Substitution of red meat with legumes and risk of primary liver cancer in 126,744 UK Biobank participants a prospective cohort study

Niels Bock[1*], Fie Langmann[1] and Christina C. Dahm[1]

[1*]Department of Public Health, Aarhus University, 8000 Aarhus C, Denmark.

*Corresponding author(s). E-mail(s): nb@ph-au.dk;

Abstract

Purpose: Lifestyle-associated primary liver cancer is on the rise worldwide. Preventative strategies are warranted. We aimed to estimate the effect of substituting unprocessed red meat, processed red meat and total red meat with legumes on primary liver cancer in a free-living population.

Methods: We analyzed data from 126,744 UK Biobank participants who completed >= 2 24-hour diet recall questionnaires. Substitution of 15g/day of legumes with 15/day of unprocessed red meat, processed red meat and total red meat was statistically modeled using the leave-one-out model. Baseline characteristics were collected from the initial assessment visit. Information on liver cancer diagnoses was collected via external linkage to inpatient hospital episodes or central cancer registries.

Results: During a median follow-up time of 11.3 years, 173 participants developed liver cancer. In the fully adjusted models, no effect of substituting 15/day of legumes with unprocessed red meat (HR: 0.97 (95% CI 0.91-1.03);), processed red meat (HR: 1.02 (95% CI 0.93-1.13)) or total red meat (HR: 0.98 (95% CI 0.93-1.04)) was observed.

Conclusion: Overall, no association between substituting red meat with legumes and liver cancer was observed. Further research with longer follow-up time is warranted to increase the number of liver cancer events.

Keywords: Food Substitutions, Liver cancer, Red meat, Legumes

Introduction

The main aim of this study was to estimate the effect of substituting unprocessed red meat, processed red meat and total red meat with legumes on primary liver cancer in a free-living population.

Research Design and Methods

Study population

The UK Biobank, a population-based prospective cohort, was initiated in 2006. [13] During 2006-2010, more than 500,000 participants, aged 40-69, were recruited and visited designated assessment centres across the UK Participants provided information about age, sex, sociodemographic factors (education, Townsend deprivation index, living alone) and lifestyle factors (smoking, alcohol consumption, physical activity) via touch screen questionnaires and computer-assisted interviews. Anthropometric data (BMI, waist circumference) were collected via physical measurements [1].

Dietary assessment

A web-based 24-hour dietary recall was administered at the end of the initial assessment visit for the last 70,000 recruited participants [3]. From February 2011 to April 2012, 320,000 participants who had provided an e-mail address were invited on four separate occasions to complete the 24-hour dietary recall, the Oxford WebQ, of which 210,947 participants completed at least one. The Oxford WebQ covered 206 food items and 32 beverage items commonly consumed in the UK. Intakes were reported in standard units of measurements, e.g., servings, cups, slices, etc. with intake categories ranging from 0 to 3+ units [10]. The Oxford WebQ has been validated against interviewer-based 24-hour dietary recalls and biomarkers [5, 9].

Researchers defined79 food groups and 14 beverage groups from the Oxford WebQ using the UK National Diet and Nutrition Survey categories [10]. These food and beverage groups were used when defining the food groups used in the substitution analyses (Supplementary Table 1). Legumes were defined as dietary pulses, baked beans, tofu-based products, peas, hummus, soy drinks, and soy-based desserts and yogurt. Red meat intake was defined as intake of beef, pork, lamb, or other meat, including offal. Processed red meat intake was defined as sausages, bacon (with and without fat), ham, or liver pate. Other food groups included were animal-based foods, unhealthy plant-based foods, healthy plant-based foods, and alcoholic beverages (Supplementary Table 1). Animal-based and healthy and unhealthy plant-based food foods were grouped based on plant-based diet indices from previous studies [7, 11, 12, 14]. An overview of included foods in each food group is displayed in Supplementary Table 1.

Due to the incapability of a single 24-hour dietary recall to properly assess habitual dietary intake and variation in diet over time [6, 15], only participants who completed two or more Oxford WebQs were eligible for inclusion in this study.

Liver cancer assessment

Liver cancer was defined according to ICD-10 diagnosis codes C22.0 for Hepatocellular carcinoma (HCC) or C22.1 for Intrahepatic cholangiocarcinoma (ICC). Incident and prevalent cases of liver cancer and corresponding diagnosis dates were obtained via linkage to central cancer registries or hospital inpatient episodes [13].

Assessment of confounders

Confounders were defined *a priori* from a literature review of the background literature and illustrated using directed acyclic graphs (Supplementary Figure 1). The following confounding variables were selected: age at baseline (years, continuous), sex (male, female), educational level (high: College or University degree, intermediate: A levels/AS levels, O levels/GCSEs, or equivalent, low: none of the previous mentioned), Townsend Deprivation Index (continuous), Living alone (yes, no), waist circumference (cm, continuous), physical activity (above/below the 2017 UK Physical activity guidelines of 150 minutes of moderate activity per week or 75 minutes of vigorous activity, or unknown), smoking (pack years as a proportion of lifespan exposed to smoking, continous), and alcohol intake (g/day, continuous). All confounders except age were selected from the initial assessment visit before the start of follow-up.

The substitution model

The substitution analyses were conducted by replacing an equal mass of meat with legumes. The size of the substitution was set to 15 g of legumes for 15 g of meat to keep the substitution size below the mean intake of any of the substituted food groups in the cohort. The substitutions were modelled using the leave-one-out-approach in which variables for every food group along with a variable for total food intake are included, except the food group that are to be substituted [8]. To estimate substitution of 15 g of all red meats (red and processed) with 15 g of legumes, the following model was defined:

$$\begin{aligned}\log(h(t;x)) = \log(h_0(t)) &+ \beta_1\text{Legumes (15g)} + \beta_2\text{Total food intake (g)}\\ &+ \beta_3\text{Other food groups (g)} + \beta_4\text{Covariates}\end{aligned} \tag{1}$$

When substituting only red meat with legumes, processed red meat was added to the model:

$$\begin{aligned}\log(h(t;x)) = \log(h_0(t)) &+ \beta_1\text{Legumes (15g)} + \beta_2\text{Processed red meat (15g)}\\ &+ \beta_3\text{Total food intake (g)} + \beta_4\text{Other food groups (g)}\\ &+ \beta_5\text{Covariates}\end{aligned} \tag{2}$$

When substituting only processed red meat with legumes, red meat was added to the model:

$$\begin{aligned}\log(h(t;x)) = \log(h_0(t)) &+ \beta_1\text{Legumes (15g)} + \beta_2\text{Red meat (15g)} \\ &+ \beta_3\text{Total food intake (g)} + \beta_4\text{Other food groups (g)} \\ &+ \beta_5\text{Covariates}\end{aligned} \tag{3}$$

Statistical analysis

Multivariable-adjusted Cox proportional hazards regression models were used to estimate hazard ratios (HR) with corresponding 95% confidence intervals (CI) with age as the underlying timescale. Participants were followed from the date of their last completed Oxford WebQ until the occurrence of the event of interest or due to right censoring, whichever came first. Participants were right censored in the event of death, loss to follow-up, or administrative end of follow-up (October 31, 2022). Two levels of adjustments were added to the substitution model. Model 1 was minimally adjusted for age, total weight of food intake, and all other food groups to fit the substitution model. Model was further adjusted for sex, educational level, Townsend Deprivation Index, living alone, physical activity, smoking, alcohol intake, and waist circumference.

In secondary analyses, each cancer type was analysed separately to evaluate if the pooling of HCC and ICC as one outcome in the main analysis was justified. Furthermore, to estimate the effect of legume intake regardsless of other dietary components, legume consumers (divided into quartiles) were compared to non-consumers.

To evaluate the robustness of the main analyses, sensitivity analyses were performed on subsamples of participants by excluding those with high alcohol intake (more than 32 grams per day for men and 24 grams per day for women), implausible energy intake (3200 or 16800 kJ/day for men and 2000 or 14000 kJ/day for women), any liver disease before baseline, any type of cancer before baseline, and fewer than 3 completed Oxford WebQs. As neither the central cancer registries nor the hospital inpatient registries were complete, liver cancer diagnoses retrieved from death registries were included in a sensitivity analysis to test for outcome misclassification bias. Lastly, one of our causal assumptions was that anthropometry confounded the causal relationship between replacing red meat with legumes and liver cancer; however, strong arguments exist giving support to anthropometry being a mediator between diet and health outcomes. Thus, to test for erroneously conditioning on a potential mediator, waist circumference was removed in a sensitivity analysis.

Sensitivity analyses were modelled like the fully adjusted models in the main analyses.

All analyses were conducted in R (version 4.1.1) with a significance level of 5 %.

Results

After excluding participants with liver cancer before baseline, participants lost to follow-up before baseline, and participants with errors in the diet data, 126,744 participants remained who had completed two or more diet questionnaires (Figure 1).

During a median follow-up time of 11.3 years, 173 participants developed liver cancer. Participants who developed liver cancer were older at baseline, had a higher waist circumference, were less physically active, fewer had never smoked, and more were male, compared to all included participants (Table 1).

Mean daily energy intake and food intake and daily intake of all specified food groups in grams are presented in table 2.

No association was found for substituting 15 g/day of legumes with 15 g/day of total red meat, unprocessed red meat, or processed red meat and risk of primary liver cancer in model 1 (Table 3: total red meat: HR: 0.98, 95% CI: 0.93, 1.04; unprocessed red meat: HR: 0.97, 95% CI: 0.91, 1.03; processed red meat: HR: 1.02, 95% CI: 0.93, 1.13). The estimated associations changed minimally but remained non-significant with further adjustments (Table 3: total red meat: HR: 1.02, 95% CI: 0.96, 1.08; unprocessed red meat: HR: 1.00, 95% CI: 0.94, 1.07; processed red meat: HR: 1.09, 95% CI: 0.98, 1.20).

In secondary analyses, when analyzing the substitution association for HCC and ICC separately, the risk of HCC was positively associated with substituting total,

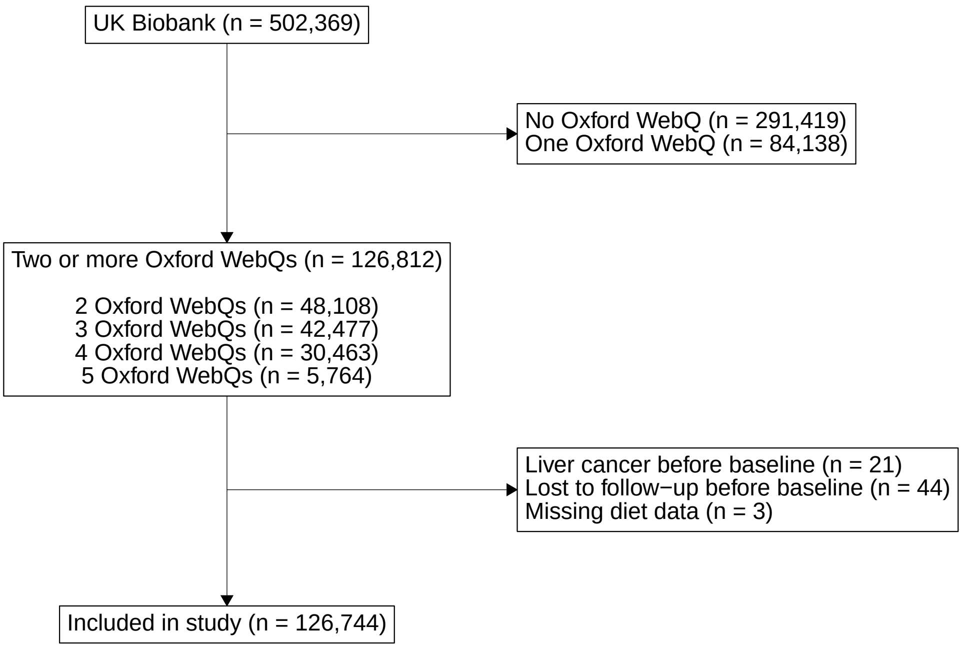

Fig. 1: Flowchart of included participants. Missing diet data was merged with loss to follow-up before baseline due to n being less than 5. It should be noted that not all UK Biobank participants were invited to complete an Oxfords WebQ. Only the last 70,000 participants to visit an assessment center were asked to complete an Oxford WebQ at the end of their visit. Further Oxfords WebQs were sent to 320,000 participants who provied an e-mail adress.

unprocessed, or processed red meat with legumes (Supplementary Table 2, total red meat: HR: 1.06, 95% CI: 0.97, 1.16; red meat: HR: 1.05, 95% CI: 0.96, 1.15; processed red meat: HR: 1.09, 95% CI: 0.95, 1.26). The association between substituting total, unprocessed, or processed red meat with legumes and ICC indicated inverse associations (Supplementary Table 2, total red meat: HR: 0.97, 95% CI: 0.90, 1.05; red meat: HR: 0.95, 95% CI: 0.87, 1.03; processed red meat: HR: 1.07, 95% CI: 0.93, 1.22). The magnitude or direction of associations were not significantly different across strata of liver cancer types.

In the adjusted non-substitution analysis, a mean intake of 6.3 grams of legumes per day was associated with a reduced risk of liver cancer, compared to no intake (HR: 0.59, 95% CI: 0.35, 0.98); however, no associations were observed with further increase in legume intake (supplementary table 3).

In sensitivity analyses, excluding participants with high alcohol intake or participants with plausible energy intake misreporting or setting inclusion criteria to 3 completed diet questionnaires did not alter the estimates in any statistically significant way (supplementary table 4).

Discussion

Summary of results

In this study of UK Biobank participants, we found no effect of replacing 15g/day of red meat with legumes on risk of primary liver cancer. The estimates did not change significantly in any of the sensitivity analyses.

Discuss results

e.g, why estimates for HCC and ICC go in opposite direction. sensitivity analyses

Contrary to our hypothesis, replacing processed red meat with legumes was associated with a non-significant increase in risk of primary liver cancer and with a greater magnitude than unprocessed red meat. This tendency was observed across all sensitivity analyses. However, the estimates for processed red meat were labeled with less confidence due to the low median intake and

Other studies: substitution analyses and foods related to disease

Discuss studies on legumes/red meat and liver cancer/other diseases. Other substitutions

Strength and limitations

This study had some limitations. First, None of the registries used to determine a diagnosis of liver cancer were complete or up to date [4]. Data from external providers, e.g., the NHS England, NHS Central Register or National Records of Scotland, were estimated to be mostly complete by the UK Biobank at various dates, ranging from 31 December 2016 for cancer data from Wales to 31 October 2022 for hospital inpatient data from England [2]. This could introduce misclassification bias of the outcome as individuals with liver cancer may not be identified. However, the estimates were robust to a sensitivity analysis where we included death registries as a source of liver cancer diagnoses in an attempt to accommodate for missing outcome events.

Second, the relative low number of events limited our efforts of adjusting for confounding as too many adjustment levels per event can affect the validity of the multivariable Cox regression model, causing potential biased estimates. Efforts were made to reach at least 10 events per variable for the main analysis to secure statistical validity by limiting the number of adjustment levels, i.e., fewer and broader food groups and fewer levels for categorical covariates. This was done while respecting our *a priori* causal assumptions. Reaching statistical validity was a trade-off with increasing residual confounding as importance of specific food groups or

Selection bias: Incomplete registries regarding cancer. Participants who completed 2 diet questionnaires are healthier than the those who didn't?

Lavt bortfald.

kun 5% deltagelse.

Information bias: Underreporting of unhealthy habits? Non-differentiated misclassification. Participants with unhealthy habits are underreporting? differentiated misclassification.

incomplete registries, information problem.

Confounding

residual confounding: broad food groups erase effect of specific foods. Are processed red meat eaten with legumes in the UK? (baked beans and bacon). Aflatoxins

Few events: trade off between statistical validity and residual confounding.

Perspectives

Conclusion

Acknowledgements

Tables

Table 1. Baseline characteristics of UK Biobank participants who completed 2 Oxford WebQ 24-hour diet recall.

	Cohort	Liver cancer
Variable	**N = 126,744**[1]	**N = 173**[1]
Typical diet yesterday[2]	73,213 (58%)	105 (61%)
Age, years	60 (53, 65)	64.0 (60.0, 68.0)
Sex		
Female	70,659 (56%)	65 (38%)
Male	56,085 (44%)	108 (62%)
Educational level[3]		
High	59,416 (47%)	76 (44%)
Intermediate	41,817 (33%)	52 (30%)
Low	25,472 (20%)	45 (26%)
Missing	39	
Townsend Deprivation Index	-2.4 (-3.8, 0.0)	-2.6 (-3.7, -0.7)
Missing	149	
Living alone	22,658 (18%)	34 (20%)
Missing	171	
Physical activity[4]		
Above	58,111 (46%)	61 (35%)
Below	50,712 (40%)	79 (46%)
Missing	17,921 (14%)	33 (19%)
Smoking		
Never	72,583 (57%)	75 (43%)
Ever	54,122 (43%)	98 (57%)
Missing	39	
Alcohol intake, g/day	11 (0, 26)	11 (0, 29)
Waist circumference, cm	88 (79, 97)	98 (89, 107)
Missing	168	

[1]Median (IQR) for continous variables; n (%) for categorical variables
[2]Participants who reported eating a typical diet yesterday for all completed diet questionnaires.
[3]High: College or University degree; Intermediate: A levels/AS levels, O levels/GCSEs, or equivalent; Low: none of the previous mentioned.
[4]Above or below the 2017 UK Physical activity guidelines of 150 minutes of moderate activity per week or 75 minutes of vigorous activity.

Table 2. Daily dietary intake of food groups, total food and total energy intake in UK Biobank participants who completed >= 2 Oxford WebQ 24-hour diet recall.

Daily food intake	Cohort N = 126,744[1]	Liver cancer N = 173[1]
Total food intake		
Energy, kJ	8,430 (7,179, 9,856)	8,579 (7,413, 10,048)
Weight, g	3,144 (2,720, 3,621)	3,162 (2,737, 3,659)
Food groups, g/day		
Legumes	11 (0, 34)	8 (0, 35)
Red and processed meat	53 (15, 86)	60 (30, 95)
Red meat	30 (0, 60)	45 (0, 73)
Processed meat	9 (0, 30)	8 (0, 31)
Other animal-based foods[2]	475 (361, 603)	448 (322, 604)
Healthy plant-based foods[3]	1,806 (1,454, 2,198)	1,791 (1,365, 2,158)
Unhealthy plant-based foods[4]	472 (324, 662)	491 (365, 698)
Alcoholic beverages	132 (0, 342)	144 (0, 375)

[1]Median (IQR)

[2]Other animal-based foods include: poultry, fish, dairy, eggs, and mixed dishes with animal products.

[3]Healthy plant-based foods include: whole grains, vegetables, fruits, nuts, plant oils, and beverages (coffee, tea, water).

[4]Unhealthy plant-based foods includes: refined grains, potatoes, mixed vegetarian dishes, sweets and snacks, fruit juice, and sugar sweetened beverages.

Table 3. Substitution of total meat, red meat and processed meat with legumes and hazard ratios and 95% confidence intervals for primary liver cancer.

	Model 1[1]	Model 2[2]
15 g/day of legumes replacing:	**HR (95% CI)**	**HR (95% CI)**
Total red meat	0.98 (0.93 to 1.04)	1.02 (0.96 to 1.08)
Unprocessed red meat	0.97 (0.91 to 1.03)	1.00 (0.94 to 1.07)
Processed red meat	1.02 (0.93 to 1.13)	1.09 (0.98 to 1.20)

[1]Adjusted for age (as underlying timescale), other food groups, and total food intake.
[2]Further adjusted for sex, educational level, Townsend deprivation index, living alone, physical activity, smoking, alcohol intake, and waist circumference.

Supplemental Material

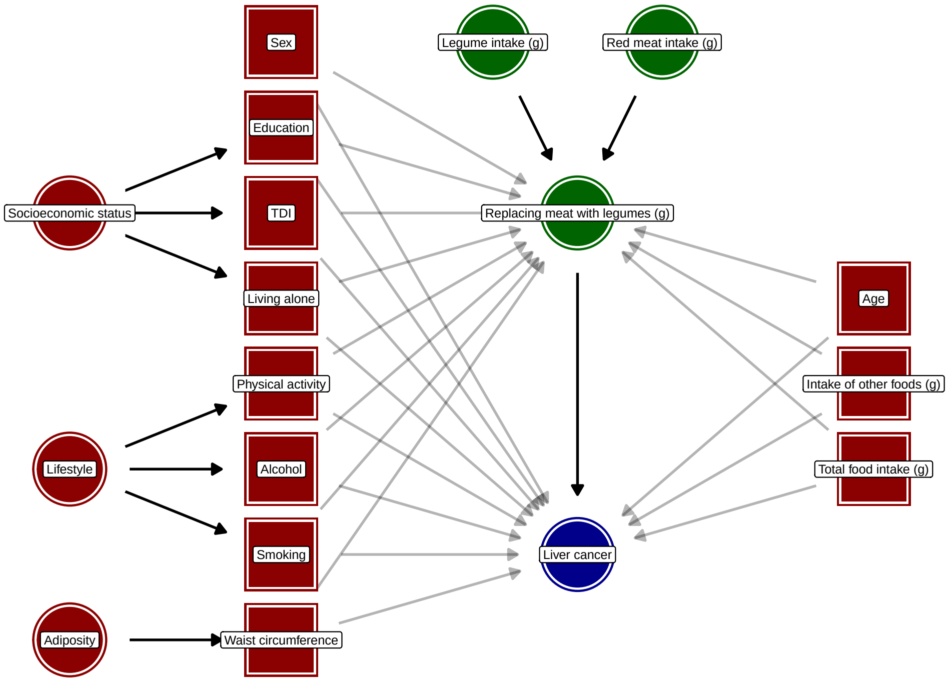

Fig. 2: Directed acyclic graph (DAG) visualizing the hypothesised causal relationship between replacing meat with legumes and liver cancer based on assumptions of biasing paths. Red nodes represent confounders. Square nodes represent the minimal sufficient adjustment set for estimating the effect of replacing meat with legumes on liver cancer. Shadowed arrows represent biasing paths.

Supplementary table 1. Summary of included foods for each food group.

Food group	Includes
Legumes	Soya-based desserts, Baked beans, pulses, Soya drinks (including calcium fortified), Tofu-based products, Hummus, Peas
Red meat	Beef, Lamb, Other meat including offal, Pork
Processed meat	Sausages, bacon (with and without fat), ham, liver pate
Animal-based foods	
Poultry	Fried poultry with batter/breadcrumbs, Poultry (with/without skin)
Fish	Fried fish with batter/breadcrumbs, Oily fish, including salmon, Prawns, lobster, crab, shellfish, Tinned tuna, white fish, other fish
Dairy	Spreadable/lower fat butter, dairy-based very low fat spread, Spreadable normal fat butter, dairy-based normal fat spread (including cholesterol lowering spread), Ice cream, milk puddings, milk-based desserts, cheesecake, Dairy-based smoothies, milk-based drinks, hot chocolate, Whole milk yogurt (plain), Cheese >17.5 g fat per 100 g, including hard cheese, soft cheese, spreadable, Blue, Feta, Mozzarella, Goats, other), Fat free and lower fat yogurt, plain or flavoured, Cheese <=17.5g fat per 100 g, including hard and spreadable lower fat cheese, Cottage, Semi-skimmed milk >1 g fat per 100 g (cow, other), Skimmed milk <1 g fat per 100 g (cow, cholesterol lowering, powdered), Whole milk >3.6 g fat per 100 g (cow, goat, sheep), Cream (cow's milk)
Eggs	Whole eggs and processed (omelette, scotch eggs, other)
Sauces	Mayonnaise, salad dressing, pesto, cheese sauce, white sauce, gravy, Yeast, chutney, olives, ketchup, brown sauce, tomato sauce
Mixed dishes	Pizza (including gluten free crust), Crisps, savoury biscuits, cheese snacks, other savoury biscuits, Soups, homemade, powdered and canned, Sushi
Healthy plant-based foods	
Whole grains	Mixed, brown or seeded bread, sliced, baguette, bap, roll, Wholemeal bread, sliced, baguette, bap, roll, Wholewheat biscuit cereal, Bran cereal, Porridge oats (including milk/dried fruit added), Oatcrunch breakfast cereal, Muesli (with or without dried fruit), Brown and wholemeal pasta and rice
Fruits	Apples and pears, Blackberries, strawberries, blueberries, raspberries, cherries, Grapefruit, orange, satsuma, Dried fruit, prunes, Bananas, mixed fruit, grapes, mango, melon, peach, pineapple, kiwi, other, Stewed fruit, plums

Nuts	Peanut-butter and chocolate-based spread, Unsalted peanuts and nuts, Salted peanuts and nuts
Plant oils	Olive oil, Olive oil based lower fat spread, plant-based lower fat margarine and soya-based lower fat spread (including cholesterol lowering spread), Olive oil based spread, plant-based soft or hard margarine and soya-based spread (including cholesterol lowering spread)
Beverages	Normal instant, filter, cappuccino, espresso coffee, Decaffeinated instant, filter, cappuccino, espresso coffee, Black, green and other tea, Decaffeinated black, herbal tea, rooibos, Plain water, sparkling water
Vegetables	Garlic, leek, onion, Broccoli, cabbage, kale, cauliflower, spinach, sprouts, Mixed side salad, lettuce, watercress, Beetroot, carrots, celery, parsnip, turnip, Fresh and tinned tomatoes, Mushrooms, mixed vegetables, avocado, broad beans, green beans, butternut squash, courgettes, peppers, other, Coleslaw, salad with added fat/mayonnaise, guacamole, sweetcorn
Unhealthy plant-based foods	
Refined cereals	Chocolate biscuits, plain biscuits, sweet biscuits and cookies, Naan, garlic bread, other bread (including gluten free), White bread, sliced, baguette, bap, roll, Oatcrunch breakfast cereal, Crisps, savoury biscuits, cheese snacks, other savoury biscuits, White pasta, rice, couscous, gluten free pasta
Potatoes	Potatoes, sweet potatoes, boiled or baked, Potatoes and chips, fried or roasted with fat, Potatoes, mashed
Fruit juice	Orange, grapefruit drink and 100% fruit juice
Mixed dishes, vegetarian	Double and single crust pies, crumble pies, Yorkshire pudding, snackpot noodles, Indian samosa, pakora snacks, Quorn-based and vegetarian burgers and products
Sweets & snacks	Table sugar, honey, jam and preserves, Chocolate bar (including white, milk and dark chocolate), chocolate-covered raisins, chocolate-covered sweets, Pancakes, croissant, Danish pastries, scones, fruitcakes, cakes, doughnuts, sponge puddings, other desserts, cereal bars, sweet snacks, Hard and soft sweets (including sugar free)
Sugar sweetened beverages	Rice and oat vegetable drinks, Low calorie fizzy drinks and squash, Fizzy sugary drinks, squash, fruit smoothies
Alcoholic beverages	Beer and cider, Spirits and other alcoholic drinks, Fortified wine, Red and rose wine, White wine

Supplementary table 2. Substitution of total meat, red meat and processed meat with legumes and hazard ratios and 95% confidence intervals for hepatocellular carcinoma and intrahepatic cholangiocarcinoma.

	Model 1[1]	Model 2[2]
15 g/day of legumes replacing:	**HR (95% CI)**	**HR (95% CI)**
Hepatocellular carcinoma		
Total red meat	1.01 (0.93 to 1.10)	1.06 (0.97 to 1.16)
Unprocessed red meat	1.01 (0.92 to 1.11)	1.05 (0.96 to 1.15)
Processed red meat	1.01 (0.88 to 1.16)	1.09 (0.95 to 1.26)
Intrahepatic cholangiocarcinoma		
Total red meat	0.94 (0.87 to 1.02)	0.97 (0.90 to 1.05)
Unprocessed red meat	0.92 (0.85 to 1.00)	0.95 (0.87 to 1.03)
Processed red meat	1.02 (0.89 to 1.17)	1.07 (0.93 to 1.22)

[1]Adjusted for age (as underlying timescale), other food groups, and total food intake.
[2]Further adjusted for sex, educational level, Townsend deprivation index, living alone, physical activity, smoking, alcohol intake, and waist circumference.

Supplementary table 3. No intake of legumes vs. quartiles of daily legume intake and hazard ratios and 95% confidence intervals for primary liver cancer.

	Model 1[1]	Model 2[2]
Characteristic	**HR (95% CI)**	**HR (95% CI)**
Categories:[3]		
No intake	—	—
Q1	0.58 (0.35 to 0.96)	0.59 (0.35 to 0.98)
Q2	0.87 (0.56 to 1.33)	0.89 (0.58 to 1.36)
Q3	0.75 (0.47 to 1.19)	0.75 (0.47 to 1.19)
Q4	0.98 (0.64 to 1.51)	1.06 (0.69 to 1.64)

[1]Adjusted for age (as underlying timescale), other food groups, and total food intake.
[2]Further adjusted for sex, educational level, Townsend deprivation index, living alone, physical activity, smoking, alcohol intake, and waist circumference.
[3]mean daily intake of legumes in grams for each quartile: Q1: 6.3, Q2: 15.7, Q3: 34.3, Q4 109.

Supplementary table 4. Sensitivity analyses

	Exclusion of participants with:						
	High alcohol intake[1]	Implausible food intake[2]	Liver disease before baseline[3]	Any cancer before baseline[4]	Fewer than 3 Oxford WebQs	Death register as source for liver cancer events	Exclusion of waist circumference from analysis
15 g/day of legumes replacing:	**HR (95% CI)**	**HR (95% CI)**	**HR (95% CI)**	**HR (95% CI)**	**HR (95% CI)**	**HR (95% CI)**	**HR (95% CI)**
Total red meat	1.00 (0.94 to 1.07)	1.02 (0.96 to 1.08)	0.99 (0.93 to 1.06)	1.04 (0.97 to 1.11)	1.04 (0.97 to 1.12)	1.02 (0.97 to 1.09)	1.00 (0.94 to 1.06)
Unprocessed red meat	0.99 (0.92 to 1.06)	1.00 (0.94 to 1.06)	0.97 (0.91 to 1.04)	1.01 (0.94 to 1.09)	1.02 (0.94 to 1.11)	1.01 (0.95 to 1.07)	0.99 (0.92 to 1.05)
Processed red meat	1.04 (0.93 to 1.17)	1.09 (0.98 to 1.20)	1.07 (0.96 to 1.20)	1.15 (1.01 to 1.30)	1.11 (0.97 to 1.27)	1.07 (0.97 to 1.18)	1.05 (0.95 to 1.16)

[1]Exclusion of the upper 10th percentile of daily alcohol intake in grams for each sex.
[2]Exclusion of the upper and lower 10th percentile of daily energy intake for each sex.
[3]ICD10 codes: K70-79, B16-19, Z94.4, I82.0, I85, I86.4, E83.0-1 and E88. ICD9 codes: 571-574, 070, V427 and 2750-2751.
[4]ICD10 codes: C00-C97 and D00-D48. ICD9 codes: 140-239.

References

[1] Biobank U (2011) Order of data collection. URL https://biobank.ndph.ox.ac.uk/ukb/ukb/docs/Orderofdatacollection.pdf

[2] Biobank U (2023) Data providers and dates of data availability. URL https://biobank.ndph.ox.ac.uk/ukb/exinfo.cgi?src=Data_providers_and_dates

[3] Biobank U (2024) 24-hour dietary recall questionnaire (oxford webq). URL https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/DietWebQ.pdf

[4] Biobank U (2024) Health outcomes overview. URL https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/HealthOutcomesOverview.pdf

[5] Greenwood DC, Hardie LJ, Frost GS, et al (2019) Validation of the oxford webq online 24-hour dietary questionnaire using biomarkers. American Journal of Epidemiology 188(10):1858–1867. https://doi.org/10.1093/aje/kwz165, URL http://dx.doi.org/10.1093/aje/kwz165

[6] Gurinović M, Zeković M, Milešević J, et al (2017) Nutritional Assessment, Elsevier. https://doi.org/10.1016/b978-0-08-100596-5.21180-3, URL http://dx.doi.org/10.1016/B978-0-08-100596-5.21180-3

[7] Heianza Y, Zhou T, Sun D, et al (2021) Healthful plant-based dietary patterns, genetic risk of obesity, and cardiovascular risk in the uk biobank study. Clinical Nutrition 40(7):4694–4701. https://doi.org/10.1016/j.clnu.2021.06.018, URL http://dx.doi.org/10.1016/j.clnu.2021.06.018

[8] Ibsen DB, Laursen ASD, Würtz AML, et al (2021) Food substitution models for nutritional epidemiology. The American Journal of Clinical Nutrition 113(2):294–303. https://doi.org/10.1093/ajcn/nqaa315, URL http://dx.doi.org/10.1093/ajcn/nqaa315

[9] Liu B, Young H, Crowe FL, et al (2011) 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 14(11):1998–2005. https://doi.org/10.1017/s1368980011000942, URL http://dx.doi.org/10.1017/S1368980011000942

[10] Piernas C, Perez-Cornago A, Gao M, et al (2021) 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 60(5):2879–2890. https://doi.org/10.1007/s00394-021-02535-x, URL http://dx.doi.org/10.1007/s00394-021-02535-x

[11] Satija A, Bhupathiraju SN, Rimm EB, et al (2016) Plant-based dietary patterns and incidence of type 2 diabetes in us men and women: Results from

three prospective cohort studies. PLOS Medicine 13(6):e1002039. https://doi.org/10.1371/journal.pmed.1002039, URL http://dx.doi.org/10.1371/journal.pmed.1002039

[12] Satija A, Bhupathiraju SN, Spiegelman D, et al (2017) Healthful and unhealthful plant-based diets and the risk of coronary heart disease in u.s. adults. Journal of the American College of Cardiology 70(4):411–422. https://doi.org/10.1016/j.jacc.2017.05.047, URL http://dx.doi.org/10.1016/j.jacc.2017.05.047

[13] Sudlow C, Gallacher J, Allen N, et al (2015) Uk biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLOS Medicine 12(3):e1001779. https://doi.org/10.1371/journal.pmed.1001779, URL http://dx.doi.org/10.1371/journal.pmed.1001779

[14] Thompson AS, Tresserra-Rimbau A, Karavasiloglou N, et al (2023) Association of healthful plant-based diet adherence with risk of mortality and major chronic diseases among adults in the uk. JAMA Network Open 6(3):e234714. https://doi.org/10.1001/jamanetworkopen.2023.4714, URL http://dx.doi.org/10.1001/jamanetworkopen.2023.4714

[15] Thompson FE, Subar AF (2013) Dietary Assessment Methodology, Elsevier, pp 5–46. https://doi.org/10.1016/b978-0-12-391884-0.00001-9, URL http://dx.doi.org/10.1016/B978-0-12-391884-0.00001-9