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, Aarhus, 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 $\geq$ 2 24-hour diet recall questionnaires. Cox proportional hazards regression models were used to estimate substitution of 15g/day of legumes with 15/day of total red meat, unprocessed red meat and processed red meat on liver cancer risk, using the leave-one-out food substitution 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 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)) was observed. The results were robust to sensitivity analyses.
Conclusion: Overall, no association between substituting red meat with legumes and liver cancer was observed. Further research with longer follow-up time is warranted.

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

Background

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. [1] 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 [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 [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 [4]. The Oxford WebQ has been validated against interviewer-based 24-hour dietary recalls and biomarkers [5, 6].

Researchers defined79 food groups and 14 beverage groups from the Oxford WebQ using the UK National Diet and Nutrition Survey categories [4]. 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 [7–10]. 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 [11, 12], 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) 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 [13, 14].

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 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 [15]. To estimate substitution of 15 g of all red meats (red and processed) with 15 g of legumes, the following model was defined:

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

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

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

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

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

The performance of the leave-one-out model when modeling equal mass substitutions has been validates against simulated data [16].

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 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 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 (exclusion of upper decile of daily alcohol intake for each sex), implausible energy intake (exclusion of upper and lower deciel of each 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 where more up-to-date, 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 modeled 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).

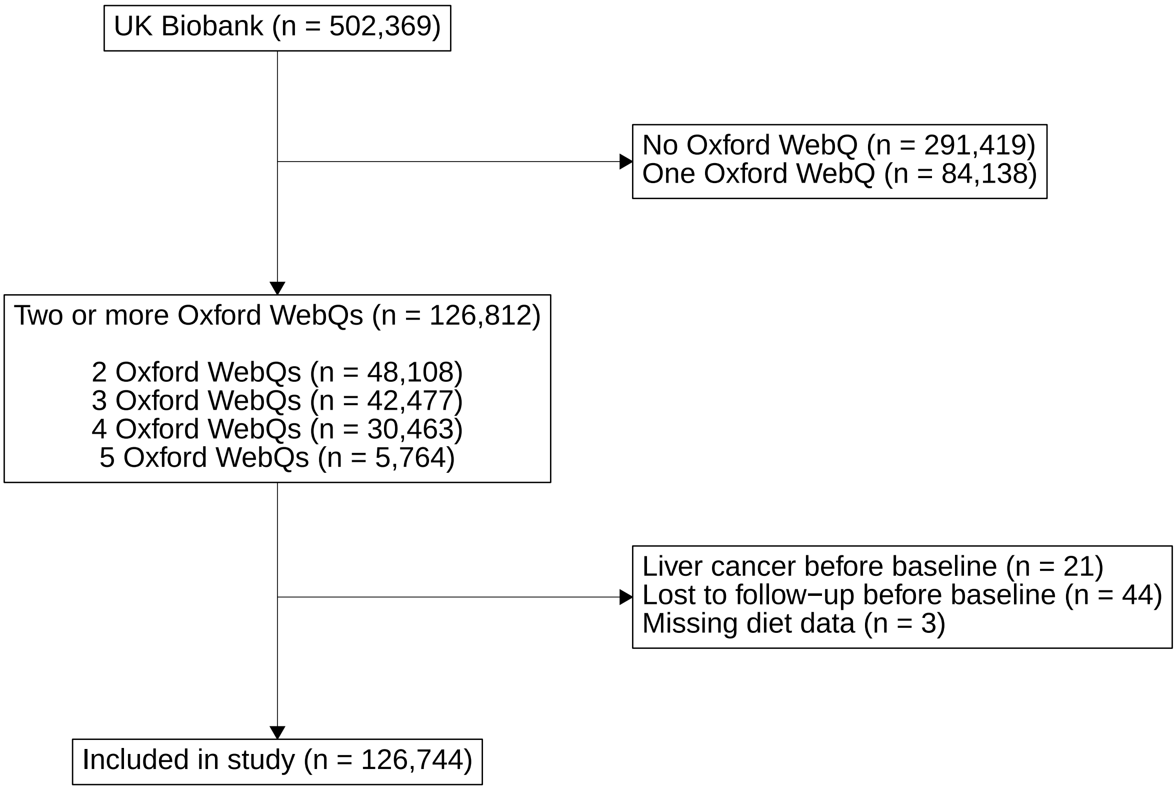

Fig. 1 Flowchart of included participants. Missing diet data were merged with loss to follow-up before baseline due to n being less than 5. It should be noted that not all UK Biobank participants were invited to complete an Oxfords WebQ. Only the last 70,000 participants to visit an assessment center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

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, 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 or unprocessed 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:

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.

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 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 in any statistically significant way. This also applied for including death registries as a source of liver cancer events and excluding waist circumference from the fully adjusted analysis (Supplementary table 4).

Discussion

In this study of UK Biobank participants, we found no effect of replacing 15g/day of red meat

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.

15 g/day of legumes replacing:	Model 1[1] HR (95% CI)	Model 2[2] 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 legumes on risk of primary liver cancer. The estimates did not change significantly in any of the sensitivity analyses. When stratifying by liver cancer type, replacing total red meat and unprocessed red meat with legumes showed an inverse association with ICC, though the results were not statistically significant.

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. These findings align with recent analyses from the UK Biobank, which indicated that 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) [17]. This supports the notion that processed meat may not be associated with liver cancer risk in this population.

Stratifying on cancer type revealed that replacing unprocessed red meat with legumes to some extent was associated with a decreased risk of ICC, contrary to HCC (Supplementary table 4). However, the robustness of these results were not tested in further analyses as they were associated with greater uncertainty.

The literature on food substitutions, particularly in relation to liver cancer, is sparse. However, 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 [18]. 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. A recent meta-analysis of observational studies showed no association between red or processed meat intake and HCC [19].

This study had some limitations. First, None of the registries used to determine a diagnosis of liver cancer were complete or up-to-date at the time of analysis [13]. 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 [14]. This could introduce misclassification bias, as individuals with liver cancer may not be identified as events. 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. Second, the relatively low number of events limited our ability to adjust for confounding factors. Too many adjustment levels 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.

Strengths of this study are the prospective longitudinal design, which establish temporality between the diet exposure and liver cancer outcome, and the large sampling size, which enhance generalizability. Though health registries may have been only partially up to date, using registries almost eliminates selection bias due to loss to follow-up. 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 estimate diet over time. Further, our specified substitution analysis have some strengths in contrast to traditional methods in nutritional epidemiology examining the effect of consuming a food or nutrient while holding all other foods constant. The substitution is easily interpretable and also reflecting 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.

Conclusion

Acknowledgements

[1] 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

[2] Biobank U (2011) Order of data collection. URL https://biobank.ndph.ox.ac.uk/ukb/ukb/docs/Orderofdatacollection.pdf, accessed 21 May 2024

[3] Biobank U (2024) 24-hour dietary recall questionnaire (oxford webq). URL https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/DietWebQ.pdf, accessed 21 May 2024

[4] 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

[5] 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

[6] 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

[7] 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

[8] 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

[9] 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

[10] 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

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

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

[13] Biobank U (2024) Health outcomes overview. URL https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/HealthOutcomesOverview.pdf, accessed 21 May 2024

[14] 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, accessed 21 May 2024

[15] 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

[16] Tomova G, Gilthorpe M, Tennant P (2022) Theory and performance of substitution models for estimating relative causal effects in nutritional epidemiology. The American Journal of Clinical Nutrition 116(5):1379–1388. https://doi.org/10.1093/ajcn/nqac188

[17] Knuppel A, Papier K, Fensom GK, et al (2020) Meat intake and cancer risk: prospective analyses in uk biobank. International Journal of Epidemiology 49(5):1540–1552. https://doi.org/10.1093/ije/dyaa142

[18] Liu K, Chen W, Zhou Y, et al (2023) Associations between food groups and liver cancer: a systematic review and meta-analysis of observational studies. Nutrition Journal 22(1). https://doi.org/10.1186/s12937-023-00858-5

[19] Di Y, Ding L, Gao L, et al (2023) Association of meat consumption with the risk of

gastrointestinal cancers: a systematic review and meta-analysis. BMC Cancer 23(1). https://doi.org/10.1186/s12885-023-11218-1