

# Development of a measure of dietary quality for the UK Biobank

Chloe Montague<sup>1,2</sup>, Stefania D'Angelo<sup>2</sup>, Nicholas Harvey<sup>2,3</sup>, Christina Vogel<sup>2,4,†</sup>, Janis Baird<sup>2,3,4,†</sup>

<sup>1</sup>Health Education England Wessex, Winchester SO21 2RU, UK

<sup>2</sup>MRC Lifecourse Epidemiology Centre, Southampton General Hospital, University of Southampton, Southampton SO16 6YD, UK

<sup>3</sup>NIHR Southampton Biomedical Research Centre, University of Southampton and University Hospital Southampton NHS Foundation Trust, Southampton SO16 6YD, UK

<sup>4</sup>NIHR Applied Research Collaboration Wessex, Southampton Science Park, Innovation Centre, 2 Venture Road, Chilworth, Southampton SO16 7NP, UK

Address correspondence to Chloe Montague, E-mail: [C.Montague@soton.ac.uk](mailto:C.Montague@soton.ac.uk)

†joint last authors

## ABSTRACT

**Background** Previous studies of the UK Biobank have examined intake of single food items and their association with health outcomes. Our aim was to develop a dietary quality score and examine the relationship between this score and markers of cardiometabolic health.

**Methods** Principal component analysis was performed on dietary data from UK Biobank participants. Linear regression was used to analyse the relationship between diet and cardiometabolic health.

**Results** The first component explained 14% of the variation in the dietary data. It was characterised by high consumption of meat and low fibre carbohydrates, and a low intake of fruit and vegetables. A higher score, indicative of healthier diet, was associated with lower systolic and diastolic blood pressure ( $\beta$   $-0.81$ , 95% CI  $-1.0$ ,  $-0.62$ ;  $\beta$   $-0.61$ , 95% CI  $-0.72$ ,  $-0.5$ ) and a healthier lipid profile (lower levels of cholesterol  $\beta$   $-0.05$ , 95% CI  $-0.06$ ,  $-0.04$ , triglycerides  $\beta$   $-0.05$ , 95% CI  $-0.06$ ,  $-0.03$ , and higher HDL cholesterol  $\beta$   $0.01$ , 95% CI  $0$ ,  $0.01$ ).

**Conclusions** The dietary quality score was a good approximation of overall dietary quality. An unhealthy diet was associated with markers of poorer cardiometabolic health.

**Keywords** food and nutrition, circulatory disease, dietary pattern

## Introduction

Poor diet, defined as one that is low in fruits, vegetables, whole grains, nuts and seeds, and high in sodium, sugar and trans-saturated fat, is the single biggest contributor to poor health globally.<sup>1,2</sup> Poor diet is a risk factor for many non-communicable diseases, in particular, cardiovascular disease, type 2 diabetes and many cancers, and often clusters with other behaviours such as smoking and physical inactivity.<sup>2,3</sup>

When examining population diet and health outcomes, there is increasing emphasis on the assessment of diet as a whole, or dietary patterns, rather than single nutrients or food groups. Measurement of dietary patterns allows for the fact that foods are eaten in combination and the balance of nutrients in a person's diet is an important determinant of disease.<sup>4</sup> Techniques for measuring dietary intake include food diaries, dietary recalls and food frequency questionnaires (FFQ).<sup>5–7</sup> Weighted food diaries are the gold-standard of

dietary assessment but they are time consuming and costly.<sup>6,7</sup> Dietary recall methods, such as FFQs, offer a less resource intensive alternative approach that is designed to capture both habitual diet and irregular consumption.<sup>8</sup> FFQs comprise a list of food and/or drink items and the participant is asked to state how often they consume each item over a period of time.<sup>5,8</sup> They can be used quickly and easily in studies with large numbers of participants.<sup>8</sup>

One of the benefits of analysing dietary data using dietary patterns is that it can provide a better understanding of overall dietary quality and, in turn, reveal stronger associations

Chloe Montague, Public Health Specialty Registrar

Stefania D'Angelo, Statistician

Nicholas Harvey, Professor

Christina Vogel, Professor

Janis Baird, Professor

with health outcomes.<sup>9–11</sup> Two approaches have been used for dietary patterns analysis. The first of these (an a priori approach) compares dietary intake to an existing framework, to assess adherence to dietary recommendations or a specific type of diet. The second (a posteriori) approach is data-driven, using dietary data to identify common patterns of consumption.<sup>12</sup> Principal component analysis (PCA) is one such approach and can identify key foods, both healthy and unhealthy, that contribute most strongly to a dietary pattern.

Several studies have investigated the feasibility and usefulness of data-driven approaches to measure dietary patterns in middle- to older-aged adults.<sup>13–15</sup> These studies, for the most part, demonstrated a positive association between a diet high in fruits and vegetables and/or a negative association between a diet high in meat and processed foods with body mass index (BMI) and/or waist circumference.<sup>13–15</sup>

The UK Biobank is a large prospective cohort study, incorporating a wide range of health, lifestyle and genetic data from over half a million participants with links to their ongoing health records.<sup>16</sup> It provides a unique opportunity to examine the determinants of health and investigate a range of diseases and health outcomes in a group of middle- to older-aged adults. The quantity of dietary data available within the UK Biobank is extensive. Previous studies have often employed a single food item or food group as an approximation of dietary quality.<sup>17–21</sup> Five studies have examined red meat and processed meat as a determinant of disease.<sup>17–21</sup> Collectively, results showed that higher levels of red and processed meat intake were associated with a higher risk of ischaemic heart disease, ischaemic stroke, pneumonia, diverticular disease, colon polyps, diabetes and colorectal cancer.<sup>17–19</sup> Applying a measure of dietary quality in such investigations could further strengthen associations between diet and health outcomes by better representing this multidimensional behaviour.

Harnessing FFQ data collected in the UK Biobank to describe dietary patterns may offer a more nuanced approach to analysing diet-disease relationships that complements approaches using single food groups. The first aim of the present study was to apply PCA to UK Biobank FFQ data to identify a dietary pattern that could describe overall dietary quality within the study. The secondary aim was to evaluate this new dietary quality score by assessing its relationship with objective markers of cardiometabolic health and other modifiable determinants of disease. The link between dietary quality and cardiovascular disease is well established.<sup>22</sup> It was therefore hypothesised that an unhealthy diet score would be associated both with markers of cardiometabolic health and other determinants of disease, such as smoking and physical inactivity.

## Methods

### Participants

This study used data from the UK Biobank.<sup>23</sup> Participants were invited by letter to take part in the UK Biobank between 2006 and 2010 if they were registered with the NHS, aged between 40 and 69 years, and living within a reasonable travel distance of one of the assessment centres.<sup>23</sup> More than 500 000 men and women have been recruited from England, Scotland and Wales.<sup>23</sup> The Greater London Foodscapes sample was comprised of 52 345 UK Biobank participants who were recruited to one of the three London centres.<sup>24</sup> Of these 52 345 participants, 33 149 provided complete dietary data and were therefore included in this analysis.

### Procedure

At the baseline assessment, UK Biobank participants completed a touchscreen questionnaire, had their physical measurements taken, and provided blood, urine and saliva samples.<sup>23</sup>

### Demographic data

Demographic information including sex, age, ethnicity, level of education and Index of Multiple Deprivation (IMD) was obtained at the baseline assessment.<sup>24</sup>

### Health behaviour data

Physical activity and smoking status data were collected at the baseline assessment. Participants were asked about the type and duration of physical activity and their current smoking status.<sup>25</sup>

### Dietary data

During the baseline assessment, participants completed an FFQ detailing their usual frequency of intake of 17 food and drink items.<sup>26</sup> They reported how many times per day they consumed fruits (fresh and dried), vegetables (cooked, salad and raw), water, tea and coffee.<sup>26</sup> Cereal and bread items were reported by type and frequency of consumption per week.<sup>26</sup> For all other items (oily and non-oily fish, processed meat, poultry, beef, lamb/mutton, pork, cheese), participants reported their usual intake over the past year.<sup>26</sup> Two further food items were derived from the data for the cereal and bread categories. Participants were allocated as either high fibre cereal consumers (e.g. bran, biscuit cereal, muesli or oat) or low fibre cereal consumers (e.g. cornflakes or Frosties). Similarly, brown/wholemeal bread and white bread were derived from the 'bread type' variable. This produced 19 food and drink items in total.

## Outcome measures

The outcomes of interest were physical and biochemical factors associated with a greater risk of cardiometabolic disease. Blood pressure and BMI were selected as physical measures of risk for cardiometabolic ill-health. Key biochemical markers relating to cardiometabolic health were selected to represent the biochemical outcomes of interest for this analysis, including: cholesterol, high-density lipoprotein (HDL) cholesterol, apolipoprotein A, apolipoprotein B, lipoprotein A, triglycerides, glycated haemoglobin (HbA1c), C-reactive Protein (CRP) and vitamin D.

## Statistical analysis

The baseline characteristics of the 33 149 participants included in this study were compared to the full Greater London foodscapes sample ( $n = 52\ 345$ ) and to the participants with incomplete dietary data ( $n = 19\ 196$ ) using either a *t*-test, a Mann–Whitney test or a chi-square test, depending on the nature of the variable.

To address the first research aim, PCA was performed on the weekly frequencies of consumption of the 19 food and drink items. PCA is a data reduction method that produces new variables that are linear combinations of the original dietary variables and maximises the explained variance within the data.<sup>27</sup> The PCA analysis was based on the correlation matrix in order to adjust for unequal variances of the original variables. The first component of the PCA, the main axis of variation in the dietary data, was used to create a summary dietary quality score for each participant. These scores were created by multiplying the PCA coefficients for each of the food and drink items by each participant's frequency of consumption, and then summing this value across all 19 items. The score was then transformed using Fisher-Yates normal scores to create a normal distribution across the sample, with a mean of 0 and a standard deviation of 1. The median of the score was used to categorise diet as healthier or less healthy. For ease of interpretation, the coefficients were multiplied by  $-1$ , so a positive score (above the median) indicated a healthier diet. The distributions of outcomes were checked by visual inspection and those that were not normally distributed were transformed using Fisher-Yates normal scores. Further details on how the PCA was performed can be found in [Table 1](#) in the Supplementary Material.

Linear regression was used to examine the association between dietary quality and markers of cardiometabolic health with estimates presented in their unadjusted form and following adjustment for pre-defined confounding factors. Potential confounding variables were pre-defined using a Directed Acyclic Graph (DAG), with the software DAGitty.<sup>28</sup>

According to the DAG (see [Figure 1](#), Supplementary Material), associations needed to be adjusted for sex, age, ethnicity, level of education, smoking and IMD.

Finally, clustering of unhealthy behaviours was analysed by cross-tabulations and chi-squared tests and included measures of smoking status, physical activity and diet quality. All analyses were performed using the statistical software Stata v15.1.

## Results

The baseline demographic characteristics of the Greater London foodscapes sample, the sample included in the analyses and participants with incomplete dietary data (who were excluded from the analysis) are shown in [Table 1](#). For our sample, under half (45%) were male and the average age was 56.2 years (SD = 8.2). The vast majority (83.8%) of the sample was of white ethnicity and almost half had a college or university degree (49.5%). Over half of participants (51.9%) were physically active and 9.7% were current smokers. The average BMI was 26.8 kg/m<sup>2</sup> (SD = 4.7). Compared with participants with incomplete dietary data, the sample included in the analyses featured a higher proportion of participants who were: male; older; of White ethnicity; physically active; non-smokers; educated to a tertiary level; living in less deprived areas; and who had a lower BMI. These differences were all statistically significant.

The second part of [Table 1](#) summarises participants' cardiometabolic health markers. Systolic and diastolic blood pressures were within the normal range.<sup>29</sup> Total cholesterol was slightly elevated although HDL cholesterol and triglycerides were both within the normal ranges.<sup>30</sup> Levels of apolipoproteins A and B and lipoprotein A in the sample were within the normal ranges.<sup>31,32</sup> HbA1c, CRP and vitamin D levels were all also within the normal ranges. Participants' mean BMI fell within the overweight range. There were some significant differences between the sample included in the analysis and the participants with incomplete dietary data. Participants with incomplete dietary data had a significantly higher average blood pressure, lipoprotein A levels, HbA1c, CRP and BMI compared with the sample analysed. Triglycerides and vitamin D were significantly lower in participants with incomplete dietary data compared with the sample.

After PCA was performed, the first component explained 14% of the variation in the dietary data. This component was characterised by a high consumption of beef, any type of meat, white bread and low-fibre cereals, and low intakes of fresh/dried fruit, salad and cooked vegetables and was defined as an 'unhealthy diet score' (see [Table 2](#), Supplementary Material).

**Table 1** Baseline characteristics of the sample populations

<i>Demographic and lifestyle factors</i>				
	<i>Greater London foodscapes sample</i>	<i>Sample included in the analysis</i>	<i>Participants with incomplete dietary data</i>	<i>P-value<sup>a</sup></i>
<i>n</i>	<i>n</i> = 52 345	<i>n</i> = 33 149	<i>n</i> = 19 196	
Sex, men (%)	44.0%	45.0%	42.3%	<0.001
Age, mean (SD)	56.0 (8.2)	56.2 (8.2)	55.6 (8.3)	<0.001
White ethnicity, (%)	80.0%	83.8%	73.2%	<0.001
More than 1 day/week vigorous activity, (%)	50.4%	51.9%	47.6%	<0.001
Current smokers, (%)	11.8%	9.7%	15.6%	<0.001
Level of education, (%)				
College or University degree	47.8%	49.5%	44.7%	<0.001
Other prof qual/HND/A levels	26.8%	26.8%	26.7%	
GCSE or less	25.4%	23.7%	28.6%	
Index Multiple Deprivation, median (IQR)	18.1 (10.0, 27.4)	16.5 (9.4, 26.3)	20.5 (12.0, 29.6)	<0.001
<b>Markers of cardiometabolic health</b>				
<i>n</i>	52 345	33 149	19 196	
Systolic blood pressure, mmHg mean (SD)	134.4 (18.1)	134.2 (17.9)	134.7 (18.5)	0.002
Diastolic blood pressure, mmHg mean (SD)	81.0 (10.1)	80.8 (10.0)	81.5 (10.3)	<0.001
Cholesterol, mmol/L mean (SD)	5.6 (1.1)	5.6 (1.1)	5.6 (1.1)	0.14
HDL cholesterol, mmol/L mean (SD)	1.5 (0.4)	1.5 (0.4)	1.5 (0.4)	0.16
Apolipoprotein A, g/L mean (SD)	1.6 (0.3)	1.6 (0.3)	1.6 (0.3)	0.31
Apolipoprotein B, g/L mean (SD)	1.0 (0.2)	1.0 (0.2)	1.0 (0.2)	0.83
Lipoprotein A, nmol/L median (IQR)	23.8 (10.2, 66.6)	22.9 (10.0, 64.1)	25.4 (10.4, 69.8)	<0.001
Triglycerides, mmol/L median (IQR)	1.4 (1.0, 2.0)	1.4 (1.0, 2.0)	1.3 (1.0, 2.0)	0.01
Glycated haemoglobin, mmol/mol median (IQR)	35.4 (32.8, 38.2)	35.3 (32.7, 38.1)	35.5 (33.0, 38.6)	<0.001
C-reactive protein, mg/L median (IQR)	1.2 (0.6, 2.5)	1.1 (0.6, 2.4)	1.2 (0.6, 2.7)	<0.001
Vitamin D, nmol/L median (IQR)	42.9 (29.4, 59.1)	44.4 (30.9, 60.1)	40.2 (27.1, 56.8)	<0.001
BMI, kg/m <sup>2</sup> mean(SD)	26.9 (4.9)	26.8 (4.7)	27.1 (5.1)	<0.001

<sup>a</sup>P-value comparing participants included and excluded from the analysis and estimated by *t*-test, Mann–Whitney ranksum test, or chi-squared test

We found an inverse association between diet quality and blood pressure: a higher dietary quality score, indicative of healthier diet, was associated with lower blood pressure (Table 2). Healthier dietary quality scores were also associated with a healthier lipid profile (lower levels of cholesterol, apolipoprotein B, triglycerides and higher HDL cholesterol), and a lower HbA1c, CRP and BMI. These associations remained statistically significant following adjustment for sex, age, ethnicity, level of education, smoking and IMD. For vitamin D, the unadjusted results identified an inverse relationship between dietary quality and vitamin D levels. After adjusting for confounders, however, this relationship showed that healthier dietary quality scores were associated with higher vitamin D levels.

The proportion of current smokers was higher among participants with an unhealthy diet score (see Table 3,

Supplementary Material). The proportion of physically active participants was higher among those with a healthy diet score (see Table 3, Supplementary Material). Both findings were statistically significant ( $P < .001$ ).

## Discussion

### Main findings of this study

This study identified that a dietary quality pattern can be created from the FFQ data in the UK Biobank. The first component that resulted when PCA was performed explained 14% of the variation in diet in the sample, and showed a dietary pattern with high intakes of meat, including processed meat, white bread and low intakes of high-fibre cereals, fruit, salad and vegetables. This dietary pattern was termed ‘unhealthy’ because it was not consistent with the

**Table 2** Association between diet quality score and markers of cardiometabolic health

	<i>Unadjusted</i>			<i>Adjusted<sup>a</sup></i>		
	<i>n</i>	<i>β</i> (95% <i>CI</i> )	<i>p</i>	<i>n</i>	<i>β</i> (95% <i>CI</i> )	<i>p</i>
Systolic blood pressure, mmHg	33 083	-1.11 (-1.31, -0.92)	<0.001	31 822	-0.81(-1.00, -0.62)	<0.001
Diastolic blood pressure, mmHg	33 083	-0.78 (-0.89, -0.67)	<0.001	31 822	-0.61(-0.72, -0.50)	<0.001
Cholesterol, mmol/L	30 967	-0.02 (-0.03, -0.01)	0.001	29 802	-0.05(-0.06, -0.04)	<0.001
HDL cholesterol, mmol/L	28 507	0.04 (0.03, 0.04)	<0.001	27 445	0.01(0.00, 0.01)	<0.001
Apolipoprotein A, g/L	28 328	0.02 (0.02, 0.02)	<0.001	27 275	-0.00(-0.00, 0.00)	0.318
Apolipoprotein B, g/L	30 821	-0.01 (-0.02, -0.01)	<0.001	29 657	-0.01(-0.02, -0.01)	<0.001
Lipoprotein A, SD	25 021	0.02 (0.01, 0.03)	0.001	27 072	-0.01(-0.02, 0.00)	0.147
Triglycerides, SD	30 946	-0.09(-0.11, -0.08)	<0.001	29 783	-0.05(-0.06, -0.03)	<0.001
Glycated haemoglobin (HbA1c), SD	30 338	-0.01 (-0.02-0.00)	0.240	29 194	-0.04(-0.05, -0.03)	<0.001
C-reactive protein, SD	30 907	-0.10 (-0.12, -0.09)	<0.001	29 744	-0.12(-0.13, -0.10)	<0.001
Vitamin D, SD	29 810	-0.02 (-0.03, -0.01)	<0.001	28 691	0.02(0.00, 0.03)	0.006
BMI, kg/m <sup>2</sup>	32 862	-0.68(-0.73, -0.63)	<0.001	31 612	-0.67(-0.73, -0.62)	<0.001

<sup>a</sup>Adjusted for sex, age, ethnicity, level of education, smoking, and IMD

recommendations within the UK Eatwell Guide.<sup>33</sup> The dietary coefficients derived from the PCA for each of the 19 food and drink items were consistent with advice contained within the Eatwell Guide.<sup>33</sup>

The dietary quality score demonstrated that a healthier diet was associated with a healthier cardiometabolic profile after adjusting for confounding variables known to influence the relationship between diet and health.<sup>34</sup> Furthermore, we found that an unhealthy diet clustered with other unhealthy lifestyle factors, namely smoking and physical inactivity. Collectively, these findings suggest that our dietary quality score is a valid measure of overall dietary quality within the UK Biobank.

### What is already known on this topic

Our measure of dietary quality for the UK Biobank compares favourably with similar analyses performed amongst older adult populations in high-income countries. Thorpe et al. used PCA to identify dietary patterns amongst Australian older adults.<sup>35</sup> The dietary patterns identified explained between 5.6 and 7.8% of variation in dietary intake.<sup>35</sup> A poorer quality diet, typified by high consumption of red and processed meat and refined grains was associated with a higher BMI, smoking and lower levels of physical activity.<sup>35</sup> Robinson et al. investigated dietary quality across the adult life course.<sup>36</sup> The dietary patterns they observed explained 7% of the variation in dietary data.<sup>36</sup> A high-quality diet was associated with a high intake of fruits, vegetables, high-fibre carbohydrates and yoghurt, and a low intake of processed meat, potato products and added sugar.<sup>36</sup> This was similar to the dietary pattern we observed.<sup>36</sup> More recently, Gao et al. have analysed dietary

patterns in the UK Biobank.<sup>37</sup> The researchers used the 24-hour online dietary assessment data, in contrast to the FFQ data that we have used here.<sup>37</sup> They described two main dietary patterns, explaining 43 and 20% of the variance in diet.<sup>37</sup> These three studies all demonstrated an association between a poorer quality diet—typified by high levels of meat, processed meat and low-fibre carbohydrates—and markers of cardiometabolic disease.<sup>35,36,37</sup>

### What this study adds

A strength of this study is the large sample size afforded by the UK Biobank database. The breadth of data collected in the UK Biobank permits adjustment for a wide range of confounding variables and the use of a variety of measures of cardiometabolic health. Our study used FFQ dietary data from UK Biobank because it provides a better representation of habitual diet than 24-hour food diary data, and includes frequency data for food and drink items—improving comparisons between participants. There is the potential for error to be introduced during the application of FFQs due to recall or response bias, however, this does not prevent these data being used to rank participants intake.<sup>42</sup> By illustration, Bradbury et al. reported high reproducibility of the FFQ administered to UK Biobank participants and found the FFQ showed greater reliability at ranking participants dietary intake than the 24-hour food diary data.<sup>42</sup>

### Limitations of this study

The UK Biobank sample is not representative of the wider UK population.<sup>38</sup> Of the 9.2 million people who were invited

to participate, 5.5% responded.<sup>38</sup> Fry et al. looked at the differences between responders and non-responders.<sup>38</sup> Responders were more likely to be female, were generally healthier and less likely to live in deprived areas. Comparing our study sample directly with the wider UK population, our sample was more ethnically diverse, more physically inactive, but contained fewer smokers.<sup>39,40</sup> The BMI of our sample was slightly lower than the mean of participants within the Health Survey for England 2019.<sup>41</sup> Overall, these comparisons present a mixed picture of how our sample compared to the wider UK population. It is likely less deprived than the whole UK population but there is not a consistent pattern in relation to demographic factors associated with poorer health outcomes. This suggests that the use of this population to create the measure of dietary quality provides reasonable general applicability to the whole UK population. Indeed, Fry et al. acknowledge that although the UK Biobank sample is not representative of the wider population, the associations observed between health behaviours and health outcomes remain useful as a reflection of patterns of disease in the adult UK population.<sup>38</sup>

Another limitation of this study relates to the selection of food and drink items for the FFQ. Frequently consumed foods such as biscuits, cakes, confectionary and crisps were not included nor was the list of healthier foods exhaustive (e.g. yoghurt, beans, pulses, eggs were not included). However, studies using data driven methods, such as PCA, have shown that dietary quality scores derived from short FFQs (~20 items) correlate strongly in the expected direction with dietary scores created from large FFQs (~100 items) and with objective biomarkers of nutritional status.<sup>42–45</sup>

## Conclusions

The dietary coefficients derived from the PCA were consistent with the Eatwell guide showing the value this dietary pattern offers to future studies using the UK Biobank. The dietary quality score developed in this study provides a useful tool for researchers wishing to incorporate diet into their investigations when using data from the UK Biobank cohort. The dietary quality score developed in this study can be easily applied to assess dietary quality as an exposure, outcome, mediating, or confounding variable in UK Biobank analyses. The ability to apply and compare this dietary quality score with other measures of morbidity and mortality within the UK Biobank dataset could be beneficial in the understanding of the relationship between diet and disease.

The findings of this large population-based study suggest that a dietary pattern with high levels of meat consumption (including processed meats, red meats and poultry) is asso-

ciated with a greater risk of poor cardiometabolic health in adults. Such findings are important for government strategy for prevention of cardiometabolic disease.

## Acknowledgments

None.

## Supplementary data

Supplementary data are available at the *Journal of Public Health* online.

## Funding

None.

## Conflicts of interest statement

CM has no conflicts of interest. SD has no conflicts of interest. NCH has no conflicts of interest. CV has a non-financial research collaboration with a UK supermarket chain. JB has a non-financial research collaboration with a UK supermarket chain and has, in the past, received grant research support from Danone Nutricia Early Life Nutrition.

## Authors' contributions

C.M. wrote the paper and had primary responsibility for the final content. S.D. performed the statistical analysis. J.B. and C.V. designed the research and contributed to the writing of the paper. All authors read and approved the final manuscript. This work was undertaken using the UK Biobank resource, access application number 17139.

## References

1. Forouzanfar M, Alexander L, Ross Anderson H *et al.* Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015;**386**:2287–323. [https://doi.org/10.1016/S0140-6736\(15\)00128-2](https://doi.org/10.1016/S0140-6736(15)00128-2).
2. Afshin A, Sur PJ, Fay KA *et al.* Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2019;**393**:1958–72. [https://doi.org/10.1016/S0140-6736\(19\)30041-8](https://doi.org/10.1016/S0140-6736(19)30041-8).
3. British Health Foundation. *Heart Statistics: BHF Statistics Factsheet – UK*. London, UK: BHF. <https://www.bhf.org.uk/what-we-do/our-research/heart-statistics> (27 April 2021, date last accessed).

4. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol* 2002;**13**:3–9. <https://doi.org/10.1097/00041433-200202000-00002>.
5. Fowles E, Sterling B, Walker L. Measuring dietary intake in nursing research. *Can J Nurs Res* 2007;**39**(2):146–65.
6. Closer. Dietary Assessment Tools (DATs).. <https://www.closer.ac.uk/cross-study-data-guides/dietary-data-guide/dietary-assessment-tools/> (13 January 2022, date last accessed).
7. Rutishauser I. Dietary intake measurements. *Public Health Nutr* 2005;**8**(7a):1100–7. <https://doi.org/10.1079/phn2005798>.
8. Cade JE, Warthon-Medina M, Albar S *et al*. DIET@NET: best practice guidelines for dietary assessment in health research. *BMC Med* 2017;**15**:202. [doi.org/10.1186/s12916-017-0962-x](https://doi.org/10.1186/s12916-017-0962-x).
9. Milte CM, McNaughton SA. Dietary patterns and successful ageing: a systematic review. *Eur J Nutr* 2016;**55**(2):423–50. <https://doi.org/10.1007/s00394-015-1123-7>.
10. Kourlaba G, Panagiotakos D. Dietary quality indices and human health: a review. *Maturitas* 2009;**62**(1):1–8. <https://doi.org/10.1016/j.maturitas.2008.11.021>.
11. Hodge A, Bassett J. What can we learn from dietary pattern analysis? *Public Health Nutr* 2016;**19**:191–4. <https://doi.org/10.1017/S1368980015003730>.
12. Schulze MB, Hoffmann K. Methodological approaches to study dietary patterns in relation to risk of coronary heart disease and stroke. *Br J Nutr* 2006;**95**(5):860–9. <https://doi.org/10.1079/bjn20061731>.
13. McNaughton S, Mishra G, Stephen A *et al*. Dietary patterns throughout adult life are associated with body mass index, waist circumference, blood pressure, and red cell folate. *J Nutr* 2007;**137**(1):99–105. <https://doi.org/10.1093/jn/137.1.99>.
14. Kesse-Guyot E, Bertrais S, Péneau S *et al*. Dietary patterns and their sociodemographic and behavioural correlates in French middle-aged adults from the SU.VI.MAX cohort. *Eur J Clin Nutr* 2009;**63**:521–8. <https://doi.org/10.1038/sj.ejcn.1602978>.
15. Muga M, Owili P, Hsu C *et al*. Dietary patterns, gender, and weight status among middle-aged and older adults in Taiwan: a cross-sectional study. *BMC Geriatr* 2017;**17**:268. <https://doi.org/10.1186/s12877-017-0664-4>.
16. UK Biobank. *Background*. <https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us> (13 January 2022, date last accessed).
17. Papier K, Fensom GK, Knuppel A *et al*. Meat consumption and risk of 25 common conditions: outcome-wide analyses in 475,000 men and women in the UK Biobank study. *BMC Med* 2021;**19**:53. [doi.org/10.1186/s12916-021-01922-9](https://doi.org/10.1186/s12916-021-01922-9).
18. Bradbury K, Murphy N, Key T. Diet and colorectal cancer in UK Biobank: a prospective study. *Int J Epidemiol* 2020;**49**(1):246–58. <https://doi.org/10.1093/ije/dyz064>.
19. Raisi-Estabragh Z, McCracken C, Gkontra P *et al*. Associations of meat and fish consumption with conventional and radiomics cardiovascular magnetic resonance phenotypes in the UK Biobank. *Front Cardiovasc Med* 2021;**8**:667849. <https://doi.org/10.3389/fcvm.2021.667849>. PMID: 34026874; PMCID: PMC8133433.
20. Anderson J, Darwis N, Mackay D *et al*. Red and processed meat consumption and breast cancer: UK Biobank cohort study and meta-analysis. *Eur J Cancer* 2018;**90**:73–82. <https://doi.org/10.1016/j.jca.2017.11.022>.
21. Zhang H, Cade J, Hadie L. Consumption of red meat is negatively associated with cognitive function: a cross-sectional analysis of UK Biobank. *Curr Dev Nutr* 2020;**4**(2):1510. [https://doi.org/10.1093/cdn/nzaa061\\_138](https://doi.org/10.1093/cdn/nzaa061_138).
22. Atkins J, Wannamethee S. Diet Quality and Cardiovascular Disease Prevention. In: Bendich A, Deckelbaum R (eds). *Preventive Nutrition. Nutrition and Health*. Cham: Springer, 2015.
23. UK Biobank. *Protocol for a large-scale prospective epidemiological resource*. <https://www.ukbiobank.ac.uk/media/gnkeyh2q/study-rationale.pdf> (13 January 2022, date last accessed).
24. UK Biobank. *Category 100094: Baseline characteristics – populations characteristics*. <https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=100094> (4 January 2022, date last accessed).
25. UK Biobank. *Category 100050: Lifestyle and environment*. <https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=100050> (5 October 2022, date last accessed).
26. UK Biobank. *UK Biobank Category 100053: Diet*. <https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=100053> (13 January 2022, date last accessed).
27. Jolliffe IT, Morgan BJ. Principal component analysis and exploratory factor analysis. *Stat Methods Med Res* 1992;**1**:69–95. <https://doi.org/10.1177/096228029200100105>.
28. Textor J, van der Zander B, Gilthorpe MS *et al*. Robust causal inference using directed acyclic graphs: the R package ‘dagitty’. *Int J Epidemiol* 2017;**45**(6):1887–94. <https://doi.org/10.1093/ije/dyw341>.
29. NICE. *Hypertension in adults: diagnosis and management*. <https://www.nice.org.uk/guidance/ng136/chapter/recommendations#treating-and-monitoring-hypertension> (13 January 2022, date last accessed).
30. NHS. *Cholesterol levels*. <https://www.nhs.uk/conditions/high-cholesterol/cholesterol-levels/> (13 January 2022, date last accessed).
31. Gloucestershire Hospitals NHS Foundation Trust. *Apolipoproteins (Apo-A1 and Apo-B)*. <https://www.gloshospitals.nhs.uk/our-services/services-we-offer/pathology/tests-and-investigations/apolipoproteins-apo-a1-and-apo-b/> (13 January 2022, date last accessed).
32. Heart UK. *Lipoprotein (a)*. <https://www.heartuk.org.uk/downloads/factsheets/lipoprotein-a.pdf> (13 January 2022, date last accessed).
33. Public Health England. *Eatwell Guide*. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/528193/Eatwell\\_guide\\_colour.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/528193/Eatwell_guide_colour.pdf) (13 January 2022, date last accessed).
34. Bleich SN, Jones-Smith J, Wolfson JA *et al*. The complex relationship between diet and health. *Health Aff* 2015;**34**:11. <https://doi.org/10.1377/hlthaff.2015.0606>.
35. Thorpe MG, Milte CM, Crawford D *et al*. A comparison of the dietary patterns derived by principal component analysis and cluster analysis in older Australians. *Int J Behav Nutr Phys Act* 2016;**13**:30. <https://doi.org/10.1186/s12966-016-0353-2>.
36. Robinson SM, Westbury LD, Cooper R *et al*. Adult lifetime diet quality and physical performance in older age: findings from a British birth cohort. *J Gerontol* 2018;**73**(11):1532–7. <https://doi.org/10.1093/gerona/glx179>.

37. Gao M, Jebb SA, Aveyard P *et al.* Associations between dietary patterns and the incidence of total and fatal cardiovascular disease and all-cause mortality in 116,806 individuals from the UK Biobank: a prospective cohort study. *BMC Med* 2021;**19**:83. <https://doi.org/10.1186/s12916-021-01958-x>.
38. Fry A, Littlejohns TJ, Sudlow C *et al.* Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol* 2017;**186**(9):1026–34. <https://doi.org/10.1093/aje/kwx246>.
39. Gov.uk. *Regional ethnic diversity*. <https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-populations/regional-ethnic-diversity/latest> (13 January 2022, date last accessed).
40. Public Health England. *Public Health Outcomes Framework*. <https://fingertips.phe.org.uk/profile/public-health-outcomes-framework/data#page/0/gid/1000042/pat/6/par/E12000008/ati/102/are/E10000014/cid/4/tbm/1> (13 January 2022, date last accessed).
41. NHS Digital *Weight*. <http://healthsurvey.hscic.gov.uk/data-visualisation/data-visualisation/explore-the-trends/weight.aspx?type=child> (13 January 2022, date last accessed).
42. Shaw S, Crozier S, Strömmer S *et al.* Development of a short food frequency questionnaire to assess diet quality in UK adolescents using the National Diet and Nutrition Survey. *Nutr J* 2021;**20**:5. <https://doi.org/10.1186/s12937-020-00658-1>.
43. Crozier S, Inskip H, Barker M *et al.* Development of a 20-item food frequency questionnaire to assess a 'prudent' dietary pattern among young women in Southampton. *Eur J Clin Nutr* 2010;**64**:99–104. <https://doi.org/10.1038/ejcn.2009.114>.
44. Jarman M, Fisk C, Ntani G *et al.* Assessing diets of 3-year-old children: evaluation of an FFQ. *Public Health Nutr* 2014;**17**(5):1069–77. <https://doi.org/10.1017/S136898001300102X>.
45. Robinson SM, Jameson KA, Bloom I *et al.* Development of a short questionnaire to assess diet quality among older community-dwelling adults. *J Nutr Health Aging* 2017;**21**:247–53. <https://doi.org/10.1007/s12603-016-0758-2>.