

Living in Ireland Survey – Technical Overview

Dorothy Watson
October 2004

This document provides a brief technical overview of the Living in Ireland Surveys and of the dataset lodged with the Irish Social Science Data Archive. Section 1 describes the Living in Ireland Surveys, while Section 2 focuses on the structure of the data files.

Section 1: The Living in Ireland Surveys

The Living in Ireland Surveys form the Irish component of the European Community Household Panel (ECHP): an EU-wide project, co-ordinated by Eurostat, to conduct harmonised longitudinal surveys dealing the social situation, financial circumstances and living standards of European individuals and households. The fact that the same set of households is interviewed each year means that it is possible to study changes in the characteristics and circumstances of particular households or individuals over time. The ECHP provides harmonised cross-sectional surveys for each year in which the survey is conducted, as well as longitudinal data, which permits dynamic analysis of changes over time.

The first wave of the ECHP was conducted in 1994, and the same individuals and households were followed each year. The survey ran for eight waves, until 2001. Twelve countries participated in 1994, with Austria and Finland joining in 1995 and 1996, respectively.

In 2000, the seventh wave, the Irish sample of individuals and households followed from Wave 1 was supplemented by the addition of 1,500 new households to the total. This was done in order to increase the overall sample size, which had declined due to attrition since 1994. A larger sample size ensures that the precision of estimates of key figures, such as the poverty rate and average equivalised household income, remained at a high level. It also allows a greater disaggregation of the data so that the situation of policy-relevant sub-groups, such as the unemployed or older adults, can be examined. These additional households, as well as the original sample, were followed in 2001.

Survey Structure

The Living in Ireland Survey involves a household questionnaire which is completed by the 'reference person' or person responsible for the accommodation, and an individual questionnaire which is completed by each adult (age 16 or over) in the household. The main items of information collected on the household questionnaire are shown in the top panel of Figure 1.

The individual questionnaires in 1994 were administered to each member born in 1977 or earlier. This 'cut-off' year was updated in each wave of the survey, so that it by 2001 all household members born in 1984 or earlier were eligible for individual interview. The main items of information collected on the individual questionnaire are shown in the lower panel of Figure 1.

Figure 1: Topics covered in the Household and Individual questionnaires of the Living in Ireland Surveys

<i>Household Questionnaire Topics</i>
Household size and composition
Housing and physical environment
Housing tenure
Rent and mortgage payments
Standard of living (things the household can afford to have or to do)
Debts and arrears
Sources of household income
Non-cash and secondary benefits

<i>Individual Questionnaire Topics</i>
Current activity status (self-defined)
Detailed information on the current job, for those working 15 or more hours per week in a job or business
More limited information on work for those working less than 15 hours per week
Some information on previous job, for those not currently working 15+ hours per week
Job search activity, for those seeking work
Other daily activities, such as caring responsibilities, social and political participation
Recent involvement in education and training
Activity in each month since the beginning of the previous calendar year
Detailed information on income in the previous calendar year from employment, self-employment, personal and occupational pensions, social welfare, education and training-related allowances and grants, property (interests, dividends, rental income), and other sources.
Health status, health service usage, and health care coverage
General outlook on life

The questionnaires were administered in a face to face interview by the ESRI's team of interviewers. On average, the household questionnaire took 12 minutes to complete, while the individual questionnaire took 30-35 minutes to complete. The average number of individual interviews per household in 1994 was 2.4.

In farm households, a farm questionnaire was also completed to collect information on the acreage farmed, and the profile of the farm in terms of crops grown, other land use, livestock held, and labour input. This information, together with data on the soil type, was used in conjunction with Teagasc's National Farm Survey to estimate the income flow (family farm income) of farm households. This approach was necessary because the nature of farm income – being a combination of market profit or loss, grants and subsidies – makes it difficult for respondents to provide the figure directly.

The core ECHP questionnaire has remained substantially the same since 1994. Modifications of the Irish Questionnaire have been kept to a minimum, but with some modules added to meet national needs.

The Living in Ireland Sample

The sample of households was originally selected for the 1994 wave of the survey. The objective of the sample design was to obtain a representative sample of private households in Ireland. Those living in institutions such as hospitals, nursing homes, convents, monasteries and prisons, are excluded from the target population, in line with the harmonised guidelines set down by Eurostat and standard practice adopted in surveys of this kind (such as the

Household Budget Survey conducted by the Central Statistics Office¹. Among those effectively excluded from the target population are a number of small groups that face a relatively high risk of poverty – such as the homeless and travellers. To do justice to the particular circumstances of groups such as these would require a different research methodology.

The sample was selected using the ESRI's RANSAM system, which was developed at the institute and has been successfully used for selecting random samples from the electoral register for over two decades. The sampling strategy allows one to pre-stratify the sampling frame according to a number of socio-demographic criteria. In selecting the sample for the Living in Ireland Survey the following strata were used:

- Province: Four categories, Dublin; rest of Leinster; Munster; Ulster/Connacht.
- Urban/Rural: Two categories: DEDs with more than 50 per cent of their population in towns with a population of 1,500 or more, versus the rest;
- Unemployment: Two categories: DEDs with an unemployment rate of 16 per cent or more versus the rest.

The target sample selected using the ESRI's RANSAM procedure was a sample of *persons*, not of *households*. Since the probability of selection is greater for households with a larger number of registered voters, this means that the resulting sample will tend to over-represent larger households. This was taken into account in reweighting the sample for analysis.

The total number of households successfully interviewed in 1994 was 4,048, representing 57 per cent of the valid sample. This response rate is as one would expect in an intensive and demanding survey of this nature, and is comparable to the response rates achieved in the Household Budget Surveys.

A total of 14,585 persons were members of the completed households. Of these, 10,418 were eligible for personal interview (i.e. born in 1997 or earlier), and 9,904 eligible respondents completed the full individual questionnaire (964 on a proxy basis). Summary details were collected on the household questionnaire on the 514 eligible individuals for whom no individual interview was obtained.

Re-interviewing households and tracing Individuals who move

The sample from the Wave 1 (1994) Living in Ireland survey was followed in subsequent years and re-interviewed. All 'sample persons' were to be followed in the survey. A sample person is someone who was a member of a household successfully interviewed in 1994. The follow-up rules for the survey meant that new households might be included in each wave where a sample person from Wave 1 moved to another household. All individuals in the Wave 1 sample were to be followed in Wave 2 and household and individual interviews were to be conducted, as long as the person still lived in a private or collective household within the EU. Table 1 summarises the wave-on-wave response rates, from Wave 1 to Wave 8.

¹ Collective households, however, were included. These are private households containing five or more unrelated persons with a looser budget-sharing relationship than in the standard private household. The main examples are boarding or lodging houses and army barracks. An individual moved living in a collective household is treated as a one-person 'sub-household'.

Table 1: Number of Completed Households in Each Wave, Number Sample Persons in Completed Households and Number Interviewed, Living in Ireland Surveys 1994-2001.

	1994	1995	1996	1997	1998	1999	2000 Original sample	2000 New sample	Total 2000	2001
<i>Households</i>										
Completed Households	4048	3584	3174	2945	2729	2378	1952	1515	3467	2865
Non-Response	3038	794	624	390	394	464	414	1146	1560	797
Total Households	7086	4378	3798	3335	3123	2842	2366	2661	5027	3662
Household Response Rate	57%	82%	84%	88%	87%	84%	83%	57%	69%	78%
Non-Sample	166	98	125	119	94	83	77	159	236	78
<i>Individuals</i>										
N. in Completed Households:	14585	12576	10889	9952	9000	7721	6276	5174	11450	9131
... followed from first wave	N/A.	12117	10241	9154	8139	6908	5530	0	5530	4820
... new since first wave	N/A.	459	648	798	861	813	746	5174	5920	4311
Eligible for Individual Interview (*)	10418	9048	7902	7255	6620	5719	4745	3952	8697	6996
Number individuals interviewed	9904	8531	7488	6868	6324	5451	4529	3527	8056	6521
Of these, N. interviewed all waves since wave 1	N/A	7942	6636	5782	5124	4329	3391	0	3391	2948
% Individual interviews completed	95%	94%	95%	95%	96%	95%	95%	89%	93%	93%

Note: * in completed Households

The total number of households eligible for inclusion in Wave 2 was 4,376, excluding non-sample households but including newly generated households². In the second wave, 3,584 household interviews were completed and 794 did not respond. The household response rate was 82 per cent. The 3,584 completed households contained 12,576 persons (12,117 from wave 1 and 459 new individuals), of whom 9,048 were eligible for individual interview (born in 1978 or earlier) and 8,531, or 94 per cent, were actually interviewed.

In the next four waves, the household response rate remained in the range from 84 to 88 per cent.

Sample Supplementation in 2000

As shown in Table 1, even with a relatively high year-on-year response rate, there was a substantial loss of respondents over time. Of the original sample individuals who were still 'in scope' in 1999 (13,964)³, only 49 per cent (6908) were in completed 1999 households, with another 813 individuals having joined the sample households at some point in the intervening years. By 2000, 5530 of the 13861 individuals still in scope (40 per cent) were in completed households.

² Generated households include (a) households generated when someone from a Wave 1 household moves out to set up a new household or (b) pre-existing households that a mover from a wave 1 household had joined by 1995. Non-sample households are those where all members are deceased, moved to an institution or outside the EU, or households from which the 'sample person' – someone who was in one of the original households in wave 1 – has moved out.

³ Of the original 14,585 individuals, 339 had died and a further 282 had moved to an institution or outside the EU by 1999. A total of 400 had died by 2000, and 324 had moved to an institution or outside the EU. This left 13861 individuals still 'in scope' by 2000.

The main reason for household non-response was refusal (ranging from 9 per cent of the eligible sample in Wave 2 to 5 per cent in Wave 5). Among the newly generated households, difficulties in obtaining forwarding addresses for those who moved also contributed to non-response.

Attrition of this magnitude is of concern for two reasons:

- To the extent that attrition is not random, it may result in a loss of representativeness in the resulting sample
- The reduction in the number of completed surveys leads to a loss of precision in the estimates derived from it.

Detailed checks on the pattern of attrition between waves of the survey are reported elsewhere (Nolan et al., 2002). In brief, these analyses suggested that the main loss was related to difficulties in tracing households that had changed address: primarily households consisting of young single adults. There was no evidence of a disproportionate loss of households from the upper or lower ends of the income distribution of the kind that would tend to bias estimates of average household incomes or poverty measures.

However, the reduced sample size still needed to be addressed. This was achieved by supplementing the sample in Wave 7. The new sample was selected using the same procedure as for the first wave of the survey in 1994, using the ESRI's RANSAM programme, based on the electoral register. The household response rate reached 57 per cent for the 2,661 new sample households contacted by interviewers. This is the same as the rate achieved in Wave 1 and is in line with the typical response rate in other surveys of a demanding nature, such as the Household Budget Survey.

A new sample of just over 1,500 completed households was added to the sample continuing from previous waves of the survey. Of the households followed from the previous year (the 'continuing sample'), 2,443 were issued to interviewers and interviews were completed in 1,952 of these (83 per cent) and with 4,745 individuals (95 per cent of those eligible). The improved economic situation in 2000 made it more difficult to establish contact with, and secure participation of, the households followed from earlier waves, since a higher proportion of the household members were working outside the home. This increased the number of call-backs required to make the initial contact and, since respondents had less free time, made refusals more likely. In addition, it created a challenge in that several experienced interviewers moved to alternative employment. A total of 290 households that had completed the Wave 6 interview could not be issued to interviewers because of these difficulties.

The sample supplementation exercise, together with the follow-up of continuing households, resulted in a completed sample in 2000 of 11,450 individuals in 3,467 households. Individual interviews were conducted with 8,056 respondents, representing 93 per cent of those who were eligible (born in 1983 or earlier).

The 2001 dataset includes 9,131 individuals, 4,820 of whom were followed from 1994 and 4311 who joined the sample since then – most of them being added when the sample was supplemented in 2000. The response rate at the household level was 78 per cent, a lower completion rate than had been achieved throughout the 1990s. The same factors that posed a

challenge to fieldwork in 2000 affected the process of data collection in 2001: higher workforce participation, making households more difficult to contact (and busier) and interviewers more difficult to recruit. Nonetheless, where the household participated in the survey, 93 per cent of adult household members were successfully interviewed⁴, resulting in 6521 personal interviews.

Cross-Sectional Sample Weights

The purpose of sample weighting is to compensate for any biases in the distribution of characteristics in the completed survey sample compared to the population of interest, whether such biases occur because of sampling error, from the nature of the sampling frame used, differential response rates or attrition.

Whatever the source of the discrepancy between the sample and population distributions, we would like to adjust the distributional characteristics of the sample in terms of factors such as age, sex, economic status and so on to match that of the population. In a cross-sectional survey, or in the first wave of a panel survey, the only way to check the distributional characteristics of the sample is to compare sample characteristics to external population figures from sources such as the Census, the Labour Force Survey, official statistics on number of social welfare recipients from the Department of Social Welfare, and so on. In waves following the first wave of a panel survey, we can also compare the characteristics of the individuals and households successfully followed to those of the individuals and households in a previous wave of the survey. In constructing the weights for the Living in Ireland Survey in Waves 2 and subsequently, both of these methods were used.

The household weights were developed in a number of steps:

- The first step involved adjusting the continuing sample for attrition.
- The second step was to calibrate the sample totals against population totals from external sources.

The external sources of information used were the Quarterly National Household Survey, the Department of Social Community and Family Affairs statistics on social welfare reciprocity levels, and figures from Teagasc on the total number of farms by farm size. The result of the weighting procedure was to ensure as close as possible a match between the sample and the population in terms of the distribution of the characteristics shown in Figure 2.

Apart from incorporating weights to control for attrition from previous waves, and the availability of new technology for constructing weights, the logic and general strategy in developing the weights for Waves 2 to 7 was very similar to that used in Wave 1. The resulting match between the weighted sample characteristics and the population characteristics for the 2001 data was highly satisfactory, confirming that the weights are effective in adjusting the achieved sample to population characteristics. Further details on the construction of sample weights are reported in Nolan et al. (2002).

⁴ As in previous years, summary information on age, sex, income, level of education and marital status was obtained on the household questionnaire for eligible adults who were not interviewed in person.

Figure 2: External Population Characteristics used in the Construction of Household Weights

Household characteristics:

- Household size (total size, number over 18 and number over 65)
- Location (Dublin, other county borough, rural)
- Number of persons at work (0, 1 and 2 or more)
- Head Age (under 25, age 25 and over)
- Number of farms in each of six size categories

Individual characteristics

- Number of males and females by 10 age categories
 - Number of males and females age 15+ by 11 age/marital status categories
 - Number of recipients of 12 major social welfare payments
 - Number of males and females by 7 economic status categories (at work (ILO), unemployed (ILO), Unemployed (not ILO), Student, home duties, retired, other)
 - Number of males and females age 20-64 by level of education (4 categories)
-

Cross-Sectional Individual Weights

As well as household weights, for use in analysing household-level data or data at the level of all individuals in a household, weights are provided for analyses of the individual questionnaires. If all individuals who were eligible for individual interview actually completed an individual questionnaire, then there would be no need for an additional set of weights (the household weights would serve for analyses at both the individual and the household level). However, since some eligible (i.e. over age 16) individuals in completed households did not complete an individual questionnaire, weights are needed to ensure that the completed individual questionnaires are representative of all eligible individuals.

- The approach in constructing the individual weights was as follows:
- The cross-sectional household weight is taken as the initial weight

The distribution of a range of characteristics of the subsample of individuals completing an individual questionnaire is grossed to that of the weighted⁵ sample of all eligible individuals. The variables used are shown in Figure 3.

Since the individual questionnaire completion rate was very high (generally in the region of 93-95 per cent), the adjustment needed tended to be minor.

⁵ With the household weight.

Figure 3: Control Characteristics Used in Construction of Individual Weights

Individual Characteristics

- 1) Age group (ten groups) by sex.
- 2) Age group by marital status by sex
- 3) Economic Status (self-defined) by sex.
- 4) Socio-economic by sex.
- 5) Education by sex.
- 6) Receipt of social welfare payments
- 7) Whether individual changed address between start wave and end wave
- 8) Whether individual was in a split household at any time between start and end wave

Household Characteristics

- 9) Household Size (6 categories)
- 10) Number of adults in household (6 categories)
- 11) Number elderly – age 65 or over – in household (none, one, 2 or more)
- 12) Number working 15+ hours in household (none, 1,2, 3 or more)
- 13) Urban/rural location of household in start wave (Dublin, Other Urban (10k+), Rural)
- 14) Poverty status of household in start wave (Scale A, 50% line)
- 15) Household equivalised income decile in start wave
- 16) Net household income decile in start wave

Characteristics of Reference Person

- 17) Sex
- 18) Age group
- 19) Marital status
- 20) Economic Status (self-defined)
- 21) Socio-economic group

Longitudinal Weights for Individual Questionnaires

As well as cross-sectional weights, longitudinal weights are provided for analysis of individual questionnaires over time.⁶ The purpose of longitudinal weighting is to ensure that the longitudinal sample (the subset of cases on whom information is available for two or more waves) is representative of the population as a whole. The longitudinal weights will differ from the cross-sectional weights to the extent that:

- i. Some individual have been lost from the sample between wave t and wave t+1
- ii. New individuals have been added to the sample between wave t and wave t+1.

Both of these have taken place in the Living in Ireland Surveys. Individuals are lost from the sample through death, leaving the EU, moving to an institution, household non-response and individual non-response. Individuals are added to the sample by becoming eligible (reaching 16 years of age, moving into a sample household) or where households are added because a sample person moves out of an original household to live with another person or persons. The longitudinal subsample excludes individuals for whom information is not available in all of the relevant waves.

⁶ Longitudinal weights are provided at the individual level (for adults age 16 and over) only, since the composition and character of households changes over time.

Choice of Base

In constructing longitudinal weights we have to choose which of the comparison years to base the weights on. That is, we need to decide whether to gross the longitudinal sample to the population totals in the beginning wave, or in the ending wave. As long as the time period is relatively short, there would be little difference in practical terms between the two strategies, since the population totals would have changed very little. The position adopted here is to gross the common subsample to the population totals for wave 1 (or the earlier wave involved in the comparison). This approach is closer to the logic of most longitudinal analysis: we generally want to look at future outcomes of experiences (unemployment, training, education, poverty status) at a particular point in time, rather than at the historical experiences leading to a particular outcome. This means, for instance, taking the cases who were unemployed in Wave 1 and looking at their employment status, psychological well-being etc. in Wave 2 to look at the impact of unemployment; rather than taking the subset of cases who were unemployed or psychologically distressed in Wave 2 to examine their employment status etc. in Wave 1.

Strategy

The strategy adopted in devising the longitudinal weights is to begin with the longitudinal sample available for each subset of waves (Wave 1-2; Waves 1, 2 and 3; Waves 1,2 3 and 4 and so on); then to re-gross the longitudinal sample totals on a range of control variables (using the cross-sectional individual weights from the earlier wave as initial weights) to the population totals for the earlier wave.

Note on Longitudinal Sample

Note that the longitudinal sample will not be representative of the population in the end year of the comparison. For instance, the Wave 1-4 longitudinal sample will not be representative of the adult population in 1997 because

- (a) it does not include individuals who were too young for individual interview in 1994, although they might be age 16 or over in 1997
- (b) the weighting scheme treats those who died (as well as those in non-completed households, or those moving outside the EU) as non-respondents so that the weighted sample contains more elderly persons than were actually in the population in 1997.

Table 2 compares the size of the longitudinal sample (the sample of individuals interviewed in all waves since wave 1) to the size of the cross-sectional sample of individual interviews.

Table 2: Size of the Longitudinal Sample (Individual Interviews)

	1994	1995	1996	1997	1998	1999	2000 Original sample	2000 New sample	Total 2000	2001
Number individuals interviewed	9904	8531	7488	6868	6324	5451	4529	3527	8056	6521
Of these, <i>N. interviewed all waves since wave 1</i>	<i>N/A</i>	7942	6636	5782	5124	4329	3391	0	3391	2948

The size of the longitudinal individual interview sample declines with the span involved. The smallest sample (2948) is for the cases interviewed individually in all 8 waves: here the cases

that are new in waves after Wave 1 (including the supplement added in 2000) are excluded, as well as those missing out on an individual interview in any of the 8 waves.

The weighting procedure weights the characteristics of the ‘common cases’ in the second row of Table 2 to the characteristics of the cases in the first row. The same list of control variables were used as for the construction of the individual weights (see Figure 3).

Section 2: The Archived Data Files

Anonymisation

Before making the *Living in Ireland Survey* data available for analyses through the *Irish Social Science Data Archive*, the data were carefully checked to remove all identifying information and to limit the amount of detail provided on certain key variables. Where this was done, additional constructed variables were added to the files in order to maximise their usefulness. The variables where anonymisation was undertaken are shown in Table 3.

Table 3: Variables Anonymised in Constructing the Living in Ireland Data Files

Variable	Anonymisation
Year of birth	Bottom coded: 1909=1909 or earlier
Occupation	Reduced detail; additional constructed variables provided
Industry	Reduced detail: NACE 1-digit code provided
Weekly income variables	Top coded (2000=£2,000 or more per week)
Value of accommodation, business	Top coded: 500000=\$500,000 or more
Additional payments (holiday bonus, profit sharing etc)	Top coded: 10000 = £10,000 or more per year
Additional earnings from overtime, commission, tips	Top coded: 20000 = £20,000 or more per year
Monthly rent/mortgage payments; Monthly rent or mortgage supplement	Top coded: 2000 = £2,000 or more

In practice, only a small number of individuals were affected by the top-coding of variables such as income. For many variables (e.g. the social welfare incomes), there were no cases where the incomes were as high as the maximum code.

For the small number of households with more than ten persons, all records in the Link file, Register files and Individual Files related to persons with a line or person number greater than 10 were deleted. This affected only 41 individuals of over 21,500 in the link file, and resulted in the deletion of only 6 individual questionnaire records.

The Main Data files

There are five sets of data files provided through the *ISSDA*. They are described briefly here. Full details of the contents are available in the accompanying codebook.

Link (LINK): This file contains one record for every person who appears in a household included in the Living in Ireland Survey. It shows the person's sex, date of birth and the ID of the household of which he or she was a member in each wave. It is used to match information on the individual across waves of the survey. The longitudinal weights for interviewed adults are included on the link file. It is sorted by the person's unique fixed id number (*zpfxid*).

Register (REG): This file contains a record for every person in a sample household in each wave. Note that movers appear twice: once associated with the household they move out of, and once associated with the household they move into. The household of which a person is currently a member can be identified using the *Zmemb* variable. There is a register file for every wave of the survey. The register file is sorted by the unique household id (*zhstdid*) and the line number (*zlineno*).

Household (HSD): This file contains household level information on completed households in each wave. There is a separate file for each of the eight waves of the survey. The file is sorted by the unique household identifier *zhstdid*.

Individual (IND): This file contains a record for every individual age 16 and over who completed an individual questionnaire. There is an individual file for every wave of the survey. The file is sorted by *zhstdid* and *zlineno*.

OTHER DATA ISSUES

Some notes on file contents

Household Files

Household files include only households that completed a household interview.

Register Files and Non-Household Members

Register files contain two records for movers: one in the household the person moved out of and one in the household the person moved into. To exclude the record in the household the person moved out of, select on the basis that *zmemb* not equal 4.

Register files also contain a record for persons who died since the last wave. These can be excluded on the basis that *zmemb* not equal 5.

To select current household members in the register file then:

Select if not any(*zmemb*,4,5).

Weights

The weights included on the files compensate for any departures from representativeness along a number of key dimensions (age, sex, marital status, labour force situation, education, household size). The weights provided on the files are designed for this purpose and also gross the sample numbers up to the corresponding population figures.

As discussed in the previous section, there are two basic kinds of weights: cross-sectional weights and longitudinal weights. The cross-sectional weights are designed to ensure that the sample is representative of the population at a particular point in time. They should be used when only one wave of data is being analysed. The

longitudinal weights are designed to gross the longitudinal sample (the sample of those for whom information was collected in more than one wave) up to the population at the time of the initial wave. Cross-sectional weights are provided at both the individual and household level for each wave.

Longitudinal weights are available at the individual level only. A very large number of different longitudinal weights is possible, depending on the waves selected for analysis. For example, there is a different number of cases available when comparing Waves 1 and 2, Waves 2 and 3 or all Waves from 1 to 8. A number of key longitudinal weights are provided for analysis of data from the individual questionnaires. These allow individuals interviewed in all waves from wave 1 to wave 8; individuals interviewed in all waves from wave 1 to wave 4; and individuals interviewed in both wave 7 and wave 8.

The weights included on the files and details of when they should be used are shown below.

Figure 4: Weights included in the Living in Ireland Data Files

Weight	Location	Use
W1to2iw	Link file	Longitudinal Weight for analyses of individuals completing an individual interview in both Waves 1 and 2 (Largest number of cases for longitudinal analysis; but short timespan)
W1to3iw	Link file	Longitudinal Weight for individuals completing an individual interview in all waves from 1 to 3.
W1to4iw	Link file	Longitudinal Weight for individuals completing an individual interview in all waves from 1 to 4.
W1to5iw	Link file	Longitudinal Weight for individuals completing an individual interview in all waves from 1 to 5.
W1to6iw	Link file	Longitudinal Weight for individuals completing an individual interview in all waves from 1 to 6.
W1to7iw	Link file	Longitudinal Weight for individuals completing an individual interview in all waves from 1 to 7.
W1to8iw	Link file	Longitudinal Weight for individuals completing an individual interview in all waves from 1 to 8. (Longest timespan; but smallest number of cases)
W7to8iw	Link file	Longitudinal Weight for individuals completing an individual interview in both Waves from 7 and 8 (Short period, but allows inclusion of 2000 Sample Supplement)
hsdwgtyy	Household Register	Household Cross-Sectional Weight in year yy (yy=94-01). Use hsdwgt94, for example, when analysing data on households in 1994.
indwgttyy	Individual Register	Individual Cross-Sectional Weight in year yy (yy=94-01). Use indwgt94, for example, when analysing data from individual interviews in 1994.

Constructed Income Variables

The main constructed income variables on the file are for current income, although a number of variables are also provided which show income for the previous calendar year. Total Gross Current Weekly Household Income from all Sources (ZHINCG) has been added to the household file. ZHINCG has been constructed from income components collected on the household and individual questionnaires. The corresponding net figure (ZHINCN) is also included. This represents gross income minus tax and social insurance contributions of all household members.

Income components (such as self-employment income, rental income, interest income etc.) that are recorded on an annual basis have been converted into weekly amounts. Information on the incomes of adults who were not interviewed is also taken into

account, so the household gross and net total income components will not equal the sum of the total incomes on individual questionnaires where there is an adult who did not complete an individual questionnaire.

Identification Numbers and Matching Files

Identification numbers (ID numbers) are needed to match individuals to their households and to link information on a given person from different waves of the survey. There are two unique identifiers: the Household ID (*ZHsdID*) and the Person’s fixed id (*PFxID*). Two supplementary identifiers are the line number (*Zlineno*) and the person number (*Zperno*). These two (*Zlineno* and *Zperno*) are not unique across waves.

Each household is given a unique identification number (*Zhsdid*) when it first appears in the sample, e.g. 4991501. The last two digits of the *zhsdid* form the split code (e.g. ‘01’ in this example). This is initially set to 01 for households selected into the sample in wave 1 (or in the wave 7 supplement). The first part of the *zhsdid* is the ‘core’ (e.g. 49915 in this example).

Each individual in the household has a two-digit line number (*Zlineno*) and a two digit person number (*Zperno*). The line number is simply the line in the Household Register on which the person is listed. In the first wave, the household reference person (HRP – the person responsible for the accommodation or the older of two or more responsible persons) was listed on the first line of the register and given a person number of 01.

Table 4: Example Household with Zperno, zpfxid, and zhsdid, Wave 1

Zhsdid=4991501 (Coleman Household)				
Zlineno	Zperno	(Name)	DOB	zpfxid
1	1	Elaine	05/09/54	49915101
2	2	John	02/10/56	49915102
3	3	Sean	10/3/79	49915103
4	4	Eithne	12/05/84	49915104
5	5	Emer	14/07/85	49915105
6	6	Joseph	09/09/76	49915106

Note: all identifying information has been removed from the data file. The fictional household in the example has names included for ease of illustration.

In the Example in Table 4, the HRP is Elaine. The *zpfxid* is constructed using the core part of the *zhsdid* (49915), the split code with the zero dropped (1) and the person number of the individual in the household in which they were first drawn into the sample (01). Person numbers (*Zperno*) are assigned sequentially. *Zperno* is unique to that individual within that household. However, if someone moves to a different (usually split-off) household, a new *zperno* is assigned. If a new person were to join the Coleman household, he or she would be given the next available person number: ‘07’. In the first wave, the line number and person number are identical.

Generated Households and ID numbers

A household splits when a household member moves out and either sets up a new household or joins another existing household. When a sample person (a member of any wave 1 household) moves out, he or she is ‘traced’ to the newly generated household. When this happens, the ‘original household’ (the one remaining at the same address) retains the split code ‘01’. The ‘generated’ household (the mover and all members of the new household) retains the core part of the *zhssid* (e.g. 49915) and gets a new split code (e.g. ‘02’). Hence, household 4991402 is a generated household linked to 49901 by virtue of the fact that somebody from 49901 moved out to 49902. For example, let’s assume that Joseph moves out of the household from wave 1 and marries. The original household retains the same *zhssid* as in the first wave. (Joseph will still appear on the register, but will be identified as someone who has moved out by *zmemb*=4.) The generated household, consisting of Joseph and his wife Ann is illustrated below. The fixed household id is 4991402.

Note that Joseph has a new *zperno* in this household (01, because he is on line number 1). Since Ann is new to the sample, her unique fixed id number (*zpfid*) is constructed from the *zhssid* (4991502) and her person number (02): 49915202.

Table 5: Example: Generated Household with Zperno, zpfid, and zhssid, Wave 2

Zhssid=4991502				
Zlineno	Zperno	(Name)	DOB	zpfid
1	1	Joseph	09/09/76	49915106
2	2	Ann	08/07/78	49915202

Note: all identifying information has been removed from the data file. The fictional household in the example has names included for ease of illustration.

The final illustration, below, shows how the *zlineno* and *zperno* may diverge after Wave 3. In the register, persons who were household members in the previous wave are listed. In our example, Joseph does not appear on the register in Wave 3 because he had moved out of the household by Wave 2. The person moving in (Liam) is given the next available person number (07 – remember 06 in this household is assigned to Joseph). His *zpfid* is based on the *zhssid* and this person number (49915107). But he appear on line 6 because this line is free.

Table 6: Example Household with Zperno, zpfid, and zhssid, Wave 3

Zhssid=4991501 (Coleman Household)				
Zlineno	Zperno	(Name)	DOB	zpfid
1	1	Elaine	05/09/54	49915101
2	2	John	02/10/56	49915102
3	3	Sean	10/3/79	49915103
4	4	Eithne	12/05/84	49915104
5	5	Emer	14/07/85	49915105
6	7	Liam (Moved in)	09/11/77	49915107

Note: all identifying information has been removed from the data file. The fictional household in the example has names included for ease of illustration.

Zperno and Zpfxid

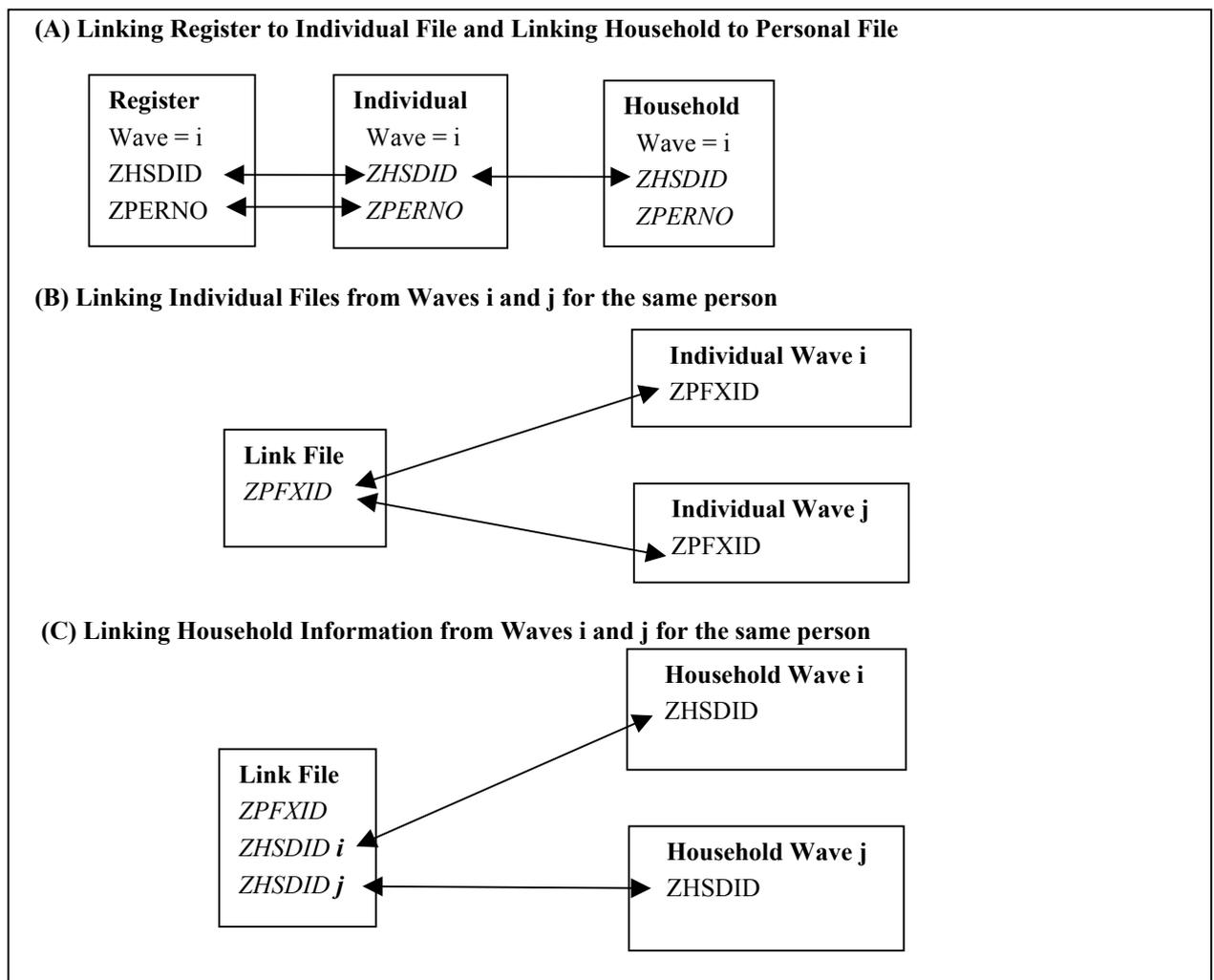
In matching the register and individual files within a wave, zhssid and zlineno are both used. However, the zhssid and zlineno (or zperno) cannot be used in matching individual (or register) files across waves, since the person may be in a different household. Zpfxid should be used in this situation: it is unique and consistent to the individual across all waves.

Matching or Linking Files

In matching information across files (household to individual, register to individual), it is important to (a) use the correct identifiers and (b) make sure both files are sorted by the identifiers.

The following diagram illustrates the appropriate identifiers to use in matching different types of files. In matching across waves, start with the link file.

Figure 5: Examples of Identifiers Used in Linking Files



Appendix: Details on Anonymisation

Appendix Table 1: Variables Anonymised in Living in Ireland Data Files

Variable		File(s)	Anonymisation
dayb	Day of birth	Register Link Individual	dropped
yearb	Year of birth	Register Link Individual	1909=1909 or earlier
zhsize	Household size	Register	10=10 or more
ZVALUE	Interviewer estimate of house value		£500,000=£500,000 or more
ZHSDSIZE	Total household size	Household	10=10 or more
ZH01_2	Since when have you lived at this address? (Year)	Household	1909=1909 or earlier
ZH05	Number of rooms in dwelling	Household	12=12 or more
ZH14	Total mortgage [Wave 1 only]	Household	£500,000=£500,000 or more
ZH17	Monthly Mortgage repayments	Household	2000=£2,000 per more
Z2H14_2	Amount of Loan 1[Waves 2-8 only]	Household	£500,000=£500,000 or more
Z2H14_5	Amount of Loan 2[Waves 2-8 only]	Household	£500,000=£500,000 or more
Z2H14_8	Amount of Loan 3[Waves 2-8 only]	Household	£500,000=£500,000 or more
ZH19	Respondent estimate of house value	Household	£500,000=£500,000 or more
ZH22	Monthly rent	Household	2000=£2,000 per more
ZH26	Monthly rent/mortgage supplement	Household	2000=£2,000 per more
ZH27_1	Monthly rent /mortgage supplement – last year	Household	2000=£2,000 per more
Z2h36_1w	Weekly cost of child care [Waves 2-8 only]	Household	500=£500 or more
ZH35_1w	Respondent estimate of weekly household income	Household	2000=£2,000 or more
ZHINCG	Gross Current Weekly Household Income from all Sources (constructed)	Household	2000=£2,000 or more
ZHINCN	Net Current Weekly Household Income from all Sources (constructed)	Household	2000=£2,000 or more
ZH48	Rental income last year	Household	50000=£50,000 or more
ZA12_1, ZA12_3, ZA12_3b	Current main job (works 15+ hours)	Individual	Codes provided are: Irish social class (CSO); 1 digit ISCO; 2-digit ISCO.
ZA13	Current Industry in main job (works 15+ hours)	Individual	Codes provided: Nace, 1-digit
PIWageg1	Current gross weekly wage in main job	Individual	2000=£2,000 or more
Piwagen1	Current net weekly wage in main job	Individual	2000=£2,000 or more
PISCHG1	Current gross weekly employment scheme allowance	Individual	2000=£2,000 or more
PISCHN1	Current net weekly employment scheme allowance	Individual	2000=£2,000 or more

Note that for variables that were top-coded (e.g. income sources), the existence of an upper limit does not necessarily imply that there were any cases receiving that upper limit.

Appendix Table 1(continued)

Variable		File(s)	Anonymisation
ZA44_1, ZA44_3, ZA44_1,	Current secondary job (works 15+ hours)	Individual	Codes provided are: Irish social class (CSO); 1 digit ISCO; 2-digit ISCO.
ZA44_3b	Current Industry in secondary job (works 15+ hours)	Individual	Codes provided: Nace, 1-digit
PIWageg2	Current gross weekly wage in secondary job	Individual	2000=£2,000 or more
PIWagen2	Current net weekly wage in secondary job	Individual	2000=£2,000 or more
PITRALG	Gross Weekly training allowance	Individual	2000=£2,000 or more
PITRALN	Net Weekly training allowance	Individual	2000=£2,000 or more
ZC2_1, ZC2_3, ZC2_3b	Current main job (works less than 15 hours)	Individual	Codes provided are: Irish social class (CSO); 1 digit ISCO; 2-digit ISCO.
ZC2_3	Current Industry in main job (works less than 15 hours)	Individual	Codes provided: Nace, 1-digit
ZE4_1, ZE4_3, ZE43b	Last main job (not working 15+ hours at present)	Individual	Codes provided are: Irish social class (CSO); 1 digit ISCO; 2-digit ISCO.
ZE5	Current Industry in last main job (not working 15+ hours at present)	Individual	Codes provided: Nace, 1-digit
ZG5	Gross weekly training allowance	Individual	2000=£2,000 or more
ZG6	New weekly training allowance	Individual	2000=£2,000 or more
ZJ2_w	Gross pay per week last year (constructed)	Individual	2000=£2,000 or more
ZJ3_w	Net pay per week last year (constructed)	Individual	2000=£2,000 or more
ZJ6_2	Extra Salary amount last year	Individual	10000 = £10,000 or more
ZJ6_4	Extra holiday amt last year	Individual	10000 = £10,000 or more
ZJ6_6	Profit/share/bonus amt last year	Individual	10000 = £10,000 or more
ZJ6_8	Company shares amt last year	Individual	10000 = £10,000 or more
ZJ6_10	Other lump sum amt last year	Individual	10000 = £10,000 or more
ZJ6_12	Other extra pmt amt last year	Individual	10000 = £10,000 or more
ZJ9_1t	Amount of additional overtime work or commissions or tips earnings last year (Constructed)	Individual	20000 = £20,000 or more
PISEMPG	Weekly profit or loss amount (constructed variable) Gross	Individual	2000=£2,000 or more
PISEMPN	Weekly profit or loss amount (constructed variable) Net	Individual	2000=£2,000 or more
ZJ27A	Estimated value of business or professional practice	Individual	500000=£500,000 or more
PIFARMG	Gross weekly farm income (constructed)	Individual	2000=£2,000 or more
PIFARMN	Net weekly farm income (constructed)	Individual	2000=£2,000 or more
ZJ32w	Average amount per week from casual work or secondary job in previous calendar year. (Constructed variable)	Individual	500=£500 or more
PITSW	Net weekly total social welfare payments, current (constructed)	Individual	2000=£2,000 or more
PITSWLY	Net weekly total social welfare payments, previous calendar year (constructed)	Individual	2000=£2,000 or more

Note that for variables that were top-coded (e.g. income sources), the existence of an upper limit does not necessarily imply that there were any cases receiving that upper limit.

Appendix Table 1: (Continued)

Variable		File(s)	Anonymisation
PICB	Weekly amount of Child Benefit (constructed) - current	Individual	2000=£2,000 or more
PICB_LY	Average Weekly Child Benefit in previous calendar year (constructed)	Individual	2000=£2,000 or more
pip	Current weekly net income from all non-social welfare pensions	Individual	2000=£2,000 or more
piotg	<i>Gross current income from other sources (constructed) per week</i>	Individual	2000=£2,000 or more
piofn	<i>Net current income from other sources (constructed) per week</i>	Individual	2000=£2,000 or more
PIGross	Persons total gross weekly income from all sources (constructed variable)	Individual	2000=£2,000 or more
PINet	Persons total net weekly income from all sources (constructed variable: Net=Gross Minus (Income Tax+PRSI))	Individual	2000=£2,000 or more
ZJ53	Lump sum receipts last year	Individual	100000 = £100,000 or more
ZM31_1s, ZM31_3, ZM31_3b	Main job of breadwinner when growing up	Individual	Codes provided are: Irish social class (CSO); 1 digit ISCO; 2-digit ISCO.

Note that for variables that were top-coded (e.g. income sources), the existence of an upper limit does not necessarily imply that there were any cases receiving that upper limit.

References

Nolan, B, Gannon, B., Layte, R., Watson, D., Whelan, C,T, Williams, James (2002), *Monitoring Poverty trends in Ireland: Results from the 2000 Living in Ireland Survey*, Dublin., Economic and Social Research Institute