A Researcher's Guide to the Born in Bradford Research Programme Data

Version 1.0

August 2025

Authors: G Santorelli, D Mason, A Watmuff, E Badrick, A Hough, T Yang, R McEachan, E Stephenson, O Shobande, J Wright

Table of Contents

1 Introduction

Born in Bradford (BiB) is an internationally recognised, multi-ethnic research programme investigating the complex factors that influence health and wellbeing. Through collaboration with local authorities and service providers, we translate our research findings into tangible initiatives to improve population health. The extensive data collected across our studies provides a rich, multifaceted resource for the research community.

Our work is structured around three primary birth cohorts, each with a distinct design:

- **The original BiB cohort (BiB)**: A longitudinal birth cohort of over 12,400 families recruited between 2007-2011. This foundational study involves regular follow-up through surveys, physical measurements, biological samples, and linkage to routine health and education records. It serves as a platform for numerous sub-studies and the ongoing 'Age of Wonder' project, which follows participants into adolescence.
- **Born in Bradford's Better Start cohort (BiBBS)**: A pioneering experimental birth cohort which has recruited over 5000 families between 2016-2024. BiBBS is designed to evaluate the impact of the Better Start Bradford programme, a comprehensive early-life intervention initiative.
- **BiB4All**: An ongoing, linkage-only cohort that began recruitment in 2019. By linking data from health, education, and social care records, BiB4All provides valuable insights into health and development trajectories without direct participant contact via questionnaires.

A visual data life course of the BiB family of studies is available in Appendix 1.

This document offers a detailed overview of the data available across each cohort to support researchers in planning their studies. For further details, please refer to our website and data dictionary. To help you navigate the data dictionary, hyperlinks to relevant datasets are provided for each section.

2 The Original Born in Bradford (BiB) Cohort

The original BiB cohort is our foundational longitudinal study, providing a platform for long-term follow-ups, nested sub-studies focussing on specific health questions, and collaborative projects such as the EU Child Cohort Network which has harmonised data across multiple European cohorts. Many of these studies also include participants from outside the original cohort (referred to as "BiB+") to broaden their scope.

The core BiB cohort recruited 12,453 pregnant women between February 2007 and March 2011[1]. This involved 13,776 pregnancies, resulting in 13,858 births (13,779 live born). The cohort comprises 13,455 singleton births, 177 sets of twins, and three sets of triplets. A total of 3,448 fathers were also recruited. Some women (1,247) participated with two pregnancies, and 38 participated with three. The average gestational age at recruitment was 26 weeks, though 448 women were recruited after their child's birth.

2.1 BiB studies

2.1.1 Longitudinal Follow-up Studies

- **Growing Up:** This follow-up focused on BiB children aged 6 – 10 years (extending to 13 years due to the COVID-19 pandemic). Recruitment took place between February 2017 and September 2021. Priority research areas included social and emotional wellbeing, growth, adiposity, cardiometabolic health, and cognitive and sensorimotor function[2]. The study recruited 5,314 mothers, 6,527 children, and 368 fathers from the original BiB cohort, and an additional 470 partners of the mothers who were not recruited to the original cohort.
- **Age of Wonder:** This ongoing study tracks adolescents from the start of secondary school until age 21, collecting information on mental and physical health, behaviours, environments, and education[3]. It includes both original BiB participants and individuals not initially part of the cohort (BiB+). Data collection will span the academic years 2021–22 to 2027–28.

2.1.2 Nested Sub-Studies

- **BiB 1000:** This study aimed to identify predictors and influences on health-related behaviours contributing to overweight and obesity[4]. Originally targeting 1,000 mothers, it recruited 1,735 mothers and 1,763 children. Questionnaires were administered at child ages of 6, 12, 18, 24, and 36 months between 2009 and 2012. 894 (5.7%) of these children were followed up at the Growing Up study.
- **ALL-IN (Allergy and Infection):** This nested study examined the 'hygiene hypothesis' by assessing the relationship between early-life infections with common herpes viruses and later immune function[5]. It recruited 2,553 mother-child pairs and questionnaires administered when the child was 12 months old (between Jul 2008 – Apr 2011) and again at 24 months (between Jul 2009 – Sep 2011). Of these, 1,655 (64.8%) were followed up at the Growing Up study.

- **Dental studies:** Two nested dental studies on oral health have been conducted.
  1. A study of 1,299 five-year-old children who had dental extractions under general anaesthetic to identify risk factors for dental caries.
  2. PLATOON (Premature Loss of bAby Teeth and its impact On Orthodontic Need)[6], a case-control study of 374 children aged 8-11, examined how premature tooth extraction affects space loss and the need for orthodontics. 332 children completed questionnaires and had an orthodontic assessment.

2.1.3 BiB+ and Cross-Cohort Studies

- **Starting School (BiB+):** Assessed key developmental areas in 6,844 children aged 4–5 (n=3,444 BiB and n=3,400 non-BiB)[7]. Conducted between March 2013 and July 2014, it evaluated literacy (BPVS, Letter ID), fine motor skills (CKAT), and social-emotional health (SDQ).
- **Primary School Years (BiB+):** Investigated cognitive development and well-being in 18,140 children aged 6-9 (n=7,006 BiB, n=11,143 non-BiB). Assessments included teacher-completed SDQs, child questionnaires, and CKAT [8].
- **BiB Breathes (BiB+):** This quasi-experimental study evaluates the impact of the Bradford Clean Zone (CAZ). Surveys were conducted on Bradford residents – including participants from the core BiB cohort – about their views on air quality and travel habits. before (2021-22, n=1,952 [1,137 BiB participants]) and after (2023, n-1994 [660 BiB participants]) the CAZ's implementation, gathering opinions on air quality and travel habits from BiB participants and the public.
- **COVID-19 studies (Cross-Cohort; BiB+):** Three phases of questionnaires to assess the impact of the pandemic were administered in April 2020 (n=2,144), September 2020 (BiB=631; BiBBS=136), and January 2021 (BiBBS=224) to BiB and BiBBS mothers, fathers, and children. Phase 3 was also open to the general public (1142 out of 1,954 completed questionnaires). In addition, BiB children completed three survey phases in June 2020 (n=970), November 2011 (n=633), and July 2021 (n=610).

2.1.4 European Research Collaborations

The BiB cohort participates in several large-scale European research collaborations. These partnerships pool data and expertise to investigate complex health issues.

2.1.4.1 MeDALL (Mechanisms of the Development of ALLergy)

A major collaboration with 23 European partners, the MeDALL project aimed to advance the understanding of how allergies develop[9]. BiB contributed significantly, recruiting 2,594 BiB mother-child pairs and 387 fathers. The study analysed biological samples – including maternal blood from pregnancy, umbilical cord blood, and paternal saliva - to identify allergy biomarkers. Further data was collected through questionnaires administered when the children were 4 – 5 years of age (Oct 2012 – May 2015). A subset of these children underwent skin prick testing, with 228 passing quality control.

2.1.5 Further European Collaborations

BiB has also contributed data to several other key European projects. The data from these collaborations are currently stored externally, and access can be requested through the respective project websites. Please see EU Child Cohort for further information.

- **HELIX (Human Early Life Exposome):** this project aimed to measure multiple environmental exposures (the 'exposome') in-utero and during childhood and assess their association with child health outcomes and molecular ('omics') markers[10]. Outdoor exposures were assessed for the entire BiB cohort using geospatial models. A specific cohort of 233 BiB mother-child pairs was recruited in 2014-2015 (at child age 6-7 years) for more intensive study. This involved questionnaires, neurodevelopment tests, physiological measurements (anthropometry, blood pressure), spirometry, and the collection of biological samples. A smaller group from this sub-study (n=43) also participated in a detailed panel study with additional monitoring for indoor/personal air pollution monitoring and further hair/biological sampling.

- **LifeCycle (Early-life stressors and LifeCycle health):** Active from 2017-2022, the LifeCycle consortium established the EU Child Cohort Network. Its goal was to create a harmonised, standardised database of exposure and health outcomes from numerous pregnancy and childhood cohorts across Europe[11]. For BiB, this involved harmonising existing data, including modelled environmental exposures from geocoded addresses, participant characteristics and health outcomes (respiratory, cardiovascular, mental) from questionnaires, physical tests, and health records.

- **ATHLETE (Advancing Tools for Human Early Lifecourse Exposome Research and Translation):** Building on LifeCycle, this consortium (2020-2025) expanded the EU Child Cohort Network by including more studies, harmonising a greater number of variables, and covering wider age ranges [12]. Within ATHLETE, 134 BiB children who previously took part in HELIX were followed up at age 13-14 (in 2021-2022). This follow-up gathered extensive data through questionnaires, sleep diaries, neurodevelopmental tests, spirometry, physical measurements, smartphone location data, activity/sleep monitors, personal air pollution, and biosamples (hair and stool).

2.2 Overview of Available Data for the BiB Cohort

Data for the original BiB cohort been compiled from the studies described above and enriched with linked routine health and education records, and geographic and environmental data.

2.2.1 Standard Data Package

Every approved data request includes a standard package of core datasets designed to provide essential information and facilitate correct data linkage. The standard package contains the following:

- **Person-Level Data:** Core demographic information for all individuals (mothers, fathers, and children).
- **Pregnancy-Level Data:** Information specific to each pregnancy recorded within the study.
- **Ethnicity Data:** Self-reported ethnicity information for participants.
- **Linkage and Relationship Files:** These datasets map relationships between participants. Their primary functions are to:
    - Enable the linkage of datasets across different participant types (e.g., linking a mother's data to her child's).
    - Correctly associated children with the specific pregnancy they belong to, which is crucial for families with multiple pregnancies or births within the cohort.

Action Required for Data Linkage: Before the linkage file can be used successfully, the user must prepare their other datasets. In most datasets, the generic identifier is `BiBPersonID`. This variable **must be renamed** to match the relevant participant identifier in the linkage file (for example, in a dataset containing mothers' information, `BiBPersonID` should be renamed to `BiBMotherID`). This allows for a successful merge with the linkage file.

2.2.2 Baseline Questionnaire

A total of 10,519 women completed the baseline questionnaire, resulting in 11,396 forms due to multiple pregnancies. Three versions of the questionnaire were used: Version 1 (1,776 completions), Version 2 (5,298 completions), and Version 3 (4,322 completions). Core questions common across all versions are consolidated in the primary baseline dataset, while separate datasets contain version-specific responses from smaller subsets of participants. In addition, 3,287 fathers completed a shorter baseline questionnaire, generating 3,387 responses.

2.2.3 Ethnicity

Ethnicity was initially recorded in the baseline questionnaire for both mothers and fathers. A child's ethnicity was initially assigned using their mother's self-reported ethnicity from the baseline questionnaire. This was later updated with the child's own ethnicity, sourced hierarchically from education records, then primary care records. If both sources were unavailable, the maternal ethnicity was used.

2.2.3 Physiological measurements

2.2.3.1 Mothers

- **Anthropometrics**

Maternal height, weight, and BMI are available for over 12,000 participants, obtained from pregnancy, the from the BiB 1000, ALL-IN, and Growing Up studies (for those who participated in these studies).

- **Blood pressure**

Blood pressure measurements during pregnancy (see eclipse and maternity records) and labour, and taken during the Growing Up study are available.

2.2.3.2 Children

- **Anthropometrics**: The BiB cohort includes over 160,000 longitudinal height and weight measurements for children, with BMI and age/sex-adjusted z-scores using UK90 reference data. These data come from various BiB studies, the National Child Measurement Programme (NCMP), and primary care records. Additional measurements include skinfold thicknesses (triceps and subscapular), head, arm, and thigh circumferences, bioimpedance, and DXA data.
- **Blood pressure:** Blood pressure in children was recorded by school nurses at Reception age (4 – 5 years; n=4,559) and later between ages 7 – 13 during the Growing Up study, resulting in 11,340 measurements from 8,529 children.

2.2.4 Obstetrics

Obstetric data was collected from several sources:

- **Eclipse**: The maternity IT system used at the time of baseline recruitment, providing pregnancy data (13,361 records) and birth data (13,525 records).
- **Maternity notes**: A 2013 project investigating maternal glucose levels and gestational diabetes extracted data from paper maternity notes for 10,939 pregnancies involving 10,041 mothers who underwent an oral glucose tolerance test (OGTT) and completed the baseline questionnaire. This dataset offers more detailed information than the Eclipse dataset, although some variables are shared between the two.
- **Ultrasound**: 45,835 scans from 11,979 participants.
- **Foetal renal measurements**: Kidney dimensions at 34 weeks gestation were recorded for 1,803 participants to explore differences between White British and Pakistani-heritage groups [13].

2.2.5 Congenital Anomalies

Initial data on congenital anomalies diagnosed from birth and age 4 were obtained from the Yorkshire and Humber Congenital Anomalies Register (YHCAR), identifying 1,201 records from 510 children. Additional anomalies were identified from primary care records up to age 5 to improve case capture (a further 428 cases) [14].

2.2.6 Biosamples and biobank

- During recruitment, pregnant women undergoing oral glucose tolerance test (OGTT) provided fasting and 2-hour post-load blood samples. Additionally, 11,703 mothers contributed an extra blood sample, and 6,979 provided a urine sample. Blood samples were analysed at BRI for glucose, lipid profiles, and vitamin D levels; urinary samples were tested for iodine levels.
- Routine obstetric blood results (e.g., haematology, biochemistry, HbA1c) are also available.
- Saliva samples were collected from 3,062 fathers.
- Umbilical cord blood samples from 9,604 babies contain data on lipids, CRP, insulin, and leptin.
- In the Growing Up study, blood samples were collected from 2,624 children and 3,033 mothers; urine samples were collected from 720 children and 51 mothers.
- Future sample collection is planned for Age of Wonder participants (ages 13-14 and 16-17).
- Biosample aliquots are stored in the University of Bristol biobank for future research.

2.2.7 Genetic and 'omics'

The following genetic and omics datasets are available. All were extracted from maternal pregnancy and cord samples. A more detailed technical summary, including participant characteristics and extraction methods is available upon request.

- **Genotype data**
  Genome-wise association study (GWAS)-ready genotype data is available for BiB participants, processed using microarray chips (Global Screening Array and HumanCoreExome). A total of 19,243 microarrays are available in raw and plink formats.

- **Imputed genotype data**
  Genotype imputation has been completed for 7,580 white European and 8,666 South Asian participants to infer unmeasured genetic variants.

- **Exome sequencing**
  Exome sequencing, a technique used to identify genetic variants altering protein sequencing has been conducted in three phases:
  - Phase I: 167 cord samples from BiB children and 2,637 BiB mothers of Pakistani heritage.
  - Phase II: 513 BiB pregnancy samples (Pakistani heritage).
  - Phase III: Samples from pregnancies, fathers, and cord samples (n=20,388 total)

- **DNA methylation (DNAm)**
  DNAm profiles from approximately 1000 White British and Pakistani mother-child pairs (pregnancy blood and cord samples). After quality control, usable data includes 450 White British mothers/457 children and 483 Pakistani mothers/493 children were retained.

- **Proteomics**
  Proteomics data has been collected through three separate projects:
  - Project 1
    A random sample of 1000 participants, comprising four groups of 250 each:
    - Pakistani with gestational diabetes mellitus (GDM)
    - Pakistani healthy controls
    - White British with GDM
    - White British healthy controls.
  - Project 2
    A further random sampled 3,000 women (1,500 White British and 1,500 Pakistani), excluding those from Project 1 was drawn. Participants were included if they had any of the following: GDM, hypertensive disorders of pregnancy, stillbirths, pre-term births, their children were born with congenital anomalies. A control group was randomly selected with none of these issues.
  - Project 3
    A third project analysed proteomics from a further 1,043 South Asian and 698 White British women.

- **Glycomics**
  Glycomics data from 976 BiB Pakistani women was collected as part of the HDRUK Multiomics Cohorts Consortium.

- **Metabolomics**
  Metabolomics data from pregnancy, neonatal, and infant samples were generated using mass spectrometry (MS) and Nuclear magnetic resonance (NMR).
  - MS data includes metabolomics involved two rounds of sampling based on participant characteristics: the first with approximately 1000 White British and Pakistani mother-child pairs (pregnancy and cord blood), and the second a case-cohort study of 2000 White British or Pakistani mothers. Flow charts detailing the sample selection methods are available upon request.
  - NMR data is available for around 10,500 mothers (pregnancy), 8,000 babies (cord blood) and ~1,700 infants at 12 months and ~1,500 at 24 months.

2.2.8 Linked Routine Data

- **Primary and Secondary Care**
  Electronic health records (EHR) from primary care are sourced from SystmOne, covering all Bradford GP practices. These include medical history, prescriptions, test results, and immunisation data. Child health and health visitor records are also linked. Secondary care data include A&E visits, inpatient and outpatient records, procedures, pharmacy, radiology, and laboratory test results. These datasets are stored separately from the main BiB datasets and are therefore not documented in the online data dictionary.

- **Asthma, eczema, and allergic rhinitis**
  We offer pre-defined datasets for asthma, eczema, and allergic rhinitis that include age at diagnosis and two validated case definition indicators (broad and specific). Using these datasets removes the need to fund and access primary care records for case identification. The methodology behind these case definitions is published in our peer-reviewed paper[15]. Researchers must consult this publication to ensure our definitions align with their study's needs before using the data.

- **Dental Records**
  Linked dental records are sourced from NHS Dental Services FP17 forms, which document single course treatments. These forms include details of assessments, examinations, and planned treatments, as well as treatment activity and associated patient charges.

- **Education records**
  Education data are primarily sourced from the City of Bradford Metropolitan District Council (CBMDC). Additional school-level information – such as Ofsted ratings and contextual characteristics – is obtained from publicly available datasets via the gov.uk data service.
  - Early Years Foundation Stage Profile (EYFSP): A teacher-led assessment conducted at the end of Reception (ages 4-5), prior to entry into Key Stage 1. It evaluates children's development across seven areas: Personal, social, and emotional development (PSED), Communication and Language (C+L), Physical Development (PD), Literacy (L), Mathematics (M), Understanding the World (UTW), Expressive Arts and Design (EAD). Data are available for 11,303 children, split across two datasets: pre-2013 (n=705) and from 2013 onwards (n=10,598).
  - Year 1 Phonics Assessment: A teacher-led assessment of children's phonics skills, typically conducted towards the end of Year 1. Children who do not meet the expected standard may retake the assessment in Year 2. Data are available for 11,081 children.
  - Key Stage 1 (KS1): This assessment, taken by children aged 5-7 (Years 1 and 2), covers subjects including English, maths, science, history, and geography, and includes SATs exams. Data are available for 10,884 children, divided into two datasets: pre-2016 (n=3,622) and 2016 onwards (n=7,262).
  - Key Stage 2 (KS2): Covering a broader curriculum for children aged 10-11 (Years 3 - 6), KS2 includes SATs exams. Data are currently available for 3,466 children who took their SATs in the 2017/2018 and 2018/2019 academic years. KS2 SATs were not conducted during the 2019/20 and 2020/21 academic years due to the COVID-19 pandemic, resulting in no data for a large proportion of the BiB cohort. Future data provision from CBMDC will begin from the 2022 academic year, including KS2 results for the final group of BiB children.

- **Geographic and Environmental Data**
  A longitudinal spatial database has been developed using participant addresses histories, neighbourhood-level data (e.g., IMD), and geographic boundaries (e.g., LSOA, Ward, LA). Addresses, traced at monthly intervals using NHS data, are geocoded to enable accurate spatial linkage. We provide data at the property level and a linkage file to participants property history over time based on research criteria.

  The *area-level data* allows for analysis of neighbourhood characteristics through:

  - Geographic boundaries: Location is defined by Lower-layer Super Output Areas (LSOA), Wards and the Bradford Local Authority.

- Deprivation Indices: A comprehensive set of measures are available, including:
  - National Indices of Multiple Deprivation (IMD) for 2010 and 2019
  - A bespoke Bradford-specific IMD to capture local variations
  - Census-based indices (Townsend and Carstairs)
  - The Income Deprivation Affecting Children Index (IDACI)

The dataset contains *property-level indicators* linked to participant addresses, including:

- Built environment and housing quality: Information on indoor housing characteristics and the immediate built environment, based on energy assessor inspections carried out between 2008 and 2021.
- Greenness and Vegetation:
- Local Amenities: Data on the food environment, including the number and proximity of food outlets and fast-food establishments.
- Transport and Mobility:
- Air quality: Data from 2018 on local concentrations of PM_{10}, $PM_{2.5}$, NO_x, sourced from the City of Bradford Metropolitan District Council CBMDC and UK Department for Environment, Food & Rural Affairs (DEFRA).

3 Born in Bradford's Better Start (BiBBS)

Between 2016 and 2024, the BiBBS cohort recruited 4,234 mothers and 402 fathers, covering 5,266 pregnancies that resulted in 5,328 children[16]. Women were mostly recruited at their OGTT appointment.

3.1 Overview of Available Data for the BiBBS Cohort

3.1.1 Baseline questionnaire

An interim dataset, containing 2,564 questionnaires from 2,378 women recruited between 2016 and 2019, is currently available [17]. The full cohort data is expected in 2025.

3.1.2 Ethnicity

Ethnicity is self-reported in the baseline questionnaire.

3.1.3 Anthropometrics

Mother's height and weight were measured at recruitment. Although triceps and subscapular skinfold thickness measurements were also initially collected, this was discontinued due to low participation. Consequently, skinfold data is available for a small number of women. Height and weight measurements for children are not part of this core dataset but are expected to be available in the linked primary care records.

3.1.4 Biosamples and biobank

OGTT data were obtained, and samples were collected at recruitment. These data are not currently available.

3.1.5 Exome Sequencing

Exome sequencing data are available for 2,516 participants (1,164 children, 1,352 mothers).

3.1.6 Linked Routine Data

Pregnancy and birth data are not yet available for research requests pending the completion of data cleaning. Like the original BiB cohort, BiBBS participant data is linked to comprehensive routine primary and secondary healthcare, dental, and geographic/environmental records. Consent has also been obtained for linkage to education records, which will become available as the children progress through the school system. For a detailed description of the available linked datasets, please refer to the corresponding sections for the original BiB cohort above.

4 BiB4All

BiB4All is an ongoing study that began in 2019. As a linkage-only cohort, it offers powerful population-level insights without the direct data collection of other BiB studies. As of June 2025, it includes approximately 9,000 women and over 9,800 births, with recruitment continuing during routine maternity care.

4.1 Overview of Available Data for the BiB4All cohort

- **Linked Data Only:** This cohort does not include questionnaires, physical measurements or biological samples collected by the research team. Its research value is derived entirely from connecting to routine administrative data.
- **Available Linkage:** BiB4All participants are linked to the same rich routine primary and secondary health care, dental, education, and geographic/environmental datasets as the other cohorts. Linkage to health and environmental data is ongoing, with education datasets to follow as the children enter school. Please refer to the detailed descriptions in the section for the original BiB cohort for a full overview of these linked datasets.

5 How to Access Born in Bradford Data

We are committed to a transparent and collaborative data access process. To ensure clarity and efficiency for both researchers and our team, data access follows these key stages:

5.1 Submission & Initial Review

To begin, submit a completed Expression of Interest (EoI) form. This will be reviewed by our Executive Committee to assess the proposal's feasibility, ethical considerations, and alignment with Born in Bradford's strategic aims. At this stage, you will not be expected to provide us with exact variable names, just the datasets that you wish to utilise.

5.2 Decision & Outcome

Following the review, you will receive a formal decision.

- **If approved**, we will proceed to the next stage to organise the necessary agreements.
- **If not approved**, we will provide constructive feedback. In some cases, you may be invited to revise and resubmit your EoI based on the committee's recommendations.

5.3 Agreements & Payment

Once your EoI is approved, we will issue a Data Sharing Contract (DSC), a Data Sharing Agreement (DSA), and an invoice for any associated costs. Your project will formally begin once we receive the signed agreements and payment has been completed.

5.4 Data Finalisation and Preparation

Your project will be assigned to one of our data analysts, marking the start of a collaborative stage to define your final variable list. Given the complexity of our data, it is essential to work closely with your assigned analyst to specify your data needs comprehensively at this point. Providing a detailed and complete list of variables upfront is essential for a smooth process and avoids delays caused by subsequent revisions or multiple small data requests.

- **Data dictionary:** The BiB Data Dictionary contains information on all available cohort data. Please note that linked routine healthcare records are stored in a separate repository and are not included in the data dictionary.
- **Healthcare Data Extraction:** We do not provide full medical records. Instead, applicants must supply a list of specific diagnostic codes relevant to their research. Extraction will be limited to these codes.

5.5 Data Extraction & Secure Transfer

Once your variable list is finalised, the data analyst will extract the required data and transfer it to you via a secure method. Upon confirmation of receipt, the request will be formally closed. All data must be handled and used in strict accordance with the terms of your signed BiB Data Access Agreement.

6 References

[1] J. Wright *et al.*, "Cohort profile: The born in bradford multi-ethnic family cohort study," *Int. J. Epidemiol.*, vol. 42, no. 4, 2013, doi: 10.1093/ije/dys112.

[2] D. . et al. McEachan, Rosemary R.C.; Santorelli, G.; Watmuff, A.; Mason, "Cohort Profile Update: Born in Bradford," *Int. J. Epidemiol.*, vol. 53, no. 2, 2024.

[3] K. A. Shire *et al.*, "Born in Bradford's Age of Wonder cohort: protocol for adolescent data collection," *Wellcome Open Res.*, vol. 9, 2024, doi: 10.12688/wellcomeopenres.20785.1.

[4] J. W. M Bryant, G Santorelli, D A Lawlor, D Farrar, D Tuffnell, R Bhopal, "Design and characteristics of a new birth cohort, to study the early origins and ethnic variation of childhood obesity: the BiB1000 study," *Longit. Life Course Stud.*, vol. 4, no. 2, 2013, doi: 10.14301/llcs.v4i2.221.

[5] L. Pembrey, D. Waiblinger, P. Griffiths, M. Patel, R. Azad, and J. Wright, "Cytomegalovirus, Epstein-Barr virus and varicella zoster virus infection in the first two years of life: A cohort study in Bradford, UK," *BMC Infect. Dis.*, vol. 17, no. 1, 2017, doi: 10.1186/s12879-017-2319-7.

[6] L. R. Brown *et al.*, "PLATOON: Premature Loss of bAby Teeth and its impact On Orthodontic Need - protocol," *J. Orthod.*, vol. 46, no. 2, 2019, doi: 10.1177/1465312519843305.

[7] K. Shire *et al.*, "Starting School: a large-scale start of school assessment within the 'Born in Bradford' longitudinal cohort," *Wellcome Open Res.*, vol. 5, 2020, doi: 10.12688/wellcomeopenres.15610.1.

[8] L. J. Hill *et al.*, "Large-scale assessment of 7-11-year-olds' cognitive and sensorimotor function within the Born in Bradford longitudinal birth cohort study," *Wellcome Open Res.*, vol. 6, 2022, doi: 10.12688/wellcomeopenres.16429.2.

[9] J. Bousquet *et al.*, "MeDALL (Mechanisms of the Development of ALLergy): An integrated approach from phenotypes to systems medicine," *Allergy: European Journal of Allergy and Clinical Immunology*, vol. 66, no. 5. 2011. doi: 10.1111/j.1398-9995.2010.02534.x.

[10] L. Maitre *et al.*, "Human Early Life Exposome (HELIX) study: A European population-based exposome cohort," *BMJ Open*, vol. 8, no. 9, 2018, doi: 10.1136/bmjopen-2017-021311.

[11] V. W. V. Jaddoe *et al.*, "The LifeCycle Project-EU Child Cohort Network: a federated analysis infrastructure and harmonized data of more than 250,000 children and parents," *Eur. J. Epidemiol.*, vol. 35, no. 7, 2020, doi: 10.1007/s10654-020-00662-z.

[12] M. Vrijheid *et al.*, "Advancing tools for human early lifecourse exposome research and translation (ATHLETE)," *Environ. Epidemiol.*, vol. 5, no. 5, 2021, doi: 10.1097/EE9.0000000000000166.

[13] P. J. Roderick, R. F. Jeffrey, H. M. Yuen, K. M. Godfrey, J. West, and J. Wright, "Smaller kidney size at birth in South Asians: Findings from the Born in Bradford birth cohort study," *Nephrol. Dial. Transplant.*, vol. 31, no. 3, 2016, doi: 10.1093/ndt/gfv274.

[14] C. Bishop *et al.*, "Improving case ascertainment of congenital anomalies: Findings from a prospective birth cohort with detailed primary care record linkage," *BMJ Paediatr. Open*, vol. 1, no. 1, 2017, doi: 10.1136/bmjpo-2017-000171.

[15] S. Souza da Cunha, G. Santorelli, and L. Pembrey, "Defining cases of asthma, eczema and allergic rhinitis using electronic health records in the Born in Bradford birth cohort," *Clinical and Experimental Allergy*, vol. 53, no. 6. 2023. doi: 10.1111/cea.14291.

[16] J. Dickerson *et al.*, "Born in Bradford's Better Start: An experimental birth cohort study to evaluate the impact of early life interventions," *BMC Public Health*, vol. 16, no. 1, 2016, doi: 10.1186/s12889-016-3318-0.

[17] J. Dickerson *et al.*, "Born in Bradford's Better Start (BiBBS) interventional birth cohort study: Interim cohort profile," *Wellcome Open Res.*, vol. 7, 2023, doi: 10.12688/wellcomeopenres.18394.2.

Appendix 1 – BiB data life course