



Leading education
and social research
Institute of Education
University of London

Millennium Cohort Study

Technical report on response in sweep 5 (age 11)

Tarek Mostafa

April 2014



Centre for Longitudinal Studies
Following lives from birth and through the adult years
www.cls.ioe.ac.uk

CLS is an ESRC Resource Centre based at the Institute of Education, University of London



**Millennium Cohort Study:
Technical report on response in sweep
5 (age 11)**

Tarek Mostafa

April 2014

First published in April 2014 by the
Centre for Longitudinal Studies
Institute of Education, University of London
20 Bedford Way
London WC1H 0AL
www.cls.ioe.ac.uk

© Centre for Longitudinal Studies

ISBN 978-1-906929-78-7

The Centre for Longitudinal Studies (CLS) is an ESRC Resource Centre based at the Institution of Education. It provides support and facilities for those using the three internationally-renowned birth cohort studies: the National Child Development Study (1958), the 1970 British Cohort Study and the Millennium Cohort Study (2000). CLS conducts research using the birth cohort study data, with a special interest in family life and parenting, family economics, youth life course transitions and basic skills.

The views expressed in this work are those of the author and do not necessarily reflect the views of the Economic and Social Research Council. All errors and omissions remain those of the author.

This document is available in alternative formats.
Please contact the Centre for Longitudinal Studies.
tel: +44 (0)20 7612 6875

email: info@cls.ioe.ac.uk

Introduction

The Millennium Cohort Study (MCS) is a multi-disciplinary, multi-purpose research project following the lives of more than 19,000 children born in the UK in 2000-01. It is the most recent of Britain's world-renowned national longitudinal birth cohort studies. The study has been tracking the Millennium children through their early childhood years and plans to follow them into adulthood. As with any longitudinal survey, the MCS is subject to attrition. Attrition takes place when respondents drop out of the survey over time. This leads to two problems: a reduction in sample size and to bias in sample composition. Sample bias arises when the likelihood of dropping out from the survey is correlated with the socio-demographic characteristics of the respondents. In this case, the survey will lose a particular type of respondent (e.g. disadvantaged families, ethnic minorities, etc) and the sample will no longer be representative of the original population.

This report explores attrition in sweep 5 (age 11) of MCS and presents the procedures used in the construction of the sweep 5 unit non-response weights. For a full description of attrition in previous waves, refer to the MCS [Technical Report on Response](#) (3rd edition, 2010). For a description of how to use the weights in Stata and SPSS refer to the respective guides ([Stata](#), [SPSS](#)). For a description of the MCS sample refer to the [Technical Report on Sampling](#) (4th edition, 2007).

Response at sweep five (MCS5)

In table 1, response and non-response rates are presented by category. The table shows that the proportion of productive cases dropped over time from 96.4 per cent in MCS1 to 69 per cent in MCS5. The proportions in all other categories rose as the proportion of non-respondents grew.

Ineligible includes child deaths, sensitive cases and temporary and permanent emigrants. *Untraced* refers to untraced movers and may include emigrants.

Table 1: Response rates in all MCS sweeps

Categories	MCS1		MCS2		MCS3		MCS4		MCS5	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Not issued	692	3.6	0	0.0	0	0.0	2,213	11.5	2,851	14.8
Productive	18,552	96.4	15,590	81.0	15,246	79.2	13,857	72.0	13,287	69.0
Ineligible	0	0.0	167	0.9	300	1.6	126	0.7	78	0.4
Untraced	0	0.0	687	3.6	547	2.8	706	3.7	388	2.0
Refusal	0	0.0	1,739	9.0	2,315	12.0	1,811	9.4	2,196	11.4
Non-contact	0	0.0	930	4.8	546	2.8	123	0.6	438	2.3
Other unproductive	0	0.0	131	0.7	290	1.5	408	2.1	6	0.0
Total	19,244	100.0	19,244	100.0	19,244	100.0	19,244	100.0	19,244	100.0

Table 2 shows that response rates were very similar across all four countries, with the highest response rate being in England.

Table 2: Response rates by country in MCS5

Categories	England		Wales		Scotland		NI	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Not issued	1,761	14.4	365	13.2	414	17.7	311	16.2
Productive	8,618	70.5	1,881	68.2	1,480	63.4	1,308	68.0
Ineligible	53	0.4	9	0.3	10	0.4	6	0.3
Untraced	200	1.6	75	2.7	84	3.6	29	1.5
Refusal	1,309	10.7	335	12.1	306	13.1	246	12.8
Non-contact	279	2.3	95	3.4	42	1.8	22	1.1
Other unproductive	5	0.0	0	0.0	0	0.0	1	0.1
Total	12,395	100.0	2,798	100.0	2,370	100.0	1,955	100.0

Table 3 shows that the response rates vary across ward types within country. Advantaged households systematically have higher response rates than disadvantaged ones while the ethnic stratum in England has a relatively high response rate.

Table 3: Response rates by stratum in MCS5

Categories	England			Wales		Scotland		NI	
	Adv.	Dis.	Ethn.	Adv.	Dis.	Adv.	Dis.	Adv.	Dis.
Unproductive	25.5	31.0	34.2	28.3	33.4	32.4	40.7	30.8	32.7
Productive	74.5	69.0	65.8	71.7	66.6	67.6	59.3	68.2	67.3

(Adv: Advantaged ward. Dis: Disadvantaged ward. Ethn: Ethnic minority ward)

Table 4 shows that 54.3 per cent of all respondents participated in all waves of MCS. In contrast, 19.6 per cent have interrupted response patterns. In other words, they participated in a number of waves then dropped out before participating again in subsequent waves. 26.1 per cent of all respondents have monotone response patterns. That is, they participated in a number of waves before dropping out for all subsequent sweeps.

Table 4: Monotone vs. non-monotone response in MCS5

Type of non-response	Freq.	%
Monotone	5,023	26.1
Non-monotone	3,773	19.6
All waves	10,448	54.3
Total	19,244	100.0

Predicting response at wave 5 for weight adjustment

The same procedure used for predicting non-response at wave 4 was again used at wave 5. Missing data for predictor variables due to non-monotone non-response or item missingness were imputed using simple and multiple imputations. Wave 5 non-response predictors were mostly the same as at wave 4. Multiple imputations were carried out using the MI command in Stata 12.

As a result of the use of simple and multiple imputations, the sample used in the logit response model consisted of 16,393 observations (i.e. the issued sample in MCS5). Weights were constructed for all respondents in MCS5. The dependent variable in the logit model is binary (1 for response and 0 otherwise) and the predictors are: the cohort member's gender; mother's age at first live birth; ethnicity; housing tenure; accommodation type; national vocational qualification; breastfeeding; main respondent's work status; whether the household is a new family which joined the survey in wave 2; and income item non-response. These variables came from all four previous waves.

Imputations were carried out in the following way:

Simple imputations: ethnicity, accommodation type and National Vocational Qualification (NVQ) were imputed using the most recent available data from previous waves with simple replacement imputations. The questions on accommodation type and NVQ were only asked if accommodation or NVQ have changed since the last wave of data collection.

Multiple imputations: main respondent's work status and housing tenure were missing for 2,744 observations. Breastfeeding was missing for the new families (617 observations). These three variables were imputed using 10 multiple imputations. Different imputation procedures were used depending on the nature of the variable: a logit procedure for work status and breastfeeding and a multinomial logit for housing tenure. The explanatory variables for the imputation of work status and housing tenure in wave 4 were the exact same variables from the previous three sweeps. For the imputation of breastfeeding I used different variables related to social class as explanatory variables: ethnicity; NVQ; number of parents in household; and type of accommodation.

It should be noted that some variables such as cohort member's gender and whether the household is a new family did not have any missing values and therefore did not require any imputation. Income item non response was constructed as a binary variable which takes the value of 1 if the respondent did not answer the income question. Mother's age at first live birth was missing for only 49 observations; these were replaced by the average age of the non-missing cases.

Table 5 shows the odds ratios of the response logit model estimated using the 10 imputed datasets. The linear predicted values were generated from this model then an inverse-logit transformation was carried out to transform the predicted values into predicted probabilities. The non-response weights at sweep 5 were constructed as the inverse of the predicted probabilities. Two overall weights were constructed by multiplying the aforementioned non-response weights with the same weights from wave 4. These overall weights adjust for both sampling and attrition. The weights are:

- EOVWT1: wave 5 overall weight for single country analysis
- EOVWT2: wave 5 overall weight for whole of UK analysis.

Table 5:logit response model

Explanatory variables	Odds Ratio	Std. Err.	t-statistic	P>t
Boy	0.89	0.039	-2.60	0.009
Mother's age at first live birth, reference: [20-30]				
Before 20	0.82	0.044	-3.71	0.000
[30-40]	1.51	0.100	6.22	0.000
After 40	0.92	0.290	-0.27	0.784
Ethnicity, reference: White				
Mixed	1.04	0.130	0.34	0.737
Indian	1.14	0.163	0.91	0.365
Pakistani, Bangladeshi	2.05	0.195	7.50	0.000
Black	0.78	0.085	-2.28	0.022
Other	1.02	0.142	0.14	0.892
Housing tenure, reference: mortgage				
Own	0.89	0.108	-0.94	0.350
Rent LA or HA	0.76	0.047	-4.48	0.000
Rent privately	0.73	0.063	-3.63	0.000
Other	0.62	0.071	-4.18	0.000
Type of accommodation, reference: house, bungalow				
Anything else (flat, studio, other)	1.33	0.086	4.41	0.000
National Vocational Qualification, reference: NVQ 1				
NVQ 2	0.95	0.079	-0.65	0.514
NVQ 3	1.01	0.095	0.14	0.886
NVQ 4	1.21	0.110	2.05	0.040
NVQ 5	1.57	0.223	3.17	0.002
NVQ 6	0.85	0.073	-1.88	0.060
Breastfeeding attempted	1.36	0.068	6.17	0.000
Respondent in work	1.09	0.058	1.61	0.109
New family	0.93	0.101	-0.67	0.505
Income item non-response	0.21	0.009	-35.51	0.000
Constant	5.21	0.611	14.09	0.000
N		16,393		

Number of imputations: 10; Minimum DoF: 82; LA and HA are local authority and housing association.

In tables 6 and 7, the means, minimums and maximums of the two weights are presented by ward type and for the UK as a whole.

Table 6: EOVWT1, Wave5 overall weight for single country analysis

Ward type	Obs	Mean	Std. Dev.	Min	Max
England - Advantaged	3,598	1.60	0.60	1.05	10.67
England - Disadvantaged	3,316	1.14	0.53	0.58	8.06
England - Ethnic	1,704	0.51	0.27	0.21	2.86
Wales - Advantaged	597	1.96	0.73	1.22	6.37
Wales - Disadvantaged	1,284	0.90	0.40	0.46	5.23
Scotland - Advantaged	774	1.34	0.72	0.48	7.05
Scotland - Disadvantaged	706	1.09	0.67	0.30	5.65
Northern Ireland - Advantaged	500	1.56	0.86	0.49	7.94
Northern Ireland - Disadvantaged	808	1.14	0.69	0.28	5.24
Total	13,287	1.22	0.69	0.21	10.67

Table 7: EOVWT2, S5 overall weight for whole of the UK analysis

Ward type	Obs	Mean	Std. Dev.	Min	Max
England - Advantaged	3,598	2.05	0.81	1.33	13.55
England - Disadvantaged	3,316	1.50	0.71	0.75	10.37
England - Ethnic	1,704	0.67	0.35	0.27	3.69
Wales - Advantaged	597	0.66	0.24	0.41	2.07
Wales - Disadvantaged	1284	0.31	0.13	0.16	1.79
Scotland - Advantaged	774	1.04	0.54	0.38	5.55
Scotland - Disadvantaged	706	0.84	0.50	0.24	4.47
Northern Ireland - Advantaged	500	0.61	0.32	0.20	3.18
Northern Ireland - Disadvantaged	808	0.43	0.24	0.11	1.87
Total	13,287	1.23	0.87	0.11	13.55

For a description of how to use the weights in Stata and SPSS refer to the respective guide: [Stata](#), [SPSS](#).

Links to supporting documents

[MCS Technical Report on Sampling \(4th edition, 2007\)](#)

[User Guide to Analysing MCS Data Using Stata \(1st edition, 2011\)](#)

[User Guide to Analysing MCS Data Using SPSS \(1st edition, 2010\)](#)

[MCS Technical Report on Response \(3rd edition, 2010\)](#)

Centre for Longitudinal Studies

Institute of Education

20 Bedford Way

London WC1H 0AL

Tel: 020 7612 6860

Fax: 020 7612 6880

Email cls@ioe.ac.uk

Web www.cls.ioe.ac.uk