Dietary metabotype modelling predicts individual responses to dietary interventions

Isabel Garcia-Perez, Joram M. Posma, Edward S. Chambers, John C. Mathers, John Draper, Manfred Beckmann, Jeremy K. Nicholson, Elaine Holmes and Gary Frost

Habitual consumption of poor quality diets is linked directly to risk factors for many non-communicable diseases. This has resulted in the vast majority of countries and the World Health Organization developing policies for healthy eating to reduce the prevalence of non-communicable diseases in the population. However, there is mounting evidence of variability in individual metabolic responses to any dietary intervention. We have developed a method for applying a pipeline for understanding interindividual differences in response to diet, based on coupling data from highly controlled dietary studies with deep metabolic phenotyping. In this feasibility study, we create an individual Dietary Metabotype Score (DMS) that embodies interindividual variability in dietary response and captures consequent dynamic changes in concentrations of urinary metabolites. We find an inverse relationship between the DMS and blood glucose concentration. There is also a relationship between the DMS and urinary metabolic energy loss. Furthermore, we use a metabolic entropy approach to visualize individual and collective responses to dietary interventions. Potentially, the DMS offers a method to target and to enhance dietary response at the individual level, thereby reducing the burden of non-communicable diseases at the population level.

One in five deaths are directly attributable to poor diet. Evidence from large interventions has shown that the risk of type 2 diabetes and coronary artery disease is reduced by a healthier lifestyle, which includes consumption of a high-quality diet, and population-based national and international dietary guidelines for non-communicable disease (NCD) prevention have been developed. However, recent evidence shows that individuals respond differently to the same diet in terms of the impact on metabolism and clinical outcomes. Exploitation of this heterogeneity in dietary response has given rise to the concept of precision nutrition, the ambition of which is to optimize individual responses to dietary intervention on the basis of understanding the determinants of these individual responses to nutritional intakes. Current tools for measuring individual dietary intake, such as food-frequency questionnaires, dietary recalls or diet diaries are subject to substantial reporting bias and are inadequate for assessing individual compliance with dietary recommendations. In addition, recent evidence indicates that matching diet to genotype produces no further benefit in dietary response over personalized dietary advice. This highlights the need for new technical approaches for understanding an individual’s metabolic response to a dietary intervention.

Metabolic phenotypes (metabotypes) are the products of interactions between an individual’s genotype and multiple environmental factors, including diet, other lifestyle factors and the gut microbiome. In this feasibility study, we develop and apply precision nutrition tools for mapping the dietary response at the individual level. Our aim is to provide a measure of metabolic variability and identify interindividual differences in the response to diet to individualize healthy eating advice, which will lead to a reduction in the risk of NCDs.

We employed a previously reported methodology to capture a systemic overview of metabolic response to each of four specific standardized diets using metabolic phenotyping of urine samples. We developed four experimental dietary interventions with stepwise degrees of concordance to the World Health Organization’s (WHO) healthy eating guidelines (increase fruits, vegetables, wholegrains and dietary fibre, and decrease fats, sugars and salt). Diet 1 is the most concordant, Diet 4 the least concordant, and there are two intermediate diets (Diets 2 and 3). Information on meals that were provided to participants in each diet are detailed in Extended Data Fig. 1, including the macronutrient content. Healthy participants attended a clinical research unit on four separate occasions and followed these diets in random order. Each intervention period lasted for three days, during which participants were observed continuously to ensure adherence to the diet. Figure 1 is a schematic of the study design. In the following series of experiments, we use this analytical framework to explore the role of individual metabolic profiles and the physiological importance of these responses.

Results

Understanding the dietary metabotype phenotype. The metabolic phenotypic response of participants to each diet, characterized by 1H NMR spectroscopy and principal component analysis of 24 h
urine samples (pooled from each void throughout the 24 h period) were found to cluster according to diet (Fig. 2a, with representative spectra shown in Fig. 1 and Supplementary Fig. 1). This indicates that participants’ urinary metabolic phenotypes were modulated systematically by changing diet. It is of interest that the urinary metabolic phenotype of participants after following Diet 1 (most concordant with WHO guidelines) are consistent with phenotypes previously associated with lean individuals, for example, higher levels of hippurate and vitamin-related compounds such as niacin; whereas the Diet 4 metabolic phenotype reflects that reported for obese individuals (for example, higher levels of glucose, carnitines, fatty acids and so on) (Supplementary Table 1). Despite the rigour of the controlled environment and participants’ adherence over each three-day intervention period, there was considerable interindividual variability in the urinary metabolic phenotype and the excretion of individual metabolites in response to each diet, as exemplified by Fig. 2b–d (hippurate associated with fruit and vegetable intake, 3-methylhistidine associated with (lean) meats).

**Fig. 1 | Schematic representation of methodology.** Four diets with different levels of concordance with WHO healthy eating guidelines were provided to each of 19 healthy participants in a random order. The macronutrient content displayed in the pie charts shows the proportions of energy from fats (red slice), carbohydrates (orange slice) and proteins (green slice) for the most (Diet 1) to the least (Diet 4) concordant diets. Fibre (blue bar), sugar (pink bar) and the DASH score (green bar, the higher the better) for each of the four diets are shown in the adjacent histograms. Twenty-four-hour urine samples were collected daily over each three-day period and analysed using 1H NMR spectroscopy to generate one metabolite profile per sample. The urinary metabolite profiles from the most and least concordant diets (Diets 1 and 4) were used to create a mathematical model to calculate a score that summarizes the response of an individual to diet (DMS). Finally, we investigated relationships between the DMS and urinary energy loss, metabolic networks and blood glucose concentration.
Development of a Dietary Metabotype Score. Urinary metabolite profiles reflect the phenotype of individuals and are particularly responsive to dietary interventions\(^2\). We developed the Dietary Metabotype Score (DMS), based on a Monte Carlo cross-validation partial least squares (MCCV-PLS) model (Methods and Supplementary Fig. 3) to quantify individual variation in urinary metabolic phenotype in response to a test diet. Thus, the DMS for any individual at any given time quantifies the similarity of that individual's urinary metabolite profile to the metabolic phenotypes of the remaining volunteers of the two extreme diets\(^\text{12,21}\) (Diets 1 and 4) with +1 reflecting a score that would represent full adherence to Diet 1 (most concordant with WHO guidelines) and −1 reflecting a spectral profile obtained after full adherence to Diet 4 (least concordant). Since the DMS is based on the whole spectral profile (16,000 variables, which includes metabolites we have identified here, as well as (presently) unknown signals that are associated with the different diets, Fig. 3), changes in both the number and the intensity of signals (relating to diversity and concentration of metabolites) will contribute to the end score; thus, a high DMS will reflect stronger changes in number and/or concentration of metabolites.

The predicted DMS of participants consuming diets more adherent to WHO guidelines\(^2\) (Diet 1) (Fig. 4a and Supplementary Fig. 4) was higher than the DMS calculated when participants were following Diet 4 (least concordance with guidelines). The DMS is presented in rank order (colour bar to the right of Fig. 4a indicates the rank of each participant) to visualize the variability in response to Diet 1. For instance, the participants with the strongest metabolic response to Diet 1 will be at the top of the ranking order (green region of the colour bar) while the one with the least pronounced metabolic response to Diet 1 will be at the bottom of the ranking order (red region of the colour bar). As expected, each participant shows a gradual increase in DMS from Diet 4 to Diet 1; the more concordant the diet is with WHO guidelines, the higher the DMS. However interindividual variability in response to each of the four diets is observed. In addition, data from days 1 and 2 can still be predictive of the diet intervention; however, they could also reflect markers related to their habitual diet at home. Therefore, we applied the same framework on the day 2 samples and found similar classification accuracies as for the day 3 model, but with a lower goodness of prediction (\(Q^2\)) value (\(Q^2 = 0.82\) (day 2), \(Q^2 = 0.85\) (day 3 model)). The important variables are very similar between these two models (Fig. 3, Supplementary Fig. 5 and Supplementary Table 1B). Combining data from days 1–3 does give a good predictive model; however, there is a reduced significance of multiple metabolites (Supplementary Fig. 6 and Supplementary Table 1B), which could relate to including data of short-term exposure (day 1 samples).

The variability in the DMS between participants for different diets was not attributed to age, BMI, fat percentage of total body weight (Supplementary Fig. 7) or sex (Supplementary Fig. 8). However, some metabolites that are excreted in different amounts in women and men (Fig. 2 and Supplementary Fig. 2) are consistent with previous literature\(^3\). Comparisons of our model (Fig. 3) with sex-specific models (Supplementary Figs. 9 and 10) reveal that with the exception of creatine in the men-specific model, all metabolites from the DMS model are significant in the models built from male participants and in the models built from female participants (Supplementary Table 1B). Also they have the same sign of association as the combined model. The sex-specific models have a smaller number of samples as part of the training set, and a larger number of samples as part of the test data, than the DMS model. Therefore, we compare our DMS model with a model where the training-to-test set ratio is similar to that of sex-specific models. The splitting (50:50 opposed to 80:20) does not impact on the associated metabolites (Supplementary Fig. 11). In addition, we calculate two regression models with the Dietary Approaches to Stop Hypertension (DASH) score calculated for the four diets and the adherence to WHO guidelines (Supplementary Figs. 12 and 13) and while these models are predictive with multiple coherence of differential metabolites between WHO concordant and discordant diets, not all metabolites previously found in the DMS model are associated with these two subsquent models. Both of these regression models are relatively
linear (Supplementary Figs. 12A and 13A); however, not all dietary components (foods and macronutrients, Extended Data Fig. 1) are linear with the diets, such as protein, and this is reflected in the urea, which does not significantly contribute to the DMS model but significantly contributes to the DASH score regression model and the WHO alignment regression model.

To validate these findings, we used a new metabolomic discovery study with a less stringent control. Ten volunteers were fed a standardized diet that reflects Diet 1 for four days with all their food being provided for breakfast, lunch and two snacks. The evening meal was provided to the volunteer to consume at home (see Methods for details). We found similar results to the initial study with a large between individual variability, despite the dietary control. As the diet was the same over four days, there was no linear change over time (Supplementary Fig. 14A). Next, we combined all the data for each person and represent this in a violin plot ordered based on the median DMS (Supplementary Fig. 14B), which shows the distribution of the DMS for each participant combining the four days. This demonstrates that some participants have very similar DMSs, whereas others differ significantly, highlighting the occurrence of interindividual variability.

Calculation of the DMS in this cohort revealed that dietary habits in the home environment are highly variable between individuals as well as within individuals (Extended Data Fig. 1). For example, participant 16 shows a high DMS (aligning with WHO guidelines) before starting all diets, while participant 10 shows a higher DMS before the admission to Diets 1 and 4 and a lower DMS before starting Diets 2 and 3. Since participants’ DMS at baseline does not determine the DMS after the test diets (Fig. 4b,c), the order of the DMS before Diet 1 does not relate to the DMS ranking order after Diet 1 (Fig. 4b) and the order before Diet 4 does not relate to the DMS ranking after Diet 4 (Fig. 4c). Moreover, Fig. 4b,c shows a clear decrease in the variability of participants’ DMS after exposure to the controlled diets for three days. It appears that, in general, the volunteers’ DMS score before the diet fell between the most and least concordant diets, with 12/19 participants increasing their score in response to the most and least concordant diets, with 12/19 participants decreasing their score in response to Diet 4 (least concordant).

**Exploring the DMS and the impact on physiology.** Individual DMS values were associated with health-related biomarkers. For example, there was an inverse association between DMS rank and blood glucose (Supplementary Table 2) for the total post-prandial glucose across all four diets (Fig. 4d, $P = 3.20 \times 10^{-3}$), and within each diet for both post-prandial (Fig. 4d, $P = 2.59 \times 10^{-2}$ to $3.42 \times 10^{-3}$) and fasting glucose (Supplementary Fig. 15, $P = 2.17 \times 10^{-2}$ to $3.30 \times 10^{-3}$). Adjustment for age, sex and BMI did not change the magnitude of these correlations appreciably. We validated these observations in a cohort of healthy middle-age overweight individuals ($n = 65$, 31 females, age $59 \pm 4$, BMI $= 29.0 \pm 2.9 \text{kg m}^{-2}$) (Methods). We demonstrated that the DMS was inversely related to the area under the glucose and insulin curves following mixed-meal tolerance tests ($P = 1.65 \times 10^{-2}$ and $P = 4.39 \times 10^{-2}$, respectively), plus fasting high-density lipoprotein cholesterol ($P = 1.33 \times 10^{-2}$), the ratio of total to high-density lipoprotein cholesterol ($P = 2.28 \times 10^{-2}$) and blood triglycerides ($P = 4.39 \times 10^{-2}$) (Supplementary Fig. 16). However, the relationship of the DMS to fasting blood glucose was
The urinary energy content was significantly related to higher excretion of multiple metabolites (Supplementary Table 3), including metabolites of microbial origin (hippurate, phenylacetylglutamine, formate, 4-cresylsulphate, 4-cresylglucuronide and 2-hydroxyisobutyrate). The formation of microbial metabolites is endothermic, requiring energy, but also represents a potentially significant energy loss to the host via phase II metabolic conversions of selected carbon sources, which renders the carbon skeleton unavailable for energy metabolism by the host. For example, the conjugation of phenylacetate and glutamine to form phenylacetylglutamine results in a loss of energy by decreasing the accessibility of glutamine for energy metabolism. However, microbial metabolites typically also: for example, hippurate is higher in urine samples from Diet 1 due to the increased intake of fruits and vegetables, whereas phenylacetylglutamine is higher in urine after Diet 4 and not found to be significant ($P=1.88 \times 10^{-1}$). The validation cohort supports the observed association of DMS with a number of important health-related markers.

We measured the energy content of the urine directly using bomb calorimetry (Methods). The range of urine energy values was similar to those previously reported in healthy humans ($P=1.88 \times 10^{-1}$). The validation cohort supports the observed association of DMS with a number of important health-related markers.

We measured the energy content of the urine directly using bomb calorimetry (Methods). The range of urine energy values was similar to those previously reported in healthy humans ($P=1.88 \times 10^{-1}$). The validation cohort supports the observed association of DMS with a number of important health-related markers.

We measured the energy content of the urine directly using bomb calorimetry (Methods). The range of urine energy values was similar to those previously reported in healthy humans ($P=1.88 \times 10^{-1}$). The validation cohort supports the observed association of DMS with a number of important health-related markers.

We measured the energy content of the urine directly using bomb calorimetry (Methods). The range of urine energy values was similar to those previously reported in healthy humans ($P=1.88 \times 10^{-1}$). The validation cohort supports the observed association of DMS with a number of important health-related markers.

We measured the energy content of the urine directly using bomb calorimetry (Methods). The range of urine energy values was similar to those previously reported in healthy humans ($P=1.88 \times 10^{-1}$). The validation cohort supports the observed association of DMS with a number of important health-related markers.
related to the increased consumption of animal protein in this diet (Supplementary Table 1A).

While the DMS ranking order and energy loss are related and share common metabolites (hippurate, phenylacetylglutamine, as well as non-microbial metabolites), there are additional microbial metabolites associated only with urinary energy content (formate, 4-creosylsulphate, 4-creosylglucuronide and 2-hydroxyisobutyrate) that reflect the influence of the gut microbiota on urinary energy loss.

Impact of dietary intervention on individual metabolic network maps. To understand the wider impact of the dietary intervention on human metabolism, we generated a urinary metabolic reaction networks map (Fig. 5a) using a previously described methodology to visualize reactions between metabolites in relation to associated metabolic pathways27. We constructed this metabolic network map using the set of 32 metabolites as anchor points, where the excretion was significantly different after following Diet 1 (most concordant with WHO dietary guidelines) and Diet 4 (least concordant) for all 19 healthy participants41. The connectivity of these 32 metabolites (that is, the minimum number of biochemical reactions required to connect all 32 metabolites) takes place through 234 metabolites (white nodes) as part of the human supra-organism metabolic reaction network27.

Compounds associated with lipid, glucose and energy metabolism explain a greater proportion of the metabolic phenotype in Diet 4 compared with Diet 1. This observation aligns with the blood chemistry measurements which, despite the short intervention period, are significantly different between Diets 1 and 4 for average post-prandial glucose ($P=0.036$) and show a trend towards a lower fasting total cholesterol ($P=0.056$) and triglycerides ($P=0.058$) (Supplementary Table 2).

All 33 metabolites show interindividual variation in excretion between participants (Fig. 2b–d and Supplementary Fig. 2), suggesting homoeostatic mechanisms are, to an extent, specific for each person. To quantify this type of metabolic variation, we calculate a measure of system disorder (metabolic entropy) based on the fold-change in excretion of each of the 33 urinary metabolites to measure the robustness of a person's metabolism under dietary challenge of Diets 1 and 4 (Methods). This entropy measure was used to map the metabolic differences (in response to Diets 1 and 4 for each participant) onto the skeleton outline of the urinary metabolic reaction networks map, as exemplified in Fig. 5b,c for participants 6 and 16 (highest and lowest metabolic entropy, respectively) and Extended Data Figs. 2 and 3 (all other participants). In each of these figures, the size of the metabolite node is proportional to the metabolite fold-change between Diets 1 and 4, and the paths connecting two metabolites are weighted by the average fold-change of the two metabolites in the respective diets (Methods) to visualize the different utilization of pathways at the individual level. The higher the associated entropy of a pathway, the larger the potential energy loss from that pathway to the urine.

Each individual metabolic network highlights the expression of the participant's metabolic phenotype and allows us to further investigate participants’ variability in response to each reference diet based on the entropies associated with specific pathways. For instance, participant 6 (highest metabolic entropy, second-highest DMS and second-highest urinary energy excretion) exhibits higher expression of most of the pathways associated with the 32 metabolites compared with the other participants (Fig. 5b and Extended Data Figs. 3 and 4). In contrast the metabolic network for participant 16 (lowest metabolic entropy, lowest DMS and lowest urinary energy excretion) indicates that specific compartments, mostly related to TCA anaplerotic and mitochondrial coenzyme A (CoA) metabolism, may contribute more to their urinary residual energy in this participant than other pathways such as histidine and beta-alanine metabolism (Fig. 5c).

Discussion

We have used individuals’ urinary metabolic phenotypes to demonstrate that even in a highly controlled environment, individuals’ metabolic responses to diet differ. Our data suggest that everyone has a unique dietary metabolotype that relates to the individuals’ physiological homeostasis, as demonstrated by glycaemic control. Our results support previous studies42–46 in highlighting the need to understand individual response to lifestyle to enhance response to dietary intervention.

In this feasibility study, we have found that among a ‘healthy’ group of subjects, the participants who have a higher loss of energy in the urine (that is, those that also are at the top of the dietary metabolotype score ranking) demonstrate greater urinary excretion of microbial metabolites. The energy content of urine in humans has so far been little studied but, as far back as 1901, Rubner observed that in addition to urea there were calorific substances in urine49. Here we describe a portfolio of microbial gut-derived

---

**Fig. 5 | Urinary metabolic reaction networks built using the MetaboNetworks software.** a. Urinary metabolic reaction networks map27 visualizing connectivities between 32 urinary metabolites, the excretion of which significantly differed between Diet 1 (most concordant with WHO dietary guidelines) and Diet 4 (least concordant) for all 19 healthy participants. The node colour indicates whether the metabolite is excreted in higher concentrations in Diet 1 (green) or Diet 4 (red). Each edge (line between nodes) indicates a reaction (substrate–product relationship) and the white nodes (only named for nodes with three or more reactions) represent metabolites that connect the identified differential biomarkers of diet/dietary response in the human supra-organism metabolic reaction network (Methods). The background shading indicates different class or type of metabolic pathways. b, c. The individualized differences in response to Diets 1 and 4 have been mapped onto the skeleton outline of the urinary metabolic reaction network (a), where the size of coloured nodes is proportional to the fold-change of that metabolite between Diets 1 and 4 for that person. The edge colours are related to the sum of fold-changes of each pair of metabolites connected via their shortest path (Methods). The metabolic entropy for each individual is calculated as the sum of the individual fold-changes40. The more disordered or metabolically flexible (high entropy), a subnetwork is, the proportion of the superorganism’s metabolic network explained by a pathway is greater. d. The different utilization of pathways in the different individuals is represented using the metabolic entropy (Methods) to visualize the different utilization of pathways at the individual level. The higher the associated entropy of a pathway, the larger the potential energy loss from that pathway to the urine.
metabolites that could contribute to the energy content of urine and which account for variation in urinary energy values between individuals. We did not collect faeces for microbiota analysis during this study as gut transit in humans is relatively long and variable and the length of dietary intervention in this study was three days, so a faeces collected on day three may not represent the microbial change induced by the diet. In humans, energy loss in the excreta so a faeces collected on day three may not represent the microbial modulation as a way of preventing weight gain remains unknown at present.

On the basis of the observed interindividual differences in urinary energy loss, we explored the variability in response to the same diet at a metabolic pathway level to gain better insight into the intricate regulatory networks shaping a given phenotypic response. By using individual urinary metabolic reaction network maps, we showed substantial variation between the composite metabolic outputs of individuals that had a similar DMS. The mechanism driving

---

**Diagram Image**: A network diagram showing metabolic pathways and their interactions. The diagram is color-coded to represent different metabolic processes such as lipid and fatty acid metabolism, mitochondrial (CoA) metabolism, and carbohydrate metabolism. The network includes various metabolites such as acetyl-CoA, malonyl-CoA, and succinyl-CoA, as well as pathways like the tricarboxylic acid cycle (TCA) and anaplerotic energy metabolism. The diagram also highlights the involvement of gut microbial (co-)metabolism. The colors range from green to purple, indicating different levels of reaction/edge entropy based on joint (normalized) entropy of biomarkers.
these observations is currently unknown, but individual genotype and/or microbial composition are likely to contribute. A limitation of these maps is that they are built using urinary metabolites alone. Nevertheless, before a metabolite ends up in the urine, it is generally in the plasma/circulatory system (or of renal origin) and thus metabolites in 24 h urine samples are a time-averaged representation of the homeostatic signatures of multiple organ and cellular systems in the body.

In summary, our feasibility study demonstrates that there are interindividual differences in response to the same diet, even when the environment is controlled. We have developed a modelling framework that could be used to monitor individual response to diet and that provides a mechanism for enhancing dietary advice with the potential to inform decision-making strategies for the prevention, risk reduction and clinical management of NCDs for precision medicine.

Methods

Study protocol, ethics and consent. The study was approved by the London–Brent Research Ethics Committee and carried out in accordance with the Declaration of Helsinki (13/LO/0078). The study protocol is available from ref. 21. All participants provided written informed consent. An external cohort (NutriTech) was used to validate associations with health-related markers (London–Brent Ethics Committee reference number: 12/LO/0339). This dataset contained 65 participants (31 female), with a mean BMI of 29.0 ± 2.9 kg m⁻² at baseline, for which urine samples were available from baseline measured before starting the main study ‘Dietary biomarker discovery using metabonomics’. The study was approved by the London–Brent Research Ethics Committee and carried out in accordance with the Declaration of Helsinki (13/LO/0078).

Validation study design. The methodology used was a four-day controlled diet coupled to a food challenge study. The aim of the challenge study is to identify and quantify potential candidate dietary biomarkers. The study we used to report the stability of the metabolite profile aimed to develop a biomarker for egg intake. Ten healthy participants were recruited (4 males and 6 females, aged 23 to 34, BMI range from 18 to 27, mean BMI 22.3 ± 3 kg m⁻²). All participants provided written informed consent before the beginning of the study. Breakfast (08:00) and lunch (13:00) were provided to the participants (details of the food consumed are given in Supplementary Table 4). The participants were given an evening meal consisting of rice, carrots, lettuce, salt and olive oil, plus the test food (eggs) in increasing doses over the four days (day 1, 0 eggs; day 2, 1 egg; day 3, 2 eggs; day 4, 4 eggs). The nutrient profile was 64% energy from carbohydrate, 26% fat and 10% protein, which aligns with Diet 1. After consuming the meal, participants collected their 12th overnight urine excretion from 20:00–08:00. Urine samples were collected into sterilized single containers overnight, and then stored in nitrogen gas and weighed on an analytical balance. The samples were dried using nitrogen gas and weighed on an analytical balance (precision < 0.001 g).

NMR spectroscopy and data pre-processing. Urine samples were prepared with a pH 7.4 phosphate buffer for ¹H NMR spectroscopy as described previously21 and NMR spectroscopy and data pre-processing. Urine samples were prepared with a pH 7.4 phosphate buffer for ¹H NMR spectroscopy as described previously21 and NMR spectroscopy and data pre-processing.
50:50 (irrespective of sex). Lastly, we used the DASH score of the diets and the adherence of each diet to WHO guidelines in two regression models (using day 3 samples) to use the urinary metabolites to predict these. The framework used for these analyses is the same as for the DMS.

The repeated measures design of the data is accounted for in the centring of the data. Any data used in the modelling is centred within-person: all data for each person (regardless of whether they are in a training or test set) is centred individually. This allows for the model to focus on changes in metabolite excretions relative to each individual’s metabolic phenotype. The scaling of variables (unit variance) was applied using all data after the centring step.

The standard deviation from the training data is used and applied on the test and validation data sets—so there is never any information from the test and validation sets in the calculation of the model (for choosing the optimal number of components we used a further splitting of the training data in each iteration to inform on the number of components).

The samples from both the NutriTech and metabolite discovery study were centred based on the average of the main study data, and scaled using the parameters from each individual model (training data). As these samples are from independent individuals, all 1,000 models from the MCCV are used to calculate the DMS.

Statistical analysis (urinary energy). The 24h urinary 1H NMR spectroscopic profiles of all participants were regressed against the calorific value using standard linear regression, adjusted for multiple testing using the Storey–Tibshirani false discovery rate (FDR) (Partial) Spearman correlation was used to compare the DMS ranking with other study outcomes (glucose, urinary energy).

Bioinformatics analysis (MetaboNetworks maps). A urinary metabolic reaction network map was constructed using the MetaboNetworks software by connecting the 33 metabolites that were identified from the MCCV-PLS model using the MetaboNetworks database, any reaction that can occur either spontaneously, due to an enzyme linked to a Homo sapiens gene, or due to an enzyme linked to a microbial gene (3,282 species were included from the phyla actinobacteria, bacteroidetes, cyanobacteria, firmicutes, fusobacteria, proteobacteria, tenericutes and verrucomycetes) was included.

Metabolic entropy calculations. The entropy was calculated using the metabolite excretion data from both Diet 1 and Diet 4 and is a single value for each person that is related to the within-individual differences associated with Diets 1 and 4. Metabolites found in higher concentrations in urine of participants after following Diet 1 (‘M1’) and those metabolites found in higher concentrations in urine after Diet 4 (‘M4’) are considered to be in the perturbed state. The corresponding baseline levels were the concentration of M1 metabolites in Diet 1 (‘B1’) and M4 metabolites in Diet 1 (‘B1’). For simplicity, two data matrices of fitted metabolite data are used, M for metabolites in the perturbed state (n = 19 participants) by p = 33 (metabolites), both M1 and M4 and B for metabolites in the baseline state (dimension 19 x 33, both B4 and B1). This allows for the calculation of a metabolic entropy contribution (w)j = logj(33)] of metabolite j and metabolite entropy contribution s for individual i and metabolite j, normalized for the weight (w) of metabolite j for each metabolite per participant based on a fold-change in excretion between Diet 1 and Diet 4 for each participant normalized by the log-normalized standard deviation (w = std(log(Rij))) of the baseline levels—alognough to similar methods.

However, metabolites that relate directly to specific foods (such as tartarate and N-acetyl-S-(12)-propenyl-cysteine-sulfoxide) will exhibit a large fold-change, thus large entropy, and dominate the total entropy over changes in endogenous metabolites. To make the comparison between metabolites more meaningful, we autoscale (mean-centred followed by division by standard deviation, s) the columns of s (sj = sij - sj) before calculating the metabolic entropy (S) per participant. S is calculated as the sum across the columns of s (S = \sum sj). As we normalized the fold-changes and then autoscaled the data, the entropy is expressed in arbitrary units.

Using the MetaboNetworks urinary metabolic reaction network as a layout, individual metabolic networks were constructed for each individual where the size of each node (metabolite) is proportional to the metabolite entropy (s). Next, for each pair of metabolites the shortest path (number of reactions) was extracted from the network and each edge (reaction) in the path was weighted by the average entropy of the two metabolites. For individuals where pairs of excreted metabolites are correlated, the reactions (and associated pathways) that connect those metabolites will have higher weights than those from pairs of metabolites that are excreted in lower amounts and/or are uncorrelated. This highlights the pathways that are associated with the individual’s metabolic profile.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
All presented data are tabulated and detailed in the main text and the Supplementary Information. The study protocol availability is detailed in the Methods. Diets provided to participants are detailed in the Supplementary Information. Quantified NMR data, DMS, AUC glucose and calorific value for Diets 1 and 4 presented here are freely available (CC BY-NC 3.0) from MendleData at https://doi.org/10.17632/xv7c7nflfd.1.

Code availability
The codes for executing the MCCV-PLS (with repeated measures) algorithm can be obtained from https://bitbucket.org/jmp111/capsl/src/. The code for executing the STORM algorithm can be obtained from https://bitbucket.org/jmp111/storm/src/. These can be executed in a MATLAB environment.

Received: 20 June 2019; Accepted: 6 May 2020; Published online: 17 June 2020

References
1. Forouhi, N. G. & Unwin, N. Global diet and health: old questions, fresh evidence, and new horizons. Lancet 393, 1916–1918 (2019).
2. GBD 2017 Diet Collaborators Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 393, 1958–1972 (2019).
3. Zeevi, D. et al. Personalized nutrition by prediction of glycemic responses. Cell 163, 1079–1094 (2015).
4. Price, N. D. et al. A well-controlled randomized study of 188 individuals using personal, dense, dynamic data clouds. Nat. Biotechnol. 35, 747–756 (2017).
5. Minot, S. et al. The human gut virome: inter-individual variation and dynamic response to diet. Genome Res. 21, 1616–1625 (2011).
6. Holmes, E. et al. Human metabolic phenotype diversity and its association with diet and blood pressure. Nature 453, 396–400 (2008).
7. Stella, C. et al. Susceptibility of human metabolic phenotypes to dietary modulation. J. Proteome Res. 5, 2780–2788 (2006).
8. Loo, R. L., Zou, X., Appel, L. J., Nicholson, J. K. & Holmes, E. Characterization of metabolic responses to healthy diets and association with blood pressure: application to the Optimal Macronutrient Intake Trial for Heart Health (OmnitHeart), a randomized controlled study. Am. J. Clin. Nutr. 107, 323–338 (2014).
9. Lefèvre, M., Champagne, C. M., Tulley, R. T., Rood, J. C. & Most, M. M. Individual variability in cardiovascular disease risk factor responses to low-fat and low-saturated-fat diets in men: body mass index, adiposity, and insulin resistance predict changes in LDL cholesterol. Am. J. Clin. Nutr. 82, 1145–1146 (2005). 957;963; 976.
10. Kirwan, L. et al. Phenotypic factors influencing the variation in response of circulating cholesterol level to personalised dietary advice in the Food4Me study. Br. J. Nutr. 116, 2011–2019 (2016).
11. Katara, M. B., Beynen, A. C., de Vries, J. H. & Nobels, A. Existence of consistent hypo- and hyperresponders to dietary cholesterol in man. Am. J. Epidemiol. 123, 221–234 (1986).
12. Heinzmann, S. S. et al. Stability and robustness of human metabolic phenotypes in response to sequential food challenges. J. Proteome Res. 11, 643–655 (2012).
13. Griffin, N. W. et al. Prior dietary practices and connections to a human gut microbial metacommunity alter responses to diet interventions. Cell Host Microbe 21, 84–96 (2017).
14. de Toro-Martin, J., Arsenault, B. J., Despres, J. P. & Vohl, M. C. Precision nutrition: a review of personalized nutritional approaches for the prevention and management of metabolic syndrome. Nutrients 9, 913 (2017).
15. Wang, D. D. & Hu, F. B. Precision nutrition for prevention and management of type 2 diabetes. Lancet Diabetes Endocrinol. 6, 416–426 (2018).
16. Poslusna, K., Ruprich, J., de Vries, J. H., Jakubikova, M. & van ‘t Veer, P. Misreporting of energy and micronutrient intake estimated by food records and 24 hour recalls, control and adjustment methods in practice. Br. J. Nutr. 101 (Suppl. 2), S73–S85 (2009).
17. Celis-Morales, C. A. et al. Effect of personalized nutrition on health-related behaviour change: evidence from the Food4Me European randomized controlled trial. Int. J. Epidemiol. 46, 578–588 (2017).
18. Cheng, S. et al. Metabolite profiling identifies pathways associated with metabolic risk in humans. Circulation 125, 2222–2231 (2012).
19. Elliott, P. et al. Urinary metabolic signatures of human adiposity. Sci. Transl. Med. 7, 285ra6 (2015).
20. Nicholson, J. K. et al. Host–gut microbiota metabolic interactions. Science 336, 1262–1267 (2012).
21. García-Perez, I. et al. Objective assessment of dietary patterns by use of metabolic phenotyping: a randomised, controlled, crossover trial. *Lancet Diabetes Endocrinol.* 5, 184–195 (2017).

22. *Global Strategy on Diet, Physical Activity and Health* (WHO, 2004).

23. Krupp, D., Doberstein, N., Shi, L. & Remer, T. Hippuric acid in 24-hour urine collections is a potential biomarker for fruit and vegetable consumption in healthy children and adolescents. *J. Nutr.* 142, 1314–1320 (2012).

24. Cheung, W. et al. A metabolomic study of biomarkers of meat and fish intake. *Am. J. Clin. Nutr.* 105, 600–608 (2017).

25. Slupsky, C. M. et al. Investigations of the effects of gender, diurnal variation, and age in human urinary metabolomic profiles. *Anal. Chem.* 79, 6995–7004 (2007).

26. Rose, C., Parker, A., Jefferson, B. & Cartmell, E. The characterization of feces and urine: a review of the literature to inform advanced treatment technology. *Crit. Rev. Environ. Sci. Technol.* 45, 1827–1879 (2015).

27. Posma, J. M., Robinette, S. L., Holmes, E. & Nicholson, J. K. MetaboNetworks, an interactive Matlab-based toolbox for creating, customizing and exploring sub-networks from KEGG. *Bioinformatics* 30, 893–895 (2014).

28. Rubner, M. The energy value of the human diet. *Z. Biol.* 42, 261–305 (1901).

29. Cummings, J. H., Beatty, E. R., Kingman, S. M., Bingham, S. A. & Englyst, H. N. Variability of gastrointestinal transit in healthy women and men. *Gut* 39, 299–305 (1996).

30. Cummings, J. H., Beatty, E. R., Kingman, S. M., Bingham, S. A. & Englyst, H. N. Digestion and physiological properties of resistant starch in the human large bowel. *Br. J. Nutr.* 75, 733–747 (1996).

31. Jumpertz, R. et al. Energy-balance studies reveal associations between gut microbes, caloric load, and nutrient absorption in humans. *Am. J. Clin. Nutr.* 94, 58–65 (2011).

32. Gibbons, H. et al. Metabolic-based identification of clusters that reflect dietary patterns. *Mol. Nutr. Food Res.* 61, 1601050 (2017).

33. Posma, J. M. et al. Subset Optimization by Reference Matching (STORM): an optimized statistical approach for recovery of metabolic biomarker structural information from 1H NMR spectra of biofluids. *Anal. Chem.* 84, 10694–10701 (2012).

34. Posma, J. M. et al. Integrated analytical and statistical two-dimensional spectroscopy strategy for metabolite identification: application to dietary biomarkers. *Anal. Chem.* 89, 3300–3309 (2017).

35. Posma, J. M. et al. Optimized phenotypic biomarker discovery and confounder elimination via covariate-adjusted projection to latent structures from metabolic spectroscopy data. *J. Proteome Res.* 17, 1586–1595 (2018).

36. Manach, C., Williamson, G., Morand, C., Scalbert, A. & Remesy, C. Bioavailability and bioefficacy of polyphenols in humans. I. Review of 97 bioavailability studies. *Am. J. Clin. Nutr.* 81, 2305–2425 (2005).

37. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* 100, 9440–9445 (2003).

38. Eckburg, P. B. et al. Diversity of the human intestinal microbial flora. *Science* 308, 1635–1638 (2005).

39. Human Microbiome Project Consortium Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214 (2012).

40. Veselkov, K. A. et al. A metabolic entropy approach for measurements of systemic metabolic disruptions in patho-physiological states. *J. Proteome Res.* 9, 3537–3544 (2010).

41. García-Perez, I. et al. An analytical pipeline for quantitative characterization of dietary intake: application to assess grape intake. *J. Agric. Food Chem.* 64, 2423–2431 (2016).

Acknowledgements

This Article presents independent research funded by the UK National Institute for Health Research (NIHR) and Medical Research Council (MRC). The views expressed are those of the authors and not necessarily those of the UK National Health Service (NHS), the NIHR or the UK Department of Health. I.G.-P. is supported by a NIHR Career Development Research Fellowship (NIHR-CDF-2017-10-032). J.M.P. is supported by a Rutherford Fund Fellowship at Health Data Research (HDR) UK (MR/R004033/1). G.F. is supported by an NIHR Senior Investigator award. J.C.M., J.D., M.B., E.H. and G.F. are supported by an MRC grant entitled ‘Metabolomics for Monitoring Dietary Exposure’ (MR/J01308/1). E.H. is supported by the Department of Jobs, Tourism, Science and Innovation, Government of Western Australian through the Premier’s Science Fellowship Program. This study was supported by the NIHR/Wellcome Trust Imperial Clinical Research Facility, infrastructure support was provided by the NIHR Imperial Biomedical Research Centre (BRC) in line with the Gut Health research theme. The Section of Investigative Medicine is funded by grants from the MRC, UK Biotechnology and Biological Sciences Research Council, NIHR and an Integrative Mammalian Biology (IMB) Capacity Building Award. We also thank IKA Werke GmbH & Co for granting access to the C1-bomb calorimeter. We gratefully acknowledge the NutriTech consortium (European Union Framework 7 programme, grant agreement ID 289511) for sharing their data with us. The samples from the validation study are part of an ongoing clinical trial and we acknowledge the Egg Nutrition Center – American Egg Board for funding (to I.G.-P.), and J. Brignardello and J. McLean for conducting the clinical trial. We acknowledge funding from the Australian Medical Research Futures Fund for the Australian National Phenome Centre.

Author contributions

I.G.-P., J.M.P., J.K.N., E.H. and G.F. conceptualized the study and wrote the manuscript. I.G.-P., J.M.P. and E.S.C. analysed the data. I.G.-P., E.S.C. and G.F. ran the clinical trial. I.G.-P., I.C.M., J.D., E.H. and G.F. designed the clinical trial. All authors read and approved the final manuscript, and approved the final submitted version. G.F. assumes responsibility for the completeness and accuracy of the data and analyses, and for adherence to the study protocol.

Competing interests

J.D. has worked on the Cook to Health project (of which Groupe SEB is a collaborator and partly funded by EIT-Health) and the FACET project (of which Abbott, Spain, is a collaborator and partly funded by EIT-Health), both outside the submitted work. G.E. is lead for the Imperial Nestlé Collaboration and reports personal fees from Unilever, both outside the submitted work. All other authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s43016-020-0092-z. Supplementary information is available for this paper at https://doi.org/10.1038/s43016-020-0092-z.

Correspondence and requests for materials should be addressed to J.K.N., E.H. or G.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020.
Extended Data Fig. 1 | Diet information for each of the four dietary intervention provided to the 19 volunteers. Detailed discretion of the meal structure for each diet. The coloured sections show the nutritional breakdown and the non-coloured give details of the foods making up the diet at each meal.
Extended Data Fig. 2 | Dietary Metabotype Score predictions for each participant prior to admission to follow the four dietary interventions. Dietary Metabotype Score (DMS) is derived from the MCCV-PLS predictions (see Methods) for each participant’s baseline sample (before starting each of the four diets). Diet 1 (dark green) is the most concordant with WHO guidelines, followed by Diet 2 (light green) and Diet 3 (orange) and Diet 4 (red) is the least concordant. Confidence regions are highlighted in green and indicate the probability of the median DMS for each sample. This shows that the dietary habits in the home environment is highly variable between and within participants as evidenced by the initial DMS. Missing data: no fasting urine sample was available for participant 4 prior to starting Diet 4.
Extended Data Fig. 3 | See next page for caption.
Extended Data Fig. 3 | Urinary metabolic reaction network maps highlighting the expression of the dietary metabotype of individual participants in response to Diet 1 and Diet 4 ordered by participant number. A, participant 1 (entropy = 90.50), B, participant 2 (entropy = 90.04), C, participant 3 (entropy = 101.76), D, participant 4 (entropy = 95.49), E, participant 5 (entropy = 96.55), F, participant 7 (entropy = 106.37), G, participant 8 (entropy = 84.46), H, participant 9 (entropy = 81.69). The size of coloured nodes is proportional to the fold-change of that metabolite for that person. The colours of the edges are related to the sum of fold-changes of each pair of metabolites connected via their shortest path (fewest number of chemical reactions – see Methods). The more disordered (high entropy), or metabolically flexible, a subnetwork is, the stronger the associated edge weights (yellow-orange) in contrast to less perturbed pathways (magenta-blue edge weights). Entropy is expressed in arbitrary units.
Extended Data Fig. 4 | See next page for caption.
Extended Data Fig. 4 | Urinary metabolic reaction network maps highlighting the expression of the dietary metabotype of individual participants in response to Diet 1 and Diet 4 ordered by participant number. A, participant 10 (entropy = 81.92), B, participant 11 (entropy = 95.31), C, participant 12 (entropy = 80.13), D, participant 13 (entropy = 84.06), E, participant 14 (entropy = 78.81), F, participant 15 (entropy = 106.53), G, participant 17 (entropy = 76.12), H, participant 18 (entropy = 95.74), I, participant 19 (entropy = 92.33). The size of coloured nodes is proportional to the fold-change of that metabolite for that person. The colours of the edges are related to the sum of fold-changes of each pair of metabolites connected via their shortest path (fewest number of chemical reactions – see Methods). The more disordered (high entropy), or metabolically flexible, a subnetwork is, the stronger the associated edge weights (yellow-orange) in contrast to less perturbed pathways (magenta-blue edge weights). Entropy is expressed in arbitrary units.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- n/a
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted. Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

Topsin 3.5 software with Icon NMR (Bruker Biospin Ltd.)

Data analysis

CODE AVAILABILITY

CA-PLS (and PLS, OSC-PLS). The code for executing the PLS, covariate-adjusted (O)PLS and simple orthogonal PLS/PLS-DA is provided in https://bitbucket.org/jmp111/caplts/src/. This can be executed in a Matlab environment.

STORM (and STOCSY): The code for executing both the STOCSY and STORM algorithms is in https://bitbucket.org/jmp111/storm/src. These can be executed in a Matlab environment.

RED-STORM - The code for executing the RED-STORM algorithm is provided in https://bitbucket.org/jmp111/redstorm/src/. This can be executed in a Matlab environment.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

DATA AVAILABILITY. All presented data are tabulated and detailed in the main text and supporting information. Codes used to analyse these data referenced in the main text and Methods section. Study protocol availability is detailed in the Methods section. Diets provided to participants are detailed in supporting information. Quantified NMR data, DMS, AUC glucose and calorific value for Diets 1 and 4 presented in here are freely available (CC BY-NC 3.0) from Mendeley Data at http://
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☑ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

The main data comes from a randomized controlled clinical trial, where the sample size was determined based on expected increase of urinary proline betaine (as marker of nutritional intake). A previous study suggested that urinary concentration of this metabolite would rise by 50 μmol/L with each incremental rise in fruit intake (i.e., pieces of fruit) in the experimental setting. With an SD of 40 μmol/L, assuming a power of 0·95 and an alpha of 0·05 to detect a difference of 50 μmol/L, we estimated that we would need 12 volunteers. Because the protocol required a high amount of volunteer time and involvement (12 inpatient days plus travelling time) and volunteers could withdraw from the study, we requested permission to recruit 30 people, with the aim of having a cohort of roughly 20 people. 19 participants completed the entire controlled clinical trial.

Data exclusions

No data was excluded from the data presented.

Replication

The data presented was validated in another population (UK) from an external cohort (NutriTech 27). This dataset contained 65 participants (31 female), with a mean BMI of 29.0±2.9 kg/m² at baseline, for which urine samples were available from baseline measurements prior to starting the clinical trial.

The data shown in figure 3 was first discovered using a food challenge study (n=3 with 4 time points) and here shown with the data from a controlled clinical trial used to validate the findings.

Randomization

Data is from a randomized controlled clinical trial where participants were given 4 diets in random order.

Blinding

Investigators responsible for data analysis could not be blinded to assigned groups for the data presented as the multivariate statistical analysis was performed in a supervised manner. However, these investigators were not responsible for assigning a randomization order for participants, this was done by independent investigators.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a Involved in the study
☑ Antibodies
☑ Eukaryotic cell lines
☑ Palaeontology
☑ Animals and other organisms
☑ Palaeontology
☑ Human research participants
☑ Clinical data

Methods

n/a Involved in the study
☑ ChIP-seq
☐ Flow cytometry
☑ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics

Data presented relates to healthy volunteers (aged 21-65 years, BMI 20–35 kg/m²) recruited from a database of a clinical research unit in the UK. These volunteers had previously been screened at the clinical research unit and had expressed an interest in being contacted regarding future research studies. Volunteers in the database who were contacted with a letter of invitation, and those who responded with an interest in participation were screened, initially by email or telephone and...
subsequently at the unit. Potential participants were excluded if they had clinically significant illnesses, if they reported weight loss or gain of 3 kg or more in the preceding 2 months, if they were taking prescription medication, if they were current smokers or substance abusers, or if they presented any abnormalities on physical examination, electrocardiography, or screening blood tests. Women were ineligible if they were pregnant or breastfeeding.

Recruitment

We recruited participants from a database of healthy volunteers at the UK National Institute for Health Research (NIHR)/Wellcome Trust Imperial Clinical Research Facility (CRF). These volunteers had previously been screened at the CRF and had expressed an interest in being contacted regarding future research studies.

Ethics oversight

Data reported is from a study approved by the London-Brent Research Ethics Committee (reference number: 13/LO/0078) and all participants provided written informed consent.

Data used for validation is from a study approved by the London Brent Ethics Committee (reference number: 12/LO/0139) and all participants provided informed consent.

Note that full information on the approval of the study protocol must also be provided in the manuscript.