








DATA NOTE

Metabolomics datasets in the Born in Bradford cohort

[version 1; peer review: awaiting peer review]

Kurt Taylor ^{1,2*}, Nancy McBride^{1-3*}, Neil J Goulding^{1,2}, Kimberley Burrows^{1,2}, Dan Mason⁴, Lucy Pembrey ⁵, Tiffany Yang ⁴, Razaq Azad⁶, John Wright ^{4,7}, Deborah A Lawlor ¹⁻³

¹Population Health Science, Bristol Medical School, University of Bristol, Bristol, BS8 2BN, UK

²MRC Integrative Epidemiology Unit, University of Bristol, Bristol, BS8 2BN, UK

³Bristol NIHR Biomedical Research Centre, University of Bristol, Bristol, BS1 2NT, UK

⁴Bradford Institute for Health Research, Bradford Hospitals National Health Service Trust, Bradford, BD9 6RJ, UK

⁵Department of Medical Statistics, London School of Hygiene & Tropical Medicine, London, UK

⁶Department of Biochemistry, Bradford Royal Infirmary, Bradford, UK

⁷Wolfson Centre for Applied Health Research, Bradford Hospitals National Health Service Trust, Bradford, BD9 6RJ, UK

* Equal contributors

V1 First published: 05 Nov 2020, 5:264
<https://doi.org/10.12688/wellcomeopenres.16341.1>

Latest published: 05 Nov 2020, 5:264
<https://doi.org/10.12688/wellcomeopenres.16341.1>

Abstract

Metabolomics is the quantification of small molecules, commonly known as metabolites. Collectively, these metabolites and their interactions within a biological system are known as the metabolome. The metabolome is a unique area of study, capturing influences from both genotype and environment. The availability of high-throughput technologies for quantifying large numbers of metabolites, as well as lipids and lipoprotein particles, has enabled detailed investigation of human metabolism in large-scale epidemiological studies. The Born in Bradford (BiB) cohort includes 12,453 women who experienced 13,776 pregnancies recruited between 2007-2011, their partners and their offspring. In this data note, we describe the metabolomic data available in BiB, profiled during pregnancy, in cord blood and during early life in the offspring. These include two platforms of metabolomic profiling: nuclear magnetic resonance and mass spectrometry. The maternal measures, taken at 26-28 weeks' gestation, can provide insight into the metabolome during pregnancy and how it relates to maternal and offspring health. The offspring cord blood measurements provide information on the fetal metabolome. These measures, alongside maternal pregnancy measures, can be used to explore how they may influence outcomes. The infant measures (taken around ages 12 and 24 months) provide a snapshot of the early life metabolome during a key phase of nutrition, environmental exposures, growth, and development. These metabolomic data can be examined alongside the BiB cohorts' extensive phenotype data from questionnaires, medical, educational and social record linkage, and

Open Peer Review

Reviewer Status *AWAITING PEER REVIEW*

Any reports and responses or comments on the article can be found at the end of the article.

other 'omics data.

Keywords

Metabolomics, mass spectrometry, nuclear magnetic resonance, pregnancy, mother, offspring health



This article is included in the [Born in Bradford gateway](#).

Corresponding authors: Kurt Taylor (kurt.taylor@bristol.ac.uk), Nancy McBride (nancy.mcbride@bristol.ac.uk)

Author roles: **Taylor K:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **McBride N:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **J Goulding N:** Writing – Review & Editing; **Burrows K:** Writing – Review & Editing; **Mason D:** Data Curation, Funding Acquisition, Writing – Review & Editing; **Pembrey L:** Data Curation, Funding Acquisition, Writing – Review & Editing; **Yang T:** Data Curation, Writing – Review & Editing; **Azad R:** Data Curation, Writing – Review & Editing; **Wright J:** Data Curation, Funding Acquisition, Writing – Review & Editing; **A Lawlor D:** Conceptualization, Data Curation, Funding Acquisition, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: DAL has received support from Medtronic LTD and Roche Diagnostics for biomarker research that is not related to work presented here. The other authors report no conflicts.

Grant information: Core funding for BiB has been provided by the Wellcome Trust [101597], a joint grant from the UK Medical Research Council (MRC) and UK Economic and Social Science Research Council (ESRC) [MR/N024397/1], the British Heart Foundation [CS/16/4/32482] and the National Institute for Health Research ARC Yorkshire and Humber [NIHR200166]. Funding for the metabolomics analyses has been provided by the US National Institutes of Health [R01 DK10324], the European Research Council (ERC) under the European Union's Seventh Framework Programme [FP7/2007-2013] / ERC grant agreement no 669545 and the MRC via the MRC Integrative Epidemiology Unit Programme to D.A.L [MC_UU_00011/6]. K.T. is supported by a British Heart Foundation Doctoral Training Program [FS/17/60/33474]. N.M. PhD studentship is funded by the National Institute for Health Research (NIHR) Biomedical Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol. K.T., N.M., N.J.G, K.B. and D.A.L. work in a unit that is supported by the University of Bristol and UK Medical Research Council [MC_UU_00011/6]. D.A.L is a NIHR Senior Investigator [NF-0616-10102]. The ALL IN sub-study was supported by the Wellcome Trust [083521]. Funding was also provided by the National Institute for Health Research (NIHR) under its Collaboration for Applied Health Research and Care (CLAHRC) for Yorkshire and Humber.

Copyright: © 2020 Taylor K *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Taylor K, McBride N, J Goulding N *et al.* **Metabolomics datasets in the Born in Bradford cohort [version 1; peer review: awaiting peer review]** Wellcome Open Research 2020, 5:264 <https://doi.org/10.12688/wellcomeopenres.16341.1>

First published: 05 Nov 2020, 5:264 <https://doi.org/10.12688/wellcomeopenres.16341.1>

Introduction

Metabolomics is the quantification of small molecules resulting from metabolic processes. The metabolome is influenced by both genotype and environment, and dynamically responds to environmental influences. Developments in high-throughput technologies have allowed the efficient and accurate quantification of metabolites. This has revolutionised our ability to understand the causes and consequences of variation in human metabolism, and the contribution that multiple metabolites can make to risk prediction, using large-scale epidemiological studies¹⁻⁴. Lipids and lipoproteins, which are measured in most high-throughput platforms used in epidemiology, are larger than the threshold used to define metabolites (<1.5k Daltons) and are therefore metabolomic traits. For simplicity in this paper we refer to these as metabolites.

Birth cohorts can be useful for exploring prenatal influences on birth and later life outcomes. Recently, studies have shown metabolomic profiling can aid us in our understanding of maternal health during pregnancy³⁻⁵ and of the influence of *in utero* exposures on subsequent offspring health^{6,7}. The Born in Bradford (BiB) study is a UK longitudinal birth cohort⁸. Nuclear magnetic resonance (NMR) and mass spectrometry (MS) data are available in BiB including measurements during pregnancy, cord blood and early life in the offspring. MS offers a truly untargeted approach with comprehensive

coverage of the metabolome (>1,000 metabolites) due to its high sensitivity. However, MS only provides relative quantification based on peak area in these approaches without comparison to a metabolite reference standard. NMR offers less coverage of the metabolome, but with absolute quantification possible in clinically meaningful units (e.g. mmol/L).

The range of metabolomics data in BiB, coupled with the substantial data obtained through questionnaires, research clinic assessments, linkage to medical records, educational and social records, genome wide (mothers, offspring and a subgroup of fathers) and epigenome wide (mother and offspring) profiling makes BiB a valuable resource for metabolomics research. This data note describes the metabolomics data currently available in BiB - how these were obtained, quantified, utilised, as well as potential future uses, strengths and limitations. [Figure 1](#) provides an illustrative summary of which type of metabolomic data have been collected on which cohort participants and when, up to 2020. Planned further metabolomic data collection is also described (see *Using the BiB metabolomic data*).

Methods

Ethical approval and consent

Ethical approval for the study was granted by the Bradford National Health Service Research Ethics Committee (ref

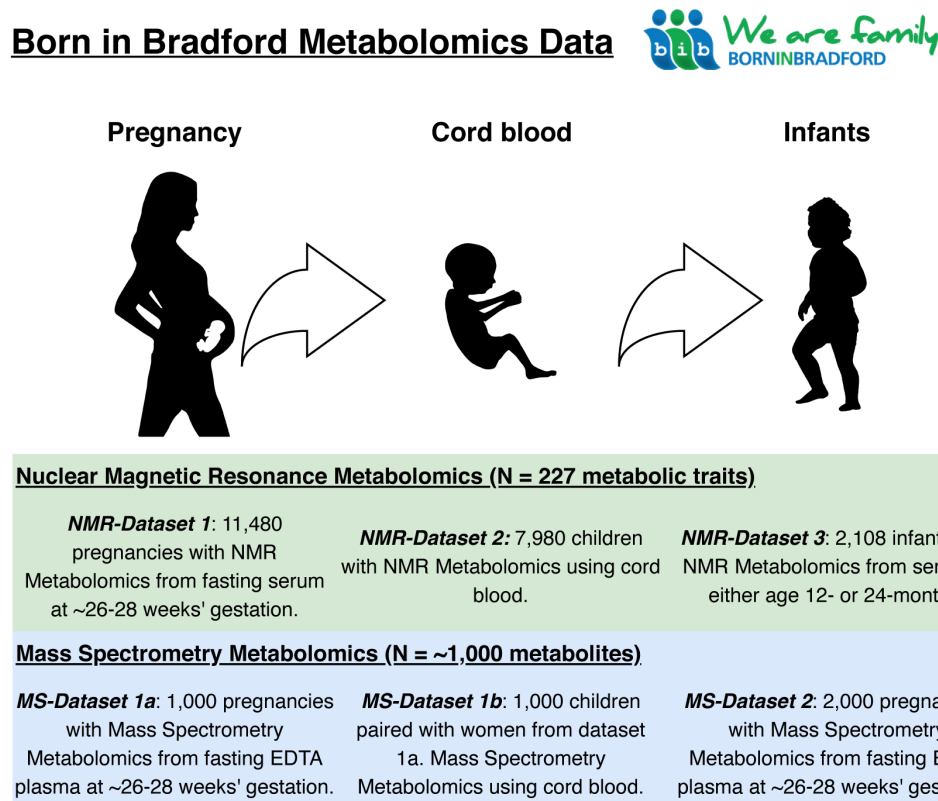


Figure 1. Summary illustration of the Born in Bradford metabolomics data. NMR, nuclear magnetic resonance; MS, mass spectrometry; EDTA, ethylenediaminetetraacetic acid.

06/Q1202/48), and all participants gave written informed consent. The ALL IN sub-study had ethical approval from the London School of Hygiene & Tropical Medicine ethics committee (ref: 5320) and the Bradford Research Ethics committee (ref: 08/H1302/21). Parents (usually the mother) gave informed, written consent to take part in the study.

Cohort

The BiB study is a population-based prospective birth cohort. In total, 12,453 women who experienced 13,776 pregnancies were recruited at their oral glucose tolerance test (OGTT) at approximately 26–28 weeks' gestation, which was offered to all women booked for delivery at Bradford Royal Infirmary (BRI). Eligible women had an expected delivery between March 2007 and December 2010. The study is unique because it includes high proportions of White European and South Asian families, all residing in Bradford, UK. Bradford is a city in the North of England with high levels of socio-economic deprivation, and the cohort was started due to a high prevalence of poor child health in the city⁸. Full details of the study methodology were reported previously⁸. The [study website](#) provides more information, including protocols, questionnaires and information on how researchers can access data and a full list of all available data. Mothers and their partners, who were recruited into the study, provided detailed interview questionnaire data, measurements, and biological samples. They also consented to the linkage of their and their child's data to routine (primary and secondary care) health and education data.

Blood sampling

Maternal overnight-fasted blood was taken during the OGTT and processed and stored at -80°C for further research and analyses. Infant cord blood samples were taken whenever possible (i.e. so long as staff were available, and collection of an umbilical vein sample did not interfere with care of the mother or infant) and immediately processed and stored at -80°C . Samples were taken in a subgroup of offspring in early childhood for a specific project on childhood viral infection⁹. We describe the processes of taking, processing, and storing samples at each time point before moving on to describe the NMR and MS metabolomic profiling.

Pregnancy blood samples. Of the 13,776 pregnancies in the BiB cohort, 11,480 had a fasting blood sample taken during the OGTT ($n = 10,574$ [92%]) between 26–28 weeks' gestation, with the remaining women being within 11–39 weeks' gestation). Samples were taken by trained phlebotomists working in the antenatal clinic of the BRI and sent immediately to the hospital laboratory.

Venous blood was collected in GEL tubes to obtain serum and plasma. The following processing steps were undertaken prior to storage at -80°C .

- 1) Storage racks were prepared.
- 2) Participant details were checked, making sure that both the BiB study ID and hospital number on the specimen bottles matched those on the participant tracking forms.

- 3) Tubes were centrifuged at 3500 rpm for 10 minutes at room temperature.
- 4) A 1 ml automatic pipette was used to aliquot samples into 2×1.5 ml vials.
- 5) Vials were labelled with appropriate BiB study labels and the duplicate barcode label was placed in the corresponding space marked on BiB tracking form.
- 6) Aliquots ($n = 3$) were then placed in racks in a -80°C freezer.

All samples were processed within 2.5 hours and then placed in -80°C freezers. There were no freeze-thaw events of the samples prior to their use for the pregnancy metabolomic profiling. Serum samples were used for NMR metabolomic profiling, except for five (0.04%) samples which were plasma. For MS pregnancy metabolomics, ethylenediaminetetraacetic acid (EDTA) (a sample tube anticoagulant) plasma samples were used. Previous work has shown that reproducibility in both serum and plasma is good. As long as the same blood sample procedures are used (as in BiB), either matrix should yield similar results¹⁰.

Cord blood samples. Venous cord blood samples were all obtained at delivery by the attending midwife at the BRI, following research protocols. Cord blood sampling was not attempted for women delivering outside of the BRI, if the attending midwife was too busy, or if attempting to collect a research cord blood sample would interfere with postnatal care. Samples were refrigerated at 4°C in EDTA tubes until collected by BRI laboratory staff within 12 hours. Samples were then spun, frozen and stored at -80°C . In total, the BiB study collected 9,604 cord blood EDTA plasma samples. There were no freeze-thaw events of the cord blood samples.

Infant blood samples. Infant metabolomics were performed on blood samples that were collected on a subsample of the BiB cohort; those enrolled into the Allergy and Infection Study (ALL IN)⁹. Children enrolled in the BiB cohort, and born on or after 1 March 2008 with a maternal baseline questionnaire were eligible to take part in ALL IN. Mothers were invited to participate in ALL IN one month before their child's first birthday. A questionnaire was completed by those who consented, and a 5ml venous blood sample was taken from the child, centrifuged, and stored at -80°C . This was repeated one year later to provide questionnaire data and serum from a ~12-month visit (mean age of 14 months, ranging from 9–18 months) and a ~24-month visit (mean age of 26 months, ranging from 23–33 months). Trained community research administrators (CRAs) recruited participants, obtained consent, and collected data, including blood samples, at each visit. They received training in phlebotomy and were assessed by the senior paediatric phlebotomist at the BRI. Ametop cream or Cryogestic spray were used to anaesthetise the venepuncture site. Only two attempts at venepuncture were permitted for each child. There was a fridge in the clinic for storing bloods before transfer to the lab. The blood samples taken on home visits were kept in a cool bag with an

ice pack and then taken straight to the laboratory at BRI within 1–2 hours. The times of each step (blood taken, arrived at lab, centrifuged, aliquoted, frozen) were recorded on the blood form and were entered onto a database (so that researchers can check distribution of times if needed). For home or clinic visits outside normal working hours, the CRA who took the blood sample would centrifuge the blood at the lab and leave it in the lab fridge for processing the next day. All infant metabolomics were performed on serum samples. There was a maximum of two freeze-thaw events prior to metabolomics analyses of the infant samples.

Metabolomic datasets in BiB

There are six metabolomics datasets including different populations and timepoints available in BiB. These are described below and summarized in [Table 1](#) and [Figure 1](#). We have divided the methods between the two main platforms (NMR and MS). We describe the methods used to generate each dataset and use flow charts to illustrate how selection was performed.

NMR metabolomics

NMR methods. We describe the NMR methods which apply to all the NMR datasets described in [Table 1](#). Profiling of circulating lipids, fatty acids, and metabolites was done by a high-throughput targeted NMR platform (Nightingale Health© (Helsinki, Finland)) at the University of Bristol, providing quantitative information on 227 metabolic traits (including ratios and other traits derived from the quantified NMR spectra)¹. Details of all 227 traits can be found in the [Extended data](#)¹¹.

The Nightingale NMR metabolite quantification was achieved through measurements of three molecular windows from each serum/plasma sample. Two of the spectra (LIPO and LMWM windows) are acquired from native serum/plasma and one spectrum from serum lipid/plasma extracts (LIPID window). The NMR spectra are measured using Bruker AVANCE III spectrometer operating at 600 MHz. Measurements of native serum/plasma samples and serum/plasma lipid extracts are conducted at 37°C and 22°C, respectively.

The NMR spectra were analysed for metabolite quantification (molar concentrations) in an automated fashion. For each metabolite, a ridge regression model was applied for quantification to overcome the problems of heavily overlapping spectral data. In the case of the lipid data, quantification models were calibrated using high-performance liquid chromatography methods, and individually cross-validated against NMR-independent lipid data. Low-molecular-weight metabolites, as well as lipid extract measures, were quantified as mmol/L based on regression modelling calibrated against a set of manually fitted metabolite measures. The calibration data were quantified based on iterative line-shape fitting analysis using PERCH NMR software (PERCH Solutions Ltd., Kuopio, Finland). Quantification could not be directly established for the lipid extract measures due to experimental variation in the lipid extraction protocol. Therefore, serum/plasma lipid extract were scaled to total a standard serum cholesterol sample from the LIPO spectrum.

Validation of the NMR platform. Quality control (QC) of the data were undertaken by Nightingale Health© prior to returning metabolite concentrations to BiB. Their QC procedures

Table 1. Metabolomics datasets in the BiB cohort separated by platform.

#	Data source	Brief description
Nuclear magnetic resonance		
1	Pregnancy NMR – Dataset 1	N = 11,480 pregnancies. Single timepoint using maternal serum taken from a fasted blood sample around 26–28 weeks' gestation. Of the 11,480, 37% are White British (40% White European) mothers and 44% Pakistani (49% South Asian).
2	Cord blood NMR – Dataset 2	N = 7,980 children. Single timepoint using cord blood, EDTA plasma.
3	Infants NMR (aged 12 or 24 months) – Dataset 3	N = 2,108 at either 12- or 24-months using serum samples. N = 1,690 at 12 months. N = 1,536 at 24 months. N = 1,118 at both timepoints.
Mass Spectrometry		
4	Pregnancy MS – Dataset 1a	N = 1,000 pregnancies. Single timepoint using EDTA plasma taken from a fasted blood sample between 26–28 weeks' gestation. Of the 1,000, 50% are White British and the other 50% are Pakistani ethnicity.
5	Cord blood MS – Dataset 1b	N = 1,000 children (paired with women from Dataset 1a). Single timepoint using cord blood, EDTA plasma.
6	Pregnancy MS – Dataset 2	N = 2,000 pregnancies within a case-cohort design. EDTA plasma sample taken between 26–28 weeks' gestation. Of the 2,000 women, 47% are White British and 53% are Pakistani.

NMR, nuclear magnetic resonance; MS, mass spectrometry; EDTA, ethylenediamine tetraacetic acid.

check various issues related to the sample integrity and the biomarker quantification. QC reports for the NMR datasets can be found in the [Extended data](#)¹¹.

We also undertook validation of some of the NMR measures by comparing concentrations of fasting glucose, total

cholesterol, high-density lipoprotein cholesterol (HDLc), low-density lipoprotein cholesterol (LDLc), and triglycerides from the NMR platform to the same measures from the same samples assessed by clinical chemistry measurements ([Figure 2](#)). Clinical chemistry measurements were completed at the BRI laboratory (fasting glucose) or Glasgow Royal Infirmary (lipids).

NMR versus Clinical Chemistry

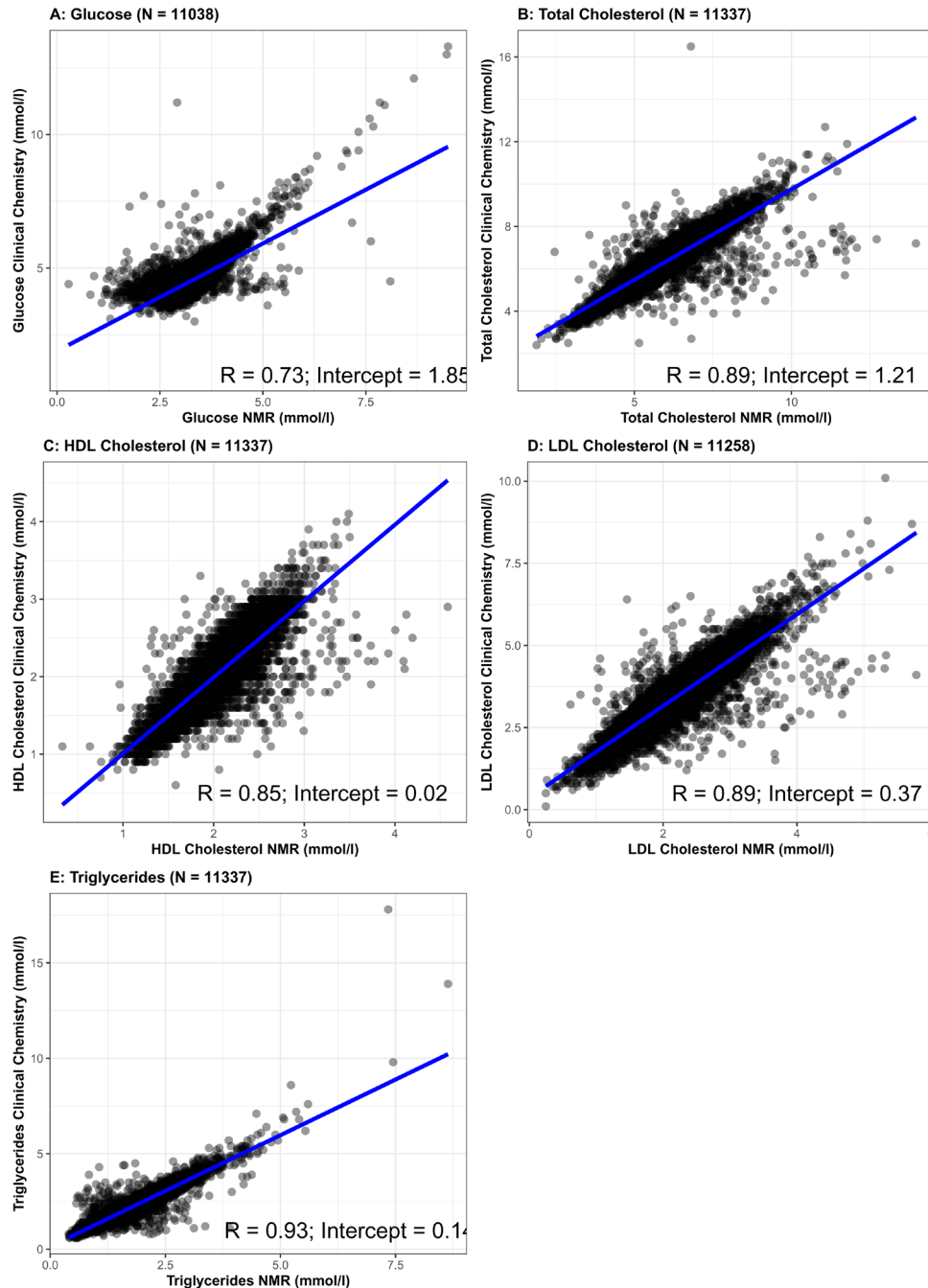


Figure 2. Comparison of glucose (**2A**), total cholesterol (**2B**), high-density lipoprotein cholesterol (**2C**), low-density lipoprotein cholesterol (**2D**) and triglycerides (**2E**) concentrations between Nightingale Health© nuclear magnetic resonance (NMR) (x-axis) and routine clinical chemistry assays (y-axis) (N = 11,036 to 11,337). R = Pearson correlation coefficient.

Glucose was measured using a glucose oxidase method that does not cross-react with insulin. Total cholesterol, HDLc and triglycerides were measured following the standard Lipid Research Clinics Protocol using enzymatic reagents. LDLc was estimated from total cholesterol, HDLc and triglycerides ($\text{LDLc} = [\text{Total cholesterol in mmol/l}] - [\text{HDLc in mmol/l}] - [\text{Triglycerides in mmol/l} \div 2.2]$). The correlation between fasting glucose measured by clinical chemistry and by NMR was 0.73 and for all four lipids was between 0.85 and 0.93, with the intercepts of the regression line close to zero for HDLc, LDLc, and triglycerides, but higher for glucose (1.85) and total cholesterol (1.21). This suggests that the NMR platform systematically underestimates glucose and total cholesterol levels. However, the high levels of correlation, particularly for the lipid measures, is reassuring and suggests association analyses would have validity. To further test the validity of the NMR measures, we compared associations of maternal early pregnancy body mass index (BMI), treated as an exposure, with fasting glucose, and the four lipid measures from clinical chemistry and NMR as the outcome. We also compared associations between the five metabolic measures (from clinical chemistry and NMR) as exposures, with hypertensive disorder of pregnancy (HDP; either gestational hypertension or pre-eclampsia, defined on the basis of international criteria applied to all measures of blood pressure and proteinuria extracted from clinical records)¹² as the outcome. Associations of BMI with the five outcomes were directionally consistent between clinical chemistry and NMR measurements. However, the NMR associations were weaker (closer to the null) and there were clear differences in magnitudes of association between the two methods for the associations of BMI with glucose and HDLc (Figure 3A). By contrast, results were concordant between the two methods for the associations of metabolites with odds ratios of HDP (Figure 3B).

Participant selection and characteristics of those with NMR data. In this subsection, we present flow charts to illustrate selection and inclusions into the NMR participant groups (Figure 4) and describe participant characteristics for the BiB NMR datasets (Table 2). All three of the samples of BiB participants with NMR data (maternal pregnancy N = 11,480, offspring cord blood N = 7,980, and offspring 12–24 months N = 2,108) had very similar distributions of maternal age, parity, early pregnancy BMI, residential area deprivation, offspring sex and birth weight to those seen in the whole cohort of 13,776 participants (Table 2).

Mass spectrometry metabolomics

Mass spectrometry methods. The untargeted MS metabolomics analysis of over 1,000 metabolites was performed at Metabolon, Inc. (Durham, North Carolina, USA). Samples were sent to Metabolon in two separate batches. Dataset 1 was completed in December 2017 and consisted of 1,000 maternal pregnancy samples and 1,000 offspring paired cord blood samples. Dataset 2 was completed in December 2018 and consisted of 2,000 maternal pregnancy samples.

At Metabolon, samples were managed by a laboratory information management system and were kept at -80°C .

Recovery standards were added to samples prior to monitor the extraction process. To remove proteins, dissociate small molecules bound to proteins, disassociate molecules trapped in the precipitated protein matrices, and to recover chemically diverse metabolites, proteins were precipitated with methanol under vigorous shaking for 2 min (Glen Mills GenoGrinder 2000) followed by centrifugation. The resulting extract was divided into five fractions: two for analysis by two separate reverse phase ultra-high-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) methods with positive ion mode electrospray ionization (ESI), one for analysis by reverse phase UPLC-MS/MS with negative ion mode ESI, one for analysis by hydrophilic interactive liquid chromatography (HILIC)/UPLC-MS/MS with negative ion mode ESI, and one sample was reserved for backup. Samples were placed on a TurboVap® (Zymark) to remove the organic solvent. The sample extracts were stored overnight under nitrogen before preparation for analysis. The four independent UPLC-MS/MS methods provide relative quantification of >1,000 metabolites in 10 key classes: amino acids, carbohydrates, lipids, nucleotides, microbiota metabolism, carbon metabolism, energy, cofactors & vitamins, xenobiotics, and unidentified metabolites. A list of metabolites defined in each of the datasets can be found in the [Extended data](#)¹¹. Metabolite concentrations were quantified using area-under-the-curve of primary MS ions and were expressed as the means of the medians (MoM) value for all batches processed on the given day.

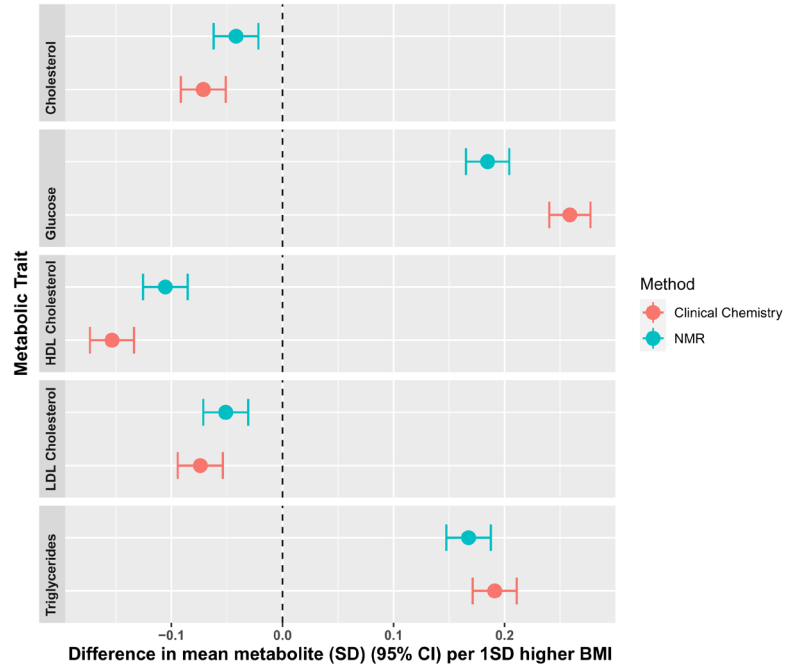
Validation of the MS platform. Details of the Metabolon QC procedures and data quality for the Metabolon BiB datasets are described in the reports found in the [Extended data](#)¹¹. In brief, procedures were conducted to: (i) assure that all aspects of the Metabolon process are operating within specifications, (ii) assess the effect of a non-plasma matrix on the Metabolon process and distinguish biological variability from process variability, (iii) assess the contribution to compound signals from the process (using Process Blank) and (iv) segregate contamination sources in the extraction (using Solvent Blank).

As an additional data QC, we explored correlations between MS and both NMR and clinical chemistry fasting glucose measures (glucose is the only common trait we have data on for MS, NMR, and clinical chemistry). Pearson's correlation coefficient comparing MS to clinical chemistry (0.65) was modest and lower than that for NMR (0.73, see above and Figure 2) and the intercept was 0.11 (Figure 5A). Correlation between Metabolon and NMR was higher (0.77) and the intercept was 0.10 (Figure 5B).

Participant selection and characteristics of those with MS data. The flow of participants into the MS datasets are illustrated in Figure 6, and the characteristics of participants included in the two MS datasets provided in Table 3. Only women of either Pakistani or White British ethnic background were included in the MS datasets because, due to cost, we were only able to do this on a subset of the cohort. As these two groups represent ~85% of BiB it was felt the numbers for any

NMR versus Clinical Chemistry

A: BMI (exposure) and metabolic traits (outcome)



B: Metabolic traits (exposure) and hypertensive disorders of pregnancy (outcome)

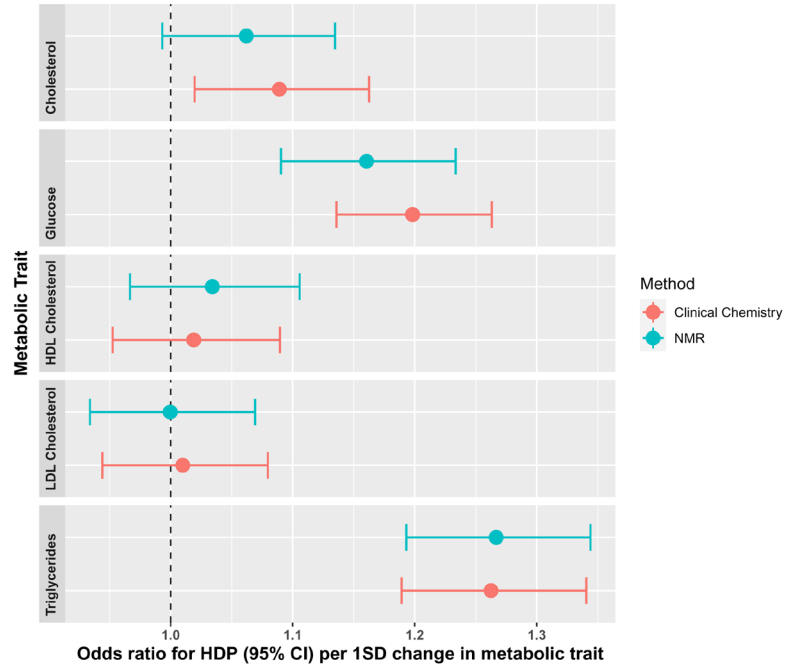


Figure 3. Comparisons of the associations of early pregnancy body mass index (BMI) with fasting glucose and lipids measured by Nightingale Health© nuclear magnetic resonance (NMR) and routine clinical chemistry assays (**3A**), fasting glucose and lipids measured by Nightingale Health© NMR and routine clinical chemistry assays with hypertensive disorder of pregnancy (HDP) (**3B**). Associations in **3A** are from unadjusted linear regression and data points show standard deviation differences in mean metabolite per one standard deviation (1SD) higher BMI. Associations in **3B** are from unadjusted logistic regression and data points show unadjusted odds ratios for HDP per 1SD higher in metabolic trait. Error bars = 95% confidence intervals.

Nuclear Magnetic Resonance Metabolomics in the Born in Bradford Cohort

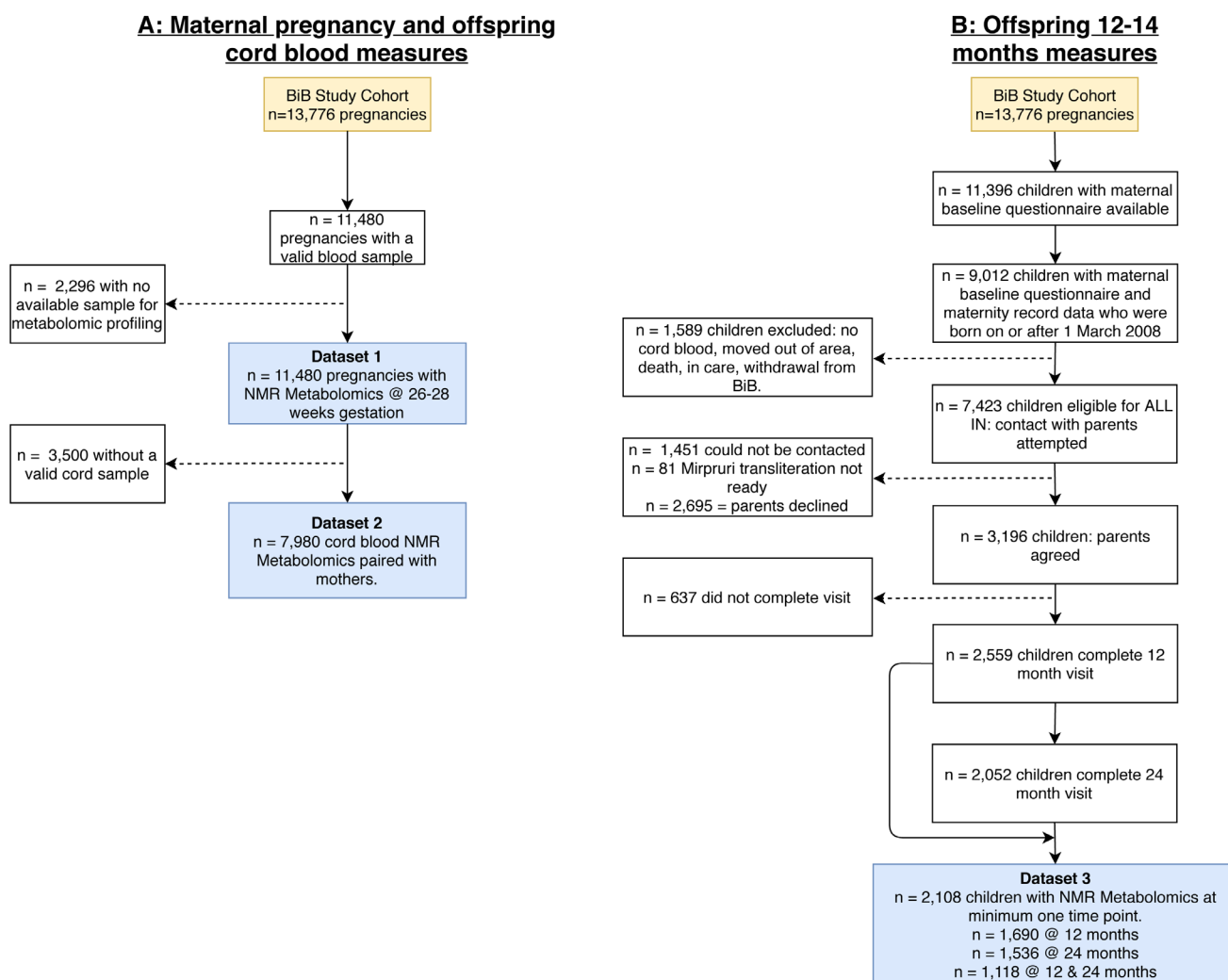


Figure 4. Illustrating the flow of participants into the NMR datasets in the Born in Bradford cohort. Figure 4A shows the maternal pregnancy (Dataset 1: NMR metabolomics at 26–28 weeks' gestation) and offspring cord blood samples (Dataset 2: NMR metabolomics taken from the umbilical vein shortly after delivery). Figure 4B shows the offspring 12–24 months NMR metabolomic sample (Dataset 3). Abbreviations: NMR, nuclear magnetic resonance; BiB, Born in Bradford; ALL IN, Allergy and Infection study.

other group would be too small for meaningful analyses. In Dataset 1, 1,000 women were selected on the basis that they had stored fasting plasma, a useable cord blood sample, genome wide data on both mother and offspring and were either of White British or Pakistani origin (Figure 6A). Following these inclusions, 500 women were selected at random from each ethnic group (White British and Pakistani). In Dataset 2, a case-cohort design was used^{13,14}. A case-cohort design consists of a cohort with an over-sampling of all cases. The BiB case-cohort consists of 2,000 women (only pregnancy samples were assayed in Dataset 2). As with Dataset 1, women were

selected based on certain characteristics shown in Figure 6B, including that they had not already had Metabolon MS analyses. From those who fulfilled these pre-specified criteria, six groups of cases were selected: women with (a) gestational diabetes; (b) gestational hypertension; (c) pre-eclampsia; (d) preterm birth; (e) congenital anomaly; (f) stillbirth. In total, 801 women had experienced one or more of these conditions. Having selected all cases these were then replaced into the eligible cohort and a sub-cohort of 1,199 women were randomly selected from the eligible cohort. Thus, the comparison group in this case-cohort study is representative of the eligible cohort

Table 2. Participant characteristics for NMR datasets in the BiB cohort.

		Maternal pregnancy NMR dataset (n=11,480)	Offspring cord blood NMR dataset (n=7,890)	Offspring 12- or 24-months NMR dataset (n = 2,108)	BiB cohort (n=13,776)
Characteristics	Unit / Category				
Maternal Age	Years	27.3 (5.6)	27.5 (5.6)	28.3 (5.7)	27.3 (5.6)
	Missing	410 (3.6)	627 (7.9)	60 (2.9)	1445 (10.5)
Maternal Parity	Nulliparous	4310 (37.5)	2765 (36.6)	819(39.9)	5101 (37.0)
	Multiparous	6428 (55.9)	5125 (65.0)	1,233(58.4)	7773 (56.4)
	Missing	742 (6.5)	344 (4.4)	56 (2.7)	902 (6.5)
Maternal BMI	kg/m ²	26.1 (5.7)	26.2 (5.7)	26.2(5.5)	26.0 (5.7)
	Missing	2160 (18.8)	1464 (18.5)	106 (5.0)	3281 (23.8)
Maternal ethnicity	White British	4268 (37.2)	2902 (37.7)	769 (49.7)	5055 (37.8)
	Pakistani	4995 (43.5)	3596 (46.7)	1048 (49.7)	6088 (45.5)
	Other	1887(16.4)	1206 (15.7)	291 (13.8)	2223 (16.6)
	Missing	330 (2.4)	186 (2.4)	0	410 (3.0)
Index of multiple deprivation	Quintile 1 (most deprived)	6646 (65.9)	4439 (65.8)	1400(66.4)	7566 (66.4)
	Quintile 2	1830 (18.2)	1220 (18.1)	355 (16.8)	2052 (18.0)
	Quintile 3	1124 (11.2)	778 (11.5)	248 (11.8)	1250 (11.0)
	Quintile 4	306 (3.0)	187 (2.8)	70 (3.3)	334 (2.9)
	Quintile 5 (least deprived)	173 (1.7)	118 (1.8)	34(1.6)	190 (1.7)
	Missing	1401 (12.2)	1148 (14.6)	1 (0.0)	2384 (17.3)
Offspring sex	Male	5705 (49.7)	4095 (51.9)	1065(50.2)	6891 (50.0)
	Female	5420 (48.7)	3795 (48.1)	1029(48.1)	6470 (48.4)
	Missing	355 (3.1)	3 (0.0)	14 (0.7)	415 (3.0)
Birth weight	Grams	3226 (565)	3266 (522)	3224 (558)	3216 (565)
	Missing	356 (3.1)	4 (0.1)	14 (0.7)	416 (3.0)

Data are means \pm SD or n (%) unless stated. ^ gestational age in weeks presented for columns 1, 2 and 3. Offspring age in weeks presented for column 4.

Abbreviations: NMR, nuclear magnetic resonance; BiB, Born in Bradford; BMI, body mass index; kg, kilogram; IMD, index of multiple deprivation (taken from 2010 national quintiles). There were 9 ethnic groups, of which White British and Pakistani were the main homogeneous groups. The 'Other' ethnicity category comprises: White Other, Mixed-White and Black, Mixed-White and South Asian, Black, Indian, Bangladeshi or Other ethnicity.

(i.e. the cohort comparison group includes some of the cases in proportions that would reflect the whole cohort). The final BiB case-cohort sample consists of three groups (Figure 6B): 1) selected as comparison group (N = 1,199), 2) selected as cases only (N = 408), and 3) selected as a case and control (N = 393). The comparison group in any analyses will vary depending on the research question.

For the MS dataset, researchers are given the option of using the 'raw' data from Metabolon or a quantified (scaled) data set, in which missing data have been imputed and the multiple of median values transformed to standard deviation-(z-) scores (by subtracting the sample mean value for each metabolite from the participant value and then dividing by the sample

standard deviation for that metabolite). This transformation helps overcome the problem of high missing data in metabolomics¹⁵. This cohort for MS profiling were sampled on their ethnicity. It is almost 50% White British and Pakistani (there are slightly more Pakistani women in Dataset 2), as opposed to around 15% of the whole BiB cohort not belonging to either of these ethnic groups. However aside from this, the sample is representative of the whole cohort (Table 3).

Overlap between metabolomics datasets

Having participants in multiple datasets (i.e. maternal pregnancy, offspring cord, offspring 12–24 months) and across the two metabolomic platforms provides scope for unique research opportunities. Figure 7 illustrates the overlap between BiB

Mass Spectrometry Glucose Correlations

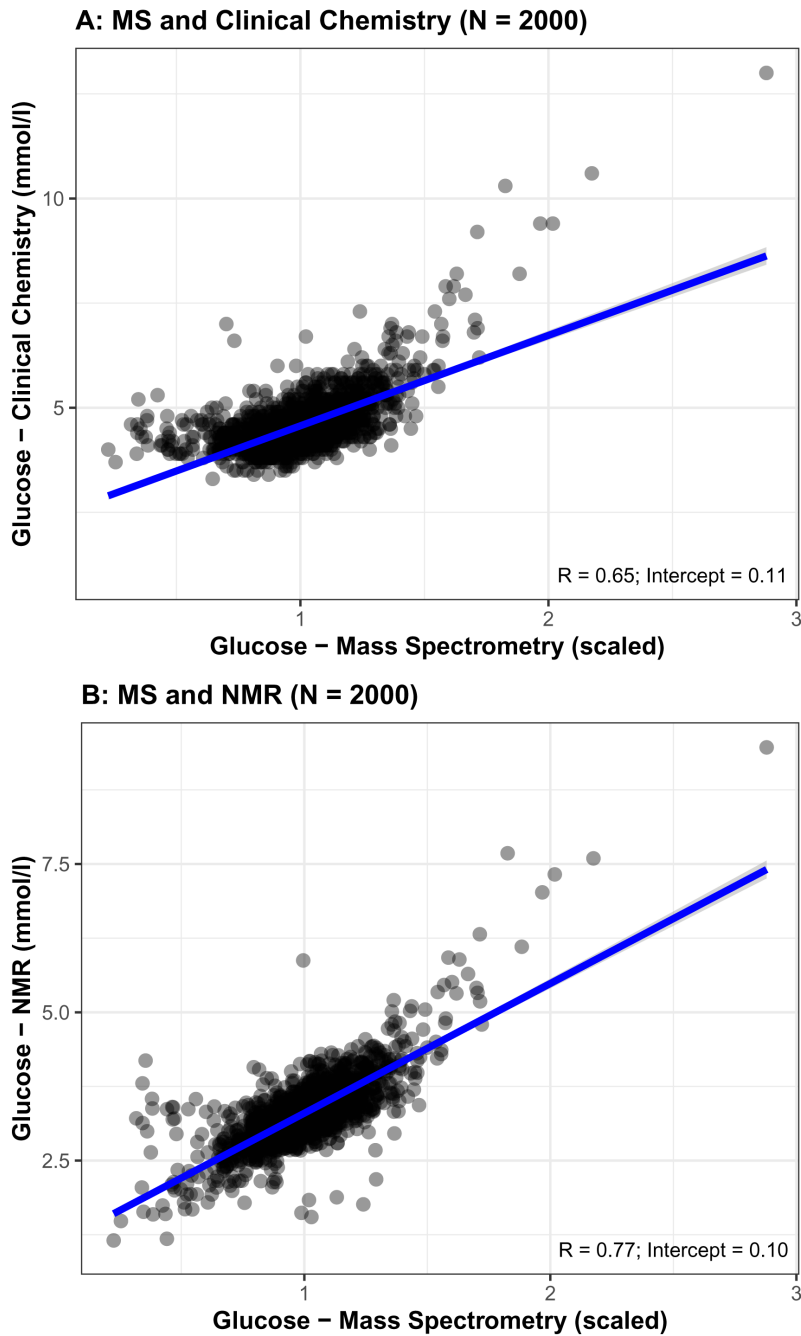


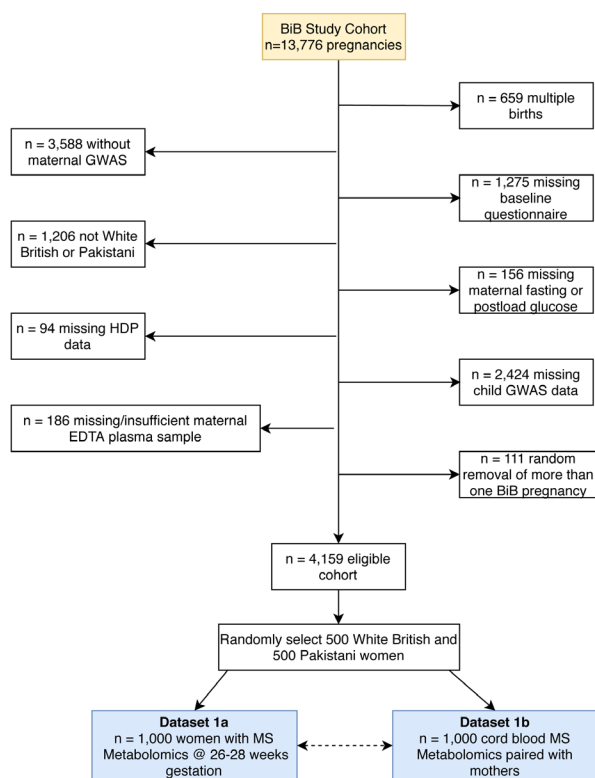
Figure 5. Comparisons of glucose concentrations for Metabolon mass spectrometry (MS) with routine glucose oxidase (**5A**) and Nightingale Health© nuclear magnetic resonance (NMR) (**5B**). R = Pearson correlation coefficient.

metabolomic datasets. The numbers are all based around the offspring, for example the number of maternal pregnancy metabolite data in any cell refer to the number of offspring

who have a mother with those samples. There were 11,557 children from 11,480 pregnancies whose mothers had a pregnancy NMR sample. Of these, 6,756 children also had a cord

Mass Spectrometry Metabolomics in the Born in Bradford Cohort

A: Dataset 1: n = 1000 women/offspring pairs



B: Dataset 2: n = 2000 women (case-cohort design)

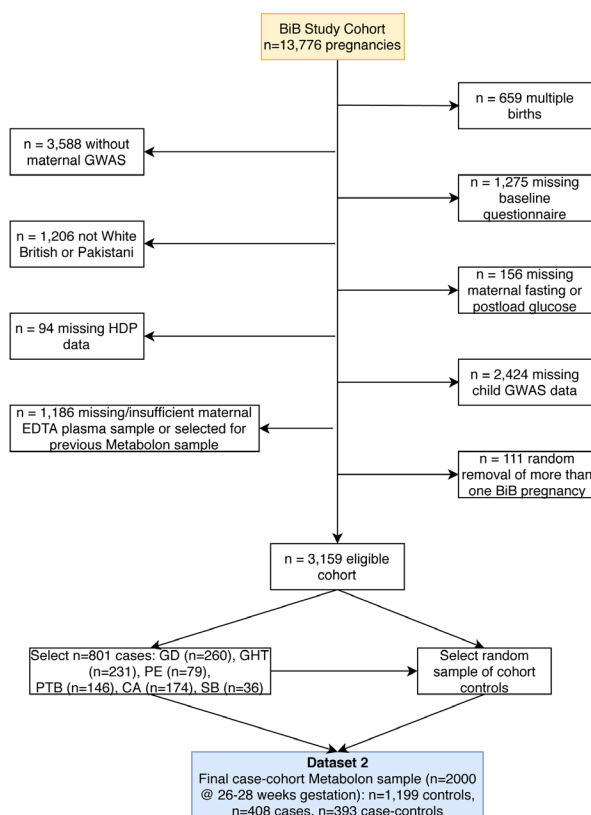


Figure 6. Illustrating the flow of participants into the Metabolon datasets in the Born in Bradford cohort. Figure 6A shows dataset 1 which includes 1,000 pregnancies and infants with MS metabolomics during pregnancy (26–28 weeks' gestation (dataset 1a)) and in cord blood (dataset 1b). Figure 6B shows dataset 2 which includes 2,000 pregnancies (26–28 weeks' gestation) with MS metabolomics within a case-cohort design. Abbreviations: MS, mass spectrometry; BiB, Born in Bradford; GWAS, genome wide association study; EDTA, ethylenediaminetetraacetic acid; HDP, hypertensive disorders of pregnancy; GD, gestational diabetes; GHT, gestational hypertension; PE, pre-eclampsia, PTB, preterm birth; CA, congenital anomaly; SB, still birth.

blood sample and 1,981 had at least one measurement from either the 12- or 24-month ALL IN subsample. All the mothers with a pregnancy MS sample (from either the first or second dataset) also have an NMR sample. There were 7,919 children in total with an NMR sample in cord blood with 1,275 of these also having at least one measure from the 12- or 24-month subsample. Of those with NMR cord blood data, 2,486 had a mother with MS pregnancy data (from either the first or second dataset) and 1,000 have MS cord blood data. There were 2,108 children with at least one NMR measure at either the 12- or 24-months assessment and of these, 690 have a mother with MS metabolite measures in pregnancy

data (from either dataset) and 229 have MS cord blood data. Although the exclusion criteria for MS dataset 2 was no prior MS metabolomics (Figure 6), there was one mother with MS metabolomics in both datasets from different pregnancies.

Using the BiB metabolomic data, including a summary of published, ongoing and future research using these data

The current BiB metabolite data have been quantified on blood samples collected during pregnancy, cord blood at birth and in the offspring at 12- and 24-months. These are critical time periods for life-course research and the combination of

Table 3. Participant characteristics of the mass spectrometry datasets in the BiB cohort.

		Dataset 1 (N = 1,000 mother/child pairs)	Dataset 2 case-cohort ^a (N = 2,000)	Dataset 2 random cohort sample only ^b (N = 1,199)	BiB cohort (N = 13,776)
Characteristics	Category	-	-	-	-
Maternal age	Years	27.5 (5.7)	27.5 (5.7)	26.91 (5.5)	27.3 (5.6)
	<i>Missing</i>	0 (0)	0 (0)	0 (0.0)	1445 (10.5)
Maternal parity	Nulliparous	359 (37.0)	745 (37.3)	433 (37.4)	5101 (37.0)
	Multiparous	611 (61.1)	1213 (60.1)	725 (60.5)	7773 (56.4)
	<i>Missing</i>	30 (3.0)	42 (2.1)	41 (3.4)	902 (6.5)
Maternal BMI	(kg/m ²)	26.7 (6.0)	26.8 (5.9)	25.9 (5.4)	26.0 (5.7)
	<i>Missing</i>	36 (3.6)	97 (4.9)	60 (5.0)	3281 (23.8)
Maternal ethnicity	White British	500 (50.0)	933 (46.7)	537 (44.8)	5055 (37.8)
	Pakistani	500 (50.0)	1067 (53.4)	662 (55.2)	6088 (45.5)
	Other	0	0	0	2223 (16.6)
	<i>Missing</i>	0	0	0	410 (3.0)
Index of multiple deprivation	Quintile 1 (most deprived)	656 (65.6)	1340 (67.0)	823 (68.6)	7566 (66.4)
	Quintile 2	175 (17.5)	358 (17.9)	203 (16.9)	2052 (18.0)
	Quintile 3	112 (11.2)	212 (10.6)	123 (10.3)	1250 (11.0)
	Quintile 4	38 (3.8)	53 (2.6)	31 (2.6)	334 (2.9)
	Quintile 5 (least deprived)	19 (1.9)	37 (1.8)	19 (1.6)	190 (1.7)
	<i>Missing</i>	0 (0)	0 (0)	0 (0)	2384 (17.3)
Offspring sex	Male	512 (51.2)	1053 (52.7)	625 (52.1)	6891 (50.0)
	Female	488 (48.8)	947 (47.3)	574 (47.9)	6470 (48.4)
	<i>Missing</i>	0 (0)	0 (0)	0 (0)	415 (3.0)
Offspring birthweight	Grams	3304 (517)	3232 (574)	3318 (486)	3216 (565)
	<i>Missing</i>	0 (0)	1 (0)	0 (0)	416 (3.0)

^a This column comprises of the full case-cohort dataset of 2,000 pregnancies. This includes 801 selected cases and the 1,199 random cohort.

^b This column includes only the 1,199 random cohort to compare to the full case-cohort with the selected cases.

Data are means \pm SD or n (%) unless stated. Abbreviations: BiB, Born in Bradford; BMI, body mass index; kg, kilogram; IMD, index of multiple deprivation (taken from 2010 national quintiles). There were nine ethnic groups, of which White British and Pakistani were the main homogeneous groups. The 'Other' ethnicity category comprises: White Other, Mixed-White and Black, Mixed-White and South Asian, Black, Indian, Bangladeshi or Other ethnicity.

these data with large amounts of genomic, epigenomic, social and health data makes BiB a platform which provides scope for unique research opportunities.

Issues for data users

1. Batch effects: The quantified NMR metabolites that have been measured in BiB are represented in clinically meaningful units, so can be compared to results from other studies. By contrast the Metabolon MS metabolites are quantified in relative abundance i.e. in relation to other quantified MS measurements that were run on the same day. The MS Dataset 1 and Dataset 2 were obtained ~2 years apart and have been

normalized to different references, so are not directly comparable. For example, the value of a specific metabolite from a maternal pregnancy sample in Dataset 1 compared to the same metabolite in Dataset 2 may differ because they are from different batches. Because of the different selection process for the two datasets (Dataset 1 is paired pregnancy-offspring cord blood samples and Dataset 2 has a case-cohort sampling frame) it is not possible to normalize them to the same reference. We recommend running analyses separately in each of the two datasets and comparing results, then meta-analyse if appropriate. In principal components analyses, 37 of the 1,000 women with a pregnancy sample in

	NMR Pregnancy Dataset 1	NMR Cord Dataset 2	NMR Infant Dataset 3	MS Pregnancy Dataset 1a	MS Cord Dataset 1b	MS Pregnancy Dataset 2
NMR Pregnancy Dataset 1	11557	6756	1982	1000	1000	2000
NMR Cord Dataset 2		7919	1275	1000	1000	1486
NMR Infant Dataset 3			2108	229	229	461
MS Pregnancy Dataset 1a				1000	1000	1000
MS Cord Dataset 1b					1000	1000
MS Pregnancy Dataset 2						2000

Figure 7. Showing the overlap between the metabolomic datasets in the Born in Bradford cohort presented at the offspring level. Abbreviations: NMR, Nuclear magnetic resonance; MS, mass spectrometry.

Dataset 1 of MS data had notably different values to those in the remaining 963 women, which appeared to be a batch effect (see Figure in [Extended data¹¹](#)). This is a new finding and in previous analyses using these data we have not treated these 37 women differently. However, for future analyses we would recommend researchers consider running analyses with all women and in a sensitivity analysis with these 37 women removed.

2. *Comparisons with clinical chemistry measurements:* We have illustrated above strong correlations between glucose and lipids measured using clinical chemistry and the NMR platform. We found weaker (though directionally consistent) associations of BMI with these outcomes measured using NMR compared to those with clinical chemistry. In a second example, results were consistent between the two methods for the associations of pre-eclampsia with glucose and lipids. Researchers considering using these data might want to check for consistency with associations using the clinical chemistry measurement available in BiB. For the MS data we were only able to explore correlations with glucose and found this to be high between clinical chemistry and MS.

Summary of published research using the BiB metabolomics data

We undertook a collaboration between BiB and the UK Pregnancies Better Eating and Activity Trial (UPBEAT), a randomised

control trial of obese pregnant women ($\text{BMI} \geq 30\text{kg/m}^2$)¹⁶. We found evidence that maternal pregnancy NMR samples can improve prediction of pregnancy related disorders¹⁶. The prediction models consisting of NMR-derived metabolomics and established risk factors (maternal age, smoking, BMI, ethnicity, and parity) performed better than established risk factors alone for gestational diabetes, hypertensive disorders of pregnancy, small/large for gestational age but not preterm birth in BiB. We found directionally consistent, but attenuated, results in UPBEAT. The attenuated results in that validation sample may reflect the differences between the studies, model overfitting in BiB or both.

In other work, we have also shown that the distributions of most of the NMR metabolic measures differed by ethnicity¹⁷. White European women had higher levels of most lipoprotein subclasses, cholesterol, glycerides and phospholipids, monosaturated fatty acids, and creatinine but lower levels of glucose, linoleic acid, omega-6 and polyunsaturated fatty acids, and most amino acids, compared with South Asian women. This suggests a more lipidomic response to pregnancy in White Europeans and a stronger glycaemic response in South Asian women. Higher BMI and having gestational diabetes were associated with higher levels of several lipoprotein subclasses, triglycerides, and other metabolites in both groups but with evidence of weaker magnitudes of association for most of these in the South Asian women.

In recent collaborations between the BiB cohort and the Pregnancy Outcome Prediction study (POPs) using Metabolon MS data, we have found evidence that 4'-hydroxyglutamate improves prediction of pre-eclampsia compared to clinical risk factors alone³ and that a ratio of four metabolites (1-(1-enyl-stearoyl)-2-oleoyl-GPC, 1,5-anhydroglucitol, 5 α -androstane-3 α ,17 α -diol disulfate and N1,N12-diacetylspermine) together with the sFlt-1:PIGF ratio is a better predictor of foetal growth restriction/small for gestational age than sFlt-1:PIGF alone⁴. Initial associations in POPs, a nulliparous, largely White European, affluent cohort from the South East of England, were validated in BiB. As we have outlined, BiB is a cohort of mixed ethnic background, with high levels of deprivation and including both nulliparous and multiparous women. The consistency of associations between POPs and BiB suggests that the prediction accuracy may be widely generalisable and that the metabolites predicting these outcomes may be causally related to them.

Furthermore, combining the MS metabolomics with genomic sequence data has enabled the establishment of metabolomic consequences of loss of functional rare variants in autozygous individuals and the health effects of this loss of function¹⁸. This has supported the development of the drug *lumasiran* for a rare kidney disease¹⁹.

Ongoing and future research

Ongoing work using both the NMR and MS metabolomics data will explore how the pregnancy metabolic environment relates to fetal growth (using repeat ultrasound scan measures and birth weight), preterm delivery, and congenital heart disease. Potential causal effects in these studies will be explored where possible by replication, the use of Mendelian randomization (MR) and triangulation with other types of data and study designs. In ongoing work, we are using data from both MS datasets to evaluate whether MS-derived metabolomics are better predictors of gestational diabetes, hypertensive disorders of pregnancy, small and large for gestational age and preterm birth, than risk factors alone (with external validation being undertaken in the POPs cohort). By combining both NMR and MS data, we are exploring the relationships between maternal pregnancy metabolites and their offspring cord blood metabolites.

BiB also contributes to metabolomic studies that are being undertaken by large collaborative efforts. This includes the European H2020 funded LifeCycle project²⁰, in which we are exploring exposure to maternal hypertensive disorder of pregnancy, gestational diabetes, small and large for gestational age and preterm delivery on offspring subsequent metabolic profile. In the Consortium of Metabolomics Studies (COMETS)² there are ongoing projects including trans-ethnic genome-wide association analyses (GWAS), and exploring effects of BMI, smoking, dietary patterns and hypertension on maternal metabolomic profiles.

Discussion and future directions for metabolomic analyses in BiB

In this data note we have described multiple datasets with NMR and MS metabolomic measures in the BiB cohort. The wealth of metabolomics data available in BiB provides opportunities for addressing a range of research questions. In this section, we discuss the strengths and limitations of the data, together with some of our insights for using these data. We also provide information on plans for future measurements of metabolomics in BiB.

A key strength of these datasets is that they are based within a cohort that has very detailed information on 13,776 pregnancies. This includes detailed socioeconomic, education, cognitive, and mental and physical health data. We have OGTT results and fasting pregnancy blood samples on most (83%) of the mothers, genomics (genome wide and sequence) data and epigenomics data in maternal pregnancy and offspring cord blood. Few studies have pregnancy metabolomics data or OGTT data in numbers of this size. We can look at metabolomics and its role in prediction of adverse pregnancy/perinatal outcomes and health and development in children. BiB has large numbers of South Asian and White European families, residing in a city with high levels of socioeconomic deprivation. The ethnic diversity allows us to try and understand ethnic differences in the developmental origins of disease, for example, why South Asian populations have a higher risk of type two diabetes and coronary heart disease.

Having access to two metabolomics profiles (NMR and MS) is beneficial. The NMR platform mostly consists of lipids and lipoproteins, but also provides quantified fatty acids, amino acids, glycolysis metabolites, ketone bodies and glycoprotein acetyl (an inflammatory marker). It provides considerably more information than clinical chemistry measures that are conventionally measured in cohorts (e.g. glucose, total cholesterol, LDLc, HDLc and triglycerides) and at not much more cost (~£20 per sample). As a result, we have been able to obtain these data on large numbers of women in pregnancy, offspring cord blood and in samples taken in offspring at 12- and 24- months assessments. By contrast, the MS data covers more of the metabolome, including being able to assess energy metabolism (which might be important in pregnancy) and markers of medications such as paracetamol. However, it is more expensive (~£80-£200, depending on how many samples are assayed at a time). By having access to both datasets here, we can have broader coverage of the metabolome²¹. There are potential uses for both platforms – ranging from disease prediction to causal analyses using methods such as MR²². Both platforms have been used in previous GWAS of metabolites²³. As such, BiB could be used to explore whether genetic instruments from GWAS can be related to NMR or MS metabolites in pregnancy.

Access to this unique metabolomic data is a big advantage in BiB. However, we have been unable to validate findings

in external cohorts. The work described above cannot be replicated because we cannot find other independent studies with relevant data¹⁷. We hope that this data note will encourage other studies to collect similar data in pregnancy, offspring cord blood, and in mothers and offspring postnatally throughout their life-course.

There are some additional important limitations of the data to consider. The impact of these limitations will depend on the research question. All the metabolomics datasets were collected on subsamples ranging from 11,480 with maternal pregnancy NMR samples (83% of the eligible 13,776 participants) to 1,000 (7%) with MS cord blood samples. Smaller sample sizes may be statistically inefficient in some analyses and the selection processes (Figure 4 and Figure 6) may result in selection bias in some analyses. It is notable and provides some reassurance that, even for the smaller samples, distributions of most characteristics are similar between participant groups with different types of metabolomics at different time points and the whole cohort (Table 2 and Table 3). As Metabolon MS data have been collected only on White British and Pakistani women it cannot be assumed that analyses with these data would generalize to other ethnic groups. BiB cohort participants were largely recruited at the OGTT (with a small number recruited after that). This was opportunistic as we had no funding for initiating the cohort. After consultation with the community and health care providers, we established that this would be a suitable time to obtain consent, interview pregnant women and collect a fasting blood sample for research. However, it means that we are likely to have missed women who did not attend the OGTT and were not captured later in pregnancy or at delivery, and those who delivered pre-term before they attended their OGTT. We have previously compared BiB participants to non-BiB births occurring between 2007–2011⁸. Summary data from obstetric and delivery records were obtained for 11,761 non-BiB births, which would include some who moved to Bradford shortly before delivery (and would not have been eligible to recruit). The comparison showed a small number of differences. BiB participants were less likely to include younger mothers (age 20–24 years) and had a higher proportion of South Asian and nulliparous mothers. There were differences in gestational age and preterm delivery that reflected recruiting BiB participants relatively late in pregnancy⁸. This selection on gestational age may introduce selection bias in some BiB analyses, including those using the metabolomics data described here.

A limitation is that BiB only has pregnancy metabolomics at a single time point and does not have pre-pregnancy measurements. Previous research suggests metabolites change upon becoming pregnant and then revert to pre-pregnancy levels⁵ and that they change during pregnancy²⁴. Earlier measures would be valuable for prediction of future adverse outcomes to enable earlier antenatal monitoring and intervention.

This data note has focused on metabolomics data that have been quantified by high throughput commercial platforms (Nightingale Health© NMR and Metabolon MS). On a small

subsample of BiB participants (N = 199) NMR urine and serum MS blood metabolites have been quantified at Imperial College, London, as part of the HELIX collaboration. HELIX aims to identify the human exposome in pregnancy and childhood. Metabolite measurements were undertaken alongside similar subsamples from five other cohorts (total N = 1,192). In all six cohorts, samples were from children aged between 6–11 years (BiB participants were mean age 6.6 years). 44 urine metabolites (24 semi-quantified) and 188 serum (56 fully quantified) metabolites were measured. We have not described these metabolomics datasets here as the assays are unique to a small subgroup of BiB participants and any research on these participants is best done together with the other HELIX cohort subgroups on whom the same metabolomic data obtained at the same time and using the same methods is available. Further information about the samples and methods used can be found elsewhere²⁵.

Up until March 2020, we were undertaking a follow-up of BiB parents and offspring, including collecting further blood samples with funding available to complete the NMR analyses on offspring and parent serum/plasma collected at this follow-up. However, that follow-up stopped on the 16th March 2020 when restrictions on normal life due to the COVID-19 pandemic began in the UK. At the time of submitting this paper we do not know when face-to-face data collection will be possible to start again and what the best plans would be for further blood sample collection. At the relevant time we will discuss different potential scenarios for completing that planned follow-up with our scientific advisory groups. Whatever the decision, we should have some participants with serum/plasma NMR measures collected ~8–9 years after birth. We are also planning to measure metabolites on the available maternal pregnancy urine samples. Urine metabolites often provide a more accurate measure of dietary intake and medicine use than serum/plasma measures and would be a valuable addition to the existing datasets described here. Any new data will be made available to the wider research community.

Data availability

Underlying data

Scientists are encouraged to make use of the BiB data, which are available through a system of managed open access.

- Before you contact BiB, please make sure you have read our [Guidance for Collaborators](#). Our BiB executive review proposals on a monthly basis and we will endeavour to respond to your request as soon as possible. You can find out about all of the different datasets which are available [here](#). If you are unsure if we have the data that you need please contact a member of the BiB team (borninbradford@bthft.nhs.uk).
- Once you have formulated your request please complete the 'Expression of Interest' form available [here](#) and email the BiB research team (borninbradford@bthft.nhs.uk).

- If your request is approved, we will ask you to sign a [collaboration agreement](#); if your request involves biological samples, we will ask you to complete a [material transfer agreement](#).

Extended data

Open Science Framework: Metabolomics data in the Born in Bradford cohort. <https://doi.org/10.17605/OSF.IO/YST7N>¹¹.

This project contains the following extended data:

- BiB_MS_Dataset1_PCA_Plot.png (Figure showing principal component analysis of dataset 1)
- MS_Metabolite_Details.xlsx (Lists the names and details of all metabolites assessed by the Metabolon platform in the BiB MS dataset 1a (sheet 1) and BiB MS dataset 2 (sheet 2).
- MS_Quality_Report_Dataset1.pdf (QC report for MS dataset 1 from Metabolon)
- MS_Quality_Report_Dataset2.pdf (QC report for MS dataset 2 from Metabolon)
- NMR_Metabolite_Details.xlsx (Lists the names and units of all metabolic traits assessed by the NMR platform)

- NMR_Quality_Report_Cord.pdf (Summarizing quality observations in the NMR cord blood dataset)
- NMR_Quality_Report_Infant.pdf (Summarizing quality observations in the NMR infant dataset)
- NMR_Quality_Report_Pregnancy.pdf (Summarizing quality observations in the NMR pregnancy dataset)

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (CC-BY 4.0).

Acknowledgements

Our thanks to all the children (and parents), teachers and school staff who were involved in the research, and staff at the Bristol Bioresource Laboratory who manage the BiB biobank, including cataloguing and shipping samples (e.g. to Metabolon) and at the Bristol Metabolomics Research who ran NMR analyses. Thanks also to the Born in Bradford (BiB) Research Assistants for data collection. Born in Bradford is only possible because of the enthusiasm and commitment of the Children and Parents in BiB. We are grateful to all the participants, practitioners and researchers who have made Born in Bradford happen.

References

1. Würtz P, Kangas AJ, Soininen P, et al.: **Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Large-Scale Epidemiology: A Primer on -Omic Technologies.** *Am J Epidemiol.* 2017; **186**(9): 1084–96. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
2. Yu B, Zanetti KA, Temprosa M, et al.: **The Consortium of Metabolomics Studies (COMETS): Metabolomics in 47 Prospective Cohort Studies.** *Am J Epidemiol.* 2019; **188**(6): 991–1012. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
3. Sovio U, McBride N, Wood AM, et al.: **4-Hydroxyglutamate is a novel predictor of pre-eclampsia.** *Int J Epidemiol.* 2020; **49**(1): 301–11. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Sovio U, Goulding N, McBride N, et al.: **A maternal serum metabolite ratio predicts fetal growth restriction at term.** *Nat Med.* 2020; **26**(3): 348–53. [PubMed Abstract](#) | [Publisher Full Text](#)
5. Wang Q, Würtz P, Auro K, et al.: **Metabolic profiling of pregnancy: cross-sectional and longitudinal evidence.** *BMC Med.* 2016; **14**(1): 205. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
6. Santos Ferreira DL, Williams DM, Kangas AJ, et al.: **Association of pre-pregnancy body mass index with offspring metabolic profile: Analyses of 3 European prospective birth cohorts.** *Ma RCW editor. PLoS Med.* 2017; **14**(8): e1002376. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. Würtz P, Wang Q, Niironen M, et al.: **Metabolic signatures of birthweight in 18 288 adolescents and adults.** *Int J Epidemiol.* 2016; **45**(5): 1539–50. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. Wright J, Small N, Raynor P, et al.: **Cohort Profile: The Born in Bradford multi-ethnic family cohort study.** *Int J Epidemiol.* 2013; **42**(4): 978–91. [PubMed Abstract](#) | [Publisher Full Text](#)
9. Pembrey L, Waiblinger D, Griffiths P, et al.: **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.* 2017; **17**(1): 220. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Yu Z, Kastenmüller G, He Y, et al.: **Differences between Human Plasma and Serum Metabolite Profiles.** Oresic M editor. *PLoS One.* 2011; **6**(7): e21230. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
11. Taylor K: **Metabolomics data in the Born in Bradford cohort.** 2020. <http://www.doi.org/10.17605/OSF.IO/YST7N>
12. Farrar D, Santorelli G, Lawlor DA, et al.: **Blood pressure change across pregnancy in white British and Pakistani women: analysis of data from the Born in Bradford cohort.** *Sci Rep.* 2019; **9**(1): 13199. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
13. Cologne J, Preston DL, Imai K, et al.: **Conventional case-cohort design and analysis for studies of interaction.** *Int J Epidemiol.* 2012; **41**(4): 1174–86. [PubMed Abstract](#) | [Publisher Full Text](#)
14. Sharp SJ, Poulaliou M, Thompson SG, et al.: **A Review of Published Analyses of Case-Cohort Studies and Recommendations for Future Reporting.** *PLoS One.* 2014; **9**(6): e101176. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
15. Shah J, Brock GN, Gaskins J: **BayesMetab: treatment of missing values in metabolomic studies using a Bayesian modeling approach.** *BMC Bioinformatics.* 2019; **20**(Suppl 24): 673. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
16. McBride N, White SL, Farrar D, et al.: **Do nuclear magnetic resonance (NMR)-based metabolomics improve the prediction of pregnancy-related disorders?** *medRxiv.* 2020; 2020.06.22.20134650. [Publisher Full Text](#)
17. Taylor K, Santos Ferreira DL, West J, et al.: **Differences in Pregnancy Metabolic Profiles and Their Determinants between White European and South Asian Women: Findings from the Born in Bradford Cohort.** *Metabolites.* 2019; **9**(9): 190. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
18. Clark DW, Okada Y, Moore KHS, et al.: **Associations of autozygosity with a broad range of human phenotypes.** *Nat Commun.* 2019; **10**(1): 4957. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

19. McGregor TL, Hunt KA, Yee E, *et al.*: **Characterising a healthy adult with a rare HAO1 knockout to support a therapeutic strategy for primary hyperoxaluria.** Goate A, Weigel D, Ryten M, Blackburn N, editors. *Elife*. 2020; **9**: e54363.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
20. Jaddoe VWW, Felix JF, Andersen A-MN, *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*. 2020; **35**(7): 709–724.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
21. Dunn WB, Broadhurst DI, Atherton HJ, *et al.*: **Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy.** *Chem Soc Rev*. 2011; **40**(1): 387–426.
[PubMed Abstract](#) | [Publisher Full Text](#)
22. Würtz P, Wang Q, Kangas AJ, *et al.*: **Metabolic Signatures of Adiposity in Young Adults: Mendelian Randomization Analysis and Effects of Weight Change.** Sheehan NA, editor. *PLoS Med*. 2014; **11**(12): e1001765.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. Lotta LA, Pietzner M, Stewart ID, *et al.*: **Cross-platform genetic discovery of small molecule products of metabolism and application to clinical outcomes.** *bioRxiv*. 2020; 2020.02.03.932541.
[Publisher Full Text](#)
24. Mills HL, Patel N, White SL, *et al.*: **The effect of a lifestyle intervention in obese pregnant women on gestational metabolic profiles: findings from the UK Pregnancies Better Eating and Activity Trial (UPBEAT) randomised controlled trial.** *BMC Med*. 2019; **17**(1): 15.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Lau C-HE, Siskos AP, Maitre L, *et al.*: **Determinants of the urinary and serum metabolome in children from six European populations.** *BMC Med*. 2018; **16**(1): 202.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)