

Millennium Cohort Study

Linked education administrative
data Scotland

User Guide (Version 1)

June 2026

Contact

Data queries: help@ukdataservice.ac.uk

Questions and feedback about this user guide: clsdata@ucl.ac.uk.

Authors

Karen Dennison, Morag Henderson, Aida Sanchez-Galvez, Andrew Peters.

How to cite this guide

Dennison K., Henderson M., Sanchez-Galvez, A., Peters A. (2026) *Millennium Cohort Study: Linked education administrative data Scotland (Version 1)*. London: UCL Centre for Longitudinal Studies.

Data citation and acknowledgement

You should cite the data and acknowledge CLS following the guidance from [Citing our data page on our website](#).

About UCL Centre for Longitudinal Studies

The [UCL Centre for Longitudinal Studies \(CLS\)](#) is an Economic and Social Research Council (ESRC) Resource Centre. It is home to a unique series of UK national cohort studies. It is part of the [UCL Social Research Institute](#), based at the [UCL Institute of Education](#).

Millennium Cohort Study is core funded by:



This document is available in alternative formats. Please email the Centre for Longitudinal Studies at clsdata@ucl.ac.uk.

Contents

About the Millennium Cohort Study	4
1. Introduction	5
2. Consent to linkage	5
3. Data linkage	6
3.1 External provider datasets	6
3.2 Matching strategy	6
3.3 Matching rates	7
4. Linked Research Data	8
4.1 Licensing and data access	8
4.2 List of datasets	8
4.3 Interpretation of the qualification codes	9
4.4 Identifiers	10
4.5 Variable description	11
4.6 Data processing	12
4.7 Missing values	12
4.8 Data inconsistencies	13
4.9 Data de-identification	13
4.10 Output Disclosure Control (for controlled data only)	13
Appendix 1: Variables included in the NRS datasets	15

About the Millennium Cohort Study

The Millennium Cohort Study (MCS) is a longitudinal birth cohort study, following a nationally representative sample of approximately 19,000 people born in the UK at the turn of the century.

The study has captured rich information about the different aspects of cohort members' lives, from birth to childhood and adolescence, and is continuing to keep up with them now they are adults.

As a multidisciplinary study, MCS is used by researchers working in a wide range of fields.

Findings from MCS have influenced policy at the highest level, and today the study remains a vital source of evidence on the major issues affecting young people's lives.

Further details of the [data available from the main MCS surveys can be found on the Millennium Cohort Study webpage](#) as well as [Exploring our data webpage on the CLS website](#).

1. Introduction

This guide describes the data linkage of education administrative (school) records for Scotland up to age 18 to survey data for cohort members in the Millennium Cohort Study (MCS). The main aim of this data linkage exercise is to enhance the research potential of the study, by combining administrative education records with the rich information collected in the surveys.

2. Consent to linkage

During the fourth (age 7) sweep, the child's parent / carer was asked for consent to link the child's education records up to age 16 to their MCS survey data. During the seventh sweep (age 17) the cohort members, then able to consent in their own right, were asked for consent to link to post-16 education records. Although consent was also sought in the eighth sweep (age 23), the records described in this guide were linked prior to completion of this sweep.

Cohort members can withdraw their or their parents' consent for further linkages at any time – consent withdrawals are applied to data before deposit at the UK Data Service. For consent withdrawals received following deposit of data, no further linkages are enacted for those individuals. Detailed information on the fieldwork and consent collection can be found in the MCS Age 6 and MCS Age 17 Technical reports and their appendices. All documents can be found under 'documentation' at:

- [MCS Age 7 Sweep page](#)
- [MCS Age 17 Sweep page](#)

3. Data linkage

3.1 External provider datasets

CLS requested linked administrative data from the Scottish Government covering school and tertiary education. These datasets provide objective, record-level information on pupils' educational trajectories from the start of primary school through to post-secondary education.

In Scotland, Primary 1 is the first year of compulsory primary education in Scotland, typically attended by children aged 4-5 years. Secondary 6 is the final year of secondary education, typically attended by pupils aged 16-17. Tertiary-level data relate to non-degree post-secondary qualifications, such as those obtained from Further Education colleges.

The following datasets linked to the MCS cohort members were provided by the Scottish Government:

- School Pupil Census – from Primary 1 through to Secondary 6
- Attendance – from Primary 1 through to Secondary 6
- Exclusions – from Primary 1 through to Secondary 6
- Leaver destinations and attainments – Secondary 6/Tertiary
- Qualifications – from Secondary 1 through to Tertiary Level Education

3.2 Matching strategy

In 2024, CLS provided National Records of Scotland (NRS) with a bespoke unique ID and direct identifiers of cohort members with consent for education linkages and who had ever been interviewed in Scotland. The following inclusion criteria were applied to cohort members:

- All those cohort members who gave consent to education data linkage at age 17, where that permission has not been revoked.
- All those cohort members whose parents gave consent to education data linkage at age 7, where that consent has not been revoked, and where there is no record for consent at age 17.

A bespoke unique ID, forename, surname, DOB, postcode and previous postcode were included in the file, as well as two data linkage consent variables –

- Education consent up to age 16 (yes/no)

- Education consent post 16 (yes/no)

NRS matched the identifiers to those that it holds and transferred linkage keys and the two consent variables to the Administrative Data Research-Scotland linkage agent in the National Safe Haven. The Scottish Government linked in the requested education variables, applying the two consent variables so that only age-specific consented education data was included in the linked data. Linked data were returned to CLS with the bespoke unique ID minus any direct identifiers.

3.3 Matching rates

Of the 3,136 cohort members ever interviewed in Scotland. 1,706 were sent to NRS with consent to education data linkage. Of these, 1,605 had consented to the 'education consent up to age 16', and 1,081 to the 'education consent post 16'.

Of the 1,465 cohort members successfully matched to their Scottish education records, 851 were matched for all years, 567 up to age 16 only, and 47 post-16 only. As consent may be withdrawn following the matching process, the number of cases included in the datasets may be slightly lower than these figures.

4. Linked Research Data

4.1 Licensing and data access

The linked education data have been processed by CLS and supplied to the UK Data Service. All data users need to be registered with the UK Data Service and to sign the UKDS End User Licence. Details of how to do this are [available on the UK Data Service website](#).

The linked Scottish education records are available as controlled data due to their potentially disclosive and/or sensitive nature, and must be accessed as controlled data from the UK Data Service SecureLab. Applicants wishing to access this data need to abide by the terms and conditions of the UK Data Service Secure Access licence.

Before gaining access, researchers must make an application detailing the intended analysis and provide a justification as to why this data is requested. [Application guidance can be found on UK Data Service website](#).

Data access will be granted once the form has been reviewed by UK Data Service and approved by the CLS Data Access Committee.

4.2 List of datasets

The majority of datasets included in this deposit are long/hierarchical format, with their own separate secondary ID to identify separate rows per cohort member.

Table 1: List of available datasets and contents

Name of the dataset	Content summary	Identifiers
mcs_nrs_pupil_census_student_restricted	Pupil's school and class situation, with one row per student stage/academic year	mcsid cnum stage_no
mcs_nrs_qualifications_restricted	Individual qualifications (year, names, results, credit points) achieved by pupils	mcsid cnum qual_no
mcs_nrs_leavers_destinations_restricted	Destinations of pupils after leaving school (3 and 9 months on)	mcsid cnum
mcs_nrs_leavers_subjects_restricted	Subjects of pupils after leaving school (3 and 9 months on)	mcsid cnum
mcs_nrs_attendance_restricted	Recorded half days of school attendance and absence (by type) incurred by the pupil each year	mcsid cnum att_no school_no
mcs_nrs_exclusions_restricted	Exclusions incurred by the pupil	mcsid cnum att_no excl_no

For a description of the identifiers please refer to section 4.5.

4.3 Interpretation of the qualification codes

Detailed information on individual qualifications achieved by pupils is available in the mcs_nrs_qualifications_restricted dataset. A summary reference guide for interpreting qualification grades (qual_result) by qualification level (qual_level) can be found in the accompanying document mcs_nrs_qualification_code_guide

(mcs_nrs_qual_level_and_qual_result_breakdown.xlsx). This guide should be consulted when interpreting grade codes, as the same codes are used across all qualification levels but their meaning varies by qualification, which can be a source of confusion in analysis.

4.4 Identifiers

Individual identifiers

MCS cohort members are identified with the research IDs **mcsid**, (family identifier) and **cnum00** (cohort member identifier), which are used for the rest of cohort data available at the UKDS. This enables the data to be merged with one another.

For MCS, researchers need to use both the MCS family identifier (**mcsid**) and the individual person identifier (**cnum00**) to merge on with other cohort data.

From sweep 8 onwards MCS survey data includes a separate household identifier **mcsidcmh**. While newer data can still be merged on original identifiers **mcsid** and **cnum00**, the **mcsidcmh** can be created in the Scottish education datasets by combining **mcsid** with **cnum00**.

Other identifiers

The majority of the Scottish education datasets released to the UKDS are structured in hierarchical (long) format, meaning that each cohort member can have multiple rows, e.g. one per qualification

Each row is uniquely identified by a combination of the cohort member identifier (**mcsid**, **cnum00**) and secondary identifiers depending on the dataset (see Table 2), which increments with each instance and is ordered by academic year. member's record, rows are ordered by academic year of instance. However, within a given academic year, records are not in chronological order, as more precise dates are not available in the data. For example, where a pupil had multiple school absences in the same academic year, the order of those absence records does not reflect the sequence in which they occurred.

Table 2: Secondary identifiers used in the MCS NRS data

Secondary identifier	Description
stage_no	Order of school stage at which student status was reported
qual_no	Qualification order for each reported qualification achieved
att_no	Attendance order for each type of absence per school/academic year
school_no	School identifier for a change in schools attendance was reported at
excl_no	Exclusion order for each exclusion reported

The NRS data do not contain geographic identifiers, or school identifiers such Unique Reference Numbers (URNs) or Local Education Authority (LEA) codes.

Some additional considerations should be made to the structure of the following datasets: The dataset containing the pupils' school and class situation

(mcs_nrs_pupil_census_student_restricted) covers Primary 1 through to Secondary 6, with additional categories for pupils attending beyond school age and those in special education.

Some cohort members have multiple rows per school stage, reflecting several 'student_status' outcomes at different schools within the same academic year. The absence dataset (mcs_nrs_attendance_restricted) contains absence records at the school level. Where a pupil changed school (school_no) during the academic year, multiple rows of the same absence type may appear (one per school attended).

4.5 Variable description

Variable order

Variables are ordered as provided by NRS. Certain datasets have been combined, both for user convenience and preserving individuals' anonymity. The classroom sizes were provided by NRS on school identifiers which have not been included with the deposit; this is instead included with the pupil census data having been combined based on school identifier and student stage (Primary 1-7).

Metadata: variable names, variable labels and value labels

With the exception of MCS identifiers, variable names are as provided by the National Records of Scotland (NRS) and are broadly self-explanatory. Likewise, variable labels are all provided by the NRS, though some have been edited with supplementary detail to add context where needed.

See Appendix 1 for a list of variables included in the NRS datasets.

Assigned values and value labels have been provided by NRS, except for the unique identifier variables created by CLS.

4.6 Data processing

Several modifications have been made to the datasets supplied by NRS to make them research-ready:

- 1) Dataset identifiers have been converted to research identifiers to protect participant confidentiality, and additional identifiers created where needed.
- 2) Data originally held at school and class level have been restructured to student level, allowing users to view individual class information without exposing school identifiers.
- 3) Exclusion records have been expanded to incorporate alternative provision data (i.e. details of alternative educational settings attended during a period of exclusion).
- 4) Qualification type details have been added to the student qualifications dataset to aid interpretation without the need to cross-reference separate lookup files.

4.7 Missing values

No specific missing values were included with the data provided by NRS. For some variables and where datasets were merged, appropriate missing values were assigned by CLS, as follows:

- -8 (No information) for fields where data were missing but expected
- -1 (Not applicable) for fields where data would not be expected

4.8 Data inconsistencies

In the absence dataset (mcs_nrs_attendance), the student_stage and student_status variables contain missing data for a number of students in the 2018/19 academic year. This is a known limitation of the data as supplied by the NRS and cannot be resolved.

The years for which data are available vary depending on the consent provided by each cohort member (see Section 3.2 for full details). Consent level is not explicitly flagged at pupil level, so researchers cannot directly identify which consent tier applies to a given individual. In practice, data will be absent for pre- and post-16 periods across all datasets for cohort members who did not consent at both age 7 and age 17. Researchers conducting longitudinal analyses spanning different age ranges should therefore be aware that missingness may reflect consent status rather than non-participation or data loss.

4.9 Data de-identification

CLS is committed to protect research participants' rights and avoid data disclosure and re-identification of individuals.

Several measures have been taken to protect the confidentiality of cohort members within the Scottish education records data. School-level geographic identifiers — such as Local Education Authority (LEA) codes and Unique Reference Numbers (URNs) — have been removed to prevent identification of the specific schools attended by cohort members.

Other potentially identifying details, including class composition of mixed-stage classrooms, first language spoken of all school pupils, and precise dates of exclusions or absences, have been withheld.

4.10 Output Disclosure Control (for controlled data only)

UK Data Service requirements

The linked education data is only available via the UK Data Service Secure Lab. Access to this data is controlled and the UK Data Service will always perform a certain level of disclosure control on the outputs generated by researchers, as outlined in their [SDC Handbook which can be downloaded from their website](#).

The two UK Data Service Secure Lab rules of thumb that will be applied to all outputs are:

- Threshold rule: No cells should contain less than 10 observations
- Dominance rule: No observation should dominate the data to a huge extent

Appendix 1: Variables included in the NRS datasets

Table A1: Variables included in mcs_nrs_pupil_census_student_restricted

Variable name	Variable label
stage_no	Stage number
ac_year	Academic Year of Census
student_stage	Student Stage
student_status	Student Status
school_funding_type	School funding type
csp_indicator	Coordinated Support Plan Indicator
iep_indicator	Individualised Education Programme (IEP) Indicator
ron_indicator	Records of Needs (RoN) Indicator
sen_indicator	Special Educational Need indicator
mainstream_integration	Number of half-days (0-10) per week of Student Mainstream Integration (for all pupils in Primary, Secondary, and Special Schools)
attendance_ssu	Student Attendance in half-days (0-10) at Special Schools/Units (for all pupils in Primary, Secondary, and Special Schools)
mode_of_attendance	Mode of Attendance
curriculum_adaptation	Access to Curriculum adaptation Required
communication_adaptation	Access to Communication adaptation Required
physical_adaptation	Access to Physical adaptation Required
free_school_meal_registered	Free School Meal - Registered
gaelic_education	Gaelic Education
level_of_english	Level of English as an additional language
two_or_more_teachers	Identify classes with 2 or more teachers present at all times
total_pupils	Total pupils in class
class_type	Mixed or single stage class

Table A2: Variables included in mcs_nrs_qualifications_restricted

Variable name	Variable label
qual_no	Qualification number
ac_year	School year that data relates to
age	Age at last birthday taken from 28/2 of academic year
entry_stage	Stage at which the student is taking the qualification
qual_code	Code denoting the specific course or unit taken by the student
qual_name	The name of the qualification
qual_product_type	This shows whether the qualification is a course or a unit, as well as being a broad description of the type of award.
qual_data_provider	The provider of the qualifications data
qual_entry_status	Status of pupil's entry for the qualification
qual_level	The level of the course or unit taken by the student. Higher level represents a higher level of attainment
qual_result	Result achieved for the course or unit
scqf_credit_points	The number of SCQF credit points awarded for the result at the level of the particular course or unit taken

Table A3: Variables included in mcs_nrs_leavers_destinations_restricted

Variable name	Variable label
ac_year	School year data relates to
simd_2016_decile	The Scottish Index of Multiple Deprivation is a relative measure of deprivation
dest_org	Initial destination, 3 months after leaving school
sdssldr_status	Initial destination, 3 months after leaving school. Sub-category of "dest_org"
foll_org	Follow-up destination, 9 months after leaving school
sdssldr_flwp_status	Follow-up destination, 9 months after leaving school. Sub-category of "foll_org"
student_stage	Stage at which a student left school

Variable name	Variable label
scqf2_n	Number of National Qualifications attained at SCQF Level 2 or better
scqf3_n	Number of National Qualifications attained at SCQF Level 3 or better
scqf4_n	Number of National Qualifications attained at SCQF Level 4 or better
scqf5_n	Number of National Qualifications attained at SCQF Level 5 or better
scqf6_n	Number of National Qualifications attained at SCQF Level 6 or better
scqf7_n	Number of National Qualifications attained at SCQF Level 7
lev2ge1	Pupil achieved one or more passes at SCQF Level 2 or better in National Qualifications
lev3ge1	Pupil achieved one or more passes at SCQF Level 3 or better in National Qualifications
lev3ge5	Pupil achieved five or more passes at SCQF Level 3 or better in National Qualifications
lev4ge1	Pupil achieved one or more passes at SCQF Level 4 or better in National Qualifications
lev4ge5	Pupil achieved five or more passes at SCQF Level 4 or better in National Qualifications
lev5ge1	Pupil achieved one or more passes at SCQF Level 5 or better in National Qualifications
lev5ge5	Pupil achieved five or more passes at SCQF Level 5 or better in National Qualifications
lev6ge1	Pupil achieved one or more passes at SCQF Level 6 or better in National Qualifications
lev6ge3	Pupil achieved three or more passes at SCQF Level 6 or better in National Qualifications
lev6ge5	Pupil achieved five or more passes at SCQF Level 6 or better in National Qualifications
lev7ge1	Pupil achieved one or more passes at SCQF Level 7 in National Qualifications
lev7ge3	Pupil achieved three or more passes at SCQF Level 7 in National Qualifications

Variable name	Variable label
bestlevel	Highest SCQF Level achieved in National Qualifications
num_scqf6	Numeracy qualifications attained at SCQF Level 6 or better
num_scqf5	Numeracy qualifications attained at SCQF Level 5 or better
num_scqf4	Numeracy qualifications attained at SCQF Level 4 or better
num_scqf3	Numeracy qualifications attained at SCQF Level 3 or better
lit_scqf6	Literacy qualifications attained at SCQF Level 6 or better
lit_scqf5	Literacy qualifications attained at SCQF Level 5 or better
lit_scqf4	Literacy qualifications attained at SCQF Level 4 or better
lit_scqf3	Literacy qualifications attained at SCQF Level 3 or better
dyw_scqf7	Number of Developing Young Workforce qualifications attained at SCQF Level 7
dyw_scqf6	Number of Developing Young Workforce qualifications attained at SCQF Level 6 or better
dyw_scqf5	Number of Developing Young Workforce qualifications attained at SCQF Level 5 or better
dyw_scqf4	Number of Developing Young Workforce qualifications attained at SCQF Level 4 or better
dyw_scqf3	Number of Developing Young Workforce qualifications attained at SCQF Level 3 or better
dyw_scqf2	Number of Developing Young Workforce qualifications attained at SCQF Level 2 or better
dyw_lev2ge1	Pupil achieved one or more passes in DYW qualifications at SCQF Level 2 or better
dyw_lev3ge1	Pupil achieved one or more passes in DYW qualifications at SCQF Level 3 or better
dyw_lev3ge5	Pupil achieved five or more passes in DYW qualifications at SCQF Level 3 or better
dyw_lev4ge1	Pupil achieved one or more passes in DYW qualifications at SCQF Level 4 or better

Variable name	Variable label
dyw_lev4ge5	Pupil achieved five or more passes in DYW qualifications at SCQF Level 4 or better
dyw_lev5ge1	Pupil achieved one or more passes in DYW qualifications at SCQF Level 5 or better
dyw_lev5ge5	Pupil achieved five or more passes in DWY qualifications at SCQF Level 5 or better
dyw_lev6ge1	Pupil achieved one or more passes in DWY qualifications at SCQF Level 6 or better
dyw_lev6ge3	Pupil achieved three or more passes in DWY qualifications at SCQF Level 6 or better
dyw_lev6ge5	Pupil achieved five or more passes in DWY qualifications at SCQF Level 6 or better
dyw_lev7ge1	Pupil achieved one or more passes in DWY qualifications at SCQF Level 7
dyw_lev7ge3	Pupil achieved three or more passes in DWY qualifications at SCQF Level 7

Table A4: Variables included in mcs_nrs_leavers_subjects_restricted

Variable name	Variable label
ac_year	School year that data relates to
qual_code	Code denoting the specific course or unit taken by the student
qual_name	Qualification name
qual_level	SQA qualification level code
qual_result	Grade achieved (Passing results only)
qual_product_type	Qualification product group
scqf_level	SCQF level for the qualification
subject	Subject code
name_grouped	Group subject in the Scottish Government publication
student_stage	Stage at which a student left school
scqf2	National Qualifications attained at SCQF Level 2 or better

Variable name	Variable label
scqf3	National Qualifications attained at SCQF Level 3 or better
scqf4	National Qualifications attained at SCQF Level 4 or better
scqf5	National Qualifications attained at SCQF Level 5 or better
scqf6	National Qualifications attained at SCQF Level 6 or better
scqf7	National Qualifications attained at SCQF Level 7

Table A5: Variables included in mcs_nrs_attendance_restricted

Variable name	Variable label
att_no	Attendance number
school_no	Change in school attendance reported for
ac_year	Academic Year of Census
code	Attendance code
half_days	Number of half-day absences due to absence reason
student_stage	Student stage in school career
student_status	Description of pupil status on school roll

Table A6: Variables included in mcs_nrs_exclusions_restricted

Variable name	Variable label
excl_no	Exclusion number
ac_year	Academic year of exclusion
excl_mth	Month exclusion started
excl_yr	Year exclusion started
removed_from_register	Removed from the Register
length_of_exclusion	Length of exclusions in half-days
appeal	Student Exclusion Appeal
no_provision_days	Number of half days with no alternative provision of education
interim_alt_provision	Interim Alternative Provision found at...