Protocol: Legume consumption and risk of primary liver cancer

Niels Bock Fie Langmann Daniel Borch Ibsen Luke W. Johnston
Christina Catherine Dahm

2023-10-04

Study information

Title

Statistical replacement of red meat with legumes in relation to incident primary liver cancer in the UK Biobank.

Description

Non-alcoholic fatty liver disease (NAFLD)-related primary liver cancer is on the rise worldwide. Preventive interventions are critical. NAFLD risk factors include sedentary lifestyles and "Western"-style diets high in fat and red meat. Legumes are now promoted in the Danish dietary guidelines as they have a low climate impact and are a good plant-based source of protein. They may additionally have anti-carcinogenic properties. Preclinical studies suggest that legumes are a promising intervention to lower risk of liver cancer. This project will use prospective data from the UK Biobank cohort as no Danish cohort consumes enough legumes to investigate. Based on a food substitution model, hazard ratios for liver cancer when substituting red meats with legumes will be estimated. It is critical that we reduce the climate impact of our diet. This study will add new knowledge that can qualify the health arguments behind the Danish dietary guideline to eat more legumes and less meat.

This research will use the UK Biobank Resource under Application Number 81520.

Objectives and hypotheses

Introduction

Food systems are estimated to cause 40% of man-made greenhouse gas emissions, and the foodstuff with the highest carbon footprint by far is red meat, particularly beef. The average adult Dane consumes ~900g/week red meat. Thus to reduce climate impacts of diet, the 2021 Danish dietary guidelines recommend eating less meat while increasing intake of legumes to ~100 g/day precooked legumes. Different species of legumes are traditional staple foods in many countries; in Denmark, however, consumption of legumes is very low. There may be great potential for downstream population health benefits of increasing legume consumption, particularly for diseases related to obesity, of which non-alcoholic fatty liver disease (NAFLD) and its sequalae primary liver cancer are of growing public health importance. Emerging evidence indicates that legumes may have anticancer effects. However, because studies on primary liver cancer to date have assessed consumption of legumes instead of any other food in relation to health outcomes, little is known of the associations at the expense of specifically red meat.

Objective

To investigate the association between statistical replacement of red meat with legumes and incidence of primary liver cancer (hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (IH-CCA)), and whether the association is mediated by NAFLD.

Hypotheses

1. Substitution of red meat (processed and unprocessed) with legumes is associated with a lower risk liver cancer.
2. The association between substitution of red meat (processed and unprocessed) with legumes and risk of liver cancer is mediated through NAFLD.
3. A higher intake of legumes is associated with a lower risk of liver cancer.

Design plan

Study type

Observational Study. Data is collected from study subjects that are not randomly assigned to a treatment. This includes surveys, natural experiments, and regression discontinuity designs.

Blinding

No blinding is involved in this study.

Study design

Because no Danish study populations consume enough legumes to investigate associations with disease, this study will use data from the UK Biobank (UKB), a prospective cohort with health information on approximately 500,000 participants, aged 40-69 at baseline, living in the UK [6]. Participants were recruited between 2006 and 2010. At baseline, participants completed touchscreen questionnaires at a designated Assessment Centre [6]. Information obtained from the touchscreen questionnaires included socioeconomic, lifestyle, and health related data, such as education level, smoking habits, physical activity, history of illness, etc. Physical, anthropometric and biomedical measurements done at the Assessment Centre included BMI, waist circumference and blood samples [6]. Previous history of diseases was collected from participants' self-reported disease status at baseline and through registration of ICD-10 codes in the National Health Services (NHS) registries.

Sampling plan

Existing data

Registration prior to analysis of the data. As of the date of submission, the data exist and you have accessed it, though no analysis has been conducted related to the research plan (including calculation of summary statistics). A common situation for this scenario when a large dataset exists that is used for many different studies over time, or when a data set is randomly split into a sample for exploratory analyses, and the other section of data is reserved for later confirmatory data analysis.

Explanation of existing data

The UK Biobank is a large national cohort of participants in the UK, with data collected in a standardized format the the sole purpose of providing a data resource for researchers to use for health research. All information about collection procedures, limitations, and sources of bias are well established and described by the UK Biobank resource.

Because of its size of data collected, it is near impossible to a priori see patterns in the data that might influence how the analysis will be conducted, unless specifically looked for or previously published on. In this way, we feel pre-analysis bias is minimal.

Data collection procedures

Exposure

The UKB was chosen as legume consumption is higher in the UK than in Denmark [7], approximately 48 grams/day [reference, Aurora]. Detailed dietary data were collected with a self-assessed, web-based 24-hour dietary recall assessment tool, Oxford WebQ. A questionnaire was added to the assessment visit at the end of recruitment. After recruitment, four additional rounds of online questionnaires were conducted online at 3-4 monthly intervals. Participants were asked to recall what they ate the previous day [8]. More than 200,000 UK Biobank participants completed

2 Oxford WebQ [9]. More than 80,000 participants completed 3 Oxford WebQ [10]. Participants who completed 2 Oxford WebQ's are included in this study. Dietary data on legumes were collected in 176,000 participants.

The exposures in this study will be red meat intake (total, processed and unprocessed) and legume intake. Legume and red meat intake will be based on total weight by food group intakes estimated from participants' responses in the Oxford WebQ. Participants who reported consuming vegetables were asked to report the intake of beans (not greens) and lentils they had the day before. Participants who reported consuming meat or poultry were asked to report the intake of different meats they had the day before.

Data from the UK Nutrient Databank Food Composition Tables were used to calculate total energy and nutrient intake by multiplying the number of portions consumed by the set quantity of each food portion size and its nutrient composition [7]. The average of the intake reported on WebQ 1 and 2 were used in this study.

Outcome

The primary outcome is liver cancer defined as either HCC or IH-CCA. Cancer diagnoses will be assessed through linkage to NHS Digital and Public Health England, and the NHS Central Register. Incident primary liver cancer diagnoses during follow-up are determined through ICD-10 codes C22.0 (Liver cell carcinoma) and C22.1 (Intrahepatic bile duct carcinoma) in the National Health Services (NHS) registries.

Covariates

Covariables are defined a priori from studies assessing diet components as exposure and liver cancer as the outcome [11] and illustrated using directed acyclic graphs (fig. 1.). Information on covariables will include all other dietary components ([g/day] or [kcal/day]), age (years), sex (male, female), socioeconomic status (educational level, Townsend Deprivation Score [quintiles], living with a spouse (yes, no), and geographical location), ethnicity (white, mixed, Asian, black, other, unknown), anthropometry (BMI [kg/m2], waist circumference [cm]), alcohol consumption, physical activity (low [0-9.9 METs/week], moderate [10-49.9 METs/week], and high [50 METs/week], unknown), smoking status (never, former, current 1-15 cigarettes per day, current 15 cigarettes per day, current but number of cigarettes per day unknown, and smoking status unknown), self-reported own or family members' diagnoses of diabetes [yes, no, unknown], and history of other liver diseases (cholelithiasis, cholecystectomy). All covariables will be baseline measures except dietary components.

Sample size

Sample size rationale

Variables

Measures variables

Covariables are defined a priori from studies assessing diet components as exposure and liver cancer as the outcome [11] and illustrated using directed acyclic graphs (fig. 1.). Information on covariables will include all other dietary components ([g/day] or [kcal/day]), age (years), sex (male, female), socioeconomic status (educational level, Townsend Deprivation Score [quintiles], living with a spouse (yes, no), and geographical location), ethnicity (white, mixed, Asian, black, other, unknown), anthropometry (BMI [kg/m2], waist circumference [cm]), alcohol consumption, physical activity (low [0-9.9 METs/week], moderate [10-49.9 METs/week], and high [50 METs/week], unknown), smoking status (never, former, current 1-15 cigarettes per day, current 15 cigarettes per day, current but number of cigarettes per day unknown, and smoking status unknown), self-reported own or family members' diagnoses of diabetes [yes, no, unknown], and history of other liver diseases (cholelithiasis, cholecystectomy). All covariables will be baseline measures except dietary components.

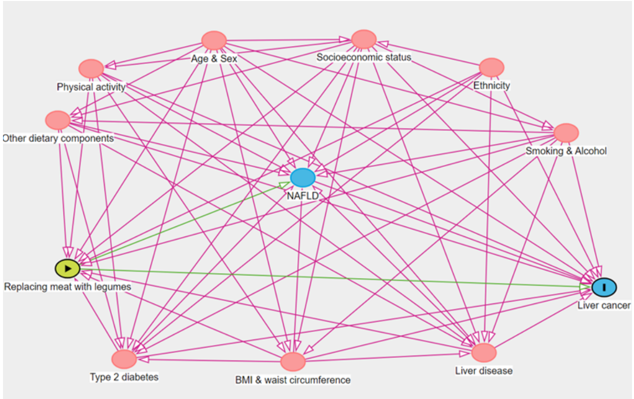

Figure 1: Figure 1. Directed acyclic graph visualizing the relationship between the exposure, replacing meat with legumes, and the outcome, liver cancer. Red circles indicate ancestors of exposure and outcome with red arrows being biasing paths. The NAFLD circle is blue to indicate that it is an ancestor of the outcome. The green arrows are causal paths.

Analysis plan

Statistical models

Multivariable-adjusted Cox proportional hazards regression model will be used to estimate the hazard ratio (HR) with corresponding 95% confidence intervals for liver cancer when substituting red meats with legumes. Substitution will be done in two ways [12]:

1. Equal-mass substitution, i.e., substituting x grams of red meats with x grams of legumes.
2. Equal-calorie substitution, i.e., substitution x calories from red meats with x calories from legumes.

Estimates will be presented for total red meat as well as stratified estimates of processed and unprocessed red meat. Age will be used as the underlying timescale. Follow-up time will start at participants' last completed Oxford WebQ. Participants will be right censured due to one of the following events, whichever comes first: most recent registry date for the full follow-up for the outcomes, date of death, date of liver cancer diagnosis, or loss to follow-up.

Different adjustment levels will be added to the substitution models. Model 1 will be minimally adjusted for age at recruitment, sex, total caloric intake, and other dietary components other than the substituted foods. When substituting grams of legumes, the unit for all dietary components will be grams/day. When substituting calories of legumes, the unit for all dietary components will be calories/day. Model 2 will be further adjusted for educational level, Townsend Deprivation Score, ethnicity, BMI, physical activity, smoking status, alcohol intake, waist circumference, type 2 diabetes, and liver diseases (cholelithiasis and cholecystectomy).

Secondary analyses

Causal mediation analysis will be used for hypothesis 2 and applied to all three levels of adjustment as in the main analyses, so that potential mediator-outcome confounding can be taken into account [13]. To evaluate the pooling of the outcome data, a stratified analysis of the liver cancer subtypes, HCC and IH-CCA, will be conducted. The association of legume intake and liver cancer risk will be estimated using quintiles of legume intake (g/day) and modelled as in model 2.

All analyses will be conducted in R with a significance level of 5%.

Transformations

Inference criteria

Data exclusion

Missing data

Exploratory analyses (optional)

Other analyses

To test for effect modification, stratified analyses will be conducted based on sex (male, female), waist circumference (above or below 94 cm for men, 80 cm for women), type 2 diabetes (yes, no), and other liver diseases (yes, no).

Sensitivity analyses will be conducted to test the robustness of the main analyses. These will include changes in exposure in the substitution model, e.g. 50 grams instead 30 grams of legumes for red meat, or substitution of 10 grams/day per 1000 kcal of legume protein with protein from red meat. Further sensitivity analyses will include imputations of missing information, inclusion of varying numbers of completed diet questionnaires, and removal of individuals highly predisposed to liver cancer, e.g. primary biliary cirrhosis, alcoholic liver disease, and viral hepatitis. All sensitivity analyses will be modelled as in the main analysis. All analyses will be conducted in R with a significance level of 5%.

References

1. Intergovernmental Panel on Climate Change, *Climate Change 2022: Mitigation of Climate Change*. 2022.

2. Poore, J. and T. Nemecek, *Reducing food's environmental impacts through producers and consumers.* Science, 2018. **360**(6392): p. 987-992.

3. Pedersen, A.m.f.N., *Danskernes kostvaner 2011-2013 : hovedresultater*. 1. udgave ed. 2015, Søborg: DTU Fødevareinstituttet, Afdeling for Ernæring. 208 sider.

4. UN. *Food and Climate Change: Healthy diets for a healthier planet*. Available from: https://www.un.org/en/climatechange/science/climate-issues/food.

5. Zhang, S., et al., *Protein foods from animal sources and risk of nonalcoholic fatty liver disease in representative cohorts from North and South China.* J Intern Med, 2022.

6. Biobank, U. *Learn more about UK Biobank*. 2022 27/07-2022 13/12-2022]; Available from: https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank.

7. Piernas, C., 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.* Eur J Nutr, 2021. **60**(5): p. 2879-2890.

8. Liu, B., 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.* Public Health Nutr, 2011. **14**(11): p. 1998-2005.

9. Perez-Cornago, A., et al., *Description of the updated nutrition calculation of the Oxford WebQ questionnaire and comparison with the previous version among 207,144 participants in UK Biobank.* Eur J Nutr, 2021. **60**(7): p. 4019-4030.

10. Sudlow, C., 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 Med, 2015. **12**(3): p. e1001779.

11. Liu, K., et al., *Associations between food groups and liver cancer: a systematic review and meta-analysis of observational studies.* Nutr J, 2023. **22**(1): p. 30.

12. Ibsen, D.B., et al., *Food substitution models for nutritional epidemiology.* Am J Clin Nutr, 2021. **113**(2): p. 294-303.

13. Nguyen, T.Q., Schmid, I., & Stuart, E. A.„ *Clarifying causal mediation analysis for the applied researcher: Defining effects based on what we want to learn.* Psychological Methods, 26(2), 2021: p. 255-271.