**Supplement for Ulla Sovio, Neil Goulding, Nancy McBride, Emma Cook, Francesca Gaccioli, D. Stephen Charnock-Jones, Deborah A. Lawlor, Gordon C. S. Smith. A maternal serum metabolite ratio predicts large for gestational age infants at term: a prospective cohort study.**

**Contents**

**Panel S1.** Analysis plan for external validation of associations between the metabolite ratio and 1) LGA and 2) birth weight z score in the Born in Bradford study

**Figure S1.** Flow diagram of the selection of LGA cases and controls from the POP study cohort (total n=281).

**Figure S2.** Birth weight z score by the log-transformed metabolite ratio z score at 36 wkGA in non-diabetic women from the POP study comparator group who delivered at term (n= 281).

**Figure S3.** Flow diagram of the selection of LGA cases and controls from the BiB study sub-group 1 (total n=909).

**Figure S4.** Flow diagram of the selection of LGA cases and controls from the BiB study sub-group 2 (total n=1457).

**Table S1.** Characteristics of the non-diabetic women from the BiB study sub-group 1 in the analysis of measurements from 24-28 wkGA by LGA status (total n=909).

**Table S2.** Characteristics of the non-diabetic women from the second BiB study sub-group 2 in the analysis of measurements from 24-28 wkGA by LGA status (total n=1457).

**Supplemental References**

**Panel S1.** Analysis plan for external validation of associations between the metabolite ratio and 1) LGA and 2) birth weight z score in the Born in Bradford study

**Outcomes**

**1.** Large for gestational age (LGA) defined as birth weight percentile >90th, applying the 1991 Hadlock formula1 to sex-adjusted weights (see Methods below). Births at any gestational age subsequent to the measurement of the metabolites are included.

**2.** Continuous birth weight z score applying the 1991 Hadlock formula to sex-adjusted weights (see Methods below).

**The main exposure**

Scaled imputed metabolite values (multiples of the median, MoMs) from maternal serum or plasma are used to calculate the following metabolite ratio which was originally developed for fetal growth restriction, as defined in our publication2 from the measurements taken at 24-28 weeks of gestation:

The ratio of two products of metabolites: **(A x B) / (C x D)**, where

**A.** 1-(1-enyl-stearoyl)-2-oleoyl-GPC (P-18:0/18:1), COMP\_ID = 52612

**B.** 1,5-anhydroglucitol (1,5-AG), COMP\_ID = 20675

**C.** 5alpha-androstan-3alpha,17alpha-diol disulfate, COMP\_ID = 37182

**D.** N1,N12-diacetylspermine, COMP\_ID = 52987

We will accept validation at *P*<0.025 (Bonferroni corrected for two outcomes, *i.e.* *P*<0.05/2), one sided, given known directionality of association being tested.An association observed in the opposite direction from the POPs study will be disregarded, given the use of a one sided test.

**Secondary exposures**

1. The product of metabolites **C** and **D** above, i.e. the denominator of main exposure.
2. The product of metabolites **A** and **B** above, i.e. the numerator of main exposure.
3. All of the other individual metabolites **A**, **B**, **C** and **D** as sole predictors.

We will accept validation at *P*<0.0042 (one sided, Bonferroni corrected for 6 comparisons with two outcomes *i.e.* *P*<0.05/12).

**Exclusions**

Exclude women who had pre-existing diabetes and women who developed gestational diabetes from the analysis.

**Transformation of the exposures**

Log-transform the exposures, e.g. if the exposure is named ratio,

log10ratio = log10(ratio).

*Separately for Born in Bradford (BiB) study sub-samples 1 and 2:*

Turn the log-transformed exposures into z scores in each sub-sample, e.g.

log10ratioZ = [log10ratio - mean(log10ratio)] / SD(log10ratio).

In the POPs, the mean and SD of log-transformed exposures for calculating z scores were obtained from the comparator group which is representative of the whole POPs cohort. Use the population mean and SD where a cohort design is used (BiB 1). Use the random sub-cohort mean and SD where a case-cohort study design is used (BiB 2), and exclude all participants outside the random sub-cohort from the BiB 2 analysis.

**Statistical analysis**

*Individual analysis of BiB 1 and BiB 2*

Fit a logistic regression model for the transformed exposure and LGA (outcome 1) in BiB 1 and BiB 2 separately, and report the odds ratio, 95%CI and *P* value (1-sided) from the analysis. Perform a ROC curve analysis and calculate AUC (95% CI) for the non-transformed exposure.

Fit a linear regression model separately for each transformed exposure and birth weight z score (outcome 2) in BiB 1 and BiB 2 separately, and report the regression estimate (beta), 95% CI and *P* value (1-sided) from each analysis.

*Pooled analysis of BiB 1 and BiB 2*

The metabolite MoMs from BiB 1 and BiB 2 cannot be pooled directly due to different normalisation. To account for differences in the two samples, calculate the pooled statistics (odds ratio and AUC or beta with 95% CI and *P* value) from the pooled data of sample-specific transformed exposures, i.e. the z scores of the log-transformed MoMs.

**Methods for calculating gestational age and fetal sex adjusted birth weight percentiles**

There were 13524 participants in the BiB dataset with information on birthweight, fetal sex and gestation length (both in weeks and days and in completed weeks). We adjusted each of these birthweights for fetal sex, applied the Hadlock 1991 formulas to these sex-adjusted weights and defined LGA as follows:

1. Participants were grouped by gestation length (in completed weeks). To get adequate numbers of participants in each group (>50), we combined the weeks 24-28, 29-31, 32-33 and 42-44. All other weeks (34-41) were analysed independently, so that there were 12 groups altogether.
2. Within each group *i* we calculated the mean birthweights for both males (*m*[*i*]) and females (*f*[*i*]), and the difference (*d*[*i*]) in the means for each group (*d*[*i*]= *m*[*i*]- *f*[*i*]).
3. We adjusted the birthweights within each group as follows:
where *m*\*[*i*] and *f*\*[*i*] are the sex-adjusted birthweights.
4. We applied the 1991 Hadlock formulas to each of the 13524 participants (using gestation length as weeks and days (in decimal form)). We then calculated z scores for each participant using the sex-adjusted birthweights defined above.
5. We defined LGA in the BiB dataset as a z score (defined above) >90th percentile.

A similar method in the POP study cohort (n=4212) was applied to obtain gestational age and fetal sex adjusted birth weight percentiles and to define LGA. This definition was used when the results from the POP and BiB study cohorts were presented together.



**Figure S1.** Flow diagram of the selection of LGA cases and controls from the POP study cohort (total n=281). BW z scores and percentiles were calculated using the UK 1990 population reference and LGA was defined as BW>90th percentile. POP, Pregnancy Outcome Prediction; wkGA, weeks of gestational age; LGA, large for gestational age; BW, birth weight.

****

**Figure S2.** Birth weight z score by the log-transformed metabolite ratio z score at 36 wkGA in non-diabetic women from the POP study comparator group who delivered at term (n= 281). The red line describes the linear association (Y = –0.14 – 0.29 X) and the black lines represent the lower and upper bounds of the 95% confidence intervals. Population based birth weight z scores using the UK 1990 reference were analysed. wkGA, weeks of gestational age.



**Figure S3.** Flow diagram of the selection of LGA cases and controls from the BiB study sub-group 1 (total n=909). BW z scores and percentiles were calculated applying the 1991 Hadlock formula to sex-adjusted birth weights. LGA was defined as BW>90th percentile. The selection of participants for this sub-group of mass spec metabolomics analyses aimed to get a random sample of 500 White British and Pakistani participants with valid oral glucose tolerance data and genome wide data on mothers and offspring so that they could be used for a wide variety of studies including the present study. The exclusion of those without birth weight and with gestational diabetes are specific to the study included in this paper. All women with pre-existing diabetes are automatically excluded as they do not attend the oral glucose tolerance test clinic. BiB, Born in Bradford; wkGA, weeks of gestational age; LGA, large for gestational age; BW, birth weight.



**Figure S4.** Flow diagram of the selection of LGA cases and controls from the BiB study sub-group 2 (total n=1457). BW z scores and percentiles were calculated applying the 1991 Hadlock formula to sex-adjusted birth weights. LGA was defined as BW>90th percentile. The selection of participants for this sub-group of mass spec metabolomics analyses aimed to get a random sample of mothers and offspring so that they could be used for a wide variety of studies including the present study. The exclusion of cases outside the sub-cohort, without birth weight and with gestational diabetes are specific to the study included in this paper. All women with pre-existing diabetes are automatically excluded as they do not attend the oral glucose tolerance test clinic. BiB, Born in Bradford; wkGA, weeks of gestational age; LGA, large for gestational age; BW, birth weight.

**Table S1.** Characteristics of the non-diabetic women from the BiB study sub-group 1 in the analysis of measurements from 24-28 wkGA by LGA status (total n=909).

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **Characteristic** | **LGA****(n=34)** | **Controls****(n=875)** |
|  |  |  |  |
| **Maternal characteristics** |  |  |
|  |  |  |
| Age, years | 29 (25 to 33) | 27 (23 to 31) |
|  |  |  |  |
| Height, cm | 166 (161 to 168) | 162 (158 to 166) |
|  | Missing | 0 (0%) | 12 (1%) |
|  |  |  |  |
| Body mass index, kg/m2 | 30.05 (26.19 to 35.49) | 25.16 (22.08 to 29.37) |
|  | Missing | 2 (6%) | 29 (3%) |
|  |  |  |  |
| Smoking | 5 (15%) | 164 (19%) |
|  | Missing | 0 (0%) | 1 (<1%) |
|  |  |  |  |
| Alcohol | 24 (71%) | 299 (34%) |
|  | Missing | 0 (0%) | 1 (<1%) |
| Education higher than A-level | 9 (26%) | 190 (22%) |
|  |  |  |  |
| Deprivation, score | 29.68 (18.23 to 44.06)  | 44.82 (26.09 to 55.65) |
| Deprivation, rank | 8369 (3457 to 15428) | 3265(1310 to 10097) |
| Deprivation rank quintile |  |  |
|  | 1 (most deprived) | 6 (18%) | 333 (38%) |
|  | 2 | 6 (18%) | 222 (25%) |
|  | 3 | 8 (24%) | 165 (19%) |
|  | 4  | 13 (38%) | 104 (12%) |
|  | 5 (least deprived) | 1 (2%) | 41 (5%) |
|  | Missing | 0 (0%) | 10 (1%) |
|  |  |  |
| Parity |  |  |
|  | 1st pregnancy | 3 (9%) |  331 (38%) |
|  | Previous pregnanciesMissing | 30 (88%)1 (3%) |  515 (59%)29 (3%) |
|  |  |  |
| White ethnicity | 31 (91%) | 444 (51%) |
|  |  |  |  |
| Married Missing | 26 (76%)0 (0%) | 556 (64%)2 (<1%) |
|  |  |  |  |
| **Birth outcomes** |  |  |
|  |  |  |
| Birth weight, g  | 4265 (4060 to 4480) | 3290 (2987 to 3589) |
| Birth weight z score, Hadlock 1991 | 1.52 (1.37 to 2.08) | -0.63 (-1.21 to -0.02)  |
| Birth weight percentile, Hadlock 1991 | 94 (92 to 98) | 26 (11 to 49) |
|  |  |  |  |
| Gestational age, weeks | 39 (38 to 40) | 40 (39 to 40) |
|  |  |  |  |
| Female fetal sex | 13 (38%) | 433 (50%) |
|  |  |  |  |
| Type and mode of delivery |  |  |
|  | Spontaneous (vaginal) | 16 (47%) | 619 (71%) |
|  | Caesarean (elective or emergency) | 13 (38%) | 82 (9%) |
|  | Induction of labour (medical or surgical) | 5 (15%) | 172 (20%) |
|  | Missing  | 0 (0%) | 2 (<1%) |

The flow diagram (**Figure S3**) describes the selection of LGA cases and controls. Data are expressed as median (interquartile range) or n (%) as appropriate. Data are complete unless ”missing” is specified. Only two ethnic groups were included in this analysis: White British and Pakistani. Alcohol measures were classified as % of women who drank alcohol during pregnancy or up to 3 months before. Smoking measures were classified as % who had smoked at all during pregnancy. Married is inclusive of those who are remarried. Maternal characteristics were taken at recruitment (24-28 weeks) or from examination of the clinical case record, or linkage to the hospital’s electronic databases. The weight measurement used in the calculation of body mass index was the mother’s first antenatal clinic (booking) weight. Birth weight z scores and percentiles were calculated applying the 1991 Hadlock formula to sex-adjusted birth weights and LGA was defined as birth weight >90th percentile (**Panel S1**). Socio-economic position was quantified using the Index of Multiple Deprivation 2010 for the local area where each woman lived at the time of recruitment.3 Deprivation score is the combined sum of the weighted, exponentially transformed domain rank of the domain score, and higher values indicate more deprivation. Conversely, the most deprived area has the lowest rank and the least deprived area has the highest rank. A national reference distribution from 2010 was used to analyse the rank in quintiles (1=most deprived, 5=least deprived). LGA, large for gestational age; wkGA, weeks of gestational age.

**Table S2.** Characteristics of the non-diabetic women from the second BiB study sub-group 2 in the analysis of measurements from 24-28 wkGA by LGA status (total n=1457).

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **Characteristic** | **LGA****(n=52)** | **Controls****(n=1405)** |
|  |  |  |  |
| **Maternal characteristics** |  |  |
|  |  |  |
| Age, years | 30 (27 to 33) | 27 (23 to 31) |
|  |  |  |  |
| Height, cm | 163 (161 to 168) | 162 (157 to 166) |
|  | Missing | 0 (0%) | 29 (2%) |
|  |  |  |  |
| Body mass index, kg/m2 | 30.62 (26.01 to 34.18) | 25.27 (22.07 to 29.31) |
|  | Missing | 2 (4%) | 64 (5%) |
|  |  |  |  |
| Smoking | 5 (10%) | 281 (20%) |
|  | Missing | 0 (0%) | 3 (<1%) |
|  |  |  |  |
| Alcohol | 19 (37%) | 447 (32%) |
|  | Missing | 0 (0%) | 2 (<1%) |
| Education higher than A- level | 15 (29%) | 285 (20%) |
|  |  |  |  |
| Deprivation, score | 35.00 (23.54 to 46.59)  | 44.65 (27.98 to 54.37) |
| Deprivation, rank | 6218(2857 to 11568) | 3307(1498 to 9133) |
| Deprivation rank quintile |  |  |
|  | 1 (most deprived) | 10 (19%) | 515 (37%) |
|  | 2 | 14 (27%) | 399 (28%) |
|  | 3 | 14 (27%) | 260 (19%) |
|  | 4  | 12 (23%) | 170 (12%) |
|  | 5 (least deprived) | 2 (4%) | 40 (3%) |
|  | Missing | 0 (0%) | 21 (1%) |
|  |  |  |
| Parity |  |  |
|  | 1st pregnancy | 10 (19%) | 519 (37%) |
|  | Previous pregnanciesMissing | 38 (73%)4 (8%) | 854 (61%)32 (2%) |
|  |  |  |
| White ethnicity | 38 (73%) | 649 (46%) |
|  |  |  |  |
| Married | 37 (71%) | 935 (67%) |
|  |  |  |  |
| **Birth outcomes** |  |  |
|  |  |  |
| Birth weight, g  | 4280 (4098 to 4497) | 3251 (2911 to 3570) |
| Birth weight z score, Hadlock 1991 | 1.64 (1.47 to 2.00) | -0.69 (-1.32 to -0.07) |
| Birth weight percentile, Hadlock 1991 | 95 (93 to 98) | 25 (9 to 47) |
|  |  |  |  |
| Gestational age, weeks | 39 (38 to 40) | 40 (39 to 40) |
|  |  |  |  |
| Female fetal sex | 24 (46%) | 670 (48%) |
|  |  |  |  |
| Type and mode of delivery |  |  |
|  | Spontaneous (vaginal) | 29 (56%) | 1027 (73%) |
|  | Caesarean (elective or emergency) | 15 (29%) | 135 (10%) |
|  | Induction of labour (medical or surgical) | 8 (15%) | 242 (17%) |
|  | Missing  | 0 (0%) | 1 (<1%) |

The flow diagram (**Figure S4**) describes the selection of LGA cases and controls. Data are expressed as median (interquartile range) or n (%) as appropriate. Data are complete unless ”missing” is specified. Only two ethnic groups were included in this analysis: White British and Pakistani. Alcohol measures were classified as % of women who drank alcohol during pregnancy or up to 3 months before. Smoking measures were classified as % who had smoked at all during pregnancy. Married is inclusive of those who are remarried. Maternal characteristics were taken at recruitment (24-28 weeks) or from examination of the clinical case record, or linkage to the hospital’s electronic databases. The weight measurement used in the calculation of body mass index was the mother’s first antenatal clinic (booking) weight. Birth weight z scores and percentiles were calculated applying the 1991 Hadlock formula to sex-adjusted birth weights and LGA was defined as birth weight >90th percentile (**Panel S1**). Socio-economic position was quantified using the Index of Multiple Deprivation 2010 for the local area where each woman lived at the time of recruitment.3 Deprivation score is the combined sum of the weighted, exponentially transformed domain rank of the domain score, and higher values indicate more deprivation. Conversely, the most deprived area has the lowest rank and the least deprived area has the highest rank. A national reference distribution from 2010 was used to analyse the rank in quintiles (1=most deprived, 5=least deprived). LGA, large for gestational age; wkGA, weeks of gestational age.

**Supplemental References**

1. Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic weight standard. *Radiology.* 1991;181(1):129-133.

2. Sovio U, Goulding N, McBride N, Cook E, Gaccioli F, Charnock-Jones DS, Lawlor DA, Smith GCS. A maternal serum metabolite ratio predicts fetal growth restriction at term. *Nat Med.* 2020;26(3):348-353.

3. McLennan D, Barnes H, Noble M, Davies J, Garratt E, Dibben C. *The English Indices of Deprivation 2010.* London: Department for Communities and Local Government; 2011.