Niels Bock

2024-03-22

# Abstract

Objective:

Research Design and Methods:

Results:

Conclusions:

# Introduction

# Research Design and Methods

**Study population**

The UK Biobank, a population-based prospective cohort, were initiated in 2006 [1]. During 2006-2010, more than 500,000 participants, aged 40-69, were recruited and assessed at designated assessment centres across the UK. Data on sociodemographic factors (education, ethnicity, Townsend deprivation Index) and lifestyle factors (smoking, alcohol consumption, physical activity) were collected via touch screen questionnaires and computer-assisted interviews. Anthropometric data (BMI, waist circumference) were collected via physical measurements (reference to UKB document here).

**Dietary assessment**

A web-based diet questionnaire was administered at the end of the initial assessment visit for the last 70,000 recruited participants (reference from UKB document here). 320,000 participants who had provided an e-mail address were invited to complete the diet questionnaire online. In the annual period February 2011 and April 2012, participants were invited on 4 separate occasions to complete the diet questionnaire online. The questionnaire comprised the Oxford WebQ, an online 24-hour dietary recall assessment tool covering 206 foods and 32 beverages commonly consumed in the UK [2]. The Oxford WebQ has been validated with interviewer-based 24-hour recalls and biomarkers [3, 4]. Intake categories ranged from 0 to +3 units of measurement (e.g. servings, cups, slices).

79 food groups and 14 beverage groups were classified using UK National Diet and Nutrition Survey (NDNS) categories [2, 3]. We used this classification to categorise intake of foods groups to match our substitution analysis. Legume intake was defined as intake of dietary pulses, baked beans, tofu-based products, peas, hummus, soya drinks, and soya-based desserts and yoghurt; red meat intake was defined as intake of beef, pork, lamb, or other meat, including offal. Processed meat intake was defined as sausages, bacon (with and without fat), ham, or liver pate. Other food groups included were poultry, fish, dairy, egg and egg dishes, whole grains, refined grains, vegetables, fruits, nuts and seeds, fats and spreads, mixed dishes, sweets and snacks, non-alcoholic beverages, and alcoholic beverages (supplemental table 1).

Due to the incapability of a single 24-hour dietary recall evaluation to properly assess variation in diet over time (reference to nutritional epi methods here?), only participants who completed two or more questionnaires were eligible for inclusion (reference to nutritional epi methods that two 24-hour recall is sufficient to capture variation in diet here?).

**Liver cancer assessment**

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

**Assessment of confounders**

Confounders were defined *a priori* from a literature review of diet components as exposure and liver cancer as the outcome and illustrated using directed acyclic graphs (supplemental fig. 1.). The following confounding variables were identified: age (years [continuous]), sex (male, female [categorical]), educational level (high (College or University degree), intermediate (A levels/AS levels, O levels/GCSEs, or equivalent), low (none of the previous mentioned) [categorical]), Townsend Deprivation Index ([continuous]), Living alone (yes, no [categorical]), ethnicity (white, other [categorical]), body mass index ( <25 kg/m2, 25-30 kg/m2, >30 kb/m2 [categorical]), waist circumference (cm [continuous]), physical activity (above/below daily MET recommendation for moderate physical activity [categorical], smoking status (never, former, current [categorical]), alcohol intake (grams/day [continuous]), history of diabetes (yes, no [categorical], Non-alcoholic liver disease (NAFLD), (yes, no [categorical], history of cholelithiasis (yes, no [categorical]), and history of cholecystectomy (yes, no [categorical]). Sociodemographic and anthropometric confounders were selected from the initial assessment visit before the start of follow-up. Diagnosis and corresponding date of diagnosis of diabetes, NAFLD, and cholelithiasis before start of follow-up were selected via linkage to ICD-10 and ICD-9 inpatient hospital episodes. For definitions of diagnoses, see supplemental table 2.

**The substitution model**

Substitutions were carried out in an equal-mass manner, i.e., substituting x grams of red meats with x grams of legumes. The size of the substitution was set to 15 grams of legumes for 15 grams of red meats to keep the substitution size below the mean intake any of the substituted food groups in the cohort. Substitutions were modelled using the leave-one-out-approach in which variables for every food group along with a variable for total food intake are included, except the food group that are to be substituted [5]. To estimate substitution of 15 grams of all red meats with 15 grams of legumes, the following model was defined:

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

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

**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 diet questionnaire until the event of interest occurred or due to right censoring, whichever came first. Participants were right censored due to one of the following events occurring: death, loss to follow-up, or administrative end of follow-up (set to Jan 31, 2022). Two levels of adjustments were added to the substitution model. Model 1 were minimally adjusted for age, sex, total weight of food intake, and all other foods groups. Model 2 were further adjusted for educational level, Townsend Deprivation Index, ethnicity, living alone, smoking status, alcohol intake, BMI, waist circumference, diabetes, NAFLD, cholelithiasis, and cholecystectomy.

A stratified analysis on each cancer type was performed to test whether pooling of HCC and ICC as an outcome was justified. Further, the following sensitivity analyses were performed to test the robustness of the main analysis:

1.       Due to the events per variable being very low in the fully adjusted analysis and thus increasing the risk of biased estimates in the cox regression model, a simpler model was defined. To reduce number of variables, food groups were polled in fewer groups and only confounders hypothesized to have the greatest impact on the estimate were included in this model.

2.       Exclusion of high alcohol consumers.

3.       Exclusion of food intake misreporters.

4.       Exclusion of participants with any other liver disease before baseline.

5.       Exclusion of participants with any form of cancer before baseline.

6.       Exclusion of the first two years of follow-up.

7.       inclusion eligibility criteria for number of completed diet questionnaires were set to ≥3, ≥4, and ≥ 5 completed questionnaires.

8.       Imputation of missing data.

All analyses were conducted in R with a significance level of 5 %

# Results

# Conclusions

# Acknowledgements

# References

# Tables

[Table 1. Main analysis](3-table-main-analysis.html)

# Figures

# Supplemental Material