



Support Materials FUNDACIÓ JAUME BOFILL

Panel Survey of Social Inequalities in Catalonia

User Manual

1st (2001/02) to 11th Waves (2012)

JUNE 2014



This User Manual sets out to be a support tool for all researchers working with the longitudinal database of the Panel on Social Inequalities in Catalonia (hereinafter PaD).

Contents

1. The PaD survey	3
2. Data files	14
3. Support materials	16
Appendices	18

1. The PaD survey.

1.1. Data on social structure and inequalities.

The primary aim of the PaD is to provide information about social inequalities in this country. Setting out from the definition of inequality as the social expression of the different possibilities open to individuals of accessing social resources and goods, the PaD focused its efforts on bringing out not only the differences among the population surveyed, but also the mechanisms which generate or reinforce social inequalities.

The PaD survey is designed to gather information about the opportunities and limitations for people living in Catalonia with regard to aspects such as having a job, having a home, finding work appropriate to their training, reaching the desired level of training, access to information, access to medical care, enjoying good health and having sufficient income to live and support children and a family, among others. Information on factors like these reveals different aspects of inequality in this country. It can also help to better understand phenomena such as functional illiteracy, educational failure, intermittent and long-term unemployment, temporary over- and under-qualification for jobs, short-term employment, dependence on public institutions, deprivation and, though to a lesser extent¹, situations of poverty and social exclusion.

To find out about the content of the PaD in greater detail, see the *Longitudinal Report on Content*.

1.2. A longitudinal survey.

The PaD is the first longitudinal survey to be run in Catalonia with an adequately representative sample of Catalan households².

From a strictly technical standpoint, a basic feature of a pure panel-type questionnaire is that it always concerns the same sample of individuals through the different waves in which it is conducted, unlike a rotating panel-type survey in which part of the sample changes each time. The frequency of the panels is not predetermined, but depends on the intentions and/or financial possibilities of each project. In this case, the frequency of the PaD was planned to be annual.

The advantages of monitoring a panel sample can be summed up in four basic aspects: higher quality, greater quantity, inclusion of the perspective of change and more meaningful data.

Quality: A panel sample means that data can be gathered much more effectively: going back to the same households year after year makes it possible to purge errors and check many of the data gathered on the individuals within them. The continuity of the survey staff makes it possible to avoid errors in surveying³ and to build trust in participation by households.

1 It is well-known that a data-gathering instrument like a survey tends to exclude the two extremes of society: the wealthiest bands (which are highly unwilling to cooperate) and the most impoverished bands (also unwilling, but above all due to major difficulties in locating and keeping them in surveys).

2 In Catalonia there are households which have formed part of samples in longitudinal surveys run by INE, the Spanish national statistics institute, such as for example the Household Budget Continuous Survey (HBCF), which provides information about the source and amount of income in Spanish and households and how it is used for different consumer expenditure, but these samples have never been sufficiently representative in statistical terms to carry out specific studies of Catalonia.

3 We are aware that the continuity of survey staff may also be a source of bias.

Quantity: The panel sample makes it possible to pose the different blocks of questions in varying depth in each of the waves (one year the topic of education is pursued in depth, another year the focus is on domestic work, another on housing issues and so on), so that much more information can be built up in each area as the different waves progress.

The perspective of change: The panel sample makes it possible to gather highly valuable data concerning change: changes in housing, changes in household infrastructure, changes in employment and so on. The fact of talking to the same individuals year after year makes it possible to find out not only what changes come about in their lives but also in which social groups changes occur, to what degree, the consequences for other aspects of personal and family life and so on, without the biases arising from talking to a new sample of individuals in each successive wave. This dynamic perspective in the data gathered represents added value which is only gained with panel samples.

Meaningfulness: The panel sample, to the extent that the same people are surveyed in each successive wave, means that in statistical terms the changes recorded are more meaningful than if a new sample of individuals were surveyed each time. The panel makes it possible to reduce the margin for error involved in any sample, and therefore to make statements about the changes which come about in the social situation, even if these are smaller in percentage terms.

The panel sample allows for a longitudinal view of reality, making it possible to go beyond the "snapshot" and look at a period of time. This reveals the progress of individuals and households, while seeing how they experience changes and respond to them within their own social and economic environment. The panel survey also offers information about particularly important events in people's lives, such as the processes of transition from school to work, work to retirement, leaving the parental home for a household of one's own, the birth of children or the dynamics of falling into or escaping poverty, among many other processes of change.

Finally, it should be pointed out that, even though the initial sample is used in the successive waves and its special interest lies precisely in this continuity, the fact that it is made up of living individuals and households also gives a certain "life" to the sample. This is why, through the different waves, the sample undergoes changes such as the addition of new individuals (births, formation of new couples, taking grandparents into the household, families coming together and so on), the loss of individuals (deaths, long stays abroad, separations, departure of grown-up children, institutionalisation of elderly people and so on), the creation of new households (when young people leave the parental household to start one of their own), the disappearance of whole households (an elderly person living alone who dies, people who leave to live outside Catalonia and so on). In this respect, therefore, the sample in a panel survey combines the dynamic dimension (of change) with that of stability and continuity.

1.3. Two units of study: the household and the individuals in it.

The PaD has the peculiarity that it deals with two units of study: households and the individuals in them. This feature has to do with the intention of broadening the explanatory potential of the data gathered as much as possible. Thus, data from the PaD can be analysed both for households themselves (geographical mobility, formation and break-up of households, strategies in the education of children under 16 and so on) and for the individuals of whom they are made up. Added to this is

the usefulness in analytical terms of having, for individuals, data concerning their households, the immediate context in which they live.

In methodology terms, this involves the following:

1) In sampling terms: sampling must be such that both the number of households and the number of individuals are adequate for the stipulated margins of error.

2) In terms of questionnaires: it was necessary to draw up two different types of questionnaire in order to gather information appropriate to the unit of study, either the household questionnaire or the individual questionnaire.

3) In terms of field work:

(a) Recruiting of households: to form part of the sample, the whole household must agree to take part in it (in years when households are recruited⁴ (1st wave, 6th to 9th waves) and households where only some of the members will take part are not accepted), and they must also agree to the successive waves of the survey.

(b) Surveying: a programme had to be developed which allowed for codification of the changes which came about in households (arrival and departure of individuals) and was able to reflect these changes in the surveys conducted immediately after reporting of the changes in the households.

4) In terms of data files: to be able to follow individuals and households over successive waves, it was necessary to create identifiers for households and individuals which would be constant over time in order to ensure that files from different waves could be put together⁵.

1.4. Objective and subjective aspects of individuals' lives.

The official statistics available concentrate in general on objective aspects of people's lives and pay little attention to subjective aspects such as motivations, perceptions, assessments, feelings and opinions. The PaD questionnaires set out to strike a balance between these two dimensions, aiming not to neglect at any time the importance of finding out how respondents experience and assess what happens to them, above and beyond simply establishing what happens.

1.5. The initial sample (2001).

The initial sample for the PaD was made up of 2,000 households in Catalonia, randomly selected following a stratified multistage sample method, with systematic random selection of the first sample units (census units) and a simple random selection of the final units (addresses).

⁴ The year in which the households were recruited for the PaD was the 1st wave, but from the 6th wave onwards resampling was begun in order to make up for losses to the sample and reinforce the sample; this resampling went on through the 7th and 8th wave of the PaD.

⁵ Detailed information is included below concerning the creation and use of the identifiers, in the section on "Individual and household identifiers" in this manual.

It was decided to use census units as the first sample units, as they are the smallest publicly-used territorial statistical units. As they are territorial groupings of an average of 1,000 - 1,200 people, or about 400 households, they have a high degree of internal coherence and the information available allows different classifications by social and territorial structure. The ideal number of households for surveying in each census unit was fixed at 10. It was assumed that, on average, the households in each census unit would involve 270 individuals in the final sample. With regard to the territorial sample fixed, 200 census units had to be selected out of the 5,186 which existed in 1996 (1991 census update).

In order to assure the representative nature of the sample, four stratification variables were combined, two of them territorial and two social:

1. province (including a sub-division into 3 sub-regions in the case of Barcelona);
2. urban-rural setting (depending on size of municipality);
3. variable constructed for social position (classification by \pm working class);
4. employment situation (according to number of active/inactive members)⁶.

The selection of census units resulted in 203 units located in 115 municipalities around Catalonia.

The sample of individuals is the result of counting an average of 2.7 individuals for each household selected for the sample, resulting in 5,445 individuals in the whole of Catalonia.

In order to ensure that the PaD sample is representative in territorial terms and make up for the high concentration of the population in the metropolitan region of Barcelona (where over three quarters of Catalan households live), a non-proportional sample was taken for the provinces.

This is why, based on $\pm 5\%$ as the acceptable margin of sampling error for overall results (with a confidence interval of 95.5%, where $k=2$, and in the case of maximum indeterminacy $p=q=50$), at least 400 households must be observed in each of the four Catalan provinces. As regards the province of Barcelona, its diversity and complexity call for a specific, larger sample, which in the end was set at 800 households, distributed in proportion to population between the city of Barcelona (261), its metropolitan region (470) and the central *comarcas* (counties) of Catalonia (69). This was how the final size of the sample came to 2,000 households, distributed non-proportionally at provincial level.

During the field work two checks were carried out as to the quality of the sample of households and individuals recruited, to ensure that there are no serious biases which would call for corrections to the sample designed.

1.5.1. Criteria for sample follow-up.

Any survey must have criteria for deciding on the sample needed to meet its objectives. In the case of PaD – a pure longitudinal survey – this is crucial, not only for recruitment, but to know with each new wave which households or individuals it should continue surveying and which not to.

Being able to form part of the sample means being considered as “eligible”, whereas not being able

⁶ The social position variable is constructed on the basis of questions 12 and 13 in the 1996 municipal census.

to form part of the sample, as “ineligible”. In each wave, therefore, an attempt has been made to survey all eligible households and individuals from the sample in the previous wave.

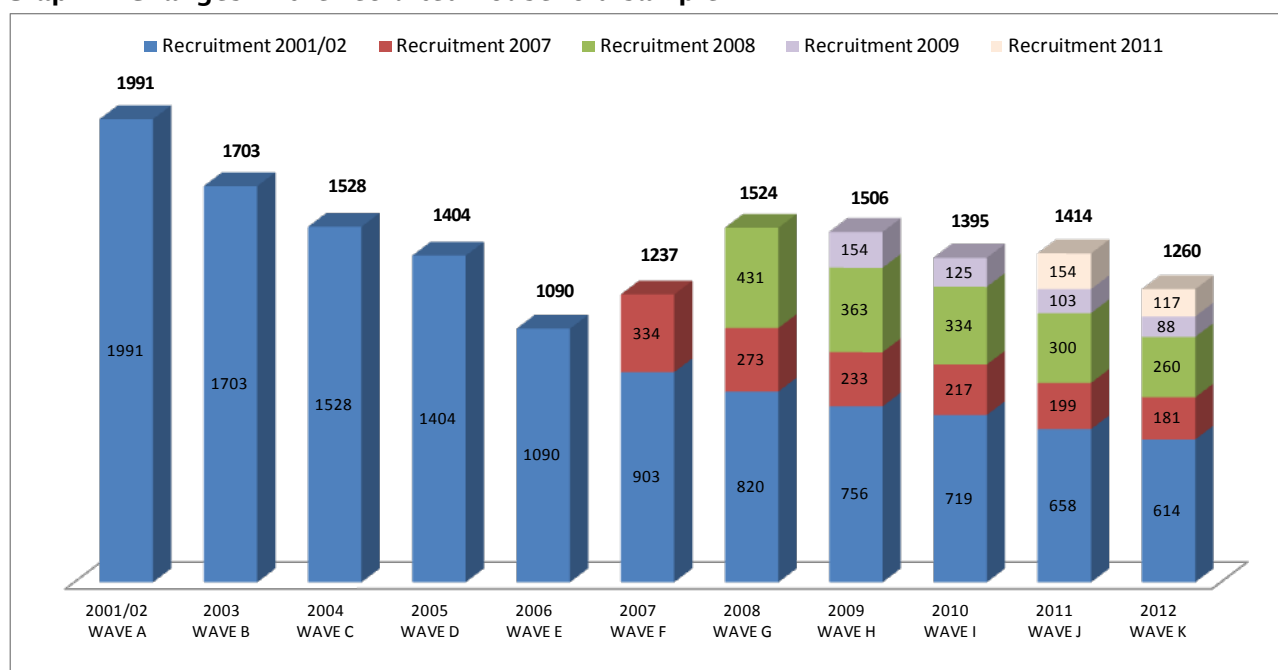
In the case of households, the eligibility criteria are that the household is established in Catalonia and is a private household that is not a school or institution. For individuals, the criteria are that they are over the age of 15 and live in an eligible household.

1.5.2. Changes in the sample.

As mentioned above, due to their periodic nature, longitudinal surveys lose cases from year to year. On the one hand, households and individuals stop being eligible (due to institutionalisation, death or moving outside Catalonia). On the other, since the response to the PaD survey is voluntary, they may refuse to participate or be impossible to locate. It has therefore been necessary to recruit new sample at different times in order to maintain margins of sampling error in statistically acceptable limits.

From the point of view of a longitudinal analysis, we need to be able to monitor changes in the sample in order to determine how many households and/or individuals can be followed up over the course of the various PaD waves. To summarise these changes, the two graphs below show the behaviour of recruited households and individuals in the waves when they were recruited (tables with the corresponding percentages are shown in the appendix). It should be noted that this is not the total number of households or individuals included in the data matrices, as new households (created due to children leaving the parental home, separation or divorce, etc.) or new individuals surveyed (after joining a household in the PaD sample) have not been included.

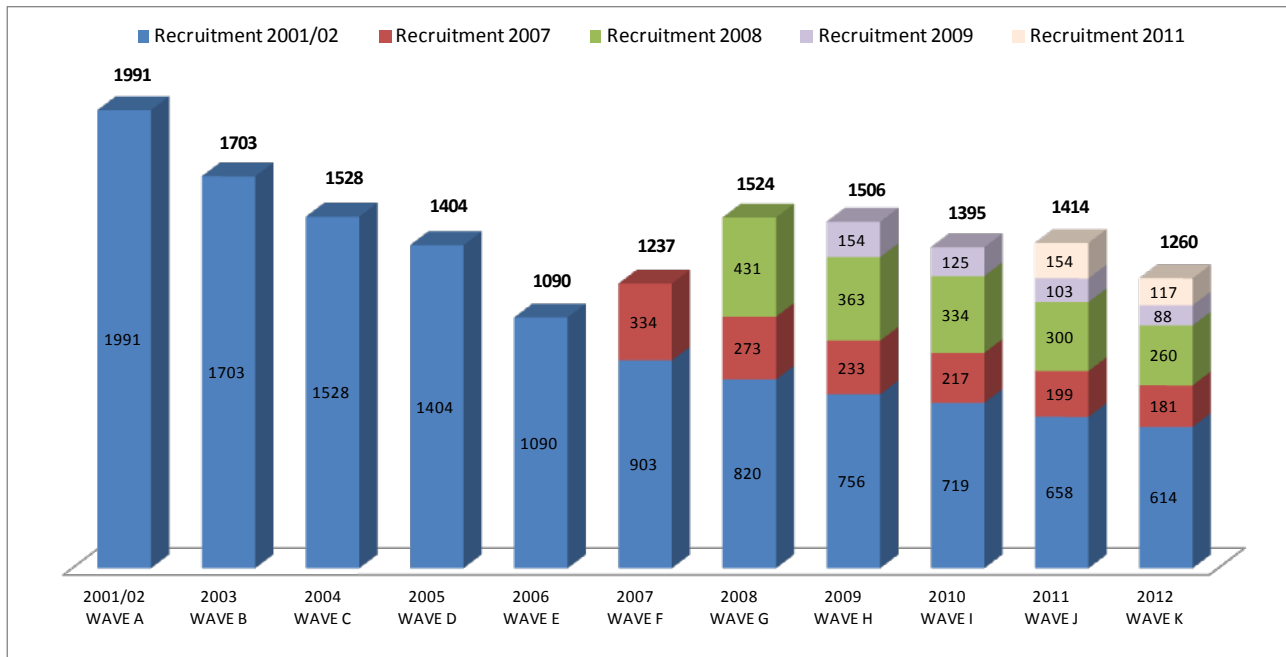
Graph 1. Changes in the recruited household sample



Source: produced by the authors using PaD data (1st to 11th waves).

This graph only includes the additions of households to the sample by recruitment, not those brought about by existing household dynamics.

Graph 2. Changes in the recruited individuals sample

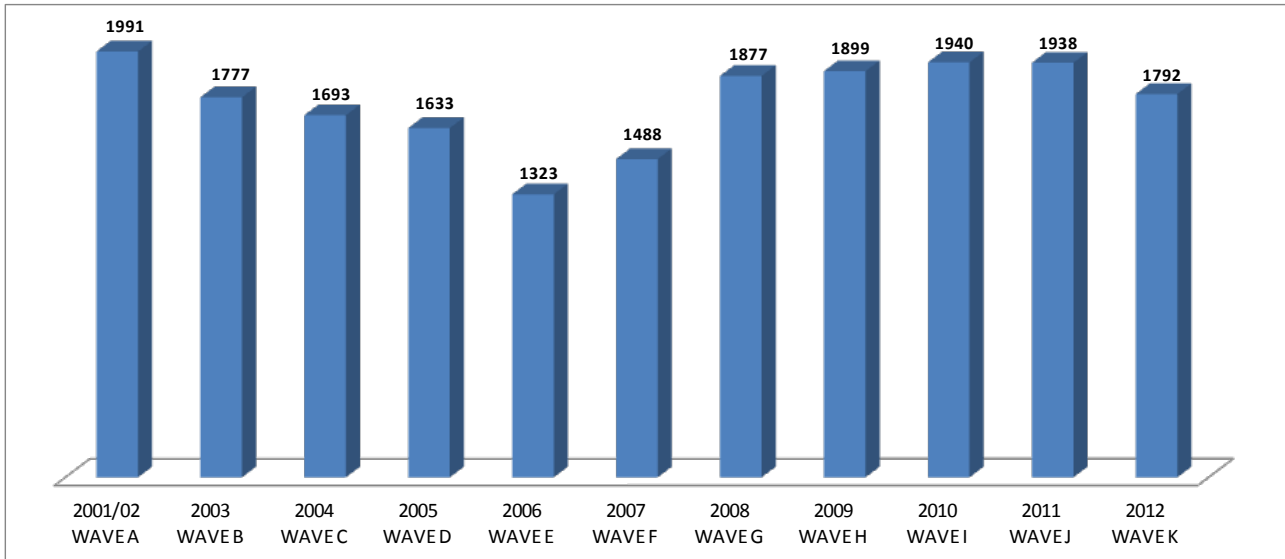


Source: produced by the authors using PaD data (1st to 11th waves).

This graph only includes the additions of individuals to the sample by recruitment, not those brought about by existing household dynamics.

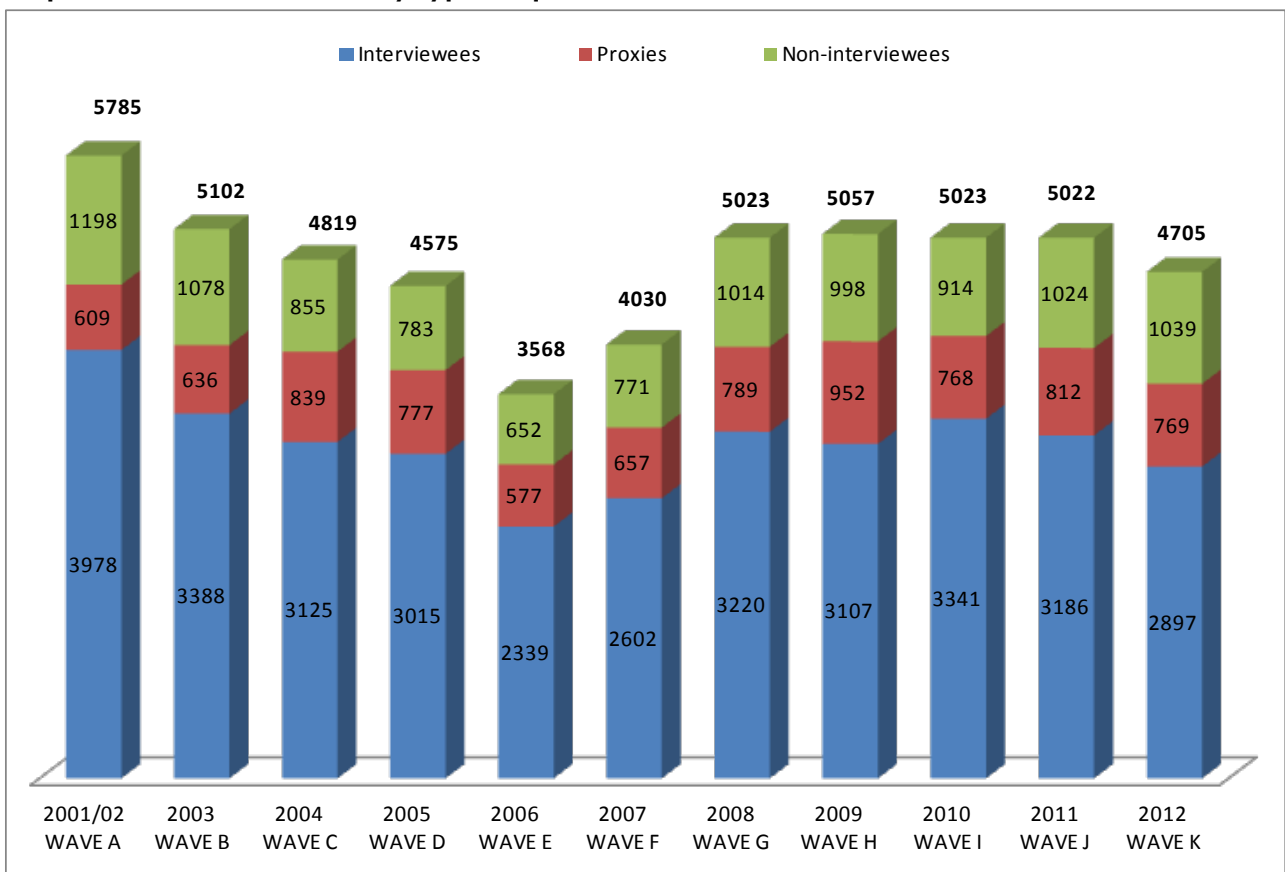
From a transversal viewpoint, all observations in each wave are of interest. The following two graphs aim to present this information clearly. In the case of individuals, we wanted to specify the type of questionnaire (individual, proxy or not interviewed) because each provides a different amount of information. More detailed information on this point can be found in the section on the questionnaire.

Graph 3. Households interviewed in PaD per wave



Source: produced by the authors using PaD data (1st to 11th waves).

Graph 4. Individuals in PaD by type of questionnaire and wave



Source: produced by the authors using PaD data (1st to 11th waves).

1.5.3. Resampling.

In the 6th wave (in 2007), in view of the accrued erosion in the number of individuals in the sample over the first five waves and thanks to the availability of economic resources, it was decided to increase the numbers in the sample.

The increase in the PaD sample was effected on the basis of resampling, following the same method as that used for the initial sampling in the 1st wave (same stratification variables, same way of selecting the final census units) but with one important difference: the census units in the initial sample with surviving households were the starting point for the new sample, which was to be completed with census units to assure it was still representative in territorial and social terms. This decision resulted from the need to limit the dispersion of survey staff over Catalan territory and make field work as easy as possible to manage. Where it was not possible to use the same census units for the results of resampling, new census units were selected at random.

The difficulties in recruiting which were encountered during field work made it necessary to prolong the process of the planned recruitment of 1,074 households during the course of the 6th wave for waves 7 (2008), 8 (2009) and 9 (2010). Prolonging recruitment over four waves meant that between the 7th and 8th waves the resampling had to be recalculated to fit the new households to be recruited as closely as possible to the Catalan situation.

1.5.4. Weighting.

While testing goodness-of-fit is a requirement in any sample (even in a sample to be followed up over the years), final data matrices always require the necessary re-balancing and use of weighting variables (also called raising or weights).

Weighting is a transformation that allows a numerical value (weight) to be given to each sample unit in the survey, representing the contribution the unit will make to the estimations based on survey data.

Each of the PaD waves has transversal weighting of variables, at household and individual level. These variables are generated by the Applied Statistics Department at the Autonomous University of Barcelona ([SEA-UAB](#)).

The basic characteristics of each of these variables are shown below. For more information, please refer to the methodological report prepared by SEA-UAB, available on the [PaD website \(english\)](#).

1.5.4.1. Households.

In the case of households, two weighting variables have been constructed. The first balances them at the time the sample was designed in 2001 with data from the 2001 Population and Housing Census (Spanish National Statistics Institute – INE). The intention of PaD was to generate it with the 2011 Census for the last two waves, but a lack of availability of data from INE meant that this was not possible. Therefore, the variable balances each of the waves with reference to the 2001 data.

This variable is presented in two formats: the first one only balances the sample (variable code

XW009001L⁷); the second balances and raises the number of households to Catalan households (XW009002L).

The variables used for weighting PaD households based on the 2001 Census are as follows:

- Census section
- Province
- Number of people living in the household
- Place of birth
- Age
- Circumstances regarding paid employment
- Housing tenure

However, it was also of interest to have a weighting variable that took into account the social reality of each of the years in which the PaD did the field work. Not only was it impossible to find universe data, but there was the added difficulty of gathering data on both households and individuals. In the end, it was decided to use a national survey, with representation from Catalonia, that was annual and harmonised at European level: the Active Population Survey (EPA). Therefore, the matrix includes a second weighting variable (XW011001L) that balances the sample to the EPA for the 4th quarter of each year.

The variables used for weighting PaD households based on the EPA are as follows:

- Province
- Number of people living in the household
- Place of birth
- Age
- Circumstances regarding paid employment
- Maximum education level

1.5.5.2. Individuals.

In the case of individuals, the reference database has been the electoral roll for each year. As with the census data for households, it is presented in the format that balances the sample (XW010001E) and in the format that, in addition to balancing, also raises it (XW010002E).

The variables used for weighting PaD individuals based on the electoral roll are as follows:

- Province
- Size of municipality
- Gender
- Age
- Place of birth

1.6. Questionnaires.

The PaD questionnaires include four different perspectives, which can be grouped in two pairs:

--> transversal versus longitudinal perspectives

--> household versus individual perspectives

⁷ X refers to the letter of the wave – 1st wave is A, 2nd is B and so on.

The first pair combines structural and dynamic perspectives. On the one hand the PaD questionnaire allows for static or transversal analysis of individuals' social circumstances at a particular time in their personal history, and consequently in the history of the country as a whole. On the other, the panel gives the option of longitudinal analysis and therefore insight into change, particularly micro-social change.

Both pairs are expressed in the questionnaire in the form of different questions which are repeated in successive waves at different frequencies depending on their content. These include the following:

Types of variable (by frequency of appearance)

CORE	<p>Identified by a T (which appear in every wave)</p> <p>Variable which appears in every wave from the time it is included in the questionnaire. It serves to analyse the household and individuals in the present and focuses on assessing micro-changes.</p>
STRUCTURAL CORE	<p>Identified by a TE (which appear in a single wave or few times in the whole panel life)</p> <p>There are two types of structural core variables: Variables with no variation over the lifetime of the respondent and are only recorded once throughout the life of the panel.</p> <p>Variables with little variation over the lifetime of the respondent and are therefore not worth recording in every wave.</p>
CYCLICAL	<p>Identified by a C (plus the frequency of appearance)</p> <p>Variable which appears cyclically, as it is felt appropriate to analyse changes in it only every X years. It is marked as C₂, C₃, C₄, etc. according to whether the variable has to appear in the questionnaire once every 2, 3, 4, etc. waves.</p>
SPORADIC	<p>Identified by a E and appearing in a single wave.</p> <p>Variable which appears sporadically in one or more of the waves of the panel and for which no date is specified for its next inclusion. Its sporadic nature means it cannot be analysed in terms of studying change.</p>

The second analytical pair combines family and individual circumstances. As pointed out above, the PaD allows for analysis of the individuals in the sample on the one hand and, on the other, for analysis of households. This is why a single topic may be divided between the household questionnaire and the individual one.

More specifically, the PaD includes 3 types of questionnaire:

- (a) The household questionnaire: this is answered by a single person who is identified as the Principal Informant. It covers information referring to the whole household.
- (b) The individual questionnaire: this is the most extensive questionnaire and is given to all the individuals aged over 15 in the household.
- (c) The proxy questionnaire: an abridged version of the individual questionnaire (it includes the objective questions); it is answered by an indirect informant in the event that a member of the household is unable or unwilling to take part.

2. Data files.

This section describes the longitudinal data file of the PaD and presents the basic features which are necessary in order to use it and makes it possible to become familiar with it quickly. This is considered to be the basic information to be taken into account in order to assess the PaD data.

2.1. Description and structure.

As a result of the two ways of gathering information from households - about the household as a whole and about each of the individuals of which it is made up - two files or data files are generated for each wave of the PaD: the file **B_K LLAR** which includes the variables supplied by the household survey and **B_K INDI** which does the same with those from the individual survey.

The longitudinal files are provided in long format, i.e. the individuals are repeated on as many rows as the number of waves in which they have taken part. There is one variable (P107001E in the individual file or P107001L in the household file), which reports on the year and therefore on the wave in which the data were gathered. They contain all the waves available except for the 1st one.

Internally, the longitudinal files include the **identifying variables** as headers. Working with the household file, the first of these are: the identifiers of the household, the identifiers of the person in question and those of the principal informant, all of them in two formats, alpha-numerical and numerical, to make it easier to work with any computer program for analysing data. Likewise, the individual file contains the individual and household identifiers. The identifiers are followed in the household file by a series of **summary or count variables**. The variables indicate such things as the number of members of the household, type of household and the number of under 16 year-olds, among others. The individual file, on the other hand, contains instead **context variables** such as the place of the member of the household, their relationship to the reference person, what type of member of the sample they are and so on.

After this, the files contain the **variables which have been harmonised and completed**. The harmonisation process involves standardising the response codes for all the waves, so that they are the same and a single longitudinal variable can be generated. The process of completion is necessary as, on the one hand, as is usual in panel-type surveys, so as not to ask the people - and households - involved the same every year, there is certain information which is referenced: respondents are reminded of it and simply asked whether there has been any change since the last

interview (e.g. their position as regards a partner). If there have been no changes, and as the original question is not repeated, the person need not give the same information again. On the other hand, there is information which is only requested in the first interview (e.g. the mother's date of birth). In both cases this procedure, in file terms, generates many missing values (of the form -8 "Filtered by base"). Thus completion means, providing there have been no changes, filling in the information available from previous waves. These two processes, therefore, give the same response codes for each of the variables in all the waves, providing the maximum number of observations of each individual.

2.2. Missing values.

So as to be able to show the variety of situations in which a question is not answered, a series of specific codes is used, all of them with negative values. However, the files are supplied with the definition as a missing value of the value -8 only, corresponding to those filtered by base or not relevant.

The possible codes for a missing response which are to be found in the files are the following:

-1 Not known

-9 No reply

-3 Not attributable. Corresponds to a few cases in particular waves in which it has not been possible to attribute income.

-8 Filtered by base (inapplicable or does not apply). This is used for cases where individual features (e.g. if the respondent states that they are retired they are not asked whether they receive a retirement pension), or the answers to previous questions (e.g. an unmarried person is not asked for the year of their marriage) mean asking the question is not relevant.

2.3. Individual and household identifiers.

Household and individual identifiers remain constant over time, i.e. they are fixed. They make it possible to follow the progress of individuals and households over different waves. They are the key variables which make it possible to link information on households and individuals. To relate household and individual information, the files must be merged on the basis of the household identifier.

The original identifiers, which are alpha-numerical in form, are assigned automatically in the course of field work for each and every household and individual in the sample. The variables which contains the household code is L017001L and the one with the individual code is L008001E.

Identifiers have their own naming system which provides information about individuals and their households. The first letter (or the first two numbers) of the household identifier indicates the wave in which the household joined the survey or was created. Thus, those starting with A (or 11) joined the PaD sample in the 1st wave (2001-2002) and so on successively. The second letter indicates, in cases where it is relevant, whether the household comes from a household which was previously part of the panel. For example, a household with EA (1511 in its numerical translation) is one created in the 5th wave (2006) out of a household in the 1st wave. In this context, households recruited subsequently have the first two letters corresponding to the wave in question (a household

recruited in the 6th wave will start with FF or 1616). Moreover, if more than one new household has been created out of an original one in a specific wave, these are distinguished with a lower-case letter at the end of the identifier (a, b, etc.). If any of these new households also generates more than one new household in a given wave, this is also provided for, by having households with two lower-case letters at the end, the first corresponding to the original household and the second to the order in which the new household was created).

The identifiers generated are subsequently created and are a numerical translation of the previous one (codes L017001LG2 i P008001EG2). The fact that the original identifier contains letters and numbers may lead to problems when it comes to merging files using some of the different statistical packages available on the market, which is why this second identifier is also provided. Its use is recommended to avoid issues arising from the nature of the original identifier.

The numerical identifier is constructed in the following manner:

- 2 digits for the first letter and two digits for the second. In both cases, the first of the two digits refers to the decade of the panel and the second to the year. This ensures that the code does not begin with zero and will be properly recognised by any computer program, as well as providing for codes generated after the 10th wave.
- the numbers of the identifier
- 1 digit for the first lower-case letter
- 1 digit for the second lower-case letter
- 2 digits for the position of the household member, in the case of individual identifiers

Thus household AA0154 will translate as 1111015400; household BA1854a as 1211185410 and household FA2342ab as 1611234212. In the case of individual identifiers, the position of the household member must be added to these codes.

3. Support materials.

A series of support materials have been developed in order to facilitate access to and use of the data. The two basic reference documents are those described below.

3.1. The longitudinal report on content.

The longitudinal report on content was produced in order to make the content of the different topic blocks within the PaD questionnaire accessible to users.

The longitudinal report on content makes it possible to view the questionnaire in a **longitudinal** manner. It contains all the PaD variables except the one-off ones, i.e. those which only appeared in a single wave of the PaD. They are organised in topic blocks and sub-blocks. This document provides basic information on variables:

- the label;
- the code of the variable in the questionnaire;

- the code of the variable in the file. The two codes are provided so as to be able to follow the questionnaire item in the data file. This code may have been transformed in some way during preparation of the information for analysis (for example, a multiple-response question is converted into more than one dichotomic variable which takes another code; an open question is closed in a categorical variable ending in R1...);
- frequency; whether the variable is core, cyclical or one-off;
- the waves in which it appeared, identified by the corresponding letter;
- whether the variable has been eliminated or absorbed by one or more other variables.

3.2. The code books.

The code books were designed as an indispensable tool for working with the data files. Each data file has its own code book. The content of code books is as follows:

- 1) The **Header**, which specifies how many records (individuals or households) it contains.
- 2) The **Body**: this shows the features of each of the variables in the same order as they appear in the data file. The information given is the following:
 - the variable code;
 - the label;
 - the base, i.e. which individuals or households have the information;
 - the response codes;
 - the number of observations;
 - the number of 'don't know'/'no answer' responses.

Appendices.

Technical details of the sample.

Universe

Private households in Catalonia

Type of sampling

- Disproportional distribution of the number of households to be interviewed by provinces and proportional to the size of municipality.
- Systematic, multi-stage random selection of the first sample units (census sections), according to social and territorial stratification variables: geographical area, size of municipality, social structure (constructed variable) and occupational circumstances (active/inactive).
- Simple random selection of the final sample units; addresses of the households to be interviewed in the first sample units.

Theoretical sample and margins of error

Households: 2,000 households in Catalonia

Margin of error: $\pm 2.79\%$ for overall results

AREAS	NO. OF HOUSEHOLDS	e* (%)
Barcelona	800	± 3.54
City	261	± 6.12
Metropolitan Region	470	± 4.61
Central Countries	69	± 12.04
Girona	400	± 5.00
Lleida	400	± 5.00
Tarragona	400	± 5.00
Total	2.000	± 2.79

*e = sampling error (confidence interval=95.5%, k=2 and p=q=50)

Individuals: 5,445 individuals in Catalonia (individuals per household: average 2.72)
 Margin of error: $\pm 1.67\%$ for overall results

AREAS	NO. OF INDIVIDUALS	e* (%)
Barcelona	2,178	± 2.14
Girona	1,054	± 3.08
Lleida	1,120	± 2.98
Tarragona	1,104	± 3.01
Total	5,456	± 1.67

*e = sampling error (confidence interval=95.5%, k=2 and p=q=50)

Real sample (end 1st wave)

1,991 households in Catalonia
 Margin of error: $\pm 2.78\%$ for overall results

AREAS	NO. OF HOUSEHOLDS	e* (%)
Barcelona	791	± 3.54
Girona	400	± 5.00
Lleida	400	± 5.00
Tarragona	400	± 5.00
Total	1,991	± 2.78

*e = sampling error (confidence interval=95.5%, k=2 and p=q=50)

5,757 individuals in Catalonia
 Margin of error: $\pm 1.32\%$ for overall results

AREAS	NO. OF INDIVIDUALS	e* (%)
Barcelona	2,213	± 2.13
Girona	1,193	± 2.89
Lleida	1,175	± 2.91
Tarragona	1,176	± 2.91
Total	5,757	± 1.32

*e = sampling error (confidence interval=95.5%, k=2 and p=q=50)

Reference tables for Graphs 1 and 2 on changes in the sample.

Households.

	Recruitment 1st wave	Interannual loss	Accumulated loss	Recruitment 6th wave	Interannual loss	Accumulated loss
1st wave 2001- 2002	1,991					
2nd wave 2003	1,703	14.47%	14.47%			
3rd wave 2004	1,528	10.28%	23.25%			
4th wave 2005	1,404	8.12%	29.48%			
5th wave 2006	1,090	22.36%	45.25%			
6th wave 2007	903	17.16%	54.65%	334		
7th wave 2008	820	9.19%	58.81%	273	18.26%	18.26%
8th wave 2009	756	7.80%	62.03%	233	14.65%	30.24%
9th wave 2010	719	4.89%	63.89%	217	6.87%	35.03%
10th wave 2011	658	8.48%	66.95%	199	8.29%	40.42%
11th wave 2012	614	6.69%	69.16%	181	9.05%	45.81%

	Recruit- ment 7th wave	Interan- nual loss	Accumu- lated loss	Recruit- ment 8th wave	Interan- nual loss	Accumu- lated loss	Total individu- als	Accumu- lated loss
1st wave 2001- 2002							1,991	
2nd wave 2003							1,703	14.47%
3rd wave 2004							1,528	23.25%
4th wave 2005							1,404	29.48%
5th wave 2006							1,090	45.25%
6th wave 2007							1,237	37.87%
7th wave 2008	431						1,524	23.46%
8th wave 2009	363	15.78%	15.78%	154			1,506	24.36%
9th wave 2010	334	7.99%	22.51%	125	18.83%	18.83%	1,395	29.93%
10th wave 2011	300	10.18%	30.39%	103	17.60%	33.12%	1,260	36.72%
11th wave 2012	260	13.33%	39.68%	88	14.56%	42.86%	1,143	42.59%

Individuals.

	Recruitment 1st wave	Interannual loss	Accumulated loss	Recruitment 6th wave	Interannual loss	Accumulated loss
1st wave 2001- 2002	4,587					
2nd wave 2003	3,686	19.64%	19.64%			
3rd wave 2004	3,246	11.94%	29.23%			
4th wave 2005	2,954	9.00%	35.60%			
5th wave 2006	2,204	25.39%	51.95%			
6th wave 2007	1,787	18.92%	61.04%	745		
7th wave 2008	1,599	10.52%	65.14%	576	22.68%	22.68%
8th wave 2009	1,451	9.26%	68.37%	486	15.63%	34.77%
9th wave 2010	1,365	5.93%	70.24%	445	8.44%	40.27%
10th wave 2011	1,245	8.79%	72.86%	382	14.16%	48.72%
11th wave 2012	1,130	9.24%	75.37%	340	10.99%	54.36%

	Recruit- ment 7th wave	Interan- nual loss	Accumu- lated loss	Recruit- ment 8th wave	Interan- nual loss	Accumu- lated loss	Total individu- als	Accumu- lated loss
1st wave 2001- 2002							4,587	
2nd wave 2003							3,686	19.64%
3rd wave 2004							3,246	29.23%
4th wave 2005							2,954	35.60%
5th wave 2006							2,204	51.95%
6th wave 2007							2,532	44.80%
7th wave 2008	957						3,132	31.72%
8th wave 2009	772	19.33%	19.33%	335			3,044	33.64%
9th wave 2010	683	11.53%	28.63%	264	21.19%	21.19%	2,757	39.90%
10th wave 2011	594	13.03%	37.93%	203	23.11%	39.40%	2,424	47.16%
11th wave 2012	495	16.67%	48.28%	168	17.24%	49.85%	2,133	53.50%

Technical details of fieldwork.

Method of conducting interviews.

Efforts are made to interview all members of the households in the original sample aged 16 or over. Personal interviews are conducted in private homes using the CAPI (Computer-Assisted personal Interviewing) system.

Three types of questionnaire are put for each household:

1. Household: this covers information referring to the whole household; it is answered by a single member of the household, the principal informant.

2. Individual: this covers information referring to each household member aged 16 or over; it is answered by each person in this category.

3. Proxy: an abridged version of the individual questionnaire (it includes the objective questions); it is answered by an indirect informant in the event that a member of the household is unable or unwilling to take part.

Sample of fieldwork (covers all the households worked on by waves).

	Initial no. of households*	No. of households worked on**	No. of households surveyed***
1st wave 2001- 2002	0	1,991	1,991
2nd wave 2003	1,991	2,149	1,777
3rd wave 2004	1,931	2,044	1,693
4th wave 2005	1,791	1,880	1,633
5th wave 2006	1,724	1,791	1,323
6th wave 2007	1,473	1,862	1,488
7th wave 2008	1,749	2,268	1,877
8th wave 2009	2,089	2,332	1,899
9th wave 2010	2,114	2,219	1,940
10th wave 2011	2,043	2,266	1,938
11th wave 2012	2,069	2,127	1,792

*INITIAL NO. OF HOUSEHOLDS The number of households at the start of the wave includes: households closed in previous wave + households not surveyed re-launched in the field (reject/lost NOT final)

NO. OF HOUSEHOLDS WORKED ON The number of households worked on includes the households at the start of the wave plus the households created DURING the wave (through the internal dynamics of the PaD (separations, individuals becoming emancipated, etc.). In recruiting years (6th, 7th, 8th, 9th) the households recruited are also included.

NO. OF HOUSEHOLDS SURVEYED The number of households surveyed includes the households worked on during the fieldwork which were successfully surveyed (this figure may differ slightly from that in the data files due to some survey lost in processing)

Percentage of interviews accumulated (per month of fieldwork).

	2nd wave 2003	3rd wave 2004	4th wave 2005	5th wave 2006	6th wave 2007	7th wave 2008	7th wave 2009	9th wave 2010	10th wave 2011	11 th wave 2012
JANUARY	5.3	11.3	0.1	-	3.4	1.8	0.4	-	-	-
FEBRUARY	29.6	40.8	28.3	24.7	21.0	23.8	15.4	-	-	0.7
MARCH	55.0	66.2	57.0	56.7	42.0	41.8	33.8	-	7.8	6.8
APRIL	72.5	79.0	77.6	68.4	55.4	54.8	50.5	26.3	37.9	35.3
MAY	83.3	90.4	87.8	81.1	65.2	65.3	68.6	68.3	70.6	66.2
JUNE	92.2	96.0	93.1	91.6	72.8	77.6	81.4	91.0	86.8	87.6
JULY	96.5	96.9	95.0	98.4	80.4	87.6	89.0	93.8	89.8	92.1
AUGUST	96.5	96.9	95.0	98.4	81.4	88.4	89.5	93.8	91.5	95.0
SEPTEMBER	96.5	96.9	95.0	99.5	84.7	92.2	93.2	93.9	95.1	97.5
OCTOBER	99.0	99.1	98.8	100.0	98.7	98.5	98.5	98.6	99.4	100.0
NOVEMBER	100.0	100.0	100.0	-	100.0	100.0	100.0	100.0	100.0	-