GUIDANCE DOCUMENT FOR RESEARCHERS CGIAD 2019/20

Contents

1.0 Introduction	3
2.0 About the component data	3
2.1 Overview	3
2.2 Collection of Data	4
2.2.1 DfC & DWP	4
2.2.2 HMRC	4
2.2.3 NIHE	4
2.3 Variables	4
2.4 What CGIAD is not	5
2.5 Caveats	5
2.5.1 Missing Data	7
2.5.2 Welfare Supplementary Payment (WSP)	7
3.0 CGIAD Summary Figures	8
4.0 Research Potential	8
5.0 Accessing the Data	9
5.1 Overview of process	9
5.2 Restricted Variables	10
6.0 CGIAD Variable List - Detail	10
7.0 Acknowledgements	13
Annex A – CGIAD Variables	15
Annex B – Glossary of Abbreviations	18

1.0 Introduction

This document provides guidance to researchers on the Cross Government Income Administrative Dataset (CGIAD). The CGIAD is an aggregate, de-identified household level pre-linked dataset for Northern Ireland (NI) and will be made available through the Administrative Data Research for Northern Ireland (ADR NI). The CGIAD was developed by the Department for Communities' Analytics Division (DfC-AD) working with Northern Ireland Statistics and Research Agency's Research Support Unit (NISRA-RSU). The dataset has been created using the Digital Economy Act 2017 as the legal gateway.

The intended audience of the CGIAD will be researchers who are accredited by ADR NI and who have a research project/proposal that is suitable for analysis from this dataset.

Included in this document is background information on the data sources that contribute to the CGIAD, a description of the concepts and intended usages of the dataset, and how researchers can access the data.

Also included is a description of variables and caveats that are essential for the understanding of the CGIAD structure and its intended usage, what it should not be used for and why. When applying to use the CGIAD data this document should be used in conjunction with the CGIAD Metadata and Data Dictionary (the variable list can be found in Annex A).

The first iteration of the CGIAD will be completed in 2024. The database may be subject to various structure changes as time progresses which is likely to be attributed to the addition of data variables. These changes will be reported in future versions of this guidance document and will aid the understanding of potential future builds.

The owners of this document are the DfC-AD and NISRA-RSU, and any changes to the document should be suggested to NISRA-RSU who will liaise with DfC-AD to make amendments if required.

2.0 About the component data

2.1 Overview

The CGIAD is an aggregate, de-identified household level pre-linked dataset containing income data from social security benefits, child benefits, housing benefits, earnings, tax credits and savings. It is predominantly an annual demographic snapshot (in this instance, reflecting the position in March/April 2020) with some annualized information for the 19/20 financial year.

The CGIAD is a standalone dataset and cannot be linked to other datasets.

2.2 Collection of Data

The CGIAD is a pre-linked dataset. Individuals are linked by National Insurance number (NINO) across the various data scans. Households are created based on address information using the first line of the address and postcode. The CGIAD for the 2019/20 year contains approximately 729,000 households. Based on comparisons with both the Family Resources Survey and NISRA's Household Projections, this indicates that it covers around 98% of NI households. Cross sectional information is collected for the dataset for the financial year 2019/20 from four main data providers: the Department for Communities (DfC), His Majesty's Revenue and Customs (HMRC), the Department for Work and Pensions (DWP) and the Northern Ireland Housing Executive (NIHE). This includes social security benefits data administered by DWP on behalf of DfC, and earnings and employment information from HMRC.

2.2.1 DfC & DWP

The Department for Work and Pensions' (DWP) provides NI social security benefit data to DfC AD for statistics, research, fraud and error prevention and management information purposes. Further information on the benefits administered can be found on the <u>Department for Communities'</u> website.

2.2.2 HMRC

His Majesty's Revenue and Customs (HMRC) data, namely earnings, savings, Child Benefit and Tax Credit information is supplied by DWP (on behalf of HMRC) to DfC.

2.2.3 NIHE

DfC AD require Housing Benefit information from the Northern Ireland Housing Executive (NIHE) for statistics, research, and operational uses.

2.3 Variables

The CGIAD contains 71 variables. A household unit is defined as one or more individuals who live in the same dwelling. The variables included cover:

Income: household income variables covering overall income, benefit income and income of
the highest earner in the household. These variables may be yearly or weekly figures as
appropriate. Income is net of tax and National Insurance contributions.

- Benefits: Number of adults in the household in receipt of each benefit.
- Geography: Lowest geography level provided is Local Government District (LGD). There are some geography variables included through deprivation quintile (from the NI Multiple Deprivation Measures 2017) and settlement type.
- Household: descriptor variables on the number, age bands and the sex of adults and dependents living in the household; and
- Other: contextual household data e.g. number of adults working or self-employed.

A full list of variables which are included in the CGIAD is presented in Annex A.

2.4 What CGIAD is not

- CGIAD is not a longitudinal database. The current version of CGIAD is for the 19/20 year.
 Demographic information is taken from March/April 2020 scans.
- CGIAD does not track changes throughout the year. It reflects the position at the point in time (March/April 2020) with some annualized information e.g., total benefit amount, income etc. Households are not tracked over time using CGIAD.
- CGIAD is most suitable and accurate when considering income research at the household level. This has been determined by the validation exercises which have generally shown the household statistics compare well with other household data.

2.5 Caveats

There are a number of caveats and notes to be aware of when considering examination of the CGIAD.

- Households where all individuals are not in receipt of income from employment, and no
 individual is in receipt of benefits are not represented in the dataset.
- If a child in the household is in receipt of Disability Living Allowance (DLA) but we have no parental information, then their data is removed.
- When compared to NISRA's household projections, CGIAD overestimates the number of one person households and underestimates the number of two person households.

- Individuals are linked to households using address and the accuracy of this is heavily dependent on the address details being accurate, complete, and up to date.
- Communal establishments are not included within the CGIAD. There is a lack of centralized information that exists regarding communal establishments:
 - DfC have been unable to source a detailed list of folds and other sheltered housing that provides information on cooking facilities available so that they can be identified as communal and therefore removed.
 - O It is possible that not all the establishments that have been removed were communal. If there are more than six adults living at an address the address is also assumed to relate to a multiple occupancy household (MoH) and therefore is excluded from analysis. However, some of these households will not be MoH; rather addresses have not been updated resulting in multiple records for a private household.
- Child Benefit data is the source of information on the number of children within a
 household. As higher earners may not receive Child Benefit, the number of children in CGIAD
 for higher earning households is likely to be underestimated due to a lack of data source for
 this cohort.
- Adults generally are identified as being those 18 and over, however there are a small
 number of cases where 16- and 17-year-olds are included on the CGIAD as adults. This will
 be the case if they are present within the data from the data providers in their own right and
 another adult in the household is not claiming Child Benefit for them. Additionally, there will
 also be instances where 18- and 19-year-olds are classified as dependents (rather than
 adults) if another adult receives Child Benefit on their behalf.
- There are other known data gaps, such as self-assessed income relating to assets
 (landlord/rental income). There is a lack of sufficient information on individual assets and
 any incomes generated from those assets including rental income, dividends, and capital
 gains.
- No information relating to tenure is available. All analysis (e.g. equivalisation of household income and poverty calculations) has been conducted on a 'Before Housing Costs' basis.

- Some variables will have their access restricted (detailed in section 5.2) in that they will not
 be able to be used in addition to other variables to minimize the risk of disclosure. In these
 instances, there are restrictions on the number of variables in a variable group that can be
 requested.
- There are certain benefits/payments that aren't included e.g., Winter Fuel Payment and Social Welfare Fund. A full list of benefits/payments that are present within the CGIAD is included within the variable list in Annex A.

2.5.1 Missing Data

The CGIAD is created from a number of datasets taken from four data providers (as detailed in section 2.2). All care is taken to ensure that all the data received is as accurate and complete as possible. However, it must be noted that there will be some level of missing information across various datasets which may result in missing information within the CGIAD. It also may not always be possible to link the data together to provide the picture at household level. It should be noted that a number of variables in the dataset, that have component variables, will not always sum to the total variable detailed due to missingness, e.g. missing ages. Age information is missing for 2.0% of adults across 3.6% of all households within the CGIAD. This missingness also impacts other age variables. For example, age information is missing for 2.4% of adults in receipt of benefits across 3.1% of all households within the CGIAD. A small percentage of adults, 0.2% across 0.3% of all households, are missing sex information.

2.5.2 Welfare Supplementary Payment (WSP)

From 2016 Welfare Supplementary Payments were introduced by the Northern Ireland Executive for the purpose of helping those who have incurred financial disadvantage through reduction, or termination, of current social security benefits as a result of welfare reform. The range of Welfare Supplementary Payment schemes include:

- The Benefit Cap;
- Time-Limiting of Contributory Employment and Support Allowance;
- Personal Independence Payment which may also affect disability payments and/or carer payments;

- Social Sector Size Criteria; and
- Universal Credit.

Depending on their circumstances, people may have been eligible for more than one Welfare Supplementary Payment.

Welfare Supplementary Payments are NI specific payments designed to offset the impacts of welfare reform on certain welfare benefits. In the CGIAD these payments are flagged withing the dataset and included in overall income, however they are not included within 'benefit' variables.

Further information on Welfare Supplementary Payments can be found online (<u>Support if you're affected by welfare changes | nidirect</u>).

3.0 CGIAD Summary Figures

The CGIAD for the 2019/20 year contains approximately 729,000 households. Based on comparisons with both the Family Resources Survey and NISRA's Household Projections this indicates that it covers around 98% of NI households. Table 1 contains summary figures for the CGIAD including the number of adults in employment, adults with benefit receipt and households in absolute/relative poverty from the CGIAD.

	Total Households	Households in Absolute Poverty		Total Adults	Adults Employed	Adults with Benefit Receipt
2019/20	729,000	112,000	140,000	1,376,000	876,000	752,000

Table 1: CGIAD 2019/20 Summary Statistics

4.0 Research Potential

The CGIAD has been developed in order to advance research across a range of themes. The CGIAD can offer an evidence base for policy development and decision making as well as the monitoring of public policy in Northern Ireland. The CGIAD will support research aligned with the Department's key outcomes which include:

The delivery of an anti-poverty strategy supported by evidence-based proposals.

- Collaborating with other Departments and delivering with partners on cross-cutting issues, including homelessness, place and deprivation-based approaches, and employability to deliver a sustainable and effective set of anti-poverty interventions.
- Enhancing citizen focused services and interventions through an ongoing process of customer insight, data analytics, research, and analysis.

This section will be updated over time as increased research potential becomes evident.

5.0 Accessing the Data

5.1 Overview of process

Access to the CGIAD dataset is limited to researchers and research teams with <u>accredited researcher</u> <u>status</u> under the Digital Economy Act 2017 (DEA) and subject to approval by the data providers. In addition, all projects must be accredited by the UK Statistics Authority Research Accreditation Panel (RAP).

In the first instance, researchers should complete the Themed Datasets Initial Research Proposal Form and submit this to rsu@nisra.gov.uk. The proposal form asks researchers to summarise the main aim and objectives of the proposed project as well as the methodological approach that they are intending to use. Researchers must also include details of the data required (including a variable list with justification) and cohort required. Details of any public engagement activities planned as part of their project must also be included. RSU, in consultation with the data providers, will determine the feasibility of the project prior to the completion of a full application by the researcher(s) to RAP. Ethical approval will also be required for each project, either through the research team's university ethics committee or via the National Statistician's Ethics Advisory Committee. RSU will support the research team throughout the application process. Further information on how to apply for approval is available here.

The CGIAD dataset may be accessed in the RSU secure research environment.

Upon approval by RAP and DfC, researchers will be provided with the relevant variables in a flat file table. The data will be made available to researchers in NISRA's secure environment. All research outputs will be cleared by RSU.

As noted previously, the CGIAD is a standalone dataset and cannot be linked to other datasets.

5.2 Restricted Variables

The following three geographic variables will not be made available together to researchers. A maximum of two of these variables can be requested together. This is to reduce the risk of a household (or individual) being identified directly or indirectly due to the specific combination of variables provided.

LGD_HH – Detailing the Local Government District (LGD) where the household is located. There are eleven LGDs in Northern Ireland. Where an LGD is not available, this will be marked as 'Unknown'.

SETT_HH – Indicating the settlement band where the household is located. This is identified using the Urban/Rural classification of the household as per NISRA methodology.

INC_DEPV – Provides the income deprivation quintile (1-5) where the household is located. This is identified from NISRA's Multiple Deprivation Measures 2017.

6.0 CGIAD Variable List - Detail

The variables in the CGIAD dataset contain information from social security benefits, child benefits, housing benefits, earnings, tax credits and savings.

Annex A contains the CGIAD variable list, with further information on a number of the variables detailed below:

Absolute Poverty – Listed as ABS_POV; An individual is considered to be in absolute poverty if they are living in a household with an equivalised income (EQI) below 60% of the (inflation adjusted) UK median income in 2010/11. In 2019/20, the absolute poverty threshold for a couple with no children is £15,698 per year. For a couple with children the EQI would be higher and for a single person (without children) the EQI would be lower.

Age Group Household – AGE_GROUP_HH; Households are defined an age group based on the composition of the adults in their respective household. Households with only working age adults are classed as working age. Those households with only pension age adults are classed as pension age. The households with both pension age and working age adults are classified as mixed age.

Age of the Individual with the Highest Income in the Household – AGE_INDV_HI; This will provide the age group of the adult individual receiving the highest income within the household.

Benefit Unit – A benefit unit is defined to be a single adult or a married or cohabiting couple and any dependent children. A household may contain multiple benefit units. It should also be noted that a benefit unit does not necessarily receive any benefits – it refers to the structure of the unit.

Dependents – Listed as DEPEN; The CGIAD will consider those between the ages 0-16 as dependents as well as those between 16 and 19 who are still in full time education. It will not include adults who may be considered dependents such as those who are disabled or vulnerable elderly people.

Employment Variables.

There are three categories of employment variables: Individuals who are employed (these are classified as employees or workers), self-employed, and those who are both employed and self-employed. These are defined each with five corresponding variables - a total count for each employment category and a breakdown into four age groups which are populated with the number of individuals fitting the employment and age criteria.

If you wish to get the sum of people categorised as being an employee/worker (in employment), you need to sum the relevant employment variable and the relevant employed and self-employed variable. Similarly, to get the number of self-employed, you need to add relevant self-employment variable and the relevant employed and self-employed variable. To count total individuals who are employed regardless of status you need to sum all three variables.

The employed variable counts individuals who have an income greater than 0. Those individuals who are self-employed are counted as long as the record for self-employment is not null. This is because people in this category can submit tax returns with negative income from self-employment and can submit returns showing their income as zero.

Highest Individual Income in the Household – Listed as AMT_INDV_HI; This is a weekly figure measuring the income garnered from the adult individual within the household who receives the highest income.

Income Deprivation Quintile – INC_DEPV; This variable is based on the income domain of NISRA's NI Multiple Deprivation Measures 2017¹. Within this, the 890 Super Output Areas (SOAs) are ranked from most deprived (1) to least deprived (890) in terms of income. Within the CGIAD, where a household is noted as being in quintile 1, this means that it is located within a SOA ranked between

¹ <u>Northern Ireland Multiple Deprivation Measure 2017 (NIMDM2017) | Northern Ireland Statistics and Research Agency (nisra.gov.uk)</u>

1 and 178 according to NISRA's income domain ranking. Similarly, quintile 2 means that the household is located within a SOA ranked between 179 and 356 and so forth.

Relative Poverty – Listed as REL_POV; An individual is considered to be in relative poverty if they are living in a household with an equivalised income below 60% of the UK median income of the year in question. For 2019/20, the relative poverty threshold for a couple with no children is £17,151 per year. For a couple with children the EQI would be higher and for a single person (without children) the EQI would be lower.

Savings – SAVINGS; This will provide a flag for households which have been detected to have savings.

Settlement Band – Listed as SETT_HH; This considers whether the address of the household is urban or rural. This is based on NISRA methodology which can be found here.

Sex of the Individual with the Highest Income in the Household – SEX_HI; This will show the sex of the adult individual in the household who receives the highest income.

Total Equivalised Household Income – EQ_INC_HH; This figure measures the annual (net) equivalised household income. This comprises total income from the available sources of all household members including dependants. Income is adjusted for household size and composition by means of equivalence scales, which reflect the extent to which households of different size and composition require a different level of income to achieve the same standard of living. For example, a household with three adults will need a higher income than a single person living alone (to enjoy a comparable standard of living). This adjusted income is referred to as equivalised income. The main equivalence scales, as used within DfC's NI Poverty and Income Inequality publication, are the modified OECD scales, which take the values shown in the following table.

Type of Household Member	OECD Modified Scale	
Head of Household	0.67	
Additional Adult	0.33	
Child aged 14 and over	0.33	
Child aged 0-13	0.20	

Equivalence scales conventionally take an adult couple without children as the reference point, with an equivalence value of one. The process then increases relatively the income of single person

households (since their incomes are divided by a value of less than one) and reduces relatively the incomes of households with three or more persons, which have an equivalence value of greater than one.

An example of this in practice can be found when considering three different households; a single person, a couple with no children, and a couple with two children aged fourteen and ten, all having unadjusted weekly household incomes of £200 (before housing costs). The process of equivalisation in each case involves dividing £200 by the appropriate equivalence scale depending on the size of the household. Consequently, this results in an equivalised income of £299 to the single person (scale of 0.67), £200 to the couple with no children (scale of 1), but only £131 to the couple with children (scale of 1.53). Further information on equivalisation can be found within the Quality and Methodology documentation for the NI Poverty and Income Inequality Report².

Total Household Income – Listed as INC_HH; This is an annualised (net) figure measuring the nominal (before equivalisation) income figure. Included within the Total Household Income is income from benefits, employment and self-employment. Not included is income from investments, savings and other capital sources/mechanisms.

In terms of rounding, all variables denoting a monetary amount are rounded to the nearest appropriate number. For most variables, this will be to the nearest £10.

7.0 Acknowledgements

When publishing or presenting findings from the CGIAD the following acknowledgement should be used:

Administrative Data Research - Northern Ireland (ADR-NI) takes privacy protection very seriously. All information that directly identifies individuals/organisations will be removed from the datasets by trusted third parties before researchers access them.

All researchers using ADR-NI are trained and accredited to use sensitive data safely and ethically, they will only access the data via a secure environment and all their findings will be vetted to ensure they comply with strict confidentiality requirements.

The help provided by the staff of the Administrative Data Research Centre - Northern Ireland (ADRC-NI) and the ADR-NI support officers within NISRA Research Support Unit (RSU) is acknowledged.

² Northern Ireland Poverty and Income Inequality report 2022-23 | Department for Communities (communities-ni.gov.uk)

ADR-NI is funded by the Economic and Social Research Council (ESRC). The authors alone are responsible for the interpretation of the data and any views or opinions presented are solely those of the author and do not necessarily represent those of ADR-NI. The data provided by the Department for Communities, His Majesty's Revenue and Customs, the Department for Work and Pensions and the Northern Ireland Housing Executive has been supplied for the sole purpose of this project.

Annex A – CGIAD Variables

Variable	Variable Name	Variable Description
Number		
1	ID_HH	Anonymised Unique Household Identifier
2	BU_HH	Number of benefit units in the household
3	TOT_ADL	Number of adults in the household
4	ADL_18_24	Count of the number of adults in the household that are aged 18- 24
5	ADL_25_49	Count of the number of adults in the household that are aged 25-49
6	ADL_50_64	Count of the number of adults in the household that are aged 50-64
7	ADL_65_OV	Count of the number of adults in the household that are aged 65+
8	MALES_ADL	Count of the number of adult males in the household
9	FEMALES ADL	Count of the number of adult females in the household
10	DEPEN	Number of dependents in household - that is those who are aged
		0-16, or 19 and under where someone is receiving Child Benefit
		for them
11	DEPEN_0_4	Dependents Age Indicator - aged 0-4
12	DEPEN_5_15	Dependents Age Indicator - aged 5-15
13	DEPEN_16_OV	Dependents Age Indicator - aged 16+ (16-19)
14	LGD_HH	Local Government District of the household home address
15	SETT_HH	Settlement Band for Urban/Rural Classification (HH)
16	INC_HH	Total annual income for household
17	AMT_INDV_HI	Highest individual income in household (weekly)
18	AGE_INDV_HI	Age band of individual with highest income in household
19	SEX_HI	Sex of individual with highest income in household
20	EQ_INC_HH	Total equivalised household income (annual)
21	INC_DEPV	Income Deprivation Quintile 1, 2, 3, 4, 5
22	ABS_POV	Absolute Poverty Indicator (Y/N)
23	REL_POV	Relative Poverty Indicator (Y/N)
24	BEN_AMT_HH	Total benefit income for the household (weekly)
25	BEN_AMT_DEPEN	Total benefit income of dependents for the household (weekly)
26	BEN_AMT_HI	The amount of benefit received by the individual in the
		household with the highest overall income
27	BEN_BU	Count of the number of benefit units in the household that are in receipt of benefits
28	TOT_BEN_ADL	Count of the number of adults in the household that are in
20	TOI_DEN_ADE	receipt of benefits
29	BEN_ADL_18_24	Count of the number of adults in the household that are in
4 3	DLIN_ADL_10_24	
		receipt of benefits aged 18-24

ehold that are in
ehold that are in
ehold that are in
ension Age, Unknown
ehold that are in
,
ehold that are in
and are aged 18-24
ehold that are in
and are aged 25-49
ehold that are in
and are aged 50-64
ehold that are in
and are aged 65+
ehold that are in self-
ehold that are in self-
ehold that are in self-
ehold that are in self-
ehold that are in self-
household that are in
ehold that are in both
e aged 18-24
ehold that are in both
e aged 25-49
ehold that are in both
e aged 50-64
ehold that are in both
e aged 65+
ehold who are in
e in receipt of
re in receipt of

53	CA_ADL	Number of adults in the household who are in receipt of Carer's Allowance
54	CB_ADL	Number of adults in the household who are in receipt of Child Benefit
55	CTC_ADL	Number of adults in the household who are in receipt of Child Tax Credit
56	DLA_ADL	Number of adults in the household who are in receipt of Disability Living Allowance
57	ESA_ADL	Number of adults in the household who are in receipt of Employment and Support Allowance
58	HB_ADL	Number of adults in the household who are in receipt of Housing Benefit
59	IB_ADL	Number of adults in the household who are in receipt of Incapacity Benefit
60	IS_ADL	Number of adults in the household who are in receipt of Income Support
61	IIB_ADL	Number of adults in the household who are in receipt of Industrial Injuries Benefit
62	JSA_ADL	Number of adults in the household who are in receipt of Jobseekers' Allowance
63	MA_ADL	Number of adults in the household who are in receipt of Maternity Allowance
64	PC_ADL	Number of adults in the household who are in receipt of Pension Credit
65	PIP_ADL	Number of adults in the household who are in receipt of Personal Independence Payment
66	RP_ADL	Number of adults in the household who are in receipt of Retirement Pension
67	SDA_ADL	Number of adults in the household who are in receipt of Severe Disablement Allowance
68	UC_ADL	Number of adults in the household who are in receipt of Universal Credit
69	WSP_ADL	Number of adults in the household who are in receipt of Welfare Supplementary Payments
70	WB_ADL	Number of adults in the household who are in receipt of Widows Benefit
71	WTC_ADL	Number of adults in the household who are in receipt of Working Tax Credit

Annex B – Glossary of Abbreviations

ABS	Absolute
ADL	Adult
AMT	Amount
BEN	Benefit
DEPEN	Dependants
DEPV	Deprivation
EMP	Employed
EQ/EQI	Equivalised Household Income
НН	Household
HI	Highest Income
INC	Income
INDV	Individual
OP	Occupational Pension
OV	Over
POV	Poverty
REL	Relative
SELF	Self-Employed
TOT	Total