National Child Development Study
User Support Group

Working Paper
No 13

RESPONSE TO A NATIONAL LONGITUDINAL STUDY: POLICY AND ACADEMIC IMPLICATIONS IN THE STUDY OF CHANGE

by

D. Hutchison

National Foundation for Educational Research
in England and Wales
The Mere, Upton Park, SLOUGH, Berks, SL1 2DQ

Social Statistics Research Unit
City University
Northampton Square
LONDON EC1V 0HB

AUGUST 1986

NCDSUSGWP13:DH;080586

National Child Development Study User Support Group Working Paper Series

This Working Paper is one of a number, available from the National Child Development Study User Support Group, which report on the background to the Study and the research that has been based on the information collected over the years. Other Working Papers in the series are listed below.

No.	Title	Author(s)	Date
1.	The National Child Development Study: an introduction to the origins of the Study and the methods of data collection	P. Shepherd	October 1985
2.	Publications arising from the National Child Development Study	NCDS User Support Group and Librarian, National Children's Bureau	October 1985
3.	After School: the education and training experiences of the 1958 cohort	K. Fogelman	October 1985
4.	A Longitudinal Study of Alcohol Consumption Amongst Young Adults In Britain: I Alcohol consumption and associated factors in young adults in Britain	C. Power	December 1985
5.	A Longitudinal Study of Alcohol Consumption Amongst Young Adults In Britain: II A national longitudinal study of Alcohol consumption between the ages of 16 and 23	M. Ghodsian and C. Power	December 1985
6.	A Longitudinal Study of Alcohol Consumption Amongst Young Adults In Britain: III Childhood and adolescent characteristics associated with drinking behaviour in early adulthood	M. Ghodsian	December 1985
7.	Report on the longitudinal exploitation of the National Child Development Study in areas of interest to DHSS	Mildred Blaxter	April 1986

RESPONSE TO A NATIONAL LONGITUDINAL STUDY
POLICY AND ACADEMIC IMPLICATIONS
IN THE STUDY OF CHANGE

BY

DOUGAL HUTCHISON
NATIONAL CHILDREN'S BUREAU
8 WAKLEY STREET
LONDON EC1V 7QE

Summary

Final sample sizes achieved in multivariate analyses of longitudinal study data can frequently be substantially smaller than the target population of the study. Only part of such sample losses is due to non response, other reasons being restriction of analyses to sub-populations and losses in the course of operationalising concepts. This paper investigates in detail the response rates, and the losses due to factors other than non-response in connection with one large longitudinal study, the National Child Developement Study NCDS.

The paper argues that the seeming weakness of longitudinal studies compared with cross-sectional studies as far as sample attrition is concerned may be as much due to the strength of a longitudinal study in identifying its target population as to losses on the achieved sample.

The paper also makes use of another strength of longitudinal studies i.e. the ability to assess the effects of sample losses. We find as usual in such investigations that the non-respondents

tend to be either more mobile or more disadvantaged than the remainder.

We also extend our investigations to look at a topic which is central to such longitudinal studies, that of relationships between subgroups or variables, and we find that these in general are considerably less affected than might be expected in comparison with the overall response rate.

Response to a national longitudinal survey: policy and academic implications in the study of change

For some time now in Great Britain, and increasingly in other countries, (Mednick and Baert 1981) longitudinal studies usually in tandem with cross-sectional studies, have been widely used in studying the development of children. There have been three major national cohort studies in Great Britain, namely the 1946 cohort, the National study of Health and Development NSHD (e.g. Douglas, Ross and Simpson 1968); the 1958 cohort, the National Child Development Study NCDS (e.g. Fogelman, ed, 1983); and the 1970 cohort, Child Health and Education in the seventies CHES (Butler, Golding, Haslum and Stewart-Brown 1982). These have all had significant impacts on policy on the welfare, health and education of children of all ages, both indirectly by acquainting the interested public with a problem so that the concern aroused was fed back to policy makers (for example when the NSHD identified the loss of talent and the uncertainty of the 11 plus selection procedure, this played an important role in the comprehensive school movement), and probably rather more often, directly, when a government department commissions one of these longitudinal studies to investigate a particular question in more detail (for example the NCDS study

of selective and non-selective secondary schools, Steedman 1980).

Such studies represent a large capital outlay on the part of sponsors. They also have a long lead time in the first instance, though in practice all three cohorts have more than compensated for this by the wide range of information they collect and their 'nose' for questions of future interest, so that after the initial data collection, it is possible to provide longitudinal results in simply the time to analyse and write up. However, the capital outlay, along with the effect that results from such studies may have render it essential that all concerned with them should ensure the validity and of the results.

This paper concentrates on one aspect of the validity of results, namely the effect of non-response, and on one of the studies, namely NCDS, the 1958 cohort. The NCDS has examined in detail the effect of non response in virtually all of its projects: some examples are published in Goldstein (1976) and Davie et al (1972)

Readers of papers relating to NCDS have upon occassion commented on the apparently relatively small sample size in some analyses: for example, the majority of analyses in the Progress in Secondary Schools Project (Steedman 1980) have an analysis N of in the region about 3000 compared with the 16000-odd pupils in the cohort. This suggests two, related questions.

Does this dramatic paring-down of the sample represent an excessive non-response rate, or are these other reasons than non-response for it? Secondly, even if the reasons are satisfactory, can we be sure that relationships within the sample are not distorted by the restrictions of the sample? This paper will investigate the components of reduction of sample size and the effect that these have on the attributes of the sample as a whole and on relationships between subgroups and variables. This will be achieved by considering one of the differences between subgroups, correlations between variables and an example of a GLM/AOV which investigated progress in maths between 11 and 16 in grammar, secondary modern and pre 1974 comprehensives, partialling out the effects of age of youngest pupil, parental social class and interest and respondents sex. The total n for this last analysis is 3087. Projects assessing the effect of non-response (including NCDS - Goldstein 1976) have generally found that non-respondents differ from the rest of the sample in a social survey, being for example more mobile (not surprisingly) or more likely to be disadvantaged. Such differences are well known and widely documented (see for example Moser & Kalton 1971) for a very readable account) and we shall not repeat them here. Neither do we propose here to look at imputation procedures to allow for the effects of non-response, such as have been described by Bailar et al (1978)

The National Child Development Study

The data on which our analyses are based is taken from the National Child Development Study. This is a longitudinal study carried out by the National Children's Bureau, looking at all children in England, Scotland and Wales born in the week 3rd-9th March,1958. Initially the survey was designed to examine factors associated with perinatal mortality (Butler & Alberman, 1969) but the opportunity arose to trace and study these children again and the cohort of some 16000 was subsequently followed up at ages 7, 11 and 16 years. Information on the children was gathered from three main sources - from schools, where teachers completed a schedule and administered attainment tests; from parents (usually mothers) who were interviewed by a health visitor; and from doctors who undertook a medical examination and completed a medical schedule. Specific tests of attainment in reading and mathematics or arithmetic administered at all three follow-ups. At 7 years the Southgate Reading Test was used and at 11 and again at 16 years a reading test constructed by the National Foundation for Educational Research to be parallel with the Watts- Vernon tests of reading comprehension. The arithmetic test at 7 years and the mathematics test at 11 years were expressly designed for this study in co-ordination with the NFER. The 16 year old test of mathematics was constructed at the University of Manchester for the NFER for their Comprehensive Schooling project. Note that the same reading test was used for 11 and 16 years but the mathematics tests were different. Additionally at 11 and 16 years the subjects provided material about themselves.

NCDS is a large multidisciplinary study as well as a longitudinal one: As explained above, information was collected on a number of occasions and using a number of instruments (mainly questionnaires) on

each of the last three occasions. Response to these instruments was not independent, in the sense that people who responded to one questionnaire at (say) 11 were more likely to respond to the others, but neither was it nearly uniform. This meant that non-response could cumulate where items from several instruments were used in one analysis. To take a hypothetical example, suppose that we were to use items from 4 instruments, each of which had the satisfacory response rate of 85 per cent, then if these non-response rates were independent, the fraction of the total sample contained in the analysis would be $(.85)^4 = 52$. However, before the reader jumps to the conclusion that this represents a fatal flow in such studies, we must remember that longitudinal studies score score here compared with purely cross-sectional studies, since we can gain some impression of the effect of these losses on the results: this will be a major feature of this paper.

The collection of data

We shall discuss data collected about the whole cohort on four occasions:

(i)	at birth	Perinatal Mortality Study
(ii)	at 7	NCDS I
(iii)	at 11	NCDS II
(iv)	at 16	NCDS III

These were not the only occasions of data collection Public exam results were collected from schools and other institutions, where appropriate. Data is at present (1983) being analysed which was collected when the members of the cohort were 23 (NCDS IV). In addition data was collected from subsets at 8½

(Adopted), 7 and later (In care) at 9 Gifted), 18 (Handicapped study), 20 (Feasibility study for NCDS IV) and from agencies at 15 (One parent), and effect of non-response on the exam results sample is discussed elsewhere (Hutchison, 1982).

The methods of data collection are described more fully else where (e.g. Fogelman, 1976) We give below brief account of the methods used.

The birth study questionnaires were distributed early in 1958 by Regional Hospital Boards and the boards of governors of teaching hospitals to maternity departments, and by medical officers of health to domiciliary midwives to be completed for all births in the week 3rd-9th March. Sets of questionnaires were also distributed to departments (such as premature baby units) where babies might be admitted and die soon after birth. Completed questionnaires were returned as soon as possible after delivery, or after death in the case of neonatal death, to the Medical Officer of Health for the appropriate County or County Borough, where they could be checked against the official notifications of birth and death before being returned to the survey unit. In the event, all local authorities bar one participated and an estimated 98 per cent of births in the week had completed questionnaires, a total of 17,416.

All schools in England Wales and Scotland were circulated asking them to inform the study team of the name and house address of all children on their registers with a relevant date of birth. Further steps were then taken to trace the remainder, including contacting Local Health authorities for children in care, writing to last-known addresses, and advertising in the press. The process is described in more detail in Fogelman (1976).

The information on the children in the first follow up was gathered from three main sources: from schools by mean of a schedule completed by the head-teachers and class-teachers and specific tests and other assessments from mothers and sometimes fathers, who were interviewed by an officer of the local authority, usually a health visitor, using a structured interview schedule and from school health services, who undertook medical examinations, carried out some special tests and completed a medical schedule. Due to shortage of time, it was not possible to put as much effort into tracing children who were not traced through schools, as would have been wished or as was in the event carried out on the 11 and 16-year sweep.

Eleven year follow up: Tracing was carried out as at seven, with the addition of a final stage in which National Health Service Executive Councils forwarded letters on behalf of the study to the remaining untraced families.

Sixteen year follow-up: Tracing was carried out as at 7.

Occasions on which data was available

A detailed examination of numbers in the target cohort together with gains and losses at each age is published elsewhere and we do not propose to repeat that here. See Goldstein (1976).

Many longitudinal studies (e.g. NSHD 1946 cohort) have aimed to follow up specified individual identified at the outset. However, up to the age of 16, the NCDS sample was contacted afresh each time data was collected. This gives rise to quite a complicated pattern of combinations of respondents with data at some ages and not others. In addition while some young people were lost to the sample by death or emigration the sample was augmented at each sweep by including information on other young people who were either new to Britain or who, though born in Britain, had somehow 'slipped the net' on earlier occasions. This independent tracing at four ages give us reasonable confidence that virtually all of the target week birth who were in the country on all four occasions will be included in our sampling frame if we include all those not known to be imigrants, or to have died.

Since we wish to compare the total numbers of NCDS respondents over the period from birth to 16 with the total numbers in the cohort of British young people during that period, those who have been lost to the British cohort either by death or emigration have generally been excluded from our comparisons:

we are only investigating the effect of 'avoidable' losses (see below), that is those individuals who were not traced, or not interviewed.

(Table 1 about here)

Table 1 shows the total number of individuals, for whom there is data on certain combinations of occasions. As stated above, deaths and emigrants are excluded. The first number of interest is the total number of such surviving individuals for whom there is data at <u>any</u> of the stages, birth, 7, 11 or 16, and there are 16,676 of these. If we exclude immigrants and those new to the study at 7 or 11 or 16 as being clearly unable to have data at all 4 points in time, the total with data at any of the four stages is 15,625. Among these 12,189 (78.0 per cent) had some data at all four stages. In general, however NCDS analyses use data from fewer points in time than this. There are 15,616 respondents with data at either birth or 7 or 11 of whom 13,494 (86.4 per cent) have data at all three stages. Of those 15,847 with data at 7 or 11 or 16 a total of 12,591 (79.5 per cent) have data at all three ages. It is at first sight surprising that this number should be larger than those with data at any of <u>birth</u> or 7 or 11 or 16, but this is because this latter total excludes those new to the study at 7, whereas the former does not.

The majority of NCDS 3 analyses (see Fogelman (ed) forthcoming) refer to development only between 11 and 16, so it is of interest to note that 15,746 have some data at either 11 or 16 and of these 13,504 (85.8 per cent) have some data at both ages.

Sources of information available at 11 and 16: completed instruments received

In the previous section we considered the occasions at which we received information from our respondents. In this section we go on to examine in greater detail the types of information available at two ages, namely 11 and 16.

At 11 information was collected in the following ways (1) From the school, via a questionnaire to the staff of the respondent's school, test booklets completed by the respondent and a Bristol Social Adjustment Guide completed by the respondent's teacher; (2) An interview with one or other parent, usually the mother and (3) From the school medical service, a medical examination and an audiogram. These we shall refer to as (1) Educational (2) Parental and (3) Medical information. All these would take place at different times, and not all 11-year respondents would have all three sets of information though, of course, most did. A child might have been absent ill for (say) the medical examination,but be present for the testing, and so on.

(Table 2 about here)

Table 2 shows the overlaps in sources of 11-year information. It might be expected that if a child can be traced at all at any given age, then he or she will be put through the entire process. Certainly the great majority (81.0 per cent) of those with 11 year data have got data from all three 'sources'. However, nearly one fifth are missing on one or more, of whom about half have information from one source, and half from two.

The situation at 16 is somewhat more complex and we consider 4 different sources: (1) Educational: questionnaire completed by the school and tests undergone by the respondent; (2) Parental: interview with one or other parent, usually the mother (3) Medical: examination by the school Health Service and (4): Individual: a questionnaire to the respondent.

(Table 3 about here)

TABLE 1

Numbers with some data on certain combination of sweeps, excluding those known to have emigrated or died

Total number of individuals for whom there is some data at

1(a)	Birth or 7 or 11 or 16	16,676
2(a)	Birth or 7 or 11 or 16 (excluding immigrants and all those new to the study at 7 or 11 or 16)	15,625
2(b)	Birth and 7 and 11 and 16	12,189
3(a)	Birth or 7 or 11 (excluding immigrants and all those new to the study at 7 or 11)	15,616
3(b)	Birth and 7 and 11	13,494
4(a)	7 or 11 or 16 (excluding immigrants and all those new to the study at 11 or 16)	15,847
4(b)	7 and 11 and 16	12,591
5(a)	11 or 16	15,746
5(b)	11 and 16	13,504

Note: that all these totals exclude those known to have emigrated or died.

Sources of 11-year information

Only one source	N	Percent of 11 year respondents
Educational	1155	7.5
Parental	112	0.7
Medical	213	1.4
	1480	9.5
Only two sources		
Educational and parental	585	3.8
Parental and medical	737	4.8
Medical and educational	112	0.7
	1433	9.3
All three sources	12444	81.0
Any 11-year data	15358	100.0
No 11-year data	3201	
Total	18559	

Note (1) 'Educational' includes the questionnaire completed by the school, the British Social Adjustment Guide completed by the teacher and the test booklet completed by the respondent. 'Medical' includes the medical examination and the Audiogram conducted by the Schools Medical staff.

(2) N s are not the same as in previous tables as we have not attempted to exclude emigrants or deaths or immigrants.

TABLE 3

Sources of 16-year information

Only one source	N	Percent of 16 year respondents
Educational	192	1.3
Parental	293	2.0
Medical	332	2.3
Individual	4	0
	821	5.6
Only two sources		
Educational and parental	93	0.6
Educational and medical	96	0.7
Educational and individual	1056	7.2
Parental and medical	1201	8.2
Parental and individual	0	0
Medical and individual	2	0
	2448	16.7
Only three sources		
Educational, Parental + Medical	394	2.7
Parental, Medical + Individual	24	0.2
Medical, Individual + Educational	1325	0.9
Educational, Parental + Individual	856	5.8
	2599	17.4
All four	8830	60.1
Any 16-year data	14698	100.0
No 16-year data	3861	
Total	18559	

Notes (1) 'Educational' includes education questionnaire completed by school and test booklet by respondent.
'Medical' includes the medical examination and the Audiogram carried out by the School Health Service
(2) We have not excluded immigrants, emigrants and deaths in this table.

Table 3 shows that not only were there overall fewer respondents at 16 (14698) than at 11 (15358) but of these there were a smaller proportion, 60.1 per cent, with all information sources. This is not entirely due to the requirement of obtaining data from an additional sources, since those with all the 3 11-year sources (Educational, Parental and Medical) are only 62.8 per cent of those with any information.

Origins and destinations

The presence of NCDS data for a respondent at one age is no guarantee that he or she will have data at any subsequent occasion, or that he or she will have supplied data on an earlier occasion. Cohort members can be lost either from the target cohort and the achieved sample because of death or emigration, or from the achieved sample because failure in tracing or contact at a subsequent sweep, or by refusal to co-operate either directly or by proxy. We shall describe these two types of losses as unavoidable (deaths and emigration) and avoidable (untraced, non-contacts, refusals), respectively.

The terms 'avoidable' and 'unavoidable' are perhaps slightly misleading as, at least in theory, it would be possible to avoid losing touch with emigrants, and conversely, it is arguable that some losses are unavoidable since no survey ever gets 100 per cent response and at least some of the respondents must be uncontactable during the period of the fieldwork, either through illness or by

TABLE 4

7-year fate of those with PMS data

		N		
Data at 7		14870		
Refused		81	}	
Untraced		1193	}	1274
Subtotal possible non-emigrant contactable		16144		
Emigrants		421		
Deaths	Stillbirths	388	}	
	Early neonatal (under 7 days)	224	}	
	Late neonatal (7-28 days)	58	}	851
	After 4th week	181	}	
		17416		

going missing. However the terms do relate to the relevant sample for development studies since they distinguish between those who on the one had did not live in Britain at the start and end of the study, and whose history is thus not material to the development of young people in Britain between the ages of (say) 11 and 16, and those on the other hand, whose history would have been material since they did live in Britain during this period.

(Table 4 about here)

Table 4 shows what had happened at 7 to the birth target population. Of the 17,416 for whom data was available at birth, 851 were known to have died, the great majority at or near birth, and 421 known to have emigrated, making a total unavoidable losses of 1,272; and of the 16,144 possible non-emigrant contactable, no data was obtained for 1,274 because of 'avoidable' refusals, non-contacts and failure in tracing.

(Table 5 about here)

Table 5 shows the destinations at the 15,407 who had provided data for the 7 year sweep. Of this total 15 were known to have died since 7 and 331 to have emigrated, making a total of 346 unavoidable losses; and of the 15,061 who were not thus identified as being uncontactable 704 (4.7 per cent) refused to co-operate and 275 (1.8 per cent) were either untraced or not contacted, making 6.5 per 'avoidable' non-response at 11 among those with 7 year data

(Table 6 about here)

TABLE 5

Fate at 11 of those with data at 7

Data at 11	14,082	
Refusals	704	Avoidable losses
No data at 11 (untraced and non-contacts)	275	Avoidable losses
Sub-total possible non-emigrant contactable	15,061	
Emigrants	331	Unavoidable losses
Deaths	15	Unavoidable losses
Total data at 7	15,407	

TABLE 6

Fate at 16 of those with data at 11

Data at 16	13,577
Refusals	834
No data at 16 (untraced and non-contacts)	722
Sub-total possible non-emigrant contactable	15,133
Emigrants	194
Deaths	31
Total data at 11	15,358

TABLE 7

Characteristics of those available for one NCDS sweep to the next compared with original total (Effect of 'unavoidable' losses 7-11)

	Available	Total	χ^2	Df
(i) Perinatal RG parental Social Class				
I,II	17.3	17.4		
III NM	9.9	9.8	1.3 (NS)	2
M,NMH	72.8	72.8		
(ii) 7 year RG parental social class				
I,II	18.3	18.4		
III NM	9.2	9.2	15.3**	2
M,NMH	72.6	72.4		
(iii) Tenure at 7				
Owner-occupier	42.0	42.2		
Council	40.2	39.8	52.4***	3
Private rented	12.4	12.4	(34.1***	2)
Other	5.5	5.6	excluding 'other'	
(iv) Number of schools 5-7				
1	80.1	79.7		
2	16.7	17.0	70.5***	2
3+	3.2	3.4		
(v) In care by 7				
Ever	2.2	2.2		
Never	97.8	97.8	.8 (NS)	1
(vi) Smoking in pregnancy				
No	68.2	68.2		
Yes	31.8	31.8	.0 (NS)	1
(vii) Facilities at 7				
Sole use of all 3	81.3	81.5	10.9***	1
Not sole	18.7	18.5		

Table 6 shows in the same way as Table 5 the destinations at 16 of the 15358 with data at 11. Of those with data at 11 years, 194 emigrants and 31 deaths represented unavoidable losses, and of the remaining 15,133 possible contacts, not known to have emigrated or died, 834 (5.5 per cent) refused and 722 (4.8 per cent) were not contacted.

Unavoidable and avoidable losses

In the previous section we made the distinction between unavoidable losses (those arising from changes in the population under study) and avoidable losses (those arising from lack of success in tracing the cohort target population). It is important to see how these losses affect the characteristics of the NCDS attained sample. We consider first the unavoidable losses, though the effect of these on the NCDS sample is of rather less interest than that of the avoidable losses.

(Table 7 about here)

Table 5 has shown the numbers of avoidable and unavoidable losses between 7 and 11, Table 7 looks at the effect of the unavoidable losses among 7-year respondents due to death or emigrations. The vast majority are the latter, so any differences found are effectively due to the qualities of emigrant families. Because of the large sample size, small differences can be statistically significant, and hence it is advisable to pay attention to the size of the difference as well.

The social class composition of the original total sample at 7 and of those still 'available' at 11 differ slightly with the former containing a higher proportion of social classes I and II and a lower proportion of those from manual families and families without a male head. Emigrants are also more likely to be owner-occupiers and less likely to be council tenants. They are also more likely at 7 to have lived in houses with sole use of all three facilities. All these differences are rather small, but do suggest that on the whole emigrants come from slightly better-off sections of the community. The other statistically significant difference is that in number of schools attended between 5 and 7, with the emigrants, unsurprisingly, already suggesting restlessness by the slightly higher mean number of schools attended.

(Table 8 about here)

Table 6 showed the fate at 16 of those with data at 11. Table 8 compares the 11 year base population with those potentially available as 16 years after unavoidable losses. Again the vast majority of the unavoidable losses were due to emigration, so any statistically significant differences between the base population and the available sample are due to the characteristics of emigrants. On the qualities compared in Table 8, there are statistically significant differences only on tenure at 11, the emigrants again being very slightly less likely to live in council house; and on number of schools

attended.

(Table 8 about here)

Table 6 showed the fate at 16 of those with data at 11. Table 8 compares the 11 year base population with those potentially available as 16 years after unavoidable losses. Again the vast majority of the unavoidable losses were due to emigration, so any statistically significant differences between the base population and the available sample are due to the characteristics of emigrants. On the qualities comapred in Table 8 there are statistically significant differences only on tenure at 11 the emigrants again being very slightly less likely to live in council houses; and on number of schools attended since 5, with the emigrants again apparently exhibiting slightly itchier feet since on average they had attended more schools.

However, as argued above, differences due to unavoidable losses, while interesting, are less important than those due to avoidable losses. Tables 9 and 10 show within the 'available' sample on Tables 7 and 8 respectively the biases due to avoidable losses.

(Table 9 about here)

Among 7-year respondents 'available' at 11, table 9 compares the 7-year characteristics three categories of outcome are considered:

TABLE 8

Characteristics of those available from one NCDS sweep to the next compared with original total (Effect of 'unavoidable' losses 11-16)

	Available	Total	χ^2	Df
(i) Overcrowding				
<1.5 persons/room	87.9	87.9	.7 (NS)	1
>1.5	12.1	12.1		
(ii) Facilities				
Sole use of all 3	87.7	87.8	.7 (NS)	1
Not sole	12.3	12.2		
(iii) 11 year RG parental social class				
I, II	22.7	22.7		
III NM	8.9	8.8	1.3 (NS)	2
M,NMH	68.4	68.4		
(iv) Tenure at 11				
Owner occupier	45.7	45.8		
Council	41.9	41.7	13.6**	3
Private rented	7.7	7.5		
Other	4.9	5.0		
(v) Number of schools since 5				
1	50.7	50.5		
2	33.2	33.2	23.4***	2
3+	16.1	16.3		
(vi) Receiving Supplementary Benefit				
Yes	7.4	7.3		
No	92.6	92.7	1.5	1

TABLE 9

Characteristics of all available respondents from 7 year sweep available compared with those supplying data at 11 year sweep (Effects of 'avoidable' losses

	Refusals	With data	Without data	Total	Df	χ^2
(i) Perinatal RG social class						
I, II	18.0	17.3	15.8	17.3		
III NM	8.9	9.9	9.8	9.9	4	1.3 NS
M, NMH	73.1	72.8	74.4	72.8	2	.6 NS
(ii) 7 year parental social class						
I, II	18.5	19.2	21.6	19.2		
III NM	10.6	9.7	9.1	9.7	4	1.5 NS
M, NMH	71.0	71.1	69.4	71.1	2	.2 NS
(iii) Tenure at 7						
Owner occupier	39.6	42.1	42.2	42.0		
Council	40.1	40.4	26.3	40.2	6	44.4***
Private rented	15.8	12.1	21.1	12.4	3	21.9***
Other	4.5	5.5	10.3	5.5		
(iv) Number of schools 5-7						
1	79.7	80.3	69.5	80.1		
2	17.7	16.6	21.2	16.7	4	34.8***
3+	2.6	3.1	9.3	3.2	2	21.1***
(v) In care by 7						
Ever	2.3	2.1	3.8	2.2	2	2.9 NS
Never	97.7	97.9	96.2	97.8	1	1.2 NS
(vi) Smoking in pregnancy						
No	68.6	66.7	66.5	66.8	2	1.0 NS
Yes	31.4	33.3	33.5	33.2	1	.6 NS
(vii) Facilities at 7						
Sole use	76.2	81.6	77.5	81.3	2	13.5***
Not sole	23.8	18.4	22.5	18.7	1	.7 NS

Details of total numbers in response categories are given Table 5

Refusals, With data and Without data (Deaths and Emigrants are excluded since they have already been covered in Table 7 and 8). Two X^2 figures are quoted for each characteristic investigated, the first comparing these three categories; and the second grouping together the first and third categories in a simple with data/ without data comparison. Again it should be emphasised that statisticially significant differences can be very small in practice. Table 9 shows no statistically significant differences between categories of perinatal or 7 year parental social class, or on percentages in care by 7 or with mothers who had smoked during their pregnancy. There were statistically significant differences in tenure category, with private rented being slightly more represented among the refusals while council tenants, perhaps because of difficulties of moving when living in this type of accommodation or because of the considerable help we received from local authorities in tracing their tenants, considerably under-represented in the 'without data' category.

There is a significant difference between the categories in proportions with the sole use of all three facilities, but the slightly higher proportion in the 'refusals' group

is largely balanced by the lower proportion in the non-contacts, so that the difference between those with some data and the total possible is not statistically significant.

Finally those who had attended a larger number of schools between the ages of 5 and 7 were slightly more likely to be among avoidable losses, possibly because of continued greater mobility. In all the difference in proportions of the populations with certain characteristics due to avoidable losses between 7 and 11 are small, less than .3 per cent.

One could expect, however, that discrepancies over the period 11 to 16 years of age might be slightly larger, because of the higher percentage of avoidable losses, as shown in Table 6.

(Table 10 about here)

TABLE 10

Characteristics of respondents from 11-year ~~sweep available at 16-year~~ sweep compared with those supplying data (Effects of 'avoidable' losses)

	Refusals	With data	Without data	Total	Df	χ^2
(i) Overcrowding						
< 1.5 persons/room	93.1	87.9	82.2	87.9	2	35.8***
≥ 1.5	6.9	12.1	17.8	12.1	1	.1 NS
(ii) Facilities						
Sole use all 3	87.8	87.9	84.3	87.7	2	6.5*
Note sole	12.2	12.1	15.7	12.3	1	2.8 NS
(iii) 11 year RG parental Social Class						
I,II	24.3	22.5	24.7	22.7		
III,NM	8.4	9.0	6.8	8.9	4	5.4 NS
M, NMH	67.3	68.5	68.5	68.4	2	4.4 NS
(iv) Tenure at 11						
Owner occupier	49.2	45.6	44.1	45.7		
Council	38.2	42.2	38.6	41.9	6	19.0***
Private rented	7.7	7.3	11.0	7.5	3	11.4***
Other	4.9	4.9	6.3	4.9		
(v) Number of schools since 5						
1	55.2	50.8	42.3	50.7		
2	30.4	33.2	36.1	33.2	4	26.5***
3+	14.4	16.0	21.6	16.1	2	2.5 NS
(vi) Receiving supplementary benefit						
Yes	4.7	7.4	10.1	7.4	2	14.6***
No	95.3	92.6	89.9	92.6	1	.0 NS

Details of total numbers in response categories are given in Table 6.

Table 10 is similar to Table 9 and compares the characteristics of respondents from the 11-year sweep who would still theoretically have been available at 16, on the basis of their response category at 16.

The differences between the 3 response categories in the proportion in the categories of tenure were statistically significant and, amalgamating the two non-response categories, there was a significant difference between those with data and those without, the council tenants particularly again being slightly more represented in the achieved sample than in the target.

The difference between the three response categories in proportion overcrowded (1.5 persons or over per room) with 6.9 per cent in refusals 12.1 in those with data and 17.8 per cent in those without, was also statistically significant but the two non-response categories are biased in opposite directions and as a result a direct comparison between those with and without data at 16 gives a non-significant difference. The picture is similar with proportions receiving supplementary benefit, with Refusals on the average being slightly better off and No-data slightly worse off than the average.

There is also a statistically significant difference between the three 16 year response categories on number of schools attended, though in contrast to the 11 year response results, the differences between the response categories balance, and the difference between those with and without data are not statistically significant.

In Tables 5-9 we have shown the effects of non-response on characteristics of a sample. However the situation in longitudinal studies is more complex and this we shall discuss in the next sections.

Sources of sample attrition

While the total number individuals available for a longitudinal analysis is frequently less as a proportion of the total number of the target population than would occur in a cross-section study, this does not give a fair assessment of the value of longitudinal studies. The first reason is to do with the nature of the target population in longitudinal studies. Since in the case of NCDS, the target population was traced independently, and successfully, at 4 separate occasions one can be assured that the entire population of those born in the target week is known to the study, with literally a handful of omissions. This means that any failures in tracing at any stage are obvious, one might almost say embarrassingly so. On the other hand, in a cross-sectional survey, while the sample population is specified by means of a sampling frame which may not correspond entirely with the target population, response rate is generally calculated as a proportion of that sampling frame. This is not simply a debating point; for example in a survey of qualified Scottish school leavers the target sample population were a sample by date of birth from those who had left during the past school year, using the

list of such pupils supplied by the schools to the SED from their records of pupils; at face value, a trustworthy sampling frame. However, comparisons with national statistics showed an under-representation of the lower-qualified, and careful investigation (McPherson & Raffe 1978) revealed the surprising finding that schools themselves did not know their leavers sufficiently accurately and tended to miss some of those leaving school in the middle of the year. Not all cross-sectional studies have the conscientiousness, or the capability, to investigate their losses in such detail , but, in contrast, in a longitudinal study such as NCDS, where, for example, pupils were enumerated at 11, losses of this nature would be obvious.

A second, perhaps more important, reason for not underestimating longitudinal studies is that the longitudinal format enables one to gain an impression of the effects of non-response on the analysis being pursued. This has been discussed earlier in connection with the occurrence of certain attributes in the population. The effect on relationships between attributes, and between subgroups will be described below: First we discuss possible sources of sample attrition.

Sample losses

Non-response is by no means the only way which sample attrition can occur. The total n quoted for a particular analysis in a particular project can differ from the overall total in the cohort for three main types of reasons

(a) non-response, of course

(b) the project is investigating a sub-group of the main cohort

(c) losses due to the method of operationalising of concepts.

I shall discuss these three in turn.

(a) Non-response proper. If we look at maths test score at 16, we find that 11,921 have data. If we correlate this score at 16 with score on maths at 11, the total N of cases (10,312) is smaller. This is not because non-response is higher at 11, since it is not, but because a different instrument (test booklet at 11) with its own distinct non-response, has been introduced. At first sight this might seem one respect in which longitudinal studies would be inferior to cross-sectional studies since the use of retrospective data should not give rise to any further loss of data.

However, restrospective data collection is well known to be subject to bias and error (see e.g. Moss & Goldstein eds. 1979, Cook &Alexander 1980) whereas I shall argue below that in longitudinal studies not only is it possible to estimate the extent of non-response bias but that it is generally small.

(b) Secondly, data may be excluded because it belongs to a different sub-population from the one being investigated. For example the NCDS Study Progress in

Secondary Schools (Steedman 1980) comparing selective and comprehensive schools only considered England, so those pupils in Wales or Scotland were excluded.

(c) Thirdly, some cases may be excluded because the data available does not fit into the conceptual framework used. For example in social research, it is customary, if not completely above criticism, to omit those respondents whose parents were in the armed forces in the attempt to operationalise the concept social class using the Registrar General's occupation categories.

While all of these can combine to reduce the sample considerably below the original NCDS, 16,000 - odd, the second and third are not non-response at all.

Example

I consider one specific example to show how the eventual sample is reached. This example comes from Progress in Secondary schools (Steedman 1980) and deals with progress in mathematics between 11 and 16 in grammar schools, secondary modern schools and comprehensive schools. It deals with an analysis for parental interest as assessed by the school at 16, RG parental social class at 11, sex of respondent and age of the youngest pupil in school.

Firstly the study was confined to those with an educational questionnaire at 16. Of the 14,698 with any information at 16 years, 12762 had had an educational questionnaire at 16.

(Table 11 about here)

The study was confined to England and type of school was one of the variables in the analysis. Of the 12,762 with an educational questionnaire at 16 years 10,561 were at school in England. Of these, all but 13 had information on type of school, not surprisingly since this information comes from the same questionnaire as was used to identify the country.

These pupils were tested in maths and reading around the same time but the same time but the sample size would be reduced by a combination of non-response -some of them might have been absent-and non-relevance - it was not felt appropriate to administer academic achievement tests to those attending special schools. Of the 10,548 possible respondents, 9,878 (93.6 per cent) had mathematics test scores at 16. (It also arose that some of the young people who had test scores had no education questionnaire and hence were excluded.) Of the same 10,548, 9,002 had maths scores at 11. A total of 8,414 had mathematics tests scores at both 11 and 16.

Of these 8,414, 6,097 had been in 11+, 12+ or 13+ secondary schools for their entire school career. This constraint excluded

TABLE 11

16 year school type X Country at 16

Country	NA	Gram	Sec Mod	Compr by 69	Compr 70-74	Direct Grant	Indep	Rest	Total
England	13	1274	2646	2124	3214	249	439	602	10561
Wales	0	40	64	264	276	3	14	59	720
Scotland	0	33	28	744	498	39	20	19	1481
Total	13	1347	2738	3132	3988	291	473	780	12762

Sample Those with 16 year education questionnaire.

those who had not received their secondary education in a single type of school. 5,454 of these 6,097 had parents who were described by the schools at 16 as being 'very interested' or quite interested' or 'not interested' at 16. This condition excluded those parents who were described as 'over concerned' with their child's education. 4,944 of these 5,454 had parents who could be classified in the Registrar General's social classes I-V or No Male Head at 16. This condition excluded those with parents in the armed forces and agricultural employment.

The final constraint introduced was to confine the had become comprehensive by 1969. This excluded for example transitional comprehensives and direct grant and independent schools and reduced the sample size to 3,087.

Next we turn our attention to the effect of the reduction in sample on one aspect of relationships in the population.

(Table 12 about here)

Table 12 shows the mean sub-population scores on the mathematics test at 16 for grammar, secondary moderns, schools which had become comprehensive by 1969 ('established' comprehensives in the terminology of Steedman, 1980), and 'transitional' comprehensives (those which became Comprehensive between 1970 and 1974). We introduced the variables singly and observed how the consequent reduction in sample size affects the mean scores. As a general rule it was seen that reducing the sample size

TABLE 12

16 year scores in maths for grammar, secondary modern and comprehensives.
Constraints introduced singly : England

Constraints	Grammar	Sec Mod	Compr.	Transitional	Maximum Increase	Range
None (1)	1.030	-.343	-.190	-.107		1.377
	(1235)	(2476)	(1920)	(2988)		
11 maths test	1.034	-.325	-.163	-.083	.026	1.359
	(1097)	(2104)	(1677)	(2599)		
11 maths test	1.042	-.296	-.140	-.064	.029	1.338
+ Age of youngest in school	(890)	(1674)	(1247)	(1878)		
11 maths test	1.052	-.274	-.139	-.047	.022	1.326
Age of youngest in school	(827)	(1455)	(1105)	(1699)		
Parental interest						
11 maths test	1.058	-.264	-.120	-.037	.019	1.322
Age of youngest in school	.776	(1332)	(979)	(1535)		
Parental interest						
RG Social class						

(1) Note that this means that those included must have an educational questionnaire and a test booklet.

leaves a group of respondents with a higher mean score. The mean score for the grammar school group increased from 1.030 with no additional constraints to 1.058 when all 4 variables (11 maths score, age of youngest in school, parent interest and RG social class) were introduced. This is in agreement with the earlier findings in this paper. However, since our main concern in social and educational research is relationships, we also followed the effect on the maximum inter-group difference, that between grammar and secondary modern. Note of course, that the mean scores for the 'No additional constraints' comparison, had the proviso that those included must have an educational questionnaire (for the school type) and a test booklet.

Since all 4 types of school show an increase in mean score as the sample size decreases, grammars from 1.030 to 1.058, secondary moderns from -.343 to -.264, established comprehensives from -.190 to -.120 and transitional comprehensives from -.107 to -.037, the change in the inter-group range (a decrease) is smaller (from 1.377 to 1.322) than for any of the subgroups except grammars, and indeed is considerably smaller when we confine our attentions to secondary moderns and comprehensives only. So far we have looked at the affect of reduction in sample size on relationships between sub-populations. What about the affect on the other main relationships of interest, namely these between variables?

(Table 13 about here)

TABLE 13

11 with 16 year maths correlation. Constraints introduced singly : England

Constraints	Correlation
None	.746 (10312)
England	.755 (8420)
England School type	.755 (8414)
England School type Age of youngest	.752 (6097)
England School type Age of youngest Parental interest	.756 (5454)
England School type Age of youngest Parental interest RG parental social class	.757 (4944)
England School type Age of youngest Parental interest RG parental social class Grammar, sec mod. or compr.	.760 (3087)

Table 13 shows the correlation between 11 and 16 year mathematics as the variables used in the NCDS study Progress in Secondary Schools were introduced, one at a time, and the resultant sample size. Because of missing data both on 11 year score and 16 year score the effective sample size for the correlation between these two variables is 10,312 and the correlation achieved was .746. Bringing in the constraint that the study is limited to England reduced the sample size to 8,420 but only changed the correlation to .775, and bringing in the other variables: Type of school, (Grammar Secondary Modern or Comprehensive by 1969), Age of Youngest Pupil, Parental Interest, Registrar General's Parental Social Class reduced the sample size to 3,087 but had virtually no effect on the correlation, which in fact only changed further in the 3rd decimal place to .760.

Analysis of covariance

The final investigation carried out in this series was to look at the effect of reduction in sample size on the results of analysis of covariance, a technique which has been widely used in NCB studies to compare the progress of different subgroups. Ideally one would set up an analysis of covariance for the entire population and then see how the results were affected as non-response on the dependent and independent variables cut down the number of cases included to the eventual analysis N, but by definition, this is not possible. A comparable approach is to take a 'minimum' analysis of covariance,

consisting of a dependent variable, a grouping variable, and a single covariate, and to replicate this minimum analysis on the sub-samples which would result if a number of covariates were introduced. The examples which we choose again relate to the study Progress in Secondary Schools (op cit)

(Table 14 and 15 about here)

Tables 14 and 15 show 'minimum' analyses with maths and reading, respectively, at 16 as dependent variable, Type of school as grouping variable, and 11 year maths and reading, again respectively, as covariates. Two sets of results are shown, the first in each instance being those for the minimum analysis, where the population has already been reduced by non-response on 11 and 16 test scores and by restricting the analysis to England and the three types of school, (Population 1) and the second those where the population is reduced by the constraints due to the further covariates used in the study, namely Age of youngest in school, RG Parental class and Parental interest (Population 2). Note that not all the reduction in the sample size in the study was due to non-response. Some cases were excluded since they did not fit the conceptual frame-work used in the study: for example young people whose parents were described (by the teachers) as 'over-interested' were excluded. By comparing population 1 and population 2 results, it can be seen that the effect of the exclusions was to raise the mean score of the remainder, by .077 standard deviations from .037 to .114 for maths and by .101

Table 14

Effect of exclusions on ANCOVA (Mathematics)

		Population 1			Population 2			Comparison with Pop 1	
Mean		.031			.114				
Variance		.959			.977				
N		5870			3087				
		Const	DF	X^2	Const	Df	X^2	Df	X^2
Overall const		.073			.099				
School type	Gram	.287			.295				
	Sec Mod	-.187	2	275.6**	-.180	2	175.5***	2	5.0(NS)
	Comp	-.100			-.115				
11 score	L	-.836			-.849				
	LM	-.471			-.452				
	M	-.019	4	35497***	.021	4	1843***	4	4.3(NS)
	HM	.383			.364				
		.943			.916				
R^2		.542			.569				
G - S		.474			.475				
G - C		.387			.410				
Interaction tested									
School type x 11 score			8	49.2***			29.9***	7	5.8(NS)

Table 15

Effect of exclusions on ANCOVA (Reading)

		Population 1			Population 2			Comparison with Pop 1	
		Const	Df	X^2	Const	Df	X^2	Df	X^2
Mean		.041			.144				
Variance		.910			.919				
N		.5897			.3098				
Overall const									
	Gram	.213			.202				
Schooltype	SecMod.	-.123	2	202.6***	-.124	2	112.5***	2	0.9(NS)
	Comp	-.090			-.078				
	L	-1.033			-1.021				
	LM	- .408			- .396				
11 score	M	.034	4	5998***	.024	4	2996.5***	4	1.4(NS)
	HM	.427			.418				
	H	.980			.975				
R^2		.614			.619				
G-S		.346			.326				
G-C		.303			.280				
Interaction tested									
School type x 11 score		39.4	8	39.4***			20.0***	8	3.2(NS)

from .041 to .144 for reading. On the other hand adjusted differences were considerably less affected, if we used the grammar-comprehensive and the grammar-secondary differences to estimate them and the changes due to interactions (labelled Comparison with Population) were not even statistically significant. The larger of the changes in these differences is of the order of .02 SD for each of Maths and Reading, compared with the .10 for the overall mean X^2 values for the significances of main effect differences reduce as the sample size decreases, and the picture for interactions is the same.

Summary and Conclusions

This paper has looked at the common accusation, as quoted by James Douglas (1976) in his more general spirited defence of longitudinal studies (1976) that longitudinal studies are particularly susceptible to non-response, compared with cross-sectional studies, in connection with the National Child Development Study (NCDS) up to the age of 16.

The paper started by establishing the total numbers of NCDS respondents with data at various combinations of ages, such as those with data at all of birth, 7, 11 and 16 or those with 7 or 11 data, Goldstein (1976) quotes overall response rates of 98.2, 91.3, 90.9 and 87.3 per cent for birth, 7, and 11 and 16 respectively. The numbers supplying data at combinations of these occasions are somewhat smaller than these as the non-response at more than one occasion cumulates.

Because of the size and multidisciplinarity of the study, data was collected in a number of separate instruments dealing with specific areas, such as a medical examination at each occasion. Not all (though the great majority of those supplying data at any one age will have supplied data for all instruments. This paper also documents the availability of combinations of questionnaires at 11 and 16, and found that the 11 year data was more 'bunched' than 16 year data, in the sense that the proportion of those respondents with any data at 11 who had data from all topic areas at the age was considerably larger than the corresponding proportion at 16.

In understanding sample attrition in longitudinal studies, it is important to appreciate the distinction between 'unavoidable' and 'avoidable' losses from the cohort, unavoidable losses arising where the composition of the cohort itself has changed, by death or emigration, and avoidable losses arising where we have failed to trace some members of the cohort or our requests for information have met with a refusal. 'Unavoidable' losses are not particularly relevant to longitudinal studies since the relevant sample has to be those living in the relevant locality throughout the period: however for completeness, we have investigated how the characteristics of those supplying data at 7 have been affecte by unavoidable losses by 11, and we have also conducted a similar investigation of the effec of unavoidable losses between 11 and 16. The great majority of unavoidable losses between 7 and 11, and between 11 and 16 were due

to emigration, so any statistically significant differences were due to characteristics of the emigrants.

Longitudinal studies have been accused of high non-response rates, but we show that the effect of this is very much exaggerated. Firstly, the format of longitudinal studies means that deficiencies are shown up which would be disguised in many cross-section studies by non-obvious deficiencies in the sampling frame.

Secondly much of the sample attrition is not due to non-response but to other unavoidable reasons as sample definition and operationalsing concepts.

Finally longitudinal studies have the strength that one is able to assess the effect of at least some of the non-response, and results on these are reported here.

In interpreting the findings of significance tests, one should bear in mind that they were originally introduced for relatively small samples, and that NCDS difference which are so small to be of no practical significance can still be adjudged to be statistically significant: to take an extreme (and non-representative) example, a difference in age of 15 minutes between two groups in the cohort would be statistically significant. In comparing proportions in a given category between respondents to an NCDS sweep (11 or 16) and the total respondents of an earlier sweep (7 or 11 respectively) we find that the largest discrepancy is .3 percent, a

relatively trival difference to anyone seeking a a picture of the population, yet a statistically significant one.

However, reassuring though such figures are, we would contend that the major focus of a longitudinal study should be on change and comparisons. Our analyses have looked at the effect of population attrition on correlation coefficients, subgroup difference, and analysis of covariance and we have found that apparently draconian reductions in sample n can actually have very little effect of the relationships involved. For example, in the correlation investigation a reduction in population size from 10312 to 3087 only alters the correlation between 11 and 16 year. From .746 to .760. A similar result was found by Goudy (1976) but he does not actually look at non-response, since he reaches his conclusion by comparing successive response waves.

Also comparing subgroup means, we find that while sample attrition biases the subgroup means, these biases are generally in the same direction so that the effect on the difference between the subgroups is smaller than the effect on any of the individual subgroups. A similar situation arise with ANCOVA while the population mean changes as a result of attrition there is no significant interaction between sample attrition and adjusted differences.

Thus this investigation of sample attrition in one large longitudinal study has shown, by using material from the sample itself,

that the kind of sample losses encountered in carrying out analyses has only negligible affect on the results, particularly affect on the results, particularly when investigating relationships, such as correlations or subgroup differences.

REFERENCES

BAILAR, B.A., BAILEY, L. & CORBY, C. (1978) 'A comparison of some adjustment and weighting procedures for survey data'. Survey Sampling and Measurement 175-198 1978.

BUTLER, N. & ALBERMAN, E. (1969) Perinatal Mobility.

BUTLER, N.R., GOLDING, J. HASLUM, M., STEWART-BROWN,S. (1982) 'Recent findings from the 1970 child health and education study: preliminary communication Jr. Roy Soc. Med. 781-784.

COOK, M.A. & ALEXANDER (1980) 'Design and substance in educational research adolescent attainments, a case in point' Sociology of Education 53, 187-202.

DAVIE, R., BUTLER, N. R. & GOLDSTEIN, H. (1972) From Birth to Seven Longmans.

DOUGLAS, J.W.B. (1976) 'The use and abuse of longitudinal studies' in M. SHIPMAN (ed) The Origin and Impact of Social Research Routlege & Kegan Paul. London.

DOUGLAS, J.W.B. Ross, J.M. & SIMPSON, H.R. (1968) All Our Future. Peter Davis.

FOGELMAN K. (ed) (1976) Britain's Sixteen-Year-Olds National Children's Bureau, London.

FOGELMAN K. (ed) (forthcoming) The National Child Development Study.

GOLDSTEIN H. (1976) 'Statistical appendix' in FOGELMAN, K. (ed) (1976) Britain's Sixteen-Year-Olds. 63-82.

GOUDY, B. (1976) 'Non response effects on relationships between variables' Public Opinion Quarterly 360-369.

HUTCHISON, D. (1982) 'Statistical appendix' in STEEDMAN, J. Examination Results in Selective/Nonselective Schools. Findings of the National Child Development Study. National Children's Bureau. London. (forthcoming)

McPHERSON, A.F. & RAFFE, D. (1978) 'Some political and social influences on a national education survey' Paper read to Royal Statistical Society Edinburgh.

MEDNICK, S.A. & BAERT, A.E. (1981) Prospective Longitudinal Research: an empirical basis for the primary prevention of psychosocial disorders. OUP.

MOSER, C.A. & KALTON, G.G. (1971) Survey methods in social investigation Heineman London

MOSS, L. & GOLDSTEIN, H. (eds) (1979) The Recall Method in Social Investigation. University of London Institute of Education, Studies in Education 9.

National Child Development Study User Support Group Working Paper Series

No.	Title	Author(s)	Date
8.	Health and social mobility during the early years of life	C. Power K. Fogelman A.J. Fox	May 1986
9.	Effects of ability grouping in secondary schools in Great Britain	A.C. Kerckhoff	June 1986
10.	Leaving the parental home: an anlysis of early housing careers	G. Jones	July 1986
11.	Stratification in youth	G. Jones	July 1986
12.	Social class changes in weight-for-height between childhood and early adulthood	C. Power and C. Moynihan	July 1986
13.	Response to a national longitudial study: policy and academic implications for the study of change	D. Hutchison	August 1986
14.	Drop out from apprenticeship: an application of survival methods to grouped data	D. Hutchison	August 1986
15.	Event history and survival analysis in the social sciences: review paper and introduction	D. Hutchison	August 1986

National Child Development Study User Support Group, Working Paper No 13, August 1986. Social Statistics Research Unit, City university, Northampton Square, LONDON EC1V OHB. Telephone (01) 253 4399

NATIONAL CHILD DEVELOPMENT STUDY

The National Child Development Study (NCDS) is a continuing longitudinal study which is seeking to follow the lives of all those living in Great Britain who were born between 3 and 9 March, 1958.

It has its origins in the Perinatal Mortality Survey (PMS). This was sponsored by the National Birthday Trust Fund and designed to examine the social and obstetric factors associated with the early death or abnormality among the 17,000 children born in England, Scotland and Wales in that one week.

To date there have been four attempts to trace all members of the birth cohort in order to monitor their physical, educational and social development. These were carried out by the National Children's Bureau in 1965 (when they were aged 7), in 1969 (when they were aged 11), in 1974 (when they were aged 16) and in 1981 (when they were aged 23). In addition, in 1978, details of public examination entry and performance were obtained from the schools, sixth-form colleges and FE colleges.

For the birth survey information was obtained from the mother and from medical records by the midwife. For the purposes of the first three NCDS surveys, information was obtained from parents (who were interviewed by health visitors), head teachers and class teachers (who completed questionnaires), the schools health service (who carried out medical examinations) and the subjects themselves (who completed tests of ability and, latterly, questionnaires). In addition the birth cohort was augmented by including immigrants born in the relevant week in the target sample for NCDS1-3.

The 1981 survey differs in that information was obtained from the subject (who was interviewed by a professional survey research interviewer) and from the 1971 and 1981 Censuses (from which variables describing area of residence were taken). Similarly, during the collection of exam data in 1978 information was obtained (by post) only from the schools attended at the time of the third follow-up in 1974 (and from sixth-form and FE colleges, when these were identified by schools). On these last two occasions case no attempt was made to include new immigrants in the survey.

All NCDS data from the surveys identified above are held by the ESRC Data Archive at the University of Essex and are available for secondary analysis by researchers in universities and elsewhere. The Archive also holds a number of NCDS-related files (for example, of data collected in the course of a special study of handicapped school-leavers, at age 18; and the data from the 5% feasibility study, conducted at age 20, which preceded the 1981 follow-up), which are similarly available for secondary analysis.

Further details about the National Child Development Study can be obtained from the NCDS User Support Group.