Northern Ireland Pooled Household Survey (NIPHS)

October 2015

A: Summary

1. The Northern Ireland Statistics and Research Agency (NISRA) have identified a user need to produce more detailed sub-regional and sub-group statistics from household surveys. Following an in-house feasibility study (with the support from Survey Methodology Division, Office for National Statistics (ONS)), it was found that by combining data from four existing core Northern Ireland household surveys such statistics could be produced. As a result, statistics on labour market and other demographic factors have been generated with both greater geographical detail and smaller sampling errors than up to now. This report summarises this work – outline results will be published in the near future.

B: Background

2. There is an on-going demand for NISRA to produce more detailed statistics including sub-regional statistics for the new 11 Local Government Districts (LGDs) and sub-group analysis (e.g. data on youth unemployment). Sample size constraints mean that detailed estimates of population changes for some small groups cannot be determined from current individual surveys.

3. This paper assesses the option to pool data from the Labour Force Survey with other core Northern Ireland annual household surveys. The resultant dataset is labeled the Northern Ireland Pooled Household Survey. A feasibility study was carried out with advice on weighting and imputation from Survey Methodology Division at ONS. The project was scoped with funding from the Department for Employment and Learning and a successful Quality Improvement Fund (QIF) bid from the UK Statistical Authority.

4. This project mirrors work in Scotland on the standardisation of 20 questions in the Scottish Health, Household and Crime & Justice Surveys, producing sufficient sample size for robust estimates of small sub-groups and sub-regional statistics. The Office for National Statistics (ONS) have also undertaken similar work with the Integrated Household Survey which is a composite survey combining a number of ONS social surveys and a core set of questions.

C: Method

5. Four of Northern Ireland's largest core household surveys were chosen for this feasibility project:
    - Labour Force Survey (LFS) – wave 1 and 5 addresses only*
    - Northern Ireland Health Survey (NIHS)
    - Continuous Household Survey (CHS); and
    - Family Resources Survey (FRS)

    **See link for an explanation of the wave structure of LFS*

6. A set of 8 questions were chosen for analysis (see table 1).

Table 1 Core Individual and Household Questions used in Pooled Household Survey

Age (age groups)	16-19 years, 20-24 years, 25-29 years, 30-34 years, 35-39 years, 40-44 years, 45-49 years, 50-54 years, 55-59 years, 60-64 years, 65-69 years, 70 and over
Sex	Male, Female
Ethnic Group	White, Other Ethnic Group, No response
Religion	Catholic, Protestant, Other religion, No religion, No response
Education	No Qualifications, Degree level or higher, All other qualifications, No response
Employment (ILO)	Working, Unemployed, Economically Inactive
Accommodation type Household	Flat, Detached House, Other, Semi Detached House, Terrace House
Tenure Household	Owned outright, Mortgage or loan, Part rent (Co-ownership), Rented, Rent free

7. The classification of these variables will be refined prior to producing statistical outputs. In addition to these core questions, data from the Pointer database was included to allow administrative variables such as Property Capital Value, Geography (Local Government District 2014) and Measures of Multiple Deprivation (MDM) to be included in the dataset. Household size, the number of adults and children along with annual time period of data collection were also included.

8. Responses on these questions from the core surveys provide a dataset (the Northern Ireland Pooled Household Survey) with a sample size in excess of 10,000 households or over 20,000 individual adults (aged 16+) responses per annum. Data for four financial years are shown in table 2.

Table 2: Household and Individual Responses by Year (Northern Ireland Pooled Household Survey)

	2010/11	2011/12	2012/13	2013/14	Total
Responding Households	**10,400**	**10,700**	**10,700**	**10,800**	**42,600**
Adults (16+)	20,100	20,500	20,400	20,500	**81,500**
Children (0-15)	5,800	5,900	5,900	5,700	**23,300**
Responding Population	**25,900**	**26,400**	**26,300**	**26,200**	**104,800**

D: Interview Types

9. There are clear distinctions in the types of data that amalgamated into the Northern Ireland Pooled Household Survey – three main categories are considered (i) **response**, (ii) **proxy** and (iii) **grid** data.

10. "Response data" is information gathered at the point of interview, asking the respondent questions around employment and other factors. For the Northern Ireland Pooled Household Survey this is around 67% of the adult dataset.

11. "Proxy data" is where the respondent is asked questions about another adult member of the household. Again questions such as employment etc… are commonly asked in the core surveys. The amount of proxy data recorded differs for each survey due to the different rules around component surveys. This represents around 18% of the adult data in the Northern Ireland Pooled Household Survey.

12. Finally there is "grid data". At each address where an interview is achieved, demographic data in a standard household grid is gathered. This is a list of who lives in the household, their age, sex and marital status (for those 16 years or more). This equates to around 14% of the adult data in the Northern Ireland Pooled Household Survey (see table 3 for types of data by survey year).

Table 3 Data types (the Northern Ireland Pooled Household Survey)

	2010/11		2011/12		2012/13		2013/14	
	Total		Total		Total		Total	
	Number	%	Number	%	Number	%	Number	%
Adults (16+)	20,100		20,500		20,400		20,500	
Data Type:								
Response	*13,400*	*67%*	*13,800*	*67%*	*13,700*	*67%*	*14,000*	*68%*
Proxy	*3,900*	*19%*	*3,700*	*18%*	*3,600*	*17%*	*3,600*	*18%*
Grid	*2,800*	*14%*	*3,000*	*15%*	*3,100*	*15%*	*2,900*	*14%*

E: Unweighted results

13. Without any further statistical intervention the final survey data can be assessed for demographic accuracy. The headline demographics compare favourably with the Northern Ireland Census results (see table 4).

Table 4 Pooled Household Survey /Census 2011

	Pooled Household Survey 2012/13 Unweighted %	Census 2011 %	Difference
Sex			
Male	47.2	48.5	1.3
Female	52.8	51.5	-1.3
Age			
16-19 years	6.6	7.0	0.6
20-24 years	7.1	8.8	1.7
25-29 years	7.7	8.7	1.0
30-34 years	7.9	8.5	0.4
35-39 years	8.0	8.6	0.6
40-44 years	9.0	9.3	0.3
45-49 years	9.0	9.3	0.3
50-54 years	8.9	8.3	-0.6
55-59 years	7.8	7.0	-0.8
60-64 years	7.7	6.6	-1.1
65-69 years	6.8	5.8	-1.0
70 and over	13.5	12.1	-0.6

Average Household Size

14. A further key indicator of the representativeness of the Northern Ireland Pooled Household Survey is average household size. The Northern Ireland Pooled Household Survey includes the number of people (adults and children) in each household. The Northern Ireland Pooled Household Survey figures compare favourably with the 2011 Census figures and the Census based household projections – see table 5.

Table 5 Average Household Size (Northern Ireland Pooled Household Survey)

	2010/11		2011/12		2012/13		2013/14	
	Total		Total		Total		Total	
	Number	%	Number	%	Number	%	Number	%
Households	10,400		10,700		10,700		10,800	
Adults (16+)	20,100	78%	20,500	78%	20,400	78%	20,500	78%
Children	5,900	23%	5,900	22%	5,900	22%	5,800	22%
Total	26,000	100%	26,400	100%	26,300	100%	26,300	100%
Average Household Size	**2.48**		**2.46**		**2.46**		**2.43**	
Census / Census based projections	**2011***		**2012****		**2013****		**2014****	
Average Household Size	**2.54**		**2.54**		**2.54**		**2.54**	

* http://www.nisra.gov.uk/Census/key_stats_bulletin_2011.pdf
** http://www.nisra.gov.uk/archive/demography/population/household/HHP12_NI.xls

Geographical Coverage

15. One of the main drivers of this feasibility study is to provide more reliable sub-regional statistics, most specifically data for the 11 new Local Government Districts (LGDs). Therefore reviewing the geographical coverage of the responses in the Northern Ireland Pooled Household Survey is essential. (For more detail on geography definitions see http://www.ninis2.nisra.gov.uk/public/documents/NISRA%20Geography%20Fact%20Sheet.pdf).

16. Shown over leaf are three figures
    - Figure 1 maps wave 1 and wave 5 LFS interviews 2012/13
    - Figure 2 maps the Northern Ireland Pooled Household Survey (1 year 2012/13)
    - Figure 3 maps the Northern Ireland Pooled Household Survey (3 years combined)

17. Adding the component surveys together provides not only more data, but more improved geographical coverage (see figure 2). When 3 years of data are combined there is extensive coverage across Northern Ireland (see figure 3).

Figure 1 LFS Interviews (wave 1 and 5 2012/13)

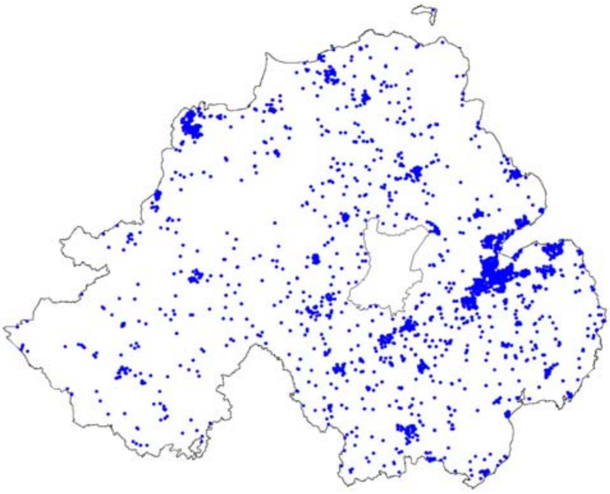


Figure 2 Pooled Household Survey Interviews (1 year 2012/13)

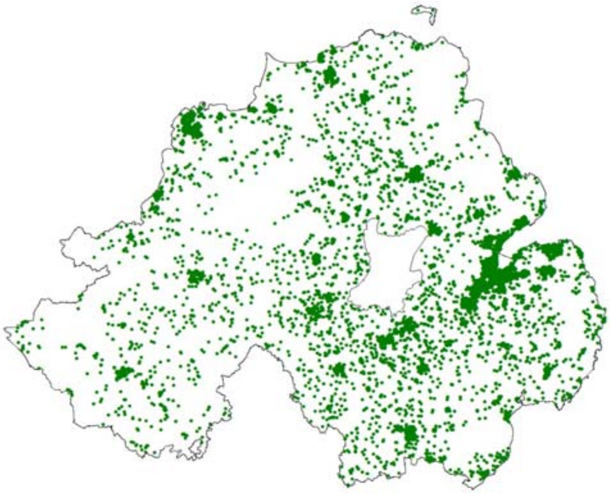
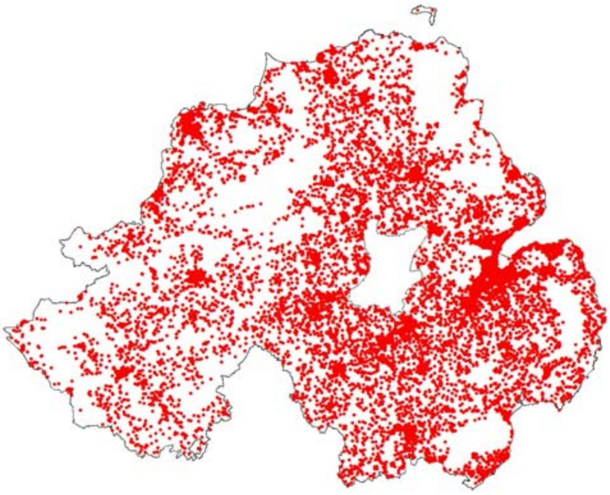


Figure 3 Pooled Household Survey Interviews (2010/11 to 2012/13)



18. As with all surveys, the Northern Ireland Pooled Household Survey requires two separate corrections. The first is that some data items for individuals are missing and a level of item imputation must be carried out on these missing cases. The second is around unit non-response when some households have either refused to be surveyed or there was no contact with the householder in the field (this issue is dealt with under the section on weighting).

19. To address item or question non-response a two-fold approach was used; "hotdecking" and within household imputation. For Education, another case from across the data pool sample with the same age and sex was used as a donor case – this case was selected on a random basis. For employment (using the International Labour Organization definition of employment – ILO*) –cases from the LFS dataset were used as donor cases by single age and sex, followed by the remainder of the data pool. (**See http://stats.oecd.org/glossary/detail.asp?ID=778 for details*). For Religion and Ethnicity, missing values were taken randomly from other members of the household were those values existed.

20. This imputation was tested. The ILO employment imputation was checked by punching holes in the complete data. In total 10% of known ILO cases were recoded as missing and item imputation carried out. This resulted in "correct" ILOs codes being imputed for 87% of known cases (see table 6).

21. For future work Tenure (as a proxy for social class) will also be used along with age and sex for imputation. This is one of the key recommendations of the Office for National Statistics research (see Appendix 1).

Table 6 Original ILO by Hot Decked ILO Crosstabulation Check

		ILOHD Employment Status (ILO) Hot Deck			Total
		Working	Unemployed	Economically Inactive	
ILO Original	Working	8,680	**15**	**80**	8,775
	Unemployed	**10**	542	**10**	562
	Economically Inactive	**80**	**7**	6,927	7,014
Total		8,770	564	7,017	16,351

Incorrect = 10 + 80 + 15 + 7 + 80 + 10 = 202
*out of 1,600 cases randomly coded as missing = **13% incorrect***
*Correct = 1600-202 = 1398 = **87% correct***

G: Weighting

22. As noted the second issue relates to unit non-response when some households either refused to be surveyed or there was no contact with the householder. This was addressed by weighting up to known population totals by single year of age and sex using the Northern Ireland level mid –year estimates of population and then removing the known communal establishment population as per Census 2011 (See link).

23. As noted a methodology report on weighting and imputation prepared under the Quality Improvement Fund is attached as Appendix 1. This report includes key recommendations around weighting which will be included in any experimental statistical outputs produced.

I: Post-weighting & item imputation Labour Market data validation

24. The provisional labour market estimates after allowing for item imputation and weighting broadly match those from the LFS datasets – see table 7 overleaf. There are however some minor differences. These differences are generally all within confidence intervals and would seem to be down to slight variations in the component survey questionnaires (especially around ordering of questions/ contextual effects/differing nature of individual surveys).

25. One consistent small difference between the LFS results and the Northern Ireland Pooled Household Survey is for economic activity and inactivity. The pooled sample consistently provides a marginally higher estimate of the number of people who are economically inactive. This difference, which is of the order of 0.7-2.5% of the population, is consistent across all age-groups. This issue should not impact on trend analysis but it is important to consider when looking at overall levels. The issue will be flagged in all outputs from the pooled sample datasets.

H: Sampling Margin of Error

26. One key issue is the margin of error of the resulting estimates. Table 7 below shows labour market estimates from the Northern Ireland Pooled Household Survey and the LFS. Not shown are the sampling margins of error. The annual sampling coefficients of variation for the Labour Force Survey are 7% for unemployment, 1% for working and 3% for economically inactive. The annual coefficients of variation for the Northern Ireland Pooled Household Survey are 3.5% for unemployment, 0.5% for working and 1.5% for economically inactive. This result is typical across the Northern Ireland Pooled Household Survey with coefficients of variation around half the size of those for the LFS. It should be noted there are other sources of error not considered most notably non-sampling errors – these can also affect the margins of error of the final estimates created.

Table 7 Levels of unemployment Pooled Household Survey and the Labour Force Survey

Pooled Household Survey (Aged 16-64)	2010/11		2011/12		2012/13		2013/14	
	Number	%	Number	%	Number	%	Number	%
Working	748,000	92.9%	749,000	92.0%	756,000	92.6%	769,000	92.8%
Unemployed	**57,000**	**7.1%**	**65,000**	**8.0%**	**61,000**	**7.4%**	**60,000**	**7.2%**
Economically Active	805,000	69.7%	814,000	70.3%	816,000	70.5%	829,000	71.6%
Economically Inactive	349,000	30.3%	343,000	29.7%	342,000	29.5%	329,000	28.4%
Total	**1,155,000**		**1,157,000**		**1,158,000**		**1,158,000**	

Labour Force Survey (Aged 16-64*)	2010/11		2011/12		2012/13		2013/14	
	Number	%	Number	%	Number	%	Number	%
Working	768,000	93.2%	780,000	92.4%	775,000	92.2%	782,000	93.1%
Unemployed	**56,000**	**6.8%**	**65,000**	**7.6%**	**65,000**	**7.8%**	**58,000**	**6.9%**
Economically Active	824,000	71.2%	845,000	72.8%	840,000	72.4%	840,000	72.3%
Economically Inactive	333,000	28.8%	315,000	27.2%	321,000	27.6%	323,000	27.8%
Total	**1,158,000**		**1,160,000**		**1,161,000**		**1,163,000**	

**Data obtained from Economic & Labour Market Statistics Branch (NISRA)*

Difference (LFS – NIPHS)	2010/11		2011/12		2012/13		2013/14	
	Number	%	Number	%	Number	%	Number	%
Working	20,000	0.3%	31,000	0.3%	19,000	-0.4%	13,000	0.3%
Unemployed	**-1,000**	**-0.3%**	**0**	**-0.3%**	**5,000**	**0.4%**	**-2,000**	**-0.3%**
Economically Active	19,000	1.5%	31,000	2.5%	24,000	1.9%	12,000	0.7%
Economically Inactive	-16,000	-1.5%	-28,000	-2.5%	-21,000	-1.9%	-6,000	-0.7%
Total	**3,000**		**3,000**		**3,000**		**5,000**	

J: Conclusion

27. This pilot work has indicated that pooling data from various household survey sources is feasible and further work on this project is ongoing. This work will ensure that a fully operational system is developed. Ultimately, this pooled sample approach will create:
    - Annual trend estimates at a Northern Ireland level for labour market statistics and other demographic data, along with crosstabulations of these socio-demographic characteristics of the population;
    - Improved precision of household survey estimates with the ability to produce survey based estimates for the 11 new Local Government Districts.

28. NISRA will release headline results from the pooled household dataset to encourage analysis, exploration and feedback by stakeholders. This process will influence future developments. Although the large-scale surveys used in the Northern Ireland Pooled Household Survey are all designated as National Statistics, results will be released under the classification "Experimental Statistics." Therefore care should be taken when analysing these data.

29. While every effort has been made to harmonise the core questions across the surveys, different context or ordering effects could lead to inconsistencies in responses across the different surveys which may skew the results. In light of the above, and despite the larger sample size, the results should <u>not</u> be considered as the primary source of data for the variables it contains. This will be reviewed as the Northern Ireland Pooled Household Survey is further tested and developed. The tables that will be made available will also highlight the preferred sources for some of the core question topics.

Northern Ireland Statistics and Research Agency
October 2015

Appendix 1:

Weighting and Imputation for the Northern Ireland Pooled Dataset

Matt Greenaway, Survey Methodology Division, Office for National Statistics

Quality Improvement Fund Report

1. General Comments – Methodology for Estimating from Multiple Datasets

In principle, there are two approaches to combining data from multiple datasets – constructing separate estimates and combining these estimates; and 'pooling' the datasets at the record level and calculating estimates using this single pooled dataset (Roberts & Binder, 2009). Given that the focus of the data pooling exercise is on creating a tool for researchers who may want to carry out a variety of analyses, it seems sensible to focus on this second approach.

The Northern Ireland pooled dataset will be 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. In this context, simple design-based estimation from the pooled dataset is straightforward, and an increase in the sample size by a factor of x will reduce standard errors by a factor of $\sqrt{x}$ - so a quadrupling of the sample size would result in a halving of the standard error.

In practice, however, we need to be concerned with non-sampling error. Different surveys may have different non-response biases (both at the household, person and item level), mode effects, and question wording effects, all of which may lead to differences in non-sampling errors between surveys. While there is evidence that mode effects when measuring economic activity tend to be negligible (Elliot et al, 2006), and question wording is identical across the surveys, there is clearly potential for differing non-response patterns and contextual effects.

The fairly standard imputation and weighting methods recommended in this report may go some way to removing differences in non-response bias. We do not recommend more complicated methods such as survey-specific non-response models or methods to account for variation in non-sampling error, as recent experiences at both the ONS and other organisations strongly suggest that more complicated methods are of limited benefit and lead to over-engineered systems which cannot be maintained.

It is therefore important to emphasise that non-sampling error will differ between the surveys, and this will have an ongoing and non-measurable impact on the pooled survey estimates. The only way to address this is to harmonise the surveys (in terms of their question ordering, interviewer instructions etc), but that is outside the scope of this report, which focuses on imputation (section 2) and weighting (section 3).

2. Imputation

2.1 Imputation: Preliminary Analysis

Imputation can be used for both item non-response (where certain items are missing for a respondent) and unit non-response (where an entire case is missing). The pooled survey contains a relatively modest amount of item non-response – approximately 16% of the file contains cases where age, sex and household-level variables are present but the key outcome variable ILODEFR is missing. ILODEFR being the ILO's categorization of an individual's labour market position. Unit non-response is also present – the response rates for all four surveys are under 60% - but very little data is available for unit non-responders. Given this paucity of data for unit non-responders, imputation should be focused on item non-response, with unit non-response adjusted for using weighting.

Imputation will remove non-response bias where –

- The data available for all cases and used in the imputation ('attribute data') is correlated to the missingness mechanism – for example, age is correlated to item non-response
- The attribute data is correlated with the variable to be imputed – for example, age is correlated with economic activity.

The variables available for all responders are – age, sex, marital status, geography, and tenure. Information for other respondents in the household is also available in some cases. The key variable to be imputed is ILODEFR – a derived variable with three categories describing employment status (employed/unemployed/economically inactive).
The tables below show, for each of the available attribute variables, the amount of missingness in ILODEFR.

Gender	Has ILO data	No ILO data	Percentage missing
Male	8,001	1,968	19.74
Female	9,718	1,338	12.1

Tenure	Has ILO data	No ILO data	Percentage missing
Own outright	6,729	1,219	15.34
Mortgage or Loan	5,967	1,270	17.55
Part Rent	59	11	15.71
Rented	4,720	755	13.79
Rent Free	232	44	15.94

Age	Has ILO data	No ILO data	Percentage missing
16-24	1,857	1,114	37.5
25-34	2,771	567	16.99
35-44	3,298	463	12.31
45-54	3,234	486	13.06
55-64	2,827	338	10.68
65+	3,732	338	8.3

NUTSIII	Has ILO data	No ILO data	Percentage missing
Belfast	2,604	2,604	17.7
Outer Belfast	3,809	699	15.51
East	4,456	4,456	15.35
North	2,829	501	15.05
West & South	4,021	738	15.51

Marital Status	Has ILO data	No ILO data	Percentage missing
Single, Never Married	4,916	1,734	26.08
Married, living with	9,880	1,447	12.77
Married, separated	691	30	4.16
Divorced	884	33	3.6
Widowed	1,324	57	4.13
Civil Partner/Same-Sex	24	4	14.29

There are clear patterns in missingess by gender (men are more likely to have missing data), age (younger people are more likely to have missing data), tenure (renters are less likely to have missing data) and marital status (single people are more likely to have missing data).

Obviously, some of these results may be linked – young people are more likely to be single, for example. To investigate further, logistic regression can be used to predict missingness, with all attribute data as independent variables. The Wald coefficients and p-values from such a model are shown below.

Effect	Wald	p-value
Age	427.4584	<.0001
Sex	164.1118	<.0001
Marital Status	129.5695	<.0001
Tenure	91.903	<.0001
NUTSIII	13.3494	0.0097

Age is by far the best predictor of missingness, followed by sex, marital status and tenure.

Regression can also be used to evaluate which variables are most predictive of employment. The table below shows Wald coefficients and p-values from a logistic regression model with employment as the dependent variable.

	Wald	p-value
Age-group	2573.248	<.0001
Sex	12.2097	0.0005
Marital Status	127.1815	<.0001
Tenure	428.7317	<.0001
NUTSIII	32.172	<.0001

Again, age is by far the best predictor, but tenure appears to be the second-most powerful variable, considerably more powerful than marital status.

2.2 Imputation: Method

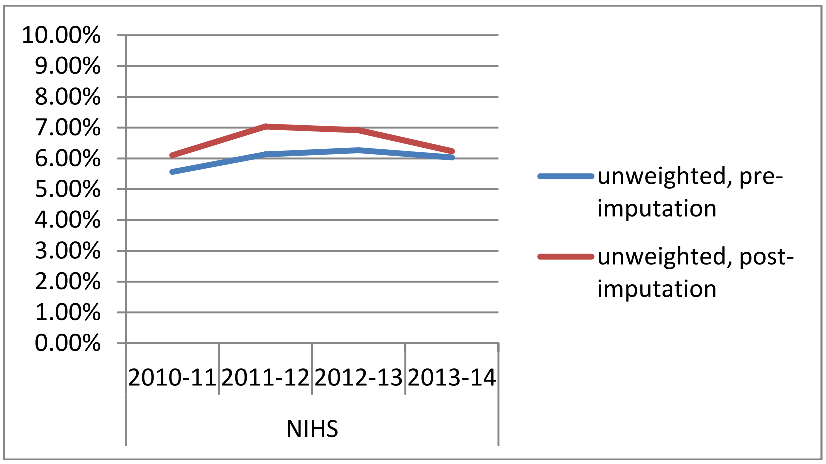
Given that we are imputing a categorical variable which appears to be well-correlated with available auxiliary information, donor imputation is the natural choice. As there is a need for simplicity, we recommend hot deck imputation, which involves constructing imputation classes using attribute data and randomly selecting a donor within an imputation class.

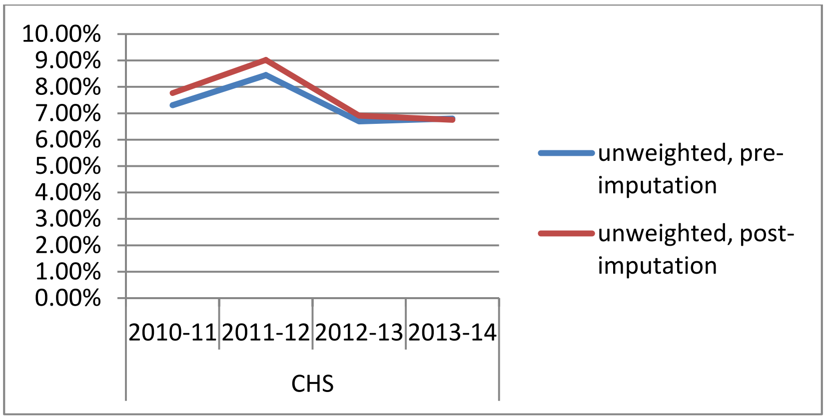
The remaining question is how these imputation classes should be defined. In order to limit imputation variance, each class should contain enough full responders to ensure that individuals are not used as a donor a large number of times. In order for the imputation to remove bias, they should also utilise variables which are well-correlated with the missingness mechanism and the outcome variables.

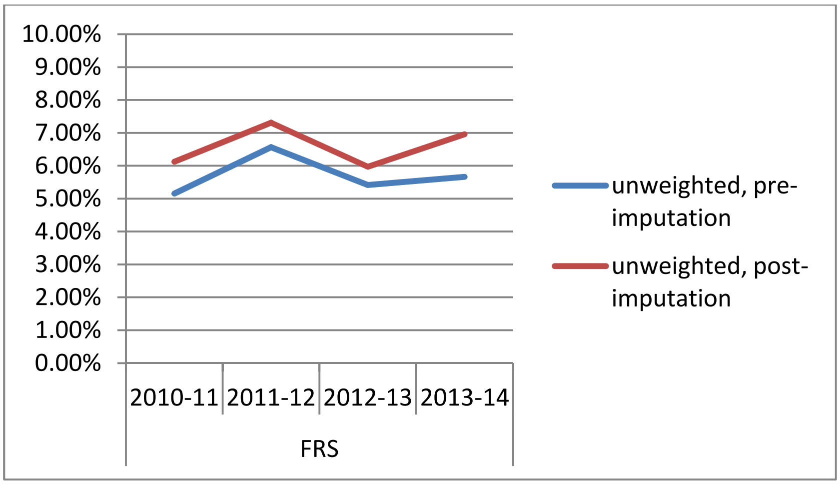
We recommend using imputation classes defined by five-year age-bands, sex, and whether an individual is an owner-occupier. The alternative would have been to utilise marital status instead of tenure – but, although marital status is more predictive of missingness, tenure is much better correlated with the outcome variable. Using finer classes incorporating both tenure and marital status would have an adverse impact on imputation variance, which is discussed in more detail in the next section.

The graphs below show the impact of imputation on the unweighted unemployment rate, separately for each survey. Imputation has a consistent positive impact on the unweighted estimates. It is somewhat larger on the FRS and somewhat smaller on the CHS, but the differences are not large.

Graphs 1-3 – impact of imputation, by survey





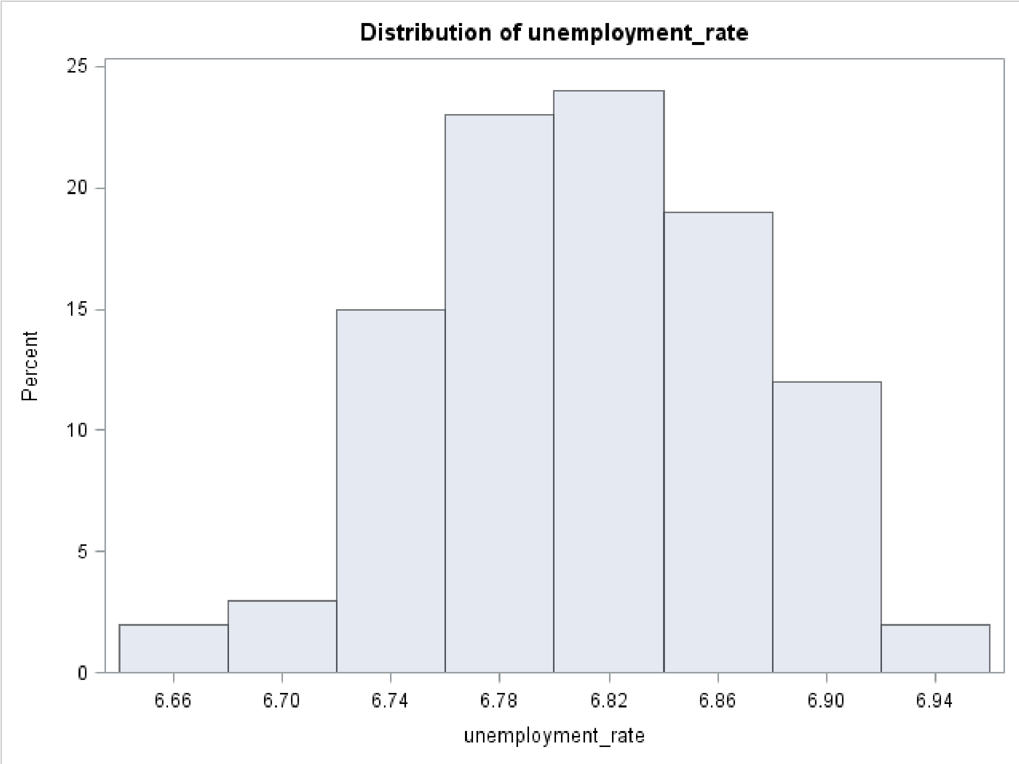


2.3 Imputation variance

It is usually advisable to attempt to quantify imputation variance, which we do by utilising 'multiple imputation' – imputing a large number of times on the same dataset. Since our recommended method is stochastic (random), each run of the imputation will produce different results.

The graph below shows unweighted unemployment rates obtained from 100 multiple imputation iterations. These are further summarised in the table. Given a mean of 6.8%, the 5th percentile lies at 6.7% and the 95th percentile lies at 6.9%. Given that the confidence intervals for the unemployment rate from the pooled dataset are likely to be in the region of plus or minus 0.5%, this does not seem like excessive extra variability from imputation.

Graph 4 – imputation variance of the unemployment rate



	Over 100 iterations			
	Mean	**5th percentile**	**95th percentile**	**CV**
Unweighted Unemployment Rate	6.80984655	6.70917	6.91073	0.87%

2.4 Alternative Imputation Methods

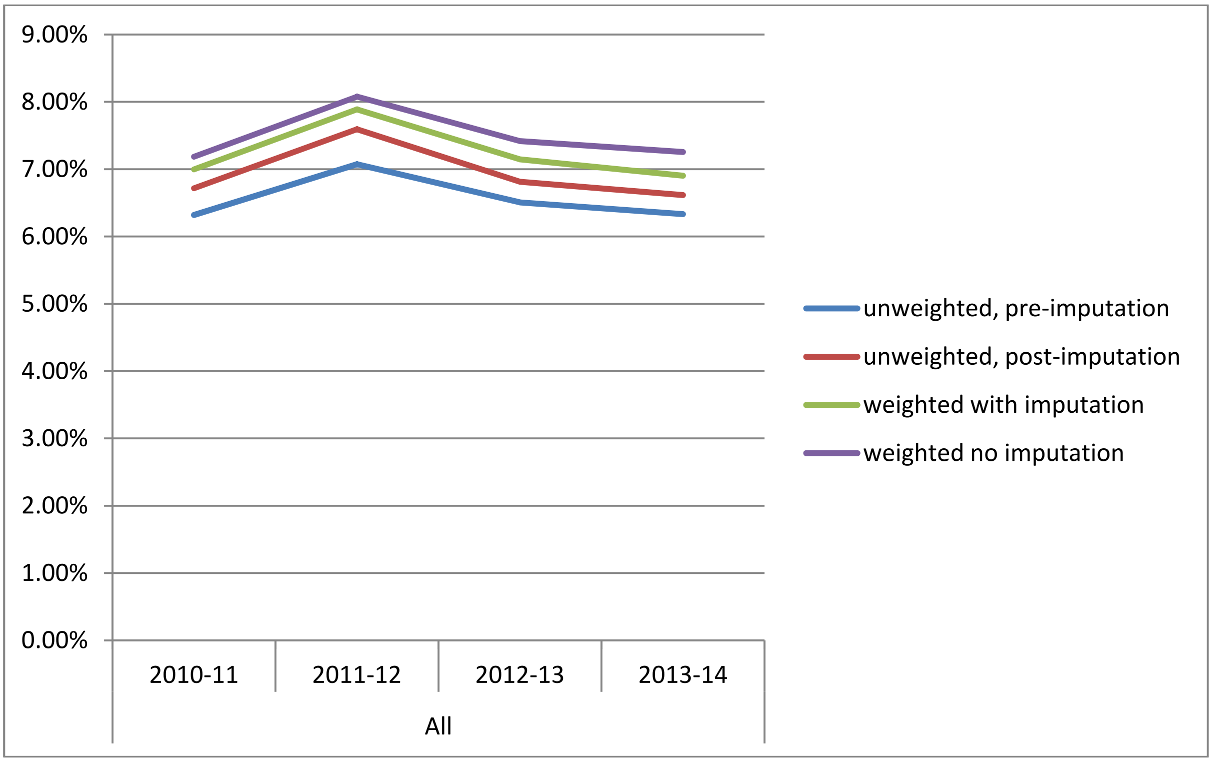
The imputation method we have recommended is very straightforward. An alternative donor imputation method would be more complicated nearest-neighbour imputation. This would involve utilising an assortment of auxiliary variables to define a distance function on which a donor is selected. This method allows more auxiliary variables to be used and therefore may better capture the missing-data mechanism.

However, nearest-neighbour imputation can quickly become prohibitively complex to implement without specialised software, particularly where limitations need to be placed on the number of times a case is used as a donor. Additionally, in the context of the NI pooled survey, where a relatively small number of variables are available, the advantages over a simple hot-deck method may be limited. There are plans for the NISRA Social Survey team to gain access CANCEIS – imputation software written by Statistics Canada – and as these expertise develop, further research into nearest-neighbour methods on the pooled dataset could begin.

Another alternative would be to utilise the recommended method, but to only use LFS cases as donors, on the grounds that the focus of the LFS is an individual's labour market outcome. However, this would increase imputation variance substantially.

A final alternative would be not to impute at all. The graph below shows the unemployment rate over time for the pooled dataset unweighted, after imputation, after imputation and weighting (using the weighting method described in the next section), and after weighted but without imputation (that is – removing cases with missing data and allowing the weights to adjust for any bias).

Graph 5 – impact of imputation



As we saw previously, imputation increases the unweighted unemployment rate. However, the impact of imputation on the weighted estimates (comparing 'weighted with imputation' and 'weighted without imputation') is fairly small. This is because our imputation method is removing bias associated with age, sex and tenure, and the weighting method described in the next section will remove bias related to age and sex.

It is probably still worth imputing – imputation does have an impact over and above weighting, and providing a dataset without missingness to users is valuable. It should be noted that one consequence of choosing to impute is that the standard errors of pooled survey estimates will be under-estimated, since the imputed values will be treated as real values in the variance estimate and the imputation variance measure in the previous section will not be accounted for.

3. Weighting

3.1 Weighting – design weights

Much of the literature on pooling surveys focuses on re-scaling the existing weights for the surveys by some factor in order to maximise precision. Under this approach, precision is maximised when the weights are scaled relative to the effective sample size of each dataset (that is, the sample size divided by the design effect) (Chu, 1997).

All four surveys in the NI pooled dataset use unclustered systematic random samples with equal probabilities of selection, meaning that the design effects can be roughly assumed to be equal, and the effective sample size is therefore proportional to the achieved sample size. The optimum way to scale the existing survey weights under the standard approach would just be to scale the separate survey weights relative to the achieved sample size of each dataset.

However, we prefer the approach of ignoring the existing weights for the separate surveys, and instead simply re-weighting the pooled dataset. This is partly for practical reasons – the component surveys use different weighting strategies and weights are not easily available – and partly because we wish to ensure that the final weight is calibrated to population totals.

Social survey weighting in official statistics is typically model-assisted – we assign design weights according to selection probabilities and calibrate these to known totals. For the pooled survey, the appropriate design weight is equal for all cases and is simply $\frac{N}{n}$, where *N* is the number of adults in the population and *n* is the number of adults in the sample. This is equivalent to scaling the separate-survey design weights by the achieved sample size of each survey, under the assumption of equal design effects. We then re-weight the entire pooled dataset using these new design weights and population totals namely NISRA Mid-Year Estimates (MYE).

3.2 Weighting – groups

The main methodological decision is then how to define the weighting groups – the groups within which weights will sum to population totals.

The main quarterly Labour Force Survey uses three sets of 'partitions', or mutually exclusive and exhaustive weighting groups –

1. local authority (old definition)
2. NI*5-year age-sex bands
3. NI*single-year of age between 16-24*sex

These should be considered separately rather than the cross-classification. For example, the weights for responders in Belfast will sum to the population total for Belfast, the weights for 16-year-old males in Northern Ireland will sum to the relevant population total, but the weight for 16-year old males in Belfast will **not** sum to the relevant population total.

Weighting in this manner – using multiple partitions – is technically complex, and requires the use of specialised software. The ONS use Statistics Canada's GES ('Generalized Estimation System').

However, if only one partition is used – if a single set of mutually exclusive and exhaustive weighting groups are defined – then weighting is straightforward to carry out by 'hand' (in Excel or similar software). This approach is known as 'post-stratification', and the weights simplify to –

$w_{ih} = d_{ih} \dfrac{T_h}{\sum_{i \in h} d_{ih}}$ where d_{ih} is the design weight for unit *i* in weighting group *h* and T_h is the known total for the weighting group

Where all design weights are identical, as in the case of the pooled dataset, the weights simplify even further to –

$w_{ih} = \dfrac{T_h}{n_h}$ where n_h is the number of responders in the weighting group

That is, where post-stratification is used and all design weights are equal, the final weight can be calculated simply as the population total divided by the number of responders, and there is no need to actually calculate the design weights.

The weighting groups *h* need to be correlated with the outcome variables and be at the correct level of detail. Choosing more detailed weighting groups will reduce standard errors and may reduce non-response bias – for example, if 16-year olds were randomly less likely to be sampled in a given dataset, or they were less likely to respond, then weighting by single-year-of age will adjust for this whereas weighting by five-year age-bands will only do so to a lesser extent. However, choosing very detailed weighting groups will increase the variance of weights and therefore increase standard errors, and may even lead to practical difficulties with empty groups. A general rule-of-thumb is that weighting groups should contain at least 30 respondents.

3.3 Weighting – recommended method

Due to the size of the pooled dataset, it is possible to use a simple post-stratification approach with very detailed weighting groups. We recommend using –

- New local authority*Age-band*Sex

Where the age-bands are defined as –

- 16-19
- 20-24
- 25-34
- 35-44
- ...
- 75+

The weights produced are not worryingly variable –

year	mean weight	minimum weight	maximum weight	CV, weights
11-12	68.2	49.5	106.7	14.02%
12-13	71.3	50.4	102.2	12.69%
13-14	71.1	51.7	129.4	14.64%

The only drawback to this method is that single-year-of age between 16-24 is not utilised, which could plausibly lead to some bias in, for example, youth unemployment. The use of 16-19 and 20-24 bands, instead of a 16-24 band, will mitigate this somewhat.

3.4 Weighting – alternative methods

Various alternative weighting methods exist. One approach would be to use multiple partitions, which would allow even finer weighting groups – perhaps, for example, single year of age between 16 and 24, in common with the standard LFS weighting. However, this would be complex to implement (requiring the use of GES software) and maintain. This may be worth investigating in the future.

Another approach would be to include an element of non-response (and, for the LFS, attrition) adjustments to the design weights, which might help account for differing unit non-response mechanisms between surveys. These adjustments would have to be based on auxiliary information which is available for both responders and non-responders – modelling non-response using output area classification (OAC) is fairly usual, since OAC is available for non-responders. However, this would again add considerable complexity – on the Integrated Household Survey, this step frequently went wrong or contained errors, and made little difference to estimates.

Finally, an interesting approach would be to utilise two-phase weighting – using survey estimates as control totals in the calibration. However, as the pooled dataset itself is likely to have the smallest sampling variability of any Northern Ireland survey, two-phase weighting would be unlikely to improve precision. It might be possible to utilise pooled survey estimates to calibrate other surveys (particularly the component surveys), and this could be the subject of further research.

4 Summary – Methods Recommended

We recommend -

- Hot-deck imputation using imputation classes defined by five-year age-bands, sex, and whether an individual is an owner-occupier; and
- Post-stratification weighting using weighting groups defined by New local authority*Age-band*Sex

While these methods are appropriate to the pooled dataset and will remove bias and reduce sampling variability, it is important to emphasise that non-sampling error will differ between the surveys in a fashion which is impossible to measure, and this may have an impact on estimates which these methods cannot remove.

References

Chu, A. (1997) “Weights for Combining Surveys across Time or Space”, available at http://www.stat.fi/isi99/proceedings/arkisto/varasto/kalt0185.pdf

Elliot, D., Rainford, L., Eldridge, J. (2006) “Interview Mode Effects in the UK Local Labour Force Survey”, paper contributed to TSMII, available at http://www.ons.gov.uk/ons/guide-method/method-quality/general-methodology/data-collection-methodology/reports-and-publications/interview-mode-effects-in-the-uk-local-labour.pdf

Roberts, G. & Binder, D. (2009) “Analysis Based on Combining Similar Information from Multiple Surveys”, available at https://www.amstat.org/sections/srms/Proceedings/y2009/Files/303934.pdf