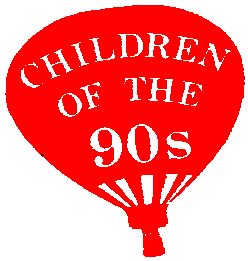
ALSPAC main documentation



Contents

I. THE ALSPAC STUDY DESIGN

ALSPAC (the Avon Longitudinal Study of Parents and Children) is an ongoing longitudinal study of a population of children born to mothers resident in a geographic area in England.

I.1 The overall aims of ALSPAC

The overall objectives of the study are to understand the ways in which the physical and social environments interact over time with genetic inheritance to affect health, behaviour and development in infancy, childhood and then into adulthood.[1,2]

A diagrammatic structure of the way in which the study has been designed is shown in Figure 1. Here it can be seen that the environmental influences considered include psychosocial conditions, features of parenting strategies, the diet and lifestyle of the family, the pollutants (including those in the air, water and food), housing circumstances, health behaviour of the family, medical and dental care, the types of day care and the schooling that the child receives. It is anticipated that the outcomes will depend on complex interactions between the environment and variation in a number of different genetic polymorphisms. Consequences that are being measured include growth, onset of obesity, respiratory function, traits relevant to adult-onset diseases, infections, motor and mental ability, educational achievements, sexual development, accidents and injuries, atopic diseases including asthma, eczema and specific allergies, mood behaviour and temperament. From late adolescence onwards ALSPAC will be able to study the genesis of type 2 diabetes, markers of increased susceptibility to coronary heart disease, schizophrenia and other psychotic disorders, criminal behaviour, ability to hold down a job, onset of drug and alcohol abuse and reproductive successes and failures.

Although the primary focus is health, the broad ranging information collected about the lives of the cohort members enables investigation of many contextual variables of interest to social scientists. For example, information is being collected at regular intervals about their physical environments (for example housing, type of neighbourhood), parental characteristics, social class, family relationships, health behaviour and psychosocial environment during first few months (for example maternal depression, use of child care facilities etc).

I.2 The study area

The study is not a national study – it is based in one geographic region and in consequence has a number of advantages. By being in one area we have been able to efficiently reinforce the commitment of the involved families through the local media; we have been able to access medical records easily (there is just one children's hospital in the area), and the link with schools is very efficient. In addition, the way in which biological samples have been able to be collected during pregnancy and delivery was made much easier by being local and fitting into the normal schedule by which biological samples were forwarded to local laboratories for clinical assays.

Avon itself is situated in England – it is an area bordering the Severn estuary and is about 120 miles due west of London (Fig. 2). The study area (total population 1 million) includes a major city of population 0.5 million (Bristol) and surrounding areas, which include small towns, villages and farming communities. Avon has the advantage of being a defined geographic area with services centralised around Bristol. It has a relatively low level of outward migration and a mixture of different social backgrounds, housing types and urban and rural areas. The study area is well-defined, consisting of that part of the county of Avon that was also within the South West Regional Health Authority. It therefore excludes Bath and district.

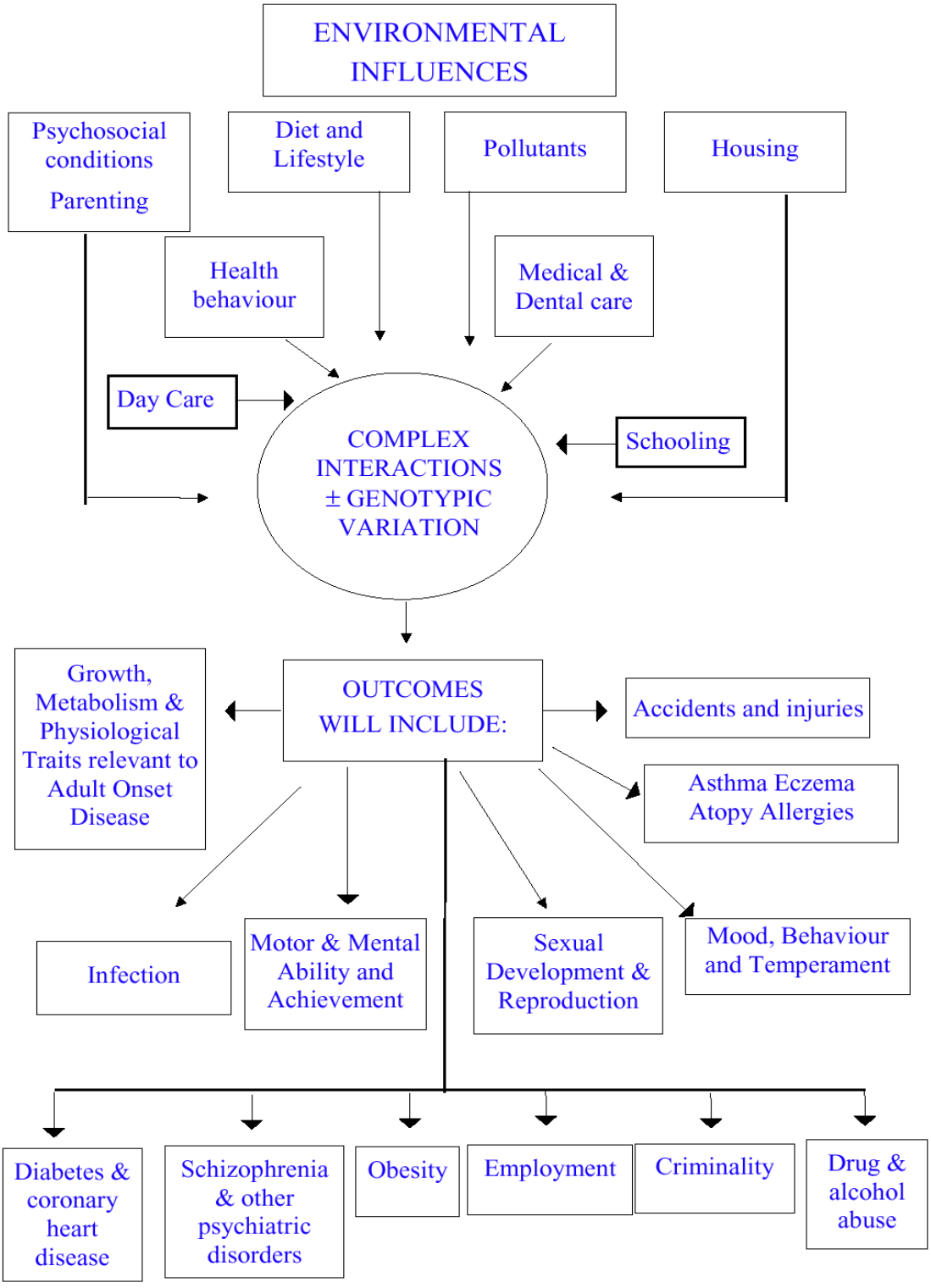


Figure 1: Diagramatic representation of the study aims and design

Figure 2: Location of Avon within Great Britain

As Eskinsmyth pointed out in her paper on the use of the major cohort studies to geographers,[3] the sample members are spread quite thickly across the county of Avon, and as a result, the data provide a rich resource for social and geographic analysis at a smaller spatial scale than the nationally based studies. ALSPAC has a spatially intensive design and can therefore contribute more widely to the debate on area and locality effects. It enables the researcher therefore, to examine the spatial distribution of a huge range of social, economic, medical, political and cultural phenomena, and provide a vast resource for straightforward geographical mapping of incidence.

I.3 Study Eligibility Criteria

To be eligible for the study, mothers had to be resident in Avon while pregnant. In addition, their expected date of delivery had to lie between 1st April 1991 and 31st December 1992 inclusive. Mothers who were resident in the area but left shortly after enrolment were omitted from further follow-up. However, those who had completed the questionnaire scheduled for the third trimester of pregnancy before leaving the study area, have been kept in the study, even if they had not delivered at the time of moving.

I.4 Enrolment strategy

Work prior to the start of enrolment in September 1990 had involved meetings with midwives and discussion with groups representing general practitioners as well as detailed discussion with obstetricians in the area.

Posters were printed for display in a variety of different places - including chemist shops, libraries, mother and toddler groups, and pre-school playgroups, general practitioner waiting-rooms, antenatal clinics and any other area where a mother in early pregnancy was likely to be. In addition, there was considerable local and national coverage in the press, radio and television.

The poster displayed the logo of the study ‘Children of the Nineties’ and asked interested pregnant mothers to get in touch with the study team. In addition, the local community midwives when interviewing the mother for the first time discussed the study with her, and gave her a card with which to send for further details.

The card that the mother completed and sent to the study office contained her full name and date of birth, her address, her last menstrual period, and expected date of delivery.

Once the card had been received at the study office, a brochure was sent to the mother. This outlined the reason for carrying out the study and explained that the mothers themselves would not benefit tangibly, but that the major benefits were likely to be for the next generation. It informed the mother that there was no compulsion for her to take part, and that even if she started within the study she was free to opt out at any point. Thirdly it emphasised the confidential nature of the information that would be collected, and promised that at no time would the names of the mother and/or child be linked to the confidential information collected. Fourthly, it explained that biological samples would be taken, but that these would not be analysed without the signed permission of the mother, and finally it stated that the information given would also be linked to information from the medical records unless the mother let us know that she did not want us to do this. The mother was told in this brochure that we would assume that she wanted to take part in the study unless she informed us otherwise.

A telephone number (the Children of the Nineties hotline) was given for parents to ring. This hotline was manned by volunteers who had been instructed in the confidential nature of the study. They are advised not to do any counselling or persuading, but rather to take messages which are then acted upon by appropriate members of the study team. In instances where parents ring to request help, they are given, if possible, the appropriate telephone number of an organisation set up to perform this type of service.

The distribution of gestations at enrolment is shown in Figure 3 for the 14105 mothers who enrolled in pregnancy. There were an additional 430 who enrolled after delivery, almost all within the first two days.

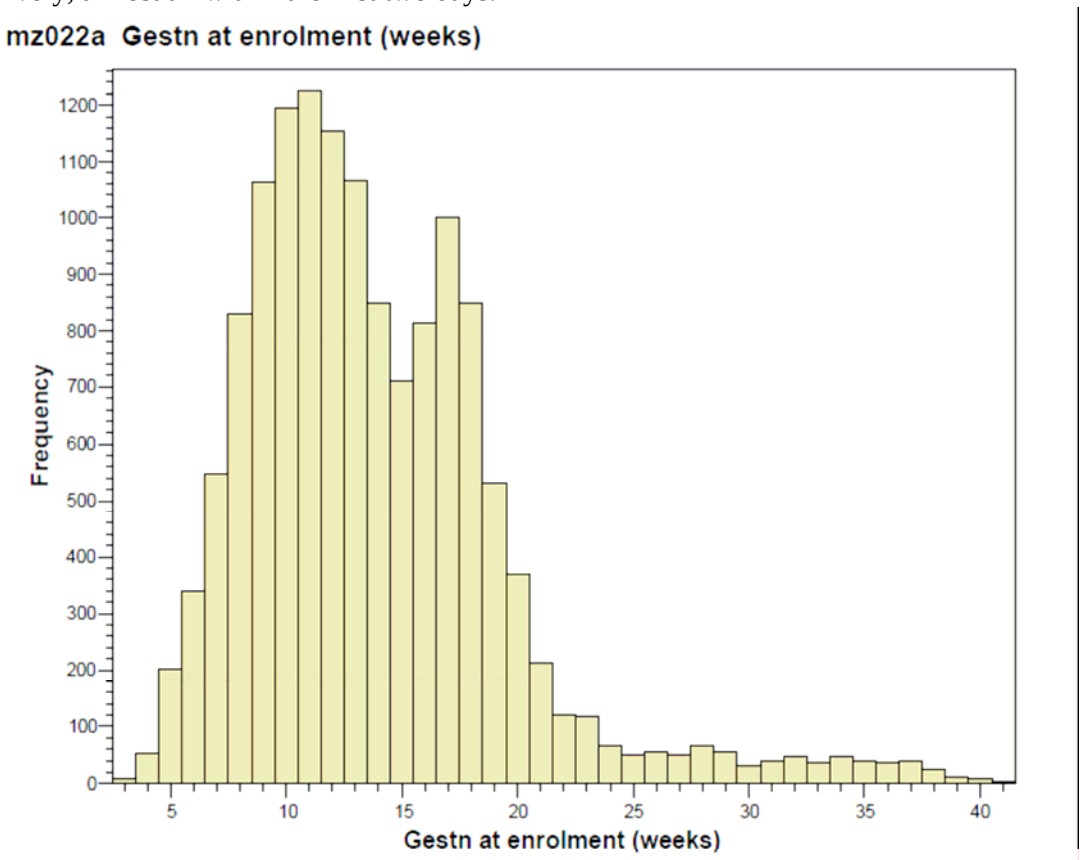


Figure 3

I.5 Data collection strategy

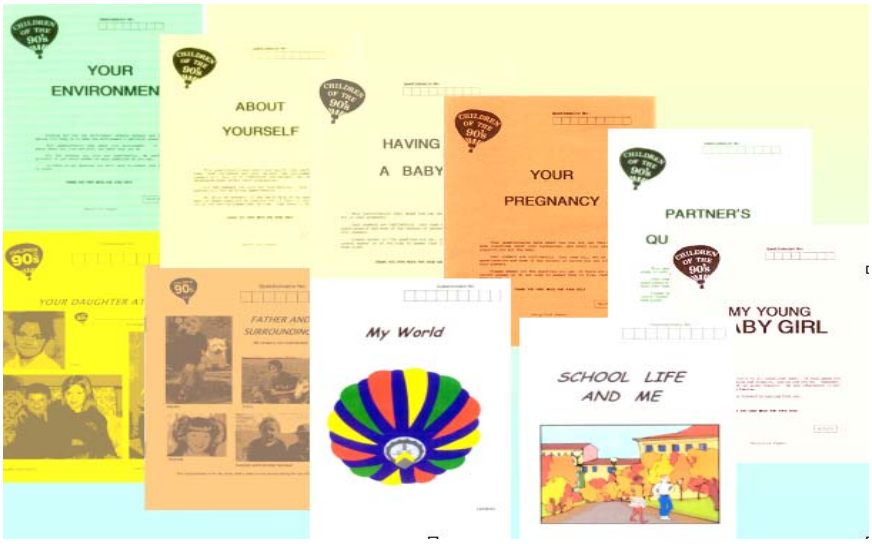
The study was designed to collect and link data together from a variety of sources, provided they fitted within the overall aims of the study, and were within its ethical constraints.[4,5]

These include:

* Self-completion questionnaires completed by the study mother
* Self-completion questionnaires completed by the mother's partner
* Assays of biological samples, including genetics
* Medical records
* Educational records
* Information from teachers and head-teachers
* Self-completion questionnaires completed by the study offspring
* Hands on assessments

For the data sets I to VI currently deposited with the UK data archive we use data from the questionnaires from the mothers and partners and for data set V we also included data from the hands-on assessments.

II DATA FROM QUESTIONNAIRES



The actual questionnaires used are available online at www.bristol.ac.uk/alpsac/sci-com/quests. It should be noted that these questionnaires do not show the boxes that were to be ticked, or the pictures included in the questionnaires.

II.1 Questionnaire Administration during Pregnancy

Approximately 7 days after the introductory brochure had been sent out (see I.4) and provided the study had not heard from the mother that she did not want to take part in the study, the first questionnaire was posted to her.

The questionnaires for the first 12-13 years of the study were designed for the participant to tick boxes as appropriate, and write text answers occasionally. Responses were

Figure 4: Gestation at which the different questionnaires were administered, depending on gestation at enrolment.

Gestation at enrolment	YE	AY	HAB	YP	YHL
<6	<6	14	18	32	-
6	6	14	18	32	-
7	7	14	18	32	-
8	8	14	18	32	-
9	9	14	18	32	-
10	10	14	18	32	-
11	11	23	18	32	-
12	12	23	18	32	-
13	13	23	18	32	-
14	14	23	18	32	-
15-18	22	26	18	32	-
19	24	28	19	32	-
20	24	28	20	32	-
21	24	28	21	32	-
22	28	36	22	32	-
23	28	36	23	32	-
24	-	33	-	29	24
25	-	34	-	30	25
26	-	35	-	31	26
27	-	36	-	32	27
28	-	36	-	32	28
29	-	36	-	32	29
30	-	37	-	33	30
31	-	X	-	31	34
32	-	X	-	32	35
33	-	X	-	33	36
34	-	X	-	34	37
35	-	X	-	35	37
36	-	X	-	36	38
37	-	X	-	37	39
38	-	X	-	38	40
39	-	X	-	39	41
40	-	X	-	40	41

X = administered 4 months after delivery; YE = Your Environment; AY= About Yourself; HAB= Having a Baby; YP= Your Pregnancy; YHL= Your Home and Lifestyle; AY was accompanied by 'You and Your Environment', HAB or YHL by 'Partner's Questionnaire'.

keyed (double keyed) by a commercial bureau. Later questionnaires were designed to be scanned.

The nature of the questionnaire depended on the woman's gestation at enrolment. The scheme for the sending of the questionnaires is illustrated in Figure 4. In brief, there were 4 questionnaires administered to the mother during pregnancy, and these were designed so that two were to be sent out at a fixed time point - 'Having a Baby' at 18 weeks gestation and 'Your Pregnancy' at 32 weeks gestation.

Provided she enrolled before 14 weeks gestation, then the questionnaire 'Your Environment' was sent to her immediately after enrolment. This questionnaire was designed in particular to identify those features of the early environment that might be responsible for effects on the fetus. The fourth questionnaire 'About Yourself' was mainly concerned with features that referred to the mother's past medical, social and environmental history, and consequently the time during pregnancy at which this was administered was relatively unimportant. If necessary, therefore, this questionnaire was sent out after the baby had been born.

For questionnaires sent during pregnancy the reminder and follow-up phase was fairly intensive. If a response had not been received within 7 days, a reminder letter was sent. If the questionnaire had still not been received after a further 10 days, a second reminder letter was sent. Finally, if no response had been received after 1 month, a member of the study team either rang the mother or visited the home, and encouraged, or helped, the mother to complete the questionnaire.

For mothers who did not enrol until long after the 18-week contact, the 'Having a Baby' questionnaire was not likely to be valid. Much of the detail in that questionnaire was concerned with attitudes, activities and emotional wellbeing at that particular point during pregnancy. Nevertheless, for those mothers who had enrolled late, there was a certain amount of information concerning the environment and lifestyle that could and should validly be collected. For these women, therefore, the appropriate information, otherwise obtained from the questionnaires 'Your Environment' and 'Having a Baby', was combined into a single questionnaire 'Your Home & Lifestyle'.

In spite of piloting, early assessments of the first questionnaire responses revealed a number of ambiguities. Later versions were printed to correct these problems. Where this has resulted in slight changes in questions, these have been documented.

The data files derived from the questionnaires depicted in Figure 4 were as follows:

* A file – YE and for later enrollers YHL
* B file – HAB and for late enrollers YHL
* C file - YP
* D file - AY

II.2 Involvement of the mother's partner

Questionnaires were also designed for the woman's current partner. The study did not, however, have the name of the mother's partner and indeed did not want to involve the partner without the mother's permission. Therefore the following procedure was invoked: Questionnaires for partners were timed to coincide with a questionnaire for the mother. Both were sent together to the mother, accompanied by a letter which explained that a questionnaire for her partner was enclosed; if she would like him/her to be involved, then she was asked to hand it to him, but if she did not want him to be involved that was fine. She was told the decision was hers. A definition of partner (i.e. male or female) was not stipulated, it was left to the mother to make such a decision. A letter was enclosed with each partner's questionnaire and also a pre-paid envelope, so that his responses could remain confidential if he so wished.

If a woman wrote and informed the study that she had no partner, or that he was not interested, then we did not include questionnaires for him from then on. Later, in the postnatal phase, the mother was informed that although the words he/him were used throughout, responses were welcomed from all partners.

During pregnancy, two questionnaires for partners were sent, 'Partner's Questionnaire' accompanied the mother's 'Having a Baby' at 18-23 weeks; and 'You and Your Environment' accompanied the mother's 'About Yourself'. Data from these questionnaires were produced to form the following files:

* PA file – You and Your Environment
* PB file – Partner's Questionnaire

II.3 Response rates to pregnancy questionnaires

One of the major issues to be considered in a cohort where a large proportion of the pregnant population have enrolled early in pregnancy is the loss due to miscarriage, termination, stillbirth or infant death. Once such an outcome had been notified, no further questionnaires (or reminders) were sent, and the ALSPAC Director wrote a personal note of condolence.

The way in which the response rate among the 14541 women who had enrolled, was affected by fetal and infant loss in regard to questionnaires in pregnancy is shown in Table II.3, together with the response rates.

Table II.3 Response rates to questionnaires resulting in the A, B, C and D files

Status of study pregnancy	A File	B File	C File	D File
Fetal or infant death	95	395	585	424
Already refused	19	15	119	56
Questionnaire sent	14392	14086	13741	13947
Questionnaire returned	13548	13194	12421	12452
Response rate	94.1%	93.7%	90.4%	89.3%

[N.B. Missing from this table are the group of women who, at the time a questionnaire was due had been lost to follow up.]

All women who had enrolled, regardless of outcome, are included in ALSPAC data files.

II.4 Questionnaire administration to the study parents once the child was born

It is statutory for births to residents of the area to be notified to the health authorities, regardless of where they were born. The study was able to use this system to identify births (and deaths) as they occurred. Information was also obtained from local neonatal intensive care units to identify babies who were seriously ill. For all livebirths with reasonable prognosis a congratulations card was sent. The date of delivery then was used to prompt the dates on which all subsequent questionnaires have been sent (Figure 5).

Figure 5: Timing of first 10 years of questionnaire administration post delivery. Those concerning the child were sent to the mother or chief carer; those for the partner were sent to the study mother to pass on

Age of child	Name of questionnaire	File	Focus of questionnaire
1m	My Young Baby Boy/Girl	KA	Child
2m	Me and My Baby	E	Mother
	Being a Father	PC	Partner
6m	My Baby Boy/Girl	KB	Child
8m	Looking after the Baby	F	Mother
	The Baby and Me	PD	Partner
15m	My Infant Son/Daughter	KC	Child
18m	Girl/Boy Toddler	KD	Child
21m	Caring for a Toddler	G	Mother
	A Toddler in the House	PE	Partner
24m	My Little Girl/Boy	KE	Child
30m	My Study Son/Daughter	KF	Child
33m	Your Health, Events and Feelings	H	Mother
	Partner's Health, Events and Feelings	PF	Partner
38m	My 3 Year Old Boy/Girl	KG	Child
42m	My Son's/Daughter's Health and Behaviour	KJ	Child
47m	Mother's New Questionnaire	J	Mother + some Child
	Partner's New Questionnaire	PG	Partner
54m	My Young 4 Year Old Boy/Girl	KK	Child
57m	Development and Health of My Son/Daughter	KL	Child
61m	Study Mother's Questionnaire	K	Mother
	Study Partner's Questionnaire	PH	Partner
65m	My Five Year Old Son/Daughter	KM	Child
69m	My School Boy/Girl	KN	Child
73m	Mother's Lifestyle	L	Mother
	Partner's Lifestyle	PJ	Partner
77m	My Daughter/Son Growing Up	KP	Child
81m	My Son/Daughter at School	KQ	Child
85m	Mother and Home	M	Mother
	Partner and Home	PK	Partner
91m	My Son/Daughter's Well-being	KR	Child
97m	Mother and Family	N	Mother
	Father and Family	PL	Partner
103m	My Son/Daughter's Health	KS	Child
	My Son/Daughter at School	KT	Child
110m	Mother of a 9 year old	P	Mother
	Father of a 9 year old	PM	Partner
115m	Your Son/Daughter at 9	KU	Child
122m	You and Your Surroundings	Q	Mother
	Father and Surroundings	PN	Partner
128m	My Son/Daughter's Health and Happiness	KV	Child

II.5 Coding of questionnaires

Most of the self-completion responses are self-coding - the ticked reply box contained a printed number that can be directly keyed. A few questions invite a textual reply, and some participants also amplify a tick response with comments.

Returned questionnaires were coded by ALSPAC staff. They checked that each question had no more than one ticked response, and that any comments did not materially affect the sense of the response. On a few occasions they also need to convert dates and similar variables to a standard format. There were rules for each variable, for how to interpret problems such as multiple ticking, or rounding of ages, e.g. where months are given and years were requested. All coding was cross-checked by a second person and then keyed and verified.

In general, where more than one box was ticked there was a rule that the 'worst' code would be used. Where no such rules could be made logically the coding supervisor made a decision.

Textual replies to questions have been dealt with separately. The range of responses to a given question is enormous, the variety of questions asked is also large, and this gives problems of maintaining coding consistency across a range of specialist areas, e.g. drugs, accidents, occupations, and environmental exposures. The problem has been resolved by keying all written responses, splitting the responses by question type, so that finally all the replies to one subject over time are available together. These can then be coded semi-automatically by a specialist in that field. Accuracy and consistency can thus be ensured without the expense of training coders in all the different disciplines required.

III. HANDS-ON ASSESSMENTS

III.1 Children in Focus sub sample

A 10% sample of children was identified from the ALSPAC births that occurred from June 1st to the end of December 1992. A standard number of children per week was chosen using a pseudo-random process from the list of births within the study. The aim of the study was to examine these children in a way that could not be done using questionnaires prior to the age of 7 when ALSPAC clinics for the whole cohort began. The sample provides both validation for certain aspects of the self-completion questionnaires, and answers several important questions in regard to outcomes that cannot be identified easily from records or questionnaires. These are related to the prevalence and aetiology of childhood growth, anaemia, otitis media with effusion, visual defects, the development of intellectual competence, speech and language, as well as motor development. The various items assessed at each time point are indicated in Table III.1.

Table III.1 Assessments undertaken at each time point: Children in Focus

	Age of child (months)									
	4	**8**	**12**	**18**	**25**	**31**	**37**	**43**	**49**	**61**
Weight	+	+	+	+	+	+	+	+	+	+
Height/length	+	+	+	+	+	+	+	+	+	+
Head circumference	+	+	+	+	+	+	+	+	+	+
Arm circumference	+	+	+	+	+	+	+	+	+	+
Waist circumference							+	+	+	+
Leg length									+	+
Skinfold thickness										+
Blood pressure							+		+	+
Vision examination	+	+	+	+	+	+	+	+		
Tympanometry		+	+	+	+	+	+	+	+	+
Hearing						+		+		+
Speech/language					+					+
Dental inspection						+		+		+
Skin examination									+	+
Cognitive function	+			+					+	
Laterality							+			
Short-term memory										+
Diet	+	+		+	+		+	+		+
Parenting observation			+							+
Day care arrangements				+	+	+	+	+	+	+
Fitness										+
Lung function										+
Allergy tests										+
Blood samples		+	+	+		+		+		+

Further details of clinic organisation is given the documentation of Group V.

III.2 Hands-on assessments of the whole sample

At birth

Although preliminary studies indicated that the hospital scales were accurate for measuring birthweight, the techniques being used by the midwives to measure birth length and head circumference were grossly inaccurate. ALSPAC therefore arranged for trained study staff to visit the two major maternity hospitals (responsible for 94% of the study births) and measure the length and head circumferences of all newborn babies provided the mothers gave permission for this to occur.

From 7 years

All 14000 children in the study were invited to a specially designed clinic where they were examined. At age 7.5 years the assessments included anthropometric data (weight, height, head, arm, chest, waist and hip circumferences, cubit width, leg length and assessment of fat using bioelectrical impedance); examination of the skin for eczema; blood pressure and pulse rates; allergy skin-prick tests; vision tests; tympanometry and audiometry; motor co-

ordination; tests of reading, spelling and phoneme awareness. A blood sample was taken and a dietary diary collected.

At age 8.5 years these children underwent tests of lung function and bronchial hyper-responsiveness; cognitive ability (IQ); speech and language; attention span; non-verbal accuracy; short-term memory as well as height and weight. In addition, interviews elicited details of bullying, antisocial activities, self-esteem and gender play.

The annual examinations continued until approximately age 13 and thereafter at about 2-year intervals.

Further details of measures taken at each age are given in:

www.bristol.ac.uk/alspac/documents/focus-clinic-sessions.pdf

IV. CORE DATA FOR UK DATA ARCHIVE GROUPS

IV.1 Format of documentation of questionnaire responses

In all the documentation of questionnaires, the following system is used:

i) Each question is quoted verbatim in italics (with differences between the questionnaire versions, where they exist).
ii) The letter denoting the questionnaire, together with the number of the question.
iii) The coding rule(s) used by the ALSPAC coders are indicated in square brackets;
iv) The editing assumptions made preparing the computer file in round brackets;
v) The variable no.
vi) A table of frequencies.

Due to extension of direct access to ALSPAC data to non-ALSPAC staff and in order to comply with guidance issued in 1996 by the ALSPAC Law & Ethics Committee regarding the confidentiality of multiple pregnancies records for triplet and quadruplet pregnancies have been put to missing.

The total sample size involves 14541 mothers although, as already noted, this includes women who dropped out of the study because their fetus/child did not survive.

IV.2 The core data

6 variables have been included in each of the 6 data sets deposited with the UK Data Archive.

(a) Parents' occupation – current or last main job as measured at 32 weeks gestation.

The questions asked of the pregnant mother at 32 weeks gestation were as follows:

The questions below ask about your current occupation and that of your partner.

As far as you can, please describe the actual job, occupation, trade or profession. (Use precise terms such as radio mechanic, woodworking machinist, tool room foreman. If the occupation is known by a special name, please use that name. If in H.M. Forces, give the rank in addition to the actual job. Please also describe the type of industry or service given: i.e. gives of what is made, materials used, or services given.)

Your present job or last main job.
Actual job, occupation, trade or profession [keyed as text]

Please tick which of the following apply to you:

Foreman
Manager
Supervisor
Leading hand
Self-employed
None of these

[when 2 boxes were ticked, the one with the lower code was used.]

The occupations were coded and classified as indicated in Appendix A. The frequencies are as shown in Table VI.2a.

Table VI.2a NS-SEC analytic class of study mother and her partner

Analytic class	Study mother (c756a)	Mother's partner (c766a)
Higher managerial and professional	796	1961
Lower managerial and professional	3260	2514
Intermediate occupations	2629	678
Small employers and own account workers	385	1494
Lower supervisory and technical	574	1949
Semi-routine occupations	1402	766
Routine occupations	748	1134
TOTAL VALID	9794	10496

(b) Ethnic orgins

Mother was asked to describe the ethnic origin of herself, her partner and her parents:

How would you describe the race or ethnic group of yourself, your partner and your parents?

(i)	*(ii)*	*(iii)*	*(iv)*
Yourself	*Partner*	*Your mother**	*Your father**

(by this we mean the mother and father figure who was mostly responsible for bringing you up)*

white
black/Caribbean
black/African
black/other (please describe below)
Indian
Pakistani
Bangladeshi
Chinese
Any other ethnic group (please describe)

[If more than 1 box was ticked, other was coded and the combinations written; these were reviewed and the categories changed as appropriate]

The child's ethnic background was derived as non-white if either parent had codes in the range 2-9. Further details are given in Group IV.

In all, using the categorisation into white/non-white, the non-white prevalence was 2.5% (312/12355) among mothers (var c800a) and 4.8% (577/12090) among study children (var c804).

(c) Maternal age

The age of the mother at the birth of the child was calculated from the mother's date of birth (obtained at enrolment) and the date of delivery (obtained from the child health computer). It is given in completed years but curtailed at either end for confidentiality reasons (<18=17; >40 = 40) (MZ028b).

For ease of use, var MZ028f provides the ages in 5 year age bands.

age	n
<20	655
20-24	2682
25-29	5369
30-34	3808
35-39	1213
40+	169
TOTAL	13896

It should be noted that a date of birth is essential for this calculation. Consequently no age is available for the women who miscarried or for whom the outcome of pregnancy is unknown. Thus for analyses including such outcomes, age at LMP is the preferred measure.

References

1. Golding J, ALSPAC Study Team. The Avon Longitudinal Study of Parents and Children (ALSPAC) – study design and collaborative opportunities. European Journal of Endocrinology 2004; 151: U119-U123.

2 Golding J, Pembrey M, Jones R, ALSPAC Study Team. ALSPAC – the Avon Longitudinal Study of Parents and Children I. Study methodology. Paediatric and Perinatal Epidemiology 2001; 15: 74-87.

3. Ekinsmyth C. Large-scale longitudinal studies; their utility for geographic enquiry. Area 1996; 28: 358-372.

4. Mumford SE. Children of the 90s: ethical guidance for a longitudinal study. Archives of Disease in Childhood Fetal and Neonatal Edition 1999; 81: F146-F151.

5. Mumford SE. Children of the 90s II: challenges for the ethics and law committee. Archives of Disease in Childhood Fetal and Neonatal Edition 1999; 81: F228-F231.

N.B. Details of all ALSPAC publications are available on the ALSPAC website:
www.bristol.ac.uk/alspac/sci-com/pubs/

Appendix A

Deriving NS-SEC

NS-SEC (National Statistics Socio-Economic Classification) is an eight class occupation based classification designed to measure employment relations and conditions of occupations. These allow delineation of the structure of socio-economic position and in turn help explain variation in social behaviour (Source:ONS). NS-SEC is currently based on the Standard Occupational Classification published in June 2000 (SOC2000).

There are three possible ways of deriving the NS-SEC categories and these depend on the level of information recorded. The 'Full' method requires the SOC2000 unit group, employment status and size of organisation. The 'Reduced' method uses SOC2000 unit group and employment status and finally, the 'Simple' method uses just the SOC2000 unit group. Due to ALSPAC questionnaires not asking about organisation size it was not possible to derive NS-SEC using the 'Full' method so the 'Reduced' method was used wherever possible instead. If it was not possible to derive NS-SEC using the 'reduced' method due to lack of information from the questionnaires the 'Simple' method was used.

For a number of files the information as to whether the respondent was self employed and employed others or not was missing as the question was not asked in the relevant questionnaire. However, it was still possible to derive NS-SEC using the reduced method.

The following table indicates which variables used which method to derive NS-SEC

File	Variable	Method
A	D3, C8	Simplified
B	E2a, C4a	Simplified
C	H4a, H4b	Reduced
C	H6a	Reduced (Missing info)
C	H5a	Simplified
F	G7b, H5b	Simplified
G	H6b, L3c, L3b	Simplified
H	H6b, J3c	Simplified
J	C45b, E5b	Simplified
PA	B2	Simplified
PB	F3	Reduced
PB	F5a	Reduced (Missing info)
PB	F4a	Simplified
PD	F6	Reduced (Missing info)
PE	G3	Reduced (Missing info)

Table 1. Derivation method used to calculate NS-SEC

The steps for deriving NS-SEC are as follows:

1) Text from the questions asking the respondent about their occupation was imported into the CASCOT (Computer Assisted Structured Coding Tool) package recommended and supplied by Peter Elias. This allowed the text to be passed through several processes both automatic and with user intervention to code it to the SOC2000 classification and produce a file of data that could be imported into SPSS.
2) Wherever possible the employment status of each individual was derived using a set of matrices available from the ONS converted to SPSS system file format and matched to the data output from CASCOT containing the SOC2000 codes. For those cases with missing information but where inferences could be made based on the text

answers the employment status variable was derived using the flow chart in Figure 1 below.

3) Once the employment status was derived the managerial/supervisory status of respondents was amended depending on answers provided within the text.
4) The appropriate NS-SEC category was then calculated using the employment status code and SOC2000 code using a further set of matrices and SPSS syntax..

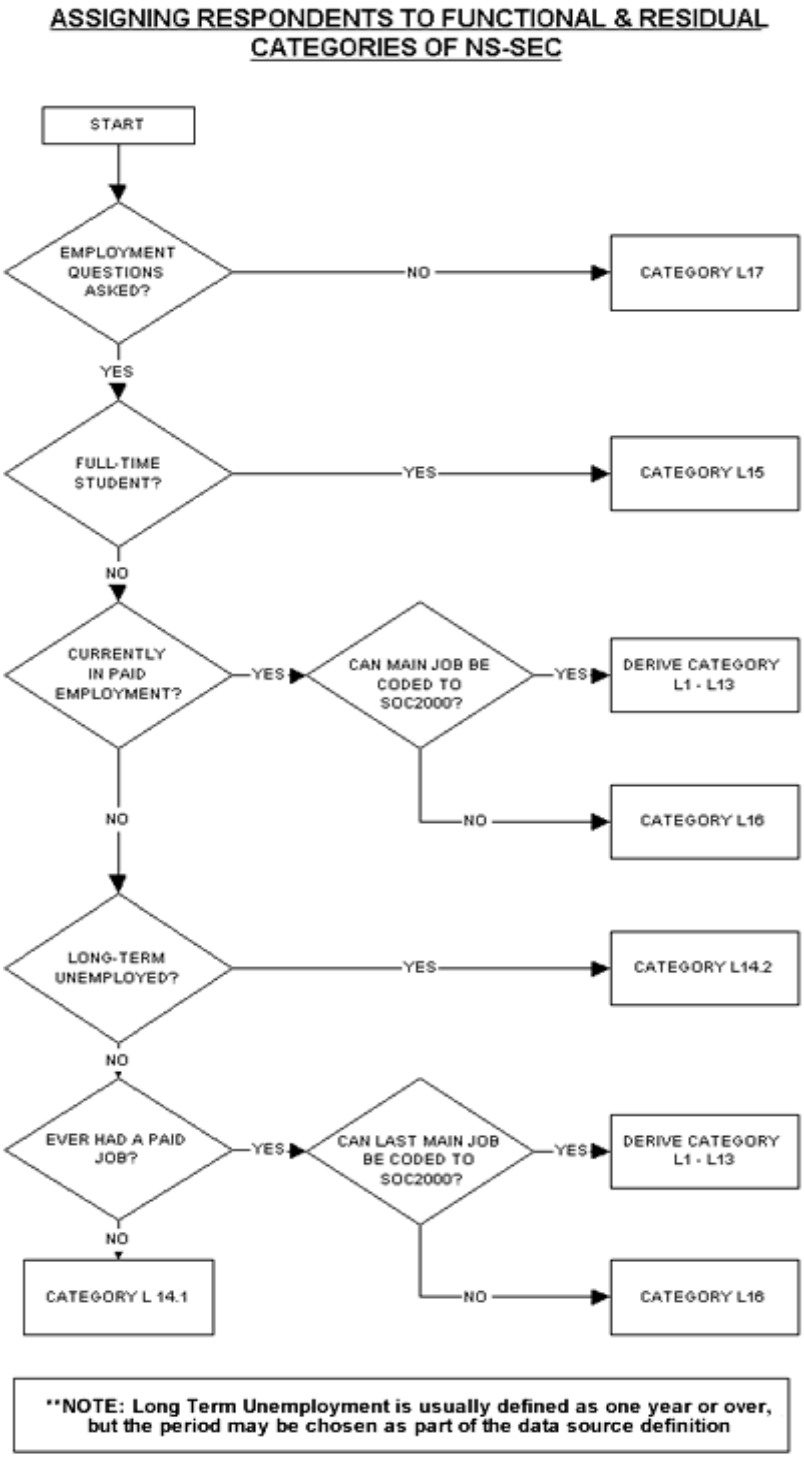


Fig.1 Employment Status Flow Chart (Source:ONS)