

# Substitution of red meat with legumes and risk of primary liver cancer in UK Biobank participants: A prospective cohort study

Niels Bock <sup>1</sup>, Fie Langmann <sup>1</sup>, Luke W. Johnston <sup>1,2</sup>, Daniel B. Ibsen <sup>1,2</sup>, and Christina C. Dahm <sup>1</sup>

<sup>1</sup>Department of Public Health, Aarhus University, Aarhus, Denmark

<sup>2</sup>Steno Diabetes Center Aarhus, Aarhus University Hospital, Aarhus N, Denmark

June 20, 2024

## Abstract

Primary liver cancer is globally on the rise, partially due to poor diets and sedentary lifestyles. Shifting to more plant-based diets may lower the risk. We aimed to estimate the effect of replacing total red meat, unprocessed red meat and processed red meat with legumes on primary liver cancer in a free-living population. We analyzed data from 126,744 UK Biobank participants who completed  $\geq 2$  24-hour diet recalls. 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. Cox proportional hazards regression models were used to estimate substitution of 15 g/day of legumes with 15 g/day of total red meat, unprocessed red meat or processed red meat on liver cancer risk, using the leave-one-out food substitution model. During a median follow-up time of 11.1 years, 173 participants developed liver cancer. In the fully adjusted models, no association was observed when substituting 15 g/day of legumes with total red meat (HR: 1.02 (95% CI 0.96-1.08)), unprocessed red meat (HR: 1.00 (95% CI 0.94-1.06)) or processed red meat (HR: 1.09 (95% CI 0.99-1.21)). Overall, little evidence of an association between replacing red meat with legumes and liver cancer was observed. Further research in other study populations with longer follow-up time is warranted.

# Introduction

Hepatocellular carcinoma (HCC) is the sixth most common cancer in the world and the third leading cause of cancer-related death with viral hepatitis being the leading risk factor [1]. In low-infection populations, modifiable risk factors, such as dietary habits, may play an increasing role in HCC pathogenesis as non-alcoholic fatty liver disease (NAFLD) has become the leading cause of liver cirrhosis [2, 3] that may in turn progress to HCC. A western dietary pattern high in fats and red meats and concurrently low in fruits, vegetables, and whole grains has been associated with NAFLD progression [4]. The prevalence of NAFLD-related HCC cases is an increasing global problem [2]. It is estimated that the prevalence of NAFLD-related HCC in the US will increase by 146%, while incident NAFLD-related HCC cases will increase by 137% in 2030 [5].

The second most common primary liver cancer is the intrahepatic cholangiocarcinoma (ICC) [6]. While HCC emerges from the liver parenchyma, ICC emerges from the bile duct. Despite being a relatively rare cancer, ICC is characterized by its aggressiveness, late diagnosis and poor survival [7]. It is estimated that the incidence of ICC is increasing in populations that are not burdened by known infectious and environmental risk factors [8]. Recent meta-analyses of observational studies and clinical trials have shown a significant adverse association between NAFLD and ICC [9, 10].

The impact of specific food groups on liver cancer risk is not well known. Observational studies suggest that intake of coffee, vegetables and whole grains may lower HCC risk [11, 12, 13, 14]. The protective properties of these foods are proposed to be their content of dietary fibers and polyphenols, which are also defining components of legumes. The health benefits of legumes extend to improved glycemic control and hypotensive and anticarcinogenic properties with observed inverse associations with cardiovascular disease and colorectal cancer [15, 16]. Two large prospective cohort studies ( $N = 130,000$ - $500,000$ ,  $n$  event = 298-940) found evidence of inverse associations between legume consumption and risk of HCC [11, 13]. However, replacement foods were not specified in these studies, which fails to reflect that a higher intake of one food is at the expense of a concomitantly lower intake of another food. On the contrary, processed red meat intake, but not unprocessed red meat intake, was associated with an increased risk of HCC in two large cohorts ( $N = 50,000$ - $120,000$ ,  $n$  events = 163), suggesting that the processing of red meat may augment the carcinogenic effects on the liver tissues [17]. Processed red meat is classified as “carcinogenic to humans” and unprocessed red meat as a possible carcinogen by the International Agency for Research on Cancer [18]. Studies on substituting animal-based proteins with

plant-based proteins are important if we are to lower the climate impacts of our diets [19]. While previous research has investigated and found protective effects when substituting animal-based proteins with plant-based proteins in relations to NAFLD [20], research on substituting meats with legumes in relation to risk of HCC and ICC is sparse. This leaves a substantial gap in the current knowledge on the beneficial effects on primary liver cancer from substituting red meat with legumes.

The low incidence of liver cancer in populations not burdened by viral hepatitis complicates observational prospective research designs; nonetheless, the prospects of the burden of liver cancer on public health warrant investigation of preventative measures. Thus, the main aim of this study was to estimate the association between replacing unprocessed red meat, processed red meat and total red meat with legumes on primary liver cancer in a large free-living population.

## Materials and Methods

The protocol for this study was written prior to conducting the analysis. It is available on the archive Zenodo [21]. Some changes were made to the final analysis plan due to lack of power to conduct subgroup and mediation analyses.

### Study population

The UK Biobank is a population-based prospective cohort initiated in 2006 [22]. 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 and lifestyle factors via touch screen questionnaires, and computer-assisted interviews. Anthropometric data were collected via physical measurements [23].

### 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 [24]. 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 [25]. The Oxford WebQ has been validated

against interviewer-based 24-hour dietary recalls showing acceptable correlations for total energy intake and most nutrients, and biomarkers showing acceptable correlations between the average values of two or more Oxford WebQs and estimated true intakes of total energy, total sugar, potassium and protein [26, 27].

A total of 79 food categories and 14 beverage categories from the Oxford WebQ has previously been defined using the UK National Diet and Nutrition Survey categories [25]. We used these food and beverage groups when defining the food groups used in the substitution analyses (Table S1). Legumes were defined as legumes and dietary pulses, baked beans, tofu-based products, peas, hummus, soy drinks, and soy-based desserts and yogurt. Unprocessed 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 paté. Total red meat was the combination of unprocessed and processed red meat. Other food groups included were animal-based foods, unhealthy plant-based foods, healthy plant-based foods, and alcoholic beverages (Table S1). Animal-based and healthy and unhealthy plant-based food foods were grouped based on plant-based diet indices from previous studies [28, 29, 30, 31].

As a single 24-hour dietary recall does not assess habitual dietary intake and variation in diet over time at an individual level [32, 33], only participants who completed two or more Oxford WebQs were eligible for inclusion in this study. Baseline food intakes were defined as average intakes from the reported 24-hour diet recalls.

## 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) and ICD-9 diagnosis codes 1550 Malignant neoplasm of liver, primary or 1551 Malignant neoplasm of intrahepatic bile ducts. Incident and prevalent cases of liver cancer and corresponding diagnosis dates were obtained via external linkage to central cancer registries or hospital inpatient episodes [34, 35].

## Assessment of confounders

Confounders were defined *a priori* from a review of the literature and illustrated using directed acyclic graphs (Figure S1). The following confounding variables were selected: age, sex, educational level, Townsend deprivation index (TDI), living alone, physical activity, smoking, alcohol intake, and waist circumference. Information on all confounders except age was collected at the initial assessment visit before the start of follow-up.

## Statistical analysis

Baseline characteristics and intake of food groups of all included participants and participants who developed liver cancer were described using standard summary statistics. Continuous variables were described with medians and interquartile range (IQR, 25th-75th percentiles) and categorical variables in total numbers (n) and percentage (%). Intake of food groups were described as g/day.

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).

The substitution analyses were conducted by modeling replacement of an equal mass of meat with legumes. The portion size of the substitution was set to 15 g/day of legumes for 15 g/day of red meat to ensure that substitutions were below the mean intake of any of the substituted food groups in the cohort. The substitutions were modeled using the leave-one-out-approach in which variables for every food group intake along with a variable for total food intake were included, except the food group that were to be substituted [36]. To estimate replacing of 15 g/day of all red meats (unprocessed and processed) with 15 g/day of legumes, the following model was defined:

$$\begin{aligned} \log(h(t; x)) = & \log(h_0(t)) + \beta_1 \text{Legumes (15 g/day)} + \\ & \beta_2 \text{Total food intake (g/day)} + \\ & \beta_3 \text{'Other food groups (g/day)} + \\ & \beta_4 \text{'Covariates} \end{aligned} \quad (1)$$

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

$$\begin{aligned} \log(h(t; x)) = & \log(h_0(t)) + \beta_1 \text{Legumes (15 g/day)} + \\ & \beta_2 \text{Processed red meat (15 g/day)} + \\ & \beta_3 \text{Total food intake (g/day)} + \\ & \beta_4 \text{'Other food groups (g/day)} + \\ & \beta_5 \text{'Covariates} \end{aligned} \quad (2)$$

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

$$\begin{aligned} \log(h(t; x)) = & \log(h_0(t)) + \beta_1 \text{Legumes (15 g/day)} + \\ & \beta_2 \text{Unprocessed red meat (15 g/day)} + \\ & \beta_3 \text{Total food intake (g/day)} + \\ & \beta_4 \text{'Other food groups (g/day)} + \\ & \beta_5 \text{'Covariates} \end{aligned} \quad (3)$$

The performance of the leave-one-out model when modeling equal mass substitutions has been validated against simulated data [37].

Two levels of adjustments were added to the substitution model. Model 1 was minimally adjusted for age (as the underlying timescale) total weight of food and beverage intake (g/day), and all other food groups (g/day) to fit the substitution model. To account for differences in baseline risks, model 1 was additionally stratified on age at recruitment (<45, 45-49, 50-54, 55-59, 60-64 and  $\geq 65$ ), attended assessment centre, and sex (male, female). Model 2 was further adjusted for 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), 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, continuous), alcohol intake (g/day, continuous), and waist circumference (cm, continuous). The included covariates, including food groups, were grouped to ensure adequate power in the analyses to discover any associations between the exposure and outcome of interest and were guided by our *a priori* assumptions [21]. Assumptions of proportional hazards were checked using Schoenfeld residuals and showed no violations.

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 association of legume intake with liver cancer, not specifying the substitution, 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 (exclusion of the upper decile of alcohol intake (g/day) by sex), implausible energy intake (exclusion of participants below the 2.5th percentile and above the 97.5th percentile of energy intake (kJ/day) by sex), any liver disease before baseline, any type of cancer before baseline, and fewer than three completed Oxford WebQs. As neither the central cancer registries nor the hospital inpatient registries were complete, liver cancer diagnoses retrieved from death registries, which were updated more recently,

were included in a sensitivity analysis to test for outcome misclassification. Further, one of the causal assumptions was that anthropometry confounded the causal relationship between replacing red meat with legumes and liver cancer; however, arguments exist giving support to anthropometry being a mediator between diet and health outcomes. Thus, to test for erroneously conditioning on a potential mediator, a sensitivity analysis was adjusted following model 2 but without waist circumference. Lastly, sensitivity analyses omitting soy milk from the estimated daily legume intake were conducted, as soy milk is unlikely to culinarily replace red meat. All sensitivity analyses were modeled as the fully adjusted models in the main analyses.

All analyses were conducted in R version 4.4.0 (2024-04-24) with a significance level of 5%. The code is structured in a reproducible manner using the targets R package [38] and is available at <https://github.com/steno-aarhus/legliv>.

## Results

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

During a median follow-up time of 11.1 years, 173 participants developed liver cancer of which 73 were HCC and 100 were ICC. Those who developed liver cancer were older at baseline, were more likely to be male, have a higher waist circumference, be less physically active, and fewer had never smoked, compared to all included participants (Table 1).

Mean daily energy and total food intakes as well as daily intake of all specified food groups in grams are presented in Table 2.

No evidence of associations 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: 0.99, 95% CI: 0.93-1.05; unprocessed red meat: HR: 0.97, 95% CI: 0.91-1.03; processed red meat: HR: 1.04, 95% CI: 0.94-1.15). The estimated associations changed minimally with further adjustments. There was weak evidence of an association between replacement of processed red meat with legumes (HR: 1.09, 95% CI: 0.99-1.21; Table 3).

In secondary analyses, when analyzing the associations between replacement of red meat with legumes and HCC or ICC separately, weak evidence of a higher risk of HCC was observed (Table S2, total red meat: HR: 1.06, 95% CI: 0.97-1.16; unprocessed red meat: HR: 1.04, 95% CI:

0.95-1.15; processed red meat: HR: 1.10, 95% CI: 0.96-1.27). This association was opposite and inverse for replacement of total red meat and unprocessed red meat and ICC (total red meat: HR: 0.97, 95% CI: 0.89-1.05; unprocessed red meat: HR: 0.94, 95% CI: 0.87-1.02) but not for processed red meat (HR: 1.07, 95% CI: 0.93-1.23, Table S2). The magnitude or direction of associations were not significantly different across strata of liver cancer types.

In the adjusted non-substitution analysis, only the first quartile of legume intake (mean intake 6.3 grams/day) was associated with a lower risk of liver cancer, compared to no intake (HR: 0.60, 95% CI: 0.36-0.99); no associations were observed for quartiles 2, 3 or 4 compared to no intake (Table S3).

In sensitivity analyses, excluding participants based on high alcohol intake, implausible energy intake, any liver disease or cancer before baseline, or fewer than 3 completed Oxford WebQs did not alter the estimates appreciably. Similar results were also found when including death registries as a source of liver cancer cases and when excluding waist circumference from the fully adjusted analysis and soy milk from the food substitutions (Table S4).

## Discussion

Contrary to our hypothesis, this study showed little evidence of an association between replacing 15 g/day of unprocessed or processed red meat with legumes on risk of primary liver cancer. The estimates were robust to sensitivity analyses. When investigating liver cancer types separately, replacing total red meat and unprocessed red meat with legumes showed some weak evidence of an inverse association with ICC but with wide confidence intervals. The results for legume intake without specified substitutions did not show a clear pattern of associations.

The prospective longitudinal design of this study established temporality between the diet exposure and liver cancer outcome, and the large sample size enabled analyses of a rare cancer. Further, our specified substitution analyses have some strengths in contrast to traditional methods in nutritional epidemiology through examining the association of consuming a food or nutrient while holding all other foods constant. The substitution is easily interpretable and reflects the implications that a higher intake of a food is at the expense of a lower intake of another food. A limitation of this research design was that the low intake of the substituted foods in this population restricted the size of the substitution, which may in turn have restricted findings of clinical relevance.

Information on dietary intake was collected using self-reported 24-hour diet recalls, which may have introduced

measurement error partly due to the limited ability of 24-hour recalls to estimate habitual dietary intake. However, estimates were robust to exclusion of participants with fewer than three completed Oxford WebQs, indicating that increasing number of dietary measurements to account for some of the natural fluctuations in dietary intake over time made little difference to our results. A validation study of the Oxford WebQ found person-specific biases for correlation with true intakes for some nutrients, particularly for participants with a higher BMI [27]. Adjustment for BMI was not included in this current study. However, adjusting for waist circumference did not change the estimates appreciably, indicating that such errors do not explain our results. Finally, by specifying that the dietary exposure was collected on at least two occasions, the study population suffered considerable attrition. This is unlikely to be completely at random, and most likely resulted in a study population with greater focus on their dietary habits compared to the general population. For example, the mean intake of processed meat was low in our study population. If a diet consisting of higher intakes of healthier plant-based foods is associated with lower liver cancer incidence, our study population may be at lower risk overall, thus reducing the power of our study to detect an association.

Registries used to determine a diagnosis of liver cancer were incomplete or not up-to-date at the time of analysis [34]. Data from external providers, such as 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 [35]. This could introduce misclassification of the outcome, as individuals with liver cancer may not be identified as cases. However, the estimates were robust in a sensitivity analysis that included death registries as an additional source of liver cancer diagnoses to accommodate missing outcome events. Incorrectly classifying non-cases as cases would lead to attenuation of our results, but this is unlikely due to register linkage. Though health registries may have been only partially up to date, using registries almost eliminates selection bias due to loss to follow-up.

The relatively low number of events limited the possibility to adjust for confounding factors. Excessive adjustment parameters per event can compromise the validity of the multivariable Cox regression model, potentially causing biased estimates. To ensure statistical validity, at least 10 events per variable were aimed for in the main analysis by limiting the number of adjustment levels, using fewer and broader food groups, and fewer levels for categorical covariates. This approach was guided by our *a priori* causal assumptions. Although this method helped maintain statistical validity, it may have increased residual confounding by diluting the importance of specific food groups. Additionally, risk factors that could

not be adjusted for, such as aflatoxin B1, a known liver carcinogen, may have contributed to additional residual confounding.

Contrary to our hypothesis, replacing processed red meat with legumes was associated with a non-significant increase in risk of primary liver cancer, with a greater effect size compared to unprocessed red meat. This pattern persisted across all sensitivity analyses. However, the estimates for processed red meat were labeled with less confidence, partly due to the low median intake. The findings of this current study align with other research in the UK Biobank, where unprocessed red meat intake was associated with a non-significant increase in liver cancer risk, with a greater effect size than processed meat (both white and red meat) [39]. This supports the notion that processed meat may not be associated with liver cancer risk in this population.

The literature on food substitutions, particularly in relation to liver cancer, is sparse. A recent meta-analysis of observational studies including approximately 350,000 individuals and 2125 liver cancer cases found a non-linear dose-response relationship between legume intake and liver cancer risk, with a protective effect observed between intakes of 8 g/day to 40 g/day [40]. This somewhat contrasts with our findings where any increase above 6.3 g/day of legumes was not associated with a lower risk of liver cancer, compared to no legume intake. One recent meta-analysis of observational studies showed no association between red or processed meat intake and HCC [41] while another found a positive association between processed meat and HCC [42]. Another study examined replacement of animal-based protein sources with plant-based protein sources and NAFLD risk in two cohorts and found a near significant decrease in NAFLD when replacing processed meat, but not unprocessed red meat, with legumes in one cohort and a near significant increase in NAFLD risk when replacing total red and processed meat with legumes in another cohort [20].

Red meat is the main source of exogenous heme iron which catalyzes lipid peroxidation of LDL-cholesterol, leading to DNA-damage [43]. Heterocyclic amines (HCAs) are formed when red meat is cooked at high temperatures and for a long time. Further, additives such as nitrate, nitrite and other N-nitroso compounds (NOCs) are often added in the processing of red meat and may, along with HCAs, constitute the carcinogenicity of processed red meat [44, 45, 46]. On the contrary, legumes are high in dietary fibres which are linked to reduced risks of cardiovascular diseases and several

cancers [47, 48]. Despite the fact that replacement of red meat with legumes will, inevitably, increase intake of dietary fibres and lower intake of possible carcinogens, this study found no association on risk of liver cancer from this food substitution. Soy milk is low in fibres and

did constitute a substantial amount of the legumes food group which would attenuate the difference in fiber intake and a possible beneficial association from replacing red meat with legumes. However, removing soy milk from food substitutions did not alter the results appreciably.

## Conclusion

Overall, little evidence of an association between replacing red meat with legumes and liver cancer was observed. These results should be interpreted with caution due to the low intake of the substituted foods and few liver cancer cases. Further research in larger study populations with longer follow-up time is warranted.

# Tables

Table 1. Baseline characteristics of UK Biobank participants who completed  $\geq 2$  Oxford WebQ dietary recalls.

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

<sup>1</sup>Median (IQR) for continuous variables; n (%) for categorical variables

<sup>2</sup>High: College or University degree; Intermediate: A levels/AS levels, O levels/GCSEs, or equivalent; Low: none of the previously mentioned.

<sup>3</sup>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  $\geq 2$  Oxford WebQ dietary recalls.**

	Cohort	Liver cancer
Daily food intake	N = 126,744 <sup>1</sup>	N = 173 <sup>1</sup>
<b>Total food intake</b>		
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)
<b>Food groups, g/day</b>		
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 <sup>2</sup>	475 (361, 603)	448 (322, 604)
Healthy plant-based foods <sup>3</sup>	1,806 (1,454, 2,198)	1,791 (1,365, 2,158)
Unhealthy plant-based foods <sup>4</sup>	472 (324, 662)	491 (365, 698)
Alcoholic beverages	132 (0, 342)	144 (0, 375)

<sup>1</sup>Median (IQR)

<sup>2</sup>Other animal-based foods include poultry, fish, dairy, eggs, and mixed dishes with animal products.

<sup>3</sup>Healthy plant-based foods include whole grains, vegetables, fruits, nuts, plant oils, and beverages (coffee, tea, water).

<sup>4</sup>Unhealthy plant-based foods include refined grains, potatoes, mixed vegetarian dishes, sweets and snacks, fruit juice, and sugar sweetened beverages.

Table 3. **Replacing 15 g/day of total red meat, unprocessed red meat, and processed meat with legumes and hazard ratios and 95% confidence intervals for primary liver cancer.**

	Model 1 <sup>1</sup>	Model 2 <sup>2</sup>
15 g/day of legumes replacing:	HR (95% CI)	HR (95% CI)
Total red meat	0.99 (0.93-1.05)	1.02 (0.96-1.08)
Unprocessed red meat	0.97 (0.91-1.03)	1.00 (0.94-1.06)
Processed red meat	1.04 (0.94-1.15)	1.09 (0.99-1.21)

<sup>1</sup>Multivariate Cox proportional hazards regression model adjusted for age (as underlying timescale), other food groups, and total food intake, and additionally stratified on sex, age, and attended assessment centre.

<sup>2</sup>Further adjusted for educational level, Townsend deprivation index, living alone, physical activity, smoking, alcohol intake, and waist circumference.



## Figures

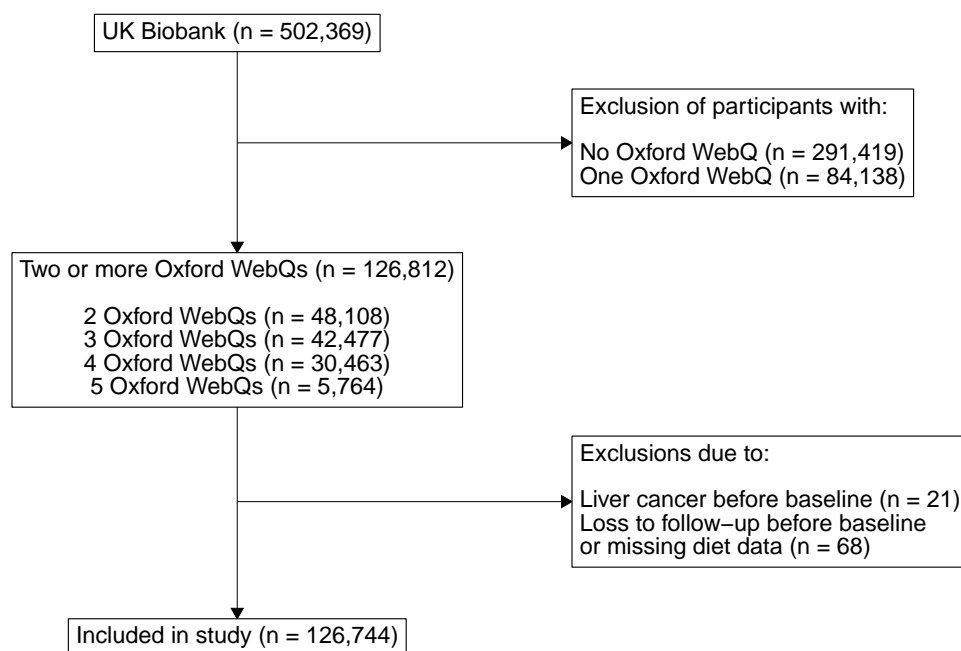


Figure 1. Flowchart of included participants. Not all UK Biobank participants were invited to complete an Oxford WebQ. Only the last 70,000 participants to visit an assessment center were asked to complete an Oxford WebQ at the end of their visit. Further Oxford WebQs were sent to 320,000 participants who provided an e-mail address.

## Supplementary materials

Table S1. Summary of included foods for each food group.

Food group	Includes
<b>Legumes</b>	Soy-based desserts, baked beans, pulses, soy drinks (including calcium fortified), tofu-based products, hummus, peas
<b>Red meat</b>	Beef, lamb, other meat including offal, pork
<b>Processed meat</b>	Sausages, bacon (with and without fat), ham, liver pate
<b>Animal-based foods</b>	Poultry, fish, dairy, eggs, mixed dishes, and sauces and condiments
<b>Healthy plant-based foods</b>	Whole grains, fruits, nuts, plant oils, beverages (water, tea and coffee), vegetables
<b>Unhealthy plant-based foods</b>	Refined cereals, potatoes, fruit juice, mixed dishes (vegetarian), sweets & snacks, and sugar sweetened beverages
<b>Alcoholic beverages</b>	Beer and cider, spirits and other alcoholic drinks, fortified wine, red and rose wine, white wine

Table S2. Replacing 15 g/day 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 <sup>1</sup>	Model 2 <sup>2</sup>
15 g/day of legumes replacing:	HR (95% CI)	HR (95% CI)
<b>Hepatocellular carcinoma (n = 87)</b>		
Total red meat	1.02 (0.94-1.11)	1.06 (0.97-1.16)
Unprocessed red meat	1.02 (0.93-1.11)	1.04 (0.95-1.15)
Processed red meat	1.04 (0.90-1.19)	1.10 (0.96-1.27)
<b>Intrahepatic cholangiocarcinoma (n = 100)</b>		
Total red meat	0.94 (0.87-1.02)	0.97 (0.89-1.05)
Unprocessed red meat	0.92 (0.85-1.00)	0.94 (0.87-1.02)
Processed red meat	1.03 (0.90-1.18)	1.07 (0.93-1.23)

<sup>1</sup>Multivariate Cox proportional hazards regression model adjusted for age (as underlying timescale), other food groups, and total food intake, and additionally stratified on sex, age, and attended assessment centre.

<sup>2</sup>Further adjusted for educational level, Townsend deprivation index, living alone, physical activity, smoking, alcohol intake, and waist circumference.

Table S3. No intake of legumes vs. quartiles of daily legume intake and hazard ratios and 95% confidence intervals for primary liver cancer.

Characteristic	Mean daily legume intake	Model 1 <sup>1</sup>	Model 2 <sup>2</sup>
		HR (95% CI)	HR (95% CI)
Categories:			
No intake	0.00	—	—
Q1	6.3	0.59 (0.35-0.98)	0.60 (0.36-0.99)
Q2	16	0.88 (0.57-1.35)	0.90 (0.58-1.38)
Q3	34	0.73 (0.46-1.17)	0.74 (0.47-1.19)
Q4	109	0.98 (0.64-1.52)	1.07 (0.69-1.66)

<sup>1</sup>Multivariate Cox proportional hazards regression model adjusted for age (as underlying timescale), other food groups, and total food intake, and additionally stratified on sex, age, and attended assessment centre.

<sup>2</sup>Further adjusted for educational level, Townsend deprivation index, living alone, physical activity, smoking, alcohol intake, and waist circumference.

Table S4. Sensitivity analyses

	Exclusion of participants with:					Exclusion of:		
	High alcohol intake <sup>1</sup>	Implausible food intake <sup>2</sup>	Liver disease before baseline <sup>3</sup>	Any cancer before baseline <sup>4</sup>	Fewer than 3 Oxford WebQs <sup>5</sup>	Death register as source of liver cancer events <sup>6</sup>	Waist circumference from analysis <sup>7</sup>	Soy milk from food substitutions <sup>8</sup>
	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)
<b>15 g/day of legumes replacing:</b>								
Total red meat	1.00 (0.94-1.06)	1.01 (0.95-1.07)	1.04 (0.96-1.12)	0.99 (0.93-1.06)	1.02 (0.96-1.08)	1.00 (0.94-1.06)	1.03 (0.96-1.11)	1.03 (0.94-1.12)
Unprocessed red meat	0.98 (0.92-1.05)	0.99 (0.93-1.05)	1.02 (0.94-1.11)	0.97 (0.90-1.04)	1.00 (0.94-1.07)	0.98 (0.92-1.05)	1.00 (0.93-1.08)	1.01 (0.92-1.11)
Processed red meat	1.06 (0.95-1.18)	1.08 (0.98-1.20)	1.11 (0.97-1.27)	1.08 (0.96-1.20)	1.07 (0.98-1.18)	1.06 (0.96-1.17)	1.15 (1.01-1.30)	1.11 (0.98-1.25)

<sup>1</sup>Exclusion of the upper decile of alcohol intake (g/day) by sex. n cases = 150.<sup>2</sup>Exclusion of participants below the 2.5th percentile and above the 97.5th percentile of energy intake (kJ/day) by sex. n cases = 164.<sup>3</sup>ICD10 codes: K70-79, Z94.4, I85, I86.4, and E83.0-1. ICD9 codes: 5710-5745, 0700-0709, V427 and 2750-2751. n cases = 151.<sup>4</sup>ICD10 codes: C00-C97 and D00-D48. ICD9 codes: 1400-2399. n cases = 129.<sup>5</sup>n cases = 109.<sup>6</sup>n cases = 183.<sup>7</sup>n cases = 173.<sup>8</sup>Soy milk was removed from the legumes food group and moved to the food group healthy plant-based foods. n cases = 173.



## References

- [1] Nader N. Massarweh and Hashem B. El-Serag. “Epidemiology of Hepatocellular Carcinoma and Intrahepatic Cholangiocarcinoma”. In: *Cancer Control* 24.3 (Sept. 1, 2017), p. 107327481772924. DOI: 10.1177/1073274817729245.
- [2] Zobair M. Younossi et al. “Global epidemiology of nonalcoholic fatty liver disease—Meta-analytic assessment of prevalence, incidence, and outcomes”. In: *Hepatology* 64.1 (Feb. 22, 2016), pp. 73–84. DOI: 10.1002/hep.28431.
- [3] Zobair M Younossi et al. “Epidemiology of chronic liver diseases in the USA in the past three decades”. In: *Gut* 69.3 (July 31, 2019), pp. 564–568. DOI: 10.1136/gutjnl-2019-318813.
- [4] Wen Guo et al. “Diet and Risk of Non-Alcoholic Fatty Liver Disease, Cirrhosis, and Liver Cancer: A Large Prospective Cohort Study in UK Biobank”. In: *Nutrients* 14.24 (Dec. 15, 2022), p. 5335. DOI: 10.3390/nu14245335. URL: <http://dx.doi.org/10.3390/nu14245335>.
- [5] Chris Estes et al. “Modeling the epidemic of non-alcoholic fatty liver disease demonstrates an exponential increase in burden of disease”. In: *Hepatology* 67.1 (Dec. 1, 2017), pp. 123–133. DOI: 10.1002/hep.29466.
- [6] Shahid A. Khan, Simona Tavolari, and Giovanni Brandi. “Cholangiocarcinoma: Epidemiology and risk factors”. In: *Liver International* 39.S1 (Mar. 24, 2019), pp. 19–31. DOI: 10.1111/liv.14095.
- [7] Martha M. Kirstein and Arndt Vogel. “Epidemiology and Risk Factors of Cholangiocarcinoma”. In: *Visceral Medicine* 32.6 (2016), pp. 395–400. DOI: 10.1159/000453013.
- [8] Annika Bergquist and Erik von Seth. “Epidemiology of cholangiocarcinoma”. In: *Best Pract Res Clin Gastroenterol* 29.2 (Apr. 2015), pp. 221–232. DOI: 10.1016/j.bpg.2015.02.003.
- [9] Nicha Wongjarupong et al. “Non-alcoholic fatty liver disease as a risk factor for cholangiocarcinoma: a systematic review and meta-analysis”. In: *BMC Gastroenterology* 17.1 (Dec. 2017). DOI: 10.1186/s12876-017-0696-4.
- [10] Salvatore Corrao, Giuseppe Natoli, and Christiano Argano. “Nonalcoholic fatty liver disease is associated with intrahepatic cholangiocarcinoma and not with extrahepatic form: definitive evidence from meta-analysis and trial sequential analysis”. In: *Eur J Gastroenterol Hepatol* 33.1 (Feb. 21, 2020), pp. 62–68. DOI: 10.1097/meg.0000000000001684.
- [11] Wei Zhang et al. “Vegetable-based dietary pattern and liver cancer risk: Results from the Shanghai Women’s and Men’s Health Studies”. In: *Cancer Science* 104.10 (Aug. 5, 2013), pp. 1353–1361. DOI: 10.1111/cas.12231. URL: <http://dx.doi.org/10.1111/cas.12231>.
- [12] Yang Yang et al. “Increased Intake of Vegetables, But Not Fruit, Reduces Risk for Hepatocellular Carcinoma: A Meta-analysis”. In: *Gastroenterology* 147.5 (Nov. 2014), pp. 1031–1042. DOI: 10.1053/j.gastro.2014.08.005. URL: <http://dx.doi.org/10.1053/j.gastro.2014.08.005>.
- [13] Xing Liu et al. “Higher intake of whole grains and dietary fiber are associated with lower risk of liver cancer and chronic liver disease mortality”. In: *Nature Communications* 12.1 (Nov. 4, 2021). DOI: 10.1038/s41467-021-26448-9. URL: <http://dx.doi.org/10.1038/s41467-021-26448-9>.
- [14] Abhishek Bhurwal et al. “Inverse Association of Coffee with Liver Cancer Development: An Updated Systematic Review and Meta-analysis”. In: *Journal of Gastrointestinal and Liver Diseases* (Aug. 19, 2020). DOI: 10.15403/jgld-805. URL: <http://dx.doi.org/10.15403/jgld-805>.
- [15] Effie Viguiouk et al. “Associations between Dietary Pulses Alone or with Other Legumes and Cardiometabolic Disease Outcomes: An Umbrella Review and Updated Systematic Review and Meta-analysis of Prospective Cohort Studies”. In: *Advances in Nutrition* 10 (Nov. 2019), S308–S319. DOI: 10.1093/advances/nmz113.
- [16] Shaoyue Jin and Youjin Je. “Nuts and legumes consumption and risk of colorectal cancer: a systematic review and meta-analysis”. In: *European Journal of Epidemiology* 37.6 (May 27, 2022), pp. 569–585. DOI: 10.1007/s10654-022-00881-6.
- [17] Yanan Ma et al. “Meat intake and risk of hepatocellular carcinoma in two large US prospective cohorts of women and men”. In: *International Journal of Epidemiology* 48.6 (July 13, 2019), pp. 1863–1871. DOI: 10.1093/ije/dyz146. URL: <http://dx.doi.org/10.1093/ije/dyz146>.
- [18] Véronique Bouvard et al. “Carcinogenicity of consumption of red and processed meat”. In: *The Lancet Oncology* 16.16 (Dec. 2015), pp. 1599–1600. DOI: 10.1016/s1470-2045(15)00444-1. URL: [http://dx.doi.org/10.1016/S1470-2045\(15\)00444-1](http://dx.doi.org/10.1016/S1470-2045(15)00444-1).
- [19] UN. *Food and Climate Change: Healthy diets for a healthier planet*. Web Page. accessed on 21 May 2024. URL: <https://www.un.org/en/climatechange/science/climate-issues/food>.

- [20] Shunming Zhang et al. “Protein foods from animal sources and risk of nonalcoholic fatty liver disease in representative cohorts from North and South China”. In: *Journal of Internal Medicine* 293.3 (Nov. 26, 2022), pp. 340–353. DOI: 10.1111/joim.13586.
- [21] Niels Bock et al. *Protocol: Legume consumption and risk of primary liver cancer*. June 2024. DOI: 10.5281/zenodo.11670569. URL: <https://doi.org/10.5281/zenodo.11670569>.
- [22] Cathie Sudlow et al. “UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age”. In: *PLOS Medicine* 12.3 (Mar. 31, 2015), e1001779. DOI: 10.1371/journal.pmed.1001779.
- [23] UK Biobank. *Order of Data Collection*. Web Page. accessed on 21 May 2024. 2011. URL: <https://biobank.ndph.ox.ac.uk/ukb/ukb/docs/Orderofdatacollection.pdf>.
- [24] UK Biobank. *24-hour dietary recall questionnaire (Oxford WebQ)*. Web Page. accessed on 21 May 2024. 2024. URL: <https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/DietWebQ.pdf>.
- [25] Carmen Piernas et al. “Describing a new food group classification system for UK biobank: analysis of food groups and sources of macro- and micronutrients in 208,200 participants”. In: *European Journal of Nutrition* 60.5 (Mar. 25, 2021), pp. 2879–2890. DOI: 10.1007/s00394-021-02535-x.
- [26] Bette Liu et al. “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”. In: *Public Health Nutrition* 14.11 (June 1, 2011), pp. 1998–2005. DOI: 10.1017/s1368980011000942.
- [27] Darren C Greenwood et al. “Validation of the Oxford WebQ Online 24-Hour Dietary Questionnaire Using Biomarkers”. In: *American Journal of Epidemiology* 188.10 (July 18, 2019), pp. 1858–1867. DOI: 10.1093/aje/kwz165.
- [28] Alysha S. Thompson et al. “Association of Healthful Plant-based Diet Adherence With Risk of Mortality and Major Chronic Diseases Among Adults in the UK”. In: *JAMA Network Open* 6.3 (Mar. 28, 2023), e234714. DOI: 10.1001/jamanetworkopen.2023.4714.
- [29] Yoriko Heianza et al. “Healthful plant-based dietary patterns, genetic risk of obesity, and cardiovascular risk in the UK biobank study”. In: *Clinical Nutrition* 40.7 (July 2021), pp. 4694–4701. DOI: 10.1016/j.clnu.2021.06.018.
- [30] Ambika Satija et al. “Healthful and Unhealthful Plant-Based Diets and the Risk of Coronary Heart Disease in U.S. Adults”. In: *Journal of the American College of Cardiology* 70.4 (July 2017), pp. 411–422. DOI: 10.1016/j.jacc.2017.05.047.
- [31] Ambika Satija et al. “Plant-Based Dietary Patterns and Incidence of Type 2 Diabetes in US Men and Women: Results from Three Prospective Cohort Studies”. In: *PLOS Medicine* 13.6 (June 14, 2016). Ed. by Steven C Moore, e1002039. DOI: 10.1371/journal.pmed.1002039.
- [32] Frances E. Thompson and Amy F. Subar. “Dietary Assessment Methodology”. In: Elsevier, 2013, pp. 5–46. DOI: 10.1016/b978-0-12-391884-0.00001-9.
- [33] Mirjana Gurinović et al. “Nutritional Assessment”. In: Elsevier, 2017. DOI: 10.1016/b978-0-08-100596-5.21180-3.
- [34] UK Biobank. *Health Outcomes Overview*. Web Page. accessed on 21 May 2024. 2024. URL: <https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/HealthOutcomesOverview.pdf>.
- [35] UK Biobank. *Data providers and dates of data availability*. Web Page. accessed on 21 May 2024. 2023. URL: [https://biobank.ndph.ox.ac.uk/ukb/exinfo.cgi?src=Data\\_providers\\_and\\_dates](https://biobank.ndph.ox.ac.uk/ukb/exinfo.cgi?src=Data_providers_and_dates).
- [36] Daniel B Ibsen et al. “Food substitution models for nutritional epidemiology”. In: *The American Journal of Clinical Nutrition* 113.2 (Feb. 2021), pp. 294–303. DOI: 10.1093/ajcn/nqaa315.
- [37] Georgia D Tomova, Mark S Gilthorpe, and Peter WG Tennant. “Theory and performance of substitution models for estimating relative causal effects in nutritional epidemiology”. In: *The American Journal of Clinical Nutrition* 116.5 (Nov. 2022), pp. 1379–1388. DOI: 10.1093/ajcn/nqac188.
- [38] William Michael Landau. “The targets R package: a dynamic Make-like function-oriented pipeline toolkit for reproducibility and high-performance computing”. In: *Journal of Open Source Software* 6.57 (2021), p. 2959. URL: <https://doi.org/10.21105/joss.02959>.
- [39] Anika Knuppel et al. “Meat intake and cancer risk: prospective analyses in UK Biobank”. In: *International Journal of Epidemiology* 49.5 (Aug. 20, 2020), pp. 1540–1552. DOI: 10.1093/ije/dyaa142.
- [40] Ke Liu et al. “Associations between food groups and liver cancer: a systematic review and meta-analysis of observational studies”. In: *Nutrition Journal* 22.1 (June 22, 2023). DOI: 10.1186/s12937-023-00858-5.

- [41] Yan Di et al. “Association of meat consumption with the risk of gastrointestinal cancers: a systematic review and meta-analysis”. In: *BMC Cancer* 23.1 (Aug. 23, 2023). DOI: 10.1186/s12885-023-11218-1.
- [42] Jinchuan Yu et al. “Meat Intake and the Risk of Hepatocellular Carcinoma: A Meta-Analysis of Observational Studies”. In: *Nutrition and Cancer* 74.9 (May 18, 2022), pp. 3340–3350. DOI: 10.1080/01635581.2022.2077386. URL: <http://dx.doi.org/10.1080/01635581.2022.2077386>.
- [43] Viktoria Jeney et al. “Pro-oxidant and cytotoxic effects of circulating heme”. In: *Blood* 100.3 (Aug. 1, 2002), pp. 879–887. DOI: 10.1182/blood.v100.3.879. URL: <http://dx.doi.org/10.1182/blood.v100.3.879>.
- [44] James S Felton et al. “Health risks of heterocyclic amines”. In: *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 376.1-2 (May 1997), pp. 37–41. DOI: 10.1016/s0027-5107(97)00023-7. URL: [http://dx.doi.org/10.1016/s0027-5107\(97\)00023-7](http://dx.doi.org/10.1016/s0027-5107(97)00023-7).
- [45] Yupeng Li and Stephen S. Hecht. “Metabolic Activation and DNA Interactions of Carcinogenic N-Nitrosamines to Which Humans Are Commonly Exposed”. In: *International Journal of Molecular Sciences* 23.9 (Apr. 20, 2022), p. 4559. DOI: 10.3390/ijms23094559. URL: <http://dx.doi.org/10.3390/ijms23094559>.
- [46] Monireh Sadat Seyyedsalehi et al. “Association of Dietary Nitrate, Nitrite, and N-Nitroso Compounds Intake and Gastrointestinal Cancers: A Systematic Review and Meta-Analysis”. In: *Toxics* 11.2 (Feb. 17, 2023), p. 190. DOI: 10.3390/toxics11020190. URL: <http://dx.doi.org/10.3390/toxics11020190>.
- [47] Christina C. Dahm et al. “Role of dietary fibres in cardiometabolic diseases”. In: *Current Opinion in Clinical Nutrition and Metabolic Care* 27.4 (May 16, 2024), pp. 355–360. DOI: 10.1097/mco.0000000000001047. URL: <http://dx.doi.org/10.1097/MCO.0000000000001047>.
- [48] Jun Hu et al. “Use of Dietary Fibers in Reducing the Risk of Several Cancer Types: An Umbrella Review”. In: *Nutrients* 15.11 (May 30, 2023), p. 2545. DOI: 10.3390/nu15112545. URL: <http://dx.doi.org/10.3390/nu15112545>.