ola en 1.323 en la 5ª. Por eso, en la 6ª Ola (año 2007), gracias a la disponibilidad de recursos económicos, se opta por aumentar los efectivos de la muestra.

El aumento de la muestra del PaD se realiza a partir de un remuestreo, siguiendo el mismo método empleado para la muestra inicial de la 1ª ola (mismas variables de estratificación, el mismo método de selección de las secciones censales finales), pero con una diferencia importante: las secciones censales de la muestra inicial con hogares supervivientes son el punto de partida del nuevo muestreo, que deberá completarse con las secciones censales que garanticen la representatividad territorial y social deseadas. Esta medida responde a la necesidad de contener la dispersión de encuestadores por el territorio catalán y facilitar al máximo la gestión del trabajo de campo. Cuando por los resultados del remuestreo no es posible utilizar las mismas secciones censales, se seleccionan de forma aleatoria nuevas secciones censales.

Las dificultades de captación que se experimentan en el trabajo de campo obligan a dilatar la captación prevista de 1.074 hogares durante la 6ª ola en las olas 7ª (2008), 8ª (2009) y 9ª (2010)⁹. La dilatación de la captación a lo largo de cuatro olas implica que, entre la 7ª y la 8ª Ola, deberá recalcularse el remuestreo para ajustar al máximo los nuevos hogares que hay que captar a las características de la sociedad catalana.

1.5.4. Ponderaciones

Aunque la muestra esté bien diseñada y las pruebas de bondad den resultados aceptables, es de todos conocido que las muestras siempre se desvían del universo a estudiar. Por este motivo, es necesario tener unas variables de ponderación o pesos.

La ponderación es una transformación que permite darle a cada unidad muestral de la encuesta un valor numérico (peso), que es la contribución que hará la unidad a las estimaciones basadas en los datos de la encuesta.

Cada una de las olas del PaD dispone de variables de ponderación transversal, a nivel de hogar y de individuo. Estas variables las genera el Servei d'Estadística Aplicada de la Universitat Autònoma de Barcelona ([SEA-UAB](#)).

A continuación presentamos las características básicas de cada una de estas variables. Para más información, os remitimos al informe metodológico elaborado por el SEA-UAB (en catalán), disponible en la [web del PaD](#).

1.5.4.1. Hogares

En el caso de los hogares, se han construido dos variables de ponderación. La primera reequilibra la muestra en el momento en el que se diseñó, el año 2001, en base a los datos del Censo de población y viviendas 2001 (Instituto Nacional de Estadística, INE). Era la intención del PaD generarla con

⁹ Los hogares captados en 2010 no se incorporaron efectivamente a la muestra hasta el año siguiente, 2011.

el Censo de 2011 para las dos últimas olas pero la disponibilidad de los datos del INE no lo ha permitido. Por esto, la variable reequilibra todas y cada una de las olas en referencia a los datos de 2001.

Esta variable se presenta en dos formatos, el primero de ellos solo reequilibra la muestra (código de la variable XW009001L¹⁰), el segundo reequilibra y eleva el número de hogares a los hogares catalanes (XW009002L).

Las variables que se utilizan para la ponderación de los hogares del PaD en base al Censo 2001 son:

- Sección censal
- Provincia
- Número de personas en el hogar
- Lugar de nacimiento
- Edad
- Situación respecto al trabajo remunerado
- Régimen de tenencia de la vivienda

Sin embargo, interesaba también disponer de una variable de ponderación que tuviera en cuenta la realidad social de cada uno de los años en los que el PaD hizo trabajo de campo. Finalmente, se optó por utilizar una encuesta de ámbito español, con representatividad para Cataluña, periodicidad anual y armonizada a nivel europeo: la Encuesta de Población Activa-EPA. Por tanto, en las matrices se incluye una segunda variable de ponderación (XW011001L) que reequilibra la muestra a la de la EPA del 4º trimestre de cada año.

Las variables que se utilizan para la ponderación de los hogares del PaD en base a la EPA son:

- Provincia
- Número de personas en el hogar
- Lugar de nacimiento
- Edad
- Situación respecto al trabajo remunerado
- Nivel máximo de estudios

1.5.5.2. Individuos

En el caso de los individuos, la base de datos de referencia ha sido el Padrón continuo de cada año. De la misma manera que con los datos del Censo para los hogares, se entregan en el formato que reequilibra la muestra (XW010001E) y en el formato que, además de reequilibrar, la eleva (XW010002E).

Las variables que se utilizan para la ponderación de los individuos del PaD en base al Padrón continuo son:

- Provincia
- Tamaño del municipio
- Sexo
- Edad
- Lugar de nacimiento

¹⁰ La X hace referencia a la letra de la ola, para la 1ª ola es A, 2ª es B i así sucesivamente.

Los cuestionarios

El hecho de que el PaD disponga de una doble unidad de análisis (hogares e individuos) ha hecho necesario desarrollar tres **tipos de cuestionarios**:

- El cuestionario de hogar. Lo responde una única persona (denominada "Informante principal"). Este cuestionario no sólo contiene preguntas generales del hogar sino que también incluye preguntas relativas a las personas menores de 16 años del hogar, a las que no entrevistaremos invididualmente. Esta información, a nivel de matrices, se incorpora a la matriz individual.

Por otro lado, hay la ficha de hogar, que no es propiamente un cuestionario y que permite la gestión de la información sobre altas y bajas de los miembros, datos de contacto, entre otros.

- El cuestionario Individual: es el cuestionario más extenso y lo contestan todas las personas de 16 o más años del hogar.
- El cuestionario de aproximación o Proxy: permite disponer de información muy básica de aquellas personas de 16 o más años del hogar a las cuales no se podrá entrevistar durante el trabajo de campo de la ola, ya sea porque rechazan participar como porque no pueden hacerlo por el motivo que sea. Este cuestionario contiene las preguntas objetivas que pueden ser contestadas por alguna otra persona del hogar, generalmente el informante principal.

Desde el punto de vista de **las preguntas** que se hacen en la encuesta, podemos distinguir diferentes tipos, según la perspectiva que tomemos.

En lo que se refiere al período temporal, el PaD incluye preguntas que recogen información transversal, sobre el momento de la entrevista, y información longitudinal. La información longitudinal se recoge en diferentes modelos temporales:

- Preguntas retrospectivas, recogen eventos sucedidos en algún momento de la vida pasada del individuo.
- Preguntas de cambio desde la última entrevista, recogen los cambios que han podido darse desde la última vez que se entrevistó al hogar o a la persona. En este tipo de preguntas se referencia la fecha de la última entrevista y la información aportada por la persona en aquel momento (por ejemplo: en la última entrevista que le hicimos, el XX de XXX de 20XX, nos dijo que trabajaba a jornada completa, desde entonces hasta ahora: ha trabajado de manera continuada, ha combinado paro con trabajo...).

Es lo que se llama "entrevista dependiente" y provoca un número alto de valores perdidos ya que, si no ha habido cambios (en el ejemplo anterior se completaba la información para los que habían trabajado de manera continuada, preguntando si había habido otros cambios, como de empresa, de tipo de contrato, de responsabilidad...), no se vuelve a preguntar la información que ya tenemos de olas anteriores (en nuestro ejemplo, la actividad de la empresa, el tipo de contrato...).

- Preguntas sobre períodos concretos en el tiempo, recogen información sobre un período ya

pasado, que es el mismo para todas las encuestas. Son, básicamente, las preguntas referidas a ingresos y gastos. En unas de ellas interesa recoger períodos que coinciden con el año natural (es el caso de los ingresos mensuales netos del hogar o del individuo que se preguntan con la referencia del año anterior), en otras, interesa recoger otros períodos temporales diferentes (por ejemplo, los gastos en educación reglada se hacen sobre el curso escolar anterior).

Sobre la periodicidad¹¹ de las preguntas, en el PaD cada pregunta tiene una periodicidad propia. Concretamente:

- Preguntas troncales estructurales. Solo se preguntan una vez porque es información invariable en el tiempo (por ejemplo, la fecha de nacimiento propia o la del padre o la madre). En el momento de la primera entrevista, se incluyen para todos los entrevistados y proxys.
- Preguntas troncales. Aparecen en cada edición de la encuesta por su naturaleza estructural pero variable en el tiempo (por ejemplo, la situación respecto al trabajo remunerado).
- Preguntas cíclicas. Dado su contenido y/o su posible variabilidad en el tiempo, se valora que no es necesario incluirlas en cada edición de la encuesta y hacerlo cada x años (por ejemplo, quien hace las tareas domésticas del hogar).

Todos los cuestionarios del PaD pueden consultarse, en català, siguiendo este [enlace](#).

Las matrices de datos

El presente apartado describe las diferentes matrices de datos del PaD, presenta los elementos básicos y necesarios para utilizarlas y permite familiarizarse con ellas de una manera ágil. Consideramos que ésta es la información básica que se debe tener en cuenta si se quieren explotar los datos de la encuesta PaD.

Tipos de matrices

Las matrices se ofrecen en dos formatos, para facilitar todo tipo de análisis: transversal, de cada ola, y longitudinal, que incluye la información de todas las olas posibles.

Matrices de ola: hay tantas como olas y unidades de análisis. Fruto de las dos formas de recogida de la información en los hogares –del hogar en su conjunto y de cada uno de los individuos que lo integran– se generan dos matrices o ficheros de datos para cada ola del PaD.

- Matriz de hogar: recoge las variables procedentes de la encuesta de hogar. Toman el nombre A_HOGAR para la primera ola (trabajo de campo entre 2001 y 2002), B_HOGAR para segunda (2003) y así sucesivamente hasta la K_HOGAR (2012).
- Matriz individual: recoge las variables de la encuesta individual y la información recogida en el cuestionario de hogar sobre los menores de 16 años. Se denominan A_INDI, B_INDI... K_INDI.

Incluyen la información estructural de la ficha del hogar, la información propia del cuestionario de la ola y, en el caso de las variables armonizadas, la información armonizada disponible.

¹¹ Hay más detalle en el anexo 3: Cuestionario

Matrices longitudinales. Se ha optado por generar dos para cada unidad de análisis. Por un lado, la serie completa de todas las olas del PaD (toman el nombre A_K INDI LONG o A_K LLAR LONG) y, por el otro, la serie completa menos la primera ola (B_K INDI LONG o B_K LLAR LONG). Estas segundas matrices han permitido incluir las variables que recogen el cambio de una ola a la otra y que empezaron a recogerse en la segunda ola (por ejemplo, situación respecto al trabajo remunerado desde la última entrevista).

Las diferencias entre las unas y las otras, es que las A_K contienen todas las olas pero no las variables de cambio (inexistentes en la 1ª ola) mientras que las B_K contienen todas las variables (incluidas las de cambio) pero no todas las olas, ya que empiezan en la 2ª.

Las matrices longitudinales se ofrecen en formato largo (*long*), es decir, los individuos o los hogares se repiten en tantas filas como el número de olas en las que han participado. Hay una variable (P107001E en la matriz individual o P107001L en la matriz de hogar), que informa sobre el año, y en consecuencia la ola, en el que se han recogido los datos.

Estructura de las matrices

Como norma general, todas las matrices –tanto las de ola como las longitudinales– tienen como cabecera las variables identificativas. Si trabajamos con la matriz de hogar, las primeras son: los identificadores de hogar, los identificadores de la persona de referencia y los del informante principal, todas ellas en formato alfanumérico y numérico. En la matriz individual encontramos los identificadores individuales y los identificadores de hogar. A éstas les siguen, en las matrices de hogar, una serie de variables resumen o de recuento. Nos referimos a variables del tipo número de miembros del hogar, tipología del hogar, número de menores de 16 años, entre otras. En el caso de la matriz individual se trata, más bien, de variables de contexto del tipo qué ubicación tiene el miembro en el hogar, su parentesco con la persona de referencia, qué tipo de miembro de la muestra es, etc., seguidas de las variables estructurales de los individuos.

A partir de aquí, en las matrices de ola siguen las variables del cuestionario propiamente dicho. En la matriz longitudinal, siguen las variables que han sido armonizadas y llenadas (en el siguiente apartado se explicita qué es el proceso de armonización y llenado).

Tipos de variables y nomenclatura

Las variables de las matrices de datos siguen un sistema de numeración lógico. Esta nomenclatura permite identificar el bloque temático del cuestionario, la dimensión de hogar / individual de la variable y si la variable es de observación directa o es generada. La suma de todos estos dígitos hace que todas las variables tengan una extensión mínima de 9 dígitos. En el anexo 3 se incluye un cuadro explicativo de la lógica de todos y cada uno de los dígitos del nombre de una variable. Para lo que interesa en el análisis de los datos, queremos destacar la lógica del comienzo y del final del nombre de la variable.

Sobre el **inicio del nombre de las variables**, podemos encontrar:

- Las variables propias de una ola, que toman como primer dígito la letra de la ola. En este

sentido, en una matriz de ola (por ejemplo, la 5ª) las variables tienen el primer dígito correspondiente (en el ejemplo, la E).

- Las variables troncales estructurales (con el identificador individual o el lugar de nacimiento), que por su carácter de invariabilidad en el tiempo, no tienen este primer dígito y empiezan por la letra que identifica el bloque temático correspondiente. Pueden encontrarse en todas las matrices, en las de ola para los individuos presentes en esa ola y en las longitudinales para todos los individuos.
- Las variables longitudinales puras, que se han generado para permitir específicamente el análisis longitudinal, tampoco tienen letra de ola. Son variables generadas con la información disponible, tanto transversal como retrospectiva, disponible durante las once olas del PaD. Son básicamente las referentes a la trayectoria educativa, aunque se dispone también de otras sobre trabajo productivo o datos de los padres. Estas variables pueden encontrarse en las matrices longitudinales pero no en las de ola.

Sobre el final del nombre de las variables, es decir las **terminaciones**, los tipos son:

- M1, M2... proceden de la partición de una multirrespuesta. Pueden ser categóricas o dicotómicas. Generalmente, el criterio para hacer una partición u otra ha sido en número de variables a generar. Se han dicotomizado las que contienen un máximo de 15 respuestas posibles (es decir, las que generan un máximo de 15 variables) y se han categorizado las que obligan a hacer más de 15 variables. Como consecuencia de esta decisión, la mayoría de multirrespuestas son dicotómicas;
- G1, G2... que se generan a partir de dos o más variables o no van aparecieron en el cuestionario de la ola en cuestión. En el proceso de depuración longitudinal, en algunos casos, ha sido posible generar una variable no existente en la ola gracias a la información existente en las olas anteriores o posteriores;
- R1, R2, R3... que son fruto del cierre de una pregunta abierta, de una recodificación de otra variable o de una agrupación de la variable original;
- I que informa sobre si el valor de la variable original ha sido imputado; y
- H1, H2, que se aplica a aquellas variables que se han armonizado y depurado longitudinalmente. El proceso de armonización supone homogeneizar los códigos de respuesta de todas las olas, para que sean los mismos y se pueda generar una única variable longitudinal. Paralelamente, es necesario el proceso de llenado. Tal y como se ha explicado anteriormente el hecho de utilizar la "encuesta dependiente", para minimizar el tiempo de encuestación y evitar el cansancio (y rechazo a seguir participando) por parte de los encuestados, genera muchos valores perdidos (de la forma -8 "Filtrado por base"). El llenado sirve para, en el caso de que no haya habido cambios, llenar la información de la que disponemos en olas anteriores. De esta manera, con la armonización y el llenado, se consigue homogeneizar los códigos de respuesta de las variables para todas las olas y disponer del máximo número posible de observaciones para cada uno de los individuos. Este tipo de variables son el contenido principal de las matrices longitudinales.

Además, estos finales de nombre de variable pueden encontrarse combinados (por ejemplo, H1M1)

cuando han pasado por dos de estos procesos (en el ejemplo, una variable armonizada y posteriormente partida como multirrespuesta).

Valores perdidos

Para poder reflejar la variedad de situaciones en las que una pregunta queda sin responder, se han adoptado una serie de códigos específicos, todos ellos con valores negativos. Sin embargo, las matrices se entregan con la definición como valor perdido solo del valor -8, correspondiente a los filtrados por base o no pertinentes.

Los códigos posibles de ausencia de respuesta que pueden encontrarse en las matrices son:

- - **1** No sabe
- - **3** No imputable. Corresponde a unos pocos casos de algunas olas en las que no ha sido posible imputar los ingresos.
- - **9** No contesta
- - **8** Filtrado por base/no pertinente. Corresponde a los casos en que la persona, o el hogar, no tiene que responder a la pregunta. Se encuentra desde el caso más sencillo, en el que, para una pregunta filtro (régimen de tenencia del hogar) no tiene que responderse una pregunta posterior (año de compra de la vivienda por personas que viven de alquiler), hasta el caso más complejo, que tiene que ver con la base de la pregunta. En el proceso de elaboración del cuestionario se definen las rutas que deben seguir las personas, y los hogares, en función de sus características personales o familiares. Esto genera lo que llamamos "la base de la pregunta", es decir, quién debe responderla, a partir de un complejo juego de filtros. Por ejemplo, el nivel máximo de estudios se pide a las personas a las que se encuesta por primera vez y a aquellas personas que ya fueron encuestadas en la última ola, declararon que estaban estudiando y ahora no estudian.

Los identificadores de individuo y de hogar

Son las variables clave que permiten vincular la información de los hogares y de los individuos o la información de diferentes olas. Para poner en relación la información de hogar y la individual es necesario fusionar los ficheros a partir del identificador de hogar. Los identificadores de los hogares y de los individuos se mantienen en el tiempo, es decir, son invariables.

Los identificadores originales, que adoptan forma alfanumérica, se asignan automáticamente en el transcurso del trabajo de campo para todos y cada uno de los hogares y de los individuos de la muestra. Las variables que recogen el código de hogar son la P017001L (en la forma alfanumérica) y la P017001LG2 (la traducción numérica), y las que recogen el individual son la P008001E y la P008001EG2. La traducción numérica del identificador se genera para permitir la fusión de los diferentes ficheros con cualquier programa informático de análisis de datos. Se recomienda su uso para evitar problemáticas derivadas de la naturaleza del identificador original. Podéis encontrar información detallada de la traducción del identificador alfanumérico al numérico en el anexo 1.

Los identificadores están dotados de una nomenclatura propia que proporciona información sobre el individuo y sus hogares. La lógica de los identificadores de hogar es la siguiente:

- la primera letra (o los dos primeros números) indica la ola en la que el hogar ha entrado a formar parte de la encuesta o se ha creado. Así, las que empiezan por A entraron en la muestra del PaD durante la 1ª ola (2001-2002), las que empiezan por B lo hicieron en la 2ª ola (2003) y así sucesivamente.
- La segunda letra indica, en los casos que corresponde, si el hogar procede de otro hogar que ya formaba parte del panel. Por ejemplo, un hogar EA es un hogar que se creó en la 5ª ola (2006) y que provenía de un hogar de 1ª ola. En este sentido, los hogares captados con posterioridad a la 1ª ola tienen las dos primeras letras coincidentes con las de la ola en cuestión (un hogar captado en la 6ª ola empezará por FF).
- Si de un hogar original, en una ola concreta, se ha creado más de un hogar nuevo, estos se distinguen con una letra pequeña al final del identificador (a, b, c, etc.; por ejemplo, DA0267a y DA0267b para dos hogares que se han creado en la 4ª ola y que provenían de un hogar de la muestra inicial). Se ha previsto que alguno de estos hogares también pudiera generar más de un hogar en una misma ola, por lo que podría haber algún hogar con dos letras pequeñas al final (la primera sería la correspondiente al hogar original y la segunda, al orden de creación del nuevo hogar).

Sin embargo, en las matrices de datos, debe tenerse en cuenta que la ola de creación de un hogar y que genera un determinado identificador de hogar, no quiere decir que se pudiera encuestar en esa ola. Es decir, que es posible que un hogar se generase, a nivel de trabajo de campo y de generación del nuevo identificador, en una ola pero que no pudiera ser efectivamente encuestado hasta una ola posterior. Por eso, se han generado las variables P040001LG1 para hogares y la P040001EG2 para individuos (variables sin letra de ola) para especificar en qué ola hay la primera entrevista efectiva, o la alta a nivel de disponibilidad de datos.

El identificador individual se crea añadiendo al de hogar la posición del individuo dentro del hogar (dos dígitos). Eso quiere decir que un individuo que se haya incorporado a la muestra en la 6ª ola (porque se ha casado con un miembro original de la muestra, por ejemplo) en un hogar AA tendrá un identificador que empezará por AA.

Los materiales de apoyo

Con el objetivo de facilitar el acceso y el uso de los datos se han desarrollado una serie de materiales de apoyo. Los dos documentos básicos de consulta son los que se detallan a continuación.

El informe longitudinal de contenidos

El informe longitudinal de contenidos se ha elaborado con el objetivo de aproximar a los usuarios los contenidos de los distintos bloques temáticos de los cuestionarios del PaD. Permite saber:

- qué variables han aparecido en las diferentes olas del PaD para cada uno de los ámbitos temáticos del cuestionario y
- identificar la correspondencia entre el código de la variable del cuestionario y el código de la variable en la matriz de datos. Esta información es básica, ya que en muchos casos la nomenclatura de las variables durante la recogida de los datos se modifica posteriormente durante la depuración y tratamiento de los datos para las matrices. También porque algunas

variables generadas no tienen una correspondencia directa con las variables de cuestionario, porque integran la información de varias de ellas.

El informe contiene todas las variables del PaD menos las que sólo aparecieron en la primera ola.

Este documento proporciona la siguiente información de las variables:

- la etiqueta;
- el código de la variable en el cuestionario;
- el código de la variable en la matriz;
- la periodicidad;
- las olas en las que ha aparecido la variable;
- si la variable ha sido eliminada o reabsorbida por otra u otras variables.

Los libros de códigos

Una vez efectuada la elección sobre los ámbitos de análisis y las variables de interés es necesario un documento más específico sobre el contenido concreto de cada matriz de datos. Los libros de códigos han sido diseñados para responder a esa necesidad y, en este sentido, son una herramienta indispensable para la explotación del PaD. Cada matriz de datos dispone de su libro de códigos propio.

La estructura de los libros de códigos es la siguiente:

Encabezamiento: en el que se especifica cuántos registros (individuos u hogares) contiene además de información básica para el tratamiento de los datos.

Cuerpo central: muestra las características de cada una de las variables en el mismo orden en el que aparecen en la matriz de datos. La información que se facilita es la siguiente:

- el código de la variable;
- la etiqueta;
- la base, es decir, qué individuos u hogares disponen de la información;
- los códigos de respuesta;
- el número de casos disponible;
- el número de no sabe o no contesta que incluye la n.

Documentos metodológicos

También disponemos de una serie de documentos metodológicos que pueden ser de interés para los usuarios, y que recogen el método empleado en diferentes procesos de preparación de los datos. Concretamente:

- Informe metodológico sobre la generación de las variables de ponderación.
- Informe metodológico sobre las variables de imputación de ingresos.
- Informe metodológico sobre la generación de las variables de trayectorias educativas.

Estos documentos son accesibles en la [web del PaD](#), en catalán. Queremos recordar, también, que en la misma web están referenciadas todas las investigaciones que se han hecho con datos del PaD, en este [enlace](#).

Anexos

Anexo 1: tablas de referencia de los gráficos números 1 y 2 sobre evolución de la muestra

Hogares

	Captación 1ª ola	Pérdida interanual	Pérdida acumulada	Captación 6ª ola	Pérdida interanual	Pérdida acumulada
Ola A (2001/02)	1.991					
Ola B (2003)	1.703	14,47%	14,47%			
Ola C (2004)	1.528	10,28%	23,25%			
Ola D (2005)	1.404	8,12%	29,48%			
Ola E (2006)	1.090	22,36%	45,25%			
Ola F (2007)	903	17,16%	54,65%	334		
Ola G (2008)	820	9,19%	58,81%	273	18,26%	18,26%
Ola H (2009)	756	7,80%	62,03%	233	14,65%	30,24%
Ola I (2010)	719	4,89%	63,89%	217	6,87%	35,03%
Ola J (2011)	658	8,48%	66,95%	199	8,29%	40,42%
Ola K (2012)	614	6,69%	69,16%	181	9,05%	45,81%

	Captación 7a ola	Pérdida interanual	Pérdida acumulada	Captación 8a ola	Pérdida interanual	Pérdida acumulada	Total individuos	Pérdida acumulada
Ola A (2001/02)							1.991	
Ola B (2003)							1.703	14,47%
Ola C (2004)							1.528	23,25%
Ola D (2005)							1.404	29,48%
Ola E (2006)							1.090	45,25%
Ola F (2007)							1.237	37,87%
Ola G (2008)	431						1.524	23,46%
Ola H (2009)	363	15,78%	15,78%	154			1.506	24,36%
Ola I (2010)	334	7,99%	22,51%	125	18,83%	18,83%	1.395	29,93%
Ola J (2011)	300	10,18%	30,39%	103	17,60%	33,12%	1.260	36,72%
Ola K (2012)	260	13,33%	39,68%	88	14,56%	42,86%	1.143	42,59%

Individuos

	Captación 1ª ola	Pérdida interanual	Pérdida acumulada	Captación 6ª ola	Pérdida interanual	Pérdida acumulada
Ola A (2001/02)	4.587					
Ola B (2003)	3.686	19,64%	19,64%			
Ola C (2004)	3.246	11,94%	29,23%			
Ola D (2005)	2.954	9,00%	35,60%			
Ola E (2006)	2.204	25,39%	51,95%			
Ola F (2007)	1.787	18,92%	61,04%	745		
Ola G (2008)	1.599	10,52%	65,14%	576	22,68%	22,68%
Ola H (2009)	1.451	9,26%	68,37%	486	15,63%	34,77%
Ola I (2010)	1.365	5,93%	70,24%	445	8,44%	40,27%
Ola J (2011)	1.245	8,79%	72,86%	382	14,16%	48,72%
Ola K (2012)	1.130	9,24%	75,37%	340	10,99%	54,36%

	Captación 7a ola	Pérdida interanual	Pérdida acumulada	Captación 8a ola	Pérdida interanual	Pérdida acumulada	Total individuos	Pérdida acumulada
Ola A (2001/02)							4.587	
Ola B (2003)							3.686	19,64%
Ola C (2004)							3.246	29,23%
Ola D (2005)							2.954	35,60%
Ola E (2006)							2.204	51,95%
Ola F (2007)							2.532	44,80%
Ola G (2008)	957						3.132	31,72%
Ola H (2009)	772	19,33%	19,33%	335			3.044	33,64%
Ola I (2010)	683	11,53%	28,63%	264	21,19%	21,19%	2.757	39,90%
Ola J (2011)	594	13,03%	37,93%	203	23,11%	39,40%	2.424	47,16%
Ola K (2012)	495	16,67%	48,28%	168	17,24%	49,85%	2.133	53,50%

NOTA: Sólo se tienen en cuenta los individuos entrevistados o proxys

Anexo 2: construcción de los identificadores numéricos

El identificador numérico se construye de la siguiente manera:

- Dos dígitos para la primera letra y dos dígitos para la segunda. El primer dígito, tanto de una letra como de la otra, hace referencia a la década del panel y el segundo al año. De esta manera, aseguramos que el código no empieza por cero y que se reconocerá bien por cualquier programa informático, además de prever los códigos que se generarán más allá de la 10ª ola.
- El número del hogar.
- Un dígito para la primera letra pequeña.
- Un dígito para la segunda letra pequeña.
- Dos dígitos para la posición del miembro, en el caso de los identificadores individuales.

De esta manera, el hogar AA0154 se traducirá en 1111015400; el hogar BA1854a en 1211185410 y el hogar FA2342ab en 1611234212. En el caso de los identificadores individuales, a estos códigos solo habrá que añadirles la posición del miembro: por ejemplo, AA015401 o 111101540001.

EJEMPLOS DE IDENTIFICADORES Y EQUIVALENCIA ENTRE LOS ALFANUMÉRICOS Y LOS NUMÉRICOS

Ejemplo	Tipo id*	Ola de incorporación del hogar	Ola de creación del hogar original	N.º del hogar	Indicativo de que de un hogar original se han derivado dos o más	Código hogar resultante	Ubicación que ocupa el individuo en el hogar	Código individual resultante (dos dígitos más)
1	O	A	A	0012	-	AA0012	02	AA001202
	G	11	11	0012	00	1111001200	02	111100120002
2	O	B	A	0365	-	BA0365	05	BA036505
	G	12	11	0365	00	1211036500	05	121103650005
3	O	D	A	0128	a	DA0128a	07	DA0128a07
	G	14	11	0128	10	1411012810	07	141101281007
4	O	F	C	1642	b	FC1642b	04	FC1642b04
	G	16	13	1642	20	1613164220	04	161316422004

* ID=identificador; ** O=original (alfanumérico); *** G=generado (numérico)
Fuente: elaboración propia

Anexo 3: Ficha técnica del cuestionario

Principi metodològic	<p>Los cuestionarios del PaD se trabajan en base a los siguientes criterios metodològicos:</p> <ul style="list-style-type: none"> ▪ Simplicidad: para minimizar los cambios a introducir en el cuestionario. ▪ Objektivaci3n y fiabilidad: para dar cuenta de los motivos finales que han motivado el cambio y la elecci3n de una determinada acci3n sobre la variable. ▪ Relevancia: para asegurar el uso potencial de la informaci3n que se recogerá. ▪ Documentaci3n: para garantizar la acreditaci3n de las modificaciones adoptadas.
Unidades de anàlisis	<p>Hogar y Individuo</p>
Tipos de Cuestionarios	<p>Cuestionario Hogar: lo contesta el informante principal (una sola persona por hogar).</p> <p>Cuestionario Individual: lo contestan todos los miembros del hogar que tienen más de 15 años.</p> <p>Cuestionario Proxy: pensado para las personas a las que no se podrá entrevistar porque durante el trabajo de campo no estàn en el hogar; no son aptas para responder; etc. El cuestionario proxy se administra a una persona dispuesta a facilitar datos básicos (preguntas objetivas) de la persona a la que no podremos entrevistar.</p>
Contenido	<p>Datos identificativos y de gesti3n; Educaci3n; Trabajo productivo, reproductivo y conciliaci3n; Usos del tiempo ; Ingresos; Vivienda; Salud y dependencia; Situaci3n personal y familiar; Posici3n y relaciones sociales; Valores y actitudes.</p>
Tipos de preguntas	<p>Combinaci3n perspectiva objetiva y subjetiva. Combinaci3n de preguntas troncales (encuesta longitudinal) y cíclicas (aparecen cada x años).</p>
Calendario de trabajo	<p>Inicio elaboraci3n cuestionarios: marzo Cierre cuestionarios de la ola: diciembre Prueba piloto: enero-febrero</p>
Actividades principales	<ul style="list-style-type: none"> ▪ Discusi3n de los contenidos ▪ Documentaci3n de los cambios de las variables ▪ Informatizaci3n de los cuestionarios ▪ Testeo del cuestionario (prueba piloto) ▪ Control de calidad de las entrevistas
Outputs	<ul style="list-style-type: none"> ▪ Cuestionarios por ola (Hogar, Individuo, Proxy) ▪ Informe longitudinal de contenidos (General y por bloques) ▪ Informe longitudinal técnic

Periodicidad de las preguntas del PaD

La **periodicidad** diferenciada de cada una de las preguntas del PaD implica que en cada ola los cuestionarios sean diferentes.

TRONCAL	<p>Preguntas que aparecen en cada ola desde el momento de su inclusión en el cuestionario. Sirven para analizar el presente de los hogares y los individuos y están enfocadas al análisis del microcambio año tras año.</p> <p>Por ejemplo: ¿Está cursando estudios reglados actualmente? ¿Cuáles? ¿Cuál es su situación respecto al trabajo remunerado?</p>
TRONCAL ESTRUCTURAL	<p>Las preguntas troncales estructurales son aquellas que recogen información con variación nula en la vida del encuestado y que, por tanto, sólo se recogen una vez en toda la vida del panel. Éste es el caso de variables como el año de nacimiento, el sexo, la edad de inicio en la vida laboral, el lugar y año de nacimiento del padre o la madre...</p> <p>Sin embargo, para poder contrastar datos y asegurar la consistencia de la información, hay preguntas troncales estructurales que se han incluido en el cuestionario en más de una ocasión.</p>
CÍCLICA	<p>Preguntas que aparecen de forma cíclica, ya que se estima que tiene sentido plantearse el análisis del cambio sólo cada X años.</p> <p>Por ejemplo: ¿Quién organiza y coordina el trabajo doméstico y familiar? ¿Qué cosas mejoraría de su vivienda? ¿Cuáles de los siguientes familiares viven cerca de usted?</p>
ESPORÁDICA	<p>Preguntas que aparecen de forma esporádica en alguna de las olas del panel y para las que no se concreta fecha de nueva inclusión. Su carácter esporádico implica que no se espera hacer ningún análisis en términos de estudio del cambio.</p> <p>Por ejemplo: Indique su grado de acuerdo o desacuerdo con las siguientes afirmaciones: en Catalunya hay demasiados inmigrantes.</p>

Descripción de la numeración lógica de las variables

DÍGITO Matrices de ola	DÍGITO Matriz long	NOMBRE DEL DÍGITO	CONTENIDOS
1		OLA DE LA VARIABLE	Cada ola se llama conforme a la letra del abecedario que toca, por orden. De esta manera todas las variables de los ficheros de la 1ª ola empiezan por la letra A, las de la 2ª ola por la letra B i así sucesivamente hasta la 11ª ola, que empieza por K.
2	1	BLOQUE DEL CUESTIONARIO	La correspondencia entre el nombre del bloque del cuestionario y la letra es la siguiente: P , portada; L , hogar; E , educación; F , fecundidad; H , vivienda; G , lengua; R , posición y relaciones sociales; S , salud; T , trabajo; D , trabajo doméstico; I , posición económica; U , tiempo de ocio; M , medios de comunicación; C , cultura política; O , inmigración; K , datos de seguimiento; W , ponderación.
3, 4, 5	2, 3, 4	DÍGITOS PARA NUMERAR LA VARIABLE	A medida que se crea una nueva variable se le asigna el número que toca, correlativamente.
6, 7, 8	5, 6, 7	CANVIO DE UNIDADES O POSIBILIDADES DE RESPUESTA QUE TIENE LA VARIABLE	En algunas preguntas y para conseguir el máximo número posible de respuestas, se da la opción de responder en unidades diferentes (por ejemplo, euros/mes y euros/año). Estas variables mantienen todos los dígitos de su nombre lógico a excepción del dígito 7, que sirve para distinguir la unidad en la que se recoge la respuesta.
9	8	TIPO CUESTIONARIO	Para diferenciar el cuestionario al que pertenece la variable: E : para las variables de los cuestionarios individual y proxy; L : para las variables de los cuestionarios de hogar. Debemos recordar la excepcionalidad de las variables que contienen información individual pero que se recogen en el cuestionario de hogar y que, en las matrices, cambian la letra final de L a E.
>9	>8	IDENTIFICADOR DEL TIPO DE VARIABLE	Además, se añaden las siguientes letras y, eventualmente, un número a las variables en los siguientes casos: - M1, M2... , provienen de la partición de una multirrespuesta; - G1 , son generadas a partir de dos o más variables; - R1, R2, R3... son fruto de una recodificación o de una agrupación de la variable original; - I , informa, dicotómicamente, sobre si el valor proviene de la variable original o ha sido imputado; - H1 , se aplica a aquellas variables que se han armonizado y llenado longitudinalmente. Pueden encontrarse combinados (por ejemplo H1M1) cuando han pasado por dos de estos procesos (en el ejemplo, una variable armonizada y posteriormente partida como multirrespuesta).

Anexo 4: Muestra

Principios metodológicos	<p>Encuesta Panel puro: seguimiento a lo largo del tiempo de una misma muestra de hogares e individuos.</p> <p>Remuestreos periódicos para garantizar el techo muestral de 2.000 hogares y 4.600 individuos encuestados.</p>																												
Universo	Hogares privados de Cataluña y los individuos que los integran.																												
Muestra original (1ª ola)	<p>→ Muestra teórica: 2.000 hogares de Cataluña; 5.456 individuos (2,7 individuos/hogar de media; 4.600 de los cuales encuestados=> mayores 15 años)</p> <p>→ Muestra captada (fin 1ª ola): 1.991 hogares; 5.757 individuos, 4.589 encuestados.</p>																												
Tipo de muestreo	<p>1ª etapa: Distribución a proporción a nivel de provincia y proporcional a nivel de tamaño de municipio, del número de hogares a entrevistar.</p> <p>2ª etapa: Selección aleatoria sistemática y polietápica de las primeras unidades muestrales (secciones censales) según variables de estratificación territorial y social:</p> <ul style="list-style-type: none"> ▪ area geográfica ▪ tamaño del municipio ▪ estructura social (=>situación ocupacional) <p>3ª etapa: Selección aleatoria simple de las unidades muestrales finales => domicilios de los hogares a entrevistar en las primeras unidades muestrales.</p>																												
Remuestreos	<p>Los remuestreos se realizan siguiendo los mismos criterios de aleatoriedad y segmentación de la muestra de la 1ª ola.</p> <table border="1" data-bbox="416 1301 1291 1608"> <thead> <tr> <th></th> <th>Nuevos hogares captados</th> <th>Hogares de seguimiento</th> <th>Total hogares muestra PaD</th> </tr> </thead> <tbody> <tr> <td>Resultados 1º Remuestreo</td> <td></td> <td></td> <td></td> </tr> <tr> <td>6ª Ola (2007):</td> <td>333</td> <td>1154</td> <td>1487</td> </tr> <tr> <td>7ª Ola (2008):</td> <td>431</td> <td>1446</td> <td>1877</td> </tr> <tr> <td>Resultados 2º Remuestreo</td> <td></td> <td></td> <td></td> </tr> <tr> <td>8ª Ola (2009):</td> <td>154</td> <td>1745</td> <td>1899</td> </tr> <tr> <td>10ª Ola (2011):</td> <td>154</td> <td>1784</td> <td>1938</td> </tr> </tbody> </table>		Nuevos hogares captados	Hogares de seguimiento	Total hogares muestra PaD	Resultados 1º Remuestreo				6ª Ola (2007):	333	1154	1487	7ª Ola (2008):	431	1446	1877	Resultados 2º Remuestreo				8ª Ola (2009):	154	1745	1899	10ª Ola (2011):	154	1784	1938
	Nuevos hogares captados	Hogares de seguimiento	Total hogares muestra PaD																										
Resultados 1º Remuestreo																													
6ª Ola (2007):	333	1154	1487																										
7ª Ola (2008):	431	1446	1877																										
Resultados 2º Remuestreo																													
8ª Ola (2009):	154	1745	1899																										
10ª Ola (2011):	154	1784	1938																										

Detalles técnicos sobre la muestra teórica

Universo

Hogares privados de Cataluña

Muestra teórica y márgenes de error

Hogares

ÁREAS	NÚMERO DE HOGARES	e* (%)
Barcelona	800	±3,54
Ciudad BCN	261	±6,12
Región Metropolitana	470	±4,61
Comarcas centrales	69	±12,04
Girona	400	±5,00
Lleida	400	±5,00
Tarragona	400	±5,00
CATALUÑA	2.000	±2,79

*e = error muestral (intervalo confianza=95,5%, k=2 i p=q=50)
Fuente: elaboración propia

Individuos

ÁREAS	NÚMERO DE INDIVIDUOS	e* (%)
Barcelona	2.178	±2,14
Girona	1.054	±3,08
Lleida	1.120	±2,98
Tarragona	1.104	±3,01
CATALUÑA	5.456	±1,67

*e = error muestral (intervalo confianza=95,5%, k=2 i p=q=50)
Fuente: elaboración propia

Anexo 5: Trabajo de campo

Ficha técnica del trabajo de campo

Principio metodológico	El principio metodológico que rige el trabajo de campo, es desarrollar un trabajo personalizado y de máxima proximidad con el equipo de encuestadores y los hogares colaboradores, para poder optimizar la calidad de los datos recogidos y la fidelización al proyecto. En el contexto de una encuesta longitudinal, la continuidad tanto de los hogares, así como de los encuestadores, es un aspecto clave.
Periodicidad de la encuesta	Anual
Calendario de realización de encuestas	<ul style="list-style-type: none"> ▪ 1ª etapa: marzo-julio: 80% del trabajo de campo ▪ 2ª etapa*: septiembre –octubre: 20% del trabajo de campo *etapa destinada a la recuperación de hogares que en la 1ª etapa no se pudieron encuestar.
Modo de administración de la encuesta	Presencial (1ª a 8ª olas). Método Mixto: Telefónico-Presencial (9ª ola en adelante): <ul style="list-style-type: none"> ▪ Muestra encuestada telefónicamente: 80%. ▪ Muestra encuestada presencialmente: 20% (incluye hogares de nueva incorporación a la encuesta).
Universo	Personas aptas mayores de 15 años considerades del hogar
Sistema de administración de la encuesta	CAPI (<i>Computer Assisted Personal Interview</i> => Entrevista asistida por ordenador)
Equipo de encuestadores/as	Formación en ciencias sociales, con alta motivación para participar en la encuesta => 80 % continuidad encuestadores (media 1ª a 8ª olas PaD)
Participación de los hogares	Voluntaria
Incentivo participación	Obsequio para el hogar
Controles de calidad de la información	<ul style="list-style-type: none"> ▪ COE: Control de calidad de la realización de encuestas (mide la calidad de la información recogida)→ <i>ver apartado Cuestionario</i> ▪ CT: Control telefónico a los hogares (comprueba la realización de la entrevista): 15% de los hogares encuestados
Principales actividades	<ul style="list-style-type: none"> ▪ Bolsa On-Line de encuestadores (oferta permanente). ▪ Identificación de los hogares a entrevistar. ▪ Gestión del trabajo de campo: encuestadores/as y hogares: formación personalizada y seguimiento via reuniones de coordinación. ▪ Ampliación de la muestra de hogares (captación). ▪ Cierre de la ola de trabajo de campo. ▪ Preparación de los datos de seguimiento para la ola siguiente.

Muestra de trabajo de campo (incluye todos los hogares trabajados en cada ola)

	HOGARES INICIO OLA*	HOGARES TRABAJADOS**	HOGARES ENTREVISTADOS***
1ª ola 2001/02	0	1991	1991
2ª ola 2003	1991	2149	1777
3ª ola 2004	1931	2044	1693
4ª ola 2005	1791	1880	1633
5ª ola 2006	1724	1791	1323
6ª ola 2007	1473	1862	1488
7ª ola 2008	1749	2268	1877
8ª ola 2009	2089	2332	1899
9ª ola 2010	2114	2219	1940
10ª ola 2011	2043	2266	1938
11ª ola 2012	2069	2127	1792

*NÚMERO DE HOGARES INICIO OLA. El número de hogares al inició de la ola incluye: hogares entrevistados en la ola anterior + hogares no entrevistados en la ola anterior pero que relanzamos al campo (rechazo o pérdida no definitivos).

** NÚMERO DE HOGARES TRABAJADOS EN LA OLA. El número de hogares trabajados incluye los hogares del inicio de la ola (hogares entrevistados en la ola anterior + hogares rechazo o pérdida no definitivos de la ola anterior) + los hogares que se crean durante la ola (por la propia dinámica de la muestra PaD: separaciones, individuos que se independizan,...). En años de captación (6ª, 7ª, 8ª i 10ª) también se incluyen los hogares captados durante la ola.

*** NÚMERO DE HOGARES ENTREVISTADOS. El número de hogares entrevistados incluye los hogares trabajados durante el trabajo de campo que se ha conseguido entrevistar.

Porcentaje acumulado de entrevistas (por mes de trabajo de campo)

	2a Ola 2003	3a Ola 2004	4a Ola 2005	5a Ola 2006	6a Ola 2007	7a Ola 2008	8a Ola 2009	9a Ola 2010	10a Ola 2010	11a Ola 2010
Enero	5,3	11,3	0,1		3,4	1,8	0,4			
Febrero	29,6	40,8	28,3	24,7	21,0	23,8	15,4			0,7
Marzo	55,0	66,2	57,0	56,7	42,0	41,8	33,8		7,8	6,8
Abril	72,5	79,0	77,6	68,4	55,4	54,8	50,5	26,3	37,9	35,3
Mayo	83,3	90,4	87,8	81,1	65,2	65,3	68,6	68,3	70,6	66,2
Junio	92,2	96,0	93,1	91,6	72,8	77,6	81,4	91,0	86,8	87,6
Julio	96,5	96,9	95,0	98,4	80,4	87,6	89,0	93,8	89,8	92,1
Agosto	96,5	96,9	95,0	98,4	81,4	88,4	89,5	93,8	91,5	95,0
Septiem- bre	96,5	96,9	95,0	99,5	84,7	92,2	93,2	93,9	95,1	97,5
Octubre	99,0	99,1	98,8	100,0	98,7	98,5	98,5	98,6	99,4	100,0
Noviembre	100,0	100,0	100,0		100,0	100,0	100,0	100,0	100,0	

Anexo 6: Datos

Ficha técnica de depuración y preparación de las matrices de datos

<p>Principio metodológico</p>	<p>El principio básico que rige el trabajo de depuración de los datos del PaD es no modificar los datos facilitados por las personas encuestadas si no es que hay un error evidente que se pueda corregir. Si una persona encuestada ofrece respuestas inconsistentes y no hay ninguna evidencia que permita hacer algún tipo de corrección, se mantiene la inconsistencia y es trabajo de cada analista decidir qué hace con ella.</p>
<p>Programas tratamiento datos</p>	<p>SPSS Microsoft Excel Barbwin</p>
<p>Calendario de depuración (una Ola)</p>	<p>Inicio depuración: octubre/noviembre (coincidiendo con el final del trabajo de campo).</p> <p>Finalización de la depuración: diciembre (año siguiente).</p> <p>El calendario está sujeto a variaciones en función de las características de la ola y de la disponibilidad de recursos humanos.</p>
<p>Principales actividades</p>	<ul style="list-style-type: none"> ▪ Consolidación de los datos. ▪ Preparación de las matrices individuales y de hogar para la depuración de las variables. ▪ Codificación de las preguntas abiertas del cuestionario. ▪ Depuración por bloques de las variables individuales y de hogar. ▪ Encargo y seguimiento de la imputación de las variables de ingresos (individuales y de hogar). ▪ Control de las bases de las variables de las matrices individuales y de hogar. ▪ Comprobación final del contenido de todas las variables a través de las frecuencias (tabulación). ▪ Controles de consistencia de las variables estructurales con otras fuentes estadísticas (para valorar la consistencia de los datos del PaD). ▪ Cálculo de las variables generadas del PaD (variables creadas por el equipo del PaD). ▪ Encargo y seguimiento de la generación de las variables de ponderación. ▪ Elaboración de los materiales de soporte.
<p>Outputs</p>	<p>Matrices de datos</p> <ul style="list-style-type: none"> • Transversales: X_IND Y X_LLar (una por cada ola) • Longitudinales: A_K INDI_LONG, B_K INDI LONG i A_K LLAR_LONG, B_K LLAR LONG <p>Libros de Códigos:</p> <ul style="list-style-type: none"> • Un libro de códigos por cada ola del PaD. • Un libro de códigos por cada una de las matrices longitudinales <p>Manual de uso de las matrices.</p>

Número de variables en las matrices

	Llar (hogar)	Indi
1a ola 2001/02	236	236
2a ola 2003	179	418
3a ola 2004	229	863
4a ola 2005	205	432
5a ola 2006	223	569
6a ola 2007	226	487
7a ola 2008	275	936
8a ola 2009	246	518
9a ola 2010	60	145
10a ola 2011	148	344
11a ola 2012	148	743
A_K LONG	42	42
B_K LONG	52	239

Anexo 7: Ficha técnica del análisis

<p>Principio metodológico</p>	<p>El principio básico que rige el análisis de datos del PaD es poner a disposición de la comunidad investigadora todos los datos del PaD de forma gratuita. La FJB impulsa el análisis de los datos (producción interna), al mismo tiempo que apoya y hace un seguimiento del uso externo de los datos (producción externa).</p>
<p>Calendario de trabajo</p>	<p>Permanente (no sujeto a fechas)</p>
<p>Principales actividades</p>	<ul style="list-style-type: none"> ▪ Atención a las demandas de información de instituciones que se interesan por los datos del PaD ▪ Cesión de los datos del PaD ▪ Encargos internos de análisis ▪ Seguimiento del uso externo de los datos ▪ Gestión grupo de usuarios del PaD ▪ Publicación de informes del PaD
<p>Outputs</p>	<p>Publicaciones:</p> <ul style="list-style-type: none"> ▪ Colección "Polítiques" ▪ Colección "Informes Breus" ▪ Documentos metodológicos (on-line) ▪ Informes de resultados (on-line) <div data-bbox="1002 938 1294 1375" style="text-align: right;"> </div> <p>Ámbitos de uso de los datos: Universidades, Administración Pública, Agentes Sociales, Medios de comunicación de Cataluña, España, Europa.</p> <p>Producción compartida: Algunos de los informes resultantes se han publicado de forma conjunta con otras instituciones.</p>

Anexo 8: Ficha técnica de difusión

<p>Principio metodológico</p>	<p>El principio básico que rige la difusión de los datos del PaD es llegar a diferentes públicos: el académico, los agentes sociales y políticos responsables del diseño e implementación de las políticas públicas y la sociedad en general.</p> <p>También se tiene una especial atención a la difusión de resultados entre los hogares colaboradores del PaD, y entre los encuestadores, como muestra de agradecimiento y como demostración de la utilidad de su participación en la encuesta.</p>
<p>Calendario de trabajo</p>	<p>Difusión en el ámbito académico y entre los agentes políticos y sociales:</p> <ul style="list-style-type: none"> ▪ Permanente (sujeto a la disponibilidad de los resultados de los estudios) <p>Difusión en el ámbito de los hogares del PaD, los encuestadores y la sociedad en general</p> <ul style="list-style-type: none"> ▪ Trípticos (inicio- mitad y final del trabajo de campo) ▪ Dossiers de prensa (marzo: inicio ola)
<p>Principales actividades</p>	<ul style="list-style-type: none"> ▪ Elaboración trípticos del PaD (comunicación con los hogares de la muestra). ▪ Elaboración de presentaciones de la encuesta y de sus resultados en diferentes formatos. ▪ Soporte a las publicaciones (pre y post edición). ▪ Organización de jornadas y seminarios para la presentación de los resultados. ▪ Mantenimiento i actualitzación de la página web del PaD. ▪ Mantenimiento i actualitzación de las bases de datos para la difusión del PaD. ▪ Elaboración de boletines electrónicos periódicos (Boletín FJB i Boletín PaD) con las novedades del PaD.
<p>Outputs</p>	<ul style="list-style-type: none"> ▪ Web del PaD => http://www.obdesigualtats.cat/ ▪ Colección trípticos del PaD => http://www.obdesigualtats.cat/sabies-que.../triptics ▪ Presentaciones, en diferentes formatos, <i>ad hoc</i>. ▪ Publicaciones => http://www.obdesigualtats.cat/publicacions