An Phríomh-Oifig Staidrimh Central Statistics Office

An Roinn Leanaí, Míchumais agus Comhionannais
Department of Children, Disability and Equality

Growing Up in Ireland

Summary Data Guide for

Cohort '98 at 25 years

Wave 5

SuirbhéannaCSO
CSOSurveys

Version 1
May 2025

Fás Aníos in Éirinn
Growing Up in Ireland

Table of Contents

Chapter 1 Introduction

1.1 Introduction

This document provides the reader with a brief summary of the fifth wave of Cohort '98 (at age 25 years) from Growing Up in Ireland (GUI), as well as an overview of the microdata files (Researcher and Anonymised) from that round of the study.

GUI - the national longitudinal study of children - is the first project of its kind undertaken in Ireland. GUI aims to describe the lives of children and young people and to identity key factors that help or hinder their development. A two-cohort longitudinal design was originally adopted. Cohort '98 recruited and interviewed 8568 nine-year-olds and their families in 2007/2008. Cohort '08 recruited and interviewed the families of 11,134 nine-month-olds in 2008. As the project is longitudinal in nature, both cohorts are being interviewed on a number of occasions and their parents / guardians were interviewed previously when the children were nine years of age, thirteen years of age and recently at seventeen/eighteen years of age, 20 years of age and 25 years of age (subject of this report). The families were interviewed when the children were nine-months, three years, five years, seven years and thirteen years of age. A series of reports, summary Key Findings and peer reviewed papers has been produced from both cohorts. In September 2024, the Central Statistics Office (CSO) and the Department of Children, Disability and Equality (DCDE) launched a third cohort, Cohort '24. Around 15,000 – 16,000 households with nine-month-old infants have been invited to take part.

The 8,568 children in the were born between 1st November 1997 and the 31st of October 1998. Data collection for the current wave of GUI Cohort '98 (age 25) took place between April 2023 and April 2024 and resulted in a complete data set of 3380 cases (plus 500 cases who completed the emigration survey).

This report describes in detail the background, design, instruments and procedures used only in respect of Cohort '98 Wave 5. The focus here is on the sample design and response rate, the nature and content of the questionnaires and other instruments, along with a broad overview of the dataset.

1.2 Background

GUI provides important input to the implementation of national strategies for children and young people such as Better Outcomes, Brighter Futures: The National Policy Framework for Children and Young People, 2014 – 2020, and Young Ireland: The National Policy Framework for Children and Young People 2023 – 2028. The principal objective of the study is to provide evidence-informed research into children and young people's well-being. This increased understanding of the determinant and drivers of well-being and its change and transformation over time will be used to assist in policy formation and in the design and delivery of services for young people and their families.

GUI is the national longitudinal study of children and young people in Ireland commissioned by the Irish Government. The study is a joint project by the CSO and the DCDE. DCDE has responsibility for the wider elements of the GUI study: engaging with policy and scientific stakeholders; consulting with children/young people; identifying research needs, data priorities and policy objectives, producing research needs reports for each new wave, and promoting the use of GUI data for research and policy development. The CSO is responsible for the GUI survey itself: translating needs into operational requirements; designing and building the survey; collecting, processing, and analysing the data, disseminating the findings in the form of statistical releases, and facilitating data access to researchers and policy makers. Working closely together ensures seamless integration of the complimenting responsibilities.

The conceptual framework for GUI draws heavily on the bio-ecological model developed by Urie Bronfenbrenner (Bronfenbrenner & Morris, 2006). Under this model, a person's development is the result of interactions between them ('bio') and their environment ('ecological'). The environment is made up of multiple 'systems' – both close to the person, such as their family and friends, and wider such as the society or culture in which they grow up. These different systems impact a person in different ways. In Bronfenbrenner's model, there are five systems:

- Microsystem – This is the person's immediate environment, and the activities, roles and relationships experienced within it. The microsystem can be the person's family, or a different setting where the individual has regular, consistent contact with other people. For 25-year-olds this could be their workplace, or their friends. At this age the microsystem might change rapidly from family to housemates to partner. Direct experience in this system has more influence on the individual than indirect experience from the other systems.
- Mesosystem – This system includes the links between different microsystems in the person's life. For example, the mesosystem could include the interaction between home and work, or between work and study.
- Exosystem –The individual is not directly involved in this system, but it still influences their experiences and affects them. For 25-year-olds, an exosystem could be a partner's job or a child's childcare.
- Macrosystem – This is mainly the cultural or societal structure for the individual which can contain multiple micro, meso, and exo systems. The macrosystem will influence the 25-year-olds interest in societal concerns, or involvement in certain types of risky behaviours.
- Chronosystem – This includes not only the ageing and maturing of the individual but also the time in which the person lives. For the 25-year-olds in Growing Up in Ireland Cohort '98, they were nine when Ireland entered recession. They were teenagers when social media entered their lives. They were 22 when the world experienced Covid-19.

GUI is a critical instrument by which policy makers and researchers can examine the factors which shape the development of children in contemporary Ireland and, through this, to contribute to the setting of responsive policies and the design of services for children and their families. In September 2024, the CSO and DCDE launched a third cohort, Cohort '24. Around 15,000 – 16,000 households with nine-month-old infants will be invited to take part over the next year.

The main policy and research domains of interest for Cohort '98 at 25 were defined via a stakeholder engagement exercise and published by DCDE in a research needs report. The themes that the final survey instrument addressed were:

1. Physical Health
2. Socio-Emotional Well-being and Key Relationships
3. Education and Income
4. Civic and Economic Participation

By providing evidence on these domains, GUI facilitates policymakers in their creation of a robust support system for 25-year-olds, fostering a generation that is healthy, emotionally resilient, well-educated, and civically engaged.

Chapter 2 The Sample and Data

2.1 Introduction

This chapter considers the methodology and sample design for Wave 5 of Cohort ’98 at 25 years of age. Consideration is given to the composition of the longitudinal sample, followed by discussion of the levels of inter-wave attrition and procedures for statistically reweighting the data to ensure that they are representative of the population

2.2 Frame and Sample Design

GUI is a longitudinal study based on a fixed panel design. Cohort '98 follows the children and their families who were recruited into the study at nine years of age for re-interview on several subsequent occasions. In the original design, the sampling frame was the register of primary schools in Ireland with a two-stage sampling design: a random sample of schools was selected from the frame, with a sample of nine-year old children then drawn from the selected schools. The design ensured that the sample was regionally representative and free from spatial bias. The 2006 Census of Population recorded 56,497 nine-year-olds in the country. The selected sample size of 8,568 accounted for approximately 14% of this population, equating to about one in every seven nine-year-olds nationwide.

After the initial sample selection at nine years of age, no additions were made to the sample. The children and their families were re-interviewed at 13, 17, and 20 years of age. There was a special Covid survey run in December 2020 when Cohort ’98 were aged 22. By 25 years of age the sample represents the respondents and their families who were resident in Ireland at nine years of age and who continued to live in the country when they were 25 years old.

At Wave 1 of Cohort ’98, 8,568 nine-year-olds and their families were interviewed. All these families were approached for re-interview when the respondent was 13 years old with 7,525 families participating in Wave 2. At the third round of interviewing, when the respondents were aged 17/18 years, 8,277 families were approached for re-interview. This included families who had been interviewed at Wave 1 but had not participated in Wave 2. There were 6,216 responding households at Wave 3. In Wave 4, questionnaires were completed by 5,190 20-year-olds representing a 65% response rate. A parent of the respondent was also interviewed if available. If the parent was not available, data from the 20-year-old was still collected. Wave 5 represents the first time that data from only the 25-year-old respondent and no other household members were included.

For the current Wave 5, all previous respondents were invited to be re-interviewed unless the family had previously definitively refused to be contacted in future waves of the study or was not eligible (i.e. the whole family had moved abroad, or the respondent was sadly deceased). 7,870 respondents were approached for re-interview in Wave 5.

2.3 Data Collection and Processing

Respondents received a letter inviting them to complete the survey online. The letter contained a personal secure code to access their questionnaire. Reminder letters were sent out in six-week intervals.

Respondents who did not self-complete the online survey were visited by an interviewer and offered an in-person interview. The online survey remained open during this time; some respondents opted to compete the interview online, others in-person with the interviewer. The final mode breakdown for the data collection was 87.1% of respondents completed the interview online and 12.9% of respondents completed the survey in-person.

2.4 Response Rate

The achieved response rate for eligible respondents was 49.5% (achieved sample of 3,380). A breakdown of the sample is provided in the below table.

Approached for interview in Wave 5	**7870**
Ineligible for interview at Wave 5 (includes emigrated, deceased and other)	1035
Completed emigration survey - 500	
Eligible Sample	**6835**
Interview Completed Complete -2836 Partial - 544	3380
Non-responders Refusal - 667 Other non-response - 2789	3346

2.5 Weighting and non-response management

In line with best practice in sample surveys the data has been reweighted or statistically adjusted to ensure that the sample is wholly representative of the population from which it has been drawn. By doing this we ensure the structure of the completed sample is in line with the structure of the population along key socio-demographic and other dimensions.

The data file contains two weighting factors. The weighting factor to be used in analysis of the 25-year-old sample is w5_Wgt_a. The weighting factor to be used with the sample of those

who have participated in all 5 waves of Cohort ’98 is w5_Wgt_b (see below for more information).

The variables xxwave1, xxwave2, xxwave3, xxwave4 and xxcovid_survey indicate if the case has data for each of the waves. A value of one indicates participation at the relevant wave.

2.5.1 Unit non-response

Unit non-response occurs when individuals who are part of the population, and are eligible for the survey, do not provide the requested information. This can lead to biased survey estimates if specific groups within the population are over- or under-represented and if these groups behave differently with respect to the survey variables. To correct for this, a non-response adjustment is used.

A threshold was applied to separate partial responses into those that were usable and suitable for imputation to address item missingness, and those where the item level missingness was deemed too high and the case was essentially one of unit non-response. If a respondent had fully completed the first section of the survey (questions relating to housing), they were included as a usable partial response. Seventy-five cases were excluded from the final dataset based on this threshold. The final dataset included 3,380 respondents (2,836 completed surveys and 544 partially completed surveys) and 3,456 non-respondents.

A non-response adjustment was then created using population figures for the entire eligible population based on the CSO administrative data and data collected from Cohort '98 Wave 1. Logistic regression models identified the variables most predictive of non-response, with adjustments made to avoid excessively large sampling variances. These variables were:

- Sex
- Highest level of education obtained.
- 2022 income, divided into quintiles.
- Wave 1 region (NUTS3)
- Wave 1 family type
- Drumcondra Reading Test score at age nine, divided into quintiles.
- Whether a respondent was still attending an institution of higher education (as under the auspices of the Higher Education Authority) for the 2022/2023 academic year.

The R package Icarus (Rebecq, 2016) was then used to create the resultant non-response weights using these variables.

As in previous waves of Cohort ’98, the base weight was the most recent weight that applied to a participant in a Cohort ’98 data collection, which in 82.7% of respondents was the weight used in Wave 4. If a respondent did not participate in Wave 4 their weight was taken from the most recent wave in which they participated – which is broken down as follows:

Last Participating Wave	Number of Respondents	% of Respondents
Wave 1 (at age 9)	51	1.5%
Wave 2 (at age 13)	168	5.0%
Wave 3 (at age 17)	367	10.9%
Wave 4 (at age 20)	2794	82.7%

As in Wave 3 and Wave 4 of Cohort '98 two separate weights, both deriving from the same population estimates, have been calculated. First is the cross-sectional weight which applies to all participants in Wave 5 (w5_Wgt_a). The second weight, referred to as the longitudinal weight, is the weight to be applied for cross-wave analysis across the cohort (w5_Wgt_b). For this reason, only respondents who have participated in every wave to date are given a longitudinal weight, which is a population now of 2,594.

As in previous waves, all weights have been truncated to avoid excessively large sampling variances – here they have been bounded to one quarter of the mean at the lower end and four times the mean at the higher end.

2.5.2 Item non-response

Item non-response occurs when a respondent does not provide an answer to a specific question in the survey. This leads to gaps in the dataset, which can be filled by the method of imputation.

Rates of item level missingness per variable ranged between less than 1% to 20%.

For the main questionnaire, single imputation using a K-nearest neighbour model was employed using the R package VIM (Kowarik & Templ,2016). All variables in the main questionnaire which were categorical, ordinal, or numeric and with no or very little dependencies had their missingness imputed. The accuracy of the models used was tested using a set of core variables against the pre-existing survey data until an accuracy rate of 90% or higher was achieved. Overall, 58 variables in the main questionnaire have imputed values. Any imputed variables have an accompanying variable named "x_imp" where x is the associated variable name – values that have been imputed are labelled as 1 and non-imputed values are labelled as 0.

Further imputation resulted from a detailed analysis of multiple related questions within the survey and/or carrying variable level data forward from previous survey waves. For example, sex, height, and household relationships if at same address and with same household structure, were carried forward from the previous wave if missing where appropriate. Similarly,

previous wave data was used to replace erroneous values entered by the respondent. Note: for analysis purposes, where the respondent sex was missing or not disclosed the record from wave 1 was carried forward for the variable c98w5_Sex_w1forward.

In the case of missing administrative data, hot deck imputation was applied based on age, sex, primary and secondary school location at age nine and age 13 respectively.

Where imputation or other methods as above to resolve item missingness were not possible or effective, the missing values have been re-coded for the purposes of this release as “Not Stated” or similar, to maximise the statistical power of the dataset.

Chapter 3 Instrument Development and Piloting

3.1 Instrument Design

This chapter describes the various groups of experts and others who had input into the development of the instruments and procedures used in Wave 5 of GUI Cohort '98. The questionnaires were developed by the CSO, the DCDE Study Team along with the input from the Scientific Advisory Group, selected policy experts and through consultations with a number of young people. Further information on the consultation processes can be found in the Research Needs Report for GUI at 25 published by DCDE.

3.2 Piloting the Instruments

The pilot of the 25-year data collection instrument and exercise was undertaken via computer assisted web interview (CAWI) and in-person computer assisted interviews (CAPI) in the respondents home. The pilot survey took place between the 15th of June and the 24th of September 2022. Letters were sent out in advance to inform the 200 pilot respondents of the data collection and invite them to take part. Along with these advance letters, reminder letters, emails and texts were issued during this period and "knock-to-nudge" exercises were carried out whereby field staff visited the respondent's home to ask them to complete the CAWI questionnaire.

To boost the pilot sample size, 200 additional cases from the main sample were invited to take part in the pilot survey. These 200 cases were reinterviewed for the main survey also. This resulted in an overall target pilot sample of 400 cases. A total of 124 cases were completed (70 CAPI and 54 CAWI) representing a 31% response rate.

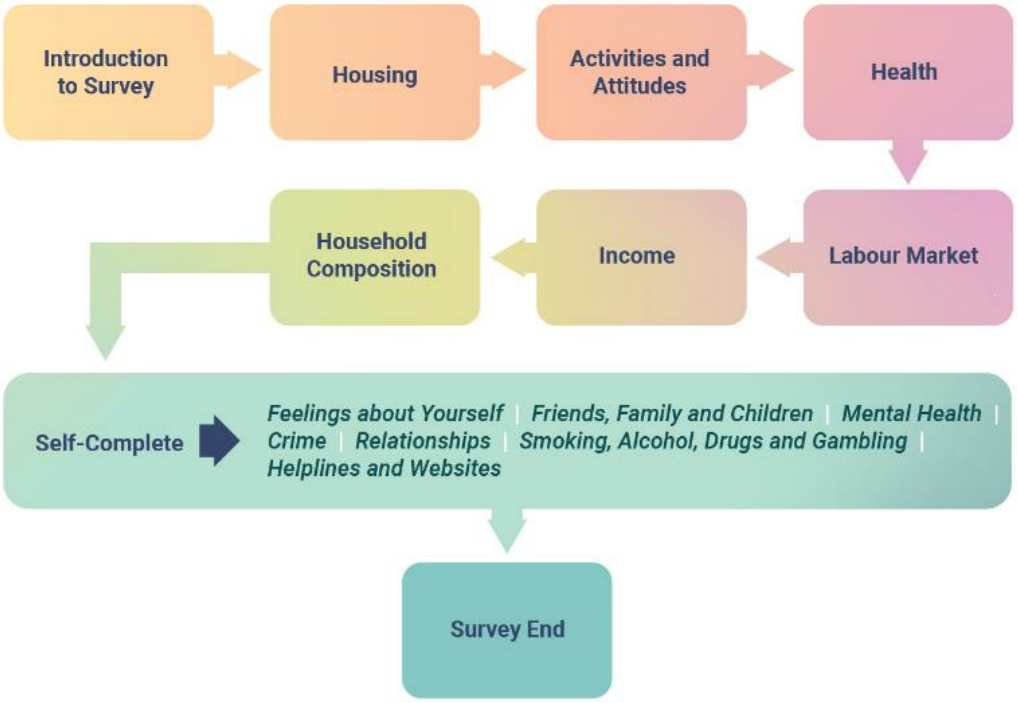
Chapter 4 Survey Structure

4.1 Questionnaire

The questionnaire used in GUI Cohort '98 at age 25 was structured in two broad sections: the main questionnaire and the sensitive questionnaire (see Figure 1 for questionnaire flow). The survey content was not randomised in order. The main questionnaire contained some initial screening questions that identified the current geographic location of the respondent; whether living in Ireland or living outside the state. If the respondent was not living in Ireland (non-resident) at the time of undertaking the survey, they were invited to complete a short emigration module. The emigration module contained questions regarding when they moved abroad, what country they moved to, the reason for emigrating and if/when they intended to return.

Those who responded they were living in Ireland proceeded directly to the main questionnaire which contained survey questions on housing; activities and attitudes; health; labour market activity/engagement, and household composition. This was then followed by the sensitive questionnaire.

The sensitive questionnaire was preceded by an information box that summarised the content of this part of the survey and gave the respondent an opportunity to 'opt out', as well as a reminder that they could skip any questions if they did not wish to answer. The sensitive questionnaire then contained survey questions on feelings; friends, family and children; how the respondent felt about things; criminal justice system experiences; bullying; identity, relationships, and sexual experiences; and smoking, alcohol, drugs, and gambling. A list of helplines was provided at the end of the sensitive module.



Source: Central Statistics Office

Figure 1. Overview of Growing Up in Ireland Cohort '98 at age 25 survey instrument.

The survey was first issued via CAWI (i.e. a web-based self-complete instrument). Those who did not respond to the CAWI instrument were then followed up with CAPI interviews whereby the interviewer called to the household. The online survey remained open during this time; some respondents opted to compete the interview online, others in-person with the interviewer. For the CAPI survey, the main questionnaire was completed with an interviewer and the sensitive questionnaire was completed by the respondent in private. The questions contained on the CAPI and CAWI survey were the same bar one exception due to a routing issue in Blaise. This exception is the Drug_use_list question ("Which of the following have you taken in the last year?"). Those who completed the CAWI survey were asked "Which of the following have you taken in the last year?" with the option to select "Not taken", "Taken less than 5 times in the last year" or "Taken 5 or more times in the last year". Those who completed the CAPI survey were asked "Which of the following have you taken in the last year?" but were not asked the frequency. To reflect this difference between the two instruments a 4th category has been added to the Drug_use_list question on the sensitive questionnaire document "Taken in the last year, frequency unknown" to reflect those who completed the CAPI instrument.

The variable "Mode" indicates if the respondent completed the CAWI or CAPI survey.

Chapter 5 Structure and Content of the Data File

5.1 Introduction

This section outlines the structure of the Research Microdata File (RMF) and Anonymised Microdata File (AMF) and provides a brief explanation of how the two data files differ in content. An overview is given of variable naming conventions and identification ('id') codes used. The variables relating to the statistical weights, derived variables, physical measurements and scaled measures are described. Finally, the coding, editing, and forward feed from previous waves are discussed.

The CSO would advise that the data are used in conjunction with the Questionnaire Documentation. This gives a broad overview of the topics included in the data file. Users should however note that for the purposes of preparation and anonymisation there may be differences in value labels between the questionnaires and the data files and not every question from the questionnaire is included in the data file. This is especially true for the AMF.

5.2 Structure of the Data Files

The AMF is a publicly available anonymised dataset, while the RMF is a more detailed dataset, access to which is subject to appointment as an Officer of Statistics by the CSO. The case-base is the 25-year-old. Blocks of variables appear in the data as follows (variable prefixes are shown in brackets):

- Main and Sensitive Questionnaires (c98w5_)
- Household Grid (c98w5_xxxp1, c98w5_xxxp2 etc.)
- Standardised Scale Scores (c98w5_)
- Physical Measurements (c98w5_)
- Derived Variables (c98w5_D_)

5.3 Differences between Anonymised (AMF) & Research (RMF) Microdata Files

To protect the anonymity of respondents, personally identifiable variables such as names and date of birth, and open text variables, were removed from both types of file. In addition, some variables with a higher risk of being statistically disclosive were either removed or had their values banded into larger groups so that frequencies with low cell counts are not visible. In some cases, this was achieved by either bottom or top coding (or both) of outlying cases. In others, continuous scores have been grouped into categories.

These anonymisation steps were especially applied to the AMF, with far more variables being top and/or bottom coded, or re-categorised. In addition, some potentially disclosive variables which appear on the RMF have been removed from the AMF, and information particularly likely to be sensitive in nature (i.e. the majority of the variables in the sensitive questionnaire) has been removed from the AMF.

The user should therefore note that not every question from the questionnaires is included in the data files, particularly in the case of the AMF. A list of variables included in each data file is available via the accompanying summary data dictionary.

5.4 Variable Naming

Variables for Wave 5 of Cohort '98 are prefixed with 'c98w5_' to indicate that the data is from Wave 5 of Cohort '98.

Exceptions to this variable naming convention are:

- 'ID' – unique respondent label
- 'w5_Wgt_' weights applied to the dataset
- 'xx' prefix - indicates if the case has data for each of the waves
- 'Mode' – indicates whether a CAPI or CAWI questionnaire was completed
- 'c98w5_D_' – derived variable prefix

Any imputed variables have an accompanying variable named "x_imp" where x is the associated variable name – values that have been imputed are labelled as 1 and non-imputed values are labelled as 0.

Note – the variable names on the associated questionnaires do not have these prefixes.

5.5 Identification Codes

Each respondent has a unique identification code, which is the same at all waves to enable matching of the data files over time.

5.6 The Household Grid

The household grid contains information on members of the household i.e. who lives in the household, their person number on the grid, gender, relationship to the 25-year-old, age and principal economic status.

For ease of reading, the household grid variables are suffixed with the person number. For example, the variable indicating the sex of the person on line 1 of the grid is 'c98w5_sexp1' where 'c98w5_' indicates Wave 5 of Cohort '98.

The reader should note that (for anonymisation purposes) exact dates of birth have been removed from the AMF and RMF and replaced with age in years.

5.5.1 Derived Variables on Household Grid

For confidentiality reasons, only some of the variables from the household grid are provided on the data files, and a number of derived variables that are based on the household grid are included instead. For example, the details of person 8 and person 9 (such as age and sex) in the household grid were too disclosive to include in the data files, however these would have been included in the household gird count variables.

In some cases, where the household grid was not completed, for example, ages of some members of the household were not filled in, these individuals would not have been included in the household grid counts for the age brackets and therefore there may be inconsistencies in the data where the household grid data was incomplete.

There was an inconsistency in response to whether respondents shared income with their children, especially small children who would not have an income. The variable c98w5_D_ShareIncomeAdultshsd was created to remove this discrepancy. No value in the derived household grid counts indicates the household grid was not complete enough to calculate.

The variables derived from the household grid are described in more detail in the derived variable document.

5.7 Data Linkage and Administrative Variables

CSO uses primary micro data sources in its statistical programmes to complement or replace survey data, to make its statistical operations more efficient or to create new insights or products. These data enable CSO to fill information needs about the Irish society, economy, and environment, reduce response burden and costs imposed by surveys, and improve data quality and timeliness. All data obtained by CSO are used solely for statistical purposes. The CSO continues to work with suppliers of primary micro data sources to ensure good quality data is available on a timely basis.

The linkage and analysis were undertaken by the CSO for statistical purposes in line with the Statistics Act, 1993 and the CSO Data Protocol.

Before using personal administrative data for statistical purposes, the CSO removes all identifying personal information including the Personal Public Service Number (PPSN). The PPSN is a unique number that enables individuals to access social welfare benefits, personal taxation, and other public services in Ireland. The CSO converts the PPSN to a Protected Identifier Key (PIK). The PIK is a unique and non-identifiable number which is internal to the CSO. Using the PIK enables the CSO to link and analyse data for statistical purposes, while protecting the security and confidentiality of the individual data. Administrative data records were linked using the PIK for this project.

The primary micro data sources which have been linked to the GUI Cohort '98 at age 25 survey data are Revenue Commissioner's Income Tax Form 11, PAYE income data, Department of Social Protection (DSP) social welfare data, Higher Education Authority (HEA) data, Quality and Qualifications Ireland (QQI) data, the State Exams Commission (SEC) and SOLAS data on apprenticeships. The education administrative data for the 2021-'22 academic year includes nearly all higher and further education courses in Ireland in both state and private bodies. It does not include many courses carried out under the Springboard program, many profession-specific qualifications which are administered by internal bodies, attendance at any secondary education institutions in Ireland that do not follow the standard national curriculum and do not

carry out the Leaving Certificate as part of their standard program, or any non-higher education related education or training which may have been undertaken outside of Ireland. Information on training undertaken abroad was collected via GUI survey questions.

If education administrative records could not be linked for a survey respondent, an assumption was applied that they had achieved a qualification up to NFQ Level 3 (Junior Certificate or equivalent). This reflected 3.4% of the weighted dataset.

Median weekly income in the "Income and Cost of Living" section was calculated broadly in line with the methodology outlined in the CSO publication Earnings Analysis using Administrative Data Sources (EAADS).

For the purposes of computing this variable, one primary employment for each respondent was utilised, defined as the highest earning employment in that calendar year. Certain types of employment were excluded:

- Employment in NACE Sector A (Agriculture, Forestry and Fishing)
- Employment in NACE Sector T (Activities of households as employers; undifferentiated goods-and services-producing activities of households for own use)
- Employment in NACE Sector U (Activities of extraterritorial organisation and bodies).
- Self-employment.
- Employees earning less than €500 per annum.
- Employments where the duration was less than two weeks in the year.
- Missing employer and employee reference numbers.

Weekly median income was calculated by dividing the gross annual wages for the primary occupation by the number of weeks worked in that employment. Weeks worked were defined as PRSI-insurable weeks. Employment that met the above criteria and occurred at any time of the year was considered eligible. This contrasts with the EAADS methodology, which primarily includes only employment active in October of the relevant year.

Data was also collected from the 2022 Census including variables on religion and whether or not the respondents spoke Irish.

For further details, see the derived variable documentation.

5.8 Derived Variables

A number of variables were derived to provide additional information on the circumstances of the household. These variables pertain to response at each wave of the survey; social class; income for the 25-year-old; and household and family composition. Other derived variables include counts of different variables (such as the number of property problems a respondent stated they have per the question c98w5_property_problem_1 to c98w5_property_problem_5). A full list of the derived variables can be found in the derived variables documentation.

5.9 Physical Measurements

5.9.1 Height

The height of the respondent (c98w5_heightcm) was recorded by the interviewer electronically on the CAPI programme and self-reported by the respondent if they completed the CAWI instrument.

The data collected was edited to remove implausible heights. If height was not collected the height from the last wave was forward fed (going back as far as wave 3; otherwise if not available the cell value was coded as 'missing').

5.9.2 Weight

The weight of the respondent (c98w5_weightkg) was recorded electronically on the CAPI programme (by the interviewer) or self-reported if the respondent completed the CAWI questionnaire. The data collected was edited to remove clearly implausible outliers.

5.9.3 Body Mass Index (BMI)

BMI scores were derived from the recorded heights and weights. The BMI score was also recoded into the following categories – underweight, normal weight, overweight and obese. These correspond to the cut-off points used in previous waves aligning to the Health Services Executive cut-offs (as of January 2025).

Note that for the AMF, the height and weight measurements were top-and-bottom coded, and BMI was subsequently re-calculated on these revised values. This has resulted in small differences between the calculated BMI for some cases on the AMF and RMF.

5.10 Scaled Measures in the Study

Scales in GUI were chosen in alignment with study objectives and established reliability and validity. These scales were scored using protocols provided by the authors and are briefly described below.

5.10.1 Rosenberg Scale Calculations for Self-Esteem Score

Self-esteem was measured using the Rosenberg Self-Esteem scale (Rosenberg, 1989). The original ten item Rosenberg Self-Esteem scale was reduced to six items rated on a four-point scale ranging from 1 to 4. Higher scores are indicative of higher global self-esteem (c98w5_D_selfesteem).

5.10.2 FAST Scale Calculations for Hazardous Levels of Drinking

The Fast Alcohol Screening Test (FAST; Hodgson, John, Alwyn, Hodgson, Waller, Thom & Newcombe, 2002) is a short screening tool for alcohol misuse. It consists of a subset of questions from the full Alcohol Use Disorders Identification Test (AUDIT). It consists of four items – females are asked how often they have six or more units of alcohol on one occasion

and males are asked how often they have eight or more units of alcohol. It produces a total score and a categorisation of alcohol misuse as 'hazardous' or 'not hazardous':

- Drinking class according to FAST (c98w5_D_fastscore)
- Total on FAST for males (c98w5_D_fastotm)
- Total on FAST for females (c98w5_D_fastotf)

5.10.3 AUDIT Scale Calculations

The AUDIT (Saunders, Aasland, Babor, de la Fuente and Grant, 1993) is a screening tool developed by the World Health Organization (WHO) to determine if a person's alcohol consumption may be harmful. A total score of the items is calculated to determine the likelihood of hazardous or harmful alcohol consumption, and alcohol dependence. In prior GUI waves, AUDIT was calculated not including those who stated they did not drink alcohol. However, in this wave these individuals were included in line with WHO scoring recommendations.

5.10.4 DASS Scale Calculations

In the GUI survey, stress at age 25 years is measured using the Depression, Anxiety and Stress Scale (DASS; Lovibond & Lovibond, 1995) which is a stress subscale from the shortened version DASS-21(Henry & Crawford, 2005). The DASS-21 stress subscale contains seven items assessing difficulty relaxing, nervous arousal and being easily upset/agitated, irritable/over-reactive and impatient. The seven items are rated on a four-point scale with responses of: did not apply to me at all; applied to me to some degree; applied to me a considerable degree and applied to me very much.

5.10.5 Everyday Discrimination Scale Calculations

The Everyday Discrimination Scale (EDS) Short Version (Sternthal, Slopen, Williams, 2011) is a 5-item measure asking participants to indicate how frequently they feel they have experienced various forms of interpersonal mistreatment in their day-to-day lives, assessed on a six-point scale (0=never, 1= less than once a year, 2= a few times a year, 3= a few times a month, 4= at least once a week, 5= almost every day). A total discrimination score was generated from the sum of all five items. Higher scores are indicative of more frequent discrimination (c98w5_D_EDS).

5.10.6 CES Depression Scale Calculations

The Centre for Epidemiological Studies Depression scale (CES-D; Radloff, 1977) is a widely used self-report measure that was developed specifically as a screening instrument for depression in the general population, as opposed to be a diagnostic tool that measures the presence of clinical depression. GUI used the 8-item short version of the CES-D 8 and provides a total score for the respondent.

Also included in the file is a variable (c98w5_D_cescat) which categorised respondents into 'depressed' or 'not depressed'. It is again noted that this is based on the CES-D 8 screening tool and does not purport to be a clinical measure.

5.11 Classifications

5.11.1 Principal Economic Status Classification

The Principal Economic Status (PES) classification, which is also used in the Labour Force Survey (LFS) and the Census of Population, is based on a single question in which respondents are asked what their usual situation is regarding employment. The answer is given in response to the following categories:

- At work
- Unemployed
- Student
- Engaged on home duties
- Retired
- Other

5.11.2 NACE Industrial Classification

Respondents to the GUI survey are asked the industrial sector of their main employment. This is converted to a two or three digit-NACE code.

5.11.3 National Framework of Qualifications (NFQ)

The NFQ was launched in 2003 and it is now the single structure for recognising all education and training in Ireland. All framework awards now have an NFQ Level, numbered from 1 to 10, which tells you about the standard of learning and an NFQ Award-Type which tells you about the purpose, volume and progression opportunities associated with a particular award. Quality and Qualifications Ireland (QQI) has responsibility to develop, promote and maintain the Irish NFQ.

In this release educational attainment results are presented using a descriptive name and the corresponding NFQ levels.

For information on the NFQ see: National Framework of Qualifications (NFQ)

5.11.4 Occupation Classification

Occupation has been recoded to the new classification SOC2010 which is also the standard used in the LFS and Census. Previous occupation classifications for GUI were ISCO-88 and SOC1990.

5.11.5 NUTS2 and NUTS3 Regions

The regional classifications in this release are based on the NUTS (Nomenclature of Territorial Units) classification used by Eurostat. Until the final quarter of 2017, the NUTS3 regions corresponded to the eight Regional Authorities established under the Local Government Act, 1991 (Regional Authorities) (Establishment) Order, 1993, which came into operation on 1 January 1994 while the NUTS2 regions, which were proposed by Government and agreed by Eurostat in 1999, were groupings of those historic NUTS3 regions.

However, the NUTS3 boundaries were amended on 21st of November 2016 under Regulation (EC) No. 2066/2016 and have come into force from the first quarter of 2018. These new groupings are reflected in the GUI data from this wave. As a result of these changes, Louth moved from the Border to the Mid-East and what was formerly South Tipperary was amalgamated with North Tipperary and moved from the South-East to the Mid-West.

The new NUTS2 and NUTS3 regions are:

Northern & Western NUTS2 Region		Southern NUTS2 Region		Eastern & Midland NUTS2 Region	
NUTS3	**Constituing Counties**	**NUTS3**	**Constituing Counties**	**NUTS3**	**Constituing Counties**
Border	Cavan Donegal Leitrim Monaghan Sligo	**Mid-West**	Clare Limerick Tipperary	**Dublin**	Dublin City Dun Laoghaire-Rathdown Fingal South Dublin
		South-East	Carlow Kilkenny Waterford Wexford	**Mid-East**	Kildare Louth Meath Wicklow
West	Galway Mayo Roscommon	**South-West**	Cork Kerry	**Midland**	Laois Longford Offaly Westmeath

5.11.6 Urban/Rural

The Census 2022 urban-rural definition was applied when classifying whether a respondent lived in an urban or rural area using the respondents Eircode. In certain instances where an Eircode was not available for the given household, this variable was set to missing. The Census definition of an urban area is a town with a total population of 1,500 or more and therefore towns with a population of less than 1,500 are included in rural areas. This definition is different to that used in previous waves and so caution is advised when comparing urban/rural variation across waves.

5.12 Coding and Editing

Regular checks were carried out on the data as they were returned from the field and inconsistencies dealt with on an on-going basis. Any anomalous or impossible responses which could not be imputed from previous waves or other sources were removed and coded as 'missing'. Specific examples are provided here:

Upon review people who said they worked full-time but said they worked less than 20 hours a week had their c98w5_hours_worked answer removed.

Upon review of the data, if certain answers appeared to be incorrect (for example, implausible sleep figures or amount of cigarettes smoked) these answers were removed and appear as missing in the dataset.

Due to a routing error in Blaise, three respondents who stated they had always lived in Ireland answered some of the living abroad questions (c98w5_lived_abroad_num , c98w5_emig_emp, c98w5_emig_family, c98w5_emig_spouse, c98w5_emig_educ, c98w5_emig_travel, c98w5_emig_other). These answers were set to NA (missing).

Note: for the questions c98w5_activities_historyC1-C6, the total number of months in the past 5 years may add up to greater than or less than 60, for example, if the respondent was in both education and employment in a given period or if the respondent took part in a different activity during this time period that wasn't captured by these answer options.

Some survey variables were presented as open questions to capture verbatim responses that would have been difficult to pre-code. Where relevant, these open-ended responses were coded into separate categorical variables after the interview. Other questions did have a pre-defined coding frame but also had an 'other specify' option for those responses which did not fit into any of the pre-coded categories - again answers were recorded on a verbatim basis by the interviewer (CAPI) or typed in directly by the respondent (CAWI). In this instance responses to these questions had to be recoded with additional categories. The newly coded responses for additional codes or variables appear in the RMF dataset. All verbatim text from the original responses has been removed from the AMF and RMF as a safeguard to protect the respondent's identity.

6 Bibliography

Bronfenbrenner, U. and Morris, P.A. (2007). The Bioecological Model of Human Development‡. In Handbook of Child Psychology (eds W. Damon, R.M. Lerner and R.M. Lerner).

Henry, J. D., & Crawford, J. R. (2005). The 21-item version of the Depression Anxiety Stress Scales (DASS–21): Normative data and psychometric evaluation in a large non-clinical sample. *British Journal of Clinical Psychology*, *44*, 227–239.

Hodgson, R.J., John, B., Alwyn, T., Hodgson, R.C., Waller, S., Thom, B. & Newcombe, R. (2002). Fast Screening for Alcohol Misuse: Manual for the FAST Alcohol Screening Test. London: Health Development Agency.

Lovibond, S.H. & Lovibond, P.F. (1995). Manual for the Depression Anxiety Stress Scales. (2nd. Ed.) Sydney: Psychology Foundation.

Kowarik A, Templ M (2016). "Imputation with the R Package VIM." *Journal of Statistical Software*, 74(7), 1–16.

Radloff LS. A Self-Report Depression Scale for Research in the General Population. Appl Psychol Meas. 1977;1(3):385–401.

Rebecq, A (2016). "Icarus: an R package for calibration in survey sampling." R package version 0.2.0.

Rosenberg, M. 1989. Society and the Adolescent Self-Image. Revised edition. Middletown, CT: Wesleyan University Press.

Saunders J.B., Aasland O.G., Babor T.F., de la Fuente J.R. and Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption II. Addiction 1993; 88:791-804.

Sternthal, M., Slopen, N., Williams, D.R. "Racial Disparities in Health: How Much Does Stress Really Matter? Du Bois Review, 2011; 8 (1): 95-113.