



## CLOSER Work Package 2:

### Harmonised Socio-Economic Measures User Guide

Brian Dodgeon  
Tim Morris  
Claire Crawford  
Samantha Parsons  
Anna Vignoles  
Zoe Oldfield

**January 2018**

# Copyright

This document is released under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) Licence. The extract below is a summary. The full terms are available from <https://creativecommons.org/licenses/by-nc/4.0/legalcode>

## You are free to:

- **Share** — copy and redistribute the material in any medium or format
- **Adapt** — remix, transform, and build upon the material

The licensor cannot revoke these freedoms as long as you follow the license terms.

## Under the following terms:

- **Attribution** — You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- **NonCommercial** — You may not use the material for [commercial purposes](#).
- **No additional restrictions** — You may not apply legal terms or [technological measures](#) that legally restrict others from doing anything the license permits.

## Notices:

- You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable [exception or limitation](#).
- No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as [publicity, privacy, or moral rights](#) may limit how you use the material.

Contents

Copyright.....2

Preface .....4

Acknowledgements .....4

Introduction .....4

Studies included .....4

Methods .....5

Data Cleaning.....6

Limitations .....6

Quality Assessment .....6

Harmonised Variables Description .....7

Datasets .....14

Appendix 1: Derivation Code .....17

    Derivation Code.....17

    Source Files.....17

References .....18

## Preface

CLOSER (Cohort & Longitudinal Studies Enhancement Resources) aims to maximise the use, value and impact of longitudinal studies, both at home and abroad. Bringing together eight leading studies, the British Library and the UK Data Service, CLOSER works to stimulate interdisciplinary research, develop shared resources, provide training, and share expertise. In this way CLOSER is helping to build the body of knowledge on how life in the UK is changing – both across generations and in comparison to the rest of the world.

CLOSER's research includes a number of work packages focused on retrospective harmonisation, their aim being to make the data from different longitudinal studies more comparable in order to find out how life in the UK is changing from generation to generation. This documentation describes the datasets produced as part of the second CLOSER harmonisation work package about social class.

## Acknowledgements

The authors would like to thank the owners of the five studies in question and of a valuable 'subsidiary' dataset connected with NCDS and BCS70:

- the MRC Unit for Lifelong Health and Ageing at UCL (NSHD);
- the Centre for Longitudinal Studies at the UCL Institute of Education (NCDS, BCS70, MCS);
- Avon Longitudinal Study of Parents and Children, University of Bristol (ALSPAC);
- Paul Gregg of the University of Bristol, who deposited the NCDS/BCS70 'Occupational Coding' dataset (SN7023) at the UK Data Archive.

We also thank Prof Paul Lambert of Stirling University, who gave invaluable advice about the feasibility of converting data between different classification systems, and produced many look-up tables downloadable from the CAMSIS website ([www.camsis.stir.ac.uk/](http://www.camsis.stir.ac.uk/)), which informed the conversion syntax we in turn produced.

Finally, individual thanks are due to Professors David Rose and David Pevalin of ISER, who produced look-up tables downloadable from the ISER website which facilitated back-coding from the SOC2010 occupational classification to earlier coding regimes.

## Introduction

The aim of this project was to harmonise and document a set of variables measuring socio-economic circumstances from the 1946, 1958, and 1970 birth cohort studies, plus ALSPAC and the Millennium Cohort Study, and publish this on the CLOSER website.

## Studies included

National Survey of Health and Development (NSHD)  
National Child Development Study (NCDS)  
British Cohort Study 1970 (BCS70)  
Millennium Cohort Study (MCS)

## Methods

The team considered a range of possible socio-economic measures before choosing one around which all five studies could be harmonised. The most difficult problem was the wide historical range of the starting-years for the five surveys, from 1946 (NSHD) to 2000/01 (MCS). During this 55-year period, different official methods of occupational classification have superseded each other every decade, as have many derived measures of socio-economic status. The list below mainly features those devised by the Office of Population Censuses and Surveys (OPCS, formerly known as the Registrar General's Office), which became the Office for National Statistics (ONS) in 1996:

### Official occupational classifications

- CO70 (OPCS 1970 Classification Of Occupations) - 223 categories, subdivided into 26 Occupational Orders
- CO80 (OPCS 1980 Operational Coding Groups) – 350 categories (or 547 'Occupational Groups' when split by supervisory status)
- SOC90 - 371 'unit' groups, with 77 Minor Groups, 22 Sub-Major Groups, 9 Major Groups
- SOC2000 – 353 unit groups, 81 Minor Groups, 25 Sub-Major groups, 9 Major Groups
- SOC2010 - relatively minor revisions to SOC2000

### Official derived measures of socio-economic status

- Registrar General's Social Class (RG Class) – revised decennially 1970 until 1990 (see 3 versions below), then superseded by NS-SEC in 2000. Always has the same six categories: I, II, III-Non-Manual, III-Manual, IV, V
- RG Class 1970 version: derived from CO70 plus employment status
- RG Class 1980 version: derived from CO80 plus employment status
- RG Class 1990 version (also known as 'Social Class Based on Occupation'): derived from SOC90 plus employment status
- Socio-Economic Group (SEG)/ Socio-Economic Class (SEC): various other classification systems from 1970-1990 combining occupation with supervisory status and size of establishment - 15, 20 or 40 categories
- NS-SEC (Office for National Statistics Socio-Economic Classification). Introduced by ONS in 2000 to supersede RG Class and SEG/SEC, incorporating distinct classes for the self-employed, & available in versions with 5, 7 or 13 categories, as well as the full 35-category version
- CAMSIS: (Cambridge Social Interaction and Stratification): a continuous measure of arithmetic standardisation based on evaluating occupations internationally, devised by Ken Prandy (ex-Cardiff & Cambridge University, now retired) & Paul Lambert (Stirling University)

Whichever coding system were to be chosen, a further challenge to bear in mind was that all the above measures rely on discerning from the respective fieldwork questionnaires:

- The unique (ie classifiable) characteristics of the person's job; and

- The degree of supervisory status the job entailed (self-employed/manager/foreman/employee), plus the establishment size (usually 7 categories).

Fieldwork questionnaires provide us with an occupational ‘text-string:’ the person’s own description of the job they do. The task of coding this to one of the above occupational classification systems can be difficult. For instance, a text-string “Petrol pump attendant” is a clear job description which is easy to code. But some respondents may describe their job in much less clear terms, e.g: “I sort out the electrical boxes at the back of the supermarket” – difficult to code.

Coding occupational text-strings into an Occ. classification system is usually a manual exercise, but can be done with software such as CASOC<sup>1</sup> or (later version) CASCOT<sup>2</sup>, developed by Warwick University.

Even CASCOT, when run in ‘fully automatic’ mode, makes some mis-codings. So ‘semi-automatic’ is preferable (i.e. manual intervention where CASCOT is struggling)

The final decision was to standardise on **RG Class (1990 version)**, which stood chronologically in the middle of the period 1946-2016, and was the measure to which, with some minor exceptions, the measures of all six studies could be converted.

Given that the six studies surveyed their respective participants at varying ages, a decision was taken to limit the harmonisation of social class to just one childhood time-point (for parent’s occupation, nearest to age 10/11) and one adult time-point (nearest to age 42). Of course in the case of MCS the cohort are still not old enough to have an adult time-point.

## Data Cleaning

For NCDS and BCS70 at age 10/11 there were a small number of cases where the desired harmonised indicator (Father’s RG Class 1990 version) could not be produced, because of the illegibility (or non-availability) of the occupational text strings to be read into the CASCOT software (see ‘Description of Derivation’ under ‘Harmonised Variables Description below).

In these cases, a categorisation was substituted from two parallel variables already available in the deposited NCDS and BCS70 datasets, which showed RG Class (1970 version) for NCDS age 11 and RG Class (1980 version) for BCS70 age 10 respectively, with a flag being provided to indicate which cases had received this substitution.

## Limitations

The data from NCDS, BCS70 and MCS can be downloaded subject to agreement to the normal End-User Licence at the UK Data Archive. For NSHD an approach should be made to the MRC Unit for Lifelong Health and Ageing at UCL).

## Quality Assessment

No quality issues were observed with any of the six datasets, other than those noted above under ‘Data Cleaning.’

## Harmonised Variables Description

Harmonised Variable Description		
<b>Variable name:</b>	NCDS2FCL	
<b>Variable Description:</b>	Father's RG Social Class (1990 version) at age 11 sweep, NCDS	
<b>Description of derivation</b>	<p>Values are based principally on variable N2SRGSC from the Paul Gregg 'Occupational Coding' dataset (SN7023), which has RG Class 1990 codes for 13,019 cases for which an occupational text string could be successfully read into the CASCOT software, which was used to derive that dataset.</p> <p>Codes for a further 2,318 cases were obtained by reference to the existing NCDS variable n1687 (RG Class 1970 version). These 2,318 cases are flagged in the variable NCDS2FLG.</p> <p>For those cases where the age 11 Parental Questionnaire was completed and there was a male head of household, but his occupation was not codifiable as a substantive RG Social Class 1.0-5.0, they are all coded 6.0 (armed forces/ unemp/ sick/ retired/ unclassifiable etc). Those interested in a more detailed breakdown of these aggregated cases may approach the Centre for Longitudinal Studies at UCL Institute of Education.</p>	
<b>Variable Code List</b>	1.0	Professional etc. Occupations
	2.0	Managerial and Technical occupations
	3.1	Skilled occupations (non-manual)
	3.2	Skilled occupations (manual)
	4.0	Partly-skilled occupations
	5.0	Unskilled occupations
	6.0	Occupation unclassifiable/ insufficient information/ Armed Forces/ carer/ unemployed/sick/retired
	10.0	No Male Head
	11.0	Age 11 Parental Questionnaire not completed

Harmonised Variable Sources Description		
Study	Age	Source Variable Name
NCDS	11	N2SRGSC
NCDS	11	N1687

Harmonised Variable Description		
<b>Variable name:</b>	NCDS6CL	
<b>Variable Description:</b>	Cohort Member's RG Social Class (1990 version) at age 42 sweep, NCDS	
<b>Description of derivation</b>	<p>Values are based on variable SC from the NCDS 1999/2000 dataset (SN5578), which has valid RG Class 1990 codes for 9,613 cases.</p> <p>For those cases where the age 40 Questionnaire was completed but the cohort member's occupation was not codifiable as a substantive RG Social Class 1.0-5.0, they are all coded 6.0 (armed forces/ carer/ unemp/ sick/ retired/ unclassifiable etc). Those requiring a more detailed breakdown of these aggregated cases may approach the Centre for Longitudinal Studies at UCL Institute of Education.</p>	
<b>Variable Code List</b>	1.0	Professional etc. Occupations
	2.0	Managerial and Technical occupations
	3.1	Skilled occupations (non-manual)
	3.2	Skilled occupations (manual)
	4.0	Partly-skilled occupations
	5.0	Unskilled occupations
	6.0	Occupation unclassifiable/ insufficient information/ armed forces/ carer/ unemployed/sick/retired
	11.0	Age 42 Questionnaire not completed

Harmonised Variable Sources Description		
Study	Age	Source Variable Name
NCDS	42	SC



Harmonised Variable Description		
<b>Variable name:</b>	BCS3FCL	
<b>Variable Description:</b>	Father's RG Social Class (1990 version) at age 10 sweep, BCS70	
<b>Description of derivation</b>	<p>Values are based principally on variable B3FSRGSC from the Paul Gregg 'Occupational Coding' dataset (SN7023), which has valid RG Class 1990 codes for 12,047 cases for which an occupational text string could be successfully read into the CASCOT software, used in the derivation of that dataset. Codes for a further 1,265 cases were obtained by reference to the existing BCS70 age 10 variable c3.4 (RG Class 1980 version). These 1,265 cases are flagged in the variable BCS3Flag.</p> <p>For those cases where the age 10 Parental Questionnaire was completed and there was a male head of household, but his occupation was not codifiable as a substantive RG Social Class 1.0-5.0, they are all coded 6.0 (armed forces/ carer/ unemp/ sick/ retired/ unclassifiable etc). Those requiring a more detailed breakdown of these aggregated cases may approach the Centre for Longitudinal Studies at UCL Institute of Education.</p>	
<b>Variable Code List</b>	1.0	Professional etc. Occupations
	2.0	Managerial and Technical occupations
	3.1	Skilled occupations (non-manual)
	3.2	Skilled occupations (manual)
	4.0	Partly-skilled occupations
	5.0	Unskilled occupations
	6.0	Occupation unclassifiable/ insufficient information/ armed forces/ carer/ unemployed/sick/retired
	11.0	Age 10 Parental Questionnaire not completed

Harmonised Variable Sources Description		
Study	Age	Source Variable Name
BCS70	10	B3FSRGSC
BCS70	10	C3.4
BCS70	10	A6.1

Harmonised Variable Description		
<b>Variable name:</b>	BCS6CL	
<b>Variable Description:</b>	Cohort member's RG Social Class (1990 version) at age 42 sweep, BCS70	
<b>Description of derivation</b>	<p>Values are based on variable B9CSC from the BCS70 Age 42 dataset (SN7473)</p> <p>For those cases where the age 42 Questionnaire was completed but the cohort member's occupation was not codifiable as a substantive RG Social Class 1.0-5.0, they are all coded 6.0 (Occup unclassifiable/ insuff. info/ armed forces/ carer/ unemp./ sick/ retired). Those requiring a more detailed breakdown of these aggregated cases may approach the Centre for Longitudinal Studies at UCL Institute of Education.</p>	
<b>Variable Code List</b>	1.0	Professional etc. Occupations
	2.0	Managerial and Technical occupations
	3.1	Skilled occupations (non-manual)
	3.2	Skilled occupations (manual)
	4.0	Partly-skilled occupations
	5.0	Unskilled occupations
	6.0	Occupation unclassifiable/ insufficient information/ armed forces/ carer unemployed/sick/retired
	11.0	Age 42 Questionnaire not completed

Harmonised Variable Sources Description		
Study	Age	Source Variable Name
BCS70	42	B9CSC

Harmonised Variable Description		
<b>Variable name:</b>	MCS5PCL	
<b>Variable Description:</b>	Parent's Social Class (1990 version) at age 11 sweep, MCS5	
<b>Description of derivation</b>	<p>Derived from Parent's Occupation (SOC2000 version) and Parent's employment status, using look-up table downloaded from ISER.</p> <p>Two MCS5 datasets needed to be merged to produce the derivation, since one variable (EPSOC200) was on the MCS5_parent_interview dataset, and the other (EDDEMP00) was on the MCS5_parental_DVs dataset.</p> <p>Because of the structure of these datasets (one line per person interviewed, which may mean both parents, but not always), we produced a 'Parent's Social Class' harmonised variable for each parent present, and coded MCS5PCLAS to be the highest of the two values in the case of both parents having a valid social class, or the only valid one if that were the case.</p>	
<b>Variable Code List</b>	1.0	Professional etc. Occupations
	2.0	Managerial and Technical occupations
	3.1	Skilled occupations (non-manual)
	3.2	Skilled occupations (manual)
	4.0	Partly-skilled occupations
	5.0	Unskilled occupations
	6.0	Occupation unclassifiable/ insufficient information/ armed forces/ carer unemployed/sick/retired
	11.0	Age 11 Parental Questionnaire not completed

Harmonised Variable Sources Description		
Study	Age	Source Variable Name
MCS	11	EPSOC200
MCS	11	EDDEMP00

Harmonised Variable Description		
<b>Variable name:</b>	NSHDFCL	
<b>Variable Description:</b>	Father's RG Social Class (1970 version) at age 11 sweep, NSHD (1957)	
<b>Description of derivation</b>	Values are based on variables FOCC57 (Father's occupation) and FSTAT57 (Father's occupational status) from the NSHD 1957 follow-up dataset. These were coded to the 1970 versions of the OPCS occupational coding scheme. The standard OPCS syntax was used for derivation of 1970 RG Social Class, as no look-up table is available to harmonise 1970 occupational codings to the 1990-version RG Class.	
<b>Variable Code List</b>	1.0	Professional etc. Occupations
	2.0	Managerial and Technical occupations
	3.1	Skilled occupations (non-manual)
	3.2	Skilled occupations (manual)
	4.0	Partly-skilled occupations
	5.0	Unskilled occupations
	6.0	Occupation unclassifiable/ insufficient information/ Armed Forces/ carer/ unemployed/sick/retired
	11.0	Age 11 Parental Questionnaire not completed

Harmonised Variable Sources Description		
Study	Age	Source Variable Name
NSHD	11	FOCC57
NSHD	11	FSTAT57

Harmonised Variable Description		
<b>Variable name:</b>	NSHDCL	
<b>Variable Description:</b>	Cohort Member's RG Social Class (1990 version) at age 43 sweep, NSHD (1989)	
<b>Description of derivation</b>	Values are based on variables OUR89 and ESR89 (Occupation and employment status at age 42 in 1989), which were coded to the 1980 OPCS occupational coding scheme. The 1990 version of RG Social Class was derived from these two variables using a look-up table provided by CAMSIS at Stirling University.	
<b>Variable Code List</b>	1.0	Professional etc. Occupations
	2.0	Managerial and Technical occupations
	3.1	Skilled occupations (non-manual)
	3.2	Skilled occupations (manual)
	4.0	Partly-skilled occupations
	5.0	Unskilled occupations
	6.0	Occupation unclassifiable/ insufficient information/ armed forces/ carer/ unemployed/sick/retired
	11.0	Age 43 Questionnaire not completed

Harmonised Variable Sources Description		
Study	Age	Source Variable Name
NSHD	43	OUR89
NSHD	43	ESR89

## Datasets

**Dataset Name:** NCDS Harmonised Childhood SES

Variable	Label	Format	No. Valid Cases
NCDSID	NCDS Case Identifier	String	18558
NCDSSEX	Sex of cohort member	Numeric	18555
NCDS2FCL	Father's RG Social Class at age 11 Sweep (1990 version), NCDS	Numeric	18558
NCDS2FLG	Flag indicating whether cases with no Father's RG Class 1990 were supplemented with Father's RG Class (1970 version) information	Numeric	18558

**Dataset Name:** NCDS Harmonised Adult SES

Variable	Label	Format	No. Valid Cases
NCDSID	NCDS Case Identifier	String	18558
NCDSSEX	Sex of cohort member	Numeric	18555
NCDS6CL	Cohort member's RG Class (1990 version) at age 42 Sweep, NCDS	Numeric	18558

**Dataset Name:** BCS70 Harmonised Childhood SES

Variable	Label	Format	No. Valid Cases
BCS70	BCS70 Case Identifier	String	14791
BCSSEX	Sex of cohort member	Numeric	14791
BCS3FCL	Father's RG Social Class (1990-version) at age 11 sweep, NCDS	Numeric	14791
BCS3FLG	Flag indicating whether cases with no Father's RG Class (1990 version) were supplemented with Father's RG Class (1980 version) information	Numeric	14791

**Dataset Name:** BCS70 Harmonised Adult SES

<b>Variable</b>	<b>Label</b>	<b>Format</b>	<b>No. Valid Cases</b>
BCS70	BCS70 Case Identifier	String	9,841
BCSSEX	Sex of cohort member	Numeric	9,841
BCS9CLAS	Cohort member's RG Class (1990-version) at age 42 sweep, BCS70	Numeric	9,841

**Dataset Name:** MCS Harmonised Childhood SES

<b>Variable</b>	<b>Label</b>	<b>Format</b>	<b>No. Valid Cases (ie MCS families)</b>
MCSID	MCS Case Identifier	String	13,287
MCS5PSEX	Sex of parent	Numeric	13,287
MCS5PCL	Parent's RG Social Class (1990-version) at age 11 sweep, MCS	Numeric	13,287

**Dataset Name:** NSHD Harmonised Childhood SES

<b>Variable</b>	<b>Label</b>	<b>Format</b>	<b>No. Valid Cases</b>
NSHD_ID	NSHD Case Identifier	String	5362
NSHDSEX	Sex of cohort member	Numeric	5362
NSHDFCL	Father's RG Social Class at age 11 Sweep (1970 version), NSHD	Numeric	5362

**Dataset Name:** NSHD Harmonised Adult SES

<b>Variable</b>	<b>Label</b>	<b>Format</b>	<b>No. Valid Cases</b>
NSHD_ID	NSHD Case Identifier	String	5362
NSHDSEX	Sex of cohort member	Numeric	5362
NSHDCL	Cohort member's RG Class (1990 version) age 43 Sweep, NSHD	Numeric	5362



## Appendix 1: Derivation Code

### Derivation Code

#### Source Files

<b>Study</b>	<b>Age</b>	<b>Data Source</b>
NSHD	11	Provided by study
NCDS	11	<a href="http://doi.org/10.5255/UKDA-SN-5565-2">http://doi.org/10.5255/UKDA-SN-5565-2</a>
	42	<a href="http://doi.org/10.5255/UKDA-SN-5578-1">http://doi.org/10.5255/UKDA-SN-5578-1</a>
BCS70	10	<a href="http://doi.org/10.5255/UKDA-SN-3723-7">http://doi.org/10.5255/UKDA-SN-3723-7</a>
	42	<a href="http://doi.org/10.5255/UKDA-SN-7473-2">http://doi.org/10.5255/UKDA-SN-7473-2</a>
MCS	11	<a href="http://doi.org/10.5255/UKDA-SN-7464-4">http://doi.org/10.5255/UKDA-SN-7464-4</a>

## References

1. Elias P, Halstead K, Prandy K: CASOC: Computer-Assisted Standard Occupational Coding. 1993, London: HMSO
2. Jones R, Elias P. CASCOT: Computer-assisted structured coding tool, Institute for Employment Research. Coventry: University of Warwick; 2005.