Leading education and social research
Institute of Education
University of London

Handling attrition and non-response in the 1970 British Cohort Study

Tarek Mostafa and Richard D. Wiggins

CLS Working Paper 2014/2

June 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

ESRC ECONOMIC & SOCIAL RESEARCH COUNCIL

Handling attrition and non-response in the 1970 British Cohort Study

Tarek Mostafa and Richard D. Wiggins

June 2014

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

ISBN 978-1-906929-79-4

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.

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

Contents

1. Abstract 4
2. Introduction 5
3. Non-response in the British Cohort Study (BCS70) 7
4. How effective are weights and imputations? 14
5. Conclusion 19
6. References 20

Abstract

The 1970 British Cohort Study (BCS70) is a continuing multi-purpose, multidisciplinary longitudinal study based on a sample of over 17,000 babies born in England, Wales and Scotland in 1970. The study has collected detailed information from the cohort members on various aspects of their lives, including their family circumstances at birth, education, employment, housing and partnership histories. There have been nine sweeps of data collection so far: at birth and at ages 5, 10, 16, 26, 30, 34, 38 and most recently age 42 (2012). This paper studies the extent of attrition in BCS70 and how it affects sample composition over time. We examine the determinants of response then construct inverse probability weights. In the last section, we use a simulation study to illustrate the effectiveness of weights and imputations in dealing with unit non-response and item missingness respectively. Our findings show that when the predictive power of the response models is weak, the efficacy of non-response weights is undermined. Further, multiple imputations are effective in reducing the bias resulting from item missingness when the magnitude of the bias is high and the imputation models are well specified.

Keywords: BCS70, attrition, unit non-response, item non-response, weights, imputation.

Introduction

Statistical description and analysis are persistently challenged by the problem of missing data (Little and Rubin 2002). Survey samples are threatened both by unit non-response and individual item missingness. In longitudinal surveys, the problem of maintaining co-operation with cohort members (CMs) over time adds another dimension to the problem of non-response. Attrition refers to situations where CMs drop out of a study and never return and situations where individual CMs have interrupted response pattern over time. These patterns are distinguished as monotone and non-monotone response, respectively.

Missing data constitutes a problem for two reasons. First, missingness leads to the loss of observations and to the reduction of sample size. For instance, in BCS70 if only CMs who have responded in all nine sweeps (since 1970) are considered, the resulting sample would represent only 20 per cent of the original sample of 17,284 CMs. Secondly, missingness leads to bias in any analysis if it is not 'completely at random' (MCAR). MCAR implies that the probability of not answering a particular question is uncorrelated with the characteristics of the respondent, and in any longitudinal survey it means that the probability of dropping-out from a sweep is uncorrelated with the characteristics of the CM. MCAR is a very strong assumption to make since missingness is more likely to be at random (MAR) or not at random (MNAR). Under MAR the probability of non-response to a question or the probability of dropping-out from a particular sweep are related to some of the observable characteristics of the respondent such as gender, social class, education, etc. Under MNAR the probability of non-response to a question or the probability of dropping-out from a particular sweep are related to unobservable factors. If missingness is related to any observable or unobservable variables then ignoring it would lead to the loss of a particular type of respondents (e.g. men, the less well educated, etc) and hence the sample will no longer be random or representative of the parent population.

Two approaches are typically adopted to tackle the problem of missing data. First, weights in longitudinal surveys are constructed to adjust or re-balance the distributions of the responders so that the relative importance of each cohort member's characteristic in any particular sweep is reweighted according to the importance of the characteristics of those who dropped out. In other words, if the survey is losing men over time, then men will be given higher weights than women. Weights are usually defined as the inverse of the probability of response (See Hawkes and Plewis 2006). The probability of response at each sweep is computed using binary logit models (response vs. non-response) or multinomial-logit models (allowing for more categories of response). These models draw upon CM characteristics in addition to external metadata as explanatory variables; Plewis (2011), Schouten and De Nooig (2005) and Micklewright et al (2012) provide further details on the use of auxiliary variables.

Even though weights are easy to use and can be made available as part of a dataset, they have a number of disadvantages:

1. Longitudinal weights are used to treat unit non-response resulting from CMs dropping-out over time. However, the application of unit non-response weights does not provide a solution to item missingness.
2. If variables x, y and z are used in predicting unit non-response and thus in the construction of weights, the results of analyses using x, y and z as dependent and independent variables will yield unbiased results. However, if other variables which are not included in the process of constructing weights are used then the results might still be biased, because the importance of these characteristics has been ignored.
3. Under conditional regression applications, if we are regressing an outcome variable from sweep t+1 on a number of independent variables collected during an earlier sweep t where attrition has possibly occurred, the weighted analysis will be constrained to using only the non-missing cases in both sweeps (Goldstein 2009, p. 64). This further undermines the efficacy of non-response weights because they will only adjust for non-response in one sweep (usually the sweep in which the dependent variable was observed).

The second approach to deal with missing data is random multiple imputation, (Little and Rubin (2002), Schafer and Olsen (1998) and Rubin (1987, 2004)). Several imputation techniques have been used in the past such as mean substitution and regression imputations (i.e. building regression models on the basis of selected predictors to predict particular items). Such techniques are known to bias the variance of imputed variables towards zero and hence are not reliable. More recently, following Rubin (1987), data augmentation and multiple imputation techniques have been developed to overcome the shortcomings of ad hoc methods. Such techniques use advanced applications such as the Markov-Chain-Monte-Carlo procedure, Gilks et al (1996). Despite the fact that multiple imputations are more complex to use than weights, they present two main advantages:

1. Multiple imputations allow the treatment of both item and unit non-response. In fact, unit missingness is a special case of item missingness where all variables are missing for the same respondent within a longitudinal record.
2. Multiple imputations can be custom-made according to the needs of the researcher. When properly specified, they are robust and generate valid inference. However, one should keep in mind that multiple imputations depend on the assumption that data is missing at random (MAR) in contrast with data being missing not at random (MNAR) (see Little and Rubin 2002). Further, multiple imputations can be designed according to the structure of the data (e.g. multilevel structure) and the nature of the variables (e.g. continuous, ordinal or multinomial variables). (Nathan (1983), Nathan and Holt (1980), and Pfeffermann (2001)).

In this paper we use the 1970 British Cohort Study (BCS70) which is a multi-purpose, multidisciplinary longitudinal study based on a sample of over 17000 babies born in England, Wales and Scotland in 1970. The study has collected detailed information from CMs on various aspects of their family circumstances at birth, and on their education, employment, housing and partnership histories over eight further sweeps of data collection at ages 5, 10, 16, 26, 30, 34, 38 and 42. The first objective of the paper is to assess the extent of unit non-response in BCS70 over the 9 available data sweeps and then to construct logit models in order to examine the determinants of response. The explanatory variables used in this analysis are the CM's birth characteristics from sweep one. These are used because they are non-missing for almost all individuals. The second objective is to assess the impact of non-response weights and imputation techniques on the bias resulting from unit non-response and item missingness respectively. We use a substantive model with variables from sweeps three and four and a simulation strategy to illustrate the efficacy of weights and imputations. The paper is organised as follows: the first section explores non-response in BCS70 and examines its determinants; the second section presents the simulation study; and the last concludes.

Non-response in the British Cohort Study (BCS70)

In what follows we use the nine available sweeps of BCS70 to examine attrition and model response. We use a number of birth characteristics as explanatory variables when modelling response, including: gender, father's social class, father's and mother's age at completion of education, mother's age at delivery, whether mother lived in London in 1970, whether mother attempted breastfeeding, marital status and number of older siblings.

In table 1, we summarise the pattern of missing data for BCS70 over the nine sweeps (1970 to 2012). 19.8per cent (Non-missing) of the CMs participated in all nine sweeps, whereas 52 per cent (non-monotone) dropped out from at least one sweep but returned to the study in a subsequent sweep, and 27.2per cent (monotone) dropped out from the survey after participating in a number of sweeps without ever returning (to date). The base sample of 17,284 CMs consists of the original birth sample (i.e. excluding immigrants who have joined the study later on).

Table 1: Patterns of missing data in BCS70 (1970 to 2008)

Pattern	Frequency	Percentage
Monotone	4,716	27.2
Non monotone	9,153	53
Non missing	3,423	19.8
Total	17,284	100

In table 2, we present the different categories of non-response. The first category presents the number of CMs who have participated in the designated sweep. The categories labelled 'contact later' and 'temporary emigrants' consist of CMs who were absent at a particular sweep but participated in a subsequent sweep. Those labelled 'dead', 'no-contact-later', and 'permanent emigrants' are CMs who have dropped out without ever returning to the survey. Refusals consist of CMs who have refused to participate in any further data collection after participating in the first four sweeps (these are basically individuals who have refused to participate after the responsibility of responding was transferred from parents to CMs after the age of 16). Finally, 'unproductive' is a miscellaneous category of non-response.

Table 2: Detailed response and non-response categories for BCS70 from 1970 to 2008[1]

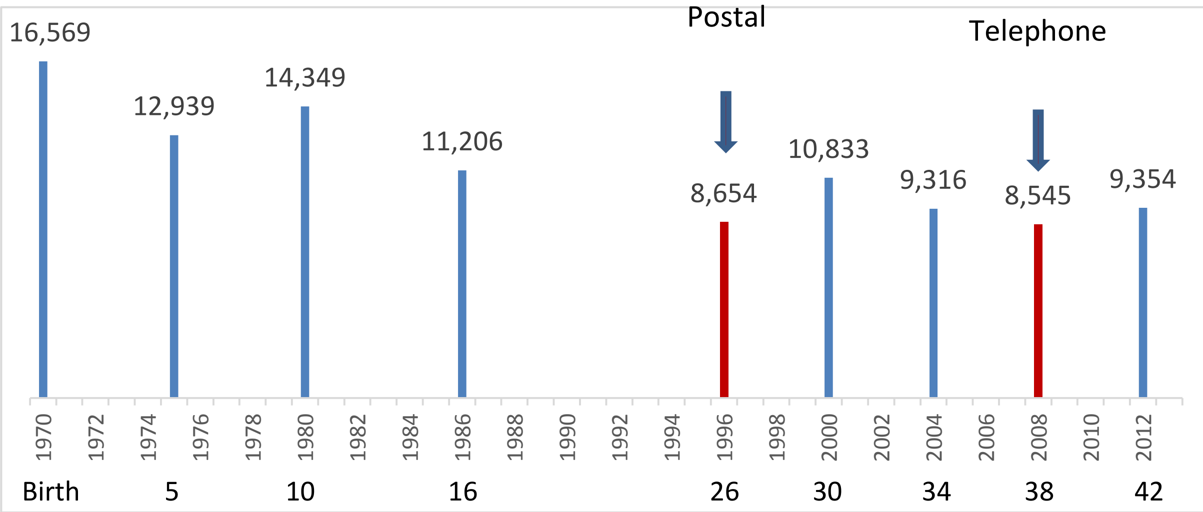
Response categories	Sweep 1	Sweep 2	Sweep 3	Sweep 4	Sweep 5	Sweep 6	Sweep 7	Sweep 8	Sweep 9
Age	Birth	5	10	16	26	30	34	38	42
Full or partial response	16,569	12,939	14,349	11,206	8,654	10,833	9,316	8,545	9,354
Dead	0	565	585	597	697	748	795	824	853
Unproductive	715	3,780	2,350	5,481	7,933	5,703	7,173	7,915	7,077
Total	17,284	17,284	17,284	17,284	17,284	17,284	17,284	17,284	17,284

Table 2 shows that over 42 years, from birth in 1970 until the ninth sweep in 2012, 7,930 CMs have dropped out for various reasons. Some have died, others have left Great Britain, while some have refused to participate or disappeared from the study to reappear again (i.e. non-monotone dropout represents about 53 per cent out of the base sample of 17,284). The category dead presents the number of cumulative deaths over the nine sweeps while the category unproductive respresents all other possibilities for dropout: permanent and temporary immigrants, refusals, non-contact, etc. One should note that dropout is not always permanent since some respondents did come back in later sweeps.

Figure 1 shows that there was a substantial drop in sample size between age 16 and age 26. The reasons for this drop are: the long period of ten years separating the two surveys, a teacher strike at age 16, the use of a self-completion postal survey at age 26, and the fact that CMs became adults by age 26 and they were required to provide consent to participate in the survey. Previously consent was sought from their parents. Furthermore, the drop in sample size at age 38 can also be partially attributed to the use of a telephone survey. Note that the sample size increased by 4.7per cent in sweep nine since some CMs were successfully traced through their NHS addresses. For further information on the data collection please see the BCS70 Technical Report.

[1] Note that the mode of data collection changed between sweep four and sweep five. In sweep five, the data was collected through postal services.

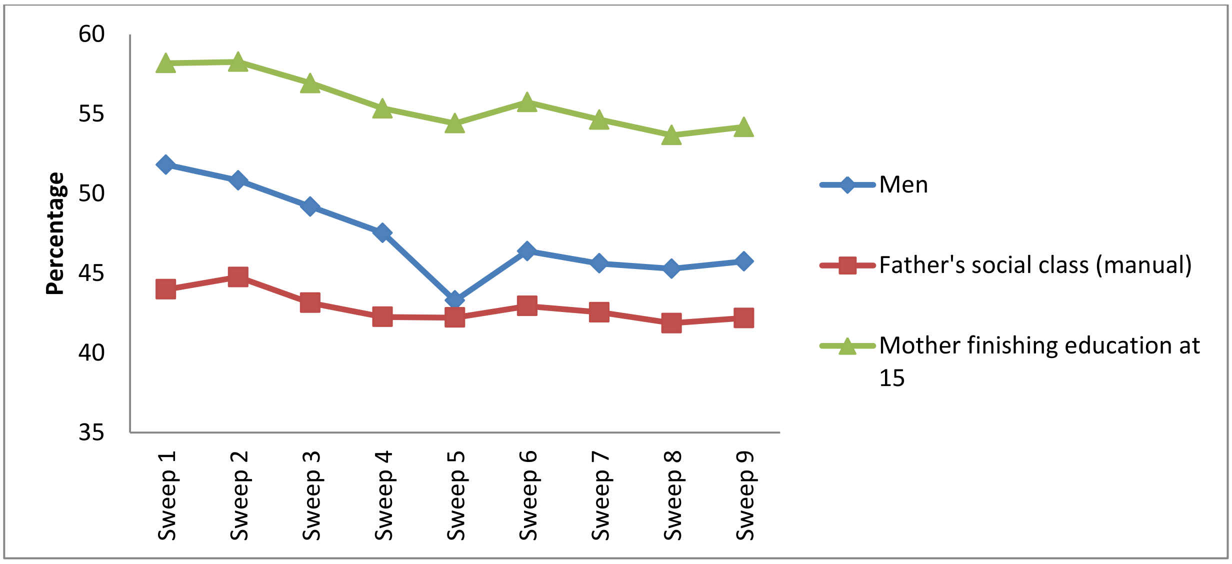
Figure 1: Sample size in the different waves of BCS70

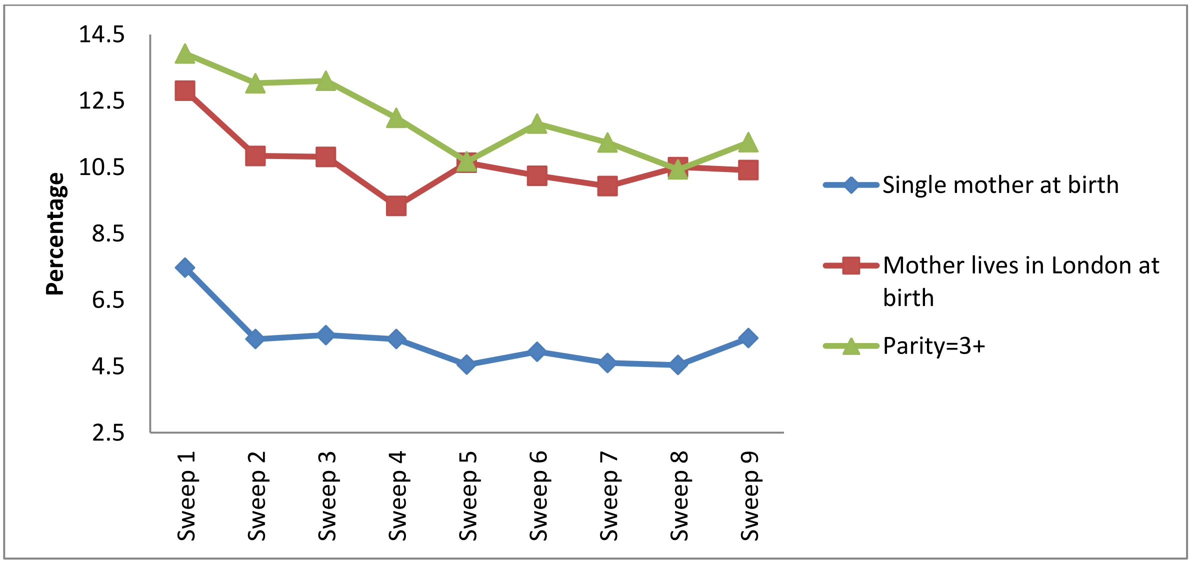


Figures 2 and 3 show the evolution of the sample composition according to CM characteristics recorded at birth. We find that the proportion of male CMs, CMs with fathers with manual occupations, and CMs with mothers with low levels of education is dropping over time. This means that these individuals are more likely to dropout than others. Similarly, the proportion of CMs whose parents were single in 1970, CMs whose mothers were living in London in 1970, and CMs who have at least three older siblings has also dropped. It is also worth noting that due to the rise in sample size in sweep nine and to the change in survey mode, the proportions have slightly converged towards their original values at birth. This indicates that the non-response bias in sweep nine is lower than that in sweep eight. Moreover, the male-female differential attained its maximum by age 26 and has declined ever since (sweep five).

In general, we can say that men from lower social backgrounds whose parents were single in 1970 are more likely to drop out from the survey. The drop out within these groups could have also been exacerbated by the lack of cohort maintenance. Based on these findings, it is obvious that dropout is not a random phenomenon. In other words, the probability of dropping out depends on a certain number of CM characteristics, and hence attrition over time will be a source of bias in any analysis because we are losing respondents of a particular type. The relative impact of these birth characteristics on response is explored in a logistic regression analysis.

Figures 2 and 3: Evolution of the BCS70 sample composition over time (1970 to 2012)





In what follows, we present the results of a logistic regression of response for each sweep using birth characteristics as explanatory variables. Response is a binary variable taking the value of 1 for those who participated (first category in table 2) and 0 for all other categories including those who died or migrated. The results are presented below in table 3. Note that sample size is relatively smaller (i.e. 15,270 instead of 17,284) than in table 2 because some CMs had missing birth characteristics.

Table 3: Odds ratios based on logistic regressions of binary response outcome for successive BCS70 sweeps

	Sweep 2	Sweep 3	Sweep 4	Sweep 5	Sweep 6	Sweep 7	Sweep 8	Sweep 9
Age	5	10	16	26	30	34	38	42
Gender (reference: men)								
Women	1.00	1.08	1.26***	1.80***	1.49***	1.48***	1.48***	1.44***
	(0.040)	(0.049)	(0.044)	(0.060)	(0.052)	(0.049)	(0.049)	(0.048)
Marital status (reference: single)								
Married	1.47***	2.18***	1.67***	1.85***	1.89***	1.89***	1.79***	1.42***
	(0.140)	(0.218)	(0.151)	(0.174)	(0.171)	(0.174)	(0.169)	(0.128)
Mother lives in London in 1970 (reference: not in London)								
In London	0.57***	0.55***	0.47***	0.71***	0.61***	0.62***	0.70***	0.67***
	(0.032)	(0.034)	(0.024)	(0.037)	(0.031)	(0.032)	(0.036)	(0.034)
Parity (reference: 0)								
1	0.97	1.02	0.87**	0.92	0.94	0.89**	0.93	0.92*
	(0.050)	(0.059)	(0.039)	(0.039)	(0.042)	(0.038)	(0.039)	(0.039)
2	0.82**	0.89	0.81***	0.79***	0.84**	0.74***	0.75***	0.81***
	(0.053)	(0.065)	(0.046)	(0.042)	(0.047)	(0.040)	(0.040)	(0.044)
3+	0.72***	0.90	0.70***	0.58***	0.65***	0.58***	0.54***	0.61***
	(0.053)	(0.076)	(0.045)	(0.036)	(0.041)	(0.036)	(0.033)	(0.038)
Breastfeeding (reference: attempted)								
Not attempted	0.82***	0.84***	0.85***	0.85***	0.92*	0.87***	0.87***	0.80***
	(0.036)	(0.041)	(0.032)	(0.031)	(0.034)	(0.031)	(0.031)	(0.029)
Mother's age at Delivery (reference: less than 20)								
[20-24]	1.42***	1.17	1.20**	1.31***	1.23**	1.33***	1.28***	1.26***
	(0.105)	(0.098)	(0.080)	(0.085)	(0.080)	(0.085)	(0.083)	(0.081)
[25-29]	1.51***	1.27**	1.28***	1.46***	1.35***	1.50***	1.45***	1.35***
	(0.121)	(0.115)	(0.092)	(0.102)	(0.096)	(0.103)	(0.101)	(0.093)
[30-34]	1.63***	1.36**	1.30**	1.62***	1.44***	1.66***	1.59***	1.39***
	(0.151)	(0.143)	(0.106)	(0.129)	(0.117)	(0.131)	(0.125)	(0.109)
35 or more	1.81***	1.56***	1.40***	1.69***	1.51***	1.81***	1.73***	1.45***
	(0.204)	(0.198)	(0.140)	(0.164)	(0.149)	(0.175)	(0.167)	(0.139)
Mother's age at completion of education (reference: 14 or less)								
15	1.56***	1.81***	1.29**	1.38***	1.20*	1.32***	1.15	1.04
	(0.141)	(0.179)	(0.106)	(0.114)	(0.098)	(0.107)	(0.094)	(0.084)
16	1.63***	1.73***	1.50***	1.50***	1.37***	1.51***	1.34***	1.21*
	(0.164)	(0.190)	(0.137)	(0.135)	(0.124)	(0.134)	(0.119)	(0.107)
17	1.47***	1.42**	1.32**	1.56***	1.26*	1.45***	1.32**	1.18
	(0.172)	(0.180)	(0.138)	(0.160)	(0.131)	(0.148)	(0.134)	(0.120)
18 or more	1.31*	1.34*	1.30*	1.48***	1.14	1.33**	1.24*	1.05
	(0.147)	(0.164)	(0.133)	(0.149)	(0.116)	(0.134)	(0.124)	(0.105)
Father's social class (reference: SC 1)								
Professional	0.94	0.98	0.85	0.94	0.93	0.99	0.95	0.97
	(0.102)	(0.116)	(0.084)	(0.087)	(0.090)	(0.092)	(0.088)	(0.090)
Clerical, non-manual	1.06	1.20	1.04	1.07	1.10	1.13	0.99	1.00
	(0.122)	(0.151)	(0.107)	(0.102)	(0.111)	(0.108)	(0.094)	(0.095)

Skilled manual	0.90	0.94	0.79*	0.79*	0.83*	0.84	0.74***	0.77**
	(0.097)	(0.109)	(0.076)	(0.071)	(0.078)	(0.076)	(0.067)	(0.070)
Unskilled manual	0.87	0.85	0.75**	0.70***	0.76**	0.75**	0.68***	0.69***
	(0.101)	(0.108)	(0.079)	(0.068)	(0.077)	(0.073)	(0.066)	(0.068)
Lowest grade workers	0.70**	0.77	0.69**	0.56***	0.64***	0.65***	0.56***	0.59***
	(0.091)	(0.111)	(0.081)	(0.063)	(0.074)	(0.072)	(0.063)	(0.065)
Other	0.34***	0.60***	0.70**	0.70**	0.69**	0.76*	0.65***	0.70**
	(0.044)	(0.085)	(0.086)	(0.082)	(0.083)	(0.088)	(0.075)	(0.081)
Father's age at completion of education (reference: 14 or less)								
15	1.20*	1.24*	1.11	1.02	1.19*	1.03	1.11	1.03
	(0.102)	(0.119)	(0.083)	(0.076)	(0.089)	(0.076)	(0.082)	(0.076)
16	1.09	1.00	1.14	1.07	1.13	1.00	1.10	0.99
	(0.107)	(0.108)	(0.098)	(0.090)	(0.096)	(0.084)	(0.092)	(0.082)
17	0.92	1.04	1.25*	1.21	1.27*	1.10	1.29*	1.08
	(0.107)	(0.136)	(0.131)	(0.122)	(0.132)	(0.111)	(0.130)	(0.108)
18 or more	0.79*	0.82	0.98	0.96	1.05	1.00	1.06	0.92
	(0.083)	(0.094)	(0.092)	(0.088)	(0.097)	(0.091)	(0.097)	(0.083)
N	15270	15270	15270	15270	15270	15270	15270	15270
pseudo R^2	0.036	0.034	0.026	0.040	0.028	0.031	0.033	0.025

Exponentiated coefficients; Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The pseudo R-squared for the regressions in table 3 above are consistently very small in magnitude, dropping from 3.6per cent in sweep one to 2.5 per cent in sweep nine. This indicates that the combined predictive power of birth characteristics is weak even for the early sweeps. This happens because a large number of variables which affect the probability of response are not accounted for in the model. Such variables may include the characteristics of interviewers and the conditions surrounding the collection of the data. However, metadata are not available in BCS70. One should also note that including explanatory variables from waves other than the first will lead to the loss of observations in a response model. In fact, only observations with complete cases can be included. Missingness in this case could be the result of previous attrition or item-non-response on a particular variable.

Turning our focus to the parameter estimates for the regression results in table 3 we obtain some indicative understanding of the response process. As expected, women are more likely to respond than men. The effect of gender becomes stronger after sweep three. This happens because the sample becomes more skewed towards women and because responsibility is transferred from parents to CMs at age 26 hence the characteristics of the CM become more important than those of the parents. Individuals whose parents were married at birth are also more likely to respond than those whose parents were single. In contrast, individuals whose mothers were living in London in 1970 and those whose mothers did not attempt breastfeeding are less likely to respond. Further, the probability of response drops with parity. The higher the number of older siblings a CM has the less likely he or she is to respond. The probability of response is strictly increasing in the age of mothers at delivery for all sweeps and is higher for CMs whose mothers had longer formal education. However, this last variable does not have a significant effect for all categories.

The higher the social class of the father, the more likely the CM will respond. However, the effect of social class is only significant for the lowest three social classes (e.g. skilled manual, unskilled manual and lowest grade workers). Father's age at completion of education does not have a significant effect on the likelihood of response. One should note that the effects of the independent variables are highly significant (at the level of 1per cent and 0.1per cent) for all variables except fathers social class and father's age at completion of education. However, the explanatory power of the model remains week.

The results from the regression analysis confirm the findings from the descriptive statistics. In other words, attrition is not a random process and dropout will most likely depend on some of the CM characteristics. Hence, working with only the productive cases from any sweep without any adjustments will lead to bias.

In order to carry out the simulation in the next section, we construct inverse probability unit non-response weights based on the response model for sweep four (4th column in table 3). These weights should adjust for unit non-response in sweep four but not for item missingness or unit non-response from other sweeps.

In the following tables we present descriptive statistics based on the birth characteristics used as explanatory variables in the response models (table 3). In sweep one, the sample consists of 15,270 CMs. In sweep four, the sample drops to 10,059 due to unit non-response. The first row gives the percentages for each category at birth (N=15,270), the second row gives the percentages at sweep four (without the use of non-response weights, N=10,059) and the third row gives the percentages at sweep four after adjustment using the non-response weights (N=10,059).[2]

Table 4: The impact of non-response weights in sweep 4

Variables	Men	Father's occupation (skilled manual)	Mother finishing education at 15	Parents are single in 1970
Sweep 1 (at birth)	51.87	59.14	45.61	4.78
Sweep 4 without weights	49.93	58.59	45.43	3.63
Sweep 4 with weights	51.91	59.17	45.59	4.78

[2] Note that the number of observations in sweep one (15,270) and sweep two (10,059) deviate from those in table 2 because some of the CMs included in the category 'participated' have missing birth characteristics. Hence, the observations included in the computation of descriptive statistics and in the logit models are those with non-missing birth characteristics.

Variables	Mother lives in London	Parity=+3	Number of observations
Sweep 1 (at birth)	12.34	13.75	15,270
Sweep 4 without weights	9.53	12.44	10,059
Sweep 4 with weights	12.31	13.72	10,059

Table 4 shows that when non-response weights are not applied, the percentages in sweep 4 deviate substantially from their original value at birth. This indicates that the sample is biased according to these characteristics. In contrast, when the non-response weights are applied, the percentages are almost identical to their original values despite the loss of observations due to non-response. Hence, non-response weights are effective in reducing bias in descriptive statistics and their efficacy is the highest when working with the variables included in the construction of these weights (i.e. the explanatory variables in the response logit models). However, one should keep in mind that the explanatory power of the response models is weak. This weakness will undermine the efficacy of weights in regression analyses which use other variables from those used in the construction of weights.

How effective are weights and imputations?

In what follows we undertake a simulation study using data from BCS70 sweeps 3 and 4 as an illustration. The purpose of the simulation is to assess the effectiveness of weights and imputation techniques in dealing with statistical bias in regression analyses (both in terms of estimates and their standard errors). The two types of bias we are trying to adjust for are unit non-response and item missingness. We use a substantive model with vocabulary scores at the age of 16 as the dependent variable, and gender, age 10 gross family income per week (measured at sweep 3) and highest parental qualification (measured at sweep 4) as the explanatory variables. Income was chosen from a different sweep than the outcome variable (sweep 3) in order to illustrate the complexity of working with longitudinal data. Our simulation consists of a number of steps:

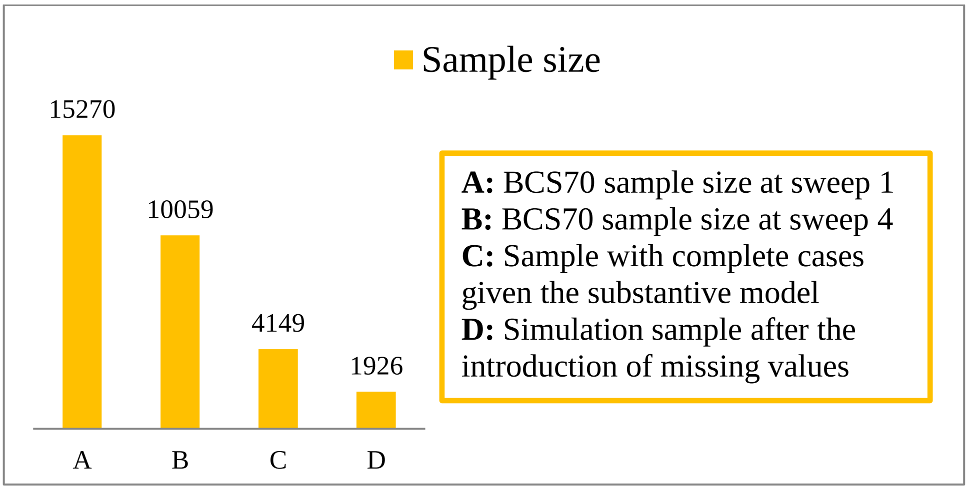
First, introduce missing values on three of the variables: literacy scores, income, and highest parental qualification according to the following rules.

1. On literacy scores, we introduce 10per cent missing values completely at random.
2. We recode the father’s occupation into a binary variable with two categories, manual and non-manual.
3. On income and highest qualification, we introduce 40per cent missing values if the father is from the lower social group and 10per cent if he is from the upper group. We do this for each of the two variables separately. Note that missing values are randomly distributed within each subsample (lower vs. upper). The difference in terms of missing values between the two categories is 30per cent. This difference is large enough to make the bias strong enough to show up in the results.

4. We do not introduce any missing values for gender since it is unlikely to suffer from any item missingness in practice (especially as recorded here on the birth of a CM).

The introduced missingness mimics the reality of working with longitudinal surveys. Such surveys suffer from both unit non-response and item missingness. Even if weights are efficient in dealing with unit non-response bias, item missingness will still lead to additional bias if it is not completely at random. This bias will vary according to the magnitude of missingness and how much it deviates from MCAR. In our simulation, we introduced item-missingness on income and highest qualification. The magnitude of missingness varied according to the father's social class in 1970 – lower vs. upper – with those from the lower group being less likely to answer the questions. Hence, the introduced missingness is not MCAR, and whether it can be assumed to be MAR or MNAR depends on whether the father's social class is treated as observable (MAR) or unobservable (MNAR) in the imputation procedures.

Figure 4: The size of the sample



The base sample consists of the 15,270 CMs in sweep one (Figure 4) with non-missing birth characteristics. The difference between A and B is due to cumulative attrition and unit non-response from sweep one to sweep four (see footnote 3). The difference between B and C is due to the fact that the variables we chose for the substantive model already contain item missingness. Finally, the difference between C and D is due to the item-missingness we introduced to create further sample bias for the purpose of the simulation.

Note that the difference between B and C is due to the combination of missing values on the three variables included in the substantive model: vocabulary scores (missing 5,021, 49.92per cent), income (missing 1,757, 17.47per cent), parents highest qualification (missing 1,188, 11.81per cent). Note that item missingness is high on vocabulary scores because not all CMs have undertaken this test.

After introducing item missingness into the data we estimate the following models:

Model 1: estimated using the sample with complete cases (C) while applying the non-response weights to adjust for the bias resulting from unit non-response (A-B).This model does not suffer from the bias resulting from the introduced item-missingness because it is estimated with the sample with complete cases. This is the benchmark model to which all other models are compared.

Model 2: estimated with the sample with complete cases (C) but without applying the non-response weights. By comparing this model to model 1, we will be able to ascertain by how much non-response weights affect the findings.

Model 3: estimated using the simulation sample (D) with listwise deletion. This model will suffer from both biases (i.e. unit non-response and item missingness) and we are not using any adjustment technique.

Model 4: estimated using the simulation sample (D) with unit non-response weights adjusting for the loss of observations (A-B).

Model 5: estimated using 20 imputed datasets that restore the sample size back from (D) to (C). The imputations adjust for the bias resulting from the introduced item missingness (C-D) but not for unit non-response (A-B).

Model 6: the most complete model and is estimated using 20 imputed datasets that restore the sample size back from (D) to (C) in conjunction with unit non-response weights. This model adjusts for both unit non-response (A-B) and the introduced item missingness (C-D).

One should note that neither of the models adjusts for the bias resulting from the already existing item missingness (B-C). The resulting bias is the same across all models and therefore does not affect their comparability.

The imputations in models 5 and 6 are carried out using multiple imputations with a Markov-Chain-Monte-Carlo procedure to produce 20 imputed datasets. In STATA we use the MI command with a linear procedure to impute literacy scores because the variable is continuous, and ordinal-logit to impute income and highest qualification (the two variables are ordinal). Following the example of Goldtsein (2006, p.69) we produce 20 imputed datasets.[3] The explanatory variables used in the imputation process are the birth characteristics from table 3: gender of the CM, parental marital status at birth, parity, breastfeeding, mother's age at delivery, mother's age at completion of education, and father's age at completion of education. One should note that we did not include father's social class as an explanatory variable. In other words, we treated father's social class as unobservable in the imputation process because the introduced item missingness is based on it. The motivation behind this decision was that item missingness is unlikely to be fully MAR. In other words, missingness will depend on observable and unobservable factors. Therefore, in order

---

[3] Note that we did the same analysis with 100 imputations. The change in the results was very limited in magnitude. This indicates that 20 imputations are enough to generate valid inference.

to mimic reality, we decided not to include father's social class and rather to include proxies for it.

We expect that models 5 and 6 will generate the closest results to model 1 in terms of estimates and in terms of standard errors. Model 3 is expected to generate the least similar results since it does not adjust from any type of bias.

Table 5: Results for the simulation study

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Gender						
Women	8.99**	8.41**	5.50	5.93	9.91**	10.4***
	(2.921)	(2.921)	(4.256)	(4.248)	(3.098)	(3.099)
Age 10 gross family income per week (reference: under £50)						
£50 - £99	3.61	2.22	0.90	2.93	1.97	2.78
	(7.464)	(7.538)	(12.054)	(11.919)	(9.508)	(9.532)
£100 - £149	9.40	7.64	2.24	4.45	8.19	8.92
	(7.443)	(7.507)	(11.994)	(11.865)	(8.989)	(9.132)
£150 - £199	14.4	12.9	5.54	7.54	10.0	10.6
	(7.941)	(7.989)	(12.641)	(12.536)	(9.913)	(10.073)
£200 - £249	13.7	11.9	10.5	12.8	17.7	18.7
	(9.184)	(9.195)	(13.898)	(13.820)	(11.240)	(11.468)
£250 or more	28.2**	27.2**	23.4	26.3	27.0*	28.0*
	(9.523)	(9.502)	(14.131)	(14.078)	(11.140)	(11.478)
Parental highest qualification (reference: no qualification)						
Other	24.3*	26.4*	30.6	29.9	20.7	19.8
	(11.816)	(11.897)	(16.687)	(16.410)	(13.989)	(13.807)
Vocational	16.9***	17.2***	23.4**	23.6***	13.8**	14.2**
	(4.502)	(4.545)	(7.256)	(7.163)	(5.262)	(5.265)
O level	32.8***	34.1***	40.1***	38.9***	28.4***	27.9***
	(4.250)	(4.246)	(6.485)	(6.465)	(5.286)	(5.235)
A level	51.3***	51.5***	56.9***	57.3***	44.1***	44.5***
	(5.533)	(5.477)	(8.135)	(8.153)	(7.180)	(7.320)
Nurse	54.5***	56.5***	46.3**	43.1**	47.6***	46.3***
	(9.536)	(9.381)	(14.076)	(14.272)	(10.854)	(11.129)
Teacher	70.3***	69.1***	74.6***	75.4***	59.1***	60.1***
	(9.115)	(8.994)	(12.006)	(12.071)	(9.470)	(9.563)
Higher degree	85.6***	84.5***	90.7***	91.1***	73.1***	74.0***
	(5.288)	(5.250)	(7.434)	(7.444)	(6.000)	(6.030)
Constant	-55.3***	-52.1***	-44.4***	-48.0***	-49.6***	-52.3***
	(8.428)	(8.517)	(13.288)	(13.142)	(9.940)	(9.885)
N	4,149	4,149	1,926	1,926	4,149	4,149

Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The findings show that models 1 and 2 generate almost identical results even though model 2 does not adjust for unit non-response. Both models use the sample with complete cases (C) and do not suffer from item missingness. This similarity between the two models is a first indication that non-response weights don't improve regression analysis by much. One reason for this is that the response logit models have very low predictive power (see pseudo R-squared in table 3). However, one should keep in mind that the weights will remove the bias from any analysis based on the variables used in the construction of weights as seen in table 4.

In terms of standard errors, model 1 generates the lowest standard errors on all estimates as expected. Model 2 generates slightly higher standard errors, this indicates that when unit non-response weights are used in model 1 we gain slightly in precision. Among the models with item missingness, models 5 and 6 generate the closest results to model 1. This is the case for all variables without exception. Moreover, the standard errors on the estimates in model 6 are almost identical to those in model 5. In contrast, models 3 and 4 generate the highest standard errors and they deviate substantially from model 1 indicating that they are less accurate. The reason for this is that models 3 and 4 suffer from a severe loss of observations due to listwise deletion while models 4 and 5 adjust for it through imputations. One should also note that even though standard errors varied in magnitude between the different models, the significance level of the estimates did not vary by much.

In terms of the estimates, the picture is mixed. Model 1 and 2 are very similar since both do not suffer from item missingness. Model 2 slightly deviates from model 1 because it does not adjust for unit non-response. Models 5 and 6 generate the closest findings to model 1 on gender and most of the modalities of income. However, for parental highest qualification models 3 and 4 generate closer results to model 1 on 3 out of 7 modalities. Hence, one can say that multiple imputations will bring the estimates closer to their original values but with some exceptions. Note that Goldstein (2006) only uses two explanatory variables: one is binary (gender) and the other is continuous. In his case, imputations reduced the size of the standard errors on both variables and improved the estimate only on gender. Our findings are similar as imputations invariably reduce the size of standard errors and improve the estimates with some exceptions.

As mentioned before, when working with longitudinal data, a researcher is confronted with unit non-response and item missingness. In our simulation we used unit non-response weights to adjust for the former and multiple imputations to adjust for the latter. The efficacy of both techniques depends on a number of conditions. The efficacy of unit non-response weights depend on the predictive power of the response models used in their construction. In our case, the predictive power of the models was weak. Further, weights are also undermined when variables from different sweeps are used in a regression analysis. In other words, weights can adjust for unit non-response in one sweep at a time. This was the case of our substantive model which used variables from sweeps three and four. It should be noted that the efficacy of weights can be improved by considering other variables that would predict unit non-response. Such variables include metadata accounting for the characteristics of interviewers and the conditions surrounding data collection. Unfortunately, in BCS70 we did not have such data.

When it comes to item missingness, our findings show that multiple imputations improve the standard errors on all variables without exception. This happens because imputations increase sample size to its former level and improve accuracy. Further, imputations improved the estimates in most cases with some exceptions on parental highest qualifications. One should note that in our simulation we introduced item missingness in a way that imputations would enhance the results. Item missingness was introduced on income and parental qualifications based on father's social class (manual vs. non-manual). The difference in terms of missing values between the two categories was 30per cent. This difference is substantial and caused the results to be sufficiently biased in a way that imputations will be better than listwise deletion. In contrast, if the difference in terms of missingness between the two categories was very small then the researcher can assume that item missingness is almost MCAR and that imputations will not make much difference. By and large, the efficacy of imputations will depend on the extent of item missingness, how much it deviates from MCAR, and whether the researcher is able to control for important determinants of item missingness in the imputation process. When such imputations are properly specified, that is when the researcher chooses variables which are correlated with the probability of missingness and when the procedure is adapted to the type of the variable to be imputed (e.g. linear procedures for continuous variables, ordinal and multinomial logit procedures for ordinal and multinomial variables, multilevel models for nested observations, etc), imputations should generate robust and valid inference.

Conclusion

In this paper we examine the extent of non-response in the British Cohort Study (BCS70) and its effect on sample composition over the eight available data sweeps (1970 to 2008). We analyse the determinants of non-response using binary logit models and birth characteristics as explanatory variables. We find that men from lower social backgrounds and with less educated parents are less likely to respond. However, despite the significance of the regression coefficients the predictive power of the models is weak. In the second section of the paper, we use a substantive model in order to examine the effect of non-response weights and imputation techniques on non-response and item missingness, respectively.

On the one hand, our findings show that non-response weights don't improve the estimates or their standard errors by much. The main reason is that the response models used in the construction of the weights have very weak predictive power. Further, in longitudinal surveys, regression analyses involve the use of variables from different sweeps with each being affected differently by unit non-response. Hence, even if the predictive power of response models is high, non-response weights can only adjust for bias in one sweep. However, in spite of this limitation, unit non-response weights are still effective in reducing bias when the variables used in the construction of the weights are being used in the analyses. On the other hand, random multiple imputations increase the number of observations available for a regression analysis and reduce the standard errors on the estimates. However, in terms of the estimates themselves, our findings show that imputations improved them but with some exceptions.

In conclusion, we can say that the efficacy of weights and imputations in dealing with bias resulting from unit non-response and item missingness depends on the extent of bias and whether variables correlated with the probability of unit and item non-response can be found. These variables can then be used in constructing weights and in imputation procedures.

References:

Centre for Longitudinal Studies, Institute of Education. (2014). Technical Report of the 1970 British Cohort Study: Age 42 Survey (2012 - 2013).

Gilks, W. Richardson, S. and Spiegelhalter, D. (1996). *Markov chain monte carlo in practice*. London, Chapman and Hall.

Goldstein, H. (2009). Handling attrition and non-reponse in longitudinal data. *Longitudinal and Life Course Studies.* Volume 1, Issue 1, pp. 63-72.

Hawkes, D. and I. Plewis (2006). Modelling non-response in the National Child Development Study. *Journal of the Royal Statistical Society*, *A.* Volume 169, p. 479-492.

Little, R. and Rubin, D. (2002). *Statistical analysis with missing data,* 2nd ed. Hoboken, N.J, Wiley.

Micklewright, J., Skinner, C. & Schnepf, S. (2012). Non-response biases in surveys of school children: the case of the English PISA samples. *Journal of the Royal Statistical Society.* Series A. Volume 175, Issue 4, p. 915-938.

Nathan, G. (1983). A simulation comparison of estimators for a regression coefficient under differential non- response. *Communications in Statistics: Theory and Methods.* Volume 11, p. 645-659.

Nathan, G. and D. Holt (1980). The effect of survey design on regression analysis. *Journal of the Royal Statistical Society* B. volume 42, p. 377-386.

Pfeffermann, D. (2001). *Multilevel modelling under informative probability sampling*. 53rd Session of International Statistical Institute.

Plewis, I. (2011). Auxiliary variables and adjustments for missingness in longitudinal studies. CCSR working paper No. 2011-04.

Rubin, D. B. (2004). “The design of a general and flexible system for handling non-response in sample surveys.” *American Statistician.* Volume 58, p. 298-302.

Rubin, D. (1987) *Multiple Imputation for Non-response in Surveys*. J. Wiley & Sons, New York.

Schafer, J. and Olsen, M. (1997). Multiple imputation for multivariate missing-data problems: a data analyst's perspective. *Multivariate Behavioural Research.* Volume 33, p. 545-571.

Schouten, B. and De Nooij, G. (2005). *Non-response adjustment using classification trees.* Voorburg, Statistics Netherlands.

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