Associations between dietary patterns and incident type 2 diabetes: prospective cohort study of 120,343 UK Biobank participants

Authors: Min Gao [1], Susan A Jebb [1,2], Paul Aveyard [1,2], Gina L Ambrosini [3], Aurora Perez-Cornago [4], Keren Papier [4], Jennifer Carter [5], Carmen Piernas [1*]

[1]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

[2]NIHR Oxford Biomedical Research Centre, Oxford University Hospitals, Oxford, UK

[3]School of Population and Global Health, University of Western Australia, 35 Stirling Highway, Crawley, 6009, Perth, Western Australia

[4] Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, UK

[5] Nuffield Department of Population Health, University of Oxford, Oxford, UK

***Correspondence to Carmen Piernas**

Carmen Piernas, MSc PhD

T: +44 (0)1865 289284 E: carmen.piernas-sanchez@phc.ox.ac.uk

Nuffield Department of Primary Care Health Sciences, University of Oxford

Radcliffe Primary Care Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG

Funding: Data access costs were funded by a grant from the British Nutrition Foundation

Abstract

OBJECTIVE: To identify dietary patterns (DPs) characterized by a set of nutrients of concern and their association with incident type 2 diabetes (T2D).

RESEARCH DESIGN AND METHODS: 120,343 participants from the UK Biobank study with at least two 24-h dietary assessments were studied. Reduced rank regression was used to derive DPs explaining variability in energy density, free sugars, saturated fat and fiber intakes. We investigated prospective associations with T2D using Cox proportional hazard models.

RESULTS: 2,878 participants developed T2D over 8.4 years of follow-up from the latest dietary assessment. Two DPs were identified that jointly explained a total of 63% variation in four nutrients. DP1 was characterized by high intakes of chocolate/confectionery, butter, low-fiber bread, sugars/preserves and low intakes of fruit/vegetables. DP1 was linearly associated with T2D in multivariable models without BMI adjustment [hazard ratio(HR) per z-score 1.11, 95%CI (1.08-1.14)] and after BMI adjustment [1.09(1.06-1.12)]. A second DP2 was characterized by high intakes of sugar-sweetened beverages, fruit juice, table sugars/preserves, and low intakes of high fat cheese/butter, but showed no clear association with T2D. There were significant interactions between both DPs and age, with increased risks among younger people in DP1 [1.13(1.09-1.18)] and DP2 [1.10(1.05-1.15)]; as well as with DP1 and BMI, with increased risks among people with obesity [1.11(1.07-1.16)].

CONCLUSIONS: A dietary pattern characterized by high intakes of chocolate confectionery, butter, low-fiber bread and added sugars, and low in fresh fruit/vegetables intakes is associated with a higher incidence of T2D, particularly among younger people and those living with obesity.

Key Words: dietary pattern; Reduced rank regression; type 2 diabetes

Introduction

The prevalence of diabetes, predominantly type 2 diabetes, has increased from 463 million in 2019 and is projected to 700 million by 2045 worldwide (1). Type 2 diabetes is associated with higher morbidity and mortality from other non-communicable diseases such as cardiovascular disease (2). In addition, the COVID-19 pandemic has revealed that people with diabetes have at least twice the risk of severe COVID-19 outcomes (3).

The risk of Type 2 diabetes can be reduced by modifying unhealthy behaviors such as a poor diet (4). According to the 2019 Global Burden of Diseases, 29.7% of disability-adjusted life-years (DALYs) of diabetes are attributable to dietary factors (5). Previous research focused on nutrients have reported associations between poor diet quality and the incidence of type 2 diabetes and its co-morbidities, particularly diets high in saturated fat (6) and low in fiber (7). Evidence with regards to the role of free sugars is less consistent (8); although high intakes have been associated with increased risk of type 2 diabetes, it is unclear whether the detrimental effects are caused by free sugars *per se*, or by their contribution to excess energy intake. Increased free sugars intake is associated with reduced insulin sensitivity (9; 10), but it has been reported that a high intake of free sugars without excess energy may not have any detrimental impact on health (11). There is also some evidence for positive associations between single foods and an increased risk of developing type 2 diabetes, such as higher intakes of unprocessed and processed meat (12; 13), fruit juices, sugar sweetened beverages (SSBs), refined grains, sweets, and desserts (8; 14) and low intakes of fresh fruit and vegetables (15).

However, the impact of individual nutrients or foods on health outcomes may not adequately reflect the health effects of a whole diet where foods are eaten in combination. Therefore, the observed associations could be driven by one or more nutrients, foods or other aspects of the diet that co-occur simultaneously. Reduced rank regression (RRR) is a data-driven approach that derives dietary patterns by using *a priori* knowledge to select nutrients that are hypothesized to be on the causal pathway to examine how specific combinations of foods are associated with disease outcomes (16; 17). This method allows specific investigations into certain diet-disease pathways (18), as opposed to the more exploratory approaches of PCA or factor analysis (19). RRR has previously been used to study the effect of dietary patterns (DPs) on pathways of glucose homeostasis (20), inflammatory biomarkers (21), plasma circulating fatty acids (22) and blood lipids (23) as biomarkers of type 2 diabetes risk, although this evidence is not particularly strong due to smaller sample sizes and low explained variance for the intermediate biomarkers. Here we use detailed dietary data from the UK Biobank participants to identify DPs explaining high variability in known dietary risk factors, such as energy density, free sugars, saturated fat and low fiber intakes; and to assess the association between DPs and incident type 2 diabetes.

RESEARCH DESIGN AND METHODS

Study Population

UK Biobank is a population-based cohort that recruited 502,536 participants (aged 37 to 73 years) from 22 sites across England, Wales, and Scotland, with baseline measures collected between 2006 and 2010 and data linked to hospital and mortality records. Extensive information on socio-demographics, health behavior and medical history was obtained via a touch screen questionnaire and face-to-face interviews. Physical measures (e.g., height, weight), blood and urine samples were performed by trained personnel conforming to a standardized protocol. All participants provided written informed consent (24; 25).

Measurement of dietary intake

A web-based 24-h dietary instrument (Oxford WebQ) was used to collect detailed dietary data. The Oxford WebQ was validated against an interviewer-administered 24-h recall questionnaire (26) and collected information on the quantities of up to 206 widely consumed food items and 32 types of drinks consumed over the previous day (27). Participants with valid email addresses were invited to complete a dietary questionnaire at baseline and were followed up to 4 times between April 2009 and June 2012 (cycle 1: February 2011 to April 2011; cycle 2: June 2011 to September 2011; cycle 3: October 2011 to December 2011; cycle 4: April 2012 to June 2012). Participants with at least two assessments were retained for analysis in order to better reflect usual intakes (28) and their mean dietary intake was calculated (17).

Food intake data were aggregated into 50 main groups aligned to the UK National Diet and Nutrition Survey and according to the similarity of their nutritional composition and culinary usage using methods described previously (17; 27). Total energy and nutrient intakes were automatically estimated by multiplying the number of portions consumed by the set quantity of each food portion size and its nutrient composition according to the UK Nutrient Databank Food Composition Tables (2012-2013 and 2013-2014) (29). Energy density (kJ/g), SFA (% total energy), free sugars (% total energy) and fiber density (g/MJ) were selected due to their significant role in the development of obesity and type 2 diabetes and its high frequency in daily life (9; 30; 31). Energy density (kJ/g) was calculated by dividing total food energy (kJ) by total food weight (g) and excluded all beverages, because of their disproportionate influence on total energy density value (32). SFA (% total energy) and free sugars (% total energy) were calculated by dividing daily energy from saturated fat or free sugars by total daily energy intake. Fiber density (g/MJ) was calculated by absolute intake of fiber (g/d) divided by total daily energy intake (MJ) then multiplying by 1000. Free sugars were defined as containing free sugars based on the Scientific Advisory Committee on Nutrition (SACN) in the UK definitions, fiber was calculated using the Englyst method (29).

Individual estimated energy requirements (EER) were calculated by the Schofield Equation from the 1985 FAO/WHO/UNU Expert Consultation Report on Human Energy Requirements, dietary misreporting was calculated by using the ratio of energy intake (EI) to EER (EI:EER) and its 95% confidence interval (CI) (33). Dietary underreporters (EI:EER < 95%CI EI:EER) and overreporter (EI:EER > 95%CI EI:EER) were further excluded.

Ascertainment of outcomes

The outcomes for this study were primary or secondary diabetes events (excluding type 1 diabetes) ascertained from hospital episode statistics (HES) and death registry data linked to the UK Biobank. We defined incident diabetes events as a hospital admission or death with the following ICD-10 codes (International Classification of Diseases, 10th revision) on the hospital or death records: E11-E14. Hospital admission data were available up to 31th Jan 2021 in England and Scotland and 28th February 2018 in Wales. Deaths were ascertained via linkage to the death certificates and were available starting from baseline up to 31th Jan 2021 in England, Wales and Scotland. Therefore, we censored diabetes analysis at the date of first incident diabetes or death or 31th Jan 2021, whichever occurred first.

Statistical analyses

Identification of dietary patterns

Reduced rank regression (RRR) was used to identify DPs that explain the maximum variation in a set of nutrient response variables hypothesized to be on the causal pathway between food groups and incident type 2 diabetes. Energy density (kJ/g), SFA (% total energy), free sugars (% total energy) and fiber density (g/MJ) were all used as response variables in the RRR model. The number of extracted patterns is in accordance with the number of response variables in the model. Respondents were scored for each DP (z score) representing the degree to which their dietary intake reflected each DP relative to other respondents. Increasing intakes of foods with positive factor loadings increase the DP z score; increasing intakes of foods with negative factor loadings decrease the DP z score. A larger factor loading value indicates that the food group makes a greater contribution to the DP. The associations between DPs and nutrients response variables were evaluated by correlation coefficients (Supplementary Table 2). DPs which individually explained more than 20% of variation in response variables were retained for subsequent analyses.

Prospective association of dietary patterns with incident type 2 diabetes

We used multivariable Cox proportional hazard models stratified for sex and regions (England, Scotland and Wales) with age (years) as timescale variable to obtain HRs (hazard ratio) with 95% CIs (confidence interval) per unit increase in DP z scores, with sequential adjustment for ethnicity (white, others, missing), Townsend index of deprivation (quintiles 1-5, with higher scores representing greater deprivation), education (higher degree [college or university degree, or professional qualifications], any school degree [A levels, AS levels, O levels, GCSEs or CSEs], vocational qualifications [NVQ, HND or HNC], other [none of the above qualifications], missing), smoking status (never, current, previous, missing), physical activity (low [<600 metabolic equivalent (MET)-minutes per week], moderate [≥600 and <3000 MET-minutes per week], high [≥ 3000 MET-minutes per week], missing), log-transformed energy intake, family history of diabetes (yes, no), hypertension (yes, no), cardiovascular disease (yes, no), high cholesterol (yes, no), menopause in women (yes, no, not applicable [men]) and BMI group (underweight [<18.5 kg/m^2], healthy weight [18.5 to <25 kg/m^2], overweight [25 to <30 kg/m^2], obese [30+ kg/m^2], missing). Detailed information on variables could be found in Supplementary Table 1. The proportional hazards assumption was assessed by Schoenfeld residuals. Trend tests were performed by including the median score of each pattern quintile as a continuous variable in the models, the lowest quintile was used as the reference. Restricted cubic splines models with the same covariate specification were computed with five knots to examine non-linear associations between DP z scores and incident type 2 diabetes.

Likelihood ratio tests were used to examine the heterogeneity of the associations of the DPs with risk of incident type 2 diabetes by age group (<60, ≥60 years), sex (female, male), smoking status (never, previous, current), physical activity (low, moderate, high), BMI group (underweight/healthy [<25], overweight [25 to <30], obesity [30+]).

Sensitivity and exploratory analyses

To investigate potential bias in relation to random variation in individual intakes, we repeated the RRR analysis to derive dietary patterns among participants providing 3+ (N=75,003), 4+ (N=34,644) or 5 (N=5,504) online dietary assessments which may better reflect usual intakes. A second sensitivity analysis excluded participants who had a diabetes event within two years after completing their last 24-h online dietary assessment.

RRR analysis was performed using SAS statistical software (version 9.4, SAS Institute, USA), the rest of the analyses were performed in Stata/MP 14.0.

Results

Of the 502,536 UK Biobank participants recruited at baseline, participants were excluded for the following reasons: a diabetes event occurred before baseline (N=26,845); pregnancy (n=375); did not complete any validated dietary assessment (n=273,434) or completed only one 24-h online dietary assessment (n=79,995); diabetes occurred before the latest dietary questionnaire (n=172); missing nutrient data (n=21); implausible energy intake (over reporters: n=130, under reporters: n=996; or missing BMI data (n=225) (Supplementary Figure 1). In this analysis, 120,343 participants who provided complete data using the 24-h online dietary assessments on at least two occasions were included.

Four DPs were derived (Supplementary Table 2). DP1 (43%) and DP2 (20%) jointly explained 63% of variation in all response variables with DP3 and DP4 only contributing to 10% and 4%, which were retained in subsequent analyses. DP1 was characterized by a high consumption of chocolate and confectionery, butter and other animal fat spreads, low-fiber bread and a low consumption of fresh fruit, vegetables and high-fiber breakfast cereals. DP2 was characterized by a high consumption of sugar-sweetened beverages (SSBs), fruit juice, table sugars and preserves, and a low consumption of high fat cheese, butter and other animal fat spreads. In sensitivity analyses we derived new DPs from RRR using subsamples of people who provided 3, 4 or 5 dietary assessments or excluding those that experienced the event within 2 years of completing their latest assessment (Supplementary Figures 2-5), yielding consistent results.

During 1,350,644 person-years of follow-up (11.2 years of median follow-up from baseline, 8.4 years of follow-up after the latest dietary assessment), 2,878 participants developed incident type 2 diabetes (including 83 fatal cases). Baseline characteristics including demographics, socioeconomic status, behavior risk factors, health history/conditions, dietary intake, main food groups across quintiles of DP scores are given in Table 1. Higher quintiles of DP1 included a higher proportion of men, people of younger age, with higher Townsend index, lower attained education level, current smokers, less physically active, and higher prevalence of obesity or hypertension. A higher proportion of participants of younger age, with lower attained education level, never smokers, more physically active, higher prevalence of hypertension, cardiovascular disease and high cholesterol had higher quintiles of DP 2.

Intakes of chocolate and confectionery, butter and other animal fat spreads, table sugars and preserves, SSBs and other sugary drinks, low-fiber bread, processed meat and high fat cheese were higher across quintiles of DP1 z scores, along with lower intakes of high-fiber bread, vegetables and fresh fruit. In DP2, intakes of chocolate and confectionery, added sugars and preserves, SSBs and other sugary drinks, fruit juice and fresh fruit were higher across quintiles of DP2 z scores, with lower intakes of low-fiber bread, processed meat, high fat cheese and vegetables.

There was positive association between adherence to DP1 and incident type 2 diabetes after sequential adjustment for demographics, socio-demographics, behavior risk factors and health history/conditions (P_{trend}<0.001), and remained statistically significant after adjustment for BMI (P_{trend}<0.001, Table 2, Supplementary Figure 6). There was no evidence of an association between adherence to DP2 and the risk of type 2 diabetes.

The exclusion of participants who had a diabetes event within two years of completing their last 24-h online dietary assessment showed that the associations between DPs and the risk of diabetes events were largely unchanged (Supplementary Table 4).

There was a significant interaction between both DPs and age, with significantly higher risks among younger people (<60 years) with DP1 [HR 1.13, 95%CI 1.09-1.18] and DP2 [HR 1.10, 95%CI 1.05-1.15] (Figure 2). A significant interaction was also found for DP1 and BMI, with higher risks among people with obesity at baseline [HR 1.11, 95%CI 1.07-1.16]. There was no evidence of any effect modification by gender, smoking status or physical activity.

Discussion

In this large cohort of British adults, we identified two main DPs. The main DP (DP1), explained almost half of the variance in nutrient intakes (43%) and was characterized by high intakes of chocolate and confectionery, butter and other animal fat spreads, low-fiber bread and low intakes of fresh fruit, vegetables and high-fiber breakfast cereals, contributing to excess energy intake, SFA and free sugars with low intake of fiber. There was a positive linear association between this DP and type 2 diabetes events, with stronger associations found among people aged <60 years and those who had obesity at baseline. A second DP (DP2), explaining much less of the variance in nutrient intakes (20%) and characterized by a very high intake of free sugars from higher intakes of SSBs, fruit juice, table sugars and preserves, together with lower intakes of high fat cheese, butter and other animal fat spreads, was not associated with the incidence of type 2 diabetes.

This study contributes new observational evidence to help understand the associations of a whole diet with incident type 2 diabetes. In practice, nutrients are not consumed separately and they may exert effects on health in a synergistic way. Analyses of DPs through RRR, which use nutrients of concern as response variables in the causal pathway, may help detect stronger associations due to the cumulative effects of different dietary components. The main DP derived in this study is characterized by food groups (and underlying nutrients) for which associations with type 2 diabetes are consistent with previous studies (19; 34). For example, excessive consumption of chocolate and confectionery (35), red and processed meat (36), added sugars and preserves (37), and low consumption of vegetables, fresh fruit (38) and high fiber bread (39) have been associated with adverse metabolic effects and higher risk of diabetes. A recent systematic literature review and meta-analysis of 16 cohorts found three patterns derived by RRR showing refined grains, SSBs, and processed meat were all significantly associated with diabetes risk, independent of the biomarkers used as response variables and study populations (40). A recent clinical trial in overweight males provided evidence that compared with a diet enriched in free sugars, a diet enriched in SFA led to increased intrahepatic triacylglycerol (IHTAG) and postprandial glycaemia, but a diet enriched in free sugars did not influence IHTAG and led to only minor metabolic changes (41).

The second DP was characterized by foods which contributed high intakes of free sugars (e.g. SSBs and fruit juice), but in the context of low SFA and energy density and adequate amounts of fiber due to high intake of fruits and vegetables. People in the highest quintile of this pattern consumed on average 17.4% of dietary energy from free sugars (more than three times the UK dietary guideline) but with only 10% from SFA (the recommended maximum level) with no significant increase in energy intake. Here there was no evidence of an overall association with incidence of diabetes. Accordingly, this analysis, which controls for differences in BMI, supports the hypothesis that free sugars increase the risk for type 2 diabetes because of their contribution to excess energy intake, rather than an direct, independent effect (42; 43). In short- or medium-term isoenergetic intervention trials, free sugars showed no effect on body weight or blood pressure and the effects on blood lipid profile were not conclusive (11). It is also consistent with the evidence that diets low in saturated fat reduce the risk of type 2 diabetes (44).

Our study found that the positive associations between DP1 and risk of type 2 diabetes were slightly attenuated but remained statistically significant after adjusting for BMI, suggesting that BMI may only partly explain the variability in type 2 diabetes risk. However, in subsequent stratified analyses, DP1 was much strongly associated with type 2 diabetes among people with obesity at baseline. Therefore these results suggest that residual confounding or mediation by adiposity cannot be ruled out and might account for some of the observed associations.

Adherence to both DP1 and DP2 was a particular risk factor for the development of type 2 diabetes in younger people (<60 years), suggesting that, when all other things are hold constant, a poor diet could increase the risk of type 2 diabetes in younger people. Also, there are also age-related differences in healthy foods consumption such as fruit and vegetable, older adults consuming more fruit and vegetables than younger adults, which have been observed in several UK national surveys. In the National Diet and Nutrition Survey (2016-17 to 2018-19), adults (19-64 years) consumed on average 4.3 portions of fruit and vegetables per day and older adults (>65 years) consumed 4.5 portions per day (45). The wider literature also has suggested that older people tend to consume more fruits and vegetables and other healthy foods (46), which may be the case of the UK Biobank participants. In our middle-aged population, 55% of type 2 diabetes events were observed in people aged ≥60 years. Our study did not showed differential associations by gender, although a previous study using RRR approach found that a high-fat, high-GI, low-fiber dietary pattern was associated with increased type 2 diabetes risk in middle-aged British women but not men (47), though the mechanistic basis for such differences is unclear. Also, differential associations across physical activity groups were not observed. DP1 was consistently associated with higher risk of type 2 diabetes regardless of the amount of physical activity reported, however this association was slightly less strong in people with high physical activity.

Strengths of this study include a large sample size and the empirical hypothesis-based DP analysis that was used to identify dietary patterns in this cohort of middle-aged British adults and their association with type 2 diabetes. Some limitations of this study should be acknowledged. First, DPs were identified and validated from at least two 24-h online self-reported dietary assessments, which may be affected by recall bias or misreporting (e.g. people at higher risk of diabetes or aware of their diabetes biomarkers may report foods differentially); and DPs were derived using data from people with higher willingness to report their diet (which may be affected by selection bias). However, our results were reassured by the sensitivity analyses which used at least 2+, 3+, 4+ or 5 repeated assessments to better capture usual intakes, showing consistent results. Second, some uncommon foods and other important covariates were not captured by the questionnaires, hence some residual confounding may remain. Third, although the identified DPs are specific to the UK population, assuming that the relationships between dietary factors and health outcomes would be similar among individuals, outcome-dependent methods, such as RRR, are expected to be relatively reproducible across different study populations (48; 49). Fourth, incident diabetes cases were ascertained based on hospital records, but this may not fully capture all cases of type 2 diabetes since many of these are reported in primary care. Finally, as usually happens with the RRR approach, the emerging dietary patterns did not explain all of the variability in the nutrient response variables that were included in the RRR model, and any remaining variability may be explained by other nutrients which may also be involved in the disease pathway. In exploratory mediation analyses (Supplementary table 4), we investigated the role of the nutrient response variables in the association between dietary patterns and the risk of diabetes. After adjustment for the nutrients showing the highest correlations with each DP, the associations for DP1 and DP2 were largely unchanged, which means that the derived dietary patterns were associated with diabetes independent of those nutrient response variables. This can be partly explained by the fact that both the DP and the nutrient response variables may explain the same variability. But this also

suggests that there might be other nutrients on the causal pathway explaining the observed risk with type 2 diabetes, for which future research is warranted. Residual confounding due to unmeasured variables (e.g. blood glucose (50), impaired fasting glycaemia (51; 52)) is another limitation of observational research which can also explain the observed results.

Current dietary guidelines in the UK and many other countries are still based predominately on nutrient recommendations. While this reflects the underlying evidence base, especially experimental and mechanistic research, it does not reflect the way people eat. Food-based dietary guidelines may help to accelerate behavior change compared to nutrient-based recommendations by providing targeted and simpler advice on foods which make it more, or less, likely to achieve an overall healthy diet (53). Moreover, it may help reduce the risk of conflicting messages regarding the relative importance of one or other nutrient, particularly saturated fat and free sugars, and recognize that there are many foods which are important sources of both.

In conclusion, this large population-based cohort study has shown clear evidence that diets high in chocolate and confectionery, butter and other animal fat spreads, low-fiber bread and low intakes of fresh fruit, vegetables and high-fiber breakfast cereals are associated with a higher risk of developing type 2 diabetes in this cohort of middle-aged British adults. The effects of a poor diet are especially pronounced in younger people and those living with obesity.

Acknowledgments

This research has been conducted using the UK Biobank resource under application number 14990. CP and MG conceived and designed the research question. MG analyzed the data and GLA contributed to dietary pattern analysis, GLA, CP, APC, JC, KP contributed to the data analysis; MG and CP wrote the first draft of the manuscript. MG and CP are guarantors of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors contributed to interpretation of the data, revised the manuscript critically for important intellectual content, and read and approved the final manuscript.

Funding

CP received a British Nutrition Foundation pump priming award which paid for the access to the data. SAJ and CP are funded by the Oxford and Thames Valley NIHR Applied Research Centre. SAJ and PA are National Institute of Health Research (NIHR) senior investigators. SAJ and JC are funded by NIHR Oxford Biomedical Research Centre. JC is supported by core grants to CTSU (Clinical Trial Service Unit) from the Medical Research Council and the British Heart Foundation (CH/1996001/9454). The funders had no role in designing the study, the analysis, or the decision to submit the paper. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Disclosures

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: support from the NIHR SPCR, ARC and BRC, as well as from the MRC and BHF for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Reference

1. Chan JCN, Lim L-L, Wareham NJ, Shaw JE, Orchard TJ, Zhang P, Lau ESH, Eliasson B, Kong APS, Ezzati M, Aguilar-Salinas CA, McGill M, Levitt NS, Ning G, So W-Y, Adams J, Bracco P, Forouhi NG, Gregory GA, Guo J, Hua X, Klatman EL, Magliano DJ, Ng B-P, Ogilvie D, Panter J, Pavkov M, Shao H, Unwin N, White M, Wou C, Ma RCW, Schmidt MI, Ramachandran A, Seino Y, Bennett PH, Oldenburg B, Gagliardino JJ, Luk AOY, Clarke PM, Ogle GD, Davies MJ, Holman RR, Gregg EW. The Lancet Commission on diabetes: using data to transform diabetes care and patient lives. The Lancet 2020;396:2019-2082
2. Shah AD, Langenberg C, Rapsomaniki E, Denaxas S, Pujades-Rodriguez M, Gale CP, Deanfield J, Smeeth L, Timmis A, Hemingway H. Type 2 diabetes and incidence of cardiovascular diseases: a cohort study in 1·9 million people. Lancet Diabetes Endocrinol 2015;3:105-113
3. Barron E, Bakhai C, Kar P, Weaver A, Bradley D, Ismail H, Knighton P, Holman N, Khunti K, Sattar N, Wareham NJ, Young B, Valabhji J. Associations of type 1 and type 2 diabetes with COVID-19-related mortality in England: a whole-population study. The Lancet Diabetes & Endocrinology 2020;8:813-822
4. Neuenschwander M, Ballon A, Weber KS, Norat T, Aune D, Schwingshackl L, Schlesinger S. Role of diet in type 2 diabetes incidence: umbrella review of meta-analyses of prospective observational studies. BMJ (Clinical research ed) 2019;366:l2368
5. Collaborators GRF. Global burden of 87 risk factors in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020;396:1223-1249
6. van Dam RM, Willett WC, Rimm EB, Stampfer MJ, Hu FB. Dietary Fat and Meat Intake in Relation to Risk of Type 2 Diabetes in Men. Diabetes Care 2002;25:417-424
7. Reynolds AN, Akerman AP, Mann J. Dietary fibre and whole grains in diabetes management: Systematic review and meta-analyses. PLoS medicine 2020;17:e1003053
8. Choo VL, Viguiliouk E, Blanco Mejia S, Cozma AI, Khan TA, Ha V, Wolever TMS, Leiter LA, Vuksan V, Kendall CWC, de Souza RJ, Jenkins DJA, Sievenpiper JL. Food sources of fructose-containing sugars and glycaemic control: systematic review and meta-analysis of controlled intervention studies. BMJ 2018;363:k4644
9. Macdonald IA. A review of recent evidence relating to sugars, insulin resistance and diabetes. Eur J Nutr 2016;55:17-23
10. Johnson RK, Lichtenstein AH, Anderson CAM, Carson JA, Després J-P, Hu FB, Kris-Etherton PM, Otten JJ, Towfighi A, Wylie-Rosett J. Low-Calorie Sweetened Beverages and Cardiometabolic Health: A Science Advisory From the American Heart Association. Circulation 2018;138:e126-e140
11. Fattore E, Botta F, Agostoni C, Bosetti C. Effects of free sugars on blood pressure and lipids: a systematic review and meta-analysis of nutritional isoenergetic intervention trials. The American Journal of Clinical Nutrition 2016;105:42-56
12. Diewertje S, Brouwer-Brolsma EM, Berendsen AAM, Vera M, Poppitt SD, Silvestre MP, Angelo T, Louis P, Claude B, Anne R. Protein intake and the incidence of pre-diabetes and diabetes in 4 population-based studies: the PREVIEW project. The American Journal of Clinical Nutrition 2019;109:1310-1318
13. Chen Z, Franco OH, Lamballais S, Ikram MA, Schoufour JD, Muka T, Voortman T. Associations of specific dietary protein with longitudinal insulin resistance, prediabetes and type 2 diabetes: The Rotterdam Study. Clinical Nutrition 2020;39:242-249
14. Imamura F, O'Connor L, Ye Z, Mursu J, Hayashino Y, Bhupathiraju SN, Forouhi NG. Consumption of sugar sweetened beverages, artificially sweetened beverages, and fruit juice and incidence of type 2 diabetes: systematic review, meta-analysis, and estimation of population attributable fraction. BMJ 2015;351:h3576
15. Zheng J-S, Sharp SJ, Imamura F, Chowdhury R, Gundersen TE, Steur M, Sluijs I, van der Schouw YT, Agudo A, Aune D, Barricarte A, Boeing H, Chirlaque M-D, Dorronsoro M, Freisling H, El-Fatouhi D, Franks PW, Fagherazzi G, Grioni S, Gunter MJ, Kyrø C, Katzke V, Kühn T, Khaw K-T, Laouali N, Masala G, Nilsson PM, Overvad K, Panico S, Papier K, Quirós JR, Rolandsson O, Redondo-Sánchez D, Ricceri F, Schulze MB, Spijkerman AMW, Tjønneland A, Tong TYN, Tumino R, Weiderpass E, Danesh J,

Butterworth AS, Riboli E, Forouhi NG, Wareham NJ. Association of plasma biomarkers of fruit and vegetable intake with incident type 2 diabetes: EPIC-InterAct case-cohort study in eight European countries. BMJ 2020;370:m2194
16. Afshin A, Sur PJ, Fay KA, Cornaby L, Ferrara G, Salama JS, Mullany EC, Abate KH, Abbafati C, Abebe Z. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet 2019;393:1958-1972
17. Gao M, Jebb SA, Aveyard P, Ambrosini GL, Perez-Cornago A, Carter J, Sun X, Piernas C. 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 Medicine 2021;19:83
18. Weikert C, Schulze MB. Evaluating dietary patterns: the role of reduced rank regression. Current opinion in clinical nutrition and metabolic care 2016;19:341-346
19. Forouhi NG, Misra A, Mohan V, Taylor R, Yancy W. Dietary and nutritional approaches for prevention and management of type 2 diabetes. BMJ 2018;361:k2234
20. McNaughton SA, Mishra GD, Brunner EJ. Dietary Patterns, Insulin Resistance, and Incidence of Type 2 Diabetes in the Whitehall II Study. Diabetes Care 2008;31:1343-1348
21. Liese AD, Weis KE, Schulz M, Tooze JA. Food Intake Patterns Associated With Incident Type 2 Diabetes. The Insulin Resistance Atherosclerosis Study 2009;32:263-268
22. Seah JY, Ong CN, Koh W-P, Yuan J-M, van Dam RM. A dietary pattern derived from reduced rank regression and fatty acid biomarkers is associated with lower risk of type 2 diabetes and coronary artery disease in Chinese adults. The Journal of nutrition 2019;149:2001-2010
23. Duan M-J, Dekker LH, Carrero J-J, Navis GJ. Blood lipids-related dietary patterns derived from reduced rank regression are associated with incident type 2 diabetes. Clinical Nutrition 2021;40:4712-4719
24. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. Plos med 2015;12:e1001779
25. Collins R. What makes UK Biobank special? Lancet 2012;379:1173-1174
26. Liu B, Young H, Crowe FL, Benson VS, Spencer EA, Key TJ, Appleby PN, Beral V. Development and evaluation of the Oxford WebQ, a low-cost, web-based method for assessment of previous 24 h dietary intakes in large-scale prospective studies. Public health nutrition 2011;14:1998-2005
27. Piernas C, Perez-Cornago A, Gao M, Young H, Pollard Z, Mulligan A, Lentjes M, Carter J, Bradbury K, Key TJ, Jebb SA. Describing a new food group classification system for UK biobank: analysis of food groups and sources of macro- and micronutrients in 208,200 participants. European Journal of Nutrition 2021;60:2879-2890
28. Carter JL, Lewington S, Piernas C, Bradbury K, Key TJ, Jebb SA, Arnold M, Bennett D, Clarke R. Reproducibility of dietary intakes of macronutrients, specific food groups, and dietary patterns in 211 050 adults in the UK Biobank study. Journal of Nutritional Science 2019;8:e34
29. Perez-Cornago A, Pollard Z, Young H, van Uden M, Andrews C, Piernas C, Key TJ, Mulligan A, Lentjes M. Description of the updated nutrition calculation of the Oxford WebQ questionnaire and comparison with the previous version among 207,144 participants in UK Biobank. European Journal of Nutrition 2021;60:4019-4030
30. Thom G, Lean M. Is There an Optimal Diet for Weight Management and Metabolic Health? Gastroenterology 2017;152:1739-1751
31. Neuenschwander M, Ballon A, Weber KS, Norat T, Aune D, Schwingshackl L, Schlesinger S. Role of diet in type 2 diabetes incidence: umbrella review of meta-analyses of prospective observational studies. BMJ 2019;366:l2368
32. Johnson L, Mander AP, Jones LR, Emmett PM, Jebb SA. A prospective analysis of dietary energy density at age 5 and 7 years and fatness at 9 years among UK children. International journal of obesity (2005) 2008;32:586-593

33. Schofield WN, Schofield C, James WPT. Basal metabolic rate: review and prediction, together with an annotated bibliography of source material. Hum Nutr 1985;39 C:5-41
34. Magliano DJ, Islam RM, Barr EL, Gregg EW, Pavkov ME, Harding JL, Tabesh M, Koye DN, Shaw JE. Trends in incidence of total or type 2 diabetes: systematic review. BMJ 2019;366:l5003
35. Curtis PJ, Sampson M, Potter J, Dhatariya K, Kroon PA, Cassidy A. Chronic ingestion of flavan-3-ols and isoflavones improves insulin sensitivity and lipoprotein status and attenuates estimated 10-year CVD risk in medicated postmenopausal women with type 2 diabetes: a 1-year, double-blind, randomized, controlled trial. Diabetes care 2012;35:226-232
36. de Koning L, Fung TT, Liao X, Chiuve SE, Rimm EB, Willett WC, Spiegelman D, Hu FB. Low-carbohydrate diet scores and risk of type 2 diabetes in men. The American journal of clinical nutrition 2011;93:844-850
37. Kmietowicz Z. Type 2 diabetes: sweetened drinks pose greater risk than other sugary foods. BMJ 2018;363:4943
38. Bazzano LA, Li TY, Joshipura KJ, Hu FB. Intake of fruit, vegetables, and fruit juices and risk of diabetes in women. Diabetes care 2008;31:1311-1317
39. Schulze MB, Schulz M, Heidemann C, Schienkiewitz A, Hoffmann K, Boeing H. Fiber and magnesium intake and incidence of type 2 diabetes: a prospective study and meta-analysis. Archives of internal medicine 2007;167:956-965
40. Jannasch F, Kröger J, Schulze MB. Dietary Patterns and Type 2 Diabetes: A Systematic Literature Review and Meta-Analysis of Prospective Studies. The Journal of Nutrition 2017;147:1174-1182
41. Parry SA, Rosqvist F, Mozes FE, Cornfield T, Hutchinson M, Piche M-E, Hülsmeier AJ, Hornemann T, Dyson P, Hodson L. Intrahepatic Fat and Postprandial Glycemia Increase After Consumption of a Diet Enriched in Saturated Fat Compared With Free Sugars. Diabetes care 2020;43:1134-1141
42. Te Morenga L, Mallard S, Mann J. Dietary sugars and body weight: systematic review and meta-analyses of randomised controlled trials and cohort studies. BMJ 2013;346:e7492
43. Choo VL, Viguiliouk E, Mejia SB, Cozma AI, Khan TA, Ha V, Wolever TM, Leiter LA, Vuksan V, Kendall CW. Food sources of fructose-containing sugars and glycaemic control: systematic review and meta-analysis of controlled intervention studies. BMJ 2018;363:k4644
44. Morio B, Fardet A, Legrand P, Lecerf JM. Involvement of dietary saturated fats, from all sources or of dairy origin only, in insulin resistance and type 2 diabetes. Nutr Rev 2016;74:33-47
45. National Diet and Nutrition Survey Rolling programme Years 9 to 11 (2016/2017 to 2018/2019) [article online], Available from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/943114/NDNS_UK_Y9-11_report.pdf. 2020
46. Reininger B, Lee M, Jennings R, Evans A, Vidoni M. Healthy eating patterns associated with acculturation, sex and BMI among Mexican Americans. Public health nutrition 2017;20:1267-1278
47. Pastorino S, Richards M, Pierce M, Ambrosini GL. A high-fat, high-glycaemic index, low-fibre dietary pattern is prospectively associated with type 2 diabetes in a British birth cohort. Br J Nutr 2016;115:1632-1642
48. Huybrechts I, Lioret S, Mouratidou T, Gunter MJ, Manios Y, Kersting M, Gottrand F, Kafatos A, De Henauw S, Cuenca-García M, Widhalm K, Gonzales-Gross M, Molnar D, Moreno LA, McNaughton SA. Using reduced rank regression methods to identify dietary patterns associated with obesity: a cross-country study among European and Australian adolescents. Br J Nutr 2017;117:295-305
49. Krebs-Smith SM, Subar AF, Reedy J. Examining Dietary Patterns in Relation to Chronic Disease: Matching Measures and Methods to Questions of Interest. Circulation 2015;132:790-793
50. Russell WR, Baka A, Björck I, Delzenne N, Gao D, Griffiths HR, Hadjilucas E, Juvonen K, Lahtinen S, Lansink M, Loon LV, Mykkänen H, Östman E, Riccardi G, Vinoy S, Weickert MO. Impact of Diet Composition on Blood Glucose Regulation. Critical reviews in food science and nutrition 2016;56:541-590

51. de Hoogh IM, Oosterman JE, Otten W, Krijger AM, Berbée-Zadelaar S, Pasman WJ, van Ommen B, Pijl H, Wopereis S. The Effect of a Lifestyle Intervention on Type 2 Diabetes Pathophysiology and Remission: The Stevenshof Pilot Study. Nutrients 2021;13
52. Kim MK, Han K, Koh ES, Hong O-K, Baek K-H, Song K-H, Kwon H-S. Cumulative exposure to impaired fasting glucose and future risk of type 2 diabetes mellitus. Diabetes Research and Clinical Practice 2021;175:108799
53. Van Horn L, Carson JAS, Appel LJ, Burke LE, Economos C, Karmally W, Lancaster K, Lichtenstein AH, Johnson RK, Thomas RJ. Recommended dietary pattern to achieve adherence to the American Heart Association/American College of Cardiology (AHA/ACC) guidelines: a scientific statement from the American Heart Association. Circulation 2016;134:505-529

Table 1 Baseline characteristics of the study population and average dietary intake

	Total (N=120,343)	Dietary pattern 1				Dietary pattern 2			
		Quintile 1(N=24,069)	Quintile 3(N=24,068)	Quintile 5(N=24,068)	*P value**	Quintile 1(N=24,069)	Quintile 3(N=24,068)	Quintile 5(N=24,068)	*P value**
Demographics									
Women (%)	56.5	70.5	59.0	35.9	<0.001	56.2	60.8	47.5	<0.001
Age (year)[†]	56.1 (7.8)	57.4 (7.3)	56.3 (7.7)	54.1 (8.1)	<0.001	56.0 (7.7)	56.3 (7.8)	55.5 (8.1)	<0.001
Whites (%)	96.8	96.6	96.9	96.7	0.089	97.7	97.2	94.7	<0.001
Socioeconomic status									
Townsend index quintile					<0.001				<0.001
Q1	20.0	20.3	20.6	18.4		19.3	20.5	19.2	
Q3	20.0	19.8	20.7	19.9		19.8	20.4	19.8	
Q5	19.9	19.2	18.6	22.5		20.9	18.6	21.4	
Missing	0.1	0.1	0.1	0.1		0.1	0.1	0.1	
Education group (%)[§]									
None of the above	6.3	5.8	5.9	7.3	<0.001	5.6	6.4	6.7	<0.001
Vocational qualification (NVQ, HND or HNC)	12.4	10.4	12.4	16.1		11.7	11.9	14.0	
Any school degree (A-level, AS-level, O-level, GCSE, CSE)	29.2	27.0	28.9	31.6		28.8	29.9	28.8	
Higher degree (college, university or professional degree/qualification)	51.8	56.5	52.5	44.6		53.6	51.6	50.0	
Missing	0.3	0.3	0.3	0.3		0.2	0.3	0.3	
Behavior risk factors									
Smoking status (%)					<0.001				<0.001
Never	57.6	60.3	59.3	52.2		54.3	58.3	59.4	
Previous	35.3	35.7	35.0	34.9		37.0	35.5	33.0	
Current	6.9	3.8	5.5	12.8		8.5	6.0	7.3	
Missing	0.2	0.2	0.2	0.2		0.2	0.2	0.2	
Physical Activity (IPAQ) (%)					<0.001				<0.001
Low	19.1	14.0	19.4	23.8		19.4	19.6	18.1	
Moderate	45.0	43.0	45.9	44.2		45.8	45.4	43.4	
High	35.8	43.0	34.6	32.0		34.8	34.9	38.5	
Missing	<0.1	<0.1	<0.1	<0.1		<0.1	<0.1	<0.1	
Health history/conditions									
Family history of diabetes (%)	15.7	15.4	15.7	16.1	0.32	15.5	15.9	16.0	0.45
Hypertension (%)	23.3	22.8	23.1	24.3	<0.001	22.3	23.1	24.6	<0.001
Cardiovascular disease (%)	5.2	5.3	5.3	5.2	0.86	4.4	5.0	6.1	<0.001

High cholesterol (%)	13.5	13.3	13.6	13.3	0.76	11.4	13.7	15.2	<0.001
Menopause in women (%)	23.2	24.1	24.3	19.4	<0.001	22.8	24.7	20.8	<0.001
BMI (kg/m^2)	26.5(4.4)	26.0(4.4)	26.4(4.3)	27.3(4.6)	<0.001	26.7(4.7)	26.5(4.4)	26.5(4.4)	<0.001
BMI group (%)					<0.001				<0.001
Underweight (<18.5)	0.6	0.9	0.5	0.5		0.7	0.5	0.5	
Healthy weight (18.5 to <25)	40.1	45.4	41.3	32.1		39.3	40.6	39.1	
Overweight (25 to <30)	41.1	38.1	41.2	44.0		39.9	41.0	42.7	
Obese (30+)	18.2	15.5	17.0	23.3		20.0	17.8	17.6	
Missing	<0.1	0.1	0.0	<0.1		<0.1	<0.1	<0.1	
Dietary intake									
Energy intake(MJ/day)	8.7 (2.2)	8.2 (2.1)	8.3 (2.0)	10.0 (2.5)	<0.001	9.3 (2.4)	8.3 (2.1)	9.0 (2.3)	<0.001
Dietary energy density (kJ/g)	6.5 (1.6)	4.8 (0.8)	6.4 (0.8)	8.4 (1.4)	<0.001	7.1 (1.6)	6.3 (1.5)	6.4 (1.6)	<0.001
Saturated fat (%E)	11.7 (3.2)	9.7 (2.6)	11.8 (2.8)	13.4 (3.3)	<0.001	14.4 (2.9)	11.3 (2.7)	10.0 (2.8)	<0.001
Free sugars (%E)	11.5 (5.2)	8.9 (4.1)	11.4 (4.5)	14.7 (6.1)	<0.001	7.7 (3.3)	10.6 (3.6)	17.4 (5.3)	<0.001
Fiber (g/day)	18.1 (6.2)	23.3 (6.7)	17.1 (5.0)	15.1 (5.3)	<0.001	18.2 (6.2)	17.8 (6.0)	18.5 (6.8)	<0.001
Fiber density (g/MJ)	2.1 (0.6)	2.9 (0.6)	2.1 (0.4)	1.5 (0.4)	<0.001	2.0 (0.6)	2.2 (0.7)	2.1 (0.7)	<0.001
Main food groups (g/day)									
Chocolate and confectionery	11.9 (21.4)	6.0 (12.4)	9.7 (16.0)	23.2 (33.5)	<0.001	8.3 (16.0)	10.4 (17.9)	18.9 (30.8)	<0.001
Butter and other normal animal fat spreads	5.1 (8.7)	1.9 (4.8)	4.3 (7.2)	10.5 (12.4)	<0.001	12.1 (12.6)	3.5 (6.4)	2.0 (5.0)	<0.001
Added sugars and preserves	8.9 (13.0)	5.4 (8.4)	7.6 (10.2)	15.7 (19.4)	<0.001	6.0 (9.4)	7.4 (10.4)	15.1 (18.6)	<0.001
SSBs and other sugary drinks	87.3 (166.1)	45.1 (107.4)	74.1 (134.8)	163.1 (245.5)	<0.001	30.8 (77.6)	58.4 (105.1)	219.5 (267.7)	<0.001
Fruit juice	110.0 (137.6)	104.8 (139.9)	114.0 (135.4)	108.1 (144.3)	<0.001	52.7 (87.6)	92.7 (108.0)	205.7 (188.9)	<0.001
Low-fiber bread	28.1 (42.3)	9.2 (22.1)	22.2 (32.2)	61.0 (58.7)	<0.001	36.6 (50.4)	25.3 (38.2)	25.9 (41.0)	<0.001
Processed meat	18.2 (28.0)	10.3 (19.7)	16.8 (24.5)	29.4 (37.5)	<0.001	25.9 (35.6)	16.8 (25.1)	14.0 (24.3)	<0.001
High fat cheese	14.8 (18.1)	10.9 (15.4)	14.4 (16.7)	19.4 (22.3)	<0.001	28.0 (24.3)	12.1 (14.1)	7.9 (12.4)	<0.001
High-fiber bread	49.0 (47.4)	53.8 (50.0)	51.2 (45.5)	39.3 (47.1)	<0.001	55.8 (52.6)	47.2 (44.2)	46.5 (47.7)	<0.001
Vegetables	191.0 (146.0)	325.6 (185.2)	170.7 (105.8)	107.1 (92.6)	<0.001	196.0 (147.7)	191.3 (140.6)	184.2 (154.9)	<0.001
Fresh fruit	197.0 (153.3)	342.7 (181.7)	177.9 (111.6)	97.4 (97.4)	<0.001	169.9 (142.6)	200.8 (148.5)	216.3 (170.0)	<0.001

Note: †Plus-minus values are means ± standard deviation (SD).§NVQ: national vocational qualification. HND: higher national diploma. HNC: higher national certificate. GCSE: general certificate of secondary education. CSE: comprehensive sex education. BMI: body mass index. *Analysis of variance or chi^2 test where appropriate.

Table 2 Sequentially adjusted hazard ratios of total incident diabetes associated with each z score increase in dietary pattern							
	Group	No. events	Sex, age, and regions adjusted	Plus Socio-demographics*	Plus Behavior risk factors †	Plus Health history/conditions ‡	Plus BMI group
Dietary pattern 1(DP1)	Total	2,878	1.16(1.13,1.19)	1.14(1.11,1.18)	1.11(1.08,1.14)	1.11(1.08,1.14)	1.09(1.06,1.12)
	Q1	452	1.00	1.00	1.00	1.00	1.00
	Q2	513	1.15(1.05,1.25)	1.15 (1.05,1.26)	1.11 (1.02,1.21)	1.10 (1.01,1.20)	1.13(1.04,1.23)
	Q3	549	1.23(1.13,1.33)	1.22 (1.12,1.33)	1.16 (1.06,1.26)	1.15 (1.06,1.25)	1.19(1.10,1.30)
	Q4	621	1.41(1.31,1.53)	1.39 (1.29,1.51)	1.28 (1.18,1.38)	1.27(1.17,1.37)	1.25(1.16,1.35)
	Q5	743	1.73(1.60,1.86)	1.65 (1.53,1.78)	1.46 (1.35,1.59)	1.45(1.34,1.57)	1.38(1.27,1.49)
	P for trend(p-value) §		<0.001	<0.001	<0.001	<0.001	<0.001
Dietary pattern 2(DP2)	Total	2,878	1.05(1.01,1.08)	1.03(0.99,1.06)	1.03(1.00,1.07)	1.02(0.98,1.05)	1.03(0.99,1.06)
	Q1	610	1.00	1.00	1.00	1.00	1.00
	Q2	545	0.98(0.90,1.06)	0.97(0.89,1.06)	0.96(0.89,1.05)	0.96 (0.88,1.04)	0.97(0.89,1.05)
	Q3	567	1.02(0.94,1.11)	1.02(0.94,1.10)	1.01(0.93,1.10)	0.99 (0.91,1.07)	1.01(0.93,1.10)
	Q4	524	0.93(0.85,1.01)	0.91(0.83,0.99)	0.91(0.84,1.00)	0.88(0.81,0.96)	0.90(0.82,0.98)
	Q5	632	1.09(1.01,1.18)	1.04(0.96,1.13)	1.06(0.98,1.14)	1.01(0.93,1.09)	1.04(0.96,1.12)
	P for trend(p-value) §		0.234	0.735	0.476	0.840	0.818

Note: Adjusted hazard ratios (HR) and 95% confidence intervals (CI) of total DP scores obtained using Cox proportional hazard regression. Age at risk adjusted for by using age during study as the underlying timescale for Cox regression. All the models were stratified by sex and regions (England, Scotland and Wales). *Socio-demographic characteristics: Ethnicity (white, others, missing), Townsend index of deprivation (quintiles 1-5, with higher scores representing greater deprivation), education (higher degree [college or university degree, or professional qualifications], any school degree [A levels, AS levels, O levels, GCSEs or CSEs], vocational qualifications [NVQ, HND or HNC], other [none of the above qualifications], missing). † Behavior risk factors: smoking status (never, current, previous, missing), physical activity (Low [<600 metabolic equivalent (MET)-minutes per week], moderate [≥600 and <3000 metabolic equivalent (MET)-minutes per week], high [≥ 3000 metabolic equivalent (MET)-minutes per week], missing), log-transformed energy intake. ‡Health history/conditions: family history of diabetes, menopause in women, hypertension, cardiovascular disease, high cholesterol. Full model was stratified by sex and regions (England, Scotland and Wales) and adjusted for demographics, socioeconomic status, behavior risk factors and health history/conditions with BMI group included.

Figure1 Factor loadings for food groups in each dietary pattern

Figure 2 The association between dietary patterns and risk of diabetes by age, sex, smoking status and BMI group (N=120,343)

Supplementary

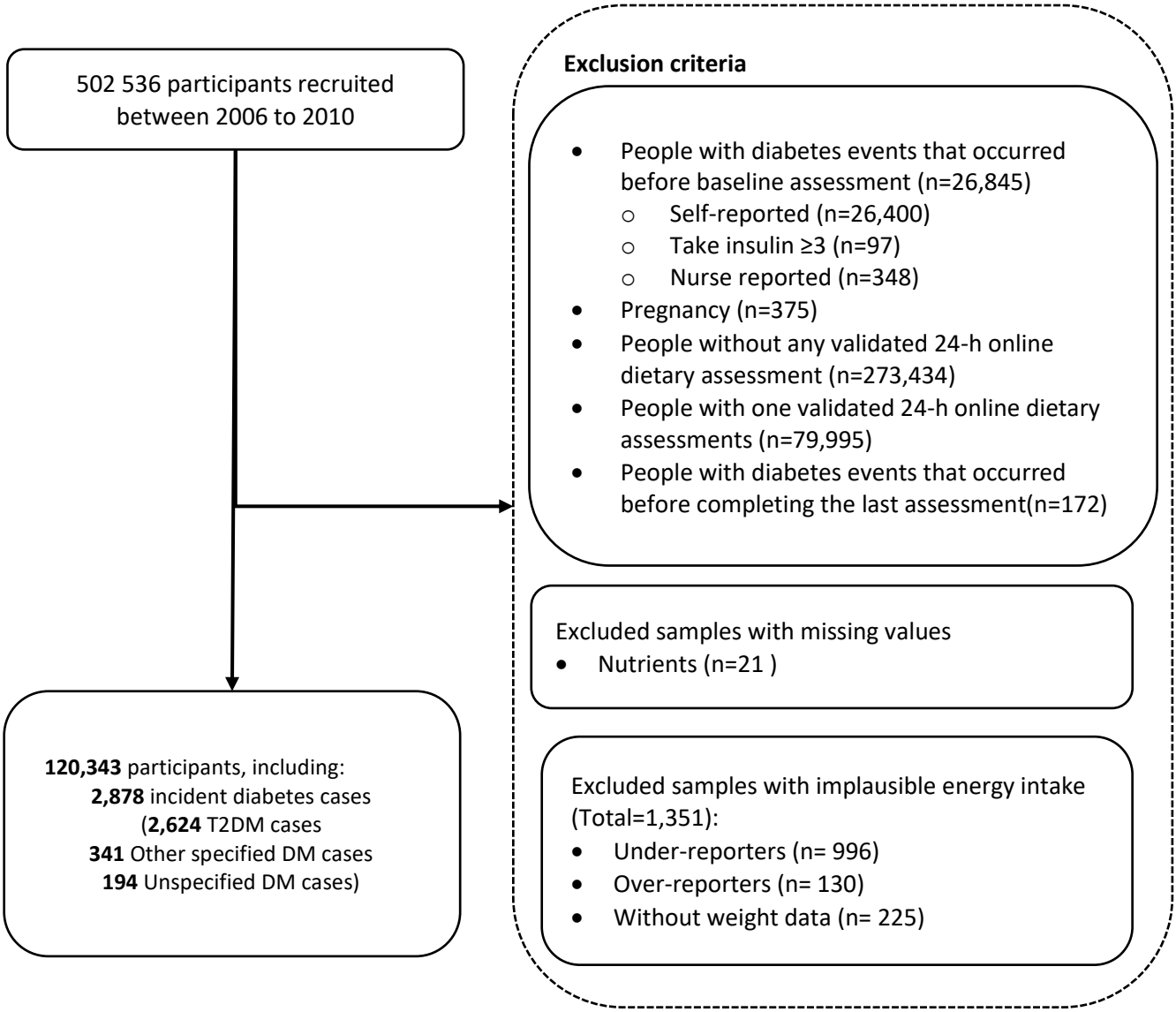
Figures

Supplementary Figure 1: Participant flow chart of the study
Supplementary Figure 2: Factor loadings for dietary patterns characterized by energy density (kJ/g), saturated fat (%total energy), free sugar (%total energy) and fibre density (g/MJ) calculated by using reduced rank regression among people with 3 or more 24-h online dietary assessments in the UK Biobank (N=75,003)
Supplementary Figure 3: Factor loadings for dietary patterns characterized by energy density (kJ/g), saturated fat (%total energy), free sugar (%total energy) and fibre density (g/MJ) calculated by using reduced rank regression among people with 4 or more 24-h online dietary assessments in the UK Biobank (N=34,644)
Supplementary Figure 4: Factor loadings for dietary patterns characterized by energy density (kJ/g), saturated fat (%total energy), free sugar (%total energy) and fibre density (g/MJ) calculated by using reduced rank regression among people with five 24-h online dietary assessments in the UK Biobank (N=5,504)
Supplementary Figure 5: Factor loadings for dietary patterns characterized by energy density (kJ/g), saturated fat (%total energy), free sugar (%total energy) and fibre density (g/MJ) calculated by using reduced rank regression excluding people who had the event within two years after completing their last 24-h online dietary assessment in the UK Biobank (N= 119,870)
Supplementary Figure 6: HRs (95% CIs) of continuous dietary pattern scores for the risk of incident diabetes (n=120,343).

Tables

Supplementary Table 1: Derivation of variables used in analysis from the UK Biobank questionnaire and interviews
Supplementary Table 2: Explained variation (%) in food intakes and response variables for each dietary pattern as assessed using reduced rank regression and correlation coefficient between DPs and response variables (N= 120 343)
Supplementary Table 3: Associations between dietary pattern scores and the risk of total diabetes cases excluding people who had the event within two years after completing their last 24-h online dietary assessment (N= 119,870)
Supplementary Table 4: Adjusted hazard ratios of total incident diabetes associated with each z score increase in dietary pattern (N= 120 343)

Supplementary Figure 1: Participant flow chart of the study



Supplementary Table 1: Derivation of variables used in analysis from the UK Biobank questionnaire and interviews

Variables	Categories used in analysis	UK Biobank variable used (question ID) and source
Health outcome		
Type 2 diabetes	No; Yes	ICD-10: Diabetes (E11 E12 E13 E14)
Demographics		
Sex	Men; Women	Sex (ID: 31)*
Age	Age	Age at recruitment (ID: 21022)*
Ethnicity	White; Others; Missing	Ethnic background (ID: 21000)*
Socioeconomic status		
Townsend index	Quintiles (high index indicates most deprivation), Missing	Townsend index (ID: 189)*
Education	Higher degree (college or university degree, or professional qualifications); Any school degree (A levels, AS levels, O levels, GCSEs or CSEs); Vocational qualifications (NVQ, HND or HNC); Other (none of the above qualifications);Missing	Qualifications (ID: 6138)†
Behavior risk factors		
Smoking status	Never; Current, Previous, Missing	Smoking status (ID: 20116)†
Physical activity (IPAQ MET scores)	Low (<600 metabolic equivalent (MET)-minutes per week); Moderate (≥600 and <3000 metabolic equivalent (MET)-minutes per week); High (≥ 3000 metabolic equivalent (MET)-minutes per week); Missing	Number of days/week of vigorous physical activity 10+ minutes (ID: 904)†; Duration of vigorous activity (ID: 914)†; Number of days/week of moderate physical activity 10+ minutes (ID: 884)†; Duration of moderate activity (ID: 894)†; Number of days/week walked 10+ minutes (ID: 864)†; Duration of walks (ID: 874)†
Health history/conditions		

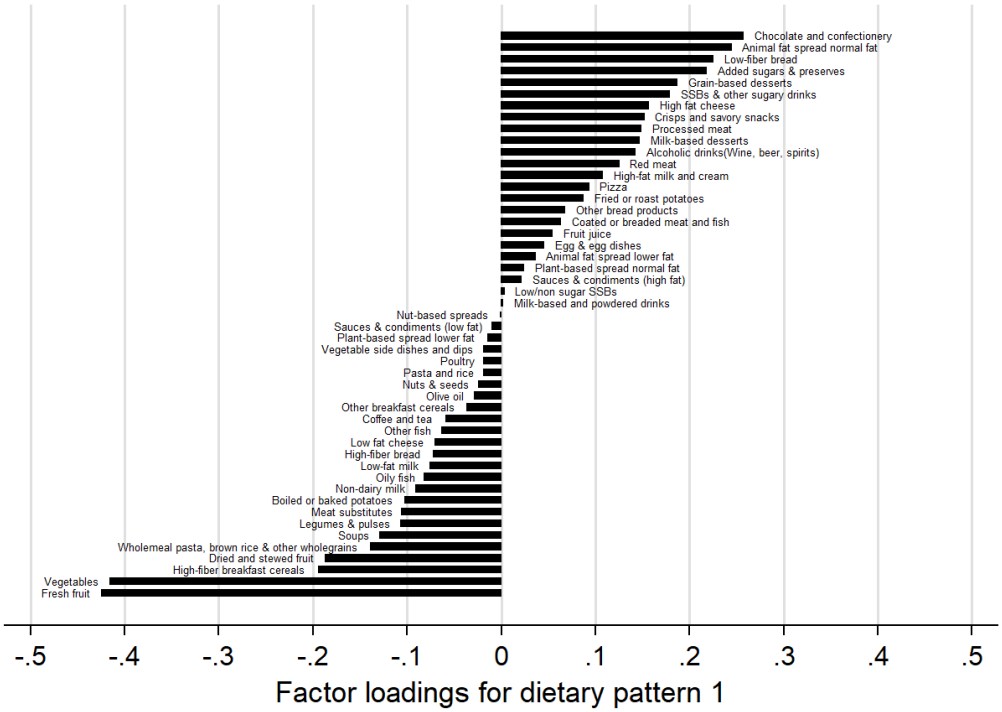
Family history of diabetes	No; Yes (father or/and mother was/were diagnosed with any types of diabetes)	Father or/and mother was/were diagnosed with any types of diabetes (ID: 20107; 20110)[†]
Menopause	N/A (if respondents are men); No; Yes (if responded to questions saying they have experienced the menopause)	Had menopause (women only) (ID: 2724)[†]
Hypertension (%)	No; Yes (if diagnosed by doctor or taking medication for blood pressure)	Vascular/heart problems diagnosed by doctor (high blood pressure is one response) (ID: 6150)[†]; Medication for cholesterol, blood pressure or diabetes (men) (ID: 6177)[†]; Medication for cholesterol, blood pressure, diabetes, or take exogenous hormones (women) (ID: 6153)[†]; Non-cancer illness diagnosed by nurses during verbal interview (ID: 20002) [#]
Cardiovascular disease (%)	No; Yes (if diagnosed by doctor or taking medication for cardiovascular disease)	Vascular/heart problems diagnosed by doctor (heart attack, angina, stoke) (ID: 6150)[†]; Non-cancer illness diagnosed by nurses during verbal interview (ID: 20002) [#]
High cholesterol (%)	No; Yes (if diagnosed by doctor or taking cholesterol lowering medication)	Medication for cholesterol, blood pressure or diabetes (men) (ID: 6177)[†]; Medication for cholesterol, blood pressure, diabetes, or take exogenous hormones (women) (ID: 6153)[†]; Non-cancer illness diagnosed by nurses during verbal interview (ID: 20002) [#]
BMI group	Underweight (<18.5); Healthy weight(18.5 to <25); Overweight (25 to <30); Obese(30+); Missing	BMI (ID: 21001)[‡]
Note: *Recruitment questions, [†]Touchscreen questions, [‡]Physical measurements, [§]24-h online dietary assessment questionnaire, [#]verbal interview		

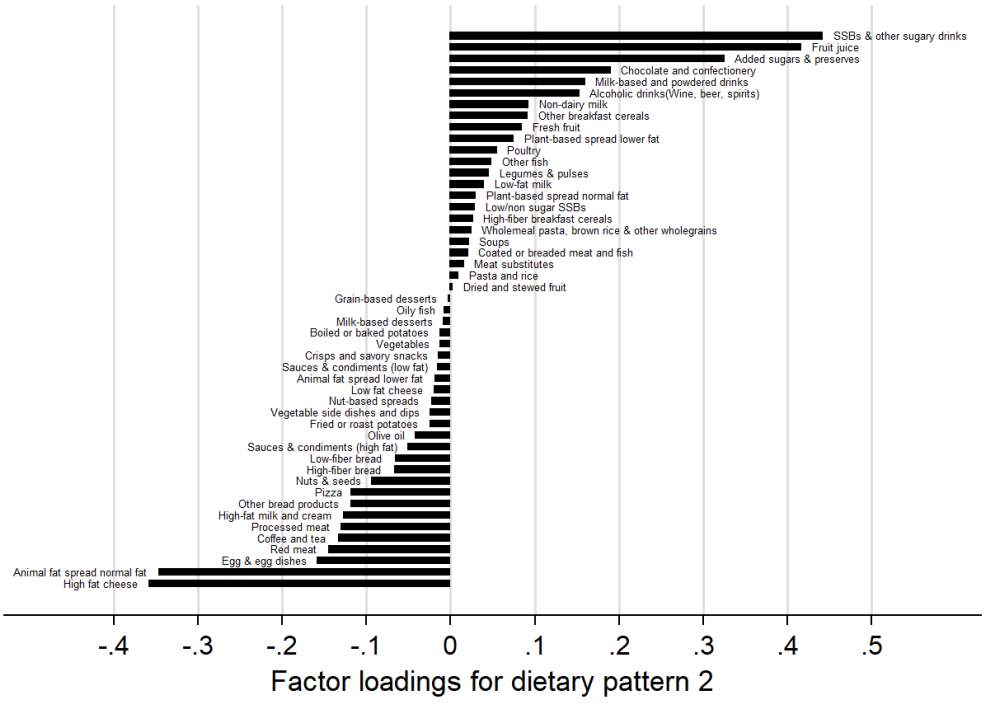
Supplementary Table 2 Explained variation (%) in food intakes and response variables for each dietary pattern as assessed using reduced rank regression and correlation coefficient between DPs and response variables (N= 120 343)

Dietary patterns	Explained variation (%)						Correlation coefficient			
	Food intakes (total)	Responses (total)	Energy density (kJ/g)	Saturated fats(%E)	Free Sugars (%E)	Fibre density (g/MJ)	Energy density (kJ/g)	Saturated fat(%E)	Free Sugars (%E)	Fibre density (g/MJ)
1	3.8	43	65	23	22	61	0.807***	0.416***	0.396***	-0.767***
2	2.0	20	1	22	54	0	-0.153***	-0.492***	0.678***	0.060***
3	2.5	10	5	23	7	5	-0.327***	0.370***	0.148***	0.372***
4	2.3	4	9	0	0	9	0.299***	0.027***	0.058***	0.264***

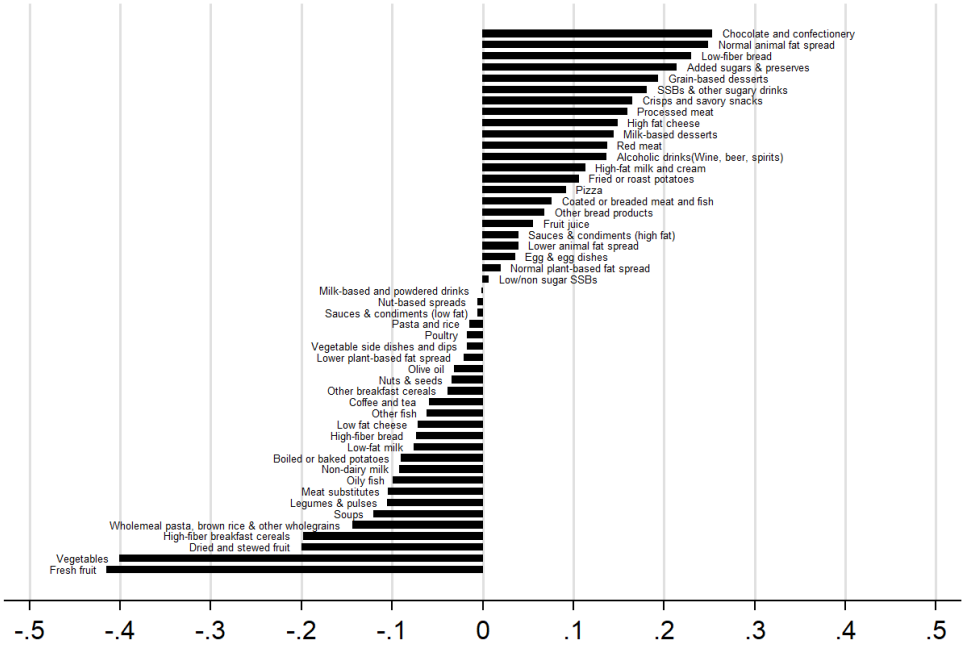
Note: %E, proportion of total energy intake. ***:P<0.001, **:P<0.01. The correlation between food groups and response variables are estimated with Pearson's r correlation coefficient.

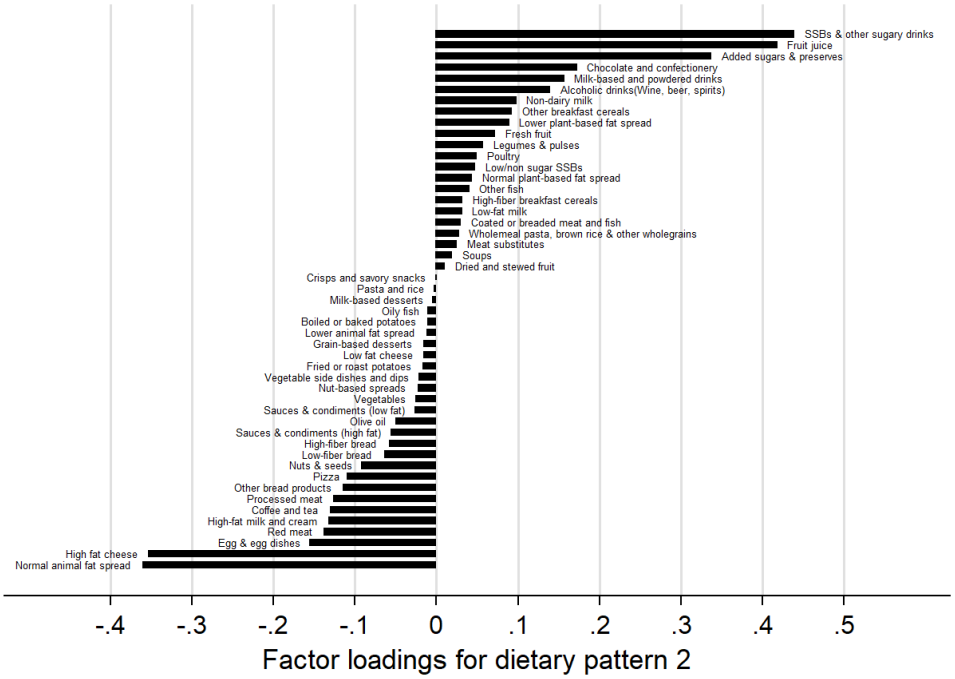
Supplementary Figure 2: Factor loadings for dietary patterns characterized by energy density (kJ/g), saturated fat (%total energy), free sugar (%total energy) and fibre density (g/MJ) calculated by using reduced rank regression among people with 3 or more 24-h online dietary assessments in the UK Biobank (N=75,003)



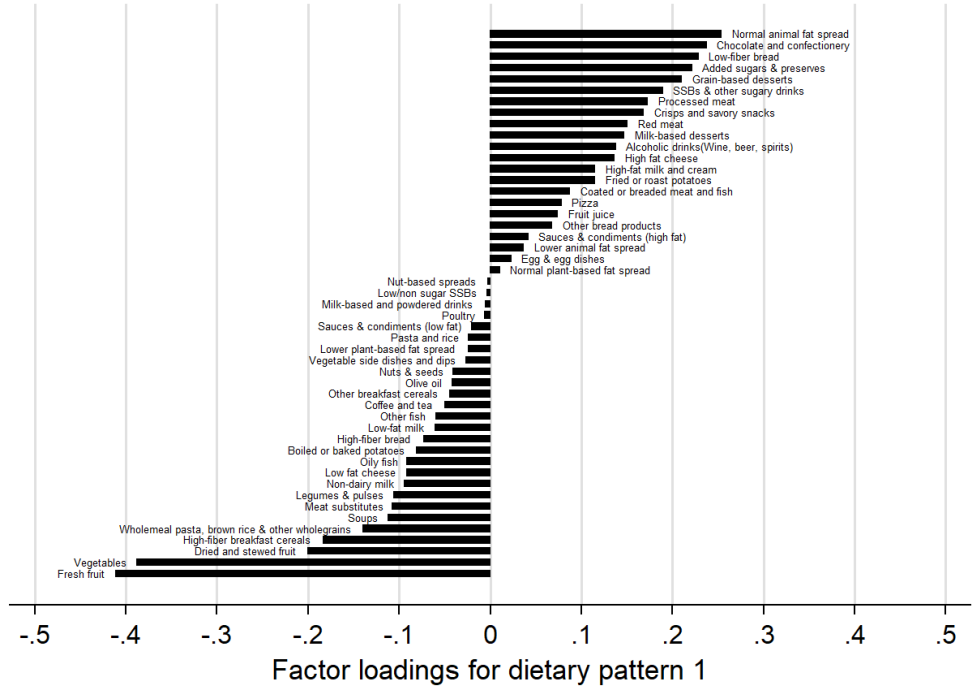


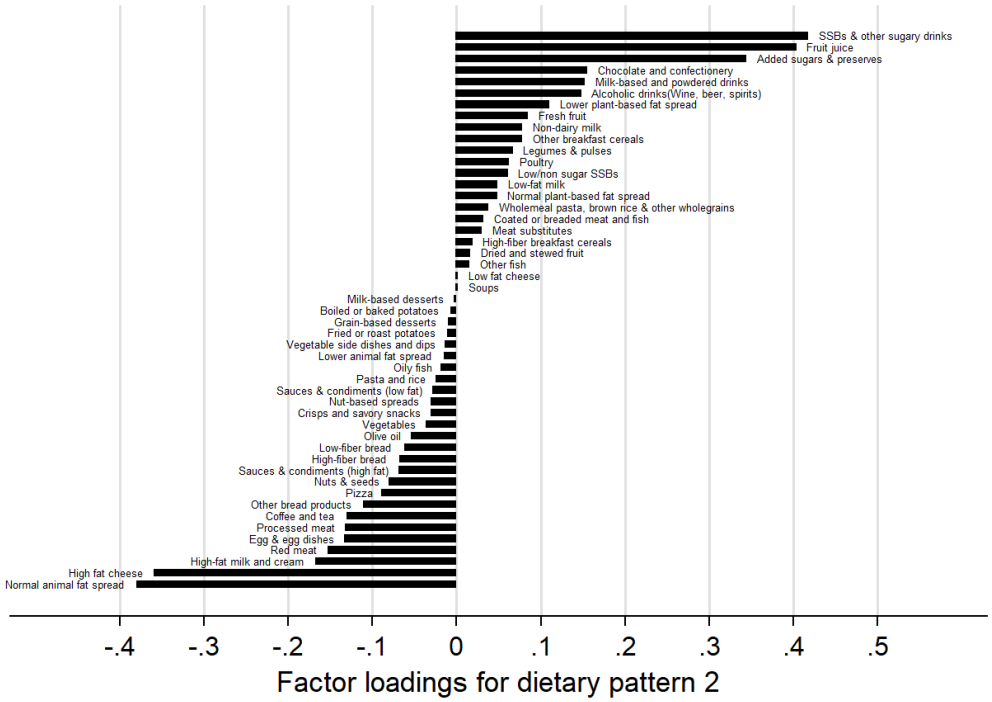
Supplementary Figure 3: Factor loadings for dietary patterns characterized by energy density (kJ/g), saturated fat (%total energy), free sugar (%total energy) and fibre density (g/MJ) calculated by using reduced rank regression among people with 4 or more 24-h online dietary assessments in the UK Biobank (N=34,644)


Chocolate and confectionery
Normal animal fat spread
Low-fiber bread
Added sugars & preserves
Grain-based desserts
SSBs & other sugary drinks
Crisps and savory snacks
Processed meat
High fat cheese
Milk-based desserts
Red meat
Alcoholic drinks(Wine, beer, spirits)
High-fat milk and cream
Fried or roast potatoes
Pizza
Coated or breaded meat and fish
Other bread products
Fruit juice
Sauces & condiments (high fat)
Lower animal fat spread
Egg & egg dishes
Normal plant-based fat spread
Low/non sugar SSBs
Milk-based and powdered drinks
Nut-based spreads
Sauces & condiments (low fat)
Pasta and rice
Poultry
Vegetable side dishes and dips
Lower plant-based fat spread
Olive oil
Nuts & seeds
Other breakfast cereals
Coffee and tea
Other fish
Low fat cheese
High-fiber bread
Low-fat milk
Boiled or baked potatoes
Non-dairy milk
Oily fish
Meat substitutes
Legumes & pulses
Soups
Wholemeal pasta, brown rice & other wholegrains
High-fiber breakfast cereals
Dried and stewed fruit
Vegetables
Fresh fruit
-.5
-.4
-.3
-.2
-.1
0
.1
.2
.3
.4
.5
Factor loadings for dietary pattern 1


SSBs & other sugary drinks
Fruit juice
Added sugars & preserves
Chocolate and confectionery
Milk-based and powdered drinks
Alcoholic drinks(Wine, beer, spirits)
Non-dairy milk
Other breakfast cereals
Lower plant-based fat spread
Fresh fruit
Legumes & pulses
Poultry
Low/non sugar SSBs
Normal plant-based fat spread
Other fish
High-fiber breakfast cereals
Low-fat milk
Coated or breaded meat and fish
Wholemeal pasta, brown rice & other wholegrains
Meat substitutes
Soups
Dried and stewed fruit
Crisps and savory snacks
Pasta and rice
Milk-based desserts
Oily fish
Boiled or baked potatoes
Lower animal fat spread
Grain-based desserts
Low fat cheese
Fried or roast potatoes
Vegetable side dishes and dips
Nut-based spreads
Vegetables
Sauces & condiments (low fat)
Olive oil
Sauces & condiments (high fat)
High-fiber bread
Low-fiber bread
Nuts & seeds
Pizza
Other bread products
Processed meat
Coffee and tea
High-fat milk and cream
Red meat
Egg & egg dishes
High fat cheese
Normal animal fat spread
-.4
-.3
-.2
-.1
0
.1
.2
.3
.4
.5
Factor loadings for dietary pattern 2

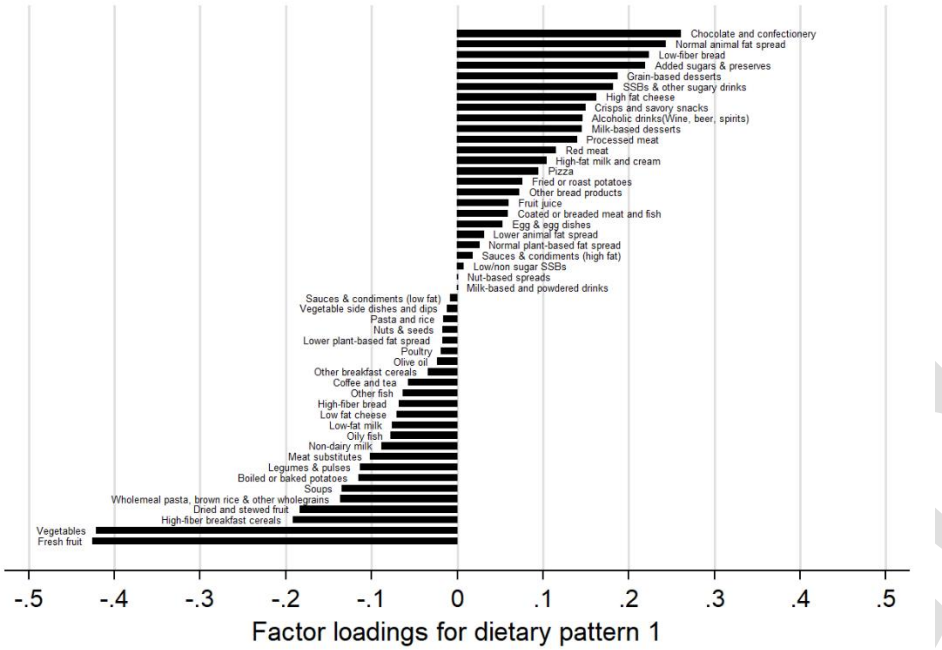
Supplementary Figure 4: Factor loadings for dietary patterns characterized by energy density (kJ/g), saturated fat (%total energy), free sugar (%total energy) and fibre density (g/MJ) calculated by using reduced rank regression among people with five 24-h online dietary assessments in the UK Biobank (N=5,504)

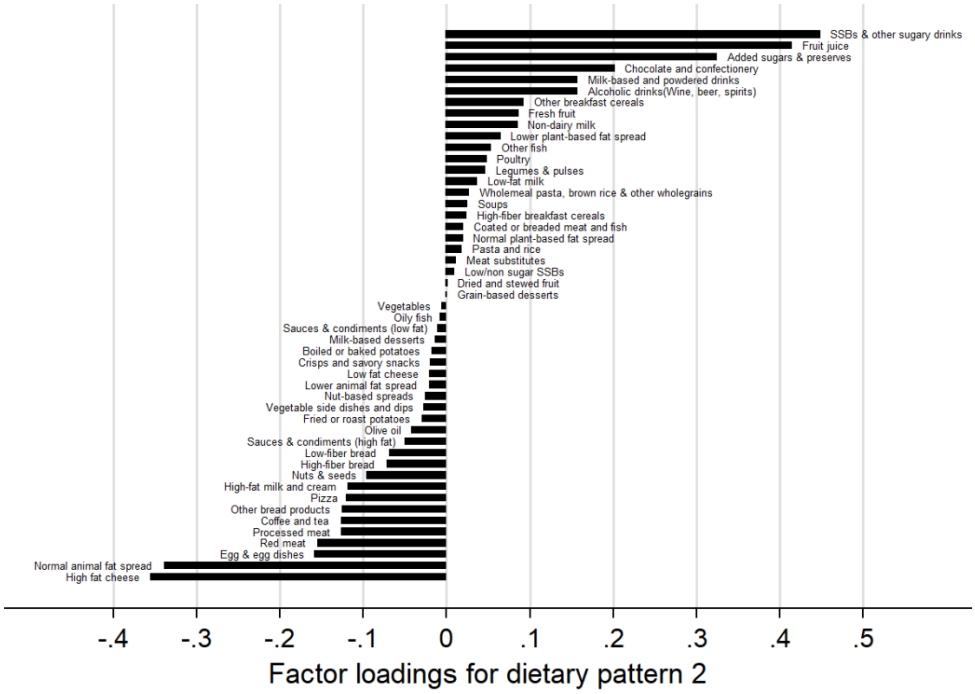


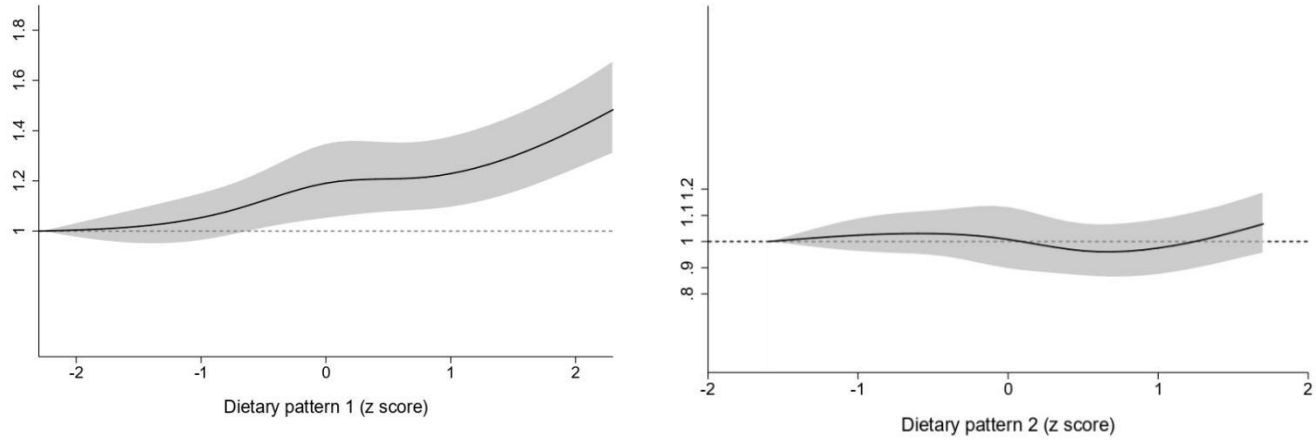


Supplementary Figure 5: Factor loadings for dietary patterns characterized by energy density (kJ/g), saturated fat (%total energy), free sugar (%total energy) and fibre density (g/MJ) calculated by using reduced rank

regression excluding people who had the event within two years after completing their last 24-h online dietary assessment in the UK Biobank (N= 119,870)







Supplementary Figure 6: HRs (95% CIs) of continuous dietary pattern scores for the risk of incident diabetes (n=120,343).

HRs (solid black lines) and 95% CIs (grey areas) were derived from spline regression models to examine the possible nonlinear relation of DP scores to diabetes events (stratified by sex and regions [England, Scotland and Wales] and adjusted for demographics, socioeconomic status, behavior risk factors, health history/conditions and energy intake). For simplicity of presentation, the reference values of DP scores were set to the 5% quantile of dietary pattern scores. Z scores for DP1 and DP2 were analyzed in mutually adjusted models to examine their independent associations with health outcomes.

Supplementary Table 3: Associations between dietary pattern scores and the risk of total diabetes cases excluding people who had the event within two years after completing their last 24-h online dietary assessment (N= 119,870)

Dietary pattern	HR	95% CI	
DP 1	1.10	1.07	1.13
DP 2	1.05	0.99	1.09
N of event		2405	

Notes: All models were stratified by sex and regions (England, Scotland and Wales). Multivariate models were further adjusted for demographics, socioeconomic status, behavior risk factors, health history/conditions and energy intake. Z scores for DP1 and DP2 were analyzed in mutually adjusted models to examine their independent associations with total diabetes.

Supplementary Table 4: Adjusted hazard ratios of total incident diabetes associated with each z score increase in dietary pattern (N= 120 343)

	Group	No. events	Final model HR(95% CI)	Final models plus nutrients adjusted HR(95% CI)
DP1	Total	2,878	1.09(1.06,1.12)	1.13(1.07,1.19)
	Q1	452	1.00	1.00
	Q2	513	1.13(1.04,1.23)	1.14(1.03,1.25)
	Q3	549	1.19(1.10,1.30)	1.19(1.10,1.29)
	Q4	621	1.25(1.16,1.35)	1.23(1.12,1.33)
	Q5	743	1.38(1.27,1.49)	1.30 (1.14,1.46)
	P for trend(p-value) §		<0.001	<0.001
DP2	Total	2,878	1.03(0.99,1.06)	1.03(0.99,1.08)
	Q1	610	1.00	1.00
	Q2	545	0.97(0.89,1.05)	0.99(0.91,1.08)
	Q3	567	1.01(0.93,1.10)	1.05(0.97,1.14)
	Q4	524	0.90(0.82,0.98)	0.95(0.87,1.04)
	Q5	632	1.04(0.96,1.12)	1.08(0.99,1.19)
	P for trend(p-value) §		0.818	0.217

Note: Adjusted hazard ratios (HR) and 95% confidence intervals (CI) of total DP scores obtained using Cox proportional hazard regression. Final model was stratified by sex and regions (England, Scotland and Wales) and adjusted for demographics, socioeconomic status, behavior risk factors, health history/conditions, BMI group, energy density (kJ/g), SFA (% total energy), free sugars (% total energy) and fiber density (g/MJ). Four loaded nutrients were adjusted for DP1, two most loaded nutrients (free sugars and saturated fat) were adjusted for DP2.