National Child Development Study
User Support Group

Working Paper
No 18

A NOTE ON HOUSEHOLD INCOME DATA IN NCDS3

by

John Micklewright

Institute of Economics and Statistics
Oxford University

This is a draft paper and comments are welcome.
It is based on work supported by ESRC grant F0023 2294.
The views expressed are those of the author only.
Please do not quote or reproduce this paper without
the permission of the author.

Social Statistics Research Unit
City University
Northampton Square
LONDON EC1V 0HB
DECEMBER 1986

NCDSUSGWP18:JM;011286

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

A Note on Household Income Data in NCDS3*

John Micklewright,

Institute of Economics and Statistics,

Oxford University.

December 1986

*Work supported by ESRC grant F0023 2294 ("Young People's Early Careers: Analysis of Employment History Data in NCDS4"). I am grateful to the NCDS User Support Group for their help with the data. Any errors in this paper, however, are my responsibility as are the suggestions in the final section.

1. Introduction

This note discusses the data on household incomes collected by the third sweep of the National Child Development Study (NCDS). This took place in 1974, a time when the subjects of the study were aged 16. The analysis is restricted to those 12,537 cases responding to the fourth sweep at age 23. The note is concerned largely with shedding light on the situation where NCDS3 income data are not present and so to provide a guide as to "useable" and "non-useable" cases in this narrow sense of quantity. It does not directly consider important questions concerning the quality of data. There are at least four sources of problem in this latter area. Firstly, income questions were not given any priority at NCDS3 and little attention was paid to their design. Secondly, the data were collected through interviews conducted by health visitors with little or no experience of obtaining sensitive survey data of this type. Thirdly, it was intended that each interview be conducted with the child's mother or mother figure with no one else present (NCDS3 Parental Questionnaire p.2). There must therefore be doubts about the quality of answers given by her regarding the income of her husband and other household members. Fourthly, a substantial proportion of interviews were conducted during the "three-day week" of early 1974 and it is NCDS lore that some respondents may have reported their reduced income levels during the three-day week rather than their normal income. Such matters are considered no further here.

The note concludes that while a substantial fraction of

cases do have missing data, the number are by no means as gloomy as they might appear from a first look at the data in the light of the relevant questionnaire and codebook. Indeed, about three-quarters of those cases in NCDS4 with parental data at NCDS3 might be considered as having "complete" income data. A single variable for total net household income at NCDS3 is constructed from the categorical data on different sources and is compared with the distribution of total household income for households with a child of compulsory school age in the 1974 Family Expenditure Survey.

2. The Income Questions and Variables

The relevant questions on the NCDS3 parental questionnaire and resulting variables were as follows:

Q30 "What have been the sources of income of the household during the past 12 months?" Answers multi-coded into N2452-N2461

Q31 "range in which the members of the household's usual net income falls" (12 ranges)

(i) "weekly net pay of father or father figure" - N2462

(ii) "weekly net pay of mother or mother figure" - N2463

(iii) "weekly net income from all other sources (e.g. FAM, earned income of other members of household, investments, private incomes, social security benefits, pensions, FIS etc..) - N2464

(iv) as (i) but monthly - N2465

(v) as (ii) but monthly - N2466

(vi) as (iii) but monthly - N2467

Q32 "If all or part of Q30 and Q31 were not answered, was this because"

- Informant did not know answer?
- Informant did not want to give answer?
- Uncertain of Reason

Answers multi-coded into N2468-9

Much of this note concerns the interpretation of the coding of the responses to these questions. Q31 instructs the interviewer that "either (i) the weekly or (ii) the monthly income is required whichever the informant finds it most convenient to give". On a strict reading this instruction, taken together with the fact that the range for each income type appears to include zero in every case, suggests that unless N2468-9 is coded all respondents should have an entry for either weekly or monthly income under each of the three headings, even where there is no income from that sources e.g. where the mother does not work.[1] There is no indication from the questionnaire that the treatment of zeros was other than this although the idea of a person deciding whether a weekly or monthly zero was more convenient makes one suspicious! The NCDS3 codebook is somewhat more informative but again the position is not entirely clear. The coding of Q31 variables includes "NA" for "No Answer" but does not indicate whether this refers just to refusals etc. or whether it may include zeros due to no income being received from that source. The coding for the bottom range - value 1 - again

---

[1] All references to "mother" or "father" in the rest of this note should be taken as implying any type of "mother figure" or "father figure" and not just the natural parents.

appears to include zero. The coding of Q32 again includes "NA" for "No Answer" but it is not transparent what this refers to. In what follows I try to throw some light on this by looking at various characteristics of those with neither weekly nor monthly income ranges coded under Q31 and who also have no explicit reason for this coded under Q32.

In order to summarise the information contained in responses to Q31, I defined a variable YESCASH to take eight values as follows:

1. Either weekly or monthly income coded into ranges 1-12 for Father's Pay (PA) and Mother's Pay (MA) and Other Income (OT)
2. Other Income (OT) missing
3. Mother's Pay (MA) missing
4. Mother's Pay (MA) and Other Income (OT) missing
5. Father's Pay (PA) missing
6. Father's Pay (PA) and Other Income (OT) missing
7. Father's Pay (PA) and Mother's Pay (MA) missing
8. Father's Pay (PA) and Mother's Pay (MA) and Other Income (OT) all missing.

Thus for YESCASH = 2 to 8, neither weekly nor monthly income is coded into ranges 1-12 for one or more of the three sources. I refer to this situation as "missing" data.

3. Breakdown of the NCDS4 Sample

12,537 cases responded to NCDS4 and for 9,411 of these there are data from a parental interview at NCDS3 (N2350 coded not missing). Table 1 crosstabulates YESCASH by N2468 for these 9,411

Table 1 YESCASH by N2468 (Reason Q30/31 not answered)

YESCASH \ N2468	Missing	1. Don't know	2. Refused	3. Uncertain	Total
1. Complete Data	3,397	35	5	7	3,444
2. OT missing	1,074	41	13	9	1,137
3. MA missing	1,646	49	11	18	1,724
4. MA and OT missing	494	61	16	16	587
5. PA missing	383	303	25	13	724
6. PA and OT missing	46	145	39	13	243
7. PA and MA missing	346	181	53	16	596
8. PA, MA, OT missing	40	364	510	42	956
Total	7,426	1,179	672	134	9,411

cases. N2468 should indicate the first reason given for missing data at either Q30 or Q31. It may thus not necessarily indicate the reason for missing data at Q31 but will be treated here as if it does (any interested user could check this out further). On the other hand it should be the case that if N2468 is not coded then neither will N2469.[2]

The first thing to strike one in Table 1 is that there appears to be "complete" income data i.e. YESCASH = 1 in only just over a third of cases (36.6 per cent). And yet for only a tenth of the sample (9.7 per cent) is there a total lack of data i.e. YESCASH = 8. The bulk of the sample, therefore (5,011 cases) has only partial data and closer inspection reveals that in 80 per cent of these cases there is no explanation coded under N2468 indicating a reason for absence.

The 1,062 cases where data is missing and N2468 is coded are, in a sense, less interesting. There is, after all, an apparent explanation and nothing more can be done about it.[3] The following remarks, however, are offered for what they're worth. Refusal accounts for over half of those cases where there are no data

---

[2]Further investigation revealed that N2469 was coded 2 (refusal) for 24 cases for whom N2468 was coded 1-3. All the remaining 9,387 cases were coded as N2469 = 0, a code which does not appear to be possible from the NCDS3 codebook. However, this would appear to indicate that N2469 is not important.

[3]Note that this interpretation hangs on N2468 referring to the reason data is missing under Q31 rather than Q30 (sources during last 12 months). This could be checked further. The fact that N2468 is coded for 47 cases with complete data (YESCASH = 1) suggests that the interpretation here may be at fault for some cases with partial data.

(YESCASH = 8), with the interviewer being uncertain of the reason for no response in only a few cases. The picture is different however where there is partial data. Interviewer uncertainty rises somewhat but "Don't know" is by far and away the most important reason, outnumbering refusals by a factor of 5 to 1. As might be expected "Don't knows" are particularly frequent where it is the father's pay which is missing, accounting for example, for 88.9 per cent of cases where YESCASH = 5 and N2468 is coded. Table A.1 in the Appendix shows that those cases with N2468 coded were more likely to have a higher father's social class at NCDS3 (where coded). For example, 33.6 per cent of cases with N2468 coded have father's social class equal to I or II compared to only 23.8 per cent of other cases (including those with any YESCASH value).

Turning to the 4,029 cases with missing data (YESCASH = 2 to 8) but no reason given at N2468, the first thing to check is whether there was simply no father or mother present to provide data. This is the subject of Table 2 with N2371 and N2375 being used to define parental absence. The table shows that absence of a mother accounts for only a small number of the cases where mother's pay is missing, but that absence of the father goes quite some way to explaining why father's pay is not coded (YESCASH $\geq$ 5), accounting for a total of 490 cases (60.1 per cent).

The major problem remains however of the large number of apparently unexplained cases where mother's pay and other income are missing - a much larger problem than that of father's pay. It is particularly odd that the former is not present when in general it will have been the mother who was interviewed. One immediately

Table 2 YESCASH by Father/Mother figure Present:

group with N2468 not coded

YESCASH	Father and Mother Present	No Father	No Mother	Total
1. Complete Data	3,397	-	-	3,397
2. OT missing	1,073	1	-	1,074
3. MA missing	1,593	2	51	1,646
4. MA and OT missing	452	-	42	494
5. PA missing	104	279	-	383
6. PA and OT missing	13	33	-	46
7. PA and MA missing	170	168	8	346
8. PA, MA and OT missing	29	10	1	40
Total	6,831	493	102	7,426

Note: No Father defined as N2375 = 10
No Mother defined as N2371 = 10

becomes suspicious about the coding of women who did not work. The obvious reading of questionnaire and codebook is that she may have been coded in the bottom range of weekly or monthly pay which, as described above appears to include zero. However, inspection of the mother's employment status variable (N2392) for cases where MA is missing suggests that the coding practice was usually different. This is shown in Table 3 which is restricted to those cases with N2484 not coded but where the mother is present. Where the mother does not work, her pay is missing in some 95 per cent of cases, compared to only 1 or 2 per cent where she is working. The implication seems clear: most zeros were not coded into the lowest pay range and we can reasonably safely treat cases with a mother not at work and with only mother's pay missing as having "complete" data. Note that one would expect the 33 cases where the mother does not work but MA is coded to have the bottom pay range (weekly or monthly coded) coded. I have not checked this. Similarly, I have not checked further to try and account for why mother's pay is missing for the 121 cases in Table 3 with YESCASH = 3, 4, 7, 8 and where she works.

Before turning to the other major problem - missing other income - the analysis of mother's employment status suggests that it would be worthwhile to look at this for the smaller group (325 cases) where father's pay is missing (YESCASH $\geq$ 5), there is no apparent explanation at N2468, and where the father is present. Again, one interpretation of the questionnaire suggests that even where the father is not at work, there should be a pay figure coded (in the bottom range). Evidence that this did not occur is shown in

Table 3 YESCASH by whether Mother works: group with N2468 not coded and Mother present

YESCASH \ N2392	Not Coded	1. Works	2. Doesn't Work	Total
1. Complete Data	-	3,376	21	3,397
2. OT missing	-	1,067	7	1,074
3. MA missing	9	68	1,518	1,595
4. MA and OT missing	3	20	429	452
5. PA missing	-	379	4	383
6. PA and OT missing	-	45	1	46
7. PA and MA missing	-	18	320	338
8. PA, MA and OT missing	6	15	18	39
Total	18	4,988	2,318	7,324

Note 1) N2392 does not give any finer breakdown of employment status than that shown above.

2) For "mother" read mother or mother-figure.

Table 4. Although the proportion of "suspect" cases (YESCASH $\geq$ 5) who are not at work is somewhat lower than for the mother's - 80.3 per cent compared to the figure of 95 per cent given above - it still makes a stark contrast with the proportion who are working - less than 5 per cent (not shown in the table).

As with the analysis of mother's employment status, the situation where the father is employed (62 cases) and pay data is missing remains suspect until proven otherwise.

Finally, there is the problem of missing other income. There are a total of 1,654 cases in Table 1 where this occurs (YESCASH = 2, 4, 6, 8) and N2468 (reasons for Q30/Q31 not answered) is not coded. Again, the problem is one of looking at other data to see if these cases have family/household circumstances which are consistent with an interpretation of the missing figures as genuine zeros, despite the apparent instructions in the questionnaire to code such cases into the bottom income range. However, getting a handle on this group is somewhat harder - the employment status of any other adults in the household, for example, is not recorded. One avenue, however, suggests itself. We do know the numbers of siblings of the study child and indeed the total numbers of other children in the household who are younger or older but aged under 21. At the time of NCDS3 (1974), Family Allowance (FAM), which should enter "other income", was payable if a family had two or more children below the age of 16 or below 19 and in full education (it was also payable in respect of an apprentice with low earnings). Most of the NCDS children should have safisfied this definition at the time of their parents' interviews for NCDS3 (conducted in the

Table 4 YESCASH by Father's Employment Status: group with N2468 not coded, Father present and Father's Pay missing

YESCASH \ N2383	Not Coded	1. No Male Head	2. Retired	3. Unemp.	4. Sick	5. Other	6. Empl.	Total
5. PA missing	2	-	13	17	46	2	24	104
6. PA and OT missing	1	1	2	-	-	-	9	13
7 PA and MA missing	3	4	41	41	69	9	11	178
8. PA, MA, OT missing	5	-	-	3	4	-	18	30
Total 5 - 8	11	5	56	61	119	11	62	325

Note 1) the reason for N2383 being coded 1 for any cases where N2375 ≠ 10 is not apparent.

2) for "father" read father or father-figure.

majority of cases in the spring before the summer in which the children were elligible to leave at age $16\frac{1}{4}$ [note that the children were affected by the raising of the minimum school leaving age in 1972]). Thus if we can identify those cases where there are other children present such that FAM should have been in receipt, we can at least say that "missing" other income does indeed mean "missing" rather than zero, provided we assume take-up of FAM was reasonably complete. (Of course we cannot infer with confidence the opposite, namely that where the child is a "single child", "missing" other income always means zero, since there could be other possible sources that have not been reported.)

To begin with we need to define children for FAM purposes in terms of NCDS3 variables. This is not easy because although we can be reasonably sure from the data where older siblings are not in the household, we do not know their employment status and thus FAM eligibility if they are still in the household. It seems wise therefore to concentrate on the younger siblings (and siblings rather than all younger household members) since where there is a younger sibling we can be reasonably confident that elligibility for a FAM payment should have occurred (ignoring the possibility that the younger sibling is not in the household which could be checked). The position is given in Table 5, for the 1,654 cases in question with missing other income and N2468 not coded. It shows that in some 35 per cent of cases there is a younger sibling and thus there would in all probability have been FAM elligibility. Here it seems we might want to conclude that the "missing" data does not represent a zero (the position is virtually identical if we look instead at

Table 5 YESCASH by whether any Younger Siblings: group with N2468 not coded and Other Income missing

YESCASH	No Younger Siblings	Younger Siblings	Total
2. OT missing	747	327	1,074
4. MA and OT missing	286	208	494
6. PA and OT missing	32	14	46
8. MA, PA and OT missing	15	25	40
Total 2, 4, 6, 8	1,080	574	1,654

Note: 1) Number of Younger Siblings given by N2368 and N2370

2) Not shown in the table is the fact that where YESCASH = 1 only 25 per cent of cases have no younger siblings compared for example with a figure of 70 per cent for YESCASH = 2 indicated by the table.

younger household members rather than younger siblings). It may be noted however that the 574 cases concerned represent only 11.8 per cent of all those cases with younger siblings and N2468 not coded. Such a figure might lead one to argue that these cases represent non take-up rather than mis-reporting. Against this is the fact that 68.1 per cent of the 574 do report FAM receipt within the last 12 months (one of N2452-61 = 13). (This is true of 93.5 per cent of other cases with younger siblings and N2468 not coded.)

What about the remaining 65 per cent or 1,080 cases in Table 5? Without exploring other avenues it is difficult to come to any safe conclusion. Even where there are no other household members at all aged under 21 (871 cases) we cannot be sure that the missing other income implies a zero because sources other than FAM would need to be considered. It seems that other income is a more problematical area than mother's or father's pay.

4. Summary and Suggestions

It will be obvious to the reader by now that this note came about largely due to the author's failure to correctly interpret questionnaire and codebook. However, in his defence, it could be said that the situation was not particularly clear.

What has been learnt? The main message is that most "unexplained" (in the sense of N2468 not coded) absences of father's or mother's pay or other income appear to be genuine zeros, in that they can be explained by looking at other NCDS3 variables. However, in my view not <u>all</u> such missing data can be safely treated in this fashion. As we have seen, there are cases where the father or

mother works or where there are younger siblings and quite possibly FAM payments, and yet no data coded. The following is suggested as a "reasonable" selection rule to obtain cases with complete household income data. (The rule is cumulative and therefore the numbers rejected at later stages might appear to be less than those suggested by Tables 1-5.)

1. Take the 9411 cases with NCDS4 interviews and parental interviews at NCDS3.

2. MINUS 1985 cases with N2468 coded. This is playing it very safe. It could be that for some cases N2468 refers only to Q30 (sources of income in last 12 months) rather than Q31, although the pattern in Table 1 is reasonably consistent with the interpretation adopted here. Could be checked further.

3. MINUS a further 121 cases with N2468 not coded, Mother Present (N2371 $\neq$ 10) N2392 = 1 (Mother works), and Mother's Pay missing (YESCASH = 3, 4, 7, 8).

4. MINUS a further 48 cases with N2468 not coded, Father Present (N2375 $\neq$ 10), N2383 = 6 (Father employed), and Father's Pay missing (YESCASH $\geq$ 5)

   = 7,257 cases, or 77.1 per cent of the total 9,411.

If one is not prepared to invoke non take-up of FAM the following addition to the rule could be made:

5. MINUS a further 547 cases with N2468 not coded, younger siblings (N2368 or N2370 > 0), and Other Income missing (YESCASH = 2, 4, 6, 8). These are the cases who "should" have FAM. There is the temptation to impute the missing FAM, call this "Other Income" and include these cases. I feel this should be resisted. These cases have "proved" themselves unreliable and other sources may have gone unreported which we cannot impute (comparison of the parental interview dates and school leaving dates might recover some cases where the former was late and the latter early). Another possibility would be to drop only those cases with FAM receipt in the last year coded (one of N2452 - 61 = 13)

   = 6,710 cases, or 71.3 per cent of the total 9,411.

IT SHOULD BE STRESSED THAT THE ABOVE "RULE" IS ONLY A SUGGESTION. FURTHER CHECKING OF THE DATA COULD LEAD TO EITHER MORE OR LESS REJECTION OF CASES.

How should we alter the coding of those apparently "genuine" zeros which we do not reject but which have PA, MA or OT missing? This depends on whether one believes that any zeros were in fact coded into the bottom ranges of N2462-7. Although this may have occured in a few cases - possibly for example the 33 cases in Table 3 where the mother does not work but MA is not missing it might be considered safe to recode the missing but "genuine" zeros to PA, MA, OT = 0, rather than 1, and to hold the view that N2462-7 = 1 excludes zero.

What about the construction of a single household income variable? First we have the problem of the mixture of weekly and monthly reporting. However, this causes little difficulty since each numbered range on one definition was intended to represent the same income per unit of time as the same numbered range on the other. e.g. range 10 is £45-49 weekly and £191-210 monthly. Thus the range codes can be translated into either measure of income per period by referring to the appropriate coding frame. In what follows I take a weekly measure and the distribution of each income source is shown in Table 6 in weekly terms for the 7,257 cases selected by parts 1-4 of the rule above.

The second problem is that of aggregating across the different income types. Clearly one cannot just sum across N2462-7 and interpret the resulting figure using either period coding frame. For if say N2463 = 2 (Mother's Weekly Pay £5-9), N2464 = 1 (Other

Table 6 Distribution of Different Income Sources:

Weekly Net Income

% (n = 7,257)

Weekly Income (£s)		Fathers Pay	Mothers Pay	Other Income
0)	zero	10.1	32.9	22.1
1)	0- 4	0.2	6.1	43.3
2)	5- 9	0.3	14.0	11.3
3)	10-14	0.8	18.8	6.8
4)	15-19	2.1	13.7	5.6
5)	20-24	9.3	8.6	4.4
6)	25-29	17.8	3.3	2.2
7)	30-34	19.6	1.4	1.6
8)	35-39	13.6	0.6	1.0
9)	40-44	9.6	0.2	0.7
10)	45-49	5.6	0.1	0.4
11)	50-59	4.9	0.1	0.2
12)	60+	6.1	0.1	0.4
		100.0	100.0	100.0

Note: Percentages may not sum to 100.0 due to rounding

Weekly Income £0-4) and N2465 = 6 (Father's Monthly Pay £106-125), then sum = 9 and although net household income will indeed be bounded from above by £44 weekly which is the top of range 9, it will be bounded from below by £30 which is the bottom of range 7. The crude solution proposed here begins by summing across the mid-points of each range (treating, for example, £7.5 as being the midpoint of the £5 - £9 range since the £9 includes all figures up to £10). "Genuine" zeros as defined above are treated as such during this summation. The mid-point is of course undefined for range 12 which is unbounded, but if for the time being we define a new categorical household income variable with a top range as for the original variables, then any case with one or more of N2462-N2467 = 12 will automatically be in this top range again. The distribution of this variable is shown in the first column of Table 7, with intervals of £10. The definition of the bounds of each range is now critical since summation of mid-points of N2462-7 will produce values of exactly 10, 15, 20 etc. The ranges in Table 7 include the top value but exclude the bottom value. Note that the range > £60 includes not only all those with one or more of N2462-N2467 = 12 (462 cases) but also those in ranges 1 - 11 of the original variables for whom the sum of the mid-points exceeds £60 (1,278 cases).

The second column in Table 7 is based on a different source. It shows the distribution of usual net household income for those households where there is a child of compulsory school age (5-15) in the 1974 Family Expenditure Survey (FES) i.e. for the same year and

Table 7 Usual Net Household Income NCDS3 and FES Compared

Weekly (£s)	NCDS3 (n=7,257)	FES 1974 (n=1,975)
zero	0.2	0.1
0-10	0.4	0.4
10-20	3.5	2.1
20-30	11.2	8.8
30-40	19.6	17.2
40-50	22.8	21.6
50-60	18.3	18.5
>60	24.0	31.3
Total	100.0	100.0

Note: 1) The ranges include the top value and exclude the bottom value

2) NCDS figures based on summing mid-points of N2362-7

3) FES figures are for Great Britain and refer only to those households with a child aged 5-15. The figures are based on the author's analysis of the microdata rather than published sources.

4) Both sources include all forms of household income e.g. state benefits, investment income and wages of all household members.

broadly the same household type as NCDS3.[4] The FES data provide a useful yardstick by which to assess the picture presented by the NCDS distribution although we should not of course expect the two sources to give identical results. Note that the comparison takes no account of the fact that the FES spreads interviewing evenly throughout the year whereas the NCDS3 interviews were concentrated in the Spring of 1974 - of some importance given wage inflation during the year.[5] This would tend to make the FES figures somewhat higher.

The NCDS figures do appear to be a little lower as one would expect but further investigation would be needed before any firm conclusions could be drawn. Note in addition to the points raised above that the summing at mid-points in the NCDS would tend to distort the comparison. (To avoid this one could compare each income type in Table 6 with equivalents in the FES.) Given the doubts that have been expressed about NCDS3 income data, the general picture in Table 7 is perhaps reasonably encouraging.

Turning to the use in practice of NCDS3 household income as, for example, an explanatory variable in multi-variate analysis, one could just create a number of dummy variables corresponding to

---

[4]The FES is an annual household survey of around 7000 households p.a. The income data are generally reckoned to be of a reasonable quality. See for example A.B. Atkinson and J. Micklewright (1981) "On the Reliability of Income Data in the Family Expenditure Survey 1970-1977", Journal of the Royal Statistical Society, Series A, Vol 146, Part 1, pp.33-61.

[5]FES interviewing in 1974 was in fact suspended during the two General Election campaigns of that year.

ranges such as those in Table 7. However if the problem of the unbounded range 12 for N2362-7 could be overcome one could use the summed values of the mid-points as a single "quasi-continuous" variable; summing over ranges 0-11 produces some 48 values in the data and it seems desirable to exploit rather than abstract from this variation. One possibility is to turn to the genuinely continuous income data in the FES and find a "representative" value for the unbounded range of each of the NCDS3 income types. These could be plugged into the NCDS data with summation then occuring over these values for range 12 and mid-points for ranges 0-11. The most natural "representative" values are probably the medians of all appropriate FES cases in the £60+ range of each income type. A less satisfactory short-cut would be to sum across ranges 0-11 as before and then to take for any case that has one or more of N2462-7 equal to 12 (462 cases in Table 7), the median FES values of total household income in the £60+ range with no further additions. Returning to Table 7 there are 618 appropriately defined FES households in the top range of income, with a median value of £73.66 per week.[6] Rather than making no further additions one could alternatively take a figure of say £72.5 as the representative value for each of the separate unbounded ranges and sum as before. There is clearly room for a more sophisticated treatment of these unbounded ranges which, as Table 6 shows, are largely a problem of father's pay.

---

[6]The mean is £83.67. Note that the mean will be more sensitive to sampling variation as well as to any non-response and under-reporting of households with very high incomes. Such considerations make the mean doubly unattractive.

Appendix

Table A.1 below shows the father's social class at NCDS3 (N2384) for the 8,649 cases in NCDS4 with parental data at NCDS3, for whom this information is coded. It thus excludes 762 cases in Table 1 for when N2384 is not coded (154 with N2468 coded and 608 not coded). The pattern of higher social class where N2468 is not coded is in fact found for all three coded values of this variable.

Table A.1 Fathers Social Class (N2384) by Reason Income Data Missing (N2468)

N2384	Per Cent N2468 Uncoded	Coded
1. Social Class I	5.2	6.9
2. Social Class II	18.6	26.7
3. Social Class III	9.6	9.7
4. Social Class III	45.2	39.3
5. Social Class IV	1.4	1.6
6. Social Class IV	13.8	9.5
7. Social Class V	5.2	4.7
8. Unclassified	1.0	1.6
	100.0	100.0
n	6,818	1,831
N2384 missing	608	154
Total	9,411	

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
16.	Transitions in young adulthoood	K. Kiernan	October 1986
17.	The NCDS5 Development Programme	Peter Shepherd (SSRU)	Oct '86
18.	A note on household income data in NCDS	John Micklewright (Inst of Economics and Statistics, Univ of Oxford)	December 1986

National Child Development Study User Support Group, Working Paper No 18, December 1986. Social Statistics Research Unit, City University, Northampton Square, LONDON EC1V OHB. Telephone (01) 253 4399

N A T I O N A L C H I L D D E V E L O P M E N T S T U D Y

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.