

# National Child Development Study: Biomedical Survey, 2002-2004

User guide (version 2)

October 2020

CENTRE FOR  
LONGITUDINAL  
STUDIES



Economic  
and Social  
Research Council

## Contact

For data queries please contact: [help@ukdataservice.ac.uk](mailto:help@ukdataservice.ac.uk)

## How to cite this guide

Centre for Longitudinal Studies. (2020) *National Child Development Study: Biomedical Survey, 2002-2004*. London: UCL Centre for Longitudinal Studies.

This guide was published by the UCL Centre for Longitudinal Studies. The 1st version was published in 2008 by Jane Elliott, Jon Johnson, Peter Shepherd et al. The 2nd version published in 2020 by Sarah Kerry-Barnard and Aida Sanchez.

UCL Social Research Institute

University College London

20 Bedford Way

London WC1H 0AL

[www.cls.ucl.ac.uk](http://www.cls.ucl.ac.uk)

The UCL Centre for Longitudinal Studies (CLS) is an Economic and Social Research Council (ESRC) Resource Centre based at the UCL Social Research Institute, University College London. It manages four internationally-renowned cohort studies: the 1958 National Child Development Study, the 1970 British Cohort Study, Next Steps, and the Millennium Cohort Study. For more information, visit [www.cls.ucl.ac.uk](http://www.cls.ucl.ac.uk).

This document is available in alternative formats. Please contact the Centre for Longitudinal Studies.

tel: +44 (0)20 7612 6875

email: [clsfeedback@ucl.ac.uk](mailto:clsfeedback@ucl.ac.uk)

# Acknowledgements

## Principal Investigators

Professor Christine Power and Professor David Strachan

Professor John Bynner, late of the Centre for Longitudinal Studies

Gillian Prior, late of the National Centre for Social Research.

## Survey Development

Peter Shepherd (CLS), Elizabeth Fuller (NatCen).

## Collaborators

Professor Adrian Davis (MRC Hearing and Communication Group, University of Manchester);

Dr Meena Kumari (Department of Epidemiology and Public Health, UCL);

Professor Gary Macfarlane (Department of Public Health, University of Aberdeen);

Dr Jugnoo Rahi (Centre for Paediatric Epidemiology and Biostatistics, UCL Institute of Child Health);

Dr Bryan Rodgers (National Centre for Epidemiology and Population Health, the Australian National University, Canberra);

Professor Stephen Stansfeld (Centre for Psychiatry, Queen Mary's School of Medicine and Dentistry, London).

## The survey doctor

Dr Melanie d'Souza

## Centre for Longitudinal Studies:

Kevin Dodwell and members of the CLS tracing team; Jon Johnson.

## National Centre for Social Research

Nafiis Boodhumeah and other programmers;

Sandra Laver and the Pink team in the Operations Department;

Christine Massett and the Telephone Unit interviewers;

Richard Boreham and Carli Lessof;

NatCen's nurses, who contributed to the survey's development and were responsible for collecting the data.

**St George's Hospital Medical School**

Florence Bunting, Dahlia Cowan and Alan Gibbons

**ALSPAC laboratory, Bristol**

Susan Ring, Wendy McArdle, Patricia Bevan and Fay Stratton at the

**Royal Victoria Infirmary, Newcastle**

Professor Ian Gibb, Gavin Spickett and other staff of the clinical biochemistry laboratory

**Royal Glasgow Infirmary**

Gordon Lowe and Ann Rumle

**Biological Psychology, Department of Psychology, University of Dresden**

Professor Clemens Kirschbaum

## Funding

The NCDS Biomedical Survey was funded under the Medical Research Council 'Health of the Public' initiative, and was carried out in collaboration with the Institute of Child Health, St George's Hospital Medical School, and the National Centre for Social Research. We would also like to acknowledge the long-standing support of the Economic and Social Research Council on previous sweeps of the NCDS.

# Contents

<b>About the National Child Development Study .....</b>	<b>1</b>
<b>1. Introduction.....</b>	<b>2</b>
1.1 NCDS Surveys.....	2
1.2 Introduction to the second edition of the 2002-2004 biomedical survey.....	3
<b>2. Development of the Biomedical Survey and Fieldwork .....</b>	<b>5</b>
2.1 Target Sample .....	5
2.2 Nurse briefings.....	5
2.3 Issue of work.....	6
<b>3. Response .....</b>	<b>7</b>
<b>4. The interview.....</b>	<b>9</b>
4.1 CAPI main interview.....	9
4.2 Paper self-completion questionnaires .....	10
4.3 Consents.....	11
4.4 Saliva sample.....	11
<b>5. Biochemical analyses of blood samples .....</b>	<b>12</b>
<b>6. Research data .....</b>	<b>14</b>
6.1 Structure of the data .....	14
6.2 End User Licence Dataset .....	14
6.3 Special Licence Dataset.....	16
6.4 Recoding of disclosive information for availability under End User Licence....	17
6.5 Data Coding and Editing .....	18
6.6 Coding of medication variables .....	22
6.7 Derived Variables.....	23
<b>Appendix 1: D-Dimer Edits .....</b>	<b>24</b>
<b>Appendix 2: Deriving Childhood Adversity .....</b>	<b>27</b>

# About the National Child Development Study

The National Child Development Study (NCDS) is a longitudinal birth cohort study, following a nationally representative sample of over 17,000 people born in Britain in a single week in March 1958.

Cohort members have been surveyed throughout their lives, since birth, creating an incredibly rich resource for a wide range of research. The study data show the very long roots of childhood, how past experiences can reverberate through the years, and the interplay between the different facets of people's lives.

NCDS has equipped policymakers with robust evidence in areas as diverse as smoking in pregnancy, educational inequalities, adult basic skills, and social mobility. Today, with the cohort now in their sixties, the study is casting light on how people experience retirement and ageing in the 21st century.

# 1. Introduction

## 1.1 NCDS Surveys

The National Child Development Study (NCDS) started life as the Perinatal Mortality Survey, which was designed to examine the social and obstetric factors associated with stillbirth and infant mortality. In the first survey, data were collected about the births and families of just over 17,000 babies born in Great Britain during one week in March 1958.

NCDS is a continuing, multi-disciplinary longitudinal study. It began as the Perinatal Mortality Survey, which was designed to examine the social and obstetric factors associated with stillbirth and infant mortality. In this first survey, data were collected about the births and families of just over 17,000 babies born in Great Britain (England, Scotland and Wales) in one week in 1958, all of whom were eligible for comprehensive follow-up.

There have been nine surveys gathering information from respondents living in England, Scotland and Wales, in order to monitor their health, education, social and economic circumstances. These surveys were carried out in 1965 (age seven), 1969 (age eleven), 1974 (age sixteen), 1981 (age 23), 1991 (age 33), 1999/2000 (age 42), 2002-2004 (44) 2004/2005 (age 46), 2008/2009 (age 50) and 2013 (age 55). The 2002-2004 follow-up was a biomedical survey conducted by nurses. As part of the 1991 survey, information was additionally collected on the children of one in three cohort members; this included assessments of the NCDS 2013 Follow-Up – User Guide 8 behaviour and cognitive development of around 5,000 co-resident children. There have also been surveys of sub-samples of the cohort: most recently, in 1995 (age 37) a 10% representative subsample was assessed for difficulties with basic skills.

Data for NCDS have so far been collected from a number of different sources; the midwife present at birth, parents, schools, health service personnel, the cohort members themselves, their spouses, cohabitantes and children. Data has been collected using a variety of methods: paper, electronic and self-completion questionnaires, clinical records, medical examinations, physical measurements,

ability tests, educational assessments and diaries. Most of the previous follow-ups have been carried out face-to face visits to cohort members in their homes, but the age 46 survey in 2004 was conducted by telephone. The Centre for Longitudinal Studies (CLS) at the UCL Institute of Education (and formerly the Social Statistics Research Unit at City University), has been responsible for the study since 1985. CLS has been funded as an ESRC resource centre since 2004.

## 1.2 Introduction to the second edition of the 2002-2004

### biomedical survey

The NCDS 2002-2004 biomedical survey was designed to obtain objective measures of ill-health and biomedical risk factors in order to address a wide range of specific hypotheses relating to anthropometry; cardiovascular, respiratory and allergic diseases; visual and hearing impairment; and mental ill-health.

This guide should be read in conjunction with the Technical Report produced by NatCen, which details the development of the survey, the conduct of the fieldwork and subsequent coding and processing of the biological samples.

This second version of the NCDS biomedical user guide contains information related to the NCDS Biomedical data available under End User Licence and the Special Licence. This is the result of the splitting of the original dataset, previously available under the UKDS Special Licence, into two datasets with different levels of sensitivity and disclosivity, which can therefore be made available under two different types of data access licences.

The majority of the NCDS 2002-4 biomedical data are now available under End User Licence (1079 variables), with the original data considered sensitive is still available under Special User Licence. This decision was the result of the Centre for Longitudinal Studies' (CLS) disclosure assessment of each variable and the broad aim to make as much data available with the lowest possible barriers.

The Special Licence dataset contains 123 variables including questions on child adversity that were collected between 2002 and 2004, but not released in the previous NCDS42-4 Biomedical dataset. The Special Licence dataset also contains specific sensitive topics including: child adversity, CIS-R specific questions on mental health and questions which contains categories with small frequencies



related to personal details such as skin colour, pregnancy, a surgical operation, specific height and unusual high number of children.

Information about the medication taken by the cohort members of the study is also deposit under End User Licence for the first time. These data were collected in 2002-4, but it was never released via the UKDS.

The number of cases in these datasets has changed from 9377 to 9293, as participants who did not provide consent to deposit their responses at the UK Data Service were excluded from the final dataset.

## 2. Development of the Biomedical Survey and Fieldwork

The survey content was proposed in detail in the MRC grant application. Practical development began in October 2001. Between then and the first pilot in March 2002 the project management team and the collaborators met on a regular basis. The content and order of the interview was determined, including the time allocated to each module, the exact measurements and questions to be included, the equipment that would be used, nurse protocols, and the way information would be recorded in CAPI (computer-assisted personal interviewing) and on paper. The issued sample was defined, and sample management procedures outlined. The CAPI was programmed and the documents were designed. Associated procedures were agreed, such as sample management, the packing and transport of equipment, the mailing of blood and saliva samples, and the exchange of information between the nurses, NatCen's Operations Department and the participating laboratories. Three field pilots were carried out, two using convenience samples recruited from the general public, and the third using a sample of cohort members. These are described in full in the Technical Report.

### 2.1 Target Sample

As in previous waves of the NCDS, the target sample for the biomedical survey was all cohort members currently living in England, Scotland or Wales,<sup>14</sup> excluding permanent refusals – 14,737 cohort members in August 2002. The sample definition was subsequently refined, and some cohort members excluded for various reasons, so that the sample issued to field (i.e. cohort members invited to take part in the study) comprised 12,037 cohort members.

### 2.2 Nurse briefings

Briefings for nurses began on 30th July 2002. Each briefing lasted three days and was attended by an average of seven nurses.

Day 1 was led by researchers from NatCen and CLS, and covered the background to the project; contact procedures; survey documents, including the ARF, consent

booklets and self-completion questionnaires; and the CAPI programme, including the CASI and the CIS-R interview.

Days 2 and 3 covered the measurement protocols in detail, including practical sessions, and were led by one of the principal investigators and the survey doctor, supported by NatCen nurse supervisors, and, on occasions, other survey collaborators.

Before leaving the briefing, each nurse carried out a dummy interview, including measurements, observed by members of the survey team and nurse supervisors. If there was any doubt about a nurse's ability to carry out the interview satisfactorily, additional training and supervision was arranged. All nurses who were new to NatCen were supervised on their first interviews.

Nineteen briefings were held in London and six regional centres. The first briefing, of nurse supervisors, took place at the end of July 2002. Six briefings were held in August, five in September and two in October 2002, to enable fieldwork to begin promptly. Subsequent briefings were held by demand in order to ensure coverage throughout the fieldwork period, in November and December of 2002, and February, May and July of 2003. In total, 122 nurses were briefed to work on the project during its lifetime.

## 2.3 Issue of work

Fieldwork took place between 9<sup>th</sup> September 2002 and 26<sup>th</sup> March 2004. The sample was issued in fifteen monthly waves, starting in September 2002 and finishing in December 2003.<sup>20</sup> It was divided into 667 assignments or 'points', each a month's workload for a nurse, clustered into geographically convenient groupings, with an average size of 18 addresses (see Chapter 2 of the Technical Report for further details).

### 3. Response

9378 cohort members were contacted and consented to the Biomedical Survey. One participant subsequently withdrew their consent and 9 interviews were lost. The dataset contains 9377 cases (the 9339 productive from the main survey and an additional 29 cases from the Dress Rehearsal.(Pilot 3) and the 9 lost interviews, for which blood was obtained).

84 participants of those 9377 cases did not provide consent for sharing the data and were excluded from the current data deposit, bringing the total number of cases to a total of 9293.

The full response is shown in the table below.

#### Summary of cohort, issued sample and response to biomedical survey

Valid	Frequency	Percent	Valid Percent
101 Productive	9339	50.3	50.3
102 Productive - dress rehearsal	29	.2	.2
103 Lost productive - some data, no CAPI	9	.0	.0
201 Issued but ineligible - dead	28	.2	.2
202 Issued but ineligible - living outside GB	65	.4	.4
301 Refused	1803	9.7	9.7
302 Data withdrawn after interview	1	.0	.0
401 Invalid address	12	.1	.1
402 No contact made at address	191	1.0	1.0
403 Moved to unknown address	494	2.7	2.7
404 Unavailable for interview	86	.5	.5
405 Other reason for no interview	12	.1	.1
501 Ineligible: dead	1196	6.4	6.4
502 Ineligible - living outside GB	1236	6.7	6.7
601 Permanent refusals	1041	5.6	5.6

Valid	Frequency	Percent	Valid Percent
901 NCDS6 by proxy	31	.2	.2
902 Not issued for another reason	2506	13.5	13.5
903 Not in sample data base	479	2.6	2.6
Total	18558	100.0	100.0

## 4. The interview

### 4.1 CAPI main interview

The CAPI interview used in the main survey included the following elements.

- Vision: measures of near vision in right and left eyes (using appropriate visual correction), with and without pinhole viewer; stereo vision; distance vision (using appropriate visual correction).
- Blood pressure and pulse: three measures of systolic and diastolic blood pressure and resting pulse.
- Prescription drugs: all prescribed drugs taken, by name and BNF code.
- Hearing: thresholds of hearing in right and left ears at 1kHz and 4kHz.
- Anthropometry: standing height, sitting height, weight, waist circumference, hip circumference.
- Lung function: three measures (from up to five attempts) of forced vital capacity (FVC), forced expiratory volume (FEV1) and peak flow (PF).
- Eye measurements using autorefractor: sphere, cylinder and axis of right and left eyes.
- Non-fasting blood sample: four tubes filled and sent by nurses to laboratories in London, Newcastle and Bristol.
- Computer Administered Self Completion (CASI):
  - AUDIT and supplementary questions about drinking alcohol
  - Adverse childhood experiences

- CIS-R<sup>1</sup> interview: modules covering appetite, fatigue, concentration and forgetfulness, sleep problems, irritability, depression, depressive ideas, anxiety, phobias, and panic.

## 4.2 Paper self-completion questionnaires

Two paper self-completion questionnaires covered the following topics.

### **Questionnaire 1 (SC1) ('yellow questionnaire', completed in advance):**

- sun exposure
- physical activity connected with work
- hearing
- eyesight
- pain
- working conditions
- household circumstances
- social support.

### **Questionnaire 2 (SC2) ('lilac questionnaire', completed during the interview):**

- general health and diet
- leisure exercise
- employment
- partnership status and children
- life events and
- (women only) contraception and HRT

Appendix A of the Technical Report includes the CAPI and self-completion questionnaires

---

<sup>1</sup> Lewis G, Pelosi AJ, Araya R, Dunn G. Measuring psychiatric disorder in the community: a standardized assessment for use by lay interviewers. *Psychological Medicine*, 1992, 22 (2), 465-86.

## 4.3 Consents

Written consents were collected for:

- all measurements;
- questions about psychological health;
- collection, testing, storage and future use of blood;
- feedback of blood test results to the cohort member;
- feedback of measurements and results to the cohort member's GP;
- analysis, storage and future use of saliva;
- deposit of data in the ESRC data archive;
- use of NHS administrative data.

## 4.4 Saliva sample

Cohort members were asked to collect and return two samples of saliva, accompanied by a short self-completion questionnaire.

Appendix B of the Technical Report includes information about the equipment used and nurse protocols.

Consent booklets and saliva forms can be found with other survey documents in Appendix C of the Technical Report.



## 5. Biochemical analyses of blood samples

### **Glycosylated haemoglobin**

Glycosylated haemoglobin (hba1c) was measured on whole citrated blood by ion exchange high performance liquid chromatography, using the Tosoh A1c 2.2 Glycohemoglobin Analyser HLC-723GHb.

### **Triglycerides, total and HDL cholesterol**

Triglycerides, total and HDL cholesterol variables (trig, chol, hdl, ldl) were measured on non-fasting serum by Olympus model AU640 autoanalyser.

LDL cholesterol was derived by the following formula:

$$\text{ldl} = \text{Total cholesterol (chol)} - (\text{hdl} + (\text{trig}/2.2))$$

### **Insulin-like growth factor**

Insulin-like growth factor (igf1) was measured on serum by chemiluminescence-immunoassay.

### **Immunoglobulin E , House Dust Mite Allergen, Cat Allergen, Grass Pollen Allergen**

Immunoglobulin E (ige) was measured on serum by the HYTEC enzyme immunoassay, with positive and negative controls. Total IgE was assayed on all specimens

Allergen-specific IgE to house dust mite (hdm), mixed grasses (grass), and cat fur (cat) were measured on specimens with a total IgE concentration above the median (30kU/L).

### **Fibrinogen**

Fibrinogen (fib) was measured on citrated plasma by the Clauss method using a MDA 180 coagulometer

### **Tissue plasminogen activator antigen and von Willebrand factor antigen**

Tissue plasminogen activator antigen (t-pa) and von Willebrand factor antigen (vwf) were measured on citrated plasma by enzyme-linked immunosorbent assays (ELISA) employing a double sandwich technique.<sup>27</sup>

### **C-reactive protein**

C-reactive protein (crp) contains data on C-reactive protein measured on citrated plasma by high-sensitivity nephelometric analysis of latex particles coated with CRP-monoclonal antibodies.

### **D-Dimer**

Data on fibrin D-dimer was received from Dr A Rumley and Professor GDO Lowe (University Department of Medicine, Royal Infirmary) on 1st September 2006. The D-dimer assays were performed at the end of the field study using two different manufacturer kits. Within in manufacture kit quality control checks of the D-dimer assays were within set limits but there was a systematic difference between the two different kits (see Appendix 1).

## 6. Research data

### 6.1 Structure of the data

The files available under the UKDS are

Name	Description
ncds_biomedical_2002_eul.sav	End User Licence data
ncds_biomedical_2002_sl.sav	Special Licence data
NCDS_biomedical_2002_technical_report.pdf	Technical report on the survey
NCDS_biomedical_2002_coded_manikins.pdf	Diagram relating to Pain coding
ncds_biomedical_2002_capi_questionnaire.pdf	Questionnaire
NCDS_biomedical_2002_data_dictionary.xls	List of variables, topics and data collection instruments (EUL)

Variables labels are prefixed to indicate which part of the survey excluding the main CAPI interview they were collected in, as follows:

- CASI – Computer Aided Self Completion
- CIS-R – CASI – CIS-R Malaise questions
- SC1 – Self Completion Booklet 1 (variable label indicates the question number)
- SC2 – Self Completion Booklet 2 (variable label indicates the question number)
- LAB – Data relating to blood sampling, biochemical measures.

### 6.2 End User Licence Dataset

The End User Licence (EUL) dataset contains 1079 variables. It includes the majority of the data previously available under Special Licence, as well as CIS-R summary scores and data on taken medications that were previously collected in the Biomedical Survey 2002-4 but never made available via the UKDS before.

This EUL dataset contains modified variables generated as a de-identified version of variables previously available under Special Licence. The modified variables can be identify with the “\_M” at the end of their name. The original variables have been kept in a dataset available under Special Licence.

A complete list of the variables is available at the UKDS website and can be downloaded with the dataset.

The contents of the EUL dataset appear in the following order:

### **CAPI**

- Interview information
- Vision Tests
- Blood Pressure Tests
- Audiometry
- Anthropometry: height, sitting height, weight, waist, hip
- Lung Function
- Eye Measurement of Refractive Error
- Blood Sampling
- Alcohol Audit (CASI)
- Mental Health Interview: CIS-R summary scores
- Medications
- Administrative and data consent variables

### **Self completion 1**

- A. Sun exposure
- B. Physical Activity connected with work
- C. Hearing
- D. Eyesight
- E. Pain
- F. Work
- G. Household circumstances
- H. Social Life

### **Self completion 2**

- A. General Health and Diet
- B. Leisure Activities
- C. Work and Home Circumstances
- D. Life Events
- E. Women Only (contraception and HRT)

#### **Blood samples /Biomedical Assays**

- Blood results
- Saliva collection

#### **Derived variables**

### **6.3 Special Licence Dataset**

The Special License dataset contains 122 biomedical variables, which can be divided into three different sections:

#### **Childhood Adversity (CASI)**

This battery of questions regarding adverse childhood conditions and upbringing (CHAD) was previously available via the UKDS (25 variables).

The summary variable adversity has been derived by Charlotte Clark at QMUL - see Appendix 2.

#### **Mental Health Interview: CIS-R questions**

The Clinical Interview Schedule-Revised (CIS-R) questions on appetite, fatigue, concentration and forgetfulness, sleeping problems and patterns, irritability, depression, depressive ideas, anxiety, phobias and panic (87 variables).

#### **Detailed personal characteristics**

A number of variables have been kept under Special Licence as they may contain potentially disclosive information such as height, skin colour, pregnancy, a surgical operation and number of children (10 variables).

## 6.4 Recoding of disclosive information for availability under End User Licence

A less disclosive version of the 10 variables potentially disclosive kept under Special Licence has been created in order to make this information available under End User Licence. The original variables names have been kept in the Special Licence dataset. The modified variables available under End User Licence have been renamed to have the “\_m” suffix at the end of the name.

<b>Original Variable (Special Licence)</b>	<b>Recoded Variable (EUL)</b>	<b>Variable label</b>	<b>Amendment</b>
bheight	bheight_m	Height	Cap to 195 if values $\geq$ 195 Cap to 145 if values $\leq$ 145
estht	estht_m	Final measured or estimated height	Cap to 195 if values $\geq$ 195 Cap to 145 if values $\leq$ 145
skincol	skincol_m	Self-reported natural skin colour	Medium and dark skin combined
childnow	childnow_m	Number of children living with you, aged 18 or less	Cap to 6 for values $\geq$ 6
childnum	childnum_m	Number of natural (biological) children	Cap to 7 if values $\geq$ 7
hadop01a	hadop01a_m	Age at the time of removal of uterus and both ovaries	Cap to 20 if values $\leq$ 20
dvhght44	dvhght44_m	Height at interview (age 44)	Cap to 194 if values $\geq$ 194 Cap to 148 if values $\leq$ 148

## 6.5 Data Coding and Editing

This section offers additional information to the coding and editing referred to in Section 7 of the Technical Report.

### **Missing values**

Where no data entry in CAPI (e.g. because of routing) cells have been left empty. Not on route is coded –1 in data from the self-completion questionnaires. Codes -8, -9 were used to label cases that did not provide an answer or cases without applicable answers.

For variables where ‘refused’ is a listed option this has been retained and has not been recoded to merge with –9 entries. Where values are measures, no response recorded as 99, 999 etc (Note: some value ranges have negative values).

### **Refused consent/not attempted**

Participants who did not provide consent to deposit their responses at the UK Data Service have been excluded, thus lowering the number of cases from 9377 to 9294.

These data have not been edited from the original CAPI, although the outcomes for particular measures may be inconsistent in cases where consent was withdrawn during measurement (e.g. blood pressure).

### **Inconsistencies, improbabilities**

The edit has taken note of soft checks within CAPI. Within the self-completion booklets, the approach has been inclusive: that is, inconsistencies and improbabilities have not always been corrected: the policy on this was decided in consultation with the scientific collaborators.

### **Sex of cohort member as key variable**

This is the “gold standard” variable for the sex of the cohort member, other variables which relate to the sex of the cohort member are left, to explain any consequent erroneous routing.

### **Lung function values**

Values have not been cleaned, and there are no 'not obtained' or missing value codes (for FVC these have been left as 9.96, for FEV 9.95, for Peak Flow 995).

### **Nurse number**

The nurse number (nurseno) has been recoded for reasons of anonymity.

### **Retinomax codes**

Where the quality score has been recorded as 0 or 99, all measures for that eye have been recoded as not obtained. Measures with unsatisfactory quality scores of 1 to 6 have been left in the data.

### **CIS-R summary scores**

The EUL dataset includes the summary scores for CIS-R modules:

totsum	CIS-R Total summary variable
overalla	CIS-R Overall stopped from doing things used to do in past 7 days
overallb	CIS-R Overall stopped from doing things more than once
overallc	CIS-R Overall things more difficult
ideasum	Depressive ideas summary score
depsum	Depression summary score
pansum	Panic summary score
phobsum	Phobias summary score
anxsum	Anxiety summary score
fatsum	Fatigue summary score
irritsum	Irritability summary score
slpsum	Sleep summary score
forgsum	Concentration and forgetfulness summary score



## **Pain Coding**

Variables [e3\_acr1 to e3\_acr10] and [e3\_man1 to e3\_man29] should be used in conjunction with the accompanying document coded\_mannikins.pdf.

## **Lab Data Batch numbers**

There are four different laboratory numbers, to be used as follows:

Newcastle batch number (nclbatch): for use with IgE and IGF-1 measurements

vWF batch number (vwfbatch): for use with vWF measurement

t-PA batch number (tpabatch): for use with t-PA measurement

Glasgow batch number (glabatch): for use with fibrinogen and CRP measurements

## **Blood Collection**

For variables clorec and bldrec the values are (-3) "Old/Insufficient" (-2) "No blood received at lab" (-1.0) "Not recorded" (0) "No" (1) "Yes".

All variables with dates for blood reception at the lab: (bldday, bldmonth, bldyear, dayrnl, monthrnl and yearrnl) are set to missing if recorded as (-2) "No blood received at lab" or (-1) "Not recorded".

## **Consent given for storing blood samples**

Variables consn, plstor, sestor refer to consent given for storing blood, plasma and serum. (-2.0) "No blood received at lab" (-1.0) "Not recorded" (0.0) "No" (1.0) "Yes".

## **Recoding of samples processed at Newcastle Laboratory**

The result was re-coded to missing if the original lab value was coded as NSR (no sample received), NOSEP (no separation), INSUFF (insufficient), UNOLD (old sample - more than 5 days old), UNUS (atypical chromatographic pattern relating to HbA1c), RNCALB (LDL result not calculable), UNSUIT (for HDL calculation as triglyceride>13), LDLUS (for LDL calculation as HDL>4.5), HAEM (?), UNAVH, UNAVL, UNCALC or coded NR (for not required, but only if IgE >30).

<b>Variable</b>	<b>Recode</b>	<b>Comments</b>
ldl	"<0.1" re-coded to MISSING	ldl <0.1 - According to Newcastle this is the default (nonsense) value reported by the computer system for a calculated result when the calculation components are non-numeric.
ldl	If hdl measured and hdl>4.5 then ldl re-coded to MISSING	According to Newcastle LDL calculation invalid if HDL>4.5
igf1	"<2" re-coded to 0	
ige	ige ">2000" re-coded to 2000	
hdm	"<0.35" re-coded to 0	
hdm	if Ige measured and Ige<=30 re-coded to 0	According to Newcastle if IgE<=30, house dust mite allergen, cat and grass allergen results are not required.
cat	"<0.35" re-coded to 0	
grass	"<0.35" re-coded to 0	
cat	if Ige measured and Ige<=30 re-coded to 0	According to Newcastle if IgE<=30, house dust mite allergen, cat and grass allergen results are not required.
grass	if Ige measured and Ige<=30 re-coded to 0	According to Newcastle if IgE<=30, house dust mite allergen, cat and grass allergen results are not required.
hdl	If triglyceride measured and triglyceride>13 then hdl re-coded to MISSING	According to Newcastle blood unsuitable for HDL analysis if triglyceride>13.

If there was a delay between sampling date and date received > 5 days (or dates missing) then all results set to MISSING. Analysis was considered not valid if a sample was more than 5 days old. (NB Delay could not be calculated for 9 study subjects as information on time blood taken missing).

### **Removal of redundant blood pressure variables**

Variables sysres1, sysres2 and sysres3 were excluded in this latest deposit because they contain exactly the same data available in the variables sys, sys2 and sys3. These variables contain information about systolic readings from blood pressure.

Similarly the variable nobp was deleted because the same information related to no BP measure due to pregnancy was already included in the variable pregmes. This later variable contained a more detailed label and is available in the Special Licence dataset.

The variable bresps2 only contained value 1 for all cases with the label 'some data'. This variable was considered redundant and removed from the new deposit.

## **6.6 Coding of medication variables**

Variables drc1\_m to drc19\_m contain data of prescribed medication taken by the cohort member. The data was self-reported by the cohort members in open-end question text format and subsequently coded using 6-digits BNF codes.

There are up to 19 drugs per person in the EUL dataset. This detailed information is sensitive and could potentially lead to data disclosure. The structure of the BNF codes is shown below as an example:

BNF Chapter	BNF Section	BNF Paragraph
05	03	05
Infections	Antiviral Drugs	Macrolides

To reduce the data disclosure risk and deposit the variables under the End User Licence the BNF codes have been truncated to 4 digits. The first two digits refer to the BNF Chapter and the second two digits refers the BNF Section. In addition, categories with less than 10 cases were grouped together in a residual category 'other' for each of the chapter.

The set of variables [ytake11 - ytake50] denote the reason for taking the drug. A taken medication can have up to three reasons.

## 6.7 Derived Variables

The derived variables included in the EUL dataset are:

<b>Name</b>	<b>Label</b>
smok42_1	Smoking at 42 codes 0 to 2
smok42_2	Smoking at 42 codes 0 to 6
sc0_1	Social Class at birth
sc42_1	Social Class at 42
reg0_58	Region at birth (based on pre-1974 regions)
reg46_58	Region at age 46 (based on pre-1974 regions)
reg46_04	Region at age 46 (based on post-1974 regions)
bresps0	Final outcome code
bresps1	Summary of cohort, issued sample and response

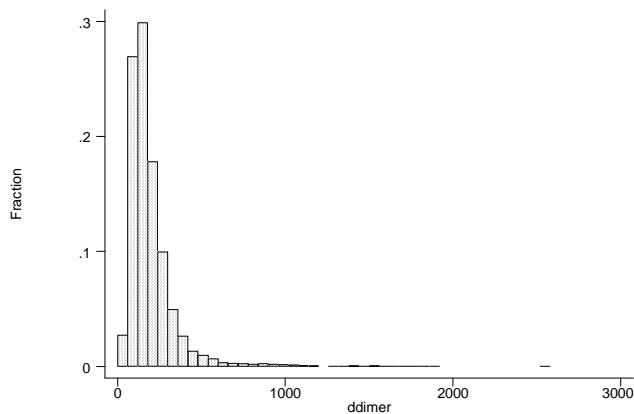
## Appendix 1: D-Dimer Edits

**Author:** Alicja R Rudnicka, St George's, University of London

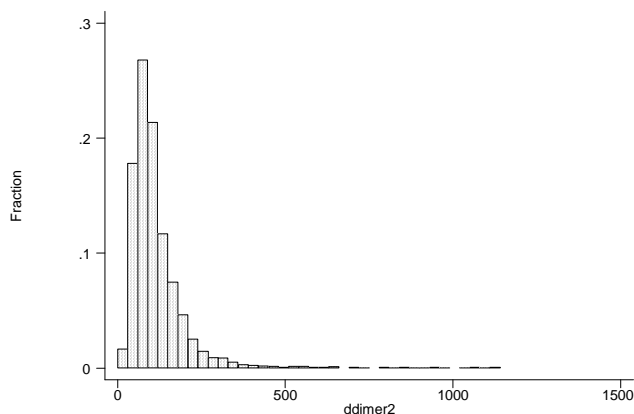
Data on fibrin D-dimer was received from Dr A Rumley and Professor GDO Lowe (University Department of Medicine, Royal Infirmary) on 1 September 2006. The D-dimer assays were performed at the end of the field study using two different manufacturer kits. Within in manufacture kit quality control checks of the D-dimer assays were within set limits but there was a systematic difference between the two different kits.

Preliminary results of the difference between two manufacture kits used to measure D-dimer

D-dimer results from Kit (1)



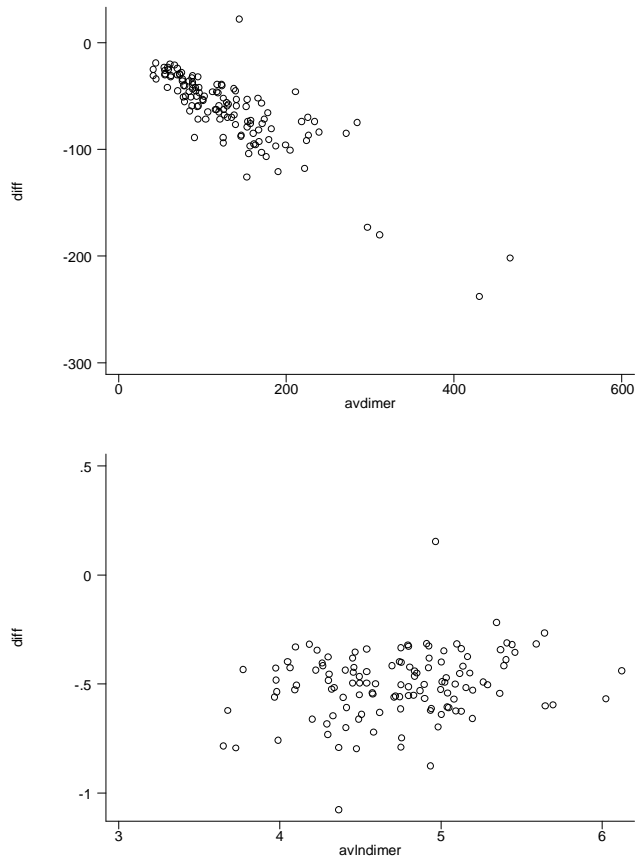
D-dimer results from Kit (2)



	D- Dimer kit (1) N=4609 (ng/mL)	D-dimer kit(2) N=3182 (ng/mL)
Median values	160	93
5 <sup>th</sup> Centile	71	40
90 <sup>th</sup> centile	325	197

There was a systematic difference between the two kits, with the second kit producing considerably lower readings.

Repeatability of repeat measures of D-dimer on same 1958 samples using two different manufacturer's kits was assessed in 124 samples. The association between the two different kits is displayed graphically below. Figure 1a shows that on the arithmetic scale the difference between the two kits decreases as values of D-dimer increase. Figure 1b shows that the difference between the two kits is constant with  $\log(\text{d-dimer})$ . There is therefore a constant proportional difference between kit (1) and kit(2). All analyses and adjustments were therefore performed using the log transformation of D-dimer



**Figure 1:** Difference in d-dimer values of kit(2) minus kit(1) on the same samples plotted against average of D-dimer values from kit(1) and kit(2) on the arithmetic scale in (a) and logarithmic scale in (b).

On average D-dimer values from kit (2) were 40% lower than with kit (1). After discussions with Dr A Rumley and Professor GDO Lowe a decision was made to adjust the values from kit(2) to be in line with that on kit (1). All values from kit (2) were increased by 40%.

## Appendix 2: Deriving Childhood Adversity

**Author:** Charlotte Clark, Barts School of Medicine & Dentistry, University of London

The scale used in the biomedical survey was taken from the Australian Path Through Life Study (Rosenman and Rodgers, 2004, 2006). This consists of 17 items which are retrospectively rated as having occurred or not (coded 0=no, 1=yes) by 16 years of age.

1.	Father or father figure was not affectionate towards me
2.	Father or father figure suffered nervous or emotional trouble or depression
3.	Father or father figure had trouble with drinking or other drug use
4.	Mother or mother figure was not affectionate towards me
5.	Mother or mother figure suffered nervous or emotional trouble or depression
6.	Mother or mother figure had trouble with drinking or other drug use
7.	There was much conflict and tension in the household whilst I was growing up
8.	I was neglected
9.	I had a strict, authoritarian or regimented childhood
10.	I grew up in poverty or financial hardship
11.	I was verbally abused by a parent
12.	I suffered humiliation, ridicule, bullying or mental cruelty from a parent
13.	I witnessed physical or sexual abuse of others in the family
14.	I was physically abused by a parent – punched, kicked, hit or beaten with an object or needed medical treatment
15.	I received too much physical punishment – hitting, smacking etc
16.	I was sexually abused by a parent



Did your parents divorce or permanently separate when you were a child –  
not asked in Biomedical Survey

**NOTE:** in the survey, the question about parental divorce by age 16 was NOT asked, but details of divorce can be obtained from childhood measures, as well as data about 33y.

The derived variable adds up the total number of adversities reported – so the scale ranges from 0 to 16, as divorce is not included.

A total score has not been provided for those who were missing one or more item on the scale (coded as -8 in the variable).

Rosenman, S., Rodgers, B., 2004. Childhood adversity in an Australian population. *Soc Psychiatry Psychiatr Epidemiol* 39, 695-702.

Rosenman, S., Rodgers, B., 2006. Childhood adversity and adult personality. *Aust N Z J Psychiatry* 40, 482-490.