



EUROPEAN COMMISSION  
EUROSTAT

Directorate D: Single market, employment and social statistics  
Unit D-2: Living conditions and social protection



DOC. PAN 168/2003-12

# **ECHP UDB manual**

*European Community Household Panel  
Longitudinal Users' Database*

**Waves 1 to 8  
Survey years 1994 to 2001**

## Table of Contents

<b>1. INTRODUCTION .....</b>	<b>3</b>
1.1. A FEW WORDS ABOUT THE ECHP .....	3
1.2. THE NEED FOR A LONGITUDINAL USERS' DATABASE .....	4
<b>2. THE ECHP DATA .....</b>	<b>5</b>
2.1. THE 'PRODUCTION' DATA BASE (PDB) .....	5
2.2. DATA CHECKING .....	5
2.3. THE 'USER' DATABASE (UDB) .....	5
2.3.1. <i>Anonymisation</i> .....	5
2.3.2. <i>User friendliness</i> .....	6
2.3.3. <i>The country file</i> .....	7
2.3.4. <i>The longitudinal link-file</i> .....	7
2.3.5. <i>The four cross-sectional files for each wave</i> .....	7
2.3.6. <i>Data description</i> .....	8
2.4. SAMPLE STRUCTURE VARIABLES .....	9
2.5. VARIABLES DEFINED AT COUNTRY LEVEL .....	10
<b>3. CONTRACTUAL ARRANGEMENTS .....</b>	<b>11</b>
<b>4. RELATED DOCUMENTS .....</b>	<b>14</b>
<b>ANNEX 1 - ADDRESSES OF NATIONAL DATA COLLECTION UNITS AND EUROSTAT .....</b>	<b>15</b>
<b>ANNEX 2 - SAMPLE WEIGHTS .....</b>	<b>17</b>
4.1. A SHORT DESCRIPTION OF ECHP WEIGHTS .....	17
4.1.1. <i>Base weights (RG003 and PG003)</i> .....	17
4.1.2. <i>Cross-sectional weights (RG002, PG002 and HG004)</i> .....	17
4.2. HOW TO USE ECHP WEIGHTS FOR ANALYSIS ? .....	17
4.2.1. <i>Cross-sectional analysis at country level</i> .....	18
4.2.2. <i>Cross-sectional multi-country analysis</i> .....	18
4.2.3. <i>Longitudinal analysis at country level</i> .....	18
4.2.4. <i>Longitudinal multi-country analysis</i> .....	19
4.2.5. <i>Advanced uses of the weights</i> .....	19
<b>ANNEX 3 - EXAMPLES OF CALCULATIONS .....</b>	<b>20</b>

## 1. INTRODUCTION

### 1.1. A few words about the ECHP...

In 1991, Eurostat, the Statistical Office of the European Communities, set up a Task Force on Household Incomes in order to respond to the strong demand for information on household and individual income. The Task Force was mandated to assess, together with EU Member States, the income data in registers and existing national household surveys, and to check whether the available outputs could be satisfactorily harmonised *ex-post*.

After the failure of this “output approach”, the decision was taken to launch a specific EU survey, the European Community Household Panel (ECHP), to adopt an input-oriented approach rather than strictly try to harmonise existing outputs. Although the questionnaire was designed centrally at Eurostat, in close consultation with the Member States, it allowed for some flexibility for adaptation to national systems.

The ECHP forms therefore the most closely co-ordinated component of the European system of social surveys. It has been given a central place in the development of comparable social statistics across Member States on income (including social transfers etc.), labour, poverty and social exclusion, housing, health, as well as various other social indicators concerning living conditions of private households and persons. The multi-dimensional and multi-purpose nature of the survey also enables the study of the interrelationships between these dimensions.

The longitudinal, ‘panel’ design of the ECHP makes it possible to follow up and interview the same set of private households and persons over several consecutive years. In contrast to a cross-sectional survey, it supplies data on EU social dynamics i.e. it provides information on relationships and transitions over time at the micro level.

ECHP data are collected by “National Data Collection Units” – “NDUs”, either National Statistical Institutes (NSIs) or research centres depending on the country (see annex 6 for addresses of NDUs). In the first wave (in 1994) a sample of some 60,500 nationally representative households - i.e. approximately 130,000 adults aged 16 years and over - were interviewed in the then 12 Member States. Austria (in 1995) and Finland (1996) have joined the project since then. From 1997 onwards, similar data is available for Sweden. In fact, ECHP UDB variables were derived from the Swedish Living Conditions Survey and are now included in the ECHP UDB. In wave 2, EU-13 samples totalled some 60,000 households and 129,000 adults.

For the fourth wave of the ECHP, i.e. in 1997, the original ECHP surveys were stopped in three countries, namely Germany, Luxembourg and in the United Kingdom. In these countries, existing national panels were then used and comparable data were derived from the German and UK survey back from 1994 onwards, and for the Luxembourg survey back from 1995 onwards. Consequently,

two sets of data are available for the years 1994 to 1996 for Germany and the UK, and 1995-1996 for Luxembourg.

Eurostat recommends the use of the original ECHIP data for any analysis covering only the years 1994-1996 for countries with two different datasets for the same year. However, for longitudinal analysis covering more years, the converted datasets should be used.

For a detailed description of the ECHIP methodology and questionnaires, please see “The European Community Household Panel (ECHIP): Volume 1 - Survey methodology and Implementation” and “The European Community Household Panel (ECHIP): Volume 1 - Survey questionnaires: Waves 1-3” - Theme 3, Series E, Eurostat, OPOCE, Luxembourg, 1996.

## 1.2. The need for a longitudinal users' database

Since the first ECHIP results became available, there has been an increasing demand from inside and outside the Commission for ECHIP based statistics. Many researchers and other users have also expressed strong interest in having direct access to the data. In view of this, Eurostat decided to develop, together with NDUs, a set of rules allowing for easier direct access to “anonymised” ECHIP micro-data, without jeopardising both the necessary conditions of data confidentiality and the value of the data.

In this context, Eurostat proposed to NDUs in November 1997 to create a user-friendly and widely documented “longitudinal users' database” (hereinafter referred to as the “users' database” or simply “UDB”) that would meet various “objective anonymisation criteria”. By “objective”, it is meant that once these criteria are applied to the various ECHIP files, there should be no risk that an individual statistical unit could be identified through “all the means that might reasonably be used by a third party to identify the said statistical unit” (EU Council regulation N° 322/97 of 17/2/97 on Community statistics, also referred to as the ‘Statistical Law’).

Provided that all of these anonymisation criteria are met, ECHIP data should thus be considered as “non-confidential” in the sense of the “Statistical Law”. However, it is essential that direct access to such anonymised micro-data be restricted by means of **contracts stipulating the strict conditions of use and access** (see section 3 on contracts; see also the related contract that links users and/or users' organisations, on the one hand, and Eurostat, on the other).

It is clearly in the interest of the ECHIP project to be widely used and visible through interesting and useful analyses and publications. In this respect Eurostat would like to thank all national ECHIP partners for supporting this major breakthrough in ECHIP data availability.

Eurostat would be grateful if users could inform the ECHIP team of any errors and omissions found in the data or in the documentation. Any information, comments or suggestions for further improvement would be welcome.

## 2. THE ECHP DATA

For each wave of the ECHP, a codebook, as well as a list of variables are available (wave 1: Doc.PAN. 15; wave 2: Doc.PAN. 30; wave 3: Doc.PAN. 65; wave 4: Doc.PAN. 81; wave 5: Doc.PAN. 97; wave 6: Doc.PAN. 112; wave 7: Doc.PAN. 151; wave 8: Doc.PAN. 159). These documents contain the "Community" question wording and corresponding variable names (questionnaire variables) which are necessary for data processing. National questionnaires are based on this common version.

The information collected by means of these questionnaires is checked by the National Data collection Units (NDUs) and by Eurostat. NDUs provide Eurostat with the results of the interviews. This information is stored in the so-called 'Production' data base (PDB). Based on this PDB, an anonymised user-friendly longitudinal user data base (UDB) is constructed.

### 2.1. The 'Production' data base (PDB)

The PDB consists of micro-data files that are sent to Eurostat in a format that is very close to the EU questionnaire. For each wave there are four cross-sectional files (D, H, R and P file).

All the work relating to data checking (both cross-sectional and longitudinal) is done in this production database.

### 2.2. Data checking

The data has been checked as thoroughly as possible, both at micro and aggregated levels, and longitudinally between waves. However, further checking and refinements are required. The ECHP is a dynamic project, i.e. the data are "never" completely final: panel data must be continuously updated using information collected in subsequent waves - it is an ongoing backwards and forwards process. Therefore it is hoped that users will promptly inform Eurostat of any errors found.

### 2.3. The 'User' database (UDB)

The PDB contains information considered 'confidential' in terms of the EU 'Statistical law'. Its structure is very complex and thus extremely difficult to use outside, but also inside Eurostat. This also increases the risk of errors through the use of the data. For these reasons, access to the original data had to be more restrictive than what would be desirable to exploit the full potential of the data.

However, in view of the increasing demand for ECHP data, Eurostat has constructed an anonymised user-friendly "longitudinal users' database", the ECHP UDB.

#### 2.3.1. Anonymisation

For confidentiality reasons, the UDB needs to meet various "objective anonymisation criteria" as described earlier. Eurostat has developed appropriate anonymisation criteria in close consultation with the NDUs. Provided that all these criteria are met, the ECHP data can be considered 'non-confidential' in terms of the

'Statistical law' and made more widely available. (see Doc.PAN.105 for more information on the anonymisation of ECHP data).

However, access to such anonymised micro-data still needs to be restricted by means of contracts stipulating the strict conditions of use (see section 3 'Contractual Arrangements').

### 2.3.2. *User friendliness*

The major changes from the 'Production' to the Users' database are as follows:

- A link file which allows tracing of individuals across waves has been set up.
- The variables have been fully reorganised, grouped together and standardised, which means that they no longer reflect the structure of the questionnaire. Analytical variables derived from original variables have also been added. One important change is that the variable names are now identical in each wave.
- For questions asked only to individuals interviewed for the first time, or to those that have undergone a significant change since the previous wave (e.g. in their labour force status), the information is forwarded to the following waves, thus permitting independent cross-sectional analysis of each wave.

Here are 2 concrete illustrations of the differences between the household and personal files in the production and users' database, as well as some short information on imputation and weighting:

- In the employment section of the PDB, one list of questions is asked to people who normally work 15 hours and more, and another one to those who worked less than 15 hours in a reference week. Although most of the information asked to those two groups is identical, it is presented in two different variables in the production database, reflecting the questionnaire structure, which depends on the amount of hours worked, i.e. 15 hours and more or less than 15 hours. These are combined into a single set of variables in the UDB covering both groups of respondents.
- In the UDB, income components have been defined at a higher level of aggregation than the detailed enumeration given in the PDB. While the latter is required to obtain as complete a picture of the household income as possible, such a degree of detail is not suitable for analytical purposes, especially for the purpose of comparative cross-country analysis. The same structure is followed at the household and person levels. Detailed items of income specified in the questionnaire are aggregated into intermediate level components for each interviewed person in the household; they are also converted to annual net amounts as required.
- Missing information on income is imputed (see Doc.PAN.164 for detailed explanation of imputation of income). Eurostat wishes to thank the Survey Research Center at the University of Michigan, and specifically Dr Raghunathan and Dr Solenberger, for their software for data imputation.

- Weights to be applied in the analysis of the data have been added to the basic data (see Annex II for a short description of sample weights and on how to use these weights, and Doc.PAN.165 for the construction of weights).

The UDB consists of the following data files:

### *2.3.3. The country file*

This file contains the following information for each wave and country :

- population figures (number of private households in the country, number of persons living in private households, number of persons aged 16+ living in private households) for grossing-up and in order to aggregate over countries.
- purchasing power parities for converting national currencies in PPS
- exchange rates for converting national currencies in ECU/EURO

The country file also contains the fixed exchange rates for the 'Eurozone' countries (after 1.1.1999).

### *2.3.4. The longitudinal link-file*

It includes data from all the waves and assigns a record to every person that ever appeared in the ECHP. The first section contains data asked only once (when the person entered the panel or when he/she became eligible). The second section, which is repeated in each wave, contains all the information required to rebuild the "longitudinal status" of the person from the beginning to the end of the panel, derived from the personal and household registers. Each person has an identification number (PID) that is fixed across waves.

### *2.3.5. The four cross-sectional files for each wave*

#### *2.3.5.1. The "register file"*

It covers all persons currently living in households with a completed household interview in each wave.

#### *2.3.5.2. The "relationship file"*

It has been derived from the relationship matrix in the household register file (R-file of the PDB). Its records have the format "person X has relationship R with person Y". It gives a record of all possible relationships of every person in a household. Hence, there is one record for each pair of persons in the same household, specifying their relationship. The following rule is used in specifying the variables corresponding to X, R and Y:

If the relationship is between an ascendant and a descendant, 'R' (variable 'Relation') always specifies the descendant side of the relationship (e.g. the child, grandchild etc.). Variable PID1 is the fixed identification number (PID) of the ascendant, and variable PID2 is the fixed identification number (PID) of the descendant.

Apart from its much simpler structure than the original relationship matrix, the relationship file has the major advantage that individuals are identified in terms of

their fixed PIDs, rather than the wave-specific 'line numbers' in the matrix, so that the consistency and evolution of relationships can be traced over waves.

#### 2.3.5.3. *The “household file”*

This file contains one record for each household with a completed household interview. The information is grouped into 7 sections.

- general information
- demographic information
- household income
- household financial situation
- accommodation
- durables, and
- children.

#### 2.3.5.4. The “personal file”

This file contains one record for each person with a completed personal interview. The information is grouped into 13 sections:

- general information
- demographic information
- current employment
- unemployment
- search for a job
- previous job
- calendar of activities
- income
- education and training
- health
- social relations
- migration, and
- satisfaction with various aspects of life.

#### 2.3.6. *Data description*

The document 'ECHP UDB description of variables' (Doc.PAN. 166) lists the variables in the ECHP UDB. It also provides the codes and labels for these variables, as well as an overview of the differences between the waves and the countries.

The document 'ECHP UDB construction of variables' (Doc.PAN. 167) describes the linkage between the questionnaire variables and the UDB variables.

Data is provided in comma separated value files (CSV-files). For more information see the 'readme'-file provided with the data.



## 2.4. Sample structure variables

To compute sampling errors, at least the following four variables are required to define the sample structure:

HG004: Sample weight (or PG002 for personal interviews, RG002 for the population, etc.)

HG005: Stratum

HG006: Primary sampling units (PSU)

HG007: Indicator of whether the PSUs were selected within strata systematically from an ordered list, and if so, the order of selection.

An additional variable could be the variable defining major domains in the country for which separate sampling error results may be required, e.g. major regions in the country such as NUTS1 .

Apart from being used for the weighting procedure (discussed in DOC.PAN. 165), variables HG005-HG007 define the essential aspects of the structure of the sample needed for the valid computation of sampling errors, which take that structure into account. These variables have been constructed from original variables in the PDB D-file (e.g. for the 1994 wave: D01POINT, D01NUTS3 and D01SMST1-D01SMST4), together with additional information provided by NDUs as necessary. Since the objective of HG005-HG007 is merely to provide structural information on the sample, with no relationship to the actual geographical locations, the original variables have been anonymised through randomisation.

The sample structure variables HG005-HG007 are strictly defined only once for a household, i.e. the first time it appears in the survey. These household variables are assigned to each member of the household, and then remain associated with each person even if the person moves to a different location or to a different household in subsequent waves. New 'non-sample' persons entering a household receive the variables from the sample persons in the household. (In the rare case when the household contains sample persons from more than one original household, these variables are defined by the person who defines the basic part (D0iHHID) of the household identification number  $HID = (D0iHHID * 100) + D0iSPLIT$ ).

Most ECHP samples use two-stage sample designs, in which case HG006 defines the sample clusters, and HG005 defines the explicit strata within each of which the clusters were selected independently. Most ECHP surveys also involve the selection of PSUs with systematic sampling from ordered lists, which amounts to additional 'implicit' stratification. Variable HG007 identifies the order in which the clusters were selected, so that this feature can be taken into account in sampling error computations. The variable is given a 'not-applicable' code (-8) if systematic sampling was not involved. For samples selected in a single stage (direct sampling of households or persons as in Denmark and Ireland), all sample structure variables receive the 'not-applicable' code.

## 2.5. Variables defined at country level

The country file provides information on the total population, number of households, exchange rates and purchasing power parities. The demographic data refer to the end of the year, and the economic data to average conditions over the stated year.

Actual counts of households and persons in various categories can be computed by multiplying the corresponding proportions estimated from the survey by the population totals. The relevant figures for this purpose are the *total population* for estimates from the personal register (Register file); the *household population* for estimates from the household questionnaire (Household file); and *population above 16* for estimates from the personal interview (Personal file).

All monetary amounts in the data are in national currency (NC) units. Normally comparisons are made in equivalent units (purchasing power standards - PPS) taking into account differences in the NC's purchasing power. This is obtained by DIVIDING the NC amounts by the purchasing power parities (PPP) provided below. For amounts referring to the Wave 6 reference year, PPP's for 1998 apply (most income information collected refers to the previous year). For current amounts referring to the survey year, those for 1999 apply.

### 3. CONTRACTUAL ARRANGEMENTS

After the coming into force of the new Commission regulation (EC) No 831/2002 on access to confidential data for scientific purposes on 7 June 2002, the arrangements for access to the ECHP UDB have changed.

Under Article 3 of this regulation, organisations that fall under the categories of:

- 1(a), i.e. universities and other higher education organisations established by Community law or by the law of a Member State; or
- 1(b), i.e. organisations or institutions for scientific research established under Community law or under the law of Member State; or
- (2) organisations (with a legal personality) that have been commissioned by departments of the Commission or of the administrations of the Member States to undertake specific research

can be granted access after the ECHP has been deemed suitable for the proposed research project, and after their research proposal has been approved by all the national authorities. Eurostat must inform the national authorities of the Member States of applications for access to the ECHP data, who have, under the regulation, up to 6 weeks in which to consider the application.

Organisations that fall under the category of:

- 1(c), other agencies, organisations and institutions, after having received the opinion of the Committee on statistical confidentiality, in accordance with the procedure laid down in Article 20(2) of Regulation (EC) No 322/97.

need the approval of the Committee on Statistical Confidentiality before they can make an application for access to the ECHP data for research purposes. If the Committee on Statistical Confidentiality grants approval, access to the data is still subject to the suitability of the data source, and the agreement of the national authorities.

After signature of a contract for access to ECHP data, any person or organisation using this CD-ROM is supposed to be **fully aware** of all of the related **contractual arrangements** that link them with Eurostat. However, Eurostat wishes to remind users that the following provisions, among others, are included in all contracts on the use of ECHP data:

- The contract signed with Eurostat is a research contract. It covers the assignment to the contractor, on the terms set out in the contract, of the right to use, in the form and according to the arrangements specified in the contract, the ECHP users' database.

The files contained in the users' database shall be used exclusively for those research purposes as specified in the contract, excluding in particular any possible administrative use. The data may be used by the contractor solely under the conditions and for the purposes described in the contract.

The contractor shall take all the necessary regulatory, administrative, technical and organisational measures to ensure that the data are used only for the research specified in the contract.

- The contractor shall also take all the necessary regulatory, administrative, technical and organisational measures to ensure that none of these data are distributed to third parties, and that there will be no attempt to identify, nor will the contractor claim to have done so, by any means whatsoever, any individual statistical unit.

In particular, the users' database shall not be connected with other data sets from any other source without Eurostat's explicit written consent.

- The contractor may not process, disseminate or otherwise allow any of the data to be made available or used for any purpose whatsoever other than the research purposes laid down in the contract. He/she shall remain bound by this obligation even after expiry or termination of the contract.

Failure to comply with these requirements shall result in termination of the contract with immediate effect and subject to legal action, non-supply of further data and prompting disciplinary action by the employer.

- The contractor undertakes to make no further use of the information made available to them by Eurostat under the contract after its expiry or termination. Failure to comply with this requirement shall result in liability to claims for damages and to penalties.

Furthermore, at the request of Eurostat, the contractor shall return or destroy all the documents and computer records related to the work to be performed under the contract.

- A copy of all the data, that is the whole CD-ROM containing the users' database, can be made only if **both the contract allows for this and explicit authorisation of the contractor's technical manager is given**. These copies must be made available for free, and may thus not be sold. The list of all the persons that are likely to receive a copy of the CD-ROM and thus the maximum number of copies of the CD-ROM must be given in the contract; only these persons may receive a copy of the CD-ROM.

The contractor's technical manager shall personally ensure that any individual researcher working with the data has signed a confidential statement promising that he/she will abide by all the provisions stated in the contract.

No copy of part of the data can be made.

After expiry or termination of the contract, the contractor will sign a declaration to the effect that he has personally ensured that any ECHP micro-data have been destroyed.

As stated in all contracts on the use of ECHP data, there are also statistical requirements to be followed in any 'publication' (including reports, papers, etc...) of results:

- Two sample size thresholds will be distinguished for ECHP cross-sectional results:
  1. below 20 observations (unweighted sample), results may not be published;
  2. from 20 to 49 observations (unweighted sample), results may be published but are to be individually identified (such as shown in brackets).

For confidentiality reasons, those publications that include sample sizes will only mention 'less than 20 observations' and '20-49 observations' (i.e. not the actual number) for these 2 thresholds respectively.

For unweighted sample sizes below 20 observations, the actual number of observations may not be derived from (combination of) other information available in the publication, e.g. column or row totals.

Exactly the same rule applies for longitudinal results, except that the thresholds are put respectively to 10 and 30 observations linked across time.

- The contractor is required to provide Eurostat with a copy of all reports (including unpublished reports/'papers'...) that have been produced using the data.
- The contractor undertakes to state the source of the data - Eurostat, European Community Household Survey (ECHP) and reference wave(s) - when disseminating the results of the work to which the contract relates.

The Luxembourg courts shall be solely competent to deal with any dispute arising between the contracting parties regarding the terms or performance of the contract.

#### 4. RELATED DOCUMENTS

Doc.PAN. 164 Imputation of income in the ECHP  
Doc.PAN. 165 Construction of weights in the ECHP  
Doc.PAN. 166 ECHP UDB Description of variables  
Doc.PAN. 167 ECHP UDB Construction of variables  
Doc.PAN. 15 Wave 1 variable list and codebook  
Doc.PAN. 30 Wave 2 variable list and codebook  
Doc.PAN. 65 Wave 3 variable list and codebook  
Doc.PAN. 81 Wave 4 variable list and codebook  
Doc.PAN. 97 Wave 5 variable list and codebook  
Doc.PAN. 112 Wave 6 variable list and codebook  
Doc.PAN. 151 Wave 7 variable list and codebook  
Doc.PAN. 159 Wave 8 variable list and codebook  
Doc.PAN. 73 ECHP classifications used in questionnaires  
Doc.PAN. 103 Imputed rent  
Doc.PAN. 105 Anonymisation criteria applied to the users' database

More information is available under

<http://forum.europa.eu.int/irc/dsis/echpanel/info/data/information.html>

**ANNEX 1 - ADDRESSES OF NATIONAL DATA COLLECTION UNITS AND EUROSTAT****EUROSTAT**

Eurostat  
Christine WIRTZ  
5, Rue Alphonse Weicker  
L-2721 Luxembourg  
Tel: (352) 4301 34994  
Fax: (352) 4301 35979  
E-mail: christine.wirtz@cec.eu.int

**Belgium**

Université de Liège - Faculté d'Economie,  
de Gestion et de Sciences Sociales  
Valérie LENOIR  
Bâtiment B31  
Boulevard du Rectorat, 7  
B-4000 Liège  
Tel.: (32) 43 66 21 85  
Fax: (32) 43 66 31 78  
E-mail: valerie.lenoir@ulg.ac.be

UIA Universiteit Antwerpen - Departement  
Politieke en Sociale Wetenschappen  
Rudy MARYNISSEN  
Universiteitsplein 1  
B-2610 Wilrijk (Antwerpen)  
Tel: (32) 38 20 28 79  
Fax: (32) 38 20 28 83  
E-mail: marijn@uia.ua.ac.be

**Denmark**

SFI Socialforskningsinstituttet  
Hans BAY  
Herluf Trolles Gade 11  
DK-1052 København K  
Tel: (45) 33 48 09 50  
Fax: (45) 33 48 08 33  
E-mail: hb@sfi.dk

**Germany**

Statistisches Bundesamt  
Institut für Forschung und Entwicklung in  
der Bundesstatistik  
Daniel RITZ  
Gustav-Stresemann-Ring 11  
Postfach 5528  
D-65189 Wiesbaden  
Tel: (49) 611 75 28 23  
Fax: (49) 611 75 39 50  
E-mail: daniel.ritz@destatis.de

**Spain**

Instituto Nacional de Estadística  
Carmen UREÑA  
Paseo de la Castellana 183  
E-28046 Madrid  
Tel: (34) 1 58 39 298  
Fax: (34) 1 58 39 488  
E-mail: curena@ine.es

**Greece**

National Statistical Service of Greece  
Giorgos NTOUROS  
Head of Unit E - Census Division  
43-45 Agisilaou Street  
GR - 10166 Athens  
Tel: (30) 1 32 89 005  
Fax: (30) 1 32 89 076  
E-mail: geodouro@statistics.gr

**France**

INSEE - Lorraine  
Jean-Christophe RINCENT  
15, rue du Général Hulot  
F- 54029 Nancy Cedex  
Tel: (33) 3 83 91 86 18  
Fax: (33) 3 83 40 45 61  
E-mail: jean-christophe.rincent@insee.fr

**Ireland**

ESRI (Economic and Social Research Institute)  
James WILLIAMS  
Burlington Road 4  
IRL-Dublin 4  
Tel: (353) 1 667 15 25  
Fax: (353) 1 668 62 31  
E-mail: james.williams@esri.ie

**Italy**

ISTAT (Istituto Nazionale di Statistica)  
Francesca GALLO  
Via A. Rava, 150  
I-00143 Roma  
Tel: (39) 6 5952 4749  
Fax: (39) 6 54 30 660  
E-mail: gallo@istat.it

**Luxembourg**

CEPS / INSTEAD  
Uwe WARNER  
Rue Emile Mark (B.P. 48)  
L-4501 Differdange  
Tel: (352) 58 58 55 554  
Fax: (352) 58 55 53  
E-mail: uwe.warner@ceps.lu

**The Netherlands**

CBS (Central Bureau voor de Statistiek)  
Ger LINDEN  
Kloosterweg 1  
P.O. Box 4481  
NL - 6401 CZ Heerlen  
Tel: (31) 45 570 7486  
Fax: (31)  
E-mail: glenn@cbs.nl

**Austria**

ICCR  
Liana GIORGI  
Schottenfeldgasse 69/1  
A-1010 Wien  
Tel: (43) 1 524 13 930  
Fax: (43) 1 524 13 93 200  
E-mail: l.giorgi@iccr-international.at

**Portugal**

INE (Instituto Nacional de Estatistica)  
Victor GARCIA  
Av. Antonio José de Almeida, 5  
P-1078 Lisboa Codex  
Tel.: (351) 1 842 61 00  
Fax: (351) 1 842 63 74  
E-mail: victor.garcia@ine.pt

**Suomi / Finland**

STATISTICS FINLAND  
Marjo PYY-MARTIKAINEN  
FIN-00022 STATISTICS FINLAND  
Tel.: (358) 9 1734 32 53  
Fax: (358) 9 1734 35 62  
E-mail: marjo.pyy-martikainen@stat.fi

**United Kingdom**

ONS  
Caroline LAKIN  
1 Drummond Gate  
UK - London SW1V 2QQ  
Tel: (44) 171 533 57 70  
Fax: (44) 171 533 57 77  
E-mail: tim.harris@ons.gov.uk

ISER (Institute for Social and Economic Research)  
Nick BUCK  
University of Essex  
Wivenhoe Park  
UK - Colchester C04 3SQ  
Tel: (44) 1206 87 30 66  
Fax: (44) 1206 87 21 05  
E-mail: nbh@essex.ac.uk



## ANNEX 2 - SAMPLE WEIGHTS

In the ECHP UDB, two sets of weights (base weights and cross-sectional weights) are available for individuals (in the R- and P-files).

For households, however, only cross-sectional weights are provided in the H-files. This is due to the fact that a household is not a stable unit over time and cannot be analysed longitudinally. Only the individuals can be followed-up over time.

This chapter provides a short description of the weights and provides information on how the weights are to be used for analysis.

For more information on the construction of weights, please see DOC.PAN 165.

### 4.1. A short description of ECHP weights

#### 4.1.1. Base weights (RG003 and PG003)

Sample persons, i.e. persons belonging to the ECHP from the initial wave onwards and new-born children born to sample mothers, are assigned a base weight. The base weight of non-sample persons is 0. These weights are scaled such that their sum over all persons equals the actual number of persons in the sample. This means that the average of the base weights of sample and non-sample persons is 1.

#### 4.1.2. Cross-sectional weights (RG002, PG002 and HG004)

Cross-sectional weights for persons (RG002 and PG002) have been derived from the base weights by sharing the weights of all persons in the same household. These weights are scaled such that their sum over all persons equals the actual number of persons in the sample. This means that the average of the cross-sectional weights is 1.

The cross-sectional household weight (HG004) corresponds to the average weight of its members. It is scaled such that the sum over the interviewed sample equals the actual number of completed household interviews. This means that the average household weight is 1.

### 4.2. How to use ECHP weights for analysis ?

The weights provided in the data files have been 'normalised', i.e. scaled to average 1.0 per unit within each country. This minimises the difference between the weighted and unweighted frequencies. For some analyses, it is however necessary to scale these weights in a different way. How the weights can be used for analytical purposes is described below.

#### 4.2.1. Cross-sectional analysis at country level

For any cross-sectional analysis at country level, the normalised cross-sectional weights can be used. These are:

PG002 The normalised cross-sectional weight for interviewed persons  
 RG002 The normalised cross-sectional weight for persons in interviewed households  
 HG004 The normalised cross-sectional weight for interviewed households

The weights should be scaled differently for the calculation of population aggregates. To do so, an 'inflation factor' has to be applied to the normalised weights. This inflation factor ( $N/n$ ) can be calculated by dividing the country population ( $N$ ) by the actual sample size ( $n$ )<sup>1</sup>. The relevant information on numbers of households and persons in the population has been provided in the 'country file' of the ECHP UDB.

The inflation factors to be applied are

For PG002: "POP16P/number of observations for this country in the P-file"  
 For RG002: "POPTOT/number of observations for this country in the R-file"  
 For HG002: "POPHHD/number of observations for this country in the H-file"

#### 4.2.2. Cross-sectional multi-country analysis

For multi-country analysis, the same weights as for the calculation of population aggregates should be used.

#### 4.2.3. Longitudinal analysis at country level

Longitudinal analysis over several waves would normally be confined to persons interviewed in all these waves or to persons living in an interviewed household in all these waves.

For longitudinal analysis upto a wave  $i$ , the normalised base weights of wave  $i$  can be used. These are:

PG003 The normalised base weight for interviewed sample persons  
 RG003 The normalised base weight for sample persons in interviewed households

For the calculation of population aggregates the base weights should be scaled in the same way as cross-sectional weights. To do so, an 'inflation factor' has to be applied to the normalised weights. This inflation factor ( $N/n$ ) can be calculated by dividing the country population ( $N$ ) by the actual sample size ( $n$ )<sup>2</sup>. The relevant information on numbers of households and persons in the population has been provided in the 'country file' of the ECHP UDB.

---

<sup>1</sup> The factor can be scaled for numerical convenience, such as to  $(0.001*N/n)$  to produce aggregates in '000s.

<sup>2</sup> The factor can be scaled for numerical convenience, such as to  $(0.001*N/n)$  to produce aggregates in '000s.

The inflation factors to be applied are

For PG003: "POP16P/number of observations for this country in the P-file"

For RG003: "POPTOT/number of observations for this country in the R-file"

#### *4.2.4. Longitudinal multi-country analysis*

For multi-country analysis, the same weights as for the calculation of population aggregates should be used.

#### *4.2.5. Advanced uses of the weights*

The ECHP weights have been constructed by taking into account non-response patterns and by calibrating to external control distributions. For specific analyses it might be appropriate to re-calibrate the weights for some subpopulations. This is however upto the descretion of the researcher analysing the data.

For longitudinal purposes, it is in principle possible to modify the base weights to take into account reduction in the sample due to incomplete longitudinal matching, especially if it is selective by age, sex, activity status and other relevant characteristics. It is believed however that, for example, the base weights for 1996 would normally suffice for longitudinal analysis over 1994-1996. Therefore, no specially constructed 'longitudinal weights' other than the base weights have been included in the ECHP UDB.

## ANNEX 3 - EXAMPLES OF CALCULATIONS

### 1. Examples on the usage of weights when analysing ECHP results (Cross-sectional weights)

#### ***Example 1.1: Results relating to the household as a whole, broken down by characteristics of the household***

Whenever the household variables are analysed alone, the household weights HG004 (available in the H-file) are used. If the units for analysis are the individuals, the household weight is to be multiplied by the size of the household. If the units for analysis are the households, the weight can be used unmodified.

#### ***Example 1.2: Results relating to the household, broken down by sex and/or age of individuals living in the household***

For such results, two ways are possible:

Option 1: The household information is merged to the file that contains one record per person (R-file). The personal weights RG002 in this file are to be used.

Option 2: Two step approach: First, the number of persons falling into the group of interest (e.g. females aged 40-60 years) is calculated from the register file (R-file). Second, the household weight HG004 is multiplied with this number.

#### ***Example 1.3: Results relating to the household or to individuals, broken down by personal characteristics obtained in the personal interview***

For such results, weights PG002 from the personal file (P-file) are to be used. [Remark: During the construction of these weights, non response for some classification variables has been taken into account. Thus, the use of P-file weights yields better results than the use of R-file weights.]

Attention: For any results that are independent from information obtained during the personal interview, the P-file weights should not be used.

	• Household information	
<i>Breakdown by ILO current activity status</i>	<i>Results = Mean equivalised total income</i>	
At work	Weighting according to 1.3	• Information obtained during the personal interview
Unemployed		
Retired		
Other economically active		
<i>All persons aged 16+</i>	Weighting according to 1.2	• Information independent from the personal interview

## 2. Examples of calculations using the ECHP

### ***Example 2.1: 'Basic' calculations***

#### *Question:*

Average number of individuals living in different types of households

#### *Solution:*

Calculate the average of variable HD001 (= household size = number of individuals living in the household) for each group of COUNTRY x HD006 (or HD006A) using as weight the variable HG004.

#### *Question:*

Percentage of individuals living in different types of households

#### *Solution:*

Merge the variable HD006 (or HD006A) from the H-file to the R-file.

Calculate the percentage of individuals for each group of COUNTRY x HD006 (or HD006A) using as weight the variable RG003.

**Example 2.2: Calculation of the 'at risk of poverty rate (after social transfers)'****Definition**

The percentage of the individuals living in households where the total equivalised income is below 60% of the national median equivalised income. Note that the definition of income includes all social transfers received.

This is a relative measure of poverty in monetary income terms. Note that this is the official 'at-risk-of-poverty rate' reported by Eurostat.

**Dataset to be used:**

All the information for calculating these indicators is stored in the H-file of the ECHP UDB.

**Calculation of equivalised income**

The total net income of each household is available in the variable HI100 (it covers the income received by all the members of the household from all sources).

For each person, the 'equivalised total net income (EQ\_INC)' is calculated as its household total net income divided by equivalised household size according to the modified OECD scale (which gives a weight of 1.0 to the first adult, 0.5 to other persons aged 14 or over who are living in the household and 0.3 to each child aged less than 14). In terms of ECHP variables, EQ\_INC is calculated as  $HI100/HD005$ .

Consequently, each person in the same household receives the same 'equivalised total net income'.

The population consists of all the persons living in private households of a country. The term person therefore includes all the members of the households, whether they are adults or children.

Persons with missing 'equivalised total net income' are excluded from the calculations (ie. people with missing household income (HI100) or households with missing composition details (HD005)).

**Weights**

For each household a household weight (HG004) is available in the H-file. In order to describe distributions of persons, this weight has to be multiplied by the number of persons belonging to the households (HD001). The result is the person weight ( $WTPERS = HG004 * HD001$ ) of the household.

**Calculation of the 'at-risk-of-poverty threshold'**

Firstly, persons have to be sorted according to their 'equivalised total net income' (sorting order: lowest to highest value).

Secondly, the median is calculated as the equivalised income of the household person for whom the cumulative sum of personal weights (WTPERS) is less than or

equal to 50% of the total sum of weights. In other words, persons in the same household are located together, on the same side of the median.

Thirdly, the 'at-risk-of-poverty threshold' is calculated as 60% of the national median.

$$\text{At risk of poverty threshold} = 60\% * EQ\_INC_i \Big|_{i=\text{person for whom the cumulated sum of weights} = 0.5 * \text{total sum of weights}}$$

To illustrate the threshold value for a one person household and for a household consisting of two adults and two children, the 'at-risk-of-poverty threshold' has to be multiplied:

\* by 1 (for a one person household) ;

\* by 2.1 (for a household consisting of two adults and two children). The factor 2.1 is obtained by reference to the 'modified-OECD equivalence scale' as the sum of 1(first adult) + 0.5(second adult) + 0.3 \* 2 (the two children).

#### Calculation of 'at-risk-of-poverty rate (after social transfers)'

The 'at-risk-of-poverty rate (after social transfers)' is calculated as the percentage of persons with an equivalised net total income below the 'at-risk-of-poverty threshold'.

$$\text{At risk of poverty rate (after social transfers)} = \frac{\sum_{\text{All persons: } EQ\_INC < \text{at risk of poverty threshold}} wtpers}{\sum_{\text{All persons}} wtpers}$$

#### Calculation of the EU average

For calculating the EU averages, population figures from the CTRYVARS-file are to be used. The EU average of the 'at-risk-of-poverty rates (after social transfers)' established for each individual country is calculated as a weighted average of the country rates, where the weighting of countries is done according to the number of persons living in private households (POPTOT) in each country.

$$\text{EU average of 'at risk of poverty rate'} = \frac{\sum_{\text{all countries}} ('at risk of poverty rate'_i * POPTOT(\text{year})_i)}{\sum_{\text{all countries}} (POPTOT(\text{year})_i)}$$

*year = year of the survey*