TTP DATA QUALITY REPORT

DATA SOURCE	
DATA SUPPLIER:	Economic and Labour Market Statistics (ELMS)
DESCRIPTION:	2011-2018 extract for individuals on the Annual Survey for Hours and Earnings (ASHE)
COVERAGE:	Northern Ireland
NUMBER OF RECORDS:	Individuals within the ASHE dataset may have more than one job within each year. The dataset will contain a record for each job for these individuals if their workplace has been selected for the survey.

Year	Records	Individuals
2011	5,770	5,528
2012	5,824	5,551
2013	5,485	5,218
2014	5,511	5,353
2015	5,793	5,623
2016	5,933	5,782
2017	5,711	5,595
2018	5,601	5,491

DATA BACKGROUND

The Annual Survey of Hours and Earnings (ASHE) is managed by the Economic and Labour Market Statistics (ELMS) branch for Northern Ireland and is a UK wide business survey that provides estimates on hourly, weekly and annual earnings for employee jobs. The ASHE is a 1% sample taken from the HM Revenue and Customs Pay as You Earn System. Employee jobs are weighted to population totals taken from the Labour Force Survey based on classes defined by occupation, region, age and sex.

ASHE does not cover the self-employed, jobs within the armed forces nor does it cover employees not paid during the reference period. So this may mean, the survey does not fully cover certain types of seasonal work, for example, employees taken on for only summer or winter work.

Validation is carried out on returned data that is regarded as incomplete or potentially inaccurate, based on automatic comparisons with data for similar jobs or against data for the same job in previous years. Selective editing is also used - a technique for prioritising suspicious values for follow-up based on their impact on published results. In these cases, respondents may be re-contacted in order to verify the information that has been provided.

ASHE uses the principal UK coding schemes for classifying occupation and industry. Occupations are coded using the Standard Occupational Classification: SOC 2010 and industry using the Standard Industrial Classification 2007: SIC 2007.

Further information on the quality and background of ASHE data is available on the ONS website.

DEMOGRAPHIC INFORMATION

Information available to the TTP for linkage purposes -

- Name information which includes Surname, Forename initial and mid-name initial
- Home and Work Postcode
- Sex
- Date of birth
- Occupation Code (SOC 2010)
- Standard Industrial Classification 2007 (SIC 2007)

DATA QUALITY & PRE-PROCESSING

Record linkage is highly dependent on the quality of the data being linked. Many key variables can be presented quite differently between datasets as the information has been collected and recorded using different methods. This can greatly complicate record linkage unless understood ahead of time. Pre-processing can be used to standardise these variables into a consistent format prior to linkage.

NAME INFORMATION- Name information is available for 99.6% of records from 2011 to 2018. The name field includes forename initial, midname initial and full surname in some but not all cases:
- 92.7% of records contain forename initial.
- 69.7% of records contain midname initial.
- Surname is 99.6% populated.

- Punctuation and characters which do not appear in the English alphabet (e.g. Á , %, $ ...) can appear in the name information. A function can be used to replace these with usable characters.

ADDRESS INFORMATION- Address information is limited to postcode i.e. House number, street name and town is not available. 96.2% of records have a Northern Ireland home postcode assigned to them. A small number of records (0.4%) had a home postcode outside of Northern Ireland. UPRN is unavailable due to address information being limited to postcode.

Work postcode is available in the ASHE dataset, with a small number of records (1.3%) having work postcode missing.

- Address and Postcode information is standardised prior to linkage.

GENDER- Information relating to the gender of the individual is available for all records.

- Gender information should be standardised prior to linkage.

DATE OF BIRTH- Information is available for 100% of records

- Date of birth will be parsed into its constituent parts (day, month and year) prior to linkage

MATCHING METHODS

Using the demographic information provided the following matching can be undertaken-

- Person matching

PREVIOUS LINKAGE MATCH RATES

The 2011 ASHE data provided by ELMS has been matched to the Census 2011 population. This produced a match rate of 85%. This match rate can be increased to 89% by supplementing the 2011 data with 2012-2018 data and using extra variables such as work postcode, SIC07 and Occ10 with additional manual matching.

EVALUATION

The 2011 ASHE data provided by Economic and Labour Market Statistics branch produces a lower match rate in comparison to other datasets used in ADR projects. Subsequent years (2012-2018) may produce slightly higher rates, but if the ASHE data is to be used for other projects manual checking and additional variables may be required to increase match rate to a satisfactory level.

Document Management

Access Limitations:	NISRA TTP, RSU, Data Suppliers, Researchers
Maintainer:	Caroline Anderson
Document Identifier:	ADR NI
Replaces:	
Review period (months):	N/A
Is related to:	N/A

Version History

Version	Notes	Last Amended
1.1	Created by NISRA TTP, 18/03/2020	