CLS-DAC-approved use of NHS Hospital Episode Statistics (HES) data linked to CLS cohort studies

Use of HES data linked to CLS cohort data

The NHS Hospital Episode Statistics (HES) data comprise Admitted Patient Care, Outpatient, Critical Care, and Accident & Emergency. CLS have obtained permission from NHS England to link HES data to the four CLS cohort studies: NCDS, BCS70, Next Steps, and MCS.

The CLS aims agreed with NHS England are the following:

1. **CLS data management** to process, store, and generate a useful linked HES data resource linked to the CLS cohort data. These processed linked HES data are suitably pseudonymised and de-identified. This work is carried out within the secure UCL server used by CLS (UCL Data Safe Haven).

2. **CLS methodological work** to describe and assess the quality of the processed linked HES data and/or their benefit to health and social care, in order to facilitate onward sharing via user training, publication of methodological papers, and promotion of policy-related research. This work is carried out within the secure UCL server used by CLS (UCL Data Safe Haven).

3. **Onward sharing via the UK Data Service Secure Lab** to enable the safe sharing of the processed linked HES data with the research community where they can be analysed in combination with the cohort survey data. Researchers, including CLS staff, who want to use the data for research purposes must apply to the UK Data Service.

   The applications are reviewed by the CLS Data Access Committee (DAC) which is an independent committee that ensures that the information is used responsibly and safely. Researchers will only be given permission to use the data if they present a strong scientific case and explain the potential impact of the research and its wider value to society.

   In addition to the standard UKDS SecureLab licence, access to these linked HES data is governed by a sub-licence between the applicant's institution and UCL, as agreed by NHS England.

CLS methodological projects (DAC-approved)

CLS DAC ref.	DAC / 113
UKDS project ref.	N/a (access via the UCL DSH)
Project title	Linkage of National Longitudinal Cohort Studies and Administrative Data: A Mutually Beneficial Arrangement
Institution	Centre for Longitudinal Studies (University College London)
Cohort(s)	NCDS and Next Steps
Plain text abstract	The overall aim of this project is to address several methodological issues at the interface between linked cohort data and administrative data that are mutually beneficial for both data sources: missing data, data quality, residual confounding. This will be undertaken using existing data linkages in health and education with UCL Centre for Longitudinal Studies (CLS) cohorts. We will address the following three methodological research questions, which will be directly reflected in the project work packages: 1. How can linked administrative data aid the handling of missing cohort data? 2. How can linked cohort data improve our understanding of the quality of administrative data? 3. How can linked cohort data help address residual confounding in analyses of administrative data? The project will utilise data across two of the world-leading national longitudinal cohort studies run by CLS (1958 National Child Development Study and Next Steps) linked to three sources of administrative data (National Pupil Database, Hospital Episode Statistics and Higher Education Statistics Agency data)
Data requested	Selected variables from

	- SN 8697 [National Child Development Study: Linked Health Administrative Datasets (Hospital Episode Statistics), England, 1997-2017: Secure Access](#) - SN 8681 [Next Steps: Linked Health Administrative Datasets (Hospital Episode Statistics), England, 1997-2017: Secure Access](#)
Date received	1 December 2020
Date approved	3 March 2021
Research outputs	*Working papers* - Silverwood RJ, Rajah N, Calderwood L, De Stavola BL, Harron K, Ploubidis GB. Examining the quality and sample representativeness of linked survey and administrative data: linking the 1958 National Child Development Study to Hospital Episode Statistics data. CLS Working Paper 2022/5. London: UCL Centre for Longitudinal Studies; 2022. [https://cls.ucl.ac.uk/wp-content/uploads/2017/02/Examining-the-quality-and-sample-representativeness-of-linked-survey-and-administrative-data-CLS-Working-Paper-2022-5.pdf](#) - Rajah N, Calderwood L, De Stavola BL, Harron K, Ploubidis GB, Silverwood RJ. Using linked Hospital Episode Statistics data to better handle non-response and restore sample representativeness in the National Child Development Study. CLS Working Paper 2023/1. London: UCL Centre for Longitudinal Studies; 2022. [https://cls.ucl.ac.uk/wp-content/uploads/2023/02/CLS-Working-Papers-2023-1-Using-linked-Hospital-Episode-Statistics-data-to-aid-the-handling-of-non-response.pdf](#) *Journal articles* - Silverwood RJ, Rajah N, Calderwood L, De Stavola BL, Harron K, Ploubidis GB. Examining the quality and target population representativeness of linked survey and administrative data: guidance and illustration using linked 1958 National Child Development Study and Hospital Episode Statistics data. Under review at Int J Popul Data Sci. - Rajah N, Calderwood L, De Stavola BL, Harron K, Ploubidis GB, Silverwood RJ. Using linked Hospital Episode Statistics

data to better handle non-response and restore sample representativeness in the National Child Development Study. Under review at BMC Medical Res Methodol.

Presentations

- Silverwood RJ. Using linked Hospital Episode Statistics data to aid the handling of missing cohort data. St George's, University of London, Population Health Research Institute Seminar (Online, March 2022)

- Rajah N. Using linked Hospital Episode Statistics data to aid the handling of missing cohort data. Health Studies User Conference (Online, July 2022)

- Silverwood RJ. Examining the quality and sample representativeness of linked 1958 National Child Development Study and Hospital Episode Statistics data. International Population Data Linkage Network Conference (Edinburgh, September 2022)

- Rajah N. Using linked Hospital Episode Statistics data to aid the handling of missing cohort data. International Population Data Linkage Network Conference (Edinburgh, September 2022)

- Silverwood RJ. Examining the quality and sample representativeness of linked 1958 National Child Development Study and Hospital Episode Statistics data. Royal Statistical Society International Conference (Aberdeen, September 2022)

- Silverwood RJ. Using linked Hospital Episode Statistics data to aid the handling of missing cohort data. Royal Statistical Society International Conference (Aberdeen, September 2022)

- Two abstracts also submitted for European Survey Research Association Conference (Milan, July 2023)

Published abstracts

- Silverwood RJ, Rajah N, Calderwood L, De Stavola BL, Harron K, Ploubidis GB. Examining the quality and sample representativeness of linked survey and administrative data: linking the 1958 National Child Development Study to Hospital Episode Statistics data. Int J Popul Data Sci, 2022; 7(3): doi: 10.23889/ijpds.v7i3.1990. https://ijpds.org/article/view/1990

- Rajah N, Calderwood L, De Stavola BL, Harron K, Ploubidis GB, Silverwood RJ. Using linked Hospital Episode Statistics data to better handle non-response and restore sample

representativeness in the National Child Development Study. Int J Popul Data Sci, 2022; 7(3): doi: 10.23889/ijpds.v7i3.1997. https://ijpds.org/article/view/1997

Training and capacity building

- Silverwood RJ. Analysing linked HES data. Part of "Using linked administrative data: Hospital Episode Statistics linked with the CLS cohorts" UCL Centre for Longitudinal Studies webinar (Online, December 2022)

CLS DAC ref.	DAC / 141
UKDS project ref.	N/a (access via the UCL DSH)
Project title	Analysis of Mental Health in Young People with Linked Data
Institution	Centre for Longitudinal Studies (University College London)
Cohort(s)	Next Steps
Plain text abstract	There has been increasing policy focus on mental health in recent years given the rising prevalence of anxiety and psychological distress, and the COVID-19 pandemic has brought it into even sharper focus than ever before. In analysing whether mental health services are reaching those most in need, this project has the potential to transform how services are best targeted. The applicants will liaise closely with policymakers to ensure findings reach those designing and delivering services. In providing an in-depth understanding of the changes in mental health in young people over time, and exactly how services are meeting their needs, this work will enable policymakers to better plan and allocate healthcare resources for this generation. The impact on society will be far reaching, with the potential to benefit the lives of several thousands of individuals and their families.
Data requested	Selected variables from:

	• SN 8681 [Next Steps: Linked Health Administrative Datasets (Hospital Episode Statistics), England, 1997-2017: Secure Access](#)
Date received	28 January 2022
Date approved	10 February 2022
Research outputs	There are no outputs associated with this project.

CLS DAC ref.	DAC / 159
UKDS project ref.	N/a (access via the UCL DSH)
Project title	Derived variables of mental health-related health services use in NCDS and BCS70
Institution	Centre for Longitudinal Studies (University College London)
Cohort(s)	NCDS and BCS70
Plain text abstract	Although early life socioeconomic indicators and adverse childhood experiences (ACEs) are related to poorer mental health outcomes during adulthood, the extent to which they are related to mental health hospitalisations in adulthood has not been yet studied. Data linkages between the British birth cohorts and data on hospital admissions and treatment length provide a unique opportunity to investigate this and other relevant research questions. However, variables capturing the cohort members' health services that are user-friendly for the scientific community are not readily available in the linked datasets but rather have to be derived. The aim of this project is precisely to derive such user-friendly variables with the main objective of adding them to the CLS research data resource for wide data sharing to the scientific community. The value of these variables for health research will be exemplified by exploring their relationship with early life socioeconomic indicators and ACEs.

Data requested	BCS and NCDS variables.
Date received	1 November 2022
Date approved	16 March 2023
Research outputs	There are no outputs associated with this project.

Research projects for access via the UK Data Service (DAC-approved)

CLS DAC ref.	DAC / 119
UKDS project ref.	203391
Project title	Socioeconomic inequalities in perinatal and maternal health care access
Institution	Imperial College London
Cohort(s)	Next Steps
Plain text abstract	In this paper, I will investigate the socioeconomic health inequalities in children's health and access to maternal healthcare. The time from conception to the age of 2 is a critical time for development and can impact physical health, mental health and opportunity throughout life. Health inequalities can be observed in early life as unfavourable birth outcomes. preterm birth, with substantially higher rates among women from more deprived areas both in the UK (de Graaf et al. (2013)). Perinatal mental health problems carry a total economic and social long-term cost to society of about £8.1 billion for each one-year cohort of births in the UK.
Data requested	Selected variables from the following NHS Hospital Episode Statistics datasets: Admitted Patient Care, Outpatient, Critical Care, and Accident & Emergency. These datasets are pseudonymised, de-identified, and available via the UK Data Service Secure Lab, as follows: • SN 8681 [Next Steps: Linked Health Administrative Datasets (Hospital Episode Statistics), England, 1997-2017: Secure Access](#)
Date received	21 June 2021
Date approved	13 July 2021

Research outputs	PhD thesis is completed and in embargo for publication since the data user wants first to publish the results in scientific journals. The outcomes of the project are part of the PhD, but as mentioned, not yet publicly available.

CLS DAC ref.	DAC / 137
UKDS project ref.	204521
Project title	Health Outcomes of Care Experienced Young People
Institution	University of Edinburgh
Cohort(s)	Next Steps
Plain text abstract	This project aims to look at what experiences affect the mental health of care experienced people. Understanding this is important given the high rates of poor mental health amongst this population compared to their non-care counterparts. Statistical tests shall be run to assess whether or not key experiences throughout the life-course e.g., parental socio-economic status, care placement, length of time in care, poverty, housing, etc. are associated with mental health outcomes. This includes an assessment of their positive and negative effects at different ages. Having an awareness of what affects the mental health of care experienced people will contribute towards our understanding of what can be done to effectively support this population. Specifically, it will allow us to develop policies, services and practices that are best placed to both support positive outcomes and protect against poor mental health for this population.
Data requested	Selected variables from • SN 8681 Next Steps: Linked Health Administrative Datasets (Hospital Episode Statistics), England, 1997-2017: Secure Access

Date received	29 October 2021
Date approved	4 October 2022
Research outputs	There are no outputs associated with this project.

CLS DAC ref.	DAC / 166
UKDS project ref.	209167
Project title	Young people's barriers to mental health services
Institution	University of Southampton
Cohort(s)	NCDS and Next Steps
Plain text abstract	On any given day in the United Kingdom 28.5% of children/young people aged 5 to 19 experience mental health problems. Only one in four of these children/young people receive formal support for these problems. There is a lack of knowledge of what happens to those young people not receiving mental health services. Mental health problems have been shown to limit economic, vocational, and social functioning and international studies have found that 50 to 70% of young people who receive services for their mental health problems continue to experience these problems in adulthood.
	In this current study we would like to explore the barriers experienced by young people in obtaining mental health services. Our aim is to use a developmental epidemiological perspective to:
	a) Explore characteristics (demographic, onset/development/progression mental health problems, environment) of those young people not receiving professional services for their mental health problems and determine if they have access to informal support.

	b) Examine resilience of young people not receiving mental health services for their problems during unpredictable challenging times (COVID pandemic)
	c) Employ a statistical technique that is novel to the field of mental health research to adjust for missing data (Missing-Not-At-Random – NMAR modelling)
Data requested	Selected variables from: • SN 5545 – Next Steps: Sweeps 1-8, 2004-2016 • SN 8681 – Next Steps: Linked Health Administrative Datasets (Hospital Episode Statistics), England, 1997-2017: Secure Access • SN 8658 – COVID-19 Survey in Five National Longitudinal Cohort Studies: Millennium Cohort Study, Next Steps, 1970 British Cohort Study and 1958 National Child Development Study, 2020-2021
Date received	3 February 2023
Date approved	10 March 2023
Research outputs	There are no outputs associated with this project.

CLS DAC ref.	DAC / 169
UKDS project ref.	233913
Project title	Physical activity, sedentary behaviour, and diet: impact on cardiometabolic and women's health
Institution	University College London (UCL)
Cohort(s)	NCDS and BCS70

| **Plain text abstract** | Many different health outcomes are influenced by lifestyle behaviours such as physical activity, sleep, and diet. Our research focuses on two main health clusters: cardiovascular related health and women's health. Lifestyle behaviours, that are often interlinked and influenced by the social environment, may have both synergistic and specific effects on health across the life course in ways we do not completely understand.

Our main aim of the proposed project will be to understand how lifestyle behaviours (physical activity, sleep, diet, alcohol, smoking) as well as specific components of each behaviour (e.g., activity type, duration, frequency, context, intensity) across the life course influence HES-data derived health outcomes including diagnoses and healthcare use. We have two main work packages in this project. First, we intend to examine cardiometabolic related conditions (diabetes, stroke, heart disease etc) and second, we intend to examine outcomes related to women's health. Cohort data on women's health is traditionally poor quality due to a greater emphasis being placed on questions relevant to both sexes and due to a historical lack of understanding of the importance of these outcomes. Within each of these, we can calculate costs associated with each HES activity using HRG grouper assigning NHS tariffs to better understand how behavioural factors such as physical activity, sedentary behaviour and diet contribute to health care costs.

We hypotheses that greater physical activity, less sedentary time, never smoking, better diet parameters, and lower alcohol consumptions will be associated with positive health outcomes. However, our aim is to quantify specific components so that we can better identify and make recommendations about health behaviours beyond what is currently known. Of particularly note, we are keen to understand how different accumulation via bouts of activity or sedentary behaviours impact health. We also want to understand what type of activity and diet may positively impact different components of menstrual health including menopause and more severe experiences across adulthood.

These findings will have direct benefits for society as it allows us to identify specific components of health behaviours that improve health, whilst having the robustness of HES and SMR data. We employ several complex epidemiological techniques such as compositional data analysis, random effects model, mediation |

	analysis, etc and have a cluster of questions with which we intend to explore the data.
Data requested	• SN 8697 – National Child Development Study: Linked Health Administrative Datasets (Hospital Episode Statistics), England, 1997-2017: Secure Access • SN 8733 – 1970 British Cohort Study: Linked Health Administrative Datasets (Hospital Episode Statistics), England, 1997-2017: Secure Access
Date received	2 March 2023
Date approved	20 April 2023
Research outputs	There are no outputs associated with this project.

CLS DAC ref.	DAC / 176
UKDS project ref.	229440
Project title	Hospital use over the life course
Institution	University College London (UCL)
Cohort(s)	NCDS and BCS70
Plain text abstract	Frailty is a valuable concept, widely used to inform and guide the clinical care of older people. Still, its relevance in people aged less than 65 is not clear. This research examines different approaches to frailty measurement that could be applied across the NHS to see if they usefully identify the risk of future poor outcomes. We will use nationally representative studies following people since birth ('birth cohorts') to make a 'frailty index'. This is a means of measuring the level of frailty and can be applied to different patient groups and over time. Frailty indices work by identifying common factors associated with frailty and determining how many are

	present in any individual. This gives rise to a frailty index, which has been shown to link to adverse outcomes in studies of older - and some younger - people. Because data have been collected since birth, we can explore the life course factors that may influence the development of frailty at any age of adulthood. In parallel, we will test a Hospital Frailty Risk Score (HFRS). This can be automatically created from NHS electronic records for all people admitted to a hospital in England. We will assess the ability of this risk score to predict how long people stay in hospital and whether they die. We will create a frailty index and an HFRS in the same individuals through linked birth cohort and NHS electronic records. We can then see how the two different approaches to frailty measurement compare for hospital outcomes such as length of stay and death, but also outcomes such as quality of life, physical or cognitive function. Once we understand the factors associated with frailty in younger adults, we can think about treatments that could be applied over the life course to slow or prevent the development of frailty. To inform and support these analyses, we will engage with lay and professional stakeholders with an interest in frailty at younger ages. Together we will consider which clinical outcomes are meaningful to younger people with a range of long-term conditions. Then, using their perspectives on experiencing frailty earlier in adulthood, we will explore the implications of the emerging findings.
Data requested	• ICD-10 codes and hospital outcomes • SN 8697 – National Child Development Study: Linked Health Administrative Datasets (Hospital Episode Statistics), England, 1997-2017: Secure Access • SN 8733 – 1970 British Cohort Study: Linked Health Administrative Datasets (Hospital Episode Statistics), England, 1997-2017: Secure Access
Date received	5 June 2023
Date approved	13 July 2023
Research outputs	There are no outputs associated with this project.

CLS DAC ref.	DAC / 178
UKDS project ref.	236210
Project title	Examining structural racism in the processes of identifying mental health needs for children and young people in the UK
Institution	University College London (UCL)
Cohort(s)	MCS
Plain text abstract	This project aims to examine inequalities in the identification of mental health difficulties for children and young people in the UK. The project aims to address the gap in research on mental health difficulties experienced by racially minoritised children and young people, specifically, exploring whether there are systematic differences across racial-ethnic groups in which children and young people are identified with mental health difficulties within administrative data systems.
Data requested	• SN 9030 – Millennium Cohort Study: Linked Health Administrative Datasets (Hospital Episode Statistics), England, 2000-2019: Secure Access • SN 8481 – Millennium Cohort Study: Linked Education Administrative Datasets (National Pupil Database), England: Secure Access
Date received	8 August 2023
Date approved	13 September 2023
Research outputs	There are no outputs associated with this project.