DATA NOTE

The ‘ALSPAC in London’ dataset: adiposity, cardiometabolic risk profiles, and the emerging arterial phenotype in young adulthood [version 1; peer review: 2 approved with reservations]

Scott T. Chiesa¹, Alicja Rapala¹,², Marietta Charakida¹,³, Kaitlin H. Wade⁴,⁵, Nicholas J. Timpson⁴,⁵, John E. Deanfield¹

¹Vascular Physiology Unit, UCL Institute of Cardiovascular Science, London, UK
²Department of Population Science and Experimental Medicine, UCL Institute of Cardiovascular Science, London, UK
³Division of Imaging Sciences and Biomedical Engineering, King’s College London, London, UK
⁴Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
⁵MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

Abstract

Rising rates of adiposity in the young pose one of the greatest threats to future population burden of cardiovascular disease. Understanding the contribution of genetic and early-life influences to adiposity profiles in young adulthood – when the first signs of subclinical cardiovascular disease commonly appear – are vital if effective lifetime prevention strategies are to be developed. This data note documents the extensive range of genotypic and phenotypic data available from a London-based sub-study of the long-running Avon Longitudinal Study of Parents and Children (ALSPAC)—the ‘ALSPAC in London’ Study—in which extensive adipose and cardiovascular phenotyping was carried out in participants recruited based on a genetic predisposition to obesity.

Keywords

Adiposity, cardiometabolic risk, arterial phenotypes, ALSPAC

This article is included in the Avon Longitudinal Study of Parents and Children (ALSPAC) gateway.
Corresponding author: Scott T. Chiesa (s.chiesa@ucl.ac.uk)

Author roles: Chiesa ST: Conceptualization, Data Curation, Methodology, Project Administration, Writing – Original Draft Preparation; Rapala A: Conceptualization, Data Curation, Methodology, Project Administration, Writing – Review & Editing; Charakida M: Conceptualization, Data Curation, Methodology, Writing – Review & Editing; Wade KH: Conceptualization, Methodology, Writing – Review & Editing; Timpson NJ: Conceptualization, Funding Acquisition, Methodology, Resources, Supervision, Writing – Review & Editing; Deanfield JE: Conceptualization, Funding Acquisition, Resources, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This publication is the work of the authors and STC will serve as guarantor for the contents of this paper. The UK Medical Research Council and Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. This research was specifically funded by the British Heart Foundation (BHF Programme Grant No. RG/10/004/28240). JED was a BHF-funded Professor during the time of this study. NJT is a Wellcome Trust Investigator (202802/Z/16/Z), is the PI of the Avon Longitudinal Study of Parents and Children (MRC & WT 102215/2/13/2), is supported by the University of Bristol NIHR Biomedical Research Centre (BRC-1215-20011), the MRC Integrative Epidemiology Unit (MC_UU_12013/3) and works within the CRUK Integrative Cancer Epidemiology Programme (C18281/A19169). KHW was supported by a Wellcome Trust Grant (096989/Z11/Z). A comprehensive list of grants funding is available on the ALSPAC website.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2018 Chiesa ST 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: Chiesa ST, Rapala A, Charakida M et al. The ‘ALSPAC in London’ dataset: adiposity, cardiometabolic risk profiles, and the emerging arterial phenotype in young adulthood [version 1; peer review: 2 approved with reservations] Wellcome Open Research 2018, 3:162 (https://doi.org/10.12688/wellcomeopenres.14942.1)

First published: 23 Dec 2018, 3:162 (https://doi.org/10.12688/wellcomeopenres.14942.1)
Introduction
The recent worldwide increase in population adiposity threatens to reverse achieved reductions in cardiovascular morbidity and mortality. The fastest rises in adiposity rates have been in children, teenagers and young adults and therefore represent one of the major growing threats to worldwide public health. The Avon Longitudinal Study of Parents and Children (ALSPAC) is a longitudinal birth cohort that recruited pregnant women living near Bristol, UK with an estimated delivery date between 1991 and 1992. The study includes extensive phenotypic, genetic, epigenetic and metabolomic data on the mothers, fathers and children from questionnaires, clinics and samples, and follow up is ongoing. Using a novel ‘recall-by-genotype’ study design to identify individuals with genetically predicted variation in body mass index (BMI)\(^1\), the ‘ALSPAC in London’ study aimed to examine the impact of adiposity on metabolic, inflammatory and autonomic disturbances, as well as their relationship to cardiovascular (CV) structure and function, with genetically-directed deep phenotyping.

Methods
ALSPAC recruited 14,541 pregnant women resident in Avon, UK (former county covering Bristol and the surrounding areas in the South West UK) with expected dates of delivery 1st April 1991 to 31st December 1992\(^2,3\). The initial number of pregnancies for which the mother enrolled in the ALSPAC study and had either returned at least one questionnaire or attended a “Children in Focus” clinic by 19/07/99 is 14,541. Of these initial pregnancies, there were a total of 14,676 fetuses, resulting in 14,062 live births and 13,988 children who were alive at 1 year of age. When the oldest children were approximately 7 years of age, an attempt was made to bolster the initial sample with eligible cases who had failed to join the study originally. As a result, when considering variables collected from the age of seven onwards (and potentially abstracted from obstetric notes) there are data available for more than the 14,541 pregnancies mentioned above.

The number of new pregnancies not in the initial sample (known as Phase I enrolment) that are currently represented in the study and reflecting enrolment status at the age of 18 is 706 (452 and 254 recruited during Phases II and III respectively), resulting in an additional 713 children being enrolled. The phases of enrolment are described in more detail in the cohort profile paper\(^3\).

The total sample size for analyses using any data collected after the age of seven is therefore 15,247 pregnancies, resulting in 15,458 foetuses. Of this total sample of 15,458 foetuses, 14,775 were live births and 14,701 were alive at 1 year of age\(^4\). The data included in this resource were collected from 436 young adults (mean age 21 ± 1 years) recalled for extensive adipose and cardiometabolic phenotyping between 2011 and 2015. Of these, 419 were recruited based on a genetic propensity for high or low BMI, with the other 17 comprising patients recruited prior to the decision to employ a recall-by-genotype study design.

Dataset 1 – genetic risk groups
Original beta-coefficients for the genome-wide genetic risk score (GRS) assembly were obtained directly from the genome-wide association study (GWAS) of BMI, conducted in 2010 by Speliotes et al.\(^5\), with the ALSPAC data removed from the release. All alleles were aligned such that the reference allele matched the BMI-increasing direction reported in the GWAS and plink (--score) was used to derive a score for each participant (N=8,350). After removing heterozygous haploid genotypes (N=46,067), direct genotypes were available for 500,527 SNPs, of which 472,208 mapped to SNPs that overlapped with the Speliotes et al.\(^5\) GWAS results and 470,667 mapped to alleles. The ALSPAC genetic dataset used was the best available data at the time of the RbG study design (2012) and was subsequently used in several papers\(^6,7\). The number of participants invited and recruited to the RbG study was orchestrated based on an a priori power calculation taking into account the predicted explained variation in BMI by the genome-wide GRS and its effect size on an example cardiovascular phenotype (SBP). Specifically, using the observed association between SBP and BMI at age 17, we calculated that we would require 450 participants in our RbG to be able to detect a difference of 3 mmHg (a difference that in young adulthood has clinically meaningful association with future cardiovascular disease (CVD) with 80% power and a two-sided p-value threshold of 0.05 and assuming an approximate 3.3kg/m\(^2\) unit difference in BMI between recalled groups. Participants were then recruited according to their appearance in nine sampling groups, from the most extreme to the least, to maximise power and difference in BMI. These samples were as follows: 3%, 6%, 9%, 12%, 15%, 18%, 21%, 24%, 27% from the lower tail of the genome-wide GRS distribution and 97%, 94%, 91%, 88%, 85%, 82%, 79%, 76%, 73% from the upper tail of the genome-wide GRS distribution. The GRS was computed for 4,602 individuals in total within these sampling groups. Of those who were invited to the study (N=2,071), 419 individuals attended from across the entire distribution of sampling groups and had both full genetic and BMI data.

Dataset 2 – participant characteristics
Height was measured in metres using a stadiometer (SECA 213, Birmingham, UK) and weight in kilograms using electronic weighing scales (Marsden M-110, Rotherham, UK). BMI was calculated as weight (kg) / height (m\(^2\)). Waist circumference was measured in centimetres at the narrowest circumference of the abdomen (roughly level to the umbilicus) using a standard flexible tape measure. Hip circumference was measured in centimetres at the widest circumference of the hips buttoks using a standard flexible tape measure. Sagittal abdominal diameter was measured in centimetres as the distance between the small of the back and the top of the abdomen, with all measurements taken in the supine position with specialist metal callipers. Numerous other measures such as medical history, recent
illnesses, smoking, alcohol consumption, physical activity and menstrual cycle were recorded via questionnaire, a copy of which is available as extended data.

Dataset 3 – adipose phenotypes

Body fat volume and distribution was measured using a 1.5T magnetic resonance imaging (MRI) system (Avanto; Siemens Medical Solutions). A graphics processing unit (GPU) implementation of the T2-IDEAL algorithm was used to measure body fat content. This technique iteratively separated MR images into fat and water components, which could then be used to measure the proportion of fat in each 36x36x10 mm voxel. Data was acquired in a continuous stack of 10 mm thick slices from the neck to the knees. To prevent motion artefact, breath holding was used for the thorax and abdomen and cardiac gating for slices containing the heart. Fat quantification in the head, arms and below the knees was impractical due to the need for participant re-positioning or specialised coils. Due to their low fat content

---

**Figure 1.** Study flow diagram.
these body parts were excluded using anatomical landmarks to ensure consistency between participants. The visceral compartment was manually separated from the fat image and the liver was excluded due to frequent artefacts at diaphragm level. Absolute volumes of fat were calculated for subcutaneous, visceral, liver, and pericardial deposits.

Dataset 4 – cardiovascular phenotypes

**Carotid intima-media thickness.** Common carotid artery intima-media thickness (cIMT) was measured using B-mode ultrasound images acquired in the ear-to-ear plane with the head rotated to 45° from the midpoint using a Zonare Z.OneUltra system equipped with a 1L0-5 linear transducer (Zonare Medical Systems, CA, US). Images were recorded in Digital 2 Imaging and Communications in Medicine (DiCOM) format as 10 second cine-loop files for offline analysis using Carotid Analyzer software (Medical Imaging Applications, Coralville, IA). Left and right cIMT were taken to be the average of three end-diastolic measurements located on the far-wall of a single segment of arterial wall 5–10 mm in length and 10 mm proximal to the bifurcation.

**Pulse wave velocity.** Pulse-wave velocity (PWV) was used to estimate arterial stiffness via applanation tonometry (SphygmoCor Vx, AtCor Medical, NSW, Australia). Two distinct indices of PWV were measured, using electrocardiograph (ECG)-gated pulse waves travelling between both carotid-radial and carotid-femoral sites. The participant was rested in a supine position and a handheld tonometer was placed over the left carotid artery in order to allow the recording of 10–12 clear and reproducible pressure waveforms. The same tonometer was then used to measure a similar number of femoral arterial pulse waveforms in the inguinal crease at the top of the right leg, or radial arterial pulse waveforms in the right wrist. Carotid-radial transit distance was measured between the upper edge of the suprasternal notch and the radial pulse measurement site using a tape measure, whereas carotid-femoral transit distance was measured between carotid and femoral sites both directly and via the umbilicus. The device software calculated the mean transit time (in milliseconds) from the recorded pulse waveforms and carotid-femoral and carotid-radial PWV were calculated as the transit distance/transit time.

**Pulse wave analysis.** Estimates of central blood pressure (BP) and indices of systemic arterial stiffness such as augmentation pressure (AP) and augmentation index (AIx) were assessed using pulse wave analysis (SphygmoCor Vx, AtCor Medical, NSW, Australia). This technique measures the arterial pressure waveform at the radial artery and applies a validated generalised transfer function to provide the central pressure waveform. Participants lay supine on a couch in ambient conditions for at least 10 mins prior to the start of the assessment. Brachial BP was measured using a digital automated sphygmomanometer (Omron M6 Comfort). With the participant’s wrist held steady, the tonometer probe was placed over the right radial artery at the level of the wrist and a recording of 10–12 pressure waveforms was saved to specialist software. Quality indices (average pulse height, pulse height variation, diastolic variation, and shape deviation) were assessed for each measurement to confirm they fell within acceptable limits (automatically calculated), otherwise the scan was repeated.

**Autonomic function.** Autonomic function was assessed using measures of baroreflex sensitivity (BRS) and heart rate variability (HRV). Participants lay supine on a couch in a darkened room in ambient conditions for at least 10 mins prior to the start of the assessment. Following this, a 3-lead ECG and 4-lead respiratory monitoring system were attached to the participant’s chest to measure heart rate and breathing cycle, while an inflatable infrared photoplethysmographic cuff (Portapres, FMS, Netherlands) was attached to the middle finger of the right hand and (following calibration) inflated in order to continuously and non-invasively measure BP. Participants remained at rest for 20 mins while measurements were recorded by the software, with all signals fed to specialist software for the calculation of BRS and HRV. A stationary period of 5 min with less than 5% atrial/ventricular ectopic beats was chosen for the temporal QT interval variability analysis using a computer algorithm. The examiner defined a template QT interval for one beat from the beginning of the QRS complex to the end of T wave, including all deflections that might relate to repolarization, including possible U waves. The algorithm found the QT intervals of all other beats by determining the stretch or compression in time of each beat that best matched the ST segment and the T wave of the template beat, whereas the QRS complex was ignored. RR interval mean and variance and QT interval mean and variance were derived from the respective time series. QTVI, which represents the log ratio between normalized QT and RR interval variability, was calculated according to the equation: QTVI = log10 ([QTv/QTm]/[RRv/RRm]), whereby a difference of 1 QTVI between two individuals implies a tenfold difference in temporal QT variability normalized to the QT interval, RR variance (HRV calculated in the time domain) and the RR interval. The squared coherence function was calculated from power spectra of the RR and QT interval time series and the cross-spectrum between these two series derived by fast Fourier Transform (Welch algorithm, five blocks, 50% overlap, Hanning window). The mean squared coherence was obtained by averaging the coherence function over the frequency band from 0 to 0.45 Hz. The coherence provides a measurement of the degree of linear interaction between the RR and QT interval fluctuations as a function of the frequency of those fluctuations. For BRS calculations, registrations were recorded at a sampling frequency of 1000 Hz and stored on a computer. The recordings were inspected offline for the removal of artefactual segments and sequences containing non-sinus beats. Ectopic beats were corrected by interpolation. The time series of systolic blood pressure (SBP) and RR interval from the entire period of recording (20 min) were scanned to identify baroreflex sequences, which were defined as three or more consecutive beats in which successive SBP and RR intervals concordantly increased or decreased, with the threshold set at 1.0 mmHg and 5.0 ms, respectively, and a shift of +1 between the BP pulse and the RR interval, resembling the classical criteria suggested by Bertinieri et al. (threshold values of 1.0 mmHg and 4.0 ms, respectively and shift 1). Linear regression was applied to each sequence and only those for
which the square of the correlation coefficient \( r^2 \) was greater than 0.85 were accepted for further analysis. The spontaneous BRS was calculated, reflecting the average regression slope for all the linear regressions. RR data were used to derive standard deviation of normal-to-normal RR intervals (SDNN) for time-domain HRV.

**Ultra-high frequency ultrasound (UHFUS).** In a randomly-selected sub-sample of participants, images of individual intima and media thicknesses (IT and MT, respectively) were obtained in the carotid, radial and dorsal pedal arteries using an ultra-high resolution ultrasound system (UHFUS; Vevo 2100, Visualsonics) with attached 40 MHz and 50 MHz transducers. Right common carotid artery scans were obtained via UHFUS and imaged longitudinally 1–2 cm proximal to the carotid bifurcation. Images were focused on the posterior (far) wall of the artery and the zoom function was used to magnify the area. Right radial artery scans were collected in the same way at a location 1–2 cm proximal to the skin fold separating the palma manus from the region antebrachii anterior. Right dorsal pedal scans were measured above the proximal first metatarsal bone in the foot. Images were recorded in DICOM format as cine loops for off-line analysis. IT echo was assessed offline and its total thickness measured with callipers within a standardised (8mm×8 mm) higher resolution zoom. The image was acquired, temporarily stored in the cine loop and consecutively zoomed when the measurements were done. The measurements of the IT were performed in systole of the vessel (determined by scrolling through the cine loop to reach the arteries largest diameter). The MT was then calculated as the difference between IMT and IT (MT = IMT−IT). IMT was defined as the distance from the leading edge of the lumen–intima interface to the leading edge of the media–adventitia interface. Lumen diameter was defined by the distance between the leading edges of the intima–lumen interface of the near wall and the lumen–intima interface of the far wall.

**Cardiovascular magnetic resonance imaging.** Cardiovascular MRI was used to assess numerous measures of cardiovascular structure and function. All measures were made using a 1.5T MR scanner (Avanto, Siemens Medical Solutions, Erlangen, Germany). Endocardial borders of the left ventricle (LV) were traced manually on short axis stacks at end-diastole and end-systole to evaluate end-diastolic volume (EDV) and end-systolic volume (ESV). Stroke volume (SV) was obtained by subtracting ESV from EDV. Epicardial borders were traced in end-diastole to calculate an epicardial volume. The EDV was subtracted from this volume, multiplied by assumed myocardial density to obtain left ventricular mass (LVM). Flow quantification was performed through-plane in a cross-section of the ascending aorta as it passes the bifurcation of the pulmonary arteries using an ECG-gated spiral phase-contrast MRI sequence. This technique allows images to be acquired within a short breath-hold (0.5 seconds) with a spatial resolution of 1.6 x 1.6 mm and a temporal resolution of 30 ms. All images were processed using in-house plug-ins for the Open source software OsiriX (OsiriX Foundation, Geneva, Switzerland). Flow images were manually segmented (using the modulus images) and SV (ml) was measured and cardiac output (CO, L/min) calculated as SV x heart rate. At the time of flow imaging, BP was simultaneously measured using MRI-compatible oscillometric sphygmomanometer (Datex Ohmeda). Systemic vascular resistance (SVR; measured in mmHg/L/min) was calculated by dividing the measured mean BP by CO. Total arterial compliance (TAC) was calculated by optimisation of the two-element Windkessel model. Briefly, the flow curves and SVR were used as inputs to the model. Pulse pressure (PP) was calculated for a series of modelled pressure curves generated using a range of TAC values from 0.1 to 5.0 mL/mmHg in increments of 0.01. The compliance value that gave the smallest error between the modelled PP and the true PP was taken to be the true compliance.

**Dataset 5 – metabolic/metabolomic phenotypes**

In order to assess cardiometabolic risk profiles, 15 ml of blood was collected at least 4–6 h post-meal using standard venepuncture techniques and divided between both serum and plasma (EDTA) vacutainers. Within 30 minutes of collection, samples were centrifuged for 12 mins at 4000 rpm, pipetted into 1 ml Eppendorf tubes and immediately frozen at –80°C. Plasma lipids (total cholesterol, triglycerides, HDL cholesterol, and direct LDL cholesterol), liver function tests (ALT, AST, and GGT), Apolipoprotein A1 and B, C-reactive protein, and glucose were measured using an automated analyser (c311, Roche Diagnostics, Burgess Hill, UK). Insulin was also measured using an automated immunoassay analyser (e411, Roche Diagnostics, Burgess Hill, UK), while adiponectin, leptin and IL-6 (high sensitivity) were measured by ELISA (R&D systems, Abingdon, UK). All measures were made using the manufacturer’s calibrators and quality control materials. Serum paraoxonase (PON-1) activities were measured by UV spectrophotometry in a 96-well plate format using paraoxon (SigmaAldrich, St Louis, Missouri). Metabolomic profiling was also performed on EDTA plasma samples by Nightingale Health (Finland) using a quantitative high-throughput NMR metabolomics platform, full details of which have been published elsewhere10.

**Dataset 6 – psychological questionnaires**

Three questionnaire-based psychological assessments were provided to participants on the day of testing in order to gauge indices of interpersonal support, happiness, and depression. The Interpersonal Support Evaluation List is a 12-item scale made up of three subscales – namely Tangible Support, Belonging Support, and Appraisal Support. All answers are given on a 4-point scale ranging from ‘Definitely True’ to ‘Definitely False’ and participants are required to rate each statement depending on its relevance to them. For analysis, the scores for negatively phrased questions are reversed and the scale totals for each subscale and the overall total are calculated10. The Oxford Happiness Questionnaire (OHQ) is a 29-item questionnaire; with each statement answered using a uniform 6-point Likert Scale. For analysis, negatively phrased questions are reversed and the sum total is divided by the number of questions to give an average score11. The Beck Depression Inventory II is a 21-question multiple choice self-report inventory used to assess the severity of depression. Questions are scored from 0–3, with overall scores of 0–13 suggesting minimal depression,
14–19 mild depression, 20–28 moderate depression, and 29–63 severe depression\(^2\).

**Ethical approval and consent**
Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Full details of the approvals obtained are available from the study website (http://www.bristol.ac.uk/alspac/researchers/research-ethics/). Written informed consent was obtained from participants prior to their clinic visit. Study members have the right to withdraw their consent for elements of the study or from the study entirely at any time.

**Data**
For ease of navigation, results include only variables that are deemed to be of particular relevance to this study and have been separated into six distinct datasets contained within tables, namely **Table 1**, genetic risk groups; **Table 2**, participant characteristics; **Table 3**, adipose phenotypes; **Table 4**, cardiovascular phenotypes; **Table 5**, metabolic/metabolomic phenotypes; and **Table 6**, psychological questionnaires. A selection of example graphs showing relationships between BMI and a number of these adipose, vascular, cardiac, and metabolic phenotypes can be found in the extended data\(^8\).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Database ID</th>
<th>n</th>
<th>Mean (SD) or %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in years</td>
<td>Age</td>
<td>436</td>
<td>21.0 (1.0)</td>
</tr>
<tr>
<td>Participant sex (M/F)</td>
<td>Gender</td>
<td>436</td>
<td>34/66</td>
</tr>
<tr>
<td>Height in m</td>
<td>Ht</td>
<td>436</td>
<td>1.71 (0.09)</td>
</tr>
<tr>
<td>Weight in kg</td>
<td>Wt</td>
<td>436</td>
<td>72.2 (19.0)</td>
</tr>
<tr>
<td>Body mass index in kg/m(^2)</td>
<td>Bmi</td>
<td>436</td>
<td>24.6 (5.7)</td>
</tr>
<tr>
<td>Waist circumference measured in cm</td>
<td>Waistcircm</td>
<td>435</td>
<td>82.5 (13.7)</td>
</tr>
<tr>
<td>Hip circumference measured in cm</td>
<td>Hipcircm</td>
<td>435</td>
<td>101.8 (11.5)</td>
</tr>
<tr>
<td>Sagittal-abdominal diameter measured in cm</td>
<td>Sad</td>
<td>435</td>
<td>15.9 (48.9)</td>
</tr>
</tbody>
</table>

**Table 1. Genetic risk groups dataset.**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Database ID</th>
<th>n</th>
<th>Mean (SD) or %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic Risk Group (H/L)</td>
<td>Genetic Group</td>
<td>419</td>
<td>54/46</td>
</tr>
</tbody>
</table>

**Table 2. Participant characteristics dataset.**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Database ID</th>
<th>n</th>
<th>Mean (SD) or %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum of visceral fat</td>
<td>Sumvf</td>
<td>409</td>
<td>20669 (20678)</td>
</tr>
<tr>
<td>Sum of subcutaneous fat</td>
<td>Sumsq</td>
<td>409</td>
<td>177688 (95361)</td>
</tr>
<tr>
<td>Sum of body volume</td>
<td>Sumbv</td>
<td>409</td>
<td>465622 (134155)</td>
</tr>
<tr>
<td>Number of voxels visceral fat</td>
<td>Vfvoxel</td>
<td>409</td>
<td>2091 (2525)</td>
</tr>
<tr>
<td>Number of voxels subcutaneous fat</td>
<td>Sqvoxel</td>
<td>411</td>
<td>17260 (9369)</td>
</tr>
<tr>
<td>Number of voxels total fat</td>
<td>Tfvoxel</td>
<td>409</td>
<td>19362 (10778)</td>
</tr>
<tr>
<td>Number of voxels body volume</td>
<td>Bvvoxel</td>
<td>409</td>
<td>43526 (15184)</td>
</tr>
<tr>
<td>Total body fat indexed to body volume</td>
<td>Tfibv</td>
<td>409</td>
<td>0.384 (0.154)</td>
</tr>
<tr>
<td>Visceral fat indexed to body volume</td>
<td>Vfibv</td>
<td>409</td>
<td>0.043 (0.036)</td>
</tr>
<tr>
<td>Subcutaneous fat indexed to body volume</td>
<td>Sqibv</td>
<td>409</td>
<td>0.415 (0.198)</td>
</tr>
<tr>
<td>Total fat indexed to body surface area</td>
<td>Tfibsa</td>
<td>409</td>
<td>10357 (4769)</td>
</tr>
<tr>
<td>Visceral fat indexed to body surface area</td>
<td>Vfibsa</td>
<td>409</td>
<td>1089 (1147)</td>
</tr>
<tr>
<td>Subcutaneous fat indexed to body surface area</td>
<td>Sqibsa</td>
<td>409</td>
<td>9310 (4189)</td>
</tr>
<tr>
<td>Subcutaneous fat indexed to visceral ratio</td>
<td>SqrV</td>
<td>409</td>
<td>12.2 (7.9)</td>
</tr>
<tr>
<td>Mean absolute liver fat</td>
<td>Livermean</td>
<td>412</td>
<td>184 (97)</td>
</tr>
<tr>
<td>Mean % liver fat</td>
<td>Liverfatperc</td>
<td>412</td>
<td>4.5 (2.4)</td>
</tr>
<tr>
<td>Pericardial fat volume</td>
<td>Pericardialvol</td>
<td>412</td>
<td>470 (411)</td>
</tr>
</tbody>
</table>
Table 4. Cardiovascular phenotypes dataset.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Database ID</th>
<th>n</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Carotid Intima-Media Thickness</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean right carotid artery diameter measured during diastole in mm</td>
<td>Rbldc123</td>
<td>434</td>
<td>6.21 (0.47)</td>
</tr>
<tr>
<td>Mean right carotid artery diameter measured during peak systole in mm</td>
<td>Rpkdc123</td>
<td>434</td>
<td>7.06 (0.53)</td>
</tr>
<tr>
<td>Mean right carotid artery far-wall IMT measured during diastole in mm</td>
<td>Rblimt123</td>
<td>435</td>
<td>0.46 (0.05)</td>
</tr>
<tr>
<td>Mean right carotid artery far-wall IMT measured during peak-systole in mm</td>
<td>Rpkimt123</td>
<td>433</td>
<td>0.44 (0.05)</td>
</tr>
<tr>
<td>Mean left carotid artery diameter measured during diastole in mm</td>
<td>Lbldc123</td>
<td>435</td>
<td>6.25 (0.48)</td>
</tr>
<tr>
<td>Mean left carotid artery diameter measured during peak systole in mm</td>
<td>Lpkdc123</td>
<td>435</td>
<td>7.04 (0.51)</td>
</tr>
<tr>
<td>Mean left carotid artery far-wall IMT measured during diastole in mm</td>
<td>Lblimt123</td>
<td>435</td>
<td>0.46 (0.05)</td>
</tr>
<tr>
<td>Mean left carotid artery far-wall IMT measured during peak-systole in mm</td>
<td>Lpkimt123</td>
<td>432</td>
<td>0.45 (0.05)</td>
</tr>
<tr>
<td><strong>Pulse Wave Velocity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carotid-Radial PWV in m/s</td>
<td>Pwvrad</td>
<td>434</td>
<td>6.79 (0.93)</td>
</tr>
<tr>
<td>Carotid-Femoral PWV when distance measured via umbilicus in m/s</td>
<td>Pwvfemviaub</td>
<td>419</td>
<td>5.59 (0.74)</td>
</tr>
<tr>
<td>Carotid-Femoral PWV when distance measured direct in m/s</td>
<td>Pwvfemdirect</td>
<td>418</td>
<td>5.37 (0.70)</td>
</tr>
<tr>
<td><strong>Flow-Mediated Dilation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline brachial artery diameter in mm prior to cuff inflation</td>
<td>Bldfmd</td>
<td>420</td>
<td>3.22 (0.52)</td>
</tr>
<tr>
<td>Peak percentage increase in arterial diameter following cuff deflation</td>
<td>Pkfmdperc</td>
<td>418</td>
<td>8.2 (4.0)</td>
</tr>
<tr>
<td>Time in seconds (from start of baseline) to reach peak percentage increase in arterial diameter following cuff deflation</td>
<td>Pkfmdtime</td>
<td>415</td>
<td>411 (17)</td>
</tr>
<tr>
<td>Mean baseline velocity-time integral in centimetres calculated from measurements 50, 30 and 10 seconds prior to cuff inflation</td>
<td>Bvti123</td>
<td>427</td>
<td>0.07 (0.03)</td>
</tr>
<tr>
<td>Peak velocity-time integral in centimetres measured following cuff deflation</td>
<td>Pkvti01</td>
<td>411</td>
<td>0.39 (0.11)</td>
</tr>
<tr>
<td><strong>Peripheral Blood Pressure</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systolic blood pressure measured immediately prior to PWA using brachial cuff occlusion in mmHg</td>
<td>Sbppwa</td>
<td>431</td>
<td>117 (10)</td>
</tr>
<tr>
<td>Diastolic blood pressure measured immediately prior to PWA using brachial cuff occlusion in mmHg</td>
<td>Dbppwa</td>
<td>431</td>
<td>67 (7)</td>
</tr>
<tr>
<td>Pulse pressure calculated immediately prior to PWA as systolic blood pressure – diastolic blood pressure</td>
<td>Pppwa</td>
<td>431</td>
<td>50 (10)</td>
</tr>
<tr>
<td>Mean arterial pressure calculated immediately prior to PWA as diastolic blood pressure + 1/3 (systolic blood pressure – diastolic blood pressure)</td>
<td>Mappwa</td>
<td>431</td>
<td>84 (7)</td>
</tr>
<tr>
<td><strong>Pulse Wave Analysis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Augmentation Pressure – extent to which central systolic pressure is augmented by the returning reflected pressure wave in mmHg</td>
<td>Augment</td>
<td>430</td>
<td>-0.51 (3.49)</td>
</tr>
<tr>
<td>Augmentation index (expressed as a percentage) calculated by dividing augmentation pressure by pulse pressure</td>
<td>Aiappp</td>
<td>430</td>
<td>-1.88 (10.38)</td>
</tr>
<tr>
<td>Augmentation index (expressed as a percentage) calculated by dividing pressure of the second peak observed within the central pressure waveform by the pressure of the first peak</td>
<td>Aip2p1</td>
<td>430</td>
<td>98.34 (11.14)</td>
</tr>
<tr>
<td>Variable</td>
<td>Database ID</td>
<td>n</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>-------------</td>
<td>----</td>
<td>------------</td>
</tr>
<tr>
<td>Aortic systolic blood pressure estimated by PWA in mmHg</td>
<td>Pwaasp</td>
<td>430</td>
<td>98 (8)</td>
</tr>
<tr>
<td>Aortic diastolic blood pressure estimated by PWA in mmHg</td>
<td>Pwaadp</td>
<td>430</td>
<td>68 (7)</td>
</tr>
<tr>
<td>Aortic pulse pressure estimated by PWA in mmHg</td>
<td>Pwaapp</td>
<td>430</td>
<td>30 (6)</td>
</tr>
<tr>
<td><strong>Autonomic Function</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRS slope</td>
<td>Brsslope</td>
<td>365</td>
<td>20.3 (9.6)</td>
</tr>
<tr>
<td>Time Domain Heart rate variability, SDNN</td>
<td>Brstdhrv</td>
<td>357</td>
<td>73.3 (32.7)</td>
</tr>
<tr>
<td>Time Domain Heart rate variability, RMSSD</td>
<td>Brstdhrv2</td>
<td>357</td>
<td>68.9 (45.8)</td>
</tr>
<tr>
<td>Time Domain Heart rate variability, Triangular Index</td>
<td>Brstdhrv3</td>
<td>357</td>
<td>17.4 (6.1)</td>
</tr>
<tr>
<td><strong>Ultra-High Frequency Ultrasound</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 MHz right radial artery diameter measured in mm</td>
<td>vhrsrrad50mhz</td>
<td>175</td>
<td>1.86 (0.39)</td>
</tr>
<tr>
<td>50 MHz right radial artery intima-media thickness measured in mm</td>
<td>vhrssraimt50mhz</td>
<td>175</td>
<td>0.15 (0.03)</td>
</tr>
<tr>
<td>50 MHz right radial artery intima thickness measured in mm</td>
<td>vhrsraint50mhz</td>
<td>175</td>
<td>0.07 (0.01)</td>
</tr>
<tr>
<td>50 MHz right radial artery media thickness measured in mm</td>
<td>vhrsramt50mhz</td>
<td>175</td>
<td>0.09 (0.03)</td>
</tr>
<tr>
<td>50 MHz right dorsal paedal artery diameter measured in mm</td>
<td>vhrsrdpad50mhz</td>
<td>145</td>
<td>1.31 (0.44)</td>
</tr>
<tr>
<td>50 MHz right dorsal paedal artery intima-media thickness measured in mm</td>
<td>vhrsrdpaimt50mhz</td>
<td>145</td>
<td>0.17 (0.11)</td>
</tr>
<tr>
<td>50 MHz right dorsal paedal artery intima thickness measured in mm</td>
<td>vhrsrdpait50mhz</td>
<td>145</td>
<td>0.07 (0.07)</td>
</tr>
<tr>
<td>50 MHz right dorsal paedal artery media thickness measured in mm</td>
<td>vhrsrdpamt50mhz</td>
<td>145</td>
<td>0.10 (0.06)</td>
</tr>
<tr>
<td><strong>Haemodynamic MRI Measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroke volume in millilitres</td>
<td>Mriao1sv</td>
<td>387</td>
<td>90 (17)</td>
</tr>
<tr>
<td>Cardiac output in litres</td>
<td>Mriao1co</td>
<td>387</td>
<td>5.7 (1.2)</td>
</tr>
<tr>
<td>Systemic vascular resistance in mmHg/L/min</td>
<td>Mriao1svr</td>
<td>387</td>
<td>15.6 (3.0)</td>
</tr>
<tr>
<td>Total arterial compliance in ml/mmHg</td>
<td>Mriao1tac</td>
<td>387</td>
<td>1.02 (0.22)</td>
</tr>
<tr>
<td><strong>Cardiac Structure MRI Measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left ventricular end diastolic volume in millilitres</td>
<td>Mrlvedv</td>
<td>387</td>
<td>135 (27)</td>
</tr>
<tr>
<td>Left ventricular end systolic volume in millilitres</td>
<td>Mrlvesv</td>
<td>387</td>
<td>46 (14)</td>
</tr>
<tr>
<td>Left ventricular stroke volume in millilitres</td>
<td>Mrlvsv</td>
<td>387</td>
<td>89 (17)</td>
</tr>
<tr>
<td>Left ventricular ejection fraction (%)</td>
<td>Mrlvef</td>
<td>387</td>
<td>66 (5)</td>
</tr>
<tr>
<td>Left ventricular mass</td>
<td>Mrlvmass</td>
<td>387</td>
<td>91 (25)</td>
</tr>
</tbody>
</table>
### Table 5. Metabolic/metabolomic phenotypes.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Database ID</th>
<th>n</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apolipoprotein B measured in g/L</td>
<td>ApoBgL</td>
<td>387</td>
<td>0.69 (0.23)</td>
</tr>
<tr>
<td>Apolipoprotein A measured in g/L</td>
<td>ApoAgL</td>
<td>387</td>
<td>1.16 (0.29)</td>
</tr>
<tr>
<td>C-reactive protein measured in mg/L</td>
<td>CRPngL</td>
<td>366</td>
<td>2.00 (3.21)</td>
</tr>
<tr>
<td>Adiponectin measured in µg/L</td>
<td>Adiponectinugml</td>
<td>386</td>
<td>27.1 (16.3)</td>
</tr>
<tr>
<td>Leptin measured in ng/mL</td>
<td>LeptinnngmL</td>
<td>387</td>
<td>38.8 (36.2)</td>
</tr>
<tr>
<td>Interleukin-6 measured in pg/mL</td>
<td>IL-6pgmL</td>
<td>375</td>
<td>1.45 (1.23)</td>
</tr>
<tr>
<td>Glucose measured in mmol/L</td>
<td>GlucosemmolL</td>
<td>385</td>
<td>3.75 (0.59)</td>
</tr>
<tr>
<td>Insulin measured in µU/mL</td>
<td>InsulinuUml</td>
<td>385</td>
<td>8.9 (6.2)</td>
</tr>
<tr>
<td>Gamma-glutamyl transferase measured in U/L</td>
<td>Ggtul</td>
<td>384</td>
<td>15.0 (14.8)</td>
</tr>
<tr>
<td>Alanine aminotransferase measured in U/L</td>
<td>Altul</td>
<td>385</td>
<td>17.0 (11.6)</td>
</tr>
<tr>
<td>Aspartate aminotransferase measured in U/L</td>
<td>Astul</td>
<td>384</td>
<td>24.8 (10.9)</td>
</tr>
<tr>
<td>Triglycerides measured in mmol/L</td>
<td>TRIGmmolL</td>
<td>384</td>
<td>0.73 (0.32)</td>
</tr>
<tr>
<td>Total cholesterol measured in mmol/L</td>
<td>CHOLmmolL</td>
<td>385</td>
<td>3.35 (1.01)</td>
</tr>
<tr>
<td>Low-density lipoprotein measured in mmol/L</td>
<td>LDLmmolL</td>
<td>384</td>
<td>1.85 (0.73)</td>
</tr>
<tr>
<td>High-density lipoprotein measured in mmol/L</td>
<td>HDLmmolL</td>
<td>384</td>
<td>1.03 (0.35)</td>
</tr>
<tr>
<td>Paraoxonase-1 Activity in µMol/P-Nitrophenol/ L/Serum/min</td>
<td>PONActivityuMolPNitrophenol</td>
<td>304</td>
<td>877 (442)</td>
</tr>
</tbody>
</table>

Metabolomics: See extended data for complete list of all 225 biomarkers measured using NMR.

### Table 6. Psychological questionnaires.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Database ID</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interpersonal Support Evaluation List</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISEL questions 1 – 12 as scored by participants</td>
<td>Iselq1 – Iselq12</td>
<td>224</td>
</tr>
<tr>
<td>Scaled and reversed totals of questions 1 – 12</td>
<td>Iselscttl1 – Iselscttl12</td>
<td>224</td>
</tr>
<tr>
<td>Appraisal subscale – calculated from total combined score of questions 2, 4, 6 and 11.</td>
<td>Iselappssubs</td>
<td>224</td>
</tr>
<tr>
<td>Belonging subscale – calculated from total combined score of questions 1, 5, 7 and 9.</td>
<td>Iselbelsubs</td>
<td>224</td>
</tr>
<tr>
<td>Tangible subscale – calculated from total combined score of questions 3, 8, 10 and 12.</td>
<td>Iseltangsubs</td>
<td>224</td>
</tr>
<tr>
<td><strong>Oxford Happiness Questionnaire</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OHQ questions 1 – 29 as score by participants</td>
<td>Oxhrqa1 – Oxhrqa29</td>
<td>224</td>
</tr>
<tr>
<td>Converted totals for negatively phrased questions</td>
<td>Oxhrct1 – Oxhrct29</td>
<td>224</td>
</tr>
<tr>
<td>Sum of converted totals</td>
<td>Oxhrsumct</td>
<td>224</td>
</tr>
<tr>
<td>Final score</td>
<td>Oxhrscore</td>
<td>224</td>
</tr>
<tr>
<td><strong>Beck Depression Inventory II</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDI-II questions 1 – 21 as scored by participant</td>
<td>Prsonqa1 – Prsonqa21</td>
<td>224</td>
</tr>
<tr>
<td>Final Score</td>
<td>Prsonscore</td>
<td>224</td>
</tr>
</tbody>
</table>
Dataset validation
In order to minimise variability inherent in the assessment of certain vascular phenotypes (e.g. carotid intima-media thickness, pulse-wave velocity, flow-mediated dilation), all study technicians working on the study were required to undergo identical training and accreditation procedures as those conducted in previous ALSPAC vascular clinics prior to scanning participants, details of which have been published elsewhere. All blood samples were analysed in a single batch upon completion of the study in order to minimise any potential inter-assay variability. As with all large-scale studies, a number of variables have different degrees of missing data. Information on the extent of missing data for each variable can be found within each of the descriptive tables provided within this data note. Of note, around 10% of participants refused to provide a blood sample due to fear of needles. In addition, ultra-high frequency resolution ultrasound and psychological questionnaires were only added to the main study protocol in 2013, and data for these variables are therefore available in only a sub-sample of participants.

Data availability
Underlying data
ALSPAC data access is through a system of managed open access. Full details of all available data can be accessed through a fully searchable data dictionary provided on the ALSPAC study website (http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary) and the steps below highlight how to apply for access to both the data included in this data note and all other ALSPAC data. The datasets presented in this data note are linked to ALSPAC project number B645; please quote this project number during your application. The ALSPAC variable codes highlighted in the dataset descriptions can be used to specify required variables.

1. Please read the ALSPAC access policy (PDF, 627kB) which describes the process of accessing the data and samples in detail and outlines the costs associated with doing so.

2. You may also find it useful to browse our fully searchable research proposals database, which lists all research projects that have been approved since April 2011.

3. Please submit your research proposal for consideration by the ALSPAC Executive Committee using the online process. You will receive a response within 10 working days to advise you whether your proposal has been approved.

If you have any questions about accessing data, please email alspac-data@bristol.ac.uk.

The ALSPAC data management plan describes in detail the policy regarding data sharing, which is through a system of managed open access.

Extended data
Extended data, including example graphs showing relationships between BMI and other phenotypes, descriptive metabolomics data and the ALSPAC in London medical history questionnaire, can be found in the figshare data repository. DOI: https://doi.org/10.6084/m9.figshare.c.4337849.v1.

Data are available under the terms of the Creative Commons Zero “No rights reserved” data waiver (CC0 1.0 Public domain dedication).

Grant information
This publication is the work of the authors and STC will serve as guarantor for the contents of this paper. The UK Medical Research Council and Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. This research was specifically funded by the British Heart Foundation (BHF Programme Grant No. RG/10/004/28240). JED was a BHF-funded Professor during the time of this study. NJT is a Wellcome Trust Investigator (202802/Z/16/Z), is the PI of the Avon Longitudinal Study of Parents and Children (MRC & WT 102215/2/13/2), is supported by the University of Bristol NIHR Biomedical Research Centre (BRC-1215-20011), the MRC Integrative Epidemiology Unit (MC_UU_12013/3) and works within the CRUK Integrative Cancer Epidemiology Programme (C18281/A19169). KHW was supported by a Wellcome Trust Grant (096989/Z/11/Z). A comprehensive list of grants funding is available on the ALSPAC website.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements
We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

References


Open Peer Review

Current Peer Review Status: ?

Version 1

Reviewer Report 10 July 2019

https://doi.org/10.21956/wellcomeopenres.16296.r35839

© 2019 Steinberger J. This is an open access peer review report 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.

Julia Steinberger
Division of Pediatric Cardiology, Department of Pediatrics, Medical School, University of Minnesota, Minneapolis, MN, USA

The authors describe the ‘ALSPAC in London’, a dataset containing a large panel of comprehensive measurements of cardiometabolic risk factors. In the current paper, they report the data available in 419 adults with complete datasets with the purpose of examining the impact of adiposity on metabolic, inflammatory and other CV risk markers, as well as the relationship with cardiovascular structure, function and genetic phenotyping. The overarching goal of this study is important in the current obesogenic environment and ALSPAC is commendable for collecting extensive information on phenotype and genetic factors.

A number of concerns reduces the enthusiasm for this report:
1. The main research question posed in the aim of this study, i.e. understanding the contribution of genetic and other life influences to adiposity profiles in young adulthood has not been answered.
2. The data presented seem to be summary statistics of the cohort with little information on how to interpret the data.
3. There is a concern that this study may be under-powered (n=419) with complete metabolic and genomic data to answer the main research question.
4. The methods section is confusing regarding the number of individuals recruited at each stage and these numbers do not clearly overlap with the numbers presented in the study flow diagram.

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

**Are the datasets clearly presented in a useable and accessible format?**
No

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Pediatric cardiology, Preventive cardiology, Metabolic syndrome, Cardiovascular risk factors

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

17 January 2019

Reviewer Report

[https://doi.org/10.21956/wellcomeopenres.16296.r34484](https://doi.org/10.21956/wellcomeopenres.16296.r34484)

© 2019 Ritchie S. This is an open access peer review report 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.

Scott C. Ritchie

Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

This article presents the ALPSPAC in London dataset, which contains genotype data, adiposity, cardiovascular, and metabolomic phenotypes, and psychological survey questionnaire data for 436 adults recruited from the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort using a recall-by-genotype study design to identify and recruit individuals genetically predicted to have low or high BMI.

To identify individuals predicted to be at the tail ends of the BMI distribution based on their genotype, a genomic risk score (GRS) for BMI was calculated using pre-existing genotype data from all ALSPAC individuals. The BMI GRS was calculated by summing the effect sizes from the 2010 BMI GWAS conducted by Speliotes et al. multiplied by their effect allele in ALSPAC participants for all directly genotyped variants.

In our experience this approach to GRS calculation is likely to lead to an underpowered BMI GRS due to the use of only directly genotyped variants and the lack of any training procedure to select the optimal set of variants.

A critical question that is not yet answered by this article is how successful the BMI GRS was at identifying participants at the tail ends of the BMI distribution at recall. The section Dataset 2 should provide details on the distribution of measured BMI in the participants in each of the defined percentile groups, as well as in the 17 participants who were not recruited by the recall-by-genotype procedure, and the text should at least provide the average and standard deviation of BMI in the lower and upper tail groups.

I regret that I lack the expertise to review the other aspects of phenotype collection or study design.
As to the question, "Are the datasets clearly presented in a useable and accessible format?", I have answered no, because the datasets presented are simply cohort-level summary statistics for the dataset, rather than accessible and useable links to the data itself. The cohort data itself is only accessible after writing a project proposal which may be considered by the ALSPAC Executive committee.

References

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Partly

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
No

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Genomic risk scores, data analysis, systems genomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.