Statistical Analysis Plan v4
Diet and risk of hip fracture in the UK Women's Cohort Study

James Webster, Darren C Greenwood, and Janet E Cade

Oct 4th, 2021

NCT number:

1. The UK Women's Cohort

The dataset will use participant data from the pre-existing UK Women's Cohort Study (UKWCS). The cohort profile has been previously published [1].

2. Data cleaning

Potential outliers will be identified by considering their distance from the mean and their biological plausibility (e.g., energy intake < 500 or > 5000 kcal/day, or BMI < 10 kg/m^2); participants with identified outliers will be excluded. Missing data will be coded consistently throughout the dataset. All data cleaning will be documented for transparency.

2.1 Exclusion criteria

Participants will be excluded from analyses for any of the following reasons:

- Not a resident of England
- Unable to link dietary and lifestyle data with hospital episode data (e.g. no NHS number provided and unable to match with records)
- Unreliable or missing nutrient or covariate data (e.g., >= 20% of food frequencies missing)
- Had a hip fracture or osteoporosis before or on the date of recruitment

3. Exposures

3.1 Primary and secondary exposures

The primary exposures of interest are data-derived dietary group (regular meat-eater, occasional meat-eater, pescatarian, vegetarian, or vegan – see Table 1 for definitions) and dietary intake of foods, beverages, and nutrients, including: 1) fruits and vegetables; 2) animal products including meat, fish, eggs and dairy products; 3) tea and coffee, and 4) nutrients associated with hip fracture in previously published studies (protein, calcium, and vitamin D).

Secondary exposures of interest include self-identified dietary group (meat-eater, pescatarian, vegetarian, or vegan), types of meat (unprocessed red, all processed meat, and unprocessed poultry) and dairy products (milk, yoghurt, cheese, cream, and dairy desserts), caffeinated and decaffeinated coffee, and nutrients with a plausible association with hip fracture (such as phosphorus intake). All relevant exposures and their derivation are listed in Supplementary Table 1.

Table 1: Data-derived dietary group categories and definitions.

Dietary group	Definition
Regular meat-eater	Total meat intake ≥ 0.8 servings/day

Occasional meat-eater	Total meat intake < 0.8 servings/day & > 0.02 servings/day
Pescatarian	Total meat intake ≤ 0.02 servings/day & total fish intake > 0.02 servings/day
Vegetarian	Total meat and fish intakes ≤ 0.02 servings/day, intake of any dairy products or eggs > 0.02 servings/day
Vegan	Total meat, total fish, dairy products, and eggs intake ≤ 0.02 servings/day

Participants with consumption frequencies ≤ 0.02 servings/day (less than once a month) will be considered non-consumers.

3.2 Exposure assessment

Dietary information of the cohort was ascertained by a self-administered 217-item food frequency questionnaire (FFQ) [1]. The FFQ was validated by comparison with 4-d weighed food diaries and a repeated FFQ on 283 women three years after baseline [1].

The baseline questionnaire asked if participants identify as a meat-eater, vegetarian, or vegan – this will constitute the self-identified dietary group variable. It also included questions about consumption of foods and beverages in the form of 'how often do you eat [specific food or beverage]?' or similar. Ten responses were possible: 0 'never', 1 '< once per month', 2 '1-3 per month', 3 'once per week', 4 '2-4 per week', 5 '5-6 per week', 6 'once per day', 7 '2-3 per day', 8 '4-5 per day', or 9 '6+ per day'. We will convert the responses to these questions into daily-based consumption frequencies as follows: 0, 0.02, 0.07, 0.14, 0.4, 0.8, 1, 2.5, 4.5, 6 times per day. Each exposure variable of interest will then be calculated in servings per day by summing daily consumptions of relevant items. For example, we will sum daily consumptions of unprocessed beef, pork, and lamb into one group titled 'red meat'; questions on unprocessed chicken or turkey into 'unprocessed poultry'; and questions around processed red or white meat or poultry into 'processed meat'. Red meat, unprocessed poultry, and processed meat will be summed to form 'total meat'. Similarly, questions on fish intake will be summed to calculate oily fish, non-oily fish, and total fish intakes; and questions on intake of dairy products will be summed to calculate milk, yogurt, cheese, cream, dairy desserts, and total dairy intakes in servings per day. Meat, fish, eggs and dairy exposures will then be used to classify subjects as regular meat-eaters, occasional meat-eaters, pescatarians, vegetarians, or vegans (Table 1). Due to the small number of vegans, vegetarians and vegans will be combined in the main analysis and considered separately in sensitivity analyses.

To explore potential associations between foods and beverages and hip fracture incidence, food and beverage intake variables in grams per day will be calculated by multiplying servings per day intake variables by standard portion weights, and summing relevant FFQ items. Food and beverage intake variables (grams/day) will be used to explore the effects of increment increases in intake on hip fracture risk, and to categorise participants into non-consumers (if > 1000 participants in this category) and levels of intake for all exposure variables, ensuring equal distribution of the population size among categories where possible. For intakes of milk, tea, coffee, and tea and coffee combined, increment increases and categories of intake will be defined by glasses and cups per day using standard sizes with the lowest exposure category as the reference group. For fruits, vegetables, and fruits and vegetables combined, categories will be defined to compare intakes below and above the recommended five-a-day (400 g/day) intake, which will be used as the reference category. All other food and beverage exposures will be split into thirds or quarters.

Nutrient variables will be derived by multiplying food and beverage intakes (g/day) by nutrient concentrations in each food or beverage item using McCance and Widdowson's Food Composition database (5th edition) [2].

4. Outcomes

Participants will be followed up via record linkage to the NHS Hospital Episode Statistics up to 31 March 2019. The primary outcome will be hip fracture incidence (International Classification of Diseases, ICD-9 code 820, and ICD-10 codes S72-S72.2). Hip fracture cases will also be identified by hip replacements (ICD-10 code Z96.64). The timeframe will be person years until hip fracture incidence (or until end of study period or death in those without a hip fracture), calculated as age at time of event (or non-event) minus age at study entry.

5. Sample size

The minimum detectable hazard ratio for the potential association between each primary exposure and hip fracture was estimated assuming a sample size of 26,300 (after applying exclusion criteria), an overall probability of hip fracture incidence of 3%, $p < 0.05$, and 80% power. For the dietary group analysis, a theoretical standard deviation of the exposure was calculated based on the percentage of participants in each category (Table 2). The minimum detectable effect of other primary exposures was calculated per standard portion size increase in intake using standard deviations of each exposure from the data in hand. Standard portion sizes were estimated from the Foods Standards Agency and previous UKWCS analyses [3, 4]. A standard portion size for total animal products was calculated by averaging standard servings of meat, fish, eggs, and dairy products.

Table 2: Minimal detectable hazard ratios for associations between primary exposures and hip fracture risk.

Exposure	SD	Minimum detectable HR
Dietary group (vegetarians vs non-vegetarians)	0.36	1.32
Fruit and vegetables (per 80 g increase)	300 g	0.98
Fruits (per 80 g increase)	227 g	0.97
Vegetables (per 80 g increase)	139 g	0.95
Total animal products (per 120 g increase)	248 g	0.95
Total meat (per 150 g increase)	81.2 g	0.84
Total fish (per 140 g increase)	28.3 g	0.61
Eggs (per 88 g increase)	18.6 g	0.63
Total dairy (per 105 g increase)	214 g	0.95

Tea and coffee (per cup increase)	2.17 cups	0.96
Tea (per cup increase)	2.01 cups	0.95
Coffee (per cup increase)	1.65 cups	1.06
Protein (per 25 g increase)	26.3 g	0.91
Calcium (per 300 mg increase)	365 mg	0.91
Vitamin D (per µg increase)	1.67 µg	0.94

1 cup of tea or coffee = 260 ml. SD = Standard deviation. HR = Hazard ratio. The SD for dietary group is theoretical and has arbitrary units.

6. Descriptive statistics

Baseline socio-demographic, anthropometric, diet, and lifestyle characteristics of the cohort were collected via questionnaire at baseline, and will be summarised (e.g. presenting their means and standard deviations) by: 1) dietary group (both for data-derived and self-identified dietary group); 2) total fruit and vegetable consumption; 3) total animal product consumption; 4) total tea and coffee consumption, and 5) protein, calcium, and vitamin D intakes, to report any differences in covariates (including co-exposures) between groups at baseline.

7. Statistical modelling

7.1 Main analyses

We will fit Cox proportional hazard regression models to estimate the hazard ratios (HR) and 95% confidence intervals (95% CI) for associations between 1) the dietary groups and hip fracture incidence, with regular meat-eaters as the reference group; and 2) between intake of: fruits, vegetables, fruits and vegetables combined, total animal products, meat, fish, dairy, eggs, tea, coffee, tea and coffee combined, total protein, calcium, and vitamin D and hip fracture incidence. Each primary food, beverage, and nutrient exposure variable will be modelled as both categorical and continuous so that models can be fit comparing the risk of hip fracture between categories of intakes (where the target estimand is the relative causal effect of the exposure on hip fracture risk) and per increment increase in exposure intake (where the target estimand is the total causal effect of the exposure on hip fracture risk). We will also assess HR's for hip fracture as a function of primary exposure intake (non-linear dose-response) via restricted cubic splines for intake of: fruits, vegetables, fruits and vegetables combined, tea, coffee, and tea and coffee combined, since previous studies have shown non-linear relationships between each of these variables and hip fracture risk [5, 6]. Proportional hazards assumptions will be assessed based on Schoenfeld residuals.

7.1.1 Accounting for confounding

For all primary analyses, we will present age-adjusted, minimally adjusted, and fully adjusted models. Minimally adjusted models will be adjusted for the minimally sufficient set of confounders. Three directed acyclic graphs (DAGs) will be constructed using the online tool DAGitty, and following available guidelines on the creation and reporting of DAG models to determine minimal adjustment sets in models of associations between hip fracture and 1) dietary patterns; 2) foods and beverages; and 3) nutrient intakes [7]. Confounders will be considered as covariates that are 1) risk factors of the outcome; 2) associated with the exposure; and 3) not caused by the exposure [8]. Confounders common to all research questions will likely include: age at baseline, ethnicity, socio-economic status, education, marriage, physical activity, smoking, alcohol, BMI (or height and weight), number of children, menopausal status, chronic disease at baseline, and fracture at baseline (non-hip). A list of likely confounders and their derivation is summarised in Supplementary Table 2. We will adjust for baseline chronic disease and fracture prevalence at sites besides the hip rather than excluding these subjects in minimally adjusted models to reduce the risk of selection bias that would result from excluding unhealthier subjects. Fully adjusted models will be further adjusted for competing exposures.

Energy intake will be adjusted for using the 'all-components' method for all analyses except dietary groups, for which energy intake could plausibly be a mediator, thus will remain unadjusted [9]. This method involves adjusting for all other individual components of energy intake besides the exposure at hand, and has been shown to produce the least biased and most precise estimand for relative and total causal effects [9]. For food and beverage exposures of interest, this will involve mutual adjustment for other major foods and beverages or food and beverage groups that might be associated with hip fracture incidence. For energy-contributing nutrient exposures of interest (macronutrients), this will involve mutual adjustment for other energy-contributing nutrients (e.g., saturated and unsaturated fat). For non-energy contributing nutrient exposures (micronutrients), this will involve adjustment for all energy-contributing nutrients, whilst avoiding over-adjustment as a result of high correlation between nutrients.

To estimate the population impact of diet on hip fracture incidence, we will determine the absolute risk difference for hip fracture between dietary groups and categories of exposure intakes, respectively. Predicted incidences and absolute risk differences will be calculated per 1000 person years and per 1000 people over 10 years using hazard ratios and 95% confidence intervals expressed as floating absolute risks [10].

7.2 Subgroup analyses

7.2.1 Dietary group

We will stratify all models by age (<60, > 60), BMI (<23.5, >23.5), SES (routine/manual, intermediate, professional/managerial), education (≥ A level, ≤ O level), smoking status (current, former/never), alcohol consumption (< 1/week, ≥ 1/week), menopausal status (pre, post), physical activity levels (inactive/low, moderate/high), and use of dietary supplements (yes/no) to determine their roles as potential effect modifiers. Effect modifiers will be defined as any covariate that meets the criteria for a confounder (previously defined) and could plausibly modify the association at varying levels of that covariate. Potential effect modifiers will also be added to all models as interaction terms. Models stratified by calcium, vitamin D, and vitamin B12 supplement use will be presented in those with phase 2 questionnaire data only, since this was not collected at baseline.

7.2.2 Foods, beverages and nutrients

We will fit additional Cox models for types of meat (unprocessed red, all processed, and unprocessed poultry), fish (oily and non-oily), dairy (milk, yoghurt, cheese, cream, and dairy desserts), and coffee (caffeinated and decaffeinated) intakes, treating each exposure intake as both categorical and continuous.

We will stratify models for combined fruits and vegetables consumption, combined tea and coffee consumption, and protein, calcium, and vitamin D intakes by age, BMI, menopausal status, and use of dietary supplements to determine their roles as potential effect modifiers, and will add these variables to fully adjusted models as interaction terms. The same cut-off points will be used as in the vegetarian diets analysis (section 3.2). The effect of protein intake on hip fracture will be further stratified by physical activity and calcium intake. We will also present models for all dietary exposures stratified by calcium and vitamin D supplement use in those with phase 2 questionnaire data only.

7.3 Sensitivity analyses

For the dietary group analysis, we will present models with vegans and vegetarians separated. We will also compare risk estimates between self-identified and data-derived dietary groups. For all analyses, we will explore the potential mediating effect of covariates identified in DAG models as being associated with hip fracture incidence and plausibly caused by the exposure by adding each of these covariates to adjusted models independently. Models will be presented with and without adjustment for BMI. To reduce risk of bias due to changes in exposure over time, phase two dietary data (collected four years after baseline) will be compared with baseline dietary data. In cases of substantial exposure change over time (e.g., > 10% of participants changing their dietary group or primary exposure categories), models will be repeated using phase two data if there is sufficient power.

Additional sensitivity analyses will involve excluding cases occurring in the first five years of follow-up (to check for reverse causality), and excluding participants on long-term treatment for illnesses. All analyses will be performed using Stata. Two-sided p values < 0.05 will be considered statistically significant. We acknowledge that the large number of planned exposure-outcome association tests increases the risk of a type-2 error. However, the alpha value will not be restricted below $p < 0.05$ since secondary analyses (including most nutrient-hip fracture associations) will be exploratory, with only the primary analyses considered hypothesis-testing. All analyses performed and effect sizes computed will be presented in resulting manuscripts to avoid bias in selection of the reported result.

8. Missing data

Participants with missing data for an FFQ item will be considered non-consumers. Participants will be excluded from each specific analysis if they have missing data for any other variable required in that analysis. Techniques to impute missing dietary or covariate data will not be performed.

9. Timescales

Target analysis completion date (excluding manuscript preparations): Jan 2021.

10. References

1. Cade JE, Burley VJ, Alwan NA, Hutchinson J, Hancock N, Morris MA, et al. Cohort Profile: The UK Women’s Cohort Study (UKWCS). Int J Epidemiol. 2015;46(2):e11-e. doi: 10.1093/ije/dyv173

2. Holland B, Welch AA, Unwin ID, Buss DH, Paul AA, Southgate DA. McCance and Widdowson's The Composition of Foods. Cambridge, UK: Royal Society of Chemistry and Ministry of Agriculture, Fisheries and Food, 1991.

3. Food Standards Agency (2002) Food Portion Sizes. London: The Stationery Office.

4. Dunneram Y, Greenwood DC, Cade JE. Diet and risk of breast, endometrial and ovarian cancer: UK Women’s Cohort Study. British Journal of Nutrition. 2019;122(5):564-74.

5. Byberg L, Bellavia A, Orsini N, Wolk A, Michaëlsson K. Fruit and vegetable intake and risk of hip fracture: a cohort study of Swedish men and women. Journal of Bone and Mineral Research. 2015;30(6):976-84.

6. Sheng J, Qu X, Zhang X, Zhai Z, Li H, Liu X, Liu G, Zhu Z, Hao Y, Qin A, Dai K. Coffee, tea, and the risk of hip fracture: a meta-analysis. Osteoporosis International. 2014 Jan;25(1):141-50.

7. Tennant PW, Murray EJ, Arnold KF, Berrie L, Fox MP, Gadd SC, Harrison WJ, Keeble C, Ranker LR, Textor J, Tomova GD. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. International journal of epidemiology. 2021;50(2):620-32.

8. McNamee R. Confounding and confounders. Occupational and environmental medicine. 2003;60(3):227-34.

9. Tomova GD, Arnold KF, Gilthorpe MS, Tennant PW. Adjustment for energy intake in nutritional research: a causal inference perspective. medRxiv. 2021.

10. Tong TY, Appleby PN, Bradbury KE, Perez-Cornago A, Travis RC, Clarke R, Key TJ. Risks of ischaemic heart disease and stroke in meat eaters, fish eaters, and vegetarians over 18 years of follow-up: results from the prospective EPIC-Oxford study. bmj. 2019;366.

11. Supplementary Material

Supplementary Figure 1: Equation used to calculate the theoretical standard deviation for the dietary group analysis.

$$SD = [p(1-p)]^{0.5}$$

Where SD = standard deviation and p = percentage of participants in the vegetarian category (as a decimal).

Supplementary Table 1: Dietary exposures and their derivation.

Dietary intake variable (categorical or continuous)	Derivation
1. *Fruits and vegetables*	
Fruits (cat)	Categorise fruits (con) as: 0 - 240 g, > 240 - 320 g, > 320 - 400 g, > 400 - 560 g, and > 560 g
Fruits (con)	Convert all specific fruit intake variables to servings/day, multiply by a standard portion weight to give g/day of each specific fruit, then sum to give g/day of total fruit intake
Vegetables (cat)	Categorise vegetables (con) as: 0 – 160 g, > 160 – 240 g, > 240 – 320 g, > 320 – 400 g, and > 400 g
Vegetables (con)	Convert all specific vegetable intake variables to servings/day then multiply by a standard portion weight to give g/day of each specific vegetable, then sum to give g/day of total vegetable intake
Fruits and vegetables (cat)	Categorise fruits and vegetables (con) as: 0 – 240 g, > 240 – 400 g, > 400 – 560 g, > 560 – 720 g, > 720 g
Fruits and vegetables (con)	Sum fruits (g/day) and vegetables (g/day)
2. *Animal products*	
Total animal products (cat)	Categorise total animal products (con) into non-consumers and quartiles of intake
Total animal products (con)	Sum total meat, total fish, eggs, and total dairy (all in g/day)
Total meat (cat)	Categorise total meat (con) into non-consumers and quartiles of intake
Total meat (con)	Sum red meat, processed meat, and unprocessed poultry (all in g/day)
Red meat (cat)	Categorise red meat (con) into non-consumers and quartiles of intake
Red meat (con)	Convert beef, pork and lamb variables to servings/day, multiply each by standard portion weights to give g/day of each variable, then sum to give g/day of total red meat intake

Processed meat (cat)	Categorise processed meat (con) into non-consumers and quartiles of intake
Processed meat (con)	Convert all processed meat variables (including processed poultry) to servings/day, multiply each by standard portion weights to give g/day of each variable, then sum to give g/day of total processed meat intake
Unprocessed poultry (cat)	Categorise unprocessed poultry (con) into non-consumers and quartiles of intake
Unprocessed poultry (con)	Convert unprocessed poultry variables (chicken and turkey) to servings/day, multiply each by standard portion weights to give g/day of each variable, then sum to give g/day of total unprocessed poultry intake
Total fish (cat)	Categorise total fish (con) into non-consumers and quartiles of intake
Total fish (con)	Sum oily and non-oily fish intakes (g/day)
Oily fish (cat)	Categorise oily fish (con) into non-consumers and quartiles of intake
Oily fish (con)	Convert oily fish variables to servings/day, multiply each by standard portion weights to give g/day of each variable, then sum to give g/day of total oily fish intake
Non-oily fish (cat)	Categorise non-oily fish (con) into non-consumers and quartiles of intake
Non-oily fish (con)	Convert non-oily fish variables to servings/day, multiply each by standard portion weights to give g/day of each variable, then sum to give g/day of total non-oily fish intake
Eggs (cat)	Categorise eggs (con) into non-consumers and quartiles of intake
Eggs (con)	Convert egg variables to servings/day, multiply each by standard portion weights to give g/day of each variable, then sum to give g/day of total egg intake
Total dairy (cat)	Categorise total dairy (con) into non-consumers and quartiles of intake
Total dairy (con)	Sum milk, yogurt, cheese and cream (all in g/day)
Milk (dairy-based) (cat)	Categorise milk (con) into four categories of intake (based on glasses/day) depending on the distribution of data.
Milk (dairy-based) (con)	Sum intakes of different types of milk (in ml/day) to total milk (ml/day)
Yoghurt (cat)	Categorise yoghurt (con) into non-consumers and quartiles of intake
Yoghurt (con)	Convert yoghurt (cat) from servings/week to servings/day, multiply by standard portion weight to give g/day of yoghurt
Cheese (cat)	Categorise cheese (con) into non-consumers and quartiles of intake

Cheese (con)	Convert cheese (cat) from servings/week to servings/day, multiply by standard portion weight to give g/day of cheese
Cream (cat)	Categorise cream (con) into non-consumers and quartiles of intake
Cream (con)	Convert cream (cat) from servings/week to servings/day, multiply by standard portion weight to give g/day of cream
3. Tea and coffee	
Tea (cat)	Categorise tea (con) into 0 – 3, > 3 – 5, > 5 – 7, and > 7 cups/day
Tea (con)	Convert tea (cat) from servings/week to servings/day
Total coffee (cat)	Categorise total coffee (con) into four categories of intake (based on cups/day) depending on the distribution of data.
Total coffee (con)	Sum caff coffee (con) and decaff coffee (con) to give servings/day of total coffee
Caff coffee (cat)	Categorise caff coffee (con) into four categories of intake (based on cups/day) depending on the distribution of data.
Caff coffee (con)	Convert caff coffee (cat) from servings/week to servings/day
Decaff coffee (cat)	Categorise decaff coffee (con) into four categories of intake (based on cups/day) depending on the distribution of data.
Decaff coffee (con)	Convert decaff coffee (cat) from servings/week to servings/day
Tea and coffee (cat)	Categorise tea and coffee (con) into 0 – 3, > 3 – 5, > 5 – 7, and > 7 cups/day
Tea and coffee (con)	Sum tea (con) and coffee (con) to give servings/day of tea and coffee combined
4. Nutrients	
Energy (con)	Multiply intake of each item on the FFQ (grams/day) by the energy content of that item (kcal), and sum
Total protein (cat)	Categorise total protein (con) into non-consumers and quartiles of intake
Total protein (con)	Multiply intake of each item on the FFQ (grams/day) by the protein content of that item, and sum to give protein (g/day)
Animal protein (con)	Multiply intake of each item on the FFQ (grams/day) by the animal protein content of that item, and sum
Vegetable protein (con)	Multiply intake of each item on the FFQ (grams/day) by the vegetable protein content of that item, and sum
Animal : vegetable protein ratio (con)	Divide animal protein (g/day) by vegetable protein (g/day)
Dairy protein (con)	Multiply intake of each item on the FFQ (grams/day) by the dairy protein content of that item, and sum
Non-dairy animal protein (con)	Animal protein (con) – dairy protein (con)

Soy protein (con)	Multiply intake of each item on the FFQ (grams/day) by the soy protein content of that item, and sum
Non-soy vegetable protein (con)	Vegetable protein (con) – soy protein (con)
Fibre (con)	Multiply intake of each item on the FFQ (grams/day) by the fibre content of that item, and sum
Fat (con)	Multiply intake of each item on the FFQ (grams/day) by the fat content of that item, and sum
SFA (con)	Multiply intake of each item on the FFQ (grams/day) by the saturated fat content of that item, and sum
MUFA (con)	Multiply intake of each item on the FFQ (grams/day) by the monounsaturated fat content of that item, and sum
PUFA (con)	Multiply intake of each item on the FFQ (grams/day) by the polyunsaturated fat content of that item, and sum
Heme iron (con)	Multiply intake of each item on the FFQ (grams/day) by the heme iron content of that item, and sum
Non-heme iron (con)	Multiply intake of each item on the FFQ (grams/day) by the non-heme content of that item, and sum
Calcium (con)	Multiply intake of each item on the FFQ (grams/day) by the calcium content of that item, and sum
Folate (con)	Multiply intake of each item on the FFQ (grams/day) by the folate content of that item, and sum
Sodium (con)	Multiply intake of each item on the FFQ (grams/day) by the sodium content of that item, and sum
Vitamin B1 (con)	Multiply intake of each item on the FFQ (grams/day) by the vitamin B1 content of that item, and sum
Vitamin B2 (con)	Multiply intake of each item on the FFQ (grams/day) by the vitamin B2 content of that item, and sum
Vitamin B6 (con)	Multiply intake of each item on the FFQ (grams/day) by the vitamin B6 content of that item, and sum
Vitamin B12 (con)	Multiply intake of each item on the FFQ (grams/day) by the vitamin B12 content of that item, and sum
Vitamin C (con)	Multiply intake of each item on the FFQ (grams/day) by the vitamin C content of that item, and sum
Vitamin D (con)	Multiply intake of each item on the FFQ (grams/day) by the vitamin D content of that item, and sum
Zinc	Multiply intake of each item on the FFQ (grams/day) by the zinc content of that item, and sum
Phosphorus	Multiply intake of each item on the FFQ (grams/day) by the phosphorus content of that item, and sum
Magnesium	Multiply intake of each item on the FFQ (grams/day) by the magnesium content of that item, and sum

Selenium	Multiply intake of each item on the FFQ (grams/day) by the selenium content of that item, and sum
Potassium	Multiply intake of each item on the FFQ (grams/day) by the potassium content of that item, and sum

Supplementary Table 2: Covariates and their derivation.

Covariate	How the variable was derived
Socio-demographic variables	
Age at baseline	Calculated as year differences between date of birth and date of recruitment and was considered a continuous variable in adjustment sets.
Ethnicity	Participants were asked to select which ethnic group they belong to of 'white, 'Bangladeshi', 'Indian', 'Chinese', 'Pakistani', 'Black-Caribbean', 'Black – other', 'other'. We will regroup ethnicity into 'White', 'Asian', 'Black', and 'Other'.
SES	Participants were asked about their occupation. Options were 'never had paid job', 'managers and administrators', 'professional', 'technical and associate professional', 'clerical and secretarial', 'craft and skilled', 'personal and protective', 'sales', 'plant and machine operatives', or 'other'. We will condense these options into 'routine/manual', 'intermediate', or 'managerial/professional'.
Education	Participants were asked what their highest educational qualification was. Options were 'no qualifications', 'O level', 'A level', 'degree', or 'missing'.
Marriage	Participants were asked 'what is your marital status?' with options of 'married or living as married', 'divorced', 'widowed', 'single', or 'separated'. We combined 'divorced' and 'separated' together, and 'widowed' and 'single' together.
Lifestyle and other variables	
Physical activity	Participants were asked how long they perform exercises that makes them sweat per week (in hours and minutes per week). This will be computed into hours per day, and categorised as 'inactive/low' and 'moderate/high'.
Smoking	Participants were asked to describe their smoking habit as 'smoke daily', 'smoke occasionally', 'ex-smoker', 'never', 'claimed to

	be non-smokers and then admitted smoking at some point. We combined daily and occasional smokers with those that admitted smoking at some point into ‘smokers’, and kept ‘ex-smoker’ and ‘never smoked’ the same.
Alcohol (cat)	Participants were asked how often they drink alcohol. Options were ‘> 1/wk’, ‘1/wk’, ‘< 1/wk’, or ‘never’. This will be regrouped as ‘>= 1/wk’, ‘< 1/wk’, or ‘never’.
Alcohol (con)	Participants were also asked how often they drink beer, wine, sherry, and spirits (glasses/wk). Each alcoholic variable will be converted to glasses/day, multiplied by a standard serving weight, and summed to give total alcohol consumption in grams/day.
Weight	Self-reported continuous variable
Height	Self-reported continuous variable
BMI	Calculated as self-reported weight divided by the square of self-reported height, considered a continuous variable
N children	Self-reported continuous variable
Menopausal status	Categorised participants as pre-menopausal or post-menopausal based on age (< or > 55 years, ovaries removed, hormone replacement therapy use, and number of periods)
HRT use	Participants were asked ‘have you ever used HRT?’ and ‘are you using HRT now?’ – based on these yes or no answers, we categorised HRT use as ‘current’, ‘ex-user’, and ‘never’.