International Journal of Population Data Science

Kids' Environment and Health Cohort: Database Protocol

Selin Akaraci[1], Alison Macfarlane[2], Amal Rammah[1], Emilie Courtin[3], Esther Lewis[4], Faith Miller[1], Jason Powell-Bavester[5], Jessica Mitchell[6], Joana Cruz[1], Matthew Lilliman[1], Niloofar Shoari[1], Samantha Hajna[1,7,8], Steven Cummins[9], Tolu Adedire[10], Vahe Nafilyan[11], and Pia Hardelid[1]

[1]Population, Policy and Practice Research and Teaching Department UCL Great Ormond Street Institute of Child Health, University College London (UCL), London, UK
[2]City University of London, London, UK
[3]Health Inequalities Lab, Department of Health Policy, London School of Economics and Political Science, London, UK
[4]Office for National Statistics, London, UK
[5]Office for National Statistics, London, UK
[6]Wolfson Institute of Population Health, Queen Mary University of London, London, UK
[7]Department of Health Sciences, Faculty of Applied Health Sciences, Brock University, St Catharines, ON, Canada
[8]MRC Epidemiology Unit, University of Cambridge, School of Clinical Medicine, Cambridge, UK
[9]Population Health Innovation Lab, Department of Public Health, Environments & Society, London School of Hygiene and Tropical Medicine, London, UK
[10]Office for National Statistics, Darlington, UK
[11]Office for National Statistics, Newport, UK

Abstract

Introduction
Environmental exposures are known to affect the health and well-being of populations throughout the life course. Children are particularly susceptible to environmental impacts on educational and health outcomes as they spend more time in their local environments compared to adults. In England, no national, longitudinal dataset linking information about the physical and social environment in and around homes and schools to children's health and education outcomes currently exists. This limits our understanding of how environments might impact the health and well-being of children as they grow up.

Objective
To establish the Kids' Environment and Health Cohort, a research-ready, de-identified and annually updated national birth cohort of all children born in England from 2006 onwards.

Methods
The Kids' Environment and Health Cohort will link birth and mortality records, health and educational attainment datasets, to maternal health (up to 12 months prior to their child's birth), and environmental data for all children born in England from 2006 – approximately 11 million children at first build. A subset of children born between 2010 and 2012, and between 2020 and 2022 will be linked to their mothers' 2011 or 2021 Census records, respectively. The cohort database will be held in, and accessed via, a trusted research environment (TRE) at the Office for National Statistics (ONS). All geographical identifiers in the cohort, allowing for linkage to further environmental data, will be securely held by the ONS, separately to the main cohort, and will be encrypted before being shared with researchers.

Conclusion
The Kids' Environment and Health Cohort will, for the first time, link administrative health and education data to longitudinal environmental exposures for children at national level in England. It will serve as a data resource to support research about the health and well-being of children via improved home and school environments.

Keywords
Paediatrics; child health; environmental factors, social determinants of health; administrative data linkage

*Corresponding Author:
Email Address: p.hardelid@ucl.ac.uk (Pia Hardelid)

Introduction

Children spend more time in their local neighbourhoods around homes and schools compared to adults [1, 2]. They are also more susceptible to environmental pollution and other exposures compared to adults due to their ongoing physical and cognitive development [3]. There is evidence that a wide range of child health and development outcomes (e.g., respiratory infection risk, cognitive development, educational attainment) are associated with features of the environments in which children live, play, and study e.g., air quality, noise, road traffic volumes, access to green and blue spaces, availability of local services (such as services (such as nurseries, schools and shops), and accessibility of fast food and alcohol outlets [4, 5]. These associations may persist across the life course [6], suggesting that supporting children's access to healthy home and school environments in childhood and adolescence may be critical for promoting positive long-term health and development outcomes.

Multiple reviews [7–11] examining the association between features of the home and school environments and children's health outcomes have reported a lack of longitudinal studies, a heavy reliance on self-reported measures of environmental exposures, and a paucity of research on potential effect modifiers (e.g., age, gender, underlying chronic conditions and family socioeconomic circumstances). Addressing these limitations is necessary to inform novel intervention strategies and to assess the impact of local and national policies aimed at improving children's health and well-being via the improvement of home and school environments.

The aim of the Kids' Environment and Health Cohort (KEHC) is to create a research-ready, de-identified, national birth cohort of all children born in England from 2006 until December 2023, to be updated annually thereafter. Children will be followed via linked administrative health and education data from pregnancy through birth and into adolescence. Children's administrative data in KEHC will be linked to their mothers' Census data, and geospatial data on environmental exposures via national health address registries. KEHC will be made available to researchers interested in studying how local environments in and around homes and schools affect children's health and education from birth to early adulthood. The cohort will help generate robust cross-sectional and longitudinal evidence on the impacts of the physical and social environment on child health and education. This evidence can in turn inform policy changes aimed at improving child health via improvements in their living and educational environments.

Several other data resources are regularly used to study associations between environmental exposures and health and/or education outcomes in children in the UK (Table 1). KEHC will complement these existing resources by enabling environmental child health research at the national scale for the first time in England; similar to what is already possible in Wales [12] and Scotland [13]. England has over 570,000 births per year [14]. KEHC will include information on a large and unselected national population of children. This will allow for the study of environmental impacts on uncommon outcomes and on the role of environments on the health of children at high-risk for adverse health outcomes, including those with long-term conditions, born prematurely, or from migrant or ethnic minority backgrounds.

Methods

Database design

The spine of KEHC will consist of linked Office for National Statistics (ONS) birth [28] and death registration [29] and National Health Service (NHS) birth notification data [30] on all children born in England since 1st January 2006 until December 2023 at first build. This spine will then be linked to administrative health and education data from multiple data providers as well as small-area level (property and postcode) environmental data (Table 2). Children will be linked to their mother's hospital and address records up to a year before their birth to allow maternal environmental exposures and health during pregnancy to be taken into account; children will also be linked to their siblings in the cohort via their mother's records. Finally, for a subset of children born between 2010 and 2012, and between 2020 and 2022, their records will additionally be linked to their mothers' 2011 or 2021 Census records, respectively.
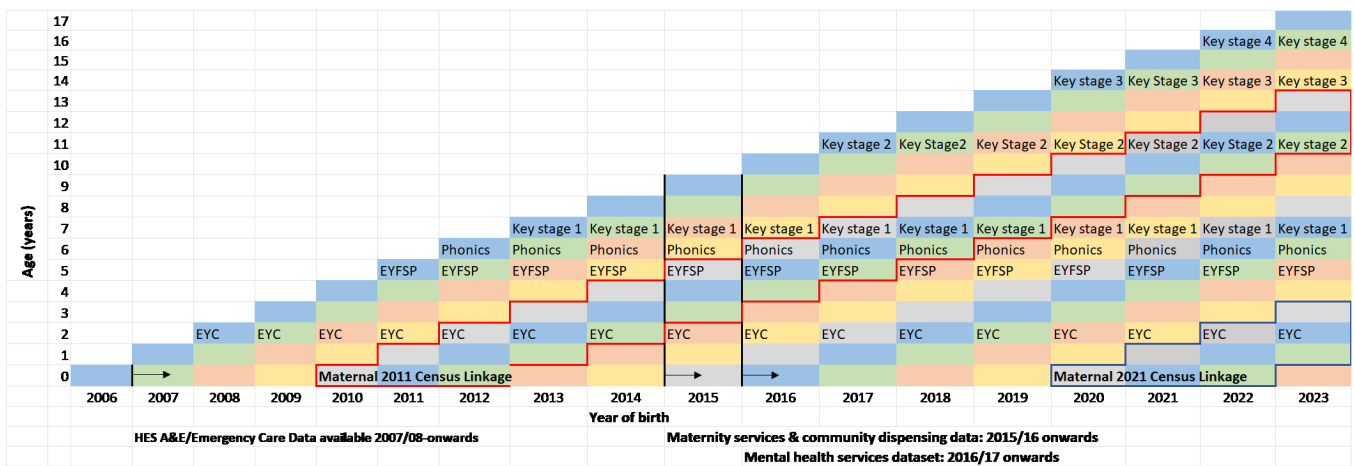
Population and follow-up

With approximately 570,000 live births occurring in England annually, the total cohort will include data for around 10.5 million children born between 2006 and 2023 (the first build of the database). In England, regardless of where a birth occurs, parents are legally required to register their child's birth within 42 days. Therefore, the cohort will include births occurring at home, at sea, on military bases or in a hospital. The cohort will also include stillbirths which occur at an rate of 3.9 per 1,000 births [14]. Children whose mothers are not resident in England (i.e., visitors or residents in other UK countries) will be included in the cohort but are unlikely to have any linked data available. Additionally, children whose parents have opted out of sharing their NHS data beyond NHS England through the National Opt-Out Programme will not have linked health or education data available, as this data is provided by NHS England. Similarly, women who have opted out through the National Data Opt-Out Programme will not have their records included. Furthermore, NHS England will exclude children with child protection flags in the PDS. Children will be followed longitudinally via linked administrative datasets from birth until death, emigration from England (as indicated in the child's PDS record), or the latest health or education data available.

Figure 1 shows the expected follow-up from birth available for cohort children in a Lexis diagram. We anticipate that KEHC will be updated annually to capture new births each year and to incorporate the most recent health and education data for children already in the cohort. Ethical approval and a legal basis have been secured for conducting these annual updates. However, some time lags are expected due to the release schedules of data providers and logistical considerations in updating the dataset.

Table 1: Examples of existing databases used to study the associations between environments and health and well-being outcomes in children in the UK

Database	Coverage & population	Example of published studies
Millennium Cohort Study	~18,000 children in the UK in 2000-2001; 10,757 participants retained in latest sweep (2018–19) [15].	Milojevic et al (2021) [16] Mueller et al (2020) [17]
The Avon Longitudinal Study of Parents and Children (ALSPAC)	~14,000 pregnant women in and around Bristol in 1991–92 [18].	Fuertes et al (2020) [19] Cai et al (2020) [20] Soares et al (2024) [21]
Born in Bradford (BiB)	~12,400 pregnant women attending the Bradford Royal Infirmary maternity unit between 2007 and 2011 [22].	Subiza-Perez et al (2024) [23] Schembari et al (2015) [24] McEachan et al (2018) [25]
SAIL Databank	Over 5 million people who have received public services in Wales [12].	Wheeler et al 2019 [26] Garret et al 2021 [27]

Figure 1: Lexis Diagram of Kids Environment and Health Cohort (KEHC) Follow-up, 2006–2023

EYC: Early Years Census; EYFSP: Early Years Foundation Stage Profile; NPD: National Pupil Database; Phonics: Student's performance in phonics screening. Key Stages indicate the different blocks of the National Curriculum in England [45]. The different colours indicate annual birth cohorts within KEHC.

Data linkage

A diagram of data flows is shown in the Appendix Figure 1. Please note the linkage processes described below is subject to agreement by each data provider.

Creation of the KEHC spine

ONS will extract identifiers for mothers and children from the live and stillbirth registration and NHS birth notification records (Table 2, Dataset 3), including mother and child NHS numbers, child's sex and date of birth, the residential postcode of the mother/child at delivery/birth, child's name and mother's names (forename, surname, and maiden name). ONS will link the birth registration data (Table 2, Dataset 1) to mortality data for liveborn children (Table 2, Dataset 2), using the NHS number, child's name, date of birth and sex. The de-identified birth notification-registration attribute data and linked mortality records will be transferred into a secure ONS Trusted Research Environment (TRE) with individual cohort ID numbers to delineate each mother and child (live or stillborn) in the cohort. The linkage between birth data and mortality data is done to capture all instances of death,

including infant deaths under 1 year of age, as well as deaths occurring later in childhood. This ensures a comprehensive record of mortality for the entire cohort, allowing us to study patterns and factors associated with early and later childhood mortality.

Individual mother IDs will be derived by ONS based on the mother's NHS number, date of birth, and postcode at delivery to allow mothers, children, and siblings to be linked in KEHC. Identifiers for mothers and liveborn children (NHS numbers, postcode at birth/delivery and child sex and date of birth) together with cohort IDs will be securely transferred by ONS to NHS England.

Linking the KEHC spine to health and education outcome data

NHS England will link the KECH identifiers to children's health data, including data from HES Admitted Patient Care and A&E/Emergency Care (see Dataset 5 in Table 2), Mental Health Services (Table 2, Dataset 7) , Maternity Services (see Dataset 8 in Table 2), Community Dispensing (Table 2, Dataset 6), and the PDS (Table 2, Dataset 4) using

Table 2: Data sources for the creation of KEHC database

ID	Data category	Dataset name	Dataset level (individual or area level)	Key variables	Data owner
1	Vital Statistics	Birth registration [14, 28]	Individual	Birth weight; multiple birth status; parity; registration type (sole or joint); marital status; ages, occupations, and countries of birth of the parents; live or still birth indicator	ONS
2	Vital Statistics	Death registration [29]	Individual	Dates and causes of death for live births & stillbirths	ONS
3	Health	NHS birth notifications [30]	Individual	Birth-related information (e.g., gestational age, child's ethnicity)	NHS England
4	Health	Personal Demographic Service (PDS) [31]	Individual	Address information from NHS registration records (e.g., current and past addresses, emigration from England)	NHS England
5	Health	Hospital Episode Statistics (HES) Admitted Patient Care, Accident & Emergency and Emergency Care Dataset [32]	Individual	Hospital admission and discharge data; emergency department attendances	NHS England
6	Health	NHS Community Dispensing Data [33]	Individual	Community dispensed medication (e.g., antibiotics, asthma medications)	NHS England
7	Health	Mental Health Services Dataset (MHSD) [34]	Individual	Mental health service use (e.g., referrals, care contact episodes)	NHS England
8	Health	Maternity Services Dataset (MSDS) [35]	Individual	Maternity care information (e.g., smoking during pregnancy)	NHS England
9	Census	2011 and 2021 Census [36]	Individual	Demographic and household information (e.g., tenure, accommodation type) – linked to cohort mothers	ONS
10	Education	National Pupil Database (NPD) [37]	Individual and School	Educational information for pupils (e.g., school test results); school level information (e.g., number of students enrolled); for children attending early years childcare (whether they receive any free hours entitlement and/or free school meals)	Department for Education
11	Education	Open schools attribute data [38]	School	School information (e.g., establishment details, performance data)	Department of Education
12	Environment	Air pollution* [39]	Area (gridded data)	Modelled annual background air pollution data within 1x1 km grids	DEFRA
13	Environment	OS Open Roads* [40]	Area	Distance to roads (e.g. residence, school) and traffic flow data	Open Street Map data and Department for Transport

Continued

Table 2: Continued

ID	Data category	Dataset name	Dataset level (individual or area level)	Key variables	Data owner
14	Environment	Energy Performance Certificate (EPC) data* [41]	Dwelling	Energy performance information and building characteristics (e.g., energy efficiency rating, insulation levels and total floor area)	Department for Levelling Up, Housing and Communities
15	Environment	Temperature* [42]	Area (gridded data)	Daily and monthly minimum and maximum temperatures at 1x1 km grids	Met Office
16	Environment	Valuation Office Agency (VOA) data* [43]	Area	Property valuation data at dwelling level (e.g., age of property, property type, number of rooms)	Valuation Office Agency
17	Environment	OS Open Greenspace* [44]	Area	Greenspace (e.g., greenspace coverage, access, and quality)	Ordnance Survey

*Environmental data will be mapped to Unique Property Reference Numbers (UPRNs) and/or postcodes before linking to KEHC.

an established, deterministic linkage algorithm (the Master Person Service - MPS) [46]. NHS England will also link children in KEHC to their mother's HES Admitted Patient Care data up to one year before birth.

The de-identified attribute health data linked to the mother and child cohort IDs will be returned to ONS and placed in the ONS TRE. PDS addresses for mothers and children will be returned to the ONS. Linkage between PDS (Table 2, Dataset 4) and NPD (National Pupil Database) (Table 2, Dataset 10), developed for the ECHILD database [47] is based on a deterministic algorithm using name, date of birth, and chronological postcode, with a linkage rate of 95% [47]. NHS England will use the established PDS-NPD link to extract anonymised pupil reference numbers (aPRNs; the individual pupil ID in the NPD) for all children in the cohort. The aPRNs will be forwarded to ONS for onward linkage to NPD attribute data via the Department for Education (DfE).

Linkage of the KEHC spine to environmental data

ONS holds PDS records from 2016 onwards. Upon agreement to this process, ONS will link the KEHC spine to the ONS-held part of PDS and join these PDS records with those received from NHS England, generating longitudinal address records for the whole cohort of children and their mothers up to a year before the child's birth. ONS will derive Unique Property Reference Numbers (UPRNs), a unique identifier which is assigned to every addressable location across Great Britain [48], and extract postcodes, all with timestamps, and store the longitudinal UPRNs and postcodes from PDS in the Data Access Platform (DAP), ONS's internal secure data linkage platform, This will allow further linkages to environmental data. ONS will also link mothers' address at delivery from birth registration to UPRNs, and store these in the DAP for researchers wanting to link environmental data to the birth dwelling only, and for validation of PDS address histories.

ONS will link the cohort to postcode-level data on air pollution (Table 2, Dataset 12) and temperature (Table 2, Dataset 15), and UPRN-level data on distance to roads (Table 2, Dataset 13), traffic flow (Table 2, Dataset 13), greenspace coverage and access (Table 2, Dataset 17), Energy Performance Certificates (EPCs) (Table 2, Dataset 14), and Valuation Office Agency data (Table 2, Dataset 16) using the longitudinal postcode and UPRN PDS histories, respectively. These UPRN-linked environmental data will be developed by the KEHC study team over the course of the project. ONS will also link the longitudinal PDS UPRNs to Index of Multiple Deprivation quintiles (a lower super output area index of deprivation) [49] and government office region indicator. All linked postcode- and dwelling -level environmental and deprivation datasets will be placed in the ONS TRE.

Linkage of the 2010-2012 and the 2020-2022 KEHC subsample to maternal Census 2011 and 2021 data

ONS will link the mothers' identifiers (forenames, surnames, maiden names, dates of birth, and postcodes) from the birth records to the mothers' 2011 and 2021 Census records (Table 2, Dataset 9) using deterministic and probabilistic linkage algorithms. Only registrations for births between March 2010 and March 2012, and March 2020 and March 2022 will be linked to the 2011 and 2021 Census, respectively. Once Census linkage is complete, the de-identified Census attribute data for the mothers in KEHC will be released to the ONS TRE. ONS will assess and provide a report on the quality of linkage required to construct the KEHC spine and link the spine to Census data.

Linkage of the KEHC spine to the NPD and school location data

ONS will forward the aPRNs received from NHS England to DfE who will extract the attribute NPD records for the children

in KEHC who are old enough to have an aPRN. DfE will return the de-identified education data to the ONS TRE for inclusion into the KEHC. ONS will link the unique school IDs in NPD (Table 2, Dataset 10) to current and historical school address information published by DfE (Table 2, Dataset 11). The school addresses and school IDs will be held separately by ONS in the DAP. Researchers may request linkage of environmental data to KEHC via the school locations or school IDs to examine the impact of environments near schools on child health and education outcomes.

Database management and documentation

We will work with ONS to clean and set up the KEHC for access by external researchers. We will document all data cleaning steps, which will be described in the KEHC user guide, and any received or derived variables will be described in the KEHC data dictionary. All code used to clean the data will be made openly available on GitHub. Although the highly sensitive data in KEHC cannot be made openly available, we will strive to make the KEHC data resource adhere to Findable, Accessible Interoperable Reusability (FAIR) principles [50], supporting open science values.

Legal and ethical considerations

The legal bases for KEHC data include the Digital Economy Act 2017 and Section 251 of the NHS Act 2006 (for processing of confidential patient information). Under the General Data Protection Regulations (GDPR), the legal bases are 6(1)e for public task and 9(2)j for archiving, research, and statistics, specifically for processing special category data. We have obtained approvals from London – City & East NHS Research Ethics Committee, National Statistician's Data Ethics Committee and the Confidentiality Advisory Group (23/CAG/0060) to establish the KEHC as a research database. We are seeking approvals from Advisory Group for Data (AGD; NHS England), Data Sharing and Approval Panel (DSAP; DfE) and Research Accreditation Panel (UK Statistics Authority; UKSA).

Data storage and access conditions

The KEHC data will be held in an ONS TRE. Access and use of the ONS TRE is based on the Five Safes Framework [51]. All researchers accessing data will require training in handling sensitive data and must be approved by ONS. Researchers requesting access to the dataset will need to comply with and follow any conditions set by UCL (the data controller) and each of the data providers. Researchers using the KEHC will also be required to write lay summaries of their projects and associated research findings. These will be published on the KEHC website (https://kenvh-cohort.org.uk/).

Demonstrator projects

We will undertake two demonstrator projects to show the research value of KEHC. In the first project, we will investigate associations between proximity to greenspaces near homes [52] and mental health outcomes in adolescents. In the second project, we will examine the association between the density, quality, and proximity to Office for Standards in Education, Children's Services and Skills (OFSTED)-registered childcare providers locally and early educational attainment.

Patient and public involvement and engagement

KEHC and KEHC-based research will be of interest to parents, expectant parents, children, young people, and the general public. The KEHC Advisory Group has three parent members, and we have set up a dedicated study website (https://kenvh-cohort.org.uk/). Following feedback from young people, we now also run Twitter and Instagram accounts where we share study updates and will inform the general public about our findings related to how local environments might impact child health, education and well-being. Furthermore, we have carried out a range of involvement and engagement activities that have involved parents, children, young people, and members of the general public in the development of the KEHC. We have spoken to a number of groups of children, young people and parents, including several meetings with the Great Ormond Street Hospital Biomedical Research Centre Young People's Advisory Group and Parents' and Carers' Advisory Group, and the Liverpool Generation R Group to discuss the use of linked administrative data for environmental health research. We are continuing to engage with families and young people as we are progressing with setting up the KEHC. Engagement with parents and children has played a crucial role in shaping the research plans for KEHC study. Through this engagement, we gained valuable insights into the concerns of families, particularly regarding the types of data being collected and the intended use of this data. The feedback obtained through engagement and involvement activities has helped us choose the name and logo for the cohort, guided our plans for sharing information about the cohort and data safeguarding, and informed our strategies for sharing research findings with the public.

Conclusion

Understanding how the environment we live in influences health and education is crucial to creating environments that will promote health and well-being of all children. The KEHC is a unique national birth cohort that will enable research into how environmental features impact children's health in-utero, during childhood, and beyond. The longitudinal nature of the database will allow for the exploration of long-term trends and developmental trajectories, offering a nuanced perspective on the dynamic interplay between environments and health. The findings that will emerge from the KEHC will guide evidence-based decision making and ultimately shape policies that promote children's health and education and foster overall health and well-being of future generations.

Acknowledgements

Ethics statement

Ethical considerations: ES/X000311/1. Research at UCL Great Ormond Street Institute of Child Health benefits from funding for the NIHR Great Ormond Street Biomedical Research Centre.

Statement on conflicts of interest

The authors declare no conflicts of interest

Data availability statement

The Kids' Environment and Health Cohort (KEHC) dataset will be available for research within the secure ONS Trusted Research Environment (TRE). The specific access conditions are currently under development, but access to de-identified, project-specific datasets will need approval from the UK Statistics Authority (UKSA) through its Research Accreditation Panel (RAP). Researchers will be able to access the data for approved projects after completing required training and becoming ONS approved researchers, under conditions set by UCL and data providers.

References

1. Eriksson M, Dahlblom K. Children's perspectives on health-promoting living environments: The significance of social capital. Social Science & Medicine. 2020;258:113059. https://doi.org/10.1016/j.socscimed.2020.113059

2. Farrow A, Taylor H, Golding J. Time Spent in the Home by Different Family Members. Environmental Technology. 1997;18(6):605-13. https://doi.org/10.1080/09593331808616578

3. World Health Organisation. Healthy Environments for Healthy Children. 2010. [Available from: https://www.who.int/publications/i/item/9789241599887]

4. Clark H, Coll-Seck AM, Banerjee A, Peterson S, Dalglish SL, Ameratunga S, et al. A future for the world's children? A WHO-UNICEF-Lancet Commission. The Lancet. 2020;395(10224):605-58. https://doi.org/10.1016/S0140-6736(19)32540-1

5. Minh A, Muhajarine N, Janus M, Brownell M, Guhn M. A review of neighborhood effects and early child development: How, where, and for whom, do neighborhoods matter? Health & Place. 2017;46:155-74. https://doi.org/10.1016/j.healthplace.2017.04.012

6. Braveman P, Barclay C. Health disparities beginning in childhood: a life-course perspective. Pediatrics. 2009;124(Supplement_3):S163-S75.

7. Jia P, Cao X, Yang H, Dai S, He P, Huang G, et al. Green space access in the neighbourhood and childhood obesity. Obesity Reviews. 2021;22(S1):e13100. https://doi.org/10.1111/obr.13100

8. Christian H, Zubrick SR, Foster S, Giles-Corti B, Bull F, Wood L, et al. The influence of the neighborhood physical environment on early child health and development: A review and call for research. Health Place. 2015;33:25-36. https://doi.org/10.1016/j.healthplace.2015.01.005

9. Gascon M, Triguero-Mas M, Martínez D, Dadvand P, Forns J, Plasència A, Nieuwenhuijsen MJ. Mental health benefits of long-term exposure to residential green and blue spaces: a systematic review. International journal of environmental research and public health. 2015;12(4):4354-79. https://doi.org/10.3390/ijerph120404354

10. Rothman L, Buliung R, Macarthur C, To T, Howard A. Walking and child pedestrian injury: a systematic review of built environment correlates of safe walking. Inj Prev. 2014;20(1):41-9. http://doi.org/10.1136/injuryprev-2012-040701

11. Tillmann S, Tobin D, Avison W, Gilliland J. Mental health benefits of interactions with nature in children and teenagers: A systematic review. J Epidemiol Community Health. 2018. https://doi.org/10.1136/jech-2018-210436

12. Jones KH, Ford DV, Thompson S, Lyons R. A profile of the SAIL databank on the UK secure research platform. International journal of population data science. 2019;4(2). https://doi.org/10.1136/jech-2018-210436

13. UNIT MCSAPHS. STUDYING PHYSICAL ACTIVITY IN CHILDREN'S ENVIRONMENTS (SPACES) [Available from: https://www.gla.ac.uk/schools/healthwellbeing/research/mrccsosocialandpublichealthsciencesunit/programmes/places/movementurbanlandscapes/spaces/

14. ONS. Births in England and Wales: 2022 2022 [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/birthsummarytablesenglandandwales/2022.

15. Connelly, Roxanne, and Lucinda Platt. "Cohort profile: UK millennium cohort study (MCS)." International journal of epidemiology 43.6 (2014): 1719-1725. https://doi.org/10.1093/ije/dyu001

16. Milojevic A, Dutey-Magni P, Dearden L, Wilkinson P. Lifelong exposure to air pollution and cognitive development in young children: the UK Millennium Cohort Study. Environmental Research Letters. 2021;16(5):055023. https://doi.org/10.1088/1748-9326/abe90c

17. Mueller MA, Flouri E. Neighbourhood greenspace and children's trajectories of self-regulation: Findings from the UK Millennium Cohort Study. Journal of environmental psychology. 2020;71:101472. https://doi.org/10.1016/j.jenvp.2020.101472

18. Fraser, Abigail, et al. "Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort." International journal of epidemiology 42.1 (2013): 97-110. https://doi.org/10.1093/ije/dys066

19. Fuertes E, Markevych I, Thomas R, Boyd A, Granell R, Mahmoud O, et al. Residential greenspace and lung function up to 24 years of age: The ALSPAC birth cohort. Environment international. 2020;140:105749. https://doi.org/10.1016/j.envint.2020.105749

20. Cai Y, Hansell AL, Granell R, Blangiardo M, Zottoli M, Fecht D, et al. Prenatal, early-life, and childhood exposure to air pollution and lung function: the ALSPAC cohort. American journal of respiratory and critical care medicine. 2020;202(1):112-23. https://doi.org/10.1164/rccm.201902-0286OC

21. Gonçalves Soares A, Santos S, Seyve E, Nedelec R, Puhakka S, Eloranta A-M, et al. Prenatal urban environment and blood pressure trajectories from childhood to early adulthood. JACC: Advances. 2024;3(2):100808. https://doi.org/10.1016/j.jacadv.2023.100808

22. Wright J, Small N, Raynor P, Tuffnell D, Bhopal R, Cameron N, et al. Cohort Profile: the Born in Bradford multi-ethnic family cohort study. International journal of epidemiology. 2013;42(4):978-91. https://doi.org/10.1093/ije/dys112

23. Subiza-Pérez M, Krenz K, Watmuff A, Yang T, Gilbody S, Vaughan L, et al. Social inequalities, residential greenness and common mental disorders in women: evidence from the Born in Bradford family cohort study. Urban Forestry & Urban Greening. 2024:128241. https://doi.org/10.1016/j.ufug.2024.128241

24. Schembari A, De Hoogh K, Pedersen M, Dadvand P, Martinez D, Hoek G, et al. Ambient air pollution and newborn size and adiposity at birth: differences by maternal ethnicity (the born in Bradford study cohort). Environmental health perspectives. 2015;123(11):1208-15. https://doi.org/10.1289/ehp.1408675

25. McEachan RR, Yang TC, Roberts H, Pickett KE, Arseneau-Powell D, Gidlow CJ, et al. Availability, use of, and satisfaction with green space, and children's mental wellbeing at age 4 years in a multicultural, deprived, urban area: results from the Born in Bradford cohort study. The Lancet Planetary Health. 2018;2(6):e244-e54. https://doi.org/10.1016/S2542-5196(18)30119-0

26. Wheeler B. Population scale data linkage methods for an 11-year natural experimental study of green/blue space availability, visits and mental health and wellbeing. Environmental Epidemiology. 2019;3:437. https://doi.org/10.1097/01.EE9.0000610920.73248.aa

27. Garrett J, Wheeler B, Akbari A, Fry R, Geary R, Lovell R, et al. Association between greenspace and time spent in nature with subjective wellbeing: a cross-sectional data linkage study. The Lancet. 2021;398:S47. https://doi.org/10.1016/S0140-6736(21)02590-3

28. ONS. Birth registrations, England and Wales Microdata Metadata. 2016. [Available from: https://doc.ukdataservice.ac.uk/doc/8043/mrdoc/pdf/8043_metadata_birth_registrations_2016.pdf]

29. ONS. Death registrations finalised in England and Wales. https://doi.org/10.57906/jygg-zn40

30. NHS England. Birth notification process. [Available from: https://digital.nhs.uk/services/national-care-records-service/birth-notification-process]

31. NHS England. Personal demographic service (PDS) 2023. [Available from: https://www.england.nhs.uk/long-read/personal-demographic-service-pds/.]

32. NHS England. Hospital Episode Statistics (HES). [Available from: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics].

33. NHS England. Prescriptions Dispensed in the Community.[Available from: https://digital.nhs.uk/data-and-information/publications/statistical/prescriptions-dispensed-in-the-community].

34. NHS England. Mental Health Services Data Set (MHSDS). [Available from: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets/mental-health-services-data-set].

35. NHS England. Maternity Services Data Set. 2024. [A]vailable from: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets/maternity-services-data-set].

36. ONS. Census data [Available from: https://www.ons.gov.uk/census].

37. Education Df. National Pupil Dataset [Available from: https://www.find-npd-data.education.gov.uk/].

38. Education Df. Get Information about Schools [Available from: https://www.get-information-schools.service.gov.uk/].

39. DEFRA (Department for Environment FaRA. Air quality and emissions statistics [Available from: https://uk-air.defra.gov.uk/].

40. contributors O. [Available from: https://www.openstreetmap.org/#map=5/54.910/-3.432].

41. Department for Levelling Up HC. Energy Performance of Buildings Data [Available from: https://epc.opendatacommunities.org/].

42. Office M. Climate variables [Available from: https://www.metoffice.gov.uk/research/climate/maps-and-data/data/haduk-grid/haduk-grid].

43. ONS. Valuation Office Agency data. [Available from: https://www.ons.gov.uk/aboutus/whatwedo/programmes andprojects/censusanddatacollectiontransformation programme/futureofpopulationandsocialstatistics/datasource overviews/valuationofficeagencydata]

44. Ordnance Survey. OS Open Greenspace [Available from: https://www.ordnancesurvey.co.uk/products/os-open-greenspace.

45. UK Government. The national curriculum [Available from: https://www.gov.uk/national-curriculum].

46. NHS Digital. Announcement of methodological change: Impact of changes to Hospital Episode Statistics (HES) processing from April 2021. 2021. [Available from: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/hospital-episode-statistics-data-changes-in-2021]

47. Mc Grath-Lone L, Libuy N, Harron K, Jay MA, Wijlaars L, Etoori D, et al. Data Resource Profile: The Education and Child Health Insights from Linked Data (ECHILD) Database. International Journal of Epidemiology. 2021. https://doi.org/10.1093/ije/dyab149

48. UK Government. Guidance: Identifying property and street information 2022 [Available from: https://www.gov.uk/government/publications/open-standards-for-government/identifying-property-and-street-information.

49. Ministry of Housing CLG. English indices of deprivation 2019. [Available from: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019]

50. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data. 2016;3(1):1-9. https://doi.org/10.1038/sdata.2016.18 (2016).

51. Office for National Statistics. The 'Five Safes' – Data Privacy at ONS 2017 [Available from: https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/].

52. Engemann K, Pedersen CB, Arge L, Tsirogiannis C, Mortensen PB, Svenning J-C. Residential green space in childhood is associated with lower risk of psychiatric disorders from adolescence into adulthood. Proceedings of the National Academy of Sciences. 2019;116(11):5188-93. https://doi.org/10.1073/pnas.1807504116
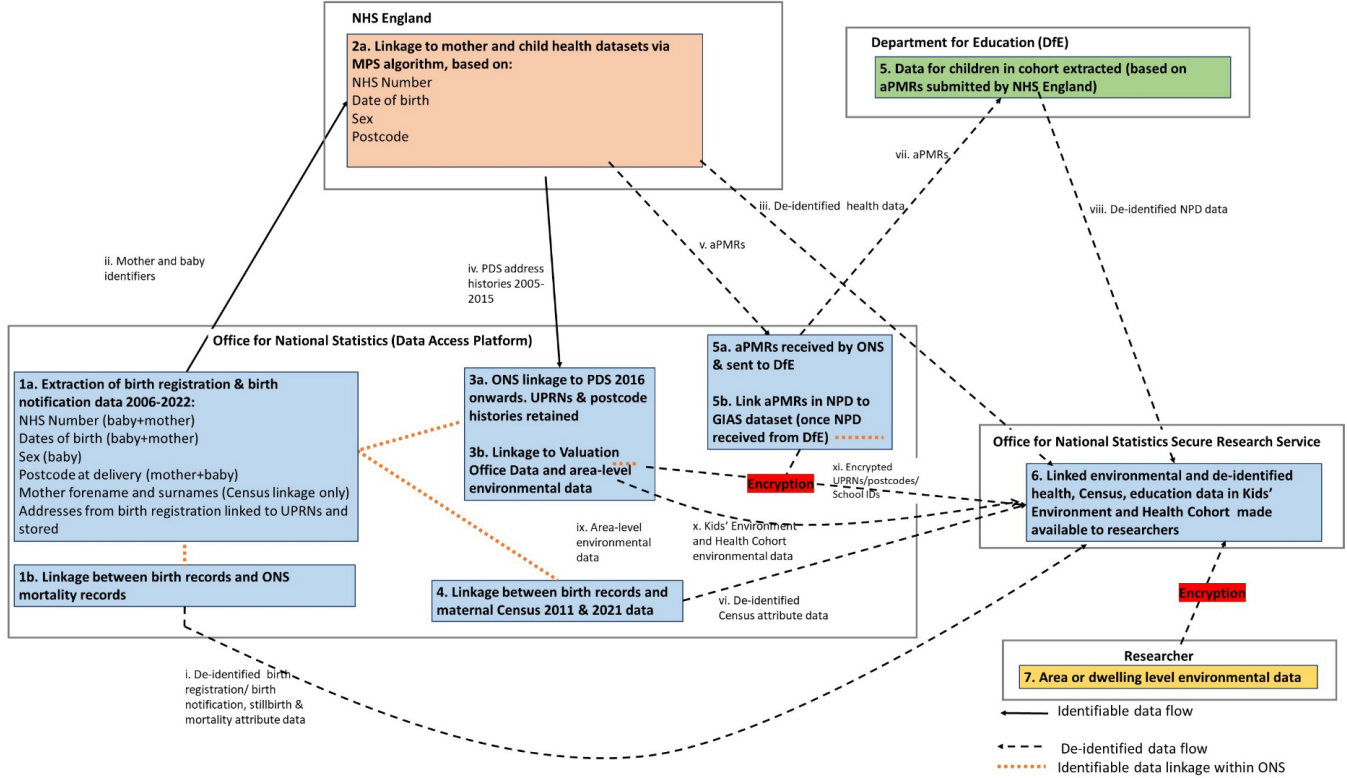
Abbreviations

ALSPAC:	Avon Longitudinal Study of Parents and Children
BiB:	Born in Bradford
DAP:	Data Access Platform
DEFRA:	Department for Environment, Food and Rural Affairs
DfE:	Department for Education
ECHILD:	Education and Child Health Insights from Linked Data
EPC:	Energy Performance Certificates
EYFS:	Early Years Foundation Stage
HES:	Hospital Episode Statistics
KEHC:	Kids' Environment and Health Cohort
MHSD:	Mental Health Services Dataset
MSDS:	Maternity Services Dataset
NHS:	National Health Service
NPD:	National Pupil Database
ONS:	Office for National Statistics
PDS:	Personal Demographic Service
UCL:	University College London
UKSA:	UK Statistics Authority
UPRN:	Unique Property Reference Numbers
VOA:	Valuation Office Agency

Supplementary Appendix

Appendix Figure 1: Kids Environment and Health Cohort (KEHC) data linkage flow chart

NHS: National Health Service; NPD: National Pupil Database; DAP: Data Access Platform; EPC: Energy Performance Certificate data; ONS: Office for National Statistics; PDS: Personal Demographic Service; TRE: Trusted Research Environment; UPRN: Unique Property Reference Number; VOA: Valuation Office Agency data.