Study Number (SN)	0019-04
Title	Growing up in Ireland Infant Cohort Wave 4 – 7/8 years, 2016
Depositor	The Economic and Social Research Institute (ESRI)
Publisher	Irish Social Science Data Archive (ISSDA)
URL	http://www.ucd.ie/issda/data/guiinfant/guiinfantwave4/
Publication Date	2018
Version	1
Type	Dataset
Suggested citation	The Economic and Social Research Institute (ESRI). (2018). *Growing up in Ireland Infant Cohort Wave 4 – 7/8 years, 2016*. [dataset]. Version 1. Irish Social Science Data Archive. SN: 0019-04. URL http://www.ucd.ie/issda/data/guiinfant/guiinfantwave4/

A Summary Guide to

Wave 4

of the Infant Cohort

(At 7/8 Years)

of

Growing Up in Ireland

Eoin McNamara, James Williams, Aisling Murray and Amanda Quail

Contents

1. Introduction

1.1 Introduction

This document provides a summary of the fourth wave of data collection with the ***Growing Up in Ireland*** (GUI) Infant Cohort (at 7/8 years), as well as an overview of the microdata files (Researcher and Anonymised) from that phase of the project.

Growing Up in Ireland - the National Longitudinal Study of Children is a landmark study of children and youth which has been running since 2006. The objectives of the study are outlined in a separate publication (Greene et al. 2010) but can be summarised as seeking to collect data on what it means to grow up a child in Ireland, with a view to informing policy on what both helps and hinders development. A two cohort, cross-sequential longitudinal design was adopted and began with one cohort (the Infant Cohort) of 11,134 infants (aged 9 months) and a second cohort (the Child Cohort) of 8,568 9-year olds. Being longitudinal in nature, the same children are followed over time. The families of the Infant Cohort have been interviewed when the children were 9 months, 3 years and subsequently 5 years of age, while the Child Cohort and their parents/guardians were interviewed at 9, 13 and 17/18 years of age. This document concerns a postal survey of the Primary Caregivers of the children in the Infant Cohort when those children were 7/8 years old.

The children of the Infant Cohort were born between 1st December 2007 and the 30th June 2008 and were aged 9 months at the time of the first data collection between September 2008 and April 2009. Over 10,000 families participated in the first wave (n=11,134) while 9,793 took part at age 3 years (Dec 2010 - July 2011), and 9,001 at age 5 years (Mar – Sep 2013). The current fourth wave of data collection, a postal questionnaire, took place in the spring of 2016, when the cohort was 7/8 years of age and was completed by 5,344 families. More details of response and attrition rates can be found in Section Two.

This report describes the design, instruments and procedures used in respect of Wave 4 of the Infant Cohort. Earlier waves of this cohort (and the Child Cohort) are the subjects of a parallel set of reports.

1.2 Background

The study began in 2006 and is funded by the Department of Children and Youth Affairs, in association with the Central Statistics Office. A contribution is also being received from the Atlantic Philanthropies. It is being carried out by a group of researchers led by the Economic and Social Research Institute (ESRI) and Trinity College, Dublin (TCD).

The central participants are the cohort members or 'Study Children' and, until they are old enough to be interviewed themselves, the principal information about them is provided by the adult acting as their Primary Caregiver (typically, but not exclusively, their mother). At earlier waves (depending on the age and as applicable to the child), information was also collected from the child's Secondary Caregiver, teacher and school principal. In the current wave, however, the postal questionnaire was sent only to the Primary Caregiver.

Growing Up in Ireland is a multi-disciplinary study which collects information on different areas of the child's life including health and physical development, socio-emotional well-being, and learning/education. It is conceptually guided to a significant degree by Bronfenbrenner's bio-

ecological model (e.g. Bronfenbrenner & Morris, 2006) which acknowledges the variety of influences on the child's development – from family to cultural context – and the importance of interactions between the individual and these different influencers. Full details on the underlying theoretical and conceptual framework can be found in Greene et al., 2010.

2. Sample Design

2.1 Introduction

In order to provide the reader with an overview of the sampling procedures used in *Growing Up in Ireland (GUI)*, this section provides a brief outline of the sample designs at the first, second and third waves of data collection with the Infant Cohort. The sample design at Wave 4, along with response and attrition rates, will be discussed in detail. The process of statistically reweighting the data to ensure that they are fully representative of the population will also be outlined.

2.2 Sample Design at Wave 1, 2 and 3

Full details on the population, sampling frame and sample design for the Infant Cohort are given in a dedicated publication entitled *Sample Design and Response in Wave 1 of the Infant Cohort of Growing Up in Ireland*; https://www.ucd.ie/t4cms/GUI-SampleDesignResponseInfants.pdf. This subsection provides a brief outline of the sampling at Wave 1, to provide the reader with a background to the sampling procedures used in **GUI**.

The Child Benefit register was used as the sampling frame to select potential respondents into the project at 9 months of age. The advantages of using this administrative database as a sampling frame were: (a) it contained a comprehensive up-to-date listing of eligible members of the relevant population; (b) had a wide range of relevant characteristic variables and (c) was already in an electronic form which could be accessed with relative ease.

There were a total of 41,185 infants registered on the Child Benefit Register as having been born between 1st December 2007 and 30th June 2008. Children for inclusion in the Study were sampled over this seven month reference period, with a view to carrying out fieldwork for Wave 1 when they were 9 months of age, between September 2008 and March/April 2009. The sample was selected on a systematic basis, pre-stratifying by mother's marital status, county of residence and nationality of payee as well as number of children in the claim - all variables which were available from the information recorded on the Benefit Register. A simple systematic selection procedure based on a random start and constant sampling fraction was used. The final completed Wave 1 sample was 11,134 infants and their families, and this formed the target sample for Wave 2. In this respect the study is based on a pure, fixed panel of children who were 9 months of age at the time of first interview. The cohort members have not been 'topped up' at any point thus far, although some individuals who missed a wave at either age 3 or 5 years (or both) have returned for the current and perhaps later waves.

At Wave 2 (aged 3 years), the target sample consisted of the 11,134 children and families who participated at Wave 1, with 9,793 completing an interview. At Wave 3 (aged 5 years), the target sample comprised both the 9,793 families who completed at Wave 2 but also any who had

participated at Wave 1 (but not Wave 2) unless the Study Team had been informed of the family's definite decision to withdraw or had firm reasons to believe the family had emigrated and hence were no longer part of the sample. The number of participating families at Wave 3 was 9,001. Further details on the sample design for previous waves can be found in the corresponding reports for that data collection (available at http://www.ucd.ie/issda/data/growingupinirelandgui/).

2.3 Sample Design at Wave 4

The target population for sampling at Wave 4 was made up of the children and families who participated in Wave 2 and/or Wave 3, as well as most of those who participated at Wave 1 but refused or otherwise did not participate at one or both of the next waves due to family circumstances at that time (e.g. due to the birth of a new baby or temporary absence from the country during the fieldwork period). Families who had moved abroad, moved within Ireland with no forwarding address, or had requested at Wave 2 or Wave 3 to be removed from the study, were not issued at Wave 4. Thus the Wave 4 sample had four components: those children and families who participated in all three earlier waves of the study; those who had participated only in Wave 1; those that participated in Waves 1 and 2; and those children and families who had participated in Waves 1 and 3. Just over 95 per cent of the families at Wave 4 had participated in all previous waves, while approximately 1 per cent had participated at Wave 1 but not at Wave 2 or 3. Two per cent of the Wave 4 study sample completed all except Wave 2, and a final 2 per cent completed all but Wave 3.

2.4 Response Rates

As noted above, the survey at the fourth wave of the study was implemented on a mixed mode (postal and phone) basis. A total of 10,317 families were selected for the first mail shot. This was made up of families who had participated in any previous round of the study, who had not requested that their names be removed from the sample and whom the Study Team understood from their most recent records to be still living in Ireland. The first mailshot to these families contained an introductory letter and questionnaire (available at http://www.esri.ie/growing-up-in-ireland/information-for-participants/participant-information-for-the-infant-cohort/). This was sent to the person identified as the Study Child's Primary Caregiver at the family's most recent face-to-face interview (mostly those which took place when the Study Child was 5 years of age).

The first mailshot in the 7/8-year survey was issued between the last week of February and first week of March 2016. A reminder letter was sent to 7,522 families between March and April 2016, the remaining 2,795 having already returned their completed questionnaire in the post in response to the first mailshot. Between April and June of 2016, a second reminder letter and questionnaire was issued to 5,444 families who had not returned the completed form at that time.

Following the postal phase, the Study Team phoned 534 non-respondent families in June and July 2016 to encourage them to participate, with a fourth copy of the questionnaire being sent to them at that time, where necessary. These families who were included for this phone phase were identified as those who were felt to be most likely to participate at that time – largely on the basis of their response histories to date.

By the end of fieldwork a total of 5,344 usable questionnaires were returned to the Study Team. This means that a crude response rate of 52 per cent was achieved in this postal round of the project. This response rate does not take account of the families who no longer lived in Ireland at

the time of the survey, and whose letters were returned by An Post as being unknown at the last address then available to the Study Team. Many of the non-respondents may no longer have been living in Ireland at the time of the survey and so should have been excluded from the response rate (from the denominator). The Study Team is not in a position to estimate how many target respondents had, in fact, left Ireland and so the estimated 52 per cent response rate is a lower bound.

2.5 Attrition

Inter-wave attrition (or non-response) is unfortunately unavoidable in panel surveys, regardless of tracking and conversion procedures employed. Attrition becomes a particular problem where it is systematically related to family or other characteristics. Watson and Wooden (2009), for example, note that it may be systematically associated with respondents': sex; age; race/ethnicity; marital status; household composition and size; educational attainment; labour force status; and family income. They found that, on average, attrition is higher among males; younger respondents; minority groups; one-parent and non-marital households; lower educated families; economically active; and low income families. It is important to understand the levels and correlates of attrition and non-response to inform reweighting procedures for statistical adjustment of the data prior to analysis.

As noted above, the target sample for the postal phase at 7/8 years of age was the continuing longitudinal sample which was initially interviewed when the Study Child was 9 months of age. That sample was further approached for interview when the Study Child was 3 and subsequently 5 years of age. This means that the postal phase was the fourth time the families had been approached for interview. The patterns of responses across the four waves is summarised in Table 2.1 below.

Table 2.1: Response patterns of participants in the 7/8-year postal survey across the four waves of the Infant Cohort, 9 months to 7/8 years of age.

Family participated at:				
9 months	3 years	5 years	7/8 years	Number
Respondents at 5 years				
Yes	Yes	Yes	Yes	5,086 (95.2%)
Yes	Yes	No	Yes	121 (2.3%)
Yes	No	Yes	Yes	97 (1.8%)
Yes	No	No	Yes	40 (0.7%)
			Total	5,344 (100%)

It is clear from the figures that the majority (95 per cent) of participants in the 7/8-year postal survey had participated in each of the previous 3 rounds of interview.

The main problem associated with inter-wave non-response is the extent to which it is systematically related to underlying characteristics of the non-respondents (as discussed, for example, in Watson and Wooden, 2009). To re-weight (or statistically adjust) the data to address systematic response bias, it is important to identify the main correlates of response in a given wave. Four key family characteristics were identified as being most strongly related to completion and return of the postal survey on the 7/8-year-olds:

- family social class
- family equivalised income
- family structure
- highest level of Primary Caregiver education

Table 2.2 summarises the likelihood (through odds ratios) of participating in the study at 7/8 years of age for the four characteristics in question.

Table 2.2: Odds ratios of participating in the postal survey of the Infant Cohort at 7/8 years of age.

Family characteristics	Odds ratio of participating at 7/8 years of age
Social Class	
Ref. category: Never worked	
Professional/Managerial	1.659**
Non manual/Skilled Manual	1.232*
Semi/Unskilled Manual	1.196
Equivalised family income	
Ref. category: Lowest income quintile	
Quintile 2	1.257**
Quintile 3	1.514**
Quintile 4	1.744**
Quintile 5	1.842**
Income missing	0.412**
Family structure	
Ref. category: One-parent family – 1 child	
One-parent family – 2+ children	1.148
Two-parent family – 1 child	0.82
Two-parent family – 2+ children	2.051**
Primary Caregiver's educational attainment	
Ref. category: Junior Certificate or less	
Leaving Certificate	1.753**
Certificate/Diploma	2.380**
Degree	2.344**

Social gradients in response (or attrition) patterns are clear from the table. For example, a Primary Caregiver in a Professional/Managerial family is 1.7 times more likely to have returned their postal questionnaire than one in the reference category (the most socially disadvantaged families). The social gradients in participation in terms of income and maternal education are equally clear from the table; for example, graduate-level mothers were 2.3 times more likely to participate than those who left school at Junior Certificate or less. The relationship between response and advantage is also reflected in family structure. Table 2.2 suggests that there is no significant difference between response patterns for the reference group (small, one-parent families) and larger one parent families or smaller two parent families. However, larger two-parent families (2 or more children) were 2.3 times more likely to participate in the postal survey than families in the reference category.

Given the strengths of these social gradients in response patterns, it was decided to use the four background characteristics in Table 2.2 to re-weight (statistically adjust) the data to ensure that they are representative of the relevant population.

2.6 Reweighting the data

All sample surveys should be re-weighted (or statistically adjusted) to ensure that design and non-response characteristics do not systematically introduce bias into the estimates derived from the sample survey and so allow researchers to draw inferences from the sample to the population. In

the case of **Growing Up in Ireland's** fourth data sweep, the population under consideration is made up of the children who were living in Ireland at 9 months of age and who continued to live here at 7/8 years, when the postal survey was carried out.

In **Growing Up in Ireland** the Study Child is the longitudinal focus. The project is concerned with the developmental trajectories of children who were living in Ireland at 9 months of age and who continued to live here 7 years later when the fourth round of data was collected. The study's longitudinal design is what is described as a 'fixed panel'. After initial recruitment of the families, no additions are made to the sample to include children who were not living in Ireland at 9 months of age but who came to live here (with their families) after that age. These children are new 'entrants' to the population of children living in Ireland when the first round of data collection was carried out in 2007/2008. In a longitudinal study with a fixed panel design, additions are not made to the sample to account for these new entrants. The only 'exits' from the population is through migration out of Ireland after 9 months of age (the child is no longer living here) or where the Study Child deceases.

The re-weighting of the data ensures that, for example, any potential bias which might result from the patterns of attrition between one wave of the study and the next are controlled for and addressed prior to analysis.

A standard iterative procedure was used to generate the weights used in all phases of **Growing Up in Ireland**. This was implemented using software (known as the GROSS[1] system) which was developed for the ESRI. The GROSS system is based on a minimum information loss algorithm which fits population marginals to sample totals, within a regression framework and adjusts the sample according to pre-specified characteristics to ensure that it produces estimates which match population totals.

The sample weights for Wave 4 of the Infant Cohort were constructed by first generating an inter-wave attrition weight to adjust the composition of the completed Wave 4 sample (7/8 years of age) to the Wave 3 sample (5-year-olds) by controlling for variations in Wave 4 response and attrition according to:

- family social class
- family equivalised income
- family structure
- Primary Caregiver's educational attainment

In developing the attrition weight, variations in response patterns according to other family and child characteristics were investigated. The four characteristics identified above, however, were found to be the most systematically predictive of response in Wave 4 (see Table 2.2).

When the Wave 4 sample was adjusted by the attrition weight in line with differential inter-wave response, a new Wave 4 weighting factor was generated by taking the product of the attrition

---

[1] See, for example, Gomulka, J., 1992. "Grossing-Up Revisited", in R. Hancock and H. Sutherland (Eds.), Microsimulation Models for Public Policy Analysis: New Frontiers, STICERD, Occasional Paper 17, LSE. Gomulka, J., 1994. "Grossing Up: A Note on Calculating Household Weights from Family Composition Totals." University of Cambridge, Department of Economics, Microsimulation Unit Research Note MU/RN/4, March 1994.

weight between Waves 3 and 4 and the Wave 3 weighting factor assigned to each child at that time. The reader is reminded that the Wave 3 weight, in turn, incorporated the differential response at Wave 2, which also incorporated response between Wave 1 and Wave 2 as well as design and response weights at Wave 1.

Normally, when reweighting data in a longitudinal study one will make an adjustment for the 'exits' from the system. In the current context this largely means adjusting for the number of children and families who emigrated from Ireland between the third and fourth waves of the study (as well as the small number of children who deceased over the same period). Because the survey at 7/8 years of age was carried out on a postal basis, the Study Team had no systematic information on the number of families who had emigrated between Wave 3 and 4. When a survey is conducted on a face-to-face basis, in situations in which they are unable to contact a family at the address provided to them interviewers attempt to establish whether or not the family has moved outside Ireland. In previous face-to-face rounds of fieldwork in *Growing Up in Ireland,* details on the number of families who emigrated outside the country were systematically recorded. This information was then used to make an adjustment for these families from the original population who had left Ireland (to account for the fixed panel design used in the study). This was used to adjust the population total from the previous round of the study. This revised population total was then used to calibrate the statistical adjustment factors to the population total, in other words to produce a population 'grossing factor' for that wave of the study.

Because the survey in Wave 4 was conducted on a postal basis, the Study Team did not have any systematic information to allow it to make any adjustment to the population to account for 'exits' between Waves 3 and 4. Accordingly, the adjusted sample at Wave 4 was calibrated to a population total of 69,300 children, the best estimate of the population as was available at 5 years of age.

The reader should note that the structure (or percentage breakdown) of the weighted' and 'grossed' samples will be the same. However, both will differ from the 'unweighted' breakdowns. The 'grossed' sample calibrates to the estimated population total of 69,300. The 'weighted' sample estimates will sum to the actual number of cases interviewed; 5,344.

2.7 A comparison of 9-month population estimates from different full and partial samples

As noted in Section 2.6 above, the completed sample from the postal survey in the fourth round of the study was 5,344. This compares with 11,134 families who participated in the first wave of the study, when the child was 9 months of age. There is clearly a substantial difference between the two sample sizes. The weight for the completed sample at 7/8 years of age incorporates differences in response between different groups of families at each round of the study. The reader may well ask how well these weights are working in adjusting for the loss of families from the completed samples between the first and fourth wave of the study. The large reduction in completed sample size at the fourth round of data collection is, of course, related to the fact that the survey in that round was based on a postal (rather than face-to-face) approach to the families.

To assess how well the weights which have been generated for Wave 4 of data collection are working, one can compare the weighted population estimates for a selection of 9-month characteristics from the full sample of 11,134 families in Wave 1 with the weighted population

estimates of the same set of 9-month characteristics based on the weighted sub-sample of 5,344. This allows the reader to ask what effect does the reduction in sample size between the 9 month face-to-face interview and the 7/8 year postal survey have on the weighted population estimates for a range of 9-month characteristics. Effectively, how well are the weights taking account of differential response across the second, third and (particularly) fourth wave of the study[2].

Table 2.3a presents a comparison of a set of Study Children's 9-month characteristics based on the full sample of 11,134 families who participated at that time (Section A of the table) and the 5,344 families who participated on a postal basis at 7/8 years of age (Section B).

Table 2.3a: A comparison of population estimates for a range of 9-month characteristics, based on (a) the full sample of 11,134 cases completed when the Study Child was 9 months of age and (b) the smaller sample of 5,344 which was completed when the child was 7/8 years.

9-month characteristic	A. Based on 9-month sample - 11,134				B. Based on 7/8-year sample - 5,344			
	Per cent	(95% CI)			Per cent	(95% CI)		
MMa5ap2 - Study Child's Gender								
Boy	51.3	50.4	52.3		51.8	50.4	53.1	
Girl	48.7	47.7	49.6		48.2	46.9	49.6	
MML10 - Primary Caregiver's Employment Status								
At work	55.9	54.9	56.8		57.4	56.1	58.8	
Student/training	1.5	1.3	1.8		1.2	0.9	1.4	
Unemployed	5.5	5.1	5.9		4.4	3.9	5.0	*
Home duties	36.0	35.1	36.9		36.1	34.8	37.4	
Other	1.0	0.8	1.1		0.9	0.6	1.1	
MML34 - Primary Caregiver's Education Status								
Junior Cert. or less	17.6	16.9	18.3		17.3	16.3	18.3	
Leaving Cert.	25.2	24.4	26.0		24.4	23.3	25.6	
Certificate/Diploma	28.0	27.1	28.8		28.5	27.3	29.7	
Degree	29.2	28.3	30.0		29.6	28.4	30.9	
bpc2J33 - Family's ability to Make Ends Meet								
With great difficulty	6.8	6.4	7.3		7.0	6.3	7.7	
With difficulty	11.2	10.6	11.8		11.9	11.0	12.8	
With some difficulty	34.8	34.0	35.7		38.5	37.2	39.8	*
Fairly easily	24.4	23.6	25.2		27.4	26.2	28.6	*
Easily	8.0	7.5	8.5		9.0	8.2	9.8	
Very easily	2.7	2.4	3.0		3.0	2.6	3.5	
hhtype4 - Household Type								
One-parent /1 child	7.3	6.8	7.7		6.3	5.6	6.9	
One-parent /2+ child	7.5	7.0	8.0		5.9	5.3	6.5	*
Two-parent /1 child	32.4	31.5	33.3		32.3	31.0	33.5	
Two-parent /2+ child	52.8	51.9	53.7		55.6	54.2	56.9	*
MME1 - Is baby currently being minded by someone else?								
Yes	39.0	38.1	39.9		39.9	38.6	41.2	
No	61.0	60.1	61.9		60.1	58.8	61.4	
MMD14 - How much is baby's sleeping pattern a problem for you?								
A large problem	3.2	2.9	3.5		3.4	2.9	3.9	
A moderate problem	7.9	7.4	8.4		7.9	7.2	8.6	
A small problem	18.8	18.1	19.5		18.6	17.6	19.7	
No problem	70.1	69.2	70.9		70.0	68.8	71.2	
MMH13 - Was baby ever breastfed?								
Yes	56.0	55.0	56.9		55.6	54.3	56.9	
No	44.0	43.1	44.9		44.3	43.0	45.7	
MMD16 - Have you used a soother with baby in the last week?								
Yes	64.8	64.0	65.7		65.1	63.9	66.4	
No	35.1	34.2	36.0		34.6	33.3	35.9	
MML32 - Proportion of total income from social welfare payments?								
Less than 5%	54.7	53.8	55.6		56.6	55.2	57.9	

[2] The reader should note, of course, that notwithstanding how comparable the population estimates are between the two samples (11,134 and 5,344 cases) the estimates based on the larger sample will be associated with smaller confidence intervals or standard errors.

5-20%	21.7	20.9	22.5		21.8	20.7	23.0	
20-50%	6.9	6.4	7.4		7.4	6.7	8.1	
50-75%	2.8	2.5	3.1		2.4	2.0	2.8	
75-99%	2.7	2.4	3.0		2.4	2.0	2.9	
100%	9.5	8.9	10.0		7.8	7.1	8.5	
MMK3 - Are you in regular contact with the baby's grandparents?								
Yes	91.2	90.7	91.7		93.3	92.7	94.0	*
No	8.8	8.3	9.3		6.6	6.0	7.3	*
MML41 - Primary Caregiver a citizen of Ireland?								
Yes	84.6	84.0	85.3		88.6	87.7	89.5	*
No	15.3	14.6	16.0		11.3	10.5	12.2	*
MML43 -Primary Caregiver born in Ireland?								
Yes	77.8	77.0	78.5		81.1	80.1	82.2	*
No	22.2	21.5	23.0		18.9	17.8	19.9	*
MML46 - Is baby a citizen of Ireland?								
Yes	96.4	96.1	96.8		98.1	97.7	98.5	*
No	3.4	3.0	3.7		1.7	1.4	2.1	*
MML48 - Was baby born in Ireland?								
Yes	99.0	98.8	99.2		99.2	98.9	99.4	
No	1.0	0.8	1.1		0.8	0.6	1.1	

Both Sections A and B present an estimate of the average, as well as the 95 per cent confidence interval for the upper and lower range for each of the characteristics in question. By considering these confidence intervals we can assess whether or not any differences in the average or mean is due to chance sampling fluctuations or whether or not we would be correct (95 times in a 100) in saying that it represents a real difference in the estimates from the two samples.

Overall, Table 2.3a suggests that there is no statistical difference (at the 95 per cent level) in the mean estimates from the two samples for the majority of variables in the table. There is a statistical difference in the estimated percentage who are unemployed at 9 months of age (5.5 per cent from the full 9-month sample compared to 4.4 per cent from the 7/8-year sample). Areas where there appears to be a more systematic trend in differences are: contact with grandparents; whether or not the PCG and Study Child were born in Ireland; and citizenship of Ireland. In general, the estimate from the sample at 7/8 years of age is higher for these variables; that is, a higher percentage of persons in the 7/8 year sample recorded that they had contact with the child's grandparents; were Irish citizens and were born in Ireland. This may, to a large degree, reflect the fixed panel design referred to above. This means that the 'exits' from the sample since the initial recruitment were concentrated among parents (and their Study Children) who were not themselves born in Ireland, but who may have migrated to Ireland in the early 2000's, when the economy was expanding rapidly and when Ireland was experiencing high levels of net in-migration. After the financial crisis of 2008 many of these families left the country. In this regard, the smaller sample at 7/8 years of age is yielding the sort of estimates and reflects outflows in the direction we would expect from a fixed panel design.

Table 2.3b: A comparison of population estimates for average birth weight and length, based on (a) the full sample of 11,134 cases completed when the Study Child was 9 months of age and (b) the smaller sample of 5,344 which was completed when the child was 7/8 years.

	MMH6gms - Baby's birth weight (grams)							
	Mean	Lower C.I.	Upper C.I.		Mean	Lower C.I.	Upper C.I.	
All 9-month olds	3,477.75	3,467.66	3,487.83		3,500.36	3,485.74	3,514.97	*
	MMH7cms - Baby's birth length (cm)							
All 9-month olds	50.72	50.56	50.89		50.72	50.49	50.95	

	Was Study Child Low Birth Weight (<2,500grams)						
Yes	6.2	*5.7*	*6.6*	5.6	*5.0*	*6.3*	
No	92.8	*92.3*	*93.3*	93.1	*92.5*	*93.8*	

Table 2.3b provides comparable estimates on birth weight, length and the percentage of children who were born at low birth weight (less than 2,500 grams). The table indicates that there is a significant, though small, difference of 23 grams in the estimated average weight of the children between the larger and smaller samples. The percentage of children classified as having a low birth weight is not significantly different between the two samples, nor is the estimated average length of the children at birth.

3. Instrument Development

3.1 Introduction

This section gives a brief outline of the consultative process for instrument development for Wave 4 of the Infant Cohort. Being a postal questionnaire, as opposed to a full household interview conducted by a trained interviewer, the instrumentation for this wave was necessarily much shorter and simpler than for previous waves. The topics were guided in large part by the development undertaken for earlier phases with an emphasis on issues particularly pertinent for children aged 7/8 years. An overview is also provided of the pilot phase of the Wave 4 data sweep.

3.2 Instrument Development

As at previous waves of the study, intensive consultation took place with various groups of experts in the development of the instruments and procedures. The policy sector input came from the funding departments of the Children and Youth Affairs; Social Protection and the Central Statistics Office as well as the Department of Education and Skills. Details of inputs prior to the current wave are detailed in the corresponding reports for individual phases.

In brief, the main guiding principles for selection of topics and/or specific questions into the postal questionnaire at age 7/8 years were:

- Policy relevance
- Longitudinal significance (and whether used in previous waves of ***Growing Up in Ireland***)
- Particular relevance to age 7/8 years
- Suitability for self-completion, on paper, by the child's Primary Caregiver
- The 'value-to-length' proportion given the need to maximise the response to a postal questionnaire

3.3 Piloting the Instrument

The pilot phase of the Wave 4 data sweep consisted of a postal questionnaire which collected information on five topics: the Study Child's family; the Study Child's health and development; the Study Child's education and after-school care; the Study Child's activities; and being a parent. The questionnaire and accompanying documents were posted (through a maximum of 3 mailshots) to the designated Primary Caregiver at the most recent round of the study (mostly when the child was 5 years old but possibly at 3 years or even 9 months).

The longitudinal pilot sample used was made up of two components, 163 families who participated in the study at 5 years of age and a further 20 who participated at earlier waves but not in the 5-year phase. In total, 86 questionnaires were returned from the pilot phase; 84 from families who participated at Wave 3 and a further 2 families who did not. This gave an overall response rate of 47%.

Although the sample obtained from the pilot was relatively small, the data recorded appeared to be very much in line with expectations. The pilot phases provided a lot of useful feedback in terms of procedures and protocols related to the postal-basis of this data sweep. No issues were encountered with regard to ethics, consent or child protection. From the pilot phase, the study team concluded that the postal questionnaire would add substantially to the study's information on the children at 7/8 years of age, as well as perform an important function in terms of cohort maintenance between 5 and 9 years of age (the intention being to interview the Study Child and their families on a face-to-face basis when they are 9 years of age).

4. Questionnaire, Fieldwork and Implementation

4.1 Questionnaire

A single postal questionnaire was sent to the home and completed by the Primary Caregiver. For more detailed information on the questionnaire used at Wave 4, see www.growingup.ie. In addition to the postal questionnaire, the Primary Caregiver was also asked to measure and record the height and weight of the Study Child. They were not provided with any equipment to conduct these measurements, instead using whatever equipment they had at their disposal in the home. The content of the questionnaire was broadly as outlined in Table 4.1.

Table 4.1 – Household-based instruments used at Wave 4

RESPONDENT	MODE OF COMPLETION	SUMMARY OF CONTENT
Primary Caregiver	Self-complete Paper Questionnaire	**Postal Questionnaire**
		A: You and your family
		B: Your child's health and development
		C: Your child's education and after-school care
		D: Your child's activities and pastimes
		E: Being a parent
Study Child	Measured by PCG	**Physical Measurements**
		- Height
		- Weight

4.2 Contacting a Household and Tracing

As with all previous waves of the study, initial contact with the family was made by way of a cover letter, sent in conjunction with the postal questionnaire. If the questionnaire was not returned by the family, two subsequent reminders and postal questionnaires were sent. If at this stage there was still no response, families were called by the Study Team.

In instances where the letter was returned by An Post, efforts were made to contact families by phone to establish their address so that the postal questionnaire could be resent, but also in preparation for the nine-year home visits.

4.3 Incidents

A detailed *Growing Up in Ireland* Child Welfare and Protection protocol was developed by the Study Team for use at previous waves of the data collection. Although there were no face-to-face interviews or home visits conducted at this wave of the study, the protocol remained operational in the event that any issues were reported by families during this fieldwork phase.

5. Structure and Content of the Data Files

5.1 Introduction

This section outlines the structure of the Researcher 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 and ordering conventions and the reweighting process. Details are provided of the derived variables, the scaled measures and the child's height and weight. Finally, the coding and editing process is outlined. The Study Team would advise that the data are used in conjunction with the questionnaire. Researchers should however note that there may be differences in value labels between the questionnaires and the data file, for the purposes of preparation and anonymisation. This is especially true for the AMF.

5.2 Anonymised (AMF) and Researcher (RMF) Microdata Files

Two data files are available for researchers: the Anonymised Microdata File (AMF) and the Researcher Microdata File (RMF). The AMF is an anonymised dataset available to researchers through application to the Irish Social Sciences Data Archive (ISSDA). The RMF is a more detailed dataset, access to which is subject to appointment as an Officer of Statistics by the Central Statistics Office.

Given the nature of this questionnaire in the wave, there are fewer differences between the AMF and the RMF than previously but some differences apply. To preserve anonymity for respondents, names, dates of birth and open text variables were removed from both the AMF and RMF. Some variables which appear on the RMF, such as the individual items for scaled measures, have been removed from the AMF. Other variables have had their values banded into larger groups so that frequencies with low cell counts are not visible. In some instances this was achieved by either bottom or top coding (or both) of outlying cases (e.g. bpc4q6, bpc4q9). In others, continuous scores have been grouped into categories (e.g. bpc4q7a, bpc4q7c, bpc4q8).

Information particularly likely to be sensitive in nature has been removed from the AMF. The user should therefore note that not every question from the questionnaires is included in the data file, particularly in the case of the AMF. For the AMF only, all individual scale items (PEDS, SDQ, SSIS, Pianta) were removed, with just subscale total scores included in the file. Whilst the individual items are present in the RMF for the PEDS (bpc4q15a-i), SSIS (bpc4q16a-z), SDQ (bpc4q32a-y) and Pianta (bpc4q37a-o) scales, individual question phrasings have been withheld due to copyright restrictions.

A list of variables included in each data file is available via the accompanying summary data dictionary.

ICD-10 codes for up to three longstanding illnesses (bpc4q11a-c) are provided in the RMF only, this information was deemed too disclosive for the AMF. For physical measurements, height, weight and the subsequent BMI calculation were all rounded to the nearest whole unit. High and low outlier (but still plausible) values were banded together to ensure that frequencies with low cell counts were not visible. These steps (rounding, banding) were not implemented for the RMF.

5.3 Structure of the data files
Both the Researcher Microdata File (RMF) and Anonymised Microdata File (AMF) are presented as a flat rectangular data file based on the single questionnaire administered to the Primary Caregiver. However, the case-base is the Study Child.

5.4 Variable naming
All variables for Wave 4 of the Infant Cohort are prefixed with a 'b' for 'birth cohort'; there are slight differences to the combination of preceding letters for the question taken directly from the questionnaire versus derived variables.

5.4.1 Naming of Questionnaire-based Variables
The prefix 'b' is followed by two letters which indicate the respondent: 'pc' for Primary Caregiver. This is followed by '4', to indicate the fourth wave of data collection, a 'q' to indicate that the variable comes directly from a question, and the question number. For example:

• Question '17a' from the Postal Questionnaire at Wave 4 will have the variable name 'bpc4q17a'

5.4.2 Naming of other variables
Exceptions to the aforementioned variable naming convention are derived variables and variables from the scaled measures, as well as direct measurements, i.e. physical measurements and cognitive tests. In these instances, the 'b' prefix is followed immediately by the '4', and then the derived variable. For example:

• The derived SDQ score for emotional difficulties has the variable name 'b4_SDQemotional'
• The derived child BMI variable is named 'b4_chdBMI'

5.5 Variable order
The order of the variables in the data file mirrors that of the postal questionnaire as closely as possible. The first variables include the household identification code, weighting factor, details of family's participation at previous waves and the Study Child's gender. Derived variables typically appear at the end of the file.

5.6 Identification Codes
Each household has a unique identification code, which is the same at all waves to enable matching of the data files where necessary. The sequence of identification codes runs from 300 to 1,113,400 and is indicated by the variable 'id'.

5.7 The Household Grid

At this wave, only summary information on the members of the Study Child's household were included on the self-complete questionnaire. A derived 'household type' variable at age 7/8 years has been included for researchers' use with the data file (b4_hhtype4). The Study Child's gender (p2sexw4) is also included in the data files.

5.8 The Respondent – Primary Caregiver

The Primary Caregiver was self-identified within the home as the person who provided most care to the Study Child and who knew most about him/her. In most cases, this was the child's mother. Note that the Primary Caregiver may change between waves.

5.9 Twins

Where there were twin or triplet Study Children, separate postal questionnaires for each child were sent to the Primary Caregiver. As with previous waves, only the data of children who were individually selected into the original main sample are archived with the main data file. This means that some non-singleton children will have the data of their twin on a separate row in the main data file and others will not. Further details on the inclusion/exclusion of twins are available in earlier guides (http://www.ucd.ie/issda/static/documentation/esri/GUI-SummaryGuideInfants.pdf).

5.10 Weighting variables

As discussed in Section 2.6 above, in line with best practice in sample surveys, the data have been re-weighted (or statistically adjusted) to ensure that the sample is representative of the population from which it has been drawn. By doing this one ensures that 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 one set of weighting and grossing factors: wgt_78yrs and gross_78yrs. The weighting factor (wgt_78yrs) incorporates the structural adjustment of the completed sample to the population, whilst maintaining the total completed sample size of 5,344 cases. The grossing factor (gross_78yrs) calibrates to the estimated population of 69,300. Both wgt_78yrs and gross_78yrs provide the user with the same structural breakdown of the data. The former can be used in significance testing and data modelling. More detail on the specifics of the weighting / grossing procedure is provided in Section Two above.

The variables xxwave1, xxwave2, xxwave3 and xxwave4 indicate if the case has data for Wave 1, Wave 2, Wave 3 and Wave 4, respectively. A value of one indicates participation at the relevant wave. In the 7 year data file xxwave1 and xxwave4 are equal to 1 for all cases, as all cases in this file have completed both Wave 1 and Wave 4. The variable of interest for selecting the appropriate sample of respondents at 3 years is xxwave2. In a small number of cases, xxwave2 will not be equal to 1, as there are 137 families who participated at Wave 1 and Wave 4 but not Wave 2. Similarly, xxwave3 will not be equal to 1, as there are 161 families who participated at Wave 1 and Wave4 but not Wave 3.

5.11 Derived Variables

In addition to some of the derived variables mentioned above (e.g. 'b4_SDQemotional'), another variable was derived to provide additional information on the circumstances of the family composition, 'b4_hhtype4'. This fourfold variable is based on whether or not the Study Child is living

in a one or two parent family as well as the number of children living in the household. This gives us a classification as follows:

- One parent, one child
- One parent, two or more children
- Two parents, one child
- Two parents, two or more children

A child is defined solely in terms of age (under 18 years) and not in terms of relationship to the Study Child or others in the household.

5.12 Scaled Measures Used in the Study

A number of scaled measures were used in the Growing Up in Ireland study and scored according to protocols provided by the authors. These are briefly described below. An indication of the reliabilities of these scaled measures, as illustrated by Cronbach's alpha, are detailed in the appendix to this report.

5.12.1 Parents' Evaluation of Developmental Status (PEDS; Glascoe, 2003)

This scale provides a measure of concerns that the Primary Caregiver may have regarding the study child. The 9-item scale can be found at question 15 on the postal questionnaire. These items make up two subscales: Developmental and Academic Concerns (b4_PEDSdevelopmental) and Mental Health Concerns (b4_PEDSmental). These subscales were modified for *Growing Up in Ireland* with consent from the original developer of the scale (Glascoe, 2003). The individual items are included in the RMF but users should note that summary labels rather than the original question phrasing is used due to copyright restrictions.

5.12.2 Social Skills Improvement System Rating Scales (SSIS_RS; Gresham & Elliot, 2008)

This scale provides a measure of the Study Child's social skills and abilities to interact positively with adults and peers. The version of the SSIS_RS used in *Growing Up in Ireland*, which appears on the postal questionnaire as question 16, comprises 26 questions. These make up four subscales: Assertion (b4_assertion), Responsibility (b4_responsibility), Empathy (b4_empathy) and Self-control (b4_selfcontrol). This measure was previously used when the child was aged 5 years.

5.12.3 Strengths & Difficulties Questionnaire (SDQ; Goodman, 1997)

The SDQ is a 25-item behavioural screening questionnaire designed to assess emotional health and problem behaviours in children. The SDQ appears on the postal questionnaire as question 32. The SDQ comprises five subscales, four of which can be combined to give a total difficulties score. The subscales and their corresponding variables names are listed in Table 5.1. This measure was previously used at ages 3 and 5 years.

Table 5.1 – Subscales of the Strength and Difficulties Questionnaire

SUBSCALE	VARIABLE NAME
Emotional	b4_SDQemotional
Conduct	b4_SDQconduct
Hyperactivity	b4_SDQhyper
Peer problems	b4_SDQpeerprobs

Prosocial	b4_SDQprosocial
Total Difficulties	b4_SDQtotaldiffs

5.12.4 The Pianta Child Parent Relationship Scale (CPRS; Pianta, 1992)
This 15-item scale assesses both the negative and positive aspects of the relationship between parent and child. The scale appears as question 37 on the postal questionnaire. The measure produces a Positive Aspects subscale (b4_PIANTApositive) and a Conflicts subscale (b4_PIANTAconflict). It has previously been used at ages 3 and 5 years

5.13 Physical Measurements

5.13.1 Height & Weight
Heights and weights of all study children were recorded and reported by their parents on the postal questionnaire. Weight could be recorded in kilograms or stones and pounds. Height could be recorded in centimetres or feet and inches.

Heights and weights recorded by the respondent were edited to remove clearly implausible values. All measurements have been converted to the metric system (centimetres and kilograms) on the data files.. For the AMF, height and weight values were rounded to the nearest whole unit. The Wave 4 measurements can be found in the following variables:

• Study Child Height (bpc4q18)

• Study Child Weight (bpc4q19)

5.13.2 Body Mass Index (BMI)
BMI scores were derived from the height and weight measurements taken by the respondent for the Study Child, producing the variable 'b4_chdBMI'. For the AMF, BMI scores were rounded to the nearest whole unit. Categorised variables are also provided, which group the Study Child as 'non-overweight', 'overweight' or 'obese' (b4_chdBMI_cat), based on World Obesity Federation (formerly IOTF) age and gender-specific guidelines.

5.14 Coding & Editing
Data are subject to checks and applicable edits where deemed necessary. In addition to quality assurance checks, the AMF in particular is subject to editing required to anonymise the data.

6. Ethical Considerations
In undertaking research with families and children, ethical considerations assumed primary importance. Procedures relating to child protection were informed by the Children First: National Guidance for the Protection and Welfare of Children (Department of Children and Youth Affairs, 2011) as well as the relevant Acts in Irish legislation. Acts of particular relevance for this Study are the Data Protection Acts 1988, 2003 and the Statistics Act, 1993. All staff working on *Growing Up in Ireland* were security vetted by An Garda Síochána (the Irish Police Service). All work in Wave 4 of the Infant Cohort was carried out under ethical approval granted by a dedicated and independent Research Ethics Committee (REC) convened by the Department of Children and Youth Affairs,

especially for *Growing Up in Ireland*. The Research Ethics Committee was very rigorous in its review and consideration of all the materials and procedures used in the project, including those in the postal survey at 7/8 years of age.

7. References

Bronfenbrenner, U., and Morris P., (2006), 'The bioecological model of human development', in R. M. V. Lerner, W. Damon, and R. M. S. Lerner (eds.), Handbook of Child Psychology, vol. 1: Theoretical Models of Human Development, 793-828 (Hoboken, NJ, Wiley).

Glascoe, F. (2003), Parents' evaluation of developmental status: how well do parents' concerns identify children with behavioural and emotional problems? *Clin Pediatr* 42 (2), 133-8.

Greene, S., Williams, J., Layte, R., Doyle, E., Harris, E., McCrory, C., Murray, A., O'Dowd, T., Quail, A., Swords, L., Thornton, M. & Whelan, C. (2010). Growing Up in Ireland: Background and Conceptual Framework. Dublin: Office of the Minister for Children and Youth Affairs.

Gresham, F. M. & Elliott, S. N. (2008). Social Skills Improvement System: Rating Scales Manual. Minneapolis, MN: NCS Pearson, Inc.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. Journal of Child Psychology and Psychiatry, 38, 581-586.

Pianta, R.C. (1992). Child-parent Relationship Scale. Unpublished measure, University of Virginia.

Watson, N, & Wooden, M. (2009). Identifying Factors Affecting Longitudinal Survey Response. In P. Lynn (Ed.), Methodology of Longitudinal Surveys. Hoboken, NJ: John Wiley & Sons, Ltd.

9. Appendix

Appendix 1 – Indicative Cronbach's alphas for the scales used in the study

SCALE	SUBSCALE	α
Parents' Evaluation of Developmental Status (PEDS)	Developmental	0.882
	Mental Health	0.854
	Total	0.917
Social Skills Improvement System Rating Scale (SSIS)	Assertion	0.784
	Responsibility	0.861
	Empathy	0.908
	Self-control	0.878
Strengths and Difficulties Questionnaire (SDQ)	Emotional	0.698
	Conduct	0.585
	Hyperactivity	0.789
	Peer problems	0.602
	Prosocial	0.703
	Total	0.719
Pianta Child-Parent Relationship Scale	Closeness	0.706
	Conflict	0.824