

1.0 Component Surveys.

The samples from four Northern Ireland household surveys have been pooled together – aggregated - to form the Northern Ireland Pooled Household Survey (NIPHS). The surveys that comprise the NIPHS are the:

- Labour Force Survey (LFS);
- Northern Ireland Health Survey (NIHS);
- Continuous Household Survey (CHS); and
- Family Resources Survey (FRS).

Returns from the surveys above were pooled together for each of the following years:

- 2010/11;
- 2011/12;
- 2012/13;
- 2013/14; and
- 2014/15.

2.0 Sample Size

Responses on the core questions from the component surveys have been pooled to provide a dataset with a sample size of 10,000 households and 20,000 adults (aged 16+) per annum. Data for the four financial years are shown in the table below.

Household and Individual Responses by Year

	2010/11	2011/12	2012/13	2013/14	2014/15	Total
Responding Households	10,500	10,700	10,700	10,800	10,100	52,800
Adults (16+)	20,200	20,500	20,400	20,500	19,200	100,800
Children (0-15)	5,800	5,900	5,900	5,700	5,400	28,700
Responding Population	26,000	26,400	26,300	26,200	24,600	129,500

3.0 Survey Design

The Northern Ireland Pooled Household Survey dataset is constructed from four surveys which all utilise the same sample design and have the same finite target population, both with respect to the 'target groups' and the reference time-period. Households were selected into the sample for each survey from the Land and Property Services Pointer dataset¹.

All four surveys in the Northern Ireland Pooled Household Survey use unclustered systematic random samples with equal probabilities of selection, meaning that the design effects can be assumed equal and no adjustments to estimates of standard errors etc. are required due to differing design effects across the component surveys.

¹ Pointer is the address database for Northern Ireland and is maintained by Land & Property Services (LPS), with input from local councils and the Royal Mail. It is the common standard address for every property in Northern Ireland. More information on Pointer is available at <http://www.nidirect.gov.uk/index/information-and-services/property-and-housing/your-neighbourhood-roads-and-streets/ordnance-survey-of-northern-ireland/product-range/digital-products/pointer.htm>

4.0 Survey Weighting

A sample should reflect the population it comes from and be representative with respect to variables measured in a survey. In order to accomplish this survey responses were weighted to pre-determined population totals. Survey weights are associated with the respondents. Weights are used to make the estimated statistics based on the gathered data more representative of the population from which the data are sourced.

For the Northern Ireland Pooled Household Survey weighting up was to known population totals by age and sex using the Northern Ireland mid –year estimates of population produced by NISRA² and then removing the population living in communal establishments. Communal establishments are not sampled in the component surveys of the Northern Ireland Pooled Household Survey and therefore population totals need adjusted to reflect this. The **annual** pooled sample datafile has around 20,000 adults in around 10,000 households. This leads to a basic estimate of the weight per adult (and household) of around 70.

However weights were constructed at both individual and household levels using a more refined method. Weights used population control totals of Local Government district populations by age-band by sex. For example:

Male population aged 25-34 years living in the Lisburn and Castleragh local government district.

The final weights produced exhibit an acceptable level of variation as shown in the table below:

Summary Statistics: Survey Weight.

year	mean weight	minimum weight	maximum weight	Coefficient of Variation, weights
10-11	69.6	46.1	118.1	16.26%
11-12	68.7	45.6	131.3	15.00%
12-13	69.7	43.9	118.8	15.95%
13-14	69.5	45.3	122.3	16.44%
14-15	74.9	40.2	152.2	18.79%

² NISRA Mid-year population estimates available at <http://www.nisra.gov.uk/demography/default.asp17.htm>

5.0 Imputation.

The Northern Ireland Pooled Household Survey requires methods to be applied to correct for missing question responses; response imputation being such a method. Item imputation, was carried out when some data for an individual was missing, for example when an individual's response relating to the qualifications he / she held was missing due to a refusal to answer the question.

To address question or item level non-response a two-fold approach to imputation was used; within household imputation and "Hotdecking". Household imputation, as the name suggests, involved replacing missing data for a household member with corresponding responses from another household member. "Hotdeck" imputation involves constructing imputation classes using attribute data and randomly selecting a donor within an imputation class. Imputation classes were defined by five-year age-bands, sex, and whether an individual was living in an owner-occupier household or not.

For religion and ethnicity, missing values were taken from a random member of the same household where those values existed. For education, another case from across the full data pool sample with the same age and sex was used as a donor case – this donor was selected on a random basis. For employment cases from the Labour Force Survey were used as donor cases by single year of age and sex.

The Northern Ireland Pooled Household Survey contains a relatively modest amount of item non-response; approximately 16 per cent of the file contains cases where age, sex and household-level variables are present but the key outcome variable labour market outcome (ILODEFR) was not.

A detailed report on weighting and imputation prepared under the Quality Improvement Fund initiative may be accessed [here](#).

6.0 Errors in Estimates Derived from the Northern Ireland Household Pooled Sample.

Errors may occur at any stage during the collection and processing of survey data. There are two main sources of survey error:

1. Sampling error (errors associated directly with the sample design and estimation methods used) and;
2. Non-sampling error (a blanket term used to cover all other errors).

Sampling error is associated with the sample design and estimation methods used. In estimating a statistic, for example the unemployment rate, a level of error is to be expected as the data is collected from a sample of the target population rather than the population as a whole.

It is possible to control sampling error through the use of high quality sampling methods and to estimate its effect using information from the sample design and sample size. All four surveys in the NIHPS use unclustered systematic random samples with equal probabilities of selection. In such a sample design the calculation of a confidence interval is straightforward using the basic equation below:

- $\pm 1.96 \sqrt{(p \cdot q / n)}$

Where: p=proportion in question for example the unemployment rate; q=(1-p) and n=relevant sample size. The use of this basic equation is employed for *ease of use*. The basic equation does not take account of issues such as:

- Dependency of respondents answers within households;
- Geographical stratification of the sample; and
- Item Imputation.

Each of the above points has a differing impact on the standard error of estimates from the NIPHS. The technically correct method of calculating standard errors would be to use re-sampling techniques. Such methods were tested and proved to provide very similar estimates of sampling error to the basic equation above.

The measure of uncertainty used in the NIPHS is the Coefficient of Variation (CV). The CV measures the variability of the estimates by expressing the standard error as a percentage of the parameter estimate. As the coefficient of variation is not measured in any specific unit, it facilitates comparison between surveys measuring different underlying variables. A larger coefficient of variation implies a larger variability. The data in the NIHPS survey tables are colour coded based on

While the sampling error may be calculated as shown previously estimates of non-sampling error are not readily available. However in delivering the four surveys NISRA Central Survey Unit undertake several methods to minimise such errors. For example by intensive interviewer training and the provision of guidance materials for interviewers. CSU also employ Computer Assisted Personnel Interviewing (CAPI) as the primary response mode. CAPI ameliorates errors in data entry and ensures as far as possible the data is of a high quality for example by employing electronic data checks.

7.0 Data Release Protocols.

In order to ensure as far as possible the confidentiality of responses to the surveys and also to ensure inferences are not made regarding unsafe data a general release rule is employed in the Northern Ireland Pooled Household Survey regarding data release for population related outputs. This rule is set out below

- (i) Any output with a weighted population/households of than 1,000 persons/households or more should be released (this will typically have a sample population of at least 15); or
- (ii) Any weighted population/household which is less 1,000 persons/households could be released **but only if** the raw sample population/household count is greater than or equal to 10 persons/households.

These rules safeguard confidentiality and provide robust statistics. In overall sample terms outputs will have a Coefficient of Variation of at worst 30% (see below).

Sample N	20,000
Threshold	10
p	0.05%
q	99.95%
Standard error of the estimate	0.016%
Coefficient of Variation	32%