Born in Bradford is a longitudinal multi-ethnic birth cohort study aiming to examine the impact of environmental, psychological and genetic factors on maternal and child health and wellbeing. Bradford is a city in the North of England with high levels of socio-economic deprivation and ethnic diversity. Women were recruited at the Bradford Royal Infirmary at 26-28 weeks gestation. For those consenting, a baseline questionnaire was completed. The full BiB cohort recruited 12,453 women comprising 13,776 pregnancies between 2007 and 2010 and the cohort is broadly characteristic of the city's maternal population. Mean age of the mothers at study recruitment was 27 years old. Researchers are looking at the links between the circumstances of a child's birth, the context in which they grow up, their health and well-being and their educational progress. Ethical approval for the data collection was granted by Bradford Research Ethics Committee (Ref 07/H1302/112).

Size of cohort

12453 mothers recruited across 13776 pregnancies

3353 fathers recruited across 3455 pregnancies

13858 babies (livebirths + stillbirths)

Mothers and fathers could sign up for more than one pregnancy, so we count pregnancies, parents and children separately.

Baseline data

10519 mothers have questionnaires across 11395 pregnancies

3287 fathers have questionnaires across 3387 pregnancies

11231 mothers have completed GTT across 12331 pregnancies

13361 pregnancies linked to pregnancy outcomes on maternity IT system data

11050 pregnancies have both questionnaire and IT system data

13525 babies linked to birth outcomes on maternity IT system data

At recruitment, pregnant women were invited to complete a questionnaire, assessing detailed personal sociodemographics and family circumstances.

They were weighed and measured and many completed a glucose tolerance test (GTT).

Their partners completed a shortened version of the questionnaire and were also weighed and measured.

Linkage to maternity IT systems was completed to obtain pregnancy measures and birth outcomes.

Other maternity healthcare record data

13248 pregnancies have ultrasound scans

10939 pregnancies have had information abstracted from patient notes, leading to:

10895 pregnancies with pregnancy blood pressures

8523 pregnancies with blood pressure during labour

10024 pregnancies with postpartum blood pressures

855 pregnancies with maternity-related hospital admissions information

5611 mothers with information abstracted about pre-BiB pregnancies and infants

All pregnancy ultrasound scans were obtained via data linkage with maternity hospital records, including reason for scan and measurements.

Blood pressures, maternity hospital admissions and other information from the written maternity notes were manually entered by clinical researchers into electronic data capture systems.

Biobank: mother samples

10587 mothers have plasma across 11503 pregnancies

10581 mothers have RBC across 11495 pregnancies

10762 mothers have serum across 11715 pregnancies

367 mothers have buffy across 370 pregnancies

9619 mothers have whole blood across 10262 pregnancies

6662 mothers have urine across 6996 pregnancies

10531 mothers have DNA across 11436 pregnancies

At recruitment, consenting women provided a blood sample and urine sample for storage in a research biobank.

DNA was extracted mainly from buffy coat (hence the reduced number remaining here) but in some cases from whole blood.

Biobank: baby samples

7909 have plasma

7739 have RBC

9303 have serum

200 have buffy

6741 have whole blood

9158 babies have DNA

At delivery of the baby, if the mother consented, a blood sample was taken from the umbilical cord for storage in a research biobank.

DNA was extracted from mainly from buffy coat but in some cases from whole blood.

Biobank: father samples

2938 fathers have DNA across 3022 pregnancies

At recruitment, consenting fathers provided a saliva sample, from which DNA was extracted.

Pregnancy biomarkers

10685 mothers at 11625 pregnancies have blood biomarkers data 1498 mothers have vitamin D data (at one pregnancy)	Insulin, cholesterol, LDL, HDL, triglycerides Subcohort vitamin D sample: albumin, calcium, phthalates, 25(OH) Vit D2, 25(OH) Vit D3

Baby blood biomarkers

7910 babies have blood biomarkers data	From cord blood sample. Adiponectin, cholesterol, HDL, LDL, insulin, leptin, triglyceride.

Metabolite profiles

10574 mothers at 11479 pregnancies have NMR metabolomics Also expected in 2018: 1000 mother-child pairs (2000 samples) to have metabolomics profiles using MS metabolomics platform Selection criteria for the 1000 pairs: • Completed GTT • 50% Pakistani, 50% white British • Singleton birth • Existing methylation data	High throughput NMR metabolomics platform to derive quantitative molecular information on ~150 metabolites. The platform focuses primarily on lipids and fatty acids, and also includes glucose, lactose, some amino acids and ketone bodies. Selected all women with fasting pregnancy serum or edta plasma sample The MS metabolomics platform provides quantified data on ~1000 metabolites, including several that reflect diet, vitamins, all essential amino acids, carbon metabolism biomarkers, a much more detailed lipid profile and fatty acids

Genome data: exome sequencing

2333 mothers at 2489 pregnancies have exome sequence data

This means there are 156 mother-level replicates

262 children have exome sequence data

A further 672 mothers have been sequenced, and data will be loaded and linked early 2018.

Exome sequencing completed on Illumina HiSeq 2000 platform. Variant calls stored for further linkage and analysis in vcf format.

Sample selection criteria:

All women who indicated their parents were related other than by marriage

Children with clinician-diagnosed congenital anomalies

The additional 672 were identified consanguineous from microarray genotype profiles

Genome data: genotyping microarrays

8610 mothers at 9099 pregnancies have microarray genotype data

This means there are 489 mother-level replicates

7157 children have microarray genotype data

6256 of these children have paired mother with genotype data, and 6131 of these pairs are from the same pregnancy

The above data is available now.

To complete the cohort, genotype data from the following samples is currently in QC and will become available in 2018:

- 2237 mother samples
- 2991 father samples
- 2005 child samples

Genotyping completed using Illumina HumanCoreExome microarrays.

Sample selection criteria:

Complete mother-offspring pairs with DNA were prioritised. Then moving to unpaired samples with good phenotypic characterisation e.g. complete GTT, complete pregnancy height/weight data.

DNA methylation expected 2018

1000 mother-offspring pairs (2000 samples)

Assays are completed and data is in QC.

Illumina MethylationEPIC (850k) chip

Selection criteria:

Completed GTT

50% Pakistani, 50% white British

Singleton birth

Existing good quality GWAS data for both in the pair

Child subcohort follow-ups

1763 children recruited to BiB 1000 (6m, 12m, 18m, 24m, 36m follow-ups)

2553 children in ALL IN 12m follow-up

2067 children in ALL IN 18m follow-up

2594 children in MeDALL 48m follow-up

2269 children in MeDALL have skinprick data

233 children in HELIX 7y follow-up

Primary research follow-ups of the children from 0-7 years were conducted through a series of subcohort. A broad range of sociodemographic, developmental and clinical measures were taken. Separate data dictionaries are available.

Starting School follow-up

3444 children in the Starting School Reception year follow-up study

3253 children have Reception year CKAT

2340 children have Reception year SDQ

3297 children have Reception year BPVS/Letter ID

Cognitive development follow-up subcohort focused on two full academic years in Bradford schools.

CKAT measures sensorimotor control

SDQ measures emotional development

BPVS measures receptive vocabulary

Letter ID measures letter identification

Education data linkage: Early Years Foundation Stage Profile

A total of 11258 children with complete EYFSP data, from two different test versions: 705 EYFSP pre-2012 10553 EYFSP post-2012 *Cohort complete*	Standard teacher-led assessment that occurs towards the end of the child's Reception year at school. It changed from the start of the 2012/13 academic year.

Education data linkage: Year 1 Phonics Assessment

11027 Y1 Phonics Assessment *Cohort complete*	Teacher-led assessment to determine whether the child has met the required phonics standard by the end of Year 1.

Education data linkage: Key Stage 1 Assessment

A total of 9347 children with complete Key Stage 1 Assessment data *Complete up to 2016/2017 academic year assessments. Expecting a further ~1600 by end of 2018.*	Standard teacher-led assessment that occurs towards the end of the child's Year 2 at school.

Education contextual data

11667 children with Education contextual data	Contextual data obtained from schools by the Local Authority. Contains child level indicators of English as an additional language, child ethnic origin, free school meals, gifted and talented, looked after child and special educational needs.

Healthcare data linkage: SystmOne GP record

12311 (98.9%) mothers matched

13776 (99.4%) children matched

3296 (98.3%) fathers matched

Clinical events are coded using a standardised terminology called CTV3 Read. This will soon migrate to SNOMED-CT

Prescription events are coded using the British National Formulary

1.1m prescription events and 5.8m clinical codes from mothers

490,000 prescription events and 2.3m clinical codes from children

200,000 prescription events and 790,000 clinical codes from fathers

All GP practices in Bradford use SystmOne to record clinical codes and prescriptions.

We extract complete GP data records from SystmOne where we find an exact match on NHS number, surname, date of birth and gender.

Healthcare data linkage: hospital admissions

12342 (99.1%) mothers have at least one event recorded, the youngest at 10 years old, the oldest at 51 years old. There are 72112 events spanning a total of 3801 days. These are mostly maternity related.

12469 (90.0%) children have at least one event recorded, the youngest at birth, the oldest at 7 years old. There are 22299 events spanning a total of 2972 days. These are mostly birth related.

We matched participants to Bradford Royal Infirmary hospital records on the basis of NHS number and extracted up to 10 years of information on admissions, including diagnostic and procedure codes.

The denominator is unknown since we cannot estimate the likelihood of a hospital admission taking place at another hospital. So, we only count events.

We do, however, know that the match rate was high, given the number of participants with detected events.

Healthcare data linkage: Accident & Emergency

7054 (56.6%) mothers have at least one event recorded, the youngest at 14 years old, the oldest at 54 years old. There are 19671 events spanning a total of 2097 days.

9686 (69.9%) children have at least one event recorded, the youngest at birth, the oldest at 7 years old. There are 28072 events spanning a total of 2097 days.

Currently only a limited slice of data is available, from September 2009 to May 2015.

We matched participants to Bradford Royal Infirmary hospital records on the basis of NHS number and extracted information A&E visits.

The denominator is unknown since we cannot estimate the likelihood of an A&E visit taking place at another hospital. So, we only count events.

The match rate is known to be high from the hospital admissions events matching.

Residential address/grid reference

Baseline registration maternal address, updated at delivery of child if necessary.

Monthly name and address searches via PDS for all mothers and children since 2010

Addresses linked to Local Land Property Gazetteer to get grid reference for GIS analysis.

Residential address was provided by the mother at recruitment and updated when the baby was born if necessary.

Following that, data linkage via NHS number to the NHS Personal Demographic Service has provided monthly updates to residential address information.

Linkage to the Local Land Property Gazetteer provides a data cleaning function and grid reference data for GIS analyses.