Word file converted to Realignment of BCS70 identifiers

This document is to inform researchers that a small number of primary identifiers (BCSID) have been changed to realign them to previous sweeps of data.

Background

Over the course of the 1970 British Cohort Study, 99 cohort members were incorrectly given new primary identifiers. A number of cohort members who had taken part in the birth survey were mistakenly thought to have joined the study at a later sweep and as such were given a new primary identifier, despite the fact they had already been assigned a primary identifier at birth. This resulted in their data in later sweeps becoming de-aligned from earlier data. This mostly occurred in the 1980 (age 10) sweep. In some cases, a wrong identifier was assigned for one or two sweeps, then the correct identifier was re-assigned afterwards. See next page for more details.

Resolution

CLS identified the de-aligned cohort members, partly by retrieving NHS numbers and partly as a result of research carried out by the CLS tracing team. The data files are available via the UK Data Service have now been re-deposited so that all cases have the correct identifier. The number of cases in each sweep where identifiers have been reassigned is shown below:

Sweep	Age at sweep	Number of cases realigned
BCS 1975	5	6
BCS 1980	10	86
BCS 1986	16	48
BCS 1996	26	38
BCS 2000	30	54
BCS 2004	34	36
BCS 2008	38	38
BCS 2012	42	47

Updating existing/transformed data sets

Researchers using British Cohort Study data are advised to replace their existing files with the re-aligned files now available on the archive. Where researchers have created their own datasets, which include derived, recoded or transformed variables, they should be able to apply existing SPSS syntax files, Stata .do files or R scripts to the revised data.

We recognise that this will not always possible. Researchers who need assistance in modifying existing derived datasets, so that all the relevant data is realigned with the correct identifier, should get in touch with the UK Data Service (contact details here). In the majority of cases, users will then be referred to CLS who will be able to re-issue users with a copy of any existing datasets but with the identifiers properly aligned.

Background: Types of de-alignment

A. Case wrongly given new identifier

Nearly all of the de-aligned cases fall into this category. Cohort members (CMs) here have been incorrectly given a new identifier, usually in the 1980 sweep. For these CMs, data from the first two sweeps cannot be added to subsequent ones, because they have been thought to be different people. See example table below.

(In the tables below a tick corresponds to productive data existing for a given identifier in a sweep)

	1970	1975	1980	1986
B1110X (original ID)	✓	✓	✗	✗
B2110X (new ID)	✗	✗	✓	✓

B. Case given new identifier, then recovered original

Cases in this second category were given a new identifier for one or two sweeps, then the CM's original identifier was recovered to their previous identifier in later sweeps. For example (see table below), the new identifier might be added in 1980, but by 1986, the CM was again using the original identifier.

	1970	1975	1980	1986
B1110X (original ID)	✓	✓	✗	✓
B2110X (new ID)	✗	✗	✓	✗

C. Case wrongly given other CM's identifier

In this last category, cohort members were given the identifier of a different cohort member in error, and so had been incorrectly linked to a different CM's data in previous sweeps. Here, the first instance of the duplicate identifier (1970 in the example below) is valid and has no correction to the original identifier. But in later sweeps (1980 onwards in the example), the data for the new identifier should refer to the original cohort member.

	1970	1975	1980	1986
B1110X (original ID)	✓	✓	✗	✗
B2110X (new ID)	✓	✗	✗	✗
B2110X (original CM using new ID)	✗	✗	✓	✓