# The role of the Mediterranean diet in breast cancer survivorship

## Study 2: A Secondary data analysis : Data analysis plan

Authors: Ge Chen, Tom Dudding, Angeliki Papadaki and Sam Leary

## 1. Research Questions

1) Is adherence to the Mediterranean diet (MD) associated with quality of life (QoL) in breast cancer (BC) survivors?

2) Is adherence to the MD associated with BC survival (e.g. all-cause mortality)?

3) Is adherence to the MD associated with health-related outcomes (e g. cancer recurrence, obesity, cardiovascular disease (CVD) incidence) in BC survivors?

## 2. Data Description

### 2.1 Overview:

| | Study design (Time of baseline data) | Total sample size | Age of participants (baseline) | Included data | Notes |
|---|---|---|---|---|---|
| **UKB** | Prospective (2006-2010) | 500,000 | 40-69 | Questionnaires, Physical measures, Sample assays, Accelerometery, Multimodal imaging, Genome-wide genotyping<br><br>Longitudinal follow-up: Death and cancer registry | People living in UK<br><br>Repeat of baseline assessment (20,000-25,000 participants) |

### 2.2 Population

2.2.1 Inclusion criteria

Adult females diagnosed with breast cancer.

- Cancer registry
  ICD 10: C50 (Malignant neoplasm of breast), D05 (Carcinoma in situ of breast)
  ICD 9: 174 (Malignant neoplasm of female breast), 2330 (Carcinoma in situ of breast)

  | UKB variable code |
  |---|
  | 40006  Type of cancer: ICD10 |
  | 40013  Type of cancer: ICD9 |
  | 40011  Histology of cancer tumour |
  | 40012  Behaviour of cancer tumour |

2.2.2 Exclusion criteria

Participants with dietary information longer than 5 years following first BC diagnosis will be excluded.

1

2.2.3 Participant subgroups for sensitivity analysis

- Participants with dietary measurement before BC diagnosis and after BC diagnosis

**2.3 Exposure**

The exposure will be the adherence to MD, expressed as an MD score, by using MD scales.

2.3.1 Dietary data in UKB overview

| Measurement time | Baseline- at recruitment (2006-2010) and repeat | Online follow-up - 2009-2012 |
|---|---|---|
| Measurement method | How often during an <u>average week</u> they consume certain foods or drinks (FFQ) | 24-hour recall dietary questionnaire (Oxford WebQ: minimum 1 time point, maximum 5 time points) |

2.3.2 Transforming to MD score

2.3.2.1 Dietary sources:

The dietary data from food frequency questionnaire (FFQ) will be used in the main analysis. The data of 24 hours recalls in UKB will be used in sensitivity analysis

2.3.2.2 MD scales

The 9-item MD scale (MDS) will be used with FFQ in the main analysis (Trichopoulou et al., 1995, Trichopoulou et al., 2003) and with 24 hours recall data in sensitivity analysis. The PREDIMED scale (Martínez-González et al., 2012) will be used with 24 hours recall respectively in sensitivity analysis. This is because more participants will have FFQ data available, but 24 hour recall data includes more detailed descriptions of dietary intake (14-items).

2.3.2.3 Dietary data will be transformed to MD scores by using the following steps:

1) FFQ data will be quantified in grams per day by using standard portion size estimation, and energy intake will be calculated.

2) Residual models will be applied to calculate energy adjusted food intakes (Willett et al., 1997). Daily food intake from dietary intake  data will be adjusted to mean energy intake of study population ( kcal/d). For UKB 24 hours recall dietary data, energy (mean energy intake of study population kcal/d) will be adjusted for each time point of measurement, and then data will be averaged across time points.

3) Foods will be grouped into relevant categories according to MD scales (MDS and PREDIMED scale). Contributing foods from the FFQs and UKB 24 hour recalls for each category will be informed by published studies on MD and health outcomes by using UKB data  (Livingstone et al., 2021, Shannon et al., 2023).

Data from FFQs will be categorised in the following potential groups: vegetables, legumes, fruit and nuts, cereals, fish and seafood, monounsaturated/saturated fats ratio, dairy products meat and meat products, and alcohol  (Livingstone et al., 2021) (Appendix 1).

Data from UKB 24 hour recalls will be categorised in the following potential groups: olive oil, vegetables, fruit, read meat, butter or margarine or cream, sweetened or carbonated drinks, wine, legumes, seafood, sweets or pastries, nuts, white meat, sofrito (Shannon et al., 2023) (Appendix 2).

4) MD scores will be generated using the 9-item MD scale for FFQ and PREDIMED scale for FFQ and 24 hours recall respectively. For the 9-item MD scale, the median intakes of the food groups mentioned above will be used as the cut-off point for scoring (Trichopoulou et al., 1995, Trichopoulou et al., 2003). For the PREDIMED scale, points will be allocated at 0 for no consumption and 1 for the recommended intake by scoring system (Martínez-González et al., 2012).

### 2.3.3 MD score application

The MD scores from FFQs and 24 hours recall generated using MDS and PREDIMED scales (see 2.3.2.2) will be categorised into tertiles to represent low, medium and high levels of MD adherence.

## 2.4 Outcomes

### 2.4.1 Outcome data in UKB and CKB

| Category of outcome | Outcome in detail | UKB variable overview |
|---|---|---|
| Mortality | All-cause mortality; BC mortality; non-BC mortality | Death register<br>Including cause and date of death, and age at death |
| BC progression | Local recurrence; Metastasis; Newly diagnosed breast cancer | Cancer register<br>Including date , type, histology and behaviour of cancer diagnosis |
| Quality of life | | Including following questionnaires:<br>EQ-5D-5L (The 5-level EQ-5D) : Qol questionnaire<br>Depression : Patient Health Questionnaire-9 (PHQ-9)<br>Fatigue: Bespoke questions on Chronic Fatigue Syndrome, plus Fatigue Severity Scale (FSS) |
| Long-term/late effect of BC treatments<br><br>(after BC diagnosis) | Comorbidities (Osteoporosis/osteopenia/fracture; Obesity/cardiovascular disease/stroke); Endometrial cancer/other cancer); Cognitive impairment; | Comorbidities: Including the date of the diseases diagnosis by ICD 10 code:<br>Osteoporosis M80-M82<br>Cardiovascular disease/stroke: I42, I43, I50, I63, I64<br>Obesity: E66<br>Other cancer will be identified from cancer registry data by ICD 10 or ICD 9 code excluding codes for breast cancer (see 1.2.1)<br>Cognitive function: Including fluid intelligence score and number of incorrect matches in round |
| Health related parameters<br>(after BC diagnosis) | Anthropometric measurements | Including BMI and Waist circumference |
| | Biochemical parameters | Including triglycerides, glucose, HbA1c, HDL and LDL cholesterol, C-reactive protein, Cholesterol, Oestradiol, SHBG and Testosterone |

### 2.4.2 Outcome variable generation

### 2.4.2.1 Mortality

All death cases (from BC diagnosis to death) will be recorded as the outcome of all-cause mortality. Furthermore, cases will be categorised into BC mortality and non-BC mortality by using recorded ICD10 codes.

### 2.4.2.2 BC progression

BC progression includes BC recurrence, BC metastasis and second primary BC.

BC progression will be recorded if there is a second BC relevant diagnosis during UKB follow- up in included participants.

2.4.2.3 Quality of life

1) Overall quality of life based on EQ-5D-5L (The 5-level EQ-5D) questionnaire: It has been developed to describe and value health across a wide range of disease areas. Variables have been derived to represent the domains of mobility, self-care, usual activities, pain/discomfort and anxiety/depression, each with five levels of severity (EuroQol, 2019); Furthermore, the visual analogue scale (VAS) can be used as a measure of the overall health status.  Therefore six variables (five domains and one overall scale) will be used for analysis.

2) Depression based on PHQ-9 (Patient Health Questionnaire-9): The total score will be calculated from the nine items which were scored from 0-3, and categorised (minimal 0-4, mild 5-9, moderate 10-14, moderately severe 15-19, severe 20-27)(Kroenke et al., 2001) for analysis.

3) Fatigue based on bespoke questions on chronic fatigue syndrome and FSS (Fatigue Severity Scale): The five categorical (*Yes/No/Do not know/ Prefer not to answer*) bespoke questions on chronic fatigue syndrome will be used separately for analysis. The mean of the nine FSS questions (originally recorded as degree of agreement from 1-7 (1 indicates strongly disagree and 7 indicates strongly agree) will be calculated and used in the analysis (Krupp et al., 1989).

2.4.2.4 Long-term/late effect of BC treatments

For the outcomes of disease incidence (osteoporosis, cardiovascular disease, obesity and other cancer), date of disease diagnosis will be checked. The date of diagnosis after BC diagnosis and dietary measurement will be recorded and varied diagnosis of each disease will be combined into binary codes (1 Yes, 0 No) to represent the incidence in analysis.

Fluid intelligence score and number of incorrect matches in round will be used separately to reflect cognitive function, which will be applied as continuous variables in analysis.

2.4.2.5 Health-related parameters

The date of health related parameters (anthropometric measurements and biochemical parameters) will be checked first. The measurements taken after BC diagnosis will be used in analysis as continuous variables.

**2.5 Potential confounding variables**

**2.5.1 Overview of potential confounding variables**

| Confounder factor | UKB Variable code |
| --- | --- |
| **Age** | 21022   Age at BC diagnosis |
| **Ethnicity** | 21000   Ethnic background |
| **Education** | 6138    Qualifications |
| **Deprivation** | 189  Townsend deprivation index at recruitment |
| **Smoking status** | 20116   Smoking status |
| **Physical activity** | 22040   Summed MET minutes per week for all activity |
| **BMI**<br>(at baseline) | 21001   BMI |
| **Female specific factors** | 2724    Had menopause<br>2814    Ever used hormone-replacement therapy (HRT)<br>2784    Ever taken oral contraceptive pill |
| **Alcohol intake** | 20117   Alcohol drinker status |

| Total energy intake | 26002   Energy |
|---|---|
| **Cancer characteristics** | Cancer stage (in situ, primary and metastasis)<br>40011   Histology of cancer tumour<br>40012   Behaviour of cancer tumour |

2.5.2 Rationale for confounding and variable generation

2.5.2.1 Age

Rationale: mortality and BC risk increase with age (CRUK, 2023) , and healthy aging in the elderly is associated with MD adherence (Mazza et al., 2021).

Age at BC diagnosis will be used as a continuous variable in analysis.

2.5.2.2 Ethnicity

Rationale: Differences were found in BC/death incidence  (CRUK, 2023) and MD adherence across ethnicities (Sotos-Prieto and Mattei, 2018)

The broad ethnic groups of ethnicity in UKB will be used and the ethnic group 2021 Census (GOV.UK, 2021) is used to inform the generation of ethnicity data. The ethnicity groups applied in analysis are: White, Black (Black or Black British), Asian ( Asian and Asian British, Chinese), Mixed, Other and Unknown (Do not know/prefer not to answer)**.**

2.5.2.3 Social factors

Rationale: Social factors influence dietary intake and associate with health status (WHO, 2023)

Deprivation: Townsend deprivation Index was recorded as a continuous variable in UKB. It will be categorised as low (quintile 1), moderate (quintiles 2–4) high (quintile 5) to represent the level of deprivation.

Education: Qualification will be categorised as higher (college/university/ other professional qualification), vocational (NVQ/HND/ HNC), upper secondary (A-levels), lower secondary (O-levels/GCSEs /CSEs), none ( none of the listed) and unknown (prefer not to answer).

2.5.2.4 Smoking status

Rationale: Smoking increase risks of BC, mortality and relevant comorbidities (CRUK, 2023) and is less likely to be reported in people at high MD adherence (MartÍNez-GonzÁLez et al., 1997).

Smoking status will be categorised by original recorded categories: never, previous (past), current and unknown (prefer not to answer)

2.5.2.5 Physical activity (PA)

Rationale: PA and MD were reported have potential interactions and join association on mortality (Hershey et al., 2022).

PA will use standard cut-offs to categorise participants as meeting PA guidelines of 150 min per week if their metabolic equivalents (METs) were ≥600 MET-min/week (0-600 MET-min/week and ≥600 MET-min/week) (IPAQ, 2005).

2.5.2.6 Body mass index (BMI)

Rationale: Being overweight and obese associate with increased risk of BC (CRUK, 2023). MD interventions were reported associate with reductions in BMI (López-Gil et al., 2023)

WHO BMI cut-offs will be used to categorise participants as underweight (<18.5), healthy weight (18.5-25), overweight (>25 and<30), and Obese (≥30)

### 2.5.2.7 Cancer stage

Rationale: Cancer stage associates with survival and cancer treatment (CRUK, 2023).

Cancer stage will be categorised based on cancer behaviour code: Stage 0 (in situ), Stage I-III (primary site) and stage IV (metastatic site). It will be replaced by UKB newly released cancer stage data in the future.

### 2.5.2.8 Female specific factors

Rationale: These factors associate with BC risk (subtypes) and treatment due to changes in estrogen levels and estrogen receptor status (CRUK, 2023).  Furthermore, MD was reported potentially links to estrogen level (Carruba et al., 2006).

Menopausal status will apply the original recorded categorises: Yes, No, Not sure (had a hysterectomy), Not sure (other reason), Unknown

HRT use and Oral contraceptive pill use will also apply the original recorded categories:  Yes, No, Unknown

Multicollinearity will be checked before including these factors in the model.

### 2.5.2.9 Total energy intake

Rationale: Energy intake is an important indicator of PA and a determinant of body lean mass, which has been associated with dietary intake and disease risk (Willett et al., 1997)

Energy intake will be adjusted at the stage of generating MD score by applying the residual models mentioned above.

### 2.5.2.10 Alcohol intake

Rationale: Alcohol intake is associated with BC risk, and comorbidities (CRUK, 2023). It is also an item in the MD scale (Trichopoulou et al., 1995, Trichopoulou et al., 2003, Martínez-González et al., 2012).

Alcohol intake will be included in the generation of MD score in the main analysis. It will be excluded from the score and applied as a covariate (confounder) in the sensitivity analysis models. The originally recorded category of intake will be used: never, previous, current and unknown (prefer not to answer)

## 3. Analysis

## 3.1 Data cleaning and variable generation

The raw data will be cleaned, which will include checking variable types, names, labels, and missing value coding. Variables will be converted into their required type.

## 3.2 Descriptive statistics

Continuous variables will be described by means and standard deviations (SDs) if normally distributed or medians and interquartile ranges (IQRs) if non-normally distributed. Categorical

variables will be described by frequencies/ percentages. Both variable types will be described stratified by the level of adherence to MD (assessed by the three versions of the MD scale).

**3.3 Potential Models**

Model 1: Age (at BC diagnosis)

Model 2: Model 1 + Ethnicity + Social factors (education, deprivation) + BMI+ Smoking status + Physical activity (BMI will be removed when testing BMI as an outcome)

Model 3:  Model 2 + Female specific factors (menopausal status, HRT use, oral contraceptive use) + Cancer stage

**3.4 Regression analyses**

3.4.1 Multivariable regression

Linear/logistic regression models will be used to compare the associations between MD adherence and the outcomes (mortality, QoL, BC progression, health replated parameters )using the set of three models described above (see 3.3).

3.4.2 Survival analysis

Cox proportional hazards regression models will be used to compare hazards of events ( mortality and BC progression), using the set of three models described above (see 3.3).

**3.5 Sensitivity analysis**

3.5.1 Stratified analysis for participants with dietary measurement before BC diagnosis and after BC diagnosis

3.5.2 Stratified analysis in individuals with different menopausal status and cancer stage

3.5.3 All analysis will be repeated by using 24 hours recalls data and applying PREDIMED scale. As well as only including participants with minimum of two 24 hours recalls data to provide a more stringent measure of habitual dietary intake and excluding participants with extreme energy intakes (<600 or >3500 kcal/d ) respectively.

3.5.4 If time permits, repeat analyses following imputation of missing dietary and covariate data using multiple imputations by chained equations. All analytic variables (covariates and outcome data) and FFQ will be used where possible as predictors in the model.

3.5.5 If appropriate covariables will be further investigated to assess if they are effect modifiers

**References**

BRION, M. J., LAWLOR, D. A., MATIJASEVICH, A., HORTA, B., ANSELMI, L., ARAÚJO, C. L., MENEZES, A. M., VICTORA, C. G. & SMITH, G. D. 2011. What are the causal effects of breastfeeding on IQ, obesity and blood pressure? Evidence from comparing high-income with middle-income cohorts. *Int J Epidemiol,* 40**,** 670-80.
CARRUBA, G., GRANATA, O. M., PALA, V., CAMPISI, I., AGOSTARA, B., CUSIMANO, R., RAVAZZOLO, B. & TRAINA, A. 2006. A traditional Mediterranean diet decreases endogenous estrogens in healthy postmenopausal women. *Nutr Cancer,* 56**,** 253-9.

CRUK. 2023. *Breast cancer* [Online]. Cancer reseach UK Available:
    https://www.cancerresearchuk.org/about-cancer/breast-cancer [Accessed 11th September
    2023].

EUROQOL. 2019. *EQ-5D User Guides* [Online]. EuroQol Reseach Fundation. Available:
    https://euroqol.org/publications/user-guides/ [Accessed 11th September 2023].

GOV.UK. 2021. *List of ethnic groups* [Online]. Available: https://www.ethnicity-facts-
    figures.service.gov.uk/style-guide/ethnic-groups [Accessed 14th September 2023].

HERSHEY, M. S., MARTÍNEZ-GONZÁLEZ, M. Á., ÁLVAREZ-ÁLVAREZ, I., MARTÍNEZ HERNÁNDEZ, J. A. &
    RUIZ-CANELA, M. 2022. The Mediterranean diet and physical activity: better together than
    apart for the prevention of premature mortality. *British Journal of Nutrition,* 128**,** 1413-1424.

IPAQ. 2005. *IPAQ scoring protocol - International physical activity questionnaire* [Online]. The IPAQ
    Group. Available:
    https://sites.google.com/site/sites/system/errors/WebspaceNotFound?path=%2Ftheipaq%2
    F [Accessed 14th September 2023].

KROENKE, K., SPITZER, R. L. & WILLIAMS, J. B. 2001. The PHQ-9: validity of a brief depression severity
    measure. *J Gen Intern Med,* 16**,** 606-13.

KRUPP, L. B., LAROCCA, N. G., MUIR-NASH, J. & STEINBERG, A. D. 1989. The Fatigue Severity Scale:
    Application to Patients With Multiple Sclerosis and Systemic Lupus Erythematosus. *Archives
    of Neurology,* 46**,** 1121-1123.

LIVINGSTONE, K. M., ABBOTT, G., BOWE, S. J., WARD, J., MILTE, C. & MCNAUGHTON, S. A. 2021. Diet
    quality indices, genetic risk and risk of cardiovascular disease and mortality: a longitudinal
    analysis of 77 004 UK Biobank participants. *BMJ Open,* 11**,** e045362.

LÓPEZ-GIL, J. F., GARCÍA-HERMOSO, A., SOTOS-PRIETO, M., CAVERO-REDONDO, I., MARTÍNEZ-
    VIZCAÍNO, V. & KALES, S. N. 2023. Mediterranean Diet-Based Interventions to Improve
    Anthropometric and Obesity Indicators in Children and Adolescents: A Systematic Review
    with Meta-Analysis of Randomized Controlled Trials. *Advances in Nutrition,* 14**,** 858-869.

MARTÍNEZ-GONZÁLEZ, M. A., GARCÍA-ARELLANO, A., TOLEDO, E., SALAS-SALVADÓ, J., BUIL-COSIALES,
    P., CORELLA, D., COVAS, M. I., SCHRÖDER, H., ARÓS, F., GÓMEZ-GRACIA, E., FIOL, M., RUIZ-
    GUTIÉRREZ, V., LAPETRA, J., LAMUELA-RAVENTOS, R. M., SERRA-MAJEM, L., PINTÓ, X.,
    MUÑOZ, M. A., WÄRNBERG, J., ROS, E. & ESTRUCH, R. 2012. A 14-item Mediterranean diet
    assessment tool and obesity indexes among high-risk subjects: the PREDIMED trial. *PLoS
    One,* 7**,** e43134.

MARTÍNEZ-GONZÁLEZ, M. Á., PÉEREZ-GUTIÉRREZ, R., MARTÍNEZ-GONZÁLEZ, J., GARCÍA-MARTÍN, M.
    & BUENO-CAVANILLAS, A. 1997. Dietary intake of some food items in smokers and non-
    smokers in a Mediterranean population. *European Journal of Public Health,* 7**,** 40-44.

MAZZA, E., FERRO, Y., PUJIA, R., MARE, R., MAUROTTI, S., MONTALCINI, T. & PUJIA, A. 2021.
    Mediterranean Diet In Healthy Aging. *J Nutr Health Aging,* 25**,** 1076-1083.

SHANNON, O. M., RANSON, J. M., GREGORY, S., MACPHERSON, H., MILTE, C., LENTJES, M.,
    MULLIGAN, A., MCEVOY, C., GRIFFITHS, A., MATU, J., HILL, T. R., ADAMSON, A., SIERVO, M.,
    MINIHANE, A. M., MUNIZ-TERERRA, G., RITCHIE, C., MATHERS, J. C., LLEWELLYN, D. J. &
    STEVENSON, E. 2023. Mediterranean diet adherence is associated with lower dementia risk,
    independent of genetic predisposition: findings from the UK Biobank prospective cohort
    study. *BMC Med,* 21**,** 81.

SOTOS-PRIETO, M. & MATTEI, J. 2018. Mediterranean Diet and Cardiometabolic Diseases in
    Racial/Ethnic Minority Populations in the United States. *Nutrients,* 10.

TRICHOPOULOU, A., COSTACOU, T., BAMIA, C. & TRICHOPOULOS, D. 2003. Adherence to a
    Mediterranean Diet and Survival in a Greek Population. *New England Journal of Medicine,*
    348**,** 2599-2608.

TRICHOPOULOU, A., KOURIS-BLAZOS, A., WAHLQVIST, M. L., GNARDELLIS, C., LAGIOU, P.,
    POLYCHRONOPOULOS, E., VASSILAKOU, T., LIPWORTH, L. & TRICHOPOULOS, D. 1995. Diet
    and overall survival in elderly people. *BMJ,* 311**,** 1457.

WHO. 2023. *Social determinants of health* [Online]. World Health Organization.  [Accessed 14th September 2023].

WILLETT, W. C., HOWE, G. R. & KUSHI, L. H. 1997. Adjustment for total energy intake in epidemiologic studies. *Am J Clin Nutr,* 65**,** 1220S-1228S; discussion 1229S-1231S.

**Appendix 1 Components and scoring of using MDS (Livingstone et al., 2021)**

| Dietary Indictor | Indicator foods[1] | Criteria for scoring |
|---|---|---|
| 1. Vegetables (excluding potatoes, legumes or fruit juice) | Mixed vegetable, vegetable pieces, avocado, beetroot, broccoli, butternut squash, cabbage/kale, carrot, cauliflower, celery, courgette, cucumber, garlic, leek, lettuce, mushroom, onion, olives, parsnip, pea, side salad, sweet pepper, spinach Sprouts, sweetcorn, fresh tomato, tinned tomato, green bean, turnip/swede, watercress, other vegetables, homemade soup (vegetables) | Sex-specific median intakes used as cut points. Intakes (for indictors 1-6) above median score 1 and intakes below the median score 0. |
| 2. Legumes | Baked bean, pulses, broad bean, homemade soup (pulses) | |
| 3. Fruit and nuts | Stewed fruit, prune, dried fruit, mixed fruit, apple, banana, berry, cherry, grapefruit, grape, mango, melon, orange, satsuma, peach/nectarine, pear intake, pineapple, plum, other fruit Orange juice, grapefruit juice, pure fruit/vegetable juice Unsalted peanuts, unsalted nuts, types of spreads/sauces consumed (Peanut butter), seeds | |
| 4. Cereals | Porridge, muesli, oat crunch, plain cereal, bran cereal, whole-wheat cereal, other cereal Bread consumed, sliced bread (mixed; wholemeal; seeded; other), baguette (mixed; wholemeal; seeded; other), bap (mixed; wholemeal; seeded; other), bread roll (mixed; wholemeal; seeded; other), other bread White pasta, wholemeal pasta, white rice, brown rice, couscous, other grain Homemade soup, ingredients in homemade soup (pasta) | |
| 5. Fish and seafood | Tinned tuna, oily fish, white fish, prawns, lobster/crab, shellfish, other fish Homemade soup, ingredients in homemade soup (fish) | |
| 6. Monounsaturated/ saturated fats ratio | Monounsaturated fats, saturated fats | |
| 7. Dairy products | Milk, milk added to cereal Low fat hard cheese, low fat cheese spread, cottage cheese Yogurt (low fat yogurt consumer; full fat yogurt consumer) Goat's cheese, hard cheese, soft cheese, blue cheese, cheese spread, feta, mozzarella, other cheese Dairy smoothie, latte, added milk to instant coffee, added milk to filtered coffee, added milk to espresso, added milk to other coffee type, added milk to standard tea, added milk to rooibos tea, cappuccino | Sex-specific median intakes used as cut points. Intakes (for indictors 7-8) below median score 1 and intakes below the median score 0. |
| 8. Meat and meat products | Beef, pork, lamb, other meat Whole egg, omelette, eggs in sandwiches, scotch egg, other egg Homemade soup, ingredients in homemade soup (meat), sausage, bacon, ham | |

| | | |
|---|---|---|
| 9. Alcohol | Red wine, rose wine, white wine<br>Beer/cider<br>Fortified wine, spirits intake, other alcohol | No more than 2 drinks/day = 1; Never drink or over 2 drinks/day = 0. |

1. As available in the UK Biobank

## Appendix 2 Components and scoring of using PREDIMED scale (Shannon et al., 2023)

| Food component | Contributing foods from the Oxford WebQ | MEDAS[1] | |
|---|---|---|---|
| | | Servings required for 0 points | Servings required for 1 point |
| Olive oil[5] | Type of fat/ oil used for cooking (20090) | Non-consumption | Consumption |
| Vegetables[3, 7] | Carrot (104170), spinach (104300), broccoli (104140), cabbage/ kale (104160), sprouts (104310), courgette (104200), cauliflower (104180), parsnip (104270), turnip/ swede (104360), leek (104230), onion (104260), garlic (104220), mushroom (104250), sweet pepper (104290), side salad (104090), lettuce (104240), cucumber (104210), celery (104190), watercress (104370), fresh tomato (104340), tin tomato (104350), sweetcorn (104320), beetroot (104130), avocado (104100), mixed vegetables (104060), vegetable pieces (104070), butternut squash (104150), other veg (104380), olives (102490), coleslaw (104080), guacamole (20088), vegetables from canned soup (102540, 20108), vegetables from homemade soup (102620, 20109) | <2/d (and/or not including 1/d raw or salad) | ≥2/d (including ≥1/d raw or salad) |
| Fruit[3] | Apple (104450), pear (104560), orange (104530), satsuma (104540), grapefruit (104490), banana (104460), grape (104500), melon (104520), peach/ nectarine (104550), plum (104580), berry (104470), dried fruit (104430), stewed fruit (104410), mixed fruit (104440), prune (104420), cherry (104480), mango (104510), pineapple (104570), other fruit (104590), fruit added to cereal (100880), grapefruit juice (100200), orange juice (100190) | <3/d | ≥3/d |
| Red meat[4] | Beef (103020), pork (103030), lamb (103040), red meat from canned soup (102540, 20108), red meat from homemade soup (102620, 20109), bacon (103070), ham (103080), sausage (103010), liver (103090), meat from Scotch egg (103090) | >1/d | <1/d |
| Butter, margarine or cream[4] | Butter/ margarine on potato (104040), baguettes with butter/ margarine (101350, 20099), baps with butter/ margarine (101390, 20100), bread rolls with butter/ margarine (101430, 20101), bread slices with butter/ margarine (101310, 20098), crackers/ crispbread with butter/ margarine (101470, 20102), oatcakes with butter/ margarine (101510, 20103), other bread with butter/ margarine (101550, 20104), butter/ margarine used in cooking (20090), cream (20088) | >1/d | <1/d |
| Sweetened or carbonated drinks[4] | Fizzy drinks (100170), low calories drinks (100160), squash intake (100180) | >1/d | <1/d |
| Wine[3] | Red wine (100590, 20096), rose wine (100630, 20097), white wine (100670, 20095) | <7/wk | ≥7/wk |
| Legumes[3] | Peas (104280), green beans (104120), broad beans (104110), baked beans (104000), pulses (104010), tofu (103270), hummus (20088), pulses from canned soup (102540, 20108), pulses from homemade soup (102620, 20109) | <3/wk | ≥3/wk |
| Seafood[3] | Battered fish (103180), breaded fish (103170), white fish (103190), oily fish (103160), shellfish (103220), other fish (103230), tinned tuna (103150), prawn (103200), lobster/ crab (103210), fish from canned soup (102540, 20108), fish from homemade soup (102620, 20109) | <3/wk | ≥3/wk |

| | | | |
|---|---|---|---|
| Sweets or pastries[4] | Chocolate biscuit (102350), chocolate covered biscuit (102340), chocolate bar (102260), chocolate sweets (102310), chocolate raisins (102300), dark chocolate (102290), milk chocolate (102280), white chocolate (102270), sweet biscuits (102360), cakes (102190), cheesecake (102220), doughnut (102200), fruitcake (102180), Danish pastry (102060), sponge pudding (102210), milk based pudding (102140), other milk based pudding (102150), other desert intake (102230), soya desert intake (102170), sweets (102330), diet sweets (102320), other sweets (102380), ice cream (102120) | >2/wk | <2/wk |
| Nuts[3] | Unsalted nuts (102440), salted nuts (102430), unsalted peanuts (102420), salted peanuts (102410), peanut butter (20088) | <3/wk | ≥3/wk |
| White meat[6] | Poultry (103060), breaded poultry (103050), white meat from canned soup (102540, 20108), white meat from homemade soup (102620, 20109) | Less white meat than red meat | More white meat than red meat |
| Sofrito[3] | Tomato-based sauce (20088) | <2/wk | ≥2/wk |

[1]Scoring for the MEDAS scale was calculated as previously described (29). A score of 1 point was awarded if participants achieved a dietary target. Otherwise, participants were awarded 0 points.

[2]Scoring for the MEDAS continuous scale was carried out as described by Shannon et al. (17). This used the same dietary components as the standard MEDAS scale. However, rather than awarding points were awarded continuously between 0 and 1 based on linear equation principles.

[3]A high intake of vegetables, fruit, wine, legumes, seafood, nuts, and sofrito was recommended. For the MEDAS continuous scale, points were allocated between 0 for no consumption and 1 for meeting the recommended intake.

[4]A low intake of red meat, butter, margarine or cream, sweetened or carbonated drinks, and sweets or pastries was recommended. For the MEDAS continuous scale, points were allocated continuously between 0 points for double the recommended intake and 1 point for below the recommended intake.

[5]For olive oil individuals who reported consumption received a score of 1 point, whilst non-consumers received 0 points. As it was not possible to quantify amount of olive oil consumed, only 1 point was allocated for olive oil intake whereas the traditional MEDAS score awards points a) using olive oil as the primary cooking fat and b) for a total olive oil intake >50 ml/d.

[6]For white meat, participants were awarded a point if the total amount of white meat consumed exceeded red meat consumption.

[7]A maximum score of 0.5 points was awarded for participants who did not also consume 1 serving per day of raw vegetables or salad as part of the MEDAS continuous scale. Conversely, 0 points were awarded for participants who did not consume 1 serving per day of raw vegetables or salad as part of the MEDAS scale, irrespective of their total vegetable intake.