Supplementary

Supplementary Figure 1: Participant flow chart of the study

502 536 participants recruited between 2006 to 2010

Exclusion criteria

- People with diabetes events that occurred before baseline assessment (n=26,845)
    - Self-reported (n=26,400)
    - Take insulin ≥3 (n=97)
    - Nurse reported (n=348)
- Pregnancy (n=375)
- People without any validated 24-h online dietary assessment (n=273,434)
- People with one validated 24-h online dietary assessments (n=79,995)
- People with diabetes events that occurred before completing the last assessment(n=172)

Excluded samples with missing values
- Nutrients (n=21)

Excluded samples with implausible energy intake (Total=1,351):
- Under-reporters (n= 996)
- Over-reporters (n= 130)
- Without weight data (n= 225)

120,343 participants, including:
2,878 incident diabetes cases
(**2,624** T2DM cases
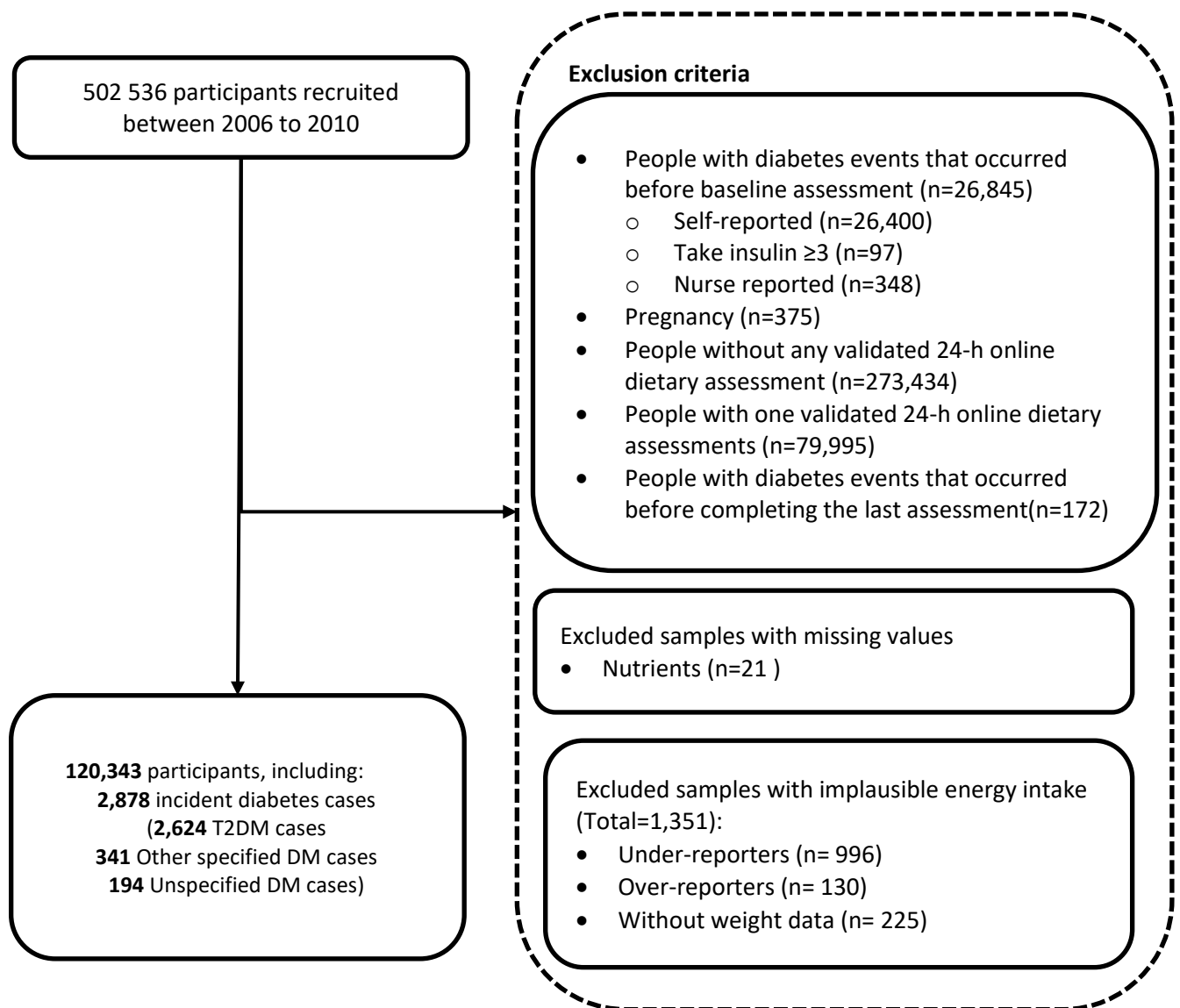**341** Other specified DM cases
194 Unspecified DM cases)

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	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

	(MET)-minutes per week); High (≥ 3000 metabolic equivalent (MET)-minutes per week); Missing	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)
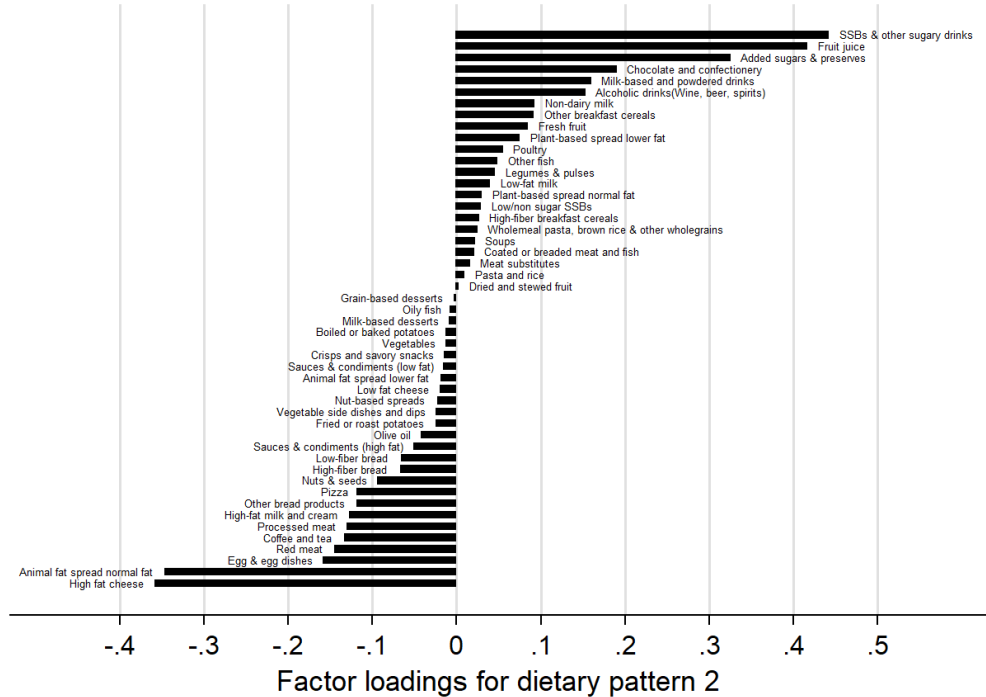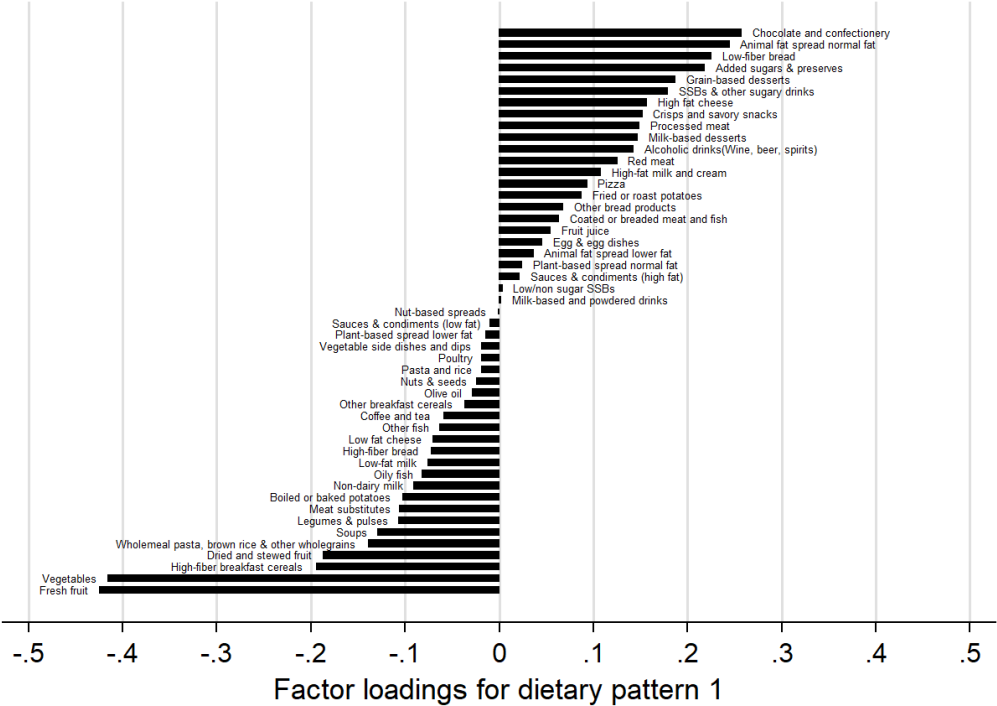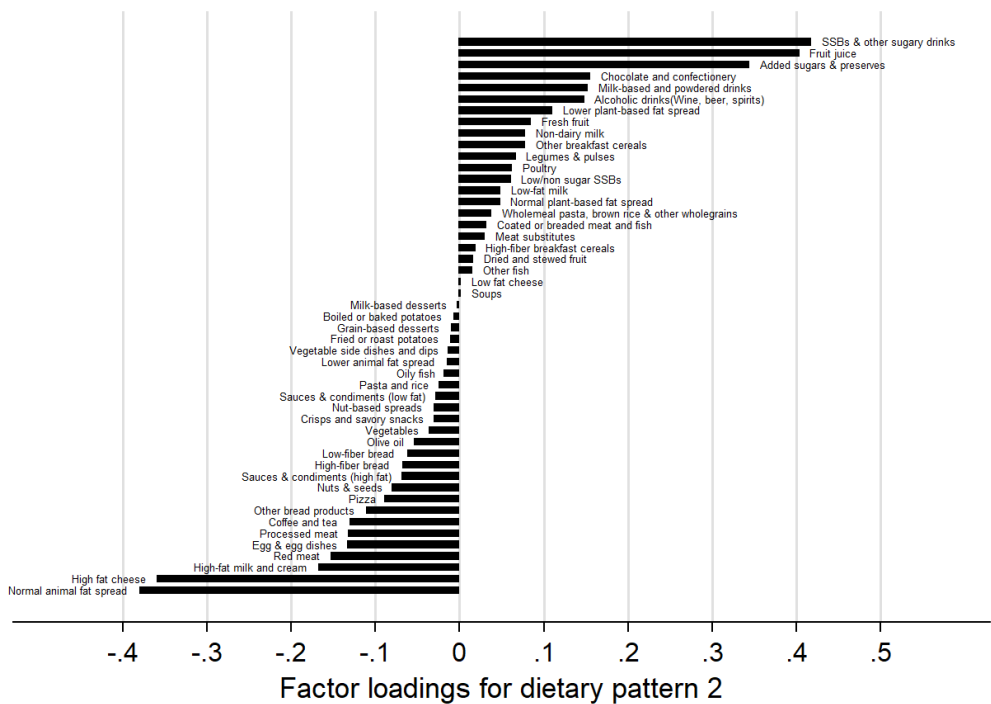
Supplementary Figure 3: Factor loadings for dietary patterns characterized by energy density (kJ/g), saturated fat (%total energy), free sugar (%total energy) and fibre density (g/MJ) calculated by using reduced rank regression among people with 4 or more 24-h online dietary assessments in the UK Biobank (N=34,644)
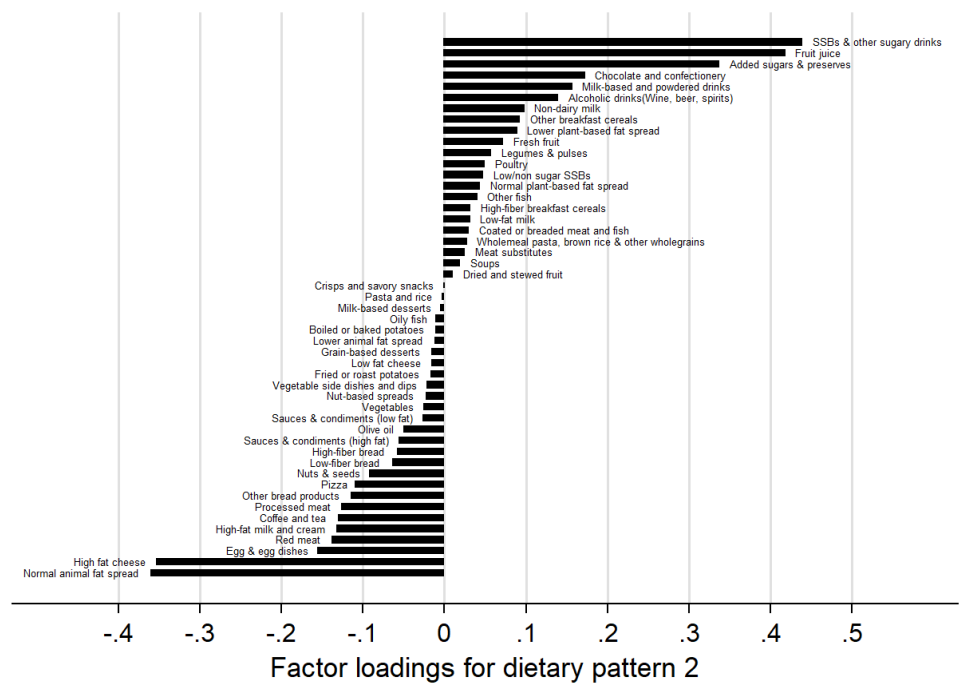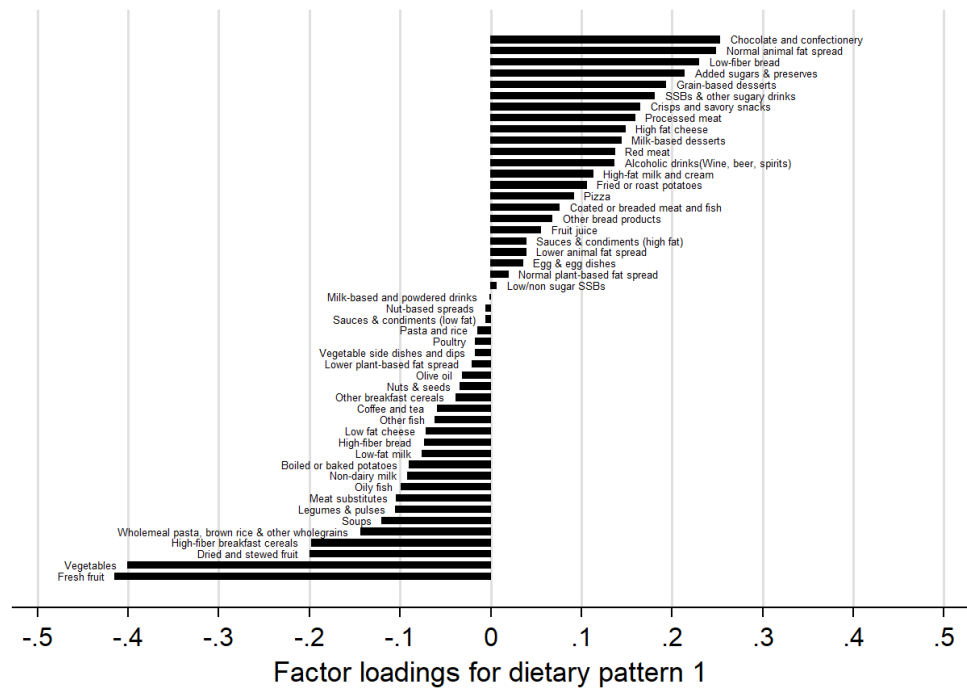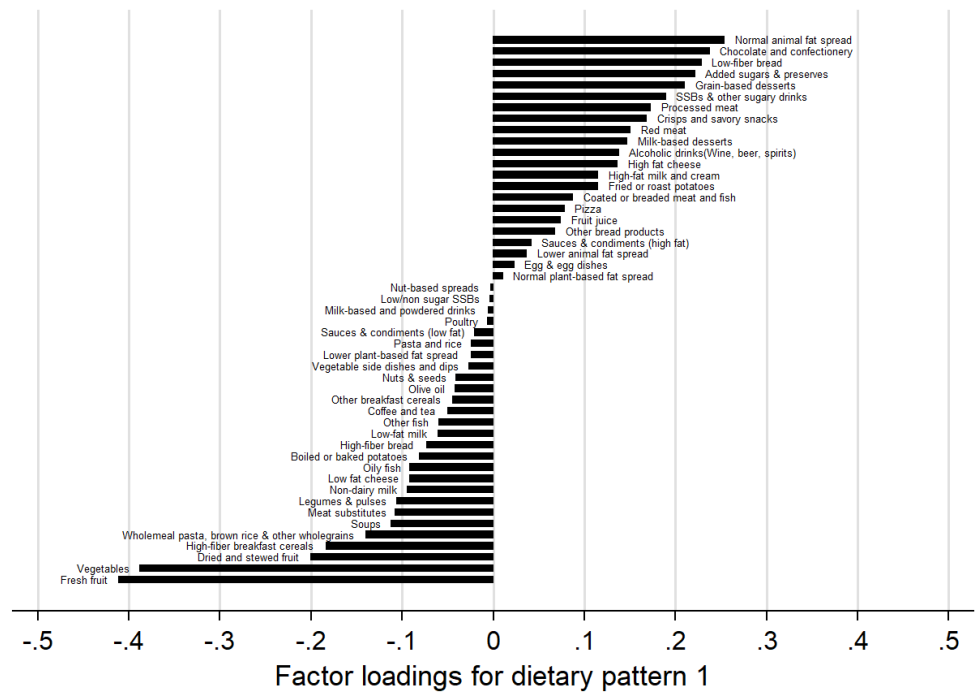
Supplementary Figure 4: Factor loadings for dietary patterns characterized by energy density (kJ/g), saturated fat (%total energy), free sugar (%total energy) and fibre density (g/MJ) calculated by using reduced rank regression among people with five 24-h online dietary assessments in the UK Biobank (N=5,504)
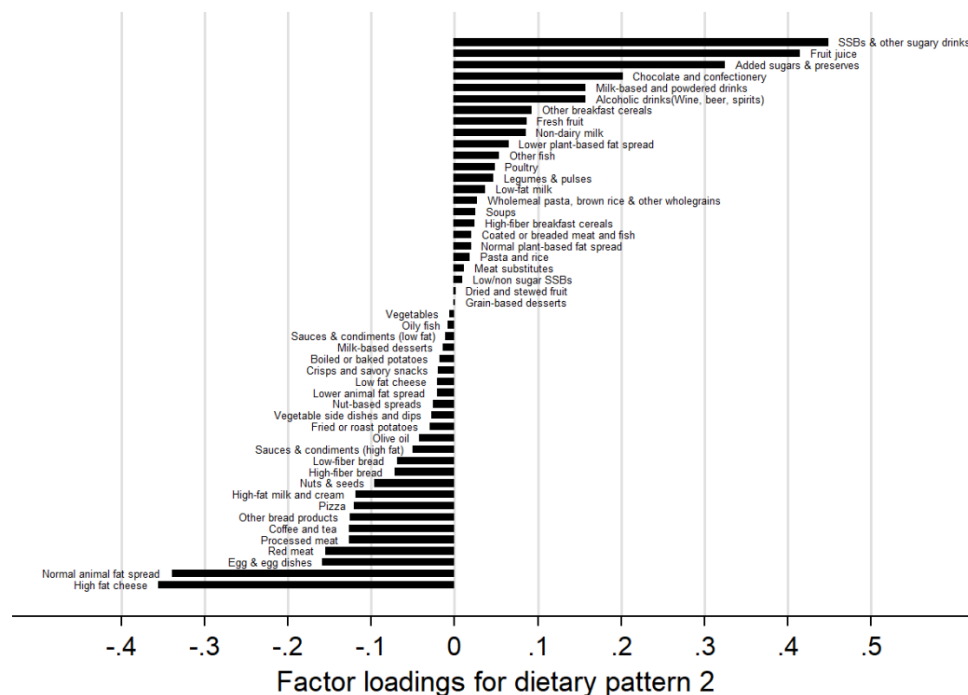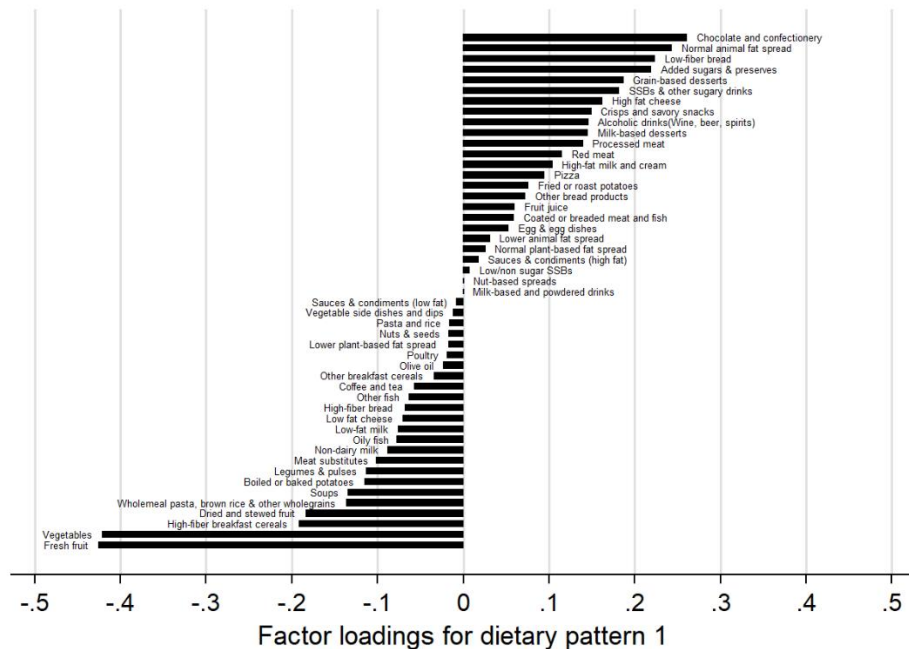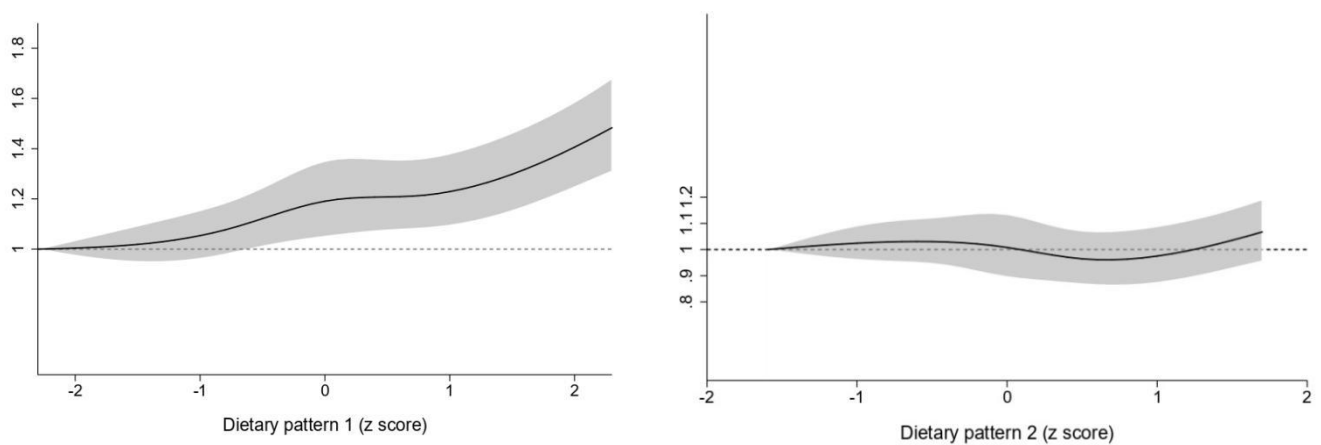
Supplementary Figure 5: Factor loadings for dietary patterns characterized by energy density (kJ/g), saturated fat (%total energy), free sugar (%total energy) and fibre density (g/MJ) calculated by using reduced rank regression excluding people who had the event within two years after completing their 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.