MDPI

Article

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], Luke W. Johnston[1,2], Daniel B. Ibsen[1,2,3,4], Christina C. Dahm[1,*]

[1] Department of Public Health, Aarhus University, Aarhus, Denmark;
[2] Steno Diabetes Center Aarhus, Aarhus University Hospital, Aarhus N, Denmark;
[3] Department of Nutrition, Exercise and Sports, University of Copenhagen, Copenhagen, Denmark;
[4] MRC Epidemiology Unit, School of Clinical Medicine, University of Cambridge, Cambridge, United Kingdom;
* Correspondence: CCD@ph.au.dk.

Abstract: Purpose: Primary liver cancer is on the rise worldwide, 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 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 $\geq$ 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 and processed red meat on liver cancer risk, using the leave-one-out food substitution model. Results: During a median follow-up time of 11.3 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: 0.98 (95% CI 0.93-1.04)), unprocessed red meat (HR: 0.97 (95% CI 0.91-1.03)) or processed red meat (HR: 1.02 (95% CI 0.93-1.13)). Conclusion: Overall, little evidence of an association between replacing red meat with legumes and liver cancer was observed. Further research in larger study populations with longer follow-up time is warranted.

Keywords: Food Substitutions; liver cancer; red meat; legumes.

Citation: Bock, N.; Langmann, F; Johnston, LW; Ibsen, DB; Dahm, CC. Substitution of red meat with legumes and risk of primary liver cancer in 126,744 UK Biobank participants: a prospective cohort study. *Nutrients* **2024**, *1*, 0. https://doi.org/

Received:
Revised:
Accepted:
Published:

This research has been conducted using the UK Biobank Resource under Application Number 81520.

1. Background

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% by 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 infectous 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 [11–14]. The protective properties of these foods are proposedly due to 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 anti-carcinogenic properties with observed inverse associations with cardiovascular disease and colorectal cancer [15,16]. Two large prospective cohort studies 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 an increase in intake of one food is at the expense of a concomitantly decreased intake of another food. Studies on substituting plant-based proteins for animal-based proteins are important if we are to lower the climate impacts of our diets [17]. Although previous research has investigated substitution of animal-based proteins with plant-based proteins in relations to NAFLD [18], 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 free-living population.

2. Research Design and Methods

2.1. Study population

The UK Biobank, a population-based prospective cohort, was initiated in 2006. [19] 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 (waist circumference) were collected via physical measurements [20].

2.2. 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 [21]. 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 [22]. The Oxford WebQ has been validated against interviewer-based 24-hour dietary recalls and biomarkers [23,24].

Researchers defined 79 food groups and 14 beverage groups from the Oxford WebQ using the UK National Diet and Nutrition Survey categories [22]. These food and beverage groups were used when defining the food groups used in the substitution analyses (Table S1). 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 (Table S1). Animal-based and healthy and unhealthy plant-based food foods were grouped based on plant-based diet indices from previous studies [25–28].

As a single 24-hour dietary recall does not assess habitual dietary intake and variation in diet over time at an individual level [29,30], only participants who completed two or more Oxford WebQs were eligible for inclusion in this study.

2.3. 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 [31,32].

2.4. Assessment of confounders

Confounders were defined *a priori* from a review of the background literature and illustrated using directed acyclic graphs (Figure **??**). 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.

2.5. The substitution model

The substitution analyses were conducted by by modelling replacement of an equal mass of meat with legumes. The portion size of the substitution was set to 15 g of legumes for 15 g 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 along with a variable for total food intake are included, except the food group that are to be substituted [33]. To estimate substitution of 15 g of all red meats (red and processed) with 15 g of legumes, the following model was defined:

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

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

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

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

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

The performance of the leave-one-out model when modeling equal mass substitutions has been validated against simulated data [34].

2.6. 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 (as the underlying timescale), total weight of food and beverage intake, and all other food groups to fit the substitution model. Model 2 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 association of legume intake with liver cancer regardless 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 (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 3 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 more up-to-date, were included in a sensitivity analysis to test for outcome misclassification. 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 modeled as the fully adjusted models in the main analyses.

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

3. 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 11.3 years, 173 participants developed liver cancer. Participants 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 **??**: 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 with further adjustments. There was weak evidence of an association between replacement of processed red meat with legumes (HR: 1.09, 95% CI: 0.98-1.20; Table **??**).

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

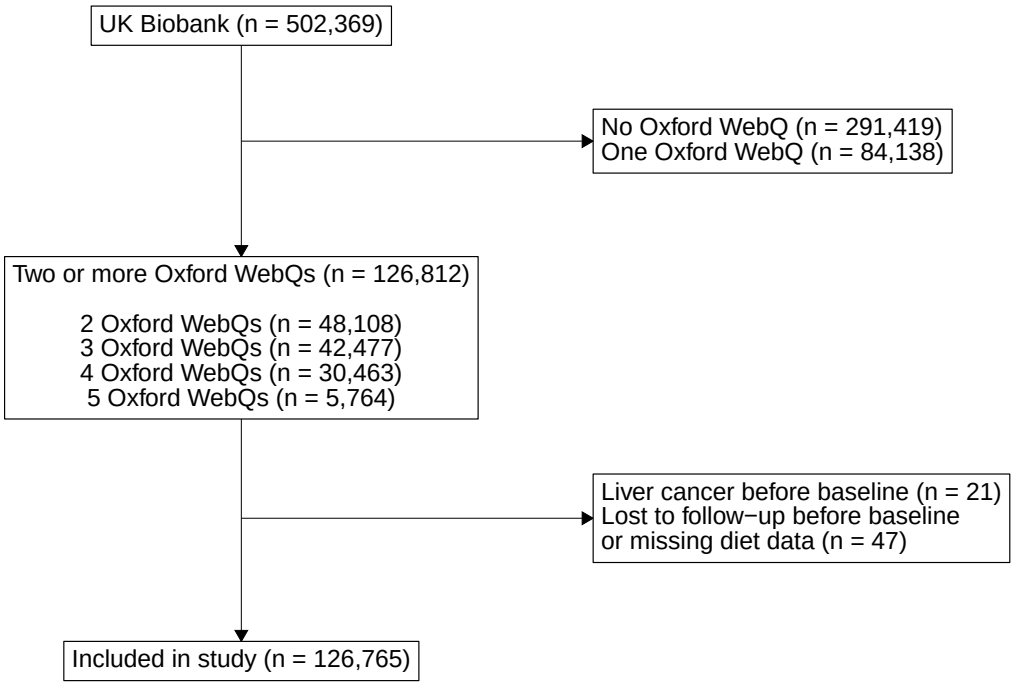

Figure 1. Flowchart of included participants. Missing diet data were merged with loss to follow-up before baseline due to n being less than 5. It should be noted that not all UK Biobank participants were invited to complete an Oxfords WebQ. Only the last 70,000 participants to visit an assessment center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

was observed (Table S2, total red meat: HR: 1.06, 95% CI: 0.97-1.16; unprocessed red meat: HR: 1.05, 95% CI: 0.96-1.15; processed red meat: HR: 1.09, 95% CI: 0.95-1.26). This was opposite for replacement of total red meat and unprocessed red meat and ICC (total red meat: HR: 0.97, 95% CI: 0.90-1.05; unprocessed red meat: HR: 0.95, 95% CI: 0.87-1.03) but not for processed red meat (HR: 1.07, 95% CI: 0.93- 1.22, 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: 0.59, 95% CI: 0.35-0.98); 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 (Table S4).

4. Discussion

Contrary to our hypothesis, this study we showed little evidence of an association between replacing 15 g/day of red or processed 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. The results for legume intake without specified substitutions did not show a clear pattern of association.

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

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

Variable	Cohort N = 126,744[1]	Liver cancer 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.

contrast to traditional methods in nutritional epidemiology through examining the effect of consuming a food or nutrient while holding all other foods constant. The substitution is easily interpretable and reflects the implications that an increased intake of a food is at the expense a decreased intake of other foods. In that sense, the food substitution model mimics some aspects of a randomized controlled design. 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 clinically significant effect sizes.

Information on dietary intake was collected using self-reported 24-hour diet recall questionnaires, which may have introduced measurement error partly due 24-hour recalls not reflecting past dietary intake. However, estimates were robust to exclusion of participants with fewer than three completed Oxford WebQs, indicating that at least two 24-hour diet recall measurements were sufficient to account for natural fluctuations in dietary intake over time. A validation study of the Oxford WebQ found some person-specific biases within participants with a higher BMI having greater disparities for correlation with true intakes for some nutrients [24]. We did not adjust for BMI; however, adjusting for waist circumference did not change our estimates significantly, lending some support to sufficient adjustment of person specific bias. Finally, by specifying that the dietary exposure was collected on at least two occasions, our study population suffered considerable attrition. This is unlikely to be completely at random, and most likely resulted in a study population

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.

	Cohort	Liver cancer
Daily food intake	**N = 126,744**[1]	**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 (0.93-1.04)	1.02 (0.96-1.08)
Unprocessed red meat	0.97 (0.91-1.03)	1.00 (0.94-1.07)
Processed red meat	1.02 (0.93-1.13)	1.09 (0.98-1.20)

[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.

with greater focus on their dietary habits. 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.

None of the registries used to determine a diagnosis of liver cancer were complete or up-to-date at the time of analysis [31]. 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 [32]. 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 our ability 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, we aimed for at least 10 events per variable 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 we could not adjust for, such as aflatoxin B1, a known liver carcinogen, may have contributed to 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) [35]. 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. Accordingly, we also conducted an analysis of legume intake without specifying food substitutions. A recent meta-anlysis of observational studies 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 [36]. This somewhat contrasts with our findings where any increase above 6.3 g/day of legumes was not associated with a decreased 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 [37] while another found a positive association between processed meat and HCC [38]. 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 [18].

5. 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.

Supplementary Materials: The following supporting information can be downloaded at:

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

NAFLD	Non-alcoholic fatty liver disease
HCC	Hepatocellular carcinoma
ICC	Intrahepatic cholangiocarcinoma

Appendix A

Table S1. 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, fish, dairy, eggs, mixed dishes, and sauces and condiments
Healthy plant-based foods	Whole grains, fruits, nuts, plant oils, beverages (water, tea and coffee), vegetables
Unhealthy plant-based foods	Refined cereals, potatoes, fruit juice, mixed dishes (vegetarian), sweets & snacks, and sugar sweetened beverages
Alcoholic beverages	Beer and cider, spirits and other alcoholic drinks, fortified wine, red and rose wine, white wine

Table S2. 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 (0.93-1.10)	1.06 (0.97-1.16)
Unprocessed red meat	1.01 (0.92-1.11)	1.05 (0.96-1.15)
Processed red meat	1.01 (0.88-1.16)	1.09 (0.95-1.26)
Intrahepatic cholangiocarcinoma		
Total red meat	0.94 (0.87-1.02)	0.97 (0.90-1.05)
Unprocessed red meat	0.92 (0.85-1.00)	0.95 (0.87-1.03)
Processed red meat	1.02 (0.89-1.17)	1.07 (0.93-1.22)

[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.

Table S3. 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	**Mean daily legume intake**	**HR (95% CI)**	**HR (95% CI)**
Categories:			
No intake	0.00	—	—
Q1	6.3	0.58 (0.35-0.96)	0.59 (0.35-0.98)
Q2	16	0.87 (0.56-1.33)	0.89 (0.58-1.36)
Q3	34	0.75 (0.47-1.19)	0.75 (0.47-1.19)
Q4	109	0.98 (0.64-1.51)	1.06 (0.69-1.64)

[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.

Table S4. 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 of 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 (0.94-1.07)	1.01 (0.94-1.08)	0.99 (0.93-1.06)	1.04 (0.97-1.11)	1.04 (0.97-1.12)	1.02 (0.97-1.09)	1.00 (0.94-1.06)
Unprocessed red meat	0.99 (0.92-1.05)	0.98 (0.91-1.06)	0.97 (0.91-1.04)	1.01 (0.94-1.09)	1.02 (0.94-1.11)	1.01 (0.95-1.07)	0.99 (0.92-1.05)
Processed red meat	1.06 (0.95-1.17)	1.10 (0.98-1.24)	1.07 (0.96-1.20)	1.15 (1.01-1.30)	1.11 (0.97-1.27)	1.07 (0.97-1.18)	1.05 (0.95-1.16)

[1]Exclusion of the upper decile of alcohol intake (g/day) for each sex.
[2]Exclusion of the upper and lower decile of energy intake (kJ/day) 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. Massarweh, N.N.; El-Serag, H.B. Epidemiology of Hepatocellular Carcinoma and Intrahepatic Cholangiocarcinoma. *Cancer Control* **2017**, *24*, 107327481772924. https://doi.org/10.1177/1073274817729245.
2. Younossi, Z.M.; Koenig, A.B.; Abdelatif, D.; Fazel, Y.; Henry, L.; Wymer, M. Global epidemiology of nonalcoholic fatty liver disease—Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* **2016**, *64*, 73–84. https://doi.org/10.1002/hep.28431.
3. Younossi, Z.M.; Stepanova, M.; Younossi, Y.; Golabi, P.; Mishra, A.; Rafiq, N.; Henry, L. Epidemiology of chronic liver diseases in the USA in the past three decades. *Gut* **2019**, *69*, 564–568. https://doi.org/10.1136/gutjnl-2019-318813.
4. Guo, W.; Ge, X.; Lu, J.; Xu, X.; Gao, J.; Wang, Q.; Song, C.; Zhang, Q.; Yu, C. Diet and Risk of Non-Alcoholic Fatty Liver Disease, Cirrhosis, and Liver Cancer: A Large Prospective Cohort Study in UK Biobank. *Nutrients* **2022**, *14*, 5335. https://doi.org/10.3390/nu14245335.
5. Estes, C.; Razavi, H.; Loomba, R.; Younossi, Z.; Sanyal, A.J. Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. *Hepatology* **2017**, *67*, 123–133. https://doi.org/10.1002/hep.29466.
6. Khan, S.A.; Tavolari, S.; Brandi, G. Cholangiocarcinoma: Epidemiology and risk factors. *Liver International* **2019**, *39*, 19–31. https://doi.org/10.1111/liv.14095.
7. Kirstein, M.M.; Vogel, A. Epidemiology and Risk Factors of Cholangiocarcinoma. *Visceral Medicine* **2016**, *32*, 395–400. https://doi.org/10.1159/000453013.
8. Bergquist, A.; von Seth, E. Epidemiology of cholangiocarcinoma. *Best Pract Res Clin Gastroenterol* **2015**, *29*, 221–232. https://doi.org/10.1016/j.bpg.2015.02.003.
9. Wongjarupong, N.; Assavapongpaiboon, B.; Susantitaphong, P.; Cheungpasitporn, W.; Treeprasertsuk, S.; Rerknimitr, R.; Chaiteerakij, R. Non-alcoholic fatty liver disease as a risk factor for cholangiocarcinoma: a systematic review and meta-analysis. *BMC Gastroenterology* **2017**, *17*. https://doi.org/10.1186/s12876-017-0696-4.
10. Corrao, S.; Natoli, G.; Argano, C. Nonalcoholic fatty liver disease is associated with intrahepatic cholangiocarcinoma and not with extrahepatic form: definitive evidence from meta-analysis and trial sequential analysis. *Eur J Gastroenterol Hepatol* **2020**, *33*, 62–68. https://doi.org/10.1097/meg.0000000000001684.
11. Zhang, W.; Xiang, Y.; Li, H.; Yang, G.; Cai, H.; Ji, B.; Gao, Y.; Zheng, W.; Shu, X. Vegetable-based dietary pattern and liver cancer risk: Results from the Shanghai Women's and Men's Health Studies. *Cancer Science* **2013**, *104*, 1353–1361. https://doi.org/10.1111/cas.12231.
12. Yang, Y.; Zhang, D.; Feng, N.; Chen, G.; Liu, J.; Chen, G.; Zhu, Y. Increased Intake of Vegetables, But Not Fruit, Reduces Risk for Hepatocellular Carcinoma: A Meta-analysis. *Gastroenterology* **2014**, *147*, 1031–1042. https://doi.org/10.1053/j.gastro.2014.08.005.
13. Liu, X.; Yang, W.; Petrick, J.L.; Liao, L.M.; Wang, W.; He, N.; Campbell, P.T.; Zhang, Z.F.; Giovannucci, E.; McGlynn, K.A.; et al. Higher intake of whole grains and dietary fiber are associated with lower risk of liver cancer and chronic liver disease mortality. *Nature Communications* **2021**, *12*. https://doi.org/10.1038/s41467-021-26448-9.
14. Bhurwal, A.; Ratta, P.; Yoshitake, S.; Pioppo, L.; Reja, D.; Dellatore, P.; Rustgi, V. Inverse Association of Coffee with Liver Cancer Development: An Updated Systematic Review and Meta-analysis. *Journal of Gastrointestinal and Liver Diseases* **2020**. https://doi.org/10.15403/jgld-805.
15. Viguiliouk, E.; Glenn, A.J.; Nishi, S.K.; Chiavaroli, L.; Seider, M.; Khan, T.; Bonaccio, M.; Iacoviello, L.; Mejia, S.B.; Jenkins, D.J.A.; 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. *Advances in Nutrition* **2019**, *10*, S308–S319. https://doi.org/10.1093/advances/nmz113.
16. Jin, S.; Je, Y. Nuts and legumes consumption and risk of colorectal cancer: a systematic review and meta-analysis. *European Journal of Epidemiology* **2022**, *37*, 569–585. https://doi.org/10.1007/s10654-022-00881-6.
17. UN. Food and Climate Change: Healthy diets for a healthier planet.
18. Zhang, S.; Yan, Y.; Meng, G.; Zhang, Q.; Liu, L.; Wu, H.; Gu, Y.; Wang, X.; Zhang, J.; Sun, S.; et al. Protein foods from animal sources and risk of nonalcoholic fatty liver disease in representative cohorts from North and South China. *Journal of Internal Medicine* **2022**, *293*, 340–353. https://doi.org/10.1111/joim.13586.
19. Sudlow, C.; Gallacher, J.; Allen, N.; Beral, V.; Burton, P.; Danesh, J.; Downey, P.; Elliott, P.; Green, J.; Landray, M.; et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine* **2015**, *12*, e1001779. https://doi.org/10.1371/journal.pmed.1001779.
20. UK Biobank. Order of Data Collection, 2011. Accessed 21 May 2024.
21. UK Biobank. 24-hour dietary recall questionnaire (Oxford WebQ), 2024. Accessed 21 May 2024.
22. Piernas, C.; Perez-Cornago, A.; Gao, M.; Young, H.; Pollard, Z.; Mulligan, A.; Lentjes, M.; Carter, J.; Bradbury, K.; Key, T.J.; 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. *European Journal of Nutrition* **2021**, *60*, 2879–2890. https://doi.org/10.1007/s00394-021-02535-x.
23. Liu, B.; Young, H.; Crowe, F.L.; Benson, V.S.; Spencer, E.A.; Key, T.J.; Appleby, P.N.; Beral, V. 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* **2011**, *14*, 1998–2005. https://doi.org/10.1017/s1368980011000942.

24. Greenwood, D.C.; Hardie, L.J.; Frost, G.S.; Alwan, N.A.; Bradbury, K.E.; Carter, M.; Elliott, P.; Evans, C.E.L.; Ford, H.E.; Hancock, N.; et al. Validation of the Oxford WebQ Online 24-Hour Dietary Questionnaire Using Biomarkers. *American Journal of Epidemiology* **2019**, *188*, 1858–1867. https://doi.org/10.1093/aje/kwz165.
25. Thompson, A.S.; Tresserra-Rimbau, A.; Karavasiloglou, N.; Jennings, A.; Cantwell, M.; Hill, C.; Perez-Cornago, A.; Bondonno, N.P.; Murphy, N.; Rohrmann, S.; et al. Association of Healthful Plant-based Diet Adherence With Risk of Mortality and Major Chronic Diseases Among Adults in the UK. *JAMA Network Open* **2023**, *6*, e234714. https://doi.org/10.1001/jamanetworkopen.2023.4714.
26. Heianza, Y.; Zhou, T.; Sun, D.; Hu, F.B.; Qi, L. Healthful plant-based dietary patterns, genetic risk of obesity, and cardiovascular risk in the UK biobank study. *Clinical Nutrition* **2021**, *40*, 4694–4701. https://doi.org/10.1016/j.clnu.2021.06.018.
27. Satija, A.; Bhupathiraju, S.N.; Spiegelman, D.; Chiuve, S.E.; Manson, J.E.; Willett, W.; Rexrode, K.M.; Rimm, E.B.; Hu, F.B. Healthful and Unhealthful Plant-Based Diets and the Risk of Coronary Heart Disease in U.S. Adults. *Journal of the American College of Cardiology* **2017**, *70*, 411–422. https://doi.org/10.1016/j.jacc.2017.05.047.
28. Satija, A.; Bhupathiraju, S.N.; Rimm, E.B.; Spiegelman, D.; Chiuve, S.E.; Borgi, L.; Willett, W.C.; Manson, J.E.; Sun, Q.; Hu, F.B. Plant-Based Dietary Patterns and Incidence of Type 2 Diabetes in US Men and Women: Results from Three Prospective Cohort Studies. *PLOS Medicine* **2016**, *13*, e1002039. https://doi.org/10.1371/journal.pmed.1002039.
29. Thompson, F.E.; Subar, A.F., Dietary Assessment Methodology; Elsevier, 2013; pp. 5–46. https://doi.org/10.1016/b978-0-12-391884-0.00001-9.
30. Gurinović, M.; Zeković, M.; Milešević, J.; Nikolić, M.; Glibetić, M., Nutritional Assessment; Elsevier, 2017. https://doi.org/10.1016/b978-0-08-100596-5.21180-3.
31. UK Biobank. Health Outcomes Overview, 2024. Accessed 21 May 2024.
32. UK Biobank. Data providers and dates of data availability, 2023. Accessed 21 May 2024.
33. Ibsen, D.B.; Laursen, A.S.D.; Würtz, A.M.L.; Dahm, C.C.; Rimm, E.B.; Parner, E.T.; Overvad, K.; Jakobsen, M.U. Food substitution models for nutritional epidemiology. *The American Journal of Clinical Nutrition* **2021**, *113*, 294–303. https://doi.org/10.1093/ajcn/nqaa315.
34. Tomova, G.; Gilthorpe, M.; Tennant, P. Theory and performance of substitution models for estimating relative causal effects in nutritional epidemiology. *The American Journal of Clinical Nutrition* **2022**, *116*, 1379–1388. https://doi.org/10.1093/ajcn/nqac188.
35. Knuppel, A.; Papier, K.; Fensom, G.K.; Appleby, P.N.; Schmidt, J.A.; Tong, T.Y.N.; Travis, R.C.; Key, T.J.; Perez-Cornago, A. Meat intake and cancer risk: prospective analyses in UK Biobank. *International Journal of Epidemiology* **2020**, *49*, 1540–1552. https://doi.org/10.1093/ije/dyaa142.
36. Liu, K.; Chen, W.; Zhou, Y.; Xu, L.; Sun, X.; Mao, Y.; Ye, D. Associations between food groups and liver cancer: a systematic review and meta-analysis of observational studies. *Nutrition Journal* **2023**, *22*. https://doi.org/10.1186/s12937-023-00858-5.
37. Di, Y.; Ding, L.; Gao, L.; Huang, H. Association of meat consumption with the risk of gastrointestinal cancers: a systematic review and meta-analysis. *BMC Cancer* **2023**, *23*. https://doi.org/10.1186/s12885-023-11218-1.
38. Yu, J.; Liu, Z.; Liang, D.; Li, J.; Ma, S.; Wang, G.; Chen, W. Meat Intake and the Risk of Hepatocellular Carcinoma: A Meta-Analysis of Observational Studies. *Nutrition and Cancer* **2022**, *74*, 3340–3350. https://doi.org/10.1080/01635581.2022.2077386.