It is made available under a CC-BY 4.0 International license .

- 1 Legumes as a substitute for red and processed meat, poultry,
- ² or fish, and the risk of non-alcoholic fatty liver disease in a large
- 3 cohort
- 4 Fie Langmann¹, Daniel B. Ibsen^{1,2}, Luke W. Johnston², Aurora Perez-Cornago³, Christina C.
- 5 $Dahm^1$
- 6 1. Aarhus University, Department of Public Health, Bartholins Allé 2, 8000 Aarhus C,
- 7 Denmark.
- 8 2. Steno Diabetes Center Aarhus, Aarhus University Hospital, Palle Juul-Jensens Boulevard
- 9 11, 8200 Aarhus C, Denmark
- 10 3. European Commission, Joint Research Centre (JRC), Ispra, Italy.

11

- 12 Author contributions: FL, DBI, APC, and CCD conceived the research question and designed
- 13 the analysis plan. FL, DBI, and LWJ conducted the data analysis. FL, DBI, and CCD drafted
- 14 the manuscript. All authors interpreted the results and critically revised the article for
- 15 important intellectual content and gave final approval of the version to publish.
- 16 Corresponding author:
- 17 Christina C. Dahm. Department of Public Health, Aarhus University, Bartholins Allé 2,
- 18 Aarhus-8000, Denmark. E-mail: <u>ccd@ph.au.dk</u>, phone: +45 2332 1875

It is made available under a CC-BY 4.0 International license .

19 Abstract

- 20 **Background**: Dietary recommendations have globally shifted towards promoting
- 21 consumption of legumes as an environmentally friendly and healthy source of protein. This
- study investigated replacement of red and processed meat, poultry, or fish for equal amounts
- 23 of legumes on the risk of non-alcoholic fatty liver disease (NAFLD).
- 24 Methods: UK Biobank participants who completed ≥ 2 dietary assessments and had complete
- covariate information were included in the analyses (N=124,194). Information on dietary
- 26 intake was collected using two to five 24-hour dietary assessments. Incident cases of NAFLD
- 27 were determined through linkage to the National Health Service registries. The rate of
- developing NAFLD when replacing 80 g/week of red and processed meat, poultry, or fish
- 29 with legumes was estimated using multivariable-adjusted Cox proportional hazards
- 30 regression.
- **Results**: During follow-up (median 10.49, IQR: 10.4-10.9 years), 1201 individuals developed
- 32 NAFLD. Replacing 80 g/week red and processed meat or poultry with legumes was
- associated with 4% and 3% lower rates of NAFLD, respectively (meat HR: 0.96, 95% CI:
- 34 0.94; 0.98; poultry HR: 0.97, 95% CI: 0.95; 0.99). Replacing 80 g/week of fish with legumes
- 35 was not associated with NAFLD (fish HR: 0.98, 95% CI: 0.96; 1.01). Results did not change
- 36 markedly after adjustment for BMI.
- 37 **Conclusions**: Consuming one serving of legumes weekly instead of red and processed meat
- or poultry was associated with a slightly lower rate of NAFLD, while consuming legumes
- 39 instead of fish did not show an association with NAFLD. Further research in cohorts with
- 40 higher legume consumption is needed to confirm these findings.

It is made available under a CC-BY 4.0 International license .

41 Highlights

- 42 Food substitution models improve interpretation of studies of dietary exposures in
- 43 observational studies.

44

- 45 Replacing red and processed meat or poultry with legumes was associated with slightly lower
- 46 rates of non-alcoholic fatty liver disease in the UK Biobank. No association was found when
- 47 replacing fish with legumes.

48

- 49 Keywords: dietary pulses, legumes, non-alcoholic fatty liver disease, NAFLD, nutritional
- 50 epidemiology, dietary substitution model

It is made available under a CC-BY 4.0 International license .

51 Introduction

Non-alcoholic fatty liver disease (NAFLD), defined as liver fat content of more than 5% 52 53 without secondary causes of hepatic fat accumulation, is the most prevalent chronic liver disease with a global prevalence of 25%.⁽¹⁾ NAFLD is associated with an increased risk of 54 cardiovascular diseases⁽¹⁾ and can furthermore progress to non-alcoholic steatohepatitis 55 (NASH), which may further predispose individuals to develop fatal liver complications.^(2; 3) 56 Long-term exposure to Western diets, obesity, physical inactivity, and smoking are all risk 57 factors for NAFLD.⁽⁴⁾ Concurrently, NAFLD and NASH increase the levels of liver enzymes 58 like alanine aminotransferase (ALT), which have been associated with increased incidence of 59 all-cause and cancer mortality.^(5; 6) 60 61 Food production and consumption contribute up to 42% of global CO_2 emissions.^(7; 8) To

reduce climate impacts of human diets, sustainable dietary guidelines have been introduced 62 emphasising greater consumption of plant-based protein sources such as legumes.^(9, 10) The 63 climate impact of legumes is undoubtedly low,⁽⁷⁾ however the scientific evidence on the 64 65 association between legume intake in isolation and health outcomes is sparse.⁽¹¹⁾ Legumes are 66 good sources of protein, low in saturated fat and energy density, and rich in dietary fibre; all components associated with a healthy diet.⁽¹²⁾ Consumption of diets including legumes has 67 been associated with overall better diet quality and greater health compared to consuming 68 Western diets high in red and processed meat, fats, and sugars in observational studies.⁽¹³⁻¹⁵⁾ 69 70 Animal studies have shown that legume consumption up-regulates lipid metabolism in liver cells, thus minimizing the risk of NAFLD by reducing the build-up of fats in the liver.^(16; 17) 71 72 However, research in humans on the association between legume consumption and NAFLD is sparse and inconclusive.^(11; 18; 19) 73 74 When individuals limit their intake of certain food groups, they will increase the intake of

certain other food groups, in an otherwise stable diet.⁽²⁰⁾ Legumes are often included in the 75 76 diet as a source of plant-based protein, and the health benefits of substituting legumes for other protein-rich foods may depend on the substitute as some sources of protein may be 77 better to replace than others.⁽¹⁹⁾ Replacing protein from animal sources with protein from 78 plant sources has previously been associated with a substantially lower mortality rate in some 79 studies.⁽²¹⁾ Furthermore, even low consumption of red and processed meats has been 80 81 associated with a greater risk of developing NAFLD, while consumption of seafood or poultry has shown mixed directions of associations.⁽¹⁻³⁾ Therefore, this study aimed to 82 investigate the association between replacing red and processed meat, poultry, or fish with 83

It is made available under a CC-BY 4.0 International license .

84 legumes and the risk of NAFLD or NASH contingent on potential confounders of the

85 associations.

86

87 Methods

88 Study population and setting

89 The UK Biobank prospective cohort was established in 2006 and invited 9.2 million people to participate in the study.⁽²²⁾ The participation rate of 5.5% resulted in approximately 90 500,000 participants aged 37-73 years in the study. At baseline, participants provided detailed 91 92 information on sociodemographic, physical, lifestyle, diet, and health-related characteristics 93 via self-completed touch-screen questionnaires and a computer-assisted personal interview.⁽²²⁾ Professionally trained staff took physical, anthropometric, and biomedical 94 measures following standardized procedures.⁽²²⁾ Participants' diets were assessed with up to 95 five 24-hour dietary assessments, using the Oxford WebQ,⁽²³⁾ which was completed at least 96

97 once by 210,950 individuals.

For the current study, only participants who completed two or more 24-hour dietary

99 assessments were included in the analysis to capture usual intake of legumes. Participants

100 with missing information on covariates were excluded apart from missing information on

101 physical activity. If participants were missing information in all these categories, they were

102 excluded from the study; if any information was available, the participant was included and

103 individual missing data points on physical activity were coded as unknown.

104 All participants gave written informed consent to participate in the study. The UK

105 Biobank was approved by the National Information Governance Board for Health and Social

106 Care and the National Health Service (NHS) North West Multicentre Research Ethics

107 Committee (ref 21/NW/0157). This research has been conducted using the UK Biobank

108 Resource under Application Number 81520.

109

110 Assessment of diet

111 Individuals recruited in the final recruitment period, between April 2009 and September

112 2010, completed their first Oxford WebQ at the assessment centre at baseline (N = 70,747).

113 Any participant in the UK Biobank cohort who provided a valid email address was invited to

114 complete the Oxford WebQ online on four separate occasions during follow-up.⁽²³⁾

115 The Oxford WebQ was designed as an internet-based 24-hour dietary assessment tool. The

116 questionnaire comprises a short set of food frequency questions on commonly eaten food

It is made available under a CC-BY 4.0 International license .

groups in the British population assessing diet in the previous 24 hours.⁽²³⁾ The Oxford

117

118	WebQ was validated against interviewer-administered 24-hour dietary recalls and recovery
119	biomarkers. ⁽²⁴⁾ The validation showed similar performance for estimating true intakes of
120	protein, potassium, and total sugars when using two or more dietary assessments compared to
121	the more burdensome interviewer-based recall. ⁽²⁴⁾ Recently, the Oxford WebQ nutrient
122	calculation was updated to provide more detailed information on nutrient intakes and to
123	incorporate new dietary variables. ⁽²⁵⁾
124	Consumption of legumes, red and processed meats, poultry, and fish was based on the
125	total weight of food group intakes estimated as the average from participants' responses in
126	the Oxford WebQs. ⁽²⁶⁾ The food groups used in this study are defined in Supplementary
127	Table 1.
128	
129	Non-alcoholic fatty liver
130	The International Classification of Diseases and Related Health Problems 10 th edition disease
131	codes (ICD-10) were used to define disease outcomes. ⁽²⁷⁾ The entity metabolic dysfunction-
132	associated steatotic liver disease (MASLD) was defined in 2023 as the new nomenclature for
133	NAFLD and NASH. However, as MASLD is not part of the ICD-10 th edition upon which the
134	outcomes are defined and diagnosed, we used NAFLD and NASH throughout this
135	manuscript. ⁽²⁸⁾ Incident cases of NASH and NAFLD were assessed through linkage to the
136	NHS registers, where diagnosis after hospital admissions were coded according to the ICD-
137	10. ⁽²⁷⁾ Incident cases of NAFLD were diagnosed with ICD-10-code K76.0 at first admission
138	to the hospital while incident cases of NASH were diagnosed with ICD-10-code K75.8. ⁽²⁹⁾
139	
140	Covariates
141	Covariates were chosen a priori based on a review of the literature and directed acyclic
142	graphs (Supplementary Figure 1).
143	Information on covariates included information on average intakes of all other food groups
144	retrieved from the Oxford WebQ (refined cereal, whole grain cereal, dairy products, dietary
145	fats, fruits, vegetables, nuts, mixed dishes, potatoes and tubers, eggs, non-alcoholic
146	beverages, alcoholic beverages, snacks, and sauces and condiments). Other covariates were
147	self-reported at the baseline visit and included information on sex, age, ethnicity, yearly
148	income, educational level, Townsend Deprivation Index (0 indicates that an individual lives
149	in an area with overall mean deprivation, positive values indicate higher material deprivation

while negative values indicate relative affluence),⁽³⁰⁾ geographical region of recruitment,⁽²³⁾

It is made available under a CC-BY 4.0 International license .

151 cohabitation, anthropometry, physical activity, smoking status, history of NAFLD-related

diseases, and family history of NAFLD-related diseases. Usual alcohol consumption (g

ethanol/day) was estimated as the average of the responses to the Oxford WebQs. Blood

- samples were drawn at baseline for all participants to measure ALT. When elevated, ALT
- serves as a proxy measure for increased risk of developing NAFLD.^(5; 6)
- 156

157 Statistical analyses

- 158 Standard summary statistics were performed to describe the distribution of participants'
- baseline characteristics and food group consumption across strata of legume consumption.
- 160 The legume consumption strata were defined as non-consumers (0 g/week), low intake
- 161 (tertile 1 of those who consumed any legumes), medium intake (tertile 2 of those who

162 consumed any legumes), and high intake (tertile 3 of those who consumed any legumes).

163 Summary statistics were also computed to describe baseline characteristics and food group

- 164 consumption across incident NAFLD.
- 165 Multivariable adjusted Cox proportional hazards regression models were used to estimate

the hazard ratio for NAFLD or NASH based on substituting equal masses of red and

- 167 processed meat, poultry, or fish with legumes. The NHS defines one serving of legumes and
- 168 pulses as three heaped tablespoons or 80 $g^{(31)}$ and the substitution magnitude was chosen to
- 169 reflect a weekly substitution where one serving of legumes or pulses (80 g) replaced an equal
- amount of either red and processed meat, poultry, or fish. Substitutions were modelled using
- 171 the leave-one-out approach, following this model:

 $log(h(t;x)) = log(h(0;t)) + \beta_1 Legumes(80g/week) + \beta_2 Total food intake (g/week)$ $+ \beta_3' Other food groups (g/week) + \beta_4' Covariates$

172 Apostrophes indicate a group of coefficients consisting of several individual variables (β 3' 173 and β 4'). Variables for intakes of each food group (β 3') and total food intake (β 2) were 174 included and held stable, while the food group that were to be substituted was left out of the model.⁽²⁰⁾ For instance, when estimating the HR for replacing red and processed meat with 175 176 legumes, 80g of legumes was specified in the model, while the intake of red and processed 177 meat was left out of β 3'. Red and processed meat still contributed to the total food intake 178 $(\beta 2)$. The estimated HR thus expressed the association of keeping the total food intake stable 179 while specifying that 80 g/week food should come from legumes instead of red and processed 180 meat.

Age was used as the underlying time scale in the analyses. Person-years at risk were calculated from the date of last completed Oxford WebQ to the date of death, loss to follow-

It is made available under a CC-BY 4.0 International license .

183 up, diagnosis of NAFLD or NASH, or right censoring, whichever occurred first. As 184 participants in UK Biobank are still followed today, participants were right censored on October 31st, 2022, when the most recent registry update of full follow-up for the outcomes 185 186 was completed. The proportional hazards assumption was evaluated using time-varying 187 covariates and Schoenfeld residuals and revealed no evidence of deviation. 188 The substitution analyses were conducted with different adjustment levels. Model 1 was 189 stratified by age at recruitment ($<45, 45-49, 50-54, 55-59, 60-64, \ge 65$ years), geographical region of recruitment (ten UK regions),⁽²³⁾ and sex, and adjusted for intakes of all other food 190 191 groups (g/week) apart from the food being replaced. When substituting g/week of legumes 192 for other foods, the standard unit for all dietary elements was also in g/week. This study modelled substitution of foods with varying energy densities.^(32; 33) (32; 33) Nevertheless, to 193 avoid estimating obscure relative effect estimands with unclear interpretations due to 194 different units of measurement in the model, total energy intake was not adjusted for.⁽³³⁾ The 195 analyses accounted for all food items based on their weight, and the analyses were adjusted 196 for total amount of food consumed in g/week.⁽²⁰⁾ Model 2 was further adjusted for alcohol 197 198 consumption (continuous; g ethanol/week as restricted cubic splines with 4 knots), ethnicity 199 (categorical; White, other), educational level (categorical; low: Certificate of Secondary 200 Education (CSE), National Vocational Qualifications, Higher National Diploma, Higher 201 National Certificates, other professional qualifications, or equivalent; intermediate: A levels, 202 O levels, General Certificate of Secondary Education, or equivalent; high: College or 203 University degree), yearly income (categorical; <18,000£, 18,000-30,999£, 31,000-51,999£, 204 52,000-100,000£, >100,000£, prefer not to answer), Townsend Deprivation Index 205 (continuous), cohabitation (categorical; alone, with spouse or partner, with non-partner, 206 prefer not to answer), physical activity (categorical; low [0-9.9 METs/week], moderate [10-207 49.9 METs/week], and high \geq 50 METs/week], unknown), smoking status (never, former, 208 current 1-15 cigarettes per day, current \geq 15 cigarettes per day, unknown), history of NAFLD-209 related diseases (categorical; yes, no or don't know), and family history of NAFLD-related 210 diseases (categorical; yes, no or don't know). As obesity may either confound or mediate the 211 association between replacing red and processed meat, poultry, or fish with legumes and the risk of NAFLD, model 3 was further adjusted for BMI (categorical; $<30 \text{ kg/m}^2$ or ≥ 30 212 kg/m^2). 213

214

215 Secondary and sensitivity analyses

It is made available under a CC-BY 4.0 International license .

216 Secondary analyses included Cox proportional hazards regression analyses modelled as the 217 main analyses in model 2, but in a study sample restricted to consumers of legumes (i.e. 218 removing non legume consumers). We also modelled legume consumption as a continuous 219 exposure to evaluate the non-substitution association between legume consumption and 220 NAFLD with an 80 g/week increase in legume consumption without changing other elements 221 of the diet. This analysis was also conducted among consumers of legumes only and followed 222 the adjustment levels of model 2 but without excluding any single food components from β 3' 223 and without adjusting for total intake of all foods $\beta 2$. 224 To evaluate the robustness of the main analyses, multiple sensitivity analyses were 225 conducted: i) Inclusion of fresh peas in the estimated weekly legume consumption, as fresh 226 peas are a legume pod though traditionally counted as a vegetable in NHS' 5 A Day recommendations for fruit and vegetables.⁽³¹⁾ The amount of peas consumed was estimated 227 228 based on the number of reported portions consumed with one portion of peas weighing 80 g/day.⁽³¹⁾ ii) Exclusion of soy milk from the estimated weekly legume consumption, as soy 229 230 milk is unlikely to culinarily replace red and processed meat, poultry, or fish. iii) Excluding participants above the 90th percentile of alcohol intake to reduce the likelihood that cases of 231 232 fatty liver disease were caused by alcohol intake. iv) Excluding participants with elevated 233 ALT levels cut-off at 40 U/L as defined by guidelines from the European Association for the Study of the Liver^(34; 35). v) Excluding participants with < 3 completed Oxford WebQs. 234 235 All secondary and sensitivity analyses followed the adjustment level in model 2 from the 236 main analysis (i.e. all covariates were included except BMI). 237 All analyses were conducted in R (Version 4.1.1, The R Foundation for Statistical 238 Computing) with a two-sided significance level of 5%. The statistical analyses followed a 239 published protocol.⁽³⁶⁾ The statistical procedures only had minor deviations from the protocol. 240 Of note, the non-specific substitution was omitted and replaced by the non-substitution 241 analyses among consumers of legumes. An additional sensitivity analysis excluding soymilk from the legume component was conducted, as replacing animal-based solid foods for a 242 243 beverage may not be feasible. The statistical analyses were structured to be reproducible using the targets R package⁽³⁷⁾ and all code is available online at https://github.com/steno-244 245 aarhus/leha/ (Accessed on October 21, 2024). 246

247 **Results**

It is made available under a CC-BY 4.0 International license .

Of the 502,369 participants in the UK Biobank cohort, 126,812 individuals had completed two or more Oxford WebQs. Of these, 2381 were excluded due to missing information on covariates, 192 had an incident event before the start of follow-up, and 45 were lost to follow-up before baseline. This resulted in a study sample of 124,194 participants (54,921 men, 69,273 women) in the analyses (Figure 1). Participants contributed a total of 1,288,233 person-years of follow-up and 1201 individuals developed NAFLD after a median of 10.49 person-years (interquartile range: 10.42-10.92 years).

255 A large proportion of participants did not report any legume consumption (N = 73,711). The median (10-90th percentile) age at inclusion was 57.0 (44.0-66.0) years and 56% of 256 257 participants were women. Individuals with higher legume consumption more often had 258 another ethnicity than White and consumed less alcohol compared to the full cohort. The 259 average weekly intakes of the different food groups were unevenly distributed among those 260 who consumed legumes (across all tertiles) compared to non-consumers. Particularly the 261 consumption of animal-based foods like red and processed meat, poultry, and fish was lower 262 among those with the highest intake of legumes, compared to the full cohort and to those with 263 lower legume consumption (Table 1).

264 When assessing baseline characteristics and food group intakes in relation to the incidence 265 of NAFLD, those who developed NAFLD had a lower deprivation compared to the full 266 cohort. Individuals who developed NAFLD more often smoked and had higher BMI 267 compared to the full cohort. The median daily alcohol intake was, however, lower among 268 those who developed NAFLD. History of NAFLD-related diseases (individual and family-269 related) was more pronounced among individuals who developed NAFLD during follow-up 270 (Supplementary Table 2). 271 Replacing 80 g/week of red and processed meat with legumes was associated with a lower

rate of developing NAFLD (HR for model 2: 0.96, 95% CI: 0.94; 0.98; Table 2). Adjusting
for BMI (Model 3) changed the magnitude of association slightly (HR: 0.97, 95% CI: 0.95;
0.99). Substituting poultry for legumes was associated with a lower rate of NAFLD (HR for
model 2: 0.97, 95% CI: 0.95; 0.99; Table 2). However, this association was not statistically

significant after adjusting for BMI. Substituting fish for legumes was not associated with

277 rates of NAFLD (Table 2).

278

279 Secondary and sensitivity analyses

280 When excluding non-consumers of legumes and reducing the sample size to 50,483

281 participants (n events = 465), the magnitude of associations changed minimally compared to

10

It is made available under a CC-BY 4.0 International license .

the main analyses, in which all participants were included (Table 2; red meat HR: 0.97, 95%)

283 CI: 0.94; 0.99; poultry HR: 0.98, 95% CI: 0.95; 1.02; fish HR: 0.99, 95% CI: 0.95; 1.02).

284 Consuming 80 g/week more legumes without substituting any other foods was not associated

with the rate of NAFLD among consumers (HR: 0.99, 95% CI: 0.98; 1.01).

- 286 When including peas in the estimated legume consumption, the magnitude and direction of
- associations remained similar as in the main analyses, albeit with wider CIs (Table 3; meat
- 288 HR: 0.95, 95% CI: 0.92; 0.99; poultry HR: 0.97, 95% CI: 0.93; 1.01; fish HR: 0.98, 95% CI:
- 289 0.94; 1.02). Excluding soy milk from the estimated legume consumption also resulted in
- similar rates of NAFLD as in the main analyses (meat HR: 0.96, 95% CI: 0.93; 0.98; poultry
- 291 HR: 0.97, 95% CI: 0.95; 1.00; fish HR: 0.98, 95% CI: 0.96; 1.01).
- When removing participants with very high alcohol intake or very high ALT circulating concentrations the magnitude and direction of associations remained almost identical to those estimated in main analyses. Exclusion of participants with fewer than three returned Oxford WebQs resulted in marginally widened CIs (Table 3).
- 296

297 **Discussion**

298 In this prospective study of British individuals, we found that replacing 80 g/week of red and 299 processed meat or poultry with legumes was associated with a modestly lower NAFLD rate. 300 There is still uncertainty about the optimal type of dietary intervention for the prevention 301 of NAFLD, and studies have investigated foods as well as sources of protein. Prospective and 302 cross-sectional studies in humans have indicated that high consumption of animal-based 303 protein particularly from red and processed meat may increase risk of NAFLD, while plantbased proteins appear beneficial.^(3; 38-40) To our knowledge, very few substitution studies, 304 305 replacing animal-based foods for legumes, have investigated the risk of NAFLD. One 306 previous study using two Chinese cohorts investigated replacement of 25 g/day of animal-307 sourced foods with legumes in relation to NAFLD-rate. They found that replacing poultry 308 with legumes was associated with a lower HR for NAFLD in the Guangzhou Nutrition and 309 Health Study (HR: 0.35, 95% CI: 0.18; 0.69), while replacing red and processed meat or fish 310 with legumes resulted in higher but non-significant rates of NAFLD (meat HR: 1.16, 95% CI: 311 0.99; 1.37; fish HR: 1.09, 95% CI: 0.92; 1.30). In the Tianjin Chronic Low-grade Systemic 312 Inflammation and Health cohort, replacing processed meat and poultry with legumes was 313 associated with a lower rate of NAFLD, while unprocessed red meat and fish was not associated with NAFLD.⁽¹⁹⁾ An Iranian cross-sectional study did not find any association 314

It is made available under a CC-BY 4.0 International license .

between substituting protein sources (animal-based or plant-based) and NAFLD.⁽⁴¹⁾ One

study using the UKB substituted 15 g/day of red and processed meat for legumes and found

317 no association with liver cancer, despite this cancer being highly associated with NAFLD.

However, these findings could be due to few cases of liver cancer. $^{(42)}$

319 The substitution effect is a result of including one food and concurrently excluding another 320 food, and it is therefore not possible to assess whether any beneficial association is caused by 321 the higher intake of one food group, or the lower intake of the other food group. By replacing 322 animal-based protein foods like red and processed meat with legumes, intakes of dietary 323 fiber, unsaturated fats, minerals, antioxidants, and phytochemicals are increased. These 324 components may have favorable health impacts and could be key factors contributing to metabolic health, potentially lowering the risk of NAFLD.⁽⁴³⁻⁴⁶⁾ In the case of red and 325 326 processed meat, this substitution may furthermore lower the intakes of saturated fats, 327 cholesterol, heme iron, and daily acid loads. These components may trigger oxidative 328 reactions and pro-inflammatory cytokines, which up-regulate inflammatory responses and potentially increase the risk of developing NAFLD.^(38; 47; 48) 329 330 A substitution size of one weekly serving has previously been argued as a realistic

substitution for the general population.⁽¹⁹⁾ Such substitutions may also be feasible for dietary recommendations for general populations due to the simple consumer message.⁽²⁰⁾ For a disease like NAFLD with no current medical treatment, it is relevant to identify the optimal diet for preventing disease onset or worsening, and substitution analyses can provide invaluable information by specifying comparisons between individual foods.⁽⁴⁹⁾ Our results indicate a modest 3-4% lower incidence rate of NAFLD when replacing a serving of red and processed meat or poultry with legumes, but this was not the case for fish.

338 Compared to research on other foods, very few studies with Western populations have investigated legume intake specifically. This may be due to a negligible legume consumption 339 among many cohorts.^(7; 50-54) Other prospective studies, such as the French Etude NutriNet-340 341 Santé and the multinational Prospective Urban Rural Epidemiology (PURE) study, also have large proportions of non-consumers of legumes or very low intakes.^(14; 55) Large variability in 342 343 legume consumption exists globally and according to the Global Dietary Database, European 344 countries represent the lowest consumption worldwide as the population in more than one third of all countries consume less than 10 g/day.⁽⁷⁾ It may also be that weak or null 345 associations in combination with low intakes and high proportions of non-consumers, have 346

led to less focus to publish studies in this research area.^(51-54; 56) To address the high

348 proportion of non-consumers in our study population, we excluded non-consumers in

It is made available under a CC-BY 4.0 International license .

349 sensitivity analyses. The results did not change markedly, but this exclusion limits the

350 generalizability of these results to the general population.

351 Our substitution models did not adjust for energy intake leading to a caloric difference in 352 the substitution, which would be related to weight and NAFLD (Supplementary Figure 1). 353 Although BMI is clinically on the causal path between food intake and NAFLD, we observed 354 only minor differences in results of our food substitution models with and without adjustment 355 for baseline BMI. We were thus unable to confirm any mediation of the association between 356 replacing red and processed meats, poultry, or fish with legumes through BMI. Similarly, the study on the two Chinese cohorts found mixed associations despite adjusting for BMI.⁽¹⁹⁾ 357 358 Formal causal mediation studies of the role of BMI after baseline on these associations in

359 other studies may provide clearer answers.

360 This study has several strengths including the prospective cohort design, large sample size 361 and detailed assessment of lifestyle. Furthermore, the food-level substitution analysis is one 362 of the first to investigate the impacts of replacing red and processed meat, poultry, and fish 363 with legumes on the rate of NAFLD. The study does however also have some limitations. 364 Participants in health research often differentiate from the underlying population on certain 365 characteristics such as health behaviour and literacy. This may introduce selection on the 366 exposure level, where individuals with a higher consumption of healthy foods, such as 367 legumes, are more prone to participate. Our study may therefore not provide accurate 368 estimates of the prevalence of legume intake in the UK due to healthy volunteer bias. 369 Despite being developed and validated against recovery biomarkers, short-term dietary 370 assessments methods such as the Oxford WebQ have some shortcomings with assessing 371 habitual dietary intake compared to a food frequency questionnaire (FFQ), particularly 372 capturing usual intake of foods that are not consumed very frequently. If diet is only assessed at one time point, the FFQ captures habitual diet better.^(57; 58) Including only participants who 373 374 had completed at least two WebQs ensured more accurate estimation of usual intake compared to relying on a single measurement.^(59; 60) This approach also allowed us to 375 376 adequately capture the range of exposure necessary to examine the association between 377 legume intake and NAFLD. Most participants completed two diet assessments, but usual 378 intake of foods not consumed daily in this population, such as legumes and fish, may not 379 have been fully captured. However, the sensitivity analyses using three dietary assessments 380 did not show a change in magnitude or direction of associations. All participants completed a 381 touchscreen dietary questionnaire at recruitment, which may capture the usual intake of meat 382 and fish intakes better than the 24-hour dietary assessments. However, as legume

It is made available under a CC-BY 4.0 International license .

consumption was not reported separately from vegetables in the touchscreen questionnaire,

384 we were unable to use these data. Furthermore, the prospective study design ensures that

participation is non-differential with regards to the outcome, and thus selection bias is

unlikely to have affected our results. Despite the thorough adjustment for confounding,

residual confounding cannot be ruled out in observational studies. The substitution analyses

388 were furthermore purely observational limiting any causal inference from the results. Future

randomized controlled trials would be needed to confirm the findings of our study in a causalsetting.

391

392 Conclusion

Replacing red and processed meat or poultry with legumes was associated with a slightly

lower rate of NAFLD in this population from the UK. No association was observed when

substituting fish for legumes. While the associations identified in this study were modest, our

findings align with general recommendations promoting increased consumption of legumes.

Further research in populations with higher legume intake is warranted to confirm thesefindings.

399 **Acknowledgments:** The authors acknowledge the participants who provided data and the

400 members of the UK Biobank cohort who collected the data.

401 **Transparency Declaration:** The lead author affirms that this manuscript is an honest,

402 accurate, and transparent account of the study being reported. The reporting of this work is

403 compliant with the Strengthening the Reporting of Observational Studies in Epidemiology

404 guidelines for nutritional epidemiology (STROBE-nut). The lead author affirms that no

405 important aspects of the study have been omitted and that any discrepancies from the study as

406 planned (<u>https://doi.org/10.5281/zenodo.11670547</u>) have been explained.

407 **Conflicts of interest:** None of the authors have any conflicts of interest to declare.

408 **Funding:** This study was funded by Aarhus University and the Steno Diabetes Center

409 Aarhus. The Graduate School of Health, Aarhus University funded the salary of Fie

410 Langmann. Access to data and data management was funded by the Steno Diabetes Center

411 Aarhus. Daniel B. Ibsen was funded by the Independent Research Fund Denmark with grant

number 1057-00016B and the Danish Diabetes Association. The funding agencies had no

role in the design and conduct of the study; the collection, management, analysis, and

It is made available under a CC-BY 4.0 International license .

- 414 interpretation of the data; the preparation, review, and approval of the manuscript, or the
- 415 decision to publish the manuscript.
- 416 **Data availability:** The study was based on data from the UK Biobank prospective cohort that
- 417 are not publicly available due to personal information. Access to data can be acquired
- through an application to the Access Management System of UK Biobank online
- 419 (https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access).

It is made available under a CC-BY 4.0 International license .

420 **References**

- 421 1. Wang RZ, Zhang WS, Jiang CQ et al. Association of fish and meat consumption with non-alcoholic
- 422 fatty liver disease: Guangzhou Biobank Cohort Study. *BMC Public Health*, 2023; 23: 2433.
- 423 2. Tan LJ, Shin S. Effects of oily fish and its fatty acid intake on non-alcoholic fatty liver disease
- 424 development among South Korean adults. *Front Nutr*, 2022; **9**: 876909.
- 425 3. Hashemian M, Merat S, Poustchi H et al. Red Meat Consumption and Risk of Nonalcoholic Fatty
- Liver Disease in a Population With Low Meat Consumption: The Golestan Cohort Study. *Am J Gastroenterol*, 2021; **116**: 1667-1675.
- 428 4. Al-Dayyat HM, Rayyan YM, Tayyem RF. Non-alcoholic fatty liver disease and associated dietary and
- 429 lifestyle risk factors. *Diabetes Metab Synd*, 2018; **12**: 569-575.
- 430 5. Wattacheril JJ, Abdelmalek MF, Lim JK et al. AGA Clinical Practice Update on the Role of
- 431 Noninvasive Biomarkers in the Evaluation and Management of Nonalcoholic Fatty Liver Disease:
- 432 Expert Review. *Gastroenterology*, 2023; **165**: 1080-1088.
- 6. European Association for the Study of the Liver (EASL), European Association for the Study of
- 434 Diabetes (EASD), European Association for the Study of Obesity (EASO). EASL-EASD-EASO Clinical
- 435 Practice Guidelines for the management of non-alcoholic fatty liver disease. *J Hepatol*, 2016; 64:
 436 1388-1402.
- 437 7. Yanni AE, Iakovidi S, Vasilikopoulou E *et al.* Legumes: A Vehicle for Transition to Sustainability.
 438 *Nutrients*, 2023; **16**: 98.
- 439 8. Food Research Collaboration (2022) *Putting climate on everyone's table: Summary of what the*
- 440 IPCC WG3 report says about food and diet. City University of London.
- 441 9. Blomhoff R, Andersen R, Arnesen EK *et al.* (2023) Nordic Nutrition Recommendations 2023.
- 442 <u>https://www.norden.org/en/publication/nordic-nutrition-recommendations-2023</u>
- 443 10. Willett W, Rockström J, Loken B *et al.* Food in the Anthropocene: the EAT-Lancet Commission on
 444 healthy diets from sustainable food systems. *Lancet*, 2019; **393**: 447-492
- 445 11. Zhao N, Jiao K, Chiu Y-H *et al.* Pulse Consumption and Health Outcomes: A Scoping Review.
- 446 *Nutrients*, 2024; **16**: 1435.
- 12. Rebello CJ, Greenway FL, Finley JW. A review of the nutritional value of legumes and their effects
 on obesity and its related co-morbidities. *Obes Rev*, 2014; 15: 392-407.
- 449 13. Dinu M, Abbate R, Gensini GF et al. Vegetarian, vegan diets and multiple health outcomes: A
- 450 systematic review with meta-analysis of observational studies. *Crit Rev Food Sci Nutr* 2017; 57: 3640451 3649.
- 452 14. Miller V, Mente A, Dehghan M et al. Fruit, vegetable, and legume intake, and cardiovascular
- disease and deaths in 18 countries (PURE): a prospective cohort study. *Lancet*, 2017; **390**: 2037-2049.
- 455 15. Mudryj AN, Yu N, Hartman TJ *et al.* Pulse consumption in Canadian adults influences nutrient 456 intakes. *Br J Nutr,* 2012; **108** S27-36.
- 457 16. Hong J, Kim S, Kim H-S. Hepatoprotective Effects of Soybean Embryo by Enhancing Adiponectin-
- 458 Mediated AMP-Activated Protein Kinase α Pathway in High-Fat and High-Cholesterol Diet-Induced
 459 Nonalcoholic Fatty Liver Disease. *J Med Food*, 2016; **19**: 549-559.
- 460 17. Son Y, Jang MK, Jung MH. Vigna nakashimae extract prevents hepatic steatosis in obese mice fed 461 high-fat diets. *J Med Food*, 2014; **17**: 1322-1331.
- 462 18. Zhang S, Kumari S, Gu Y et al. Soy Food Intake Is Inversely Associated with Newly Diagnosed
- 463 Nonalcoholic Fatty Liver Disease in the TCLSIH Cohort Study. J Nutr, 2020; **150**: 3280-3287.
- 464 19. Zhang S, Yan Y, Meng G 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* 2023; **293**: 340353.
- 20. Ibsen DB, Laursen ASD, Würtz AML *et al.* Food substitution models for nutritional epidemiology. *Am J Clin Nutr* 2020; **113**: 294-303.
- 469 21. Song M, Fung TT, Hu FB *et al.* Association of Animal and Plant Protein Intake With All-Cause and
- 470 Cause-Specific Mortality. *JAMA Intern Med*, 2016; **176**: 1453-1463.

It is made available under a CC-BY 4.0 International license .

- 471 22. Sudlow C, Gallacher J, Allen N *et al.* UK biobank: an open access resource for identifying the
- 472 causes of a wide range of complex diseases of middle and old age. *PLoS Med*, 2015; **12**: e1001779.
- 473 23. Watling CZ, Kelly RK, Dunneram Y *et al.* Associations of intakes of total protein, protein from
- dairy sources, and dietary calcium with risks of colorectal, breast, and prostate cancer: a prospective
 analysis in UK Biobank. *Br J Cancer*, 2023; **129**: 636-647.
- 476 24. Greenwood DC, Hardie LJ, Frost GS *et al.* Validation of the Oxford WebQ Online 24-Hour Dietary
 477 Questionnaire Using Biomarkers. *Am J Epidemiol*, 2019; **188**: 1858-1867.
- 478 25. Perez-Cornago A, Pollard Z, Young H *et al.* Description of the updated nutrition calculation of the
- 479 Oxford WebQ questionnaire and comparison with the previous version among 207,144 participants
- 480 in UK Biobank. *Eur J Nutr*, 2021; **60**: 4019-4030.
- 481 26. Piernas C, Perez-Cornago A, Gao M *et al.* Describing a new food group classification system for
- 482 UK biobank: analysis of food groups and sources of macro- and micronutrients in 208,200
- 483 participants. *Eur J Nutr*, 2021; **60**: 2879-2890.
- 484 27. NHS Digital (2021) NHS Data Model and Dictionary International Classification of Diseases (ICD).
- 485 <u>https://datadictionary.nhs.uk/supporting information/international classification of diseases icd</u>
 486 _.html (accessed 14 October 2021)
- 487 28. Rinella ME, Lazarus JV, Ratziu V *et al.* A multisociety Delphi consensus statement on new fatty
- 488 liver disease nomenclature. *Hepatology*, 2023; **78**: 1966-1986.
- 489 29. UK Biobank (2020) *Hospital inpatient data*. UK Biobank.
- 490 30. National Centre for Research Methods (n.d.) Geographical Referencing Learning Resources -
- 491 Townsend Deprivation Index. <u>https://www.restore.ac.uk/geo-refer/36229dtuks00y19810000.php</u>
- 492 (accessed 23 April 2024)
- 493 31. UK National Health Services (2022) 5 A Day portion sizes. <u>https://www.nhs.uk/live-well/eat-</u>
 494 well/5-a-day/portion-sizes/ (accessed 16 December 2022)
- 495 32. Tomova GD, Arnold KF, Gilthorpe MS *et al*. Adjustment for energy intake in nutritional research:
- 496 a causal inference perspective. *Am J Clin Nutr,* 2022; **115**: 189-198.
- 497 33. Tomova GD, Gilthorpe MS, Tennant PWG. Theory and performance of substitution models for
- 498 estimating relative causal effects in nutritional epidemiology. *The American Journal of Clinical*499 *Nutrition*, 2022; **116**: 1379-1388.
- 34. Radonjić T, Milićević O, Jovanović I *et al.* Elevated Transaminases as Predictors of COVID-19
 Pneumonia Severity. *Medicina (Kaunas)*, 2022; **58**.
- 502 35. Claus M, Antoni C, Hofmann B. Factors associated with elevated alanine aminotransferase in
- employees of a German chemical company: results of a large cross-sectional study. BMC
 Gastroenterol. 2021: 21: 25.
- 505 36. Langmann F, Ibsen DB, Johnston LW et al. Protocol: Substituting red and processed meat,
- 506 poultry, and fish with legumes and risk of non-alcoholic fatty liver disease in a large prospective 507 cohort. *Zenodo*, 2024.
- 508 37. Landau W. The targets R package: a dynamic Make-like function-oriented pipeline toolkit for
- reproducibility and high-performance computing. *Journal of Open Source Software*, 2021; **6**: 2959.
- 510 38. Alferink LJ, Kiefte-de Jong JC, Erler NS *et al*. Association of dietary macronutrient composition 511 and non-alcoholic fatty liver disease in an ageing population: the Rotterdam Study. *Gut*, 2019; **68**:
- 512 1088-1098
- 513 39. Noureddin M, Zelber-Sagi S, Wilkens LR et al. Diet Associations With Nonalcoholic Fatty Liver
- 514 Disease in an Ethnically Diverse Population: The Multiethnic Cohort. *Hepatology*, 2020; **71**: 1940515 1952.
- 516 40. Rietman A, Sluik D, Feskens EJM et al. Associations between dietary factors and markers of
- 517 NAFLD in a general Dutch adult population. *Eur J Clin Nutr*, 2018; **72**: 117-123.
- 518 41. Amirkalali B, Khoonsari M, Sohrabi MR *et al.* Relationship between dietary macronutrient
- 519 composition and non-alcoholic fatty liver disease in lean and non-lean populations: a cross-sectional
- 520 study. *Public Health Nutr* 2021; **24**: 6178-6190.

It is made available under a CC-BY 4.0 International license .

- 521 42. Bock N, Langmann F, Johnston LW *et al.* The Association between the Substitution of Red Meat
- with Legumes and the Risk of Primary Liver Cancer in the UK Biobank: A Cohort Study. *Nutrients*,
 2024; 16: 2383.
- 43. Satija A, Bhupathiraju SN, Spiegelman D *et al.* Healthful and Unhealthful Plant-Based Diets and
- the Risk of Coronary Heart Disease in U.S. Adults. *J Am Coll Cardiol*, 2017; **70**: 411-422.
- 526 44. Ferreira H, Vasconcelos M, Gil AM et al. Benefits of pulse consumption on metabolism and
- health: A systematic review of randomized controlled trials. *Crit Rev Food Sci Nutr,* 2021; **61**: 85-96.
- 528 45. Bouchenak M, Lamri-Senhadji M. Nutritional quality of legumes, and their role in
- 529 cardiometabolic risk prevention: a review. *J Med Food*, 2013; **16**: 185-198.
- 46. Mudryj AN, Yu N, Aukema HM. Nutritional and health benefits of pulses. *Appl Physiol Nutr*
- 531 *Metab,* 2014; **39**: 1197-1204.
- 532 47. Ma J, Hennein R, Liu C et al. Improved Diet Quality Associates With Reduction in Liver Fat,
- Particularly in Individuals With High Genetic Risk Scores for Nonalcoholic Fatty Liver Disease.
 Gastroenterology, 2018; 155: 107-117.
- 48. Sivasubramanian BP, Dave M, Panchal V *et al.* Comprehensive Review of Red Meat Consumption and the Risk of Cancer. *Cureus*, 2023; **15**: e45324.
- 537 49. Song M, Giovannucci E. Substitution analysis in nutritional epidemiology: proceed with caution.
- 538 *Eur J Epidemiol*, 2018; **33**: 137-140.
- 539 50. Faber I, Henn K, Brugarolas M et al. Relevant characteristics of food products based on

540 alternative proteins according to European consumers. J Sci Food Agric, 2021.

- 541 51. Mertens E, Kuijsten A, Dofková M et al. Geographic and socioeconomic diversity of food and
- nutrient intakes: a comparison of four European countries. *Eur J Nutr,* 2019; **58**: 1475-1493.
- 543 52. Rawal V, Charrondiere R, Xipsiti M *et al.* (2019) *The Global Economy of Pulses*. Food and
- 544 Agriculture Organization of the United Nations.
- 545 53. Figueira N, Curtain F, Beck E *et al.* Consumer Understanding and Culinary Use of Legumes in 546 Australia. *Nutrients*, 2019; **11**: 1575.
- 547 54. Perera T, Russo C, Takata Y *et al.* Legume Consumption Patterns in US Adults: National Health
- and Nutrition Examination Survey (NHANES) 2011–2014 and Beans, Lentils, Peas (BLP) 2017 Survey.
 Nutrients, 2020; 12: 1237.
- 550 55. Berthy F, Brunin J, Allès B *et al*. Association between adherence to the EAT-Lancet diet and risk of
- cancer and cardiovascular outcomes in the prospective NutriNet-Santé cohort. *The American Journal* of Clinical Nutrition, 2022; **116**: 980-991.
- 553 56. Sharma H, Verma S. Is positive publication bias really a bias, or an intentionally created
- discrimination toward negative results? *Saudi J Anaesth*, 2019; **13**: 352-355.
- 555 57. Freedman LS, Midthune D, Arab L et al. Combining a Food Frequency Questionnaire With 24-
- 556 Hour Recalls to Increase the Precision of Estimation of Usual Dietary Intakes-Evidence From the
- 557 Validation Studies Pooling Project. *Am J Epidemiol*, 2018; **187**: 2227-2232.
- 558 58. Park Y, Dodd KW, Kipnis V *et al.* Comparison of self-reported dietary intakes from the Automated
- 559 Self-Administered 24-h recall, 4-d food records, and food-frequency questionnaires against recovery
- biomarkers. *The American Journal of Clinical Nutrition*, 2018; **107**: 80-93.
- 561 59. Carter JL, Lewington S, Piernas C et al. Reproducibility of dietary intakes of macronutrients,
- specific food groups, and dietary patterns in 211 050 adults in the UK Biobank study. J Nutr Sci, 2019;
 8: e34.
- 564 60. Galante J, Adamska L, Young A *et al.* The acceptability of repeat Internet-based hybrid diet
- assessment of previous 24-h dietary intake: administration of the Oxford WebQ in UK Biobank.
- 566 British Journal of Nutrition, 2016; **115**: 681-686.
- 567

		Legume consumption strata			
Characteristic	All participants	Non-consumers	Low intake	Medium intake	High intake
Characteristic	N = 124,194	N = 73,711	N = 15,571	N = 18,055	N = 16,857
Legume intake, g/week	0 (0, 473)	0 (0, 0)	81.7 (40.8, 145.8)	245 (163.3, 315)	590.6 (419.1, 1,540)
NAFLD ^a	1,201 (1.0%)	736 (1.0%)	146 (0.9%)	160 (0.9%)	159 (0.9%)
Age	57.0 (44.0, 66.0)	58.0 (45.0, 66.0)	57.0 (44.0, 66.0)	57.0 (44.0, 66.0)	56.0 (44.0, 66.0)
Sex					
Female	69,273 (56%)	41,178 (56%)	9,061 (58%)	9,601 (53%)	9,433 (56%)
Male	54,921 (44%)	32,533 (44%)	6,510 (42%)	8,454 (47%)	7,424 (44%)
Yearly income, £					
>100,000	8,904 (7.2%)	5,325 (7.2%)	1,371 (8.8%)	1,269 (7.0%)	939 (5.6%)
52,000-100,000	28,831 (23%)	17,048 (23%)	3,899 (25%)	4,269 (24%)	3,615 (21%)
31,000-51,999	32,819 (26%)	19,419 (26%)	4,085 (26%)	4,756 (26%)	4,559 (27%)
18,000-30,999	26,854 (22%)	15,983 (22%)	3,146 (20%)	3,984 (22%)	3,741 (22%)
<18,000	15,602 (13%)	9,134 (12%)	1,749 (11%)	2,245 (12%)	2,474 (15%)
Unknown	11,184 (9%)	6802 (9.3%)	1321 (8.4%)	1532 (8.4%)	1529 (9.1%)
Educational level ^b					
High	58,435 (47%)	33,499 (45%)	8,285 (53%)	8,790 (49%)	7,861 (47%)
Intermediate	41,127 (33%)	25,250 (34%)	4,636 (30%)	5,735 (32%)	5,506 (33%)
Low	24,632 (20%)	14,962 (20%)	2,650 (17%)	3,530 (20%)	3,490 (21%)
Deprivation score ^c	-2.4 (-4.6, 2.6)	-2.5 (-4.7, 2.4)	-2.3 (-4.6, 2.7)	-2.3 (-4.6, 2.8)	-2.1 (-4.6, 3.0)
cohabitation					
Alone	22,134 (18%)	12,920 (18%)	2,692 (17%)	3,198 (18%)	3,324 (20%)
With spouse/partner	92,515 (75%)	55,185 (75%)	11,698 (75%)	13,491 (75%)	12,141 (72%)
Other non-partner	9,366 (7.5%)	5,516 (7.5%)	1,153 (7.4%)	1,343 (7.4%)	1,354 (8.0%)
Unknown	179 (0.1%)	90 (0.1%)	28 (0.2%)	23 (0.1%)	38 (0.2%)
Ethnicity					
White	119,968 (97%)	71,771 (97%)	14,879 (96%)	17,280 (96%)	16,038 (95%)
Other	4,226 (3.4%)	1,940 (2.6%)	692 (4.4%)	775 (4.3%)	819 (4.9%)

Table 1. Baseline characteristics across consumption of legumes in the UK Biobank cohort across legume consumption strata (N = 124, 194)

Physical activity ^d					
High	21,596 (17%)	12,408 (17%)	2,685 (17%)	3,216 (18%)	3,287 (19%)
Moderate	55,037 (44%)	32,174 (44%)	7,102 (46%)	8,163 (45%)	7,598 (45%)
Low	30,167 (24%)	18,460 (25%)	3,735 (24%)	4,271 (24%)	3,701 (22%)
Unknown	17,394 (14%)	10,669 (14%)	2,049 (13%)	2,405 (13%)	2,271 (13%)
Smoking status					
Current, >15	1,797 (1.4%)	1,173 (1.6%)	189 (1.2%)	239 (1.3%)	196 (1.2%)
cigarettes/day					
Current, <15	3,379 (2.7%)	2,076 (2.8%)	397 (2.5%)	481 (2.7%)	425 (2.5%)
cigarettes/day					
Former	44,379 (36%)	25,939 (35%)	5,617 (36%)	6,617 (37%)	6,206 (37%)
Never	70,978 (57%)	42,400 (58%)	8,918 (57%)	10,158 (56%)	9,502 (56%)
Unknown	3,661 (2.9%)	2,123 (2.9%)	450 (2.9%)	560 (3.1%)	528 (3.1%)
Alcohol consumption, g/day	11.4 (0.0, 44.0)	11.8 (0.0, 44.9)	12.6 (0.0, 44.4)	11.3 (0.0, 44.1)	8.5 (0.0, 39.5)
$BMI \ge 30 \text{ kg/m}^2$	24,556 (20%)	14,871 (20%)	2,894 (19%)	3,577 (20%)	3,214 (19%)
History of NAFLD-related	45,747 (37%)	27,390 (37%)	5,713 (37%)	6,521 (36%)	6,123 (36%)
diseases ^e					
Family history of related	96,936 (78%)	57,428 (78%)	12,127 (78%)	14,084 (78%)	13,297 (79%)
diseases ^f					
Weekly food group intakes,					
g/week ^g					
Red and processed meat	371 (0, 844)	396 (0, 847)	373 (0, 840)	368 (0, 866)	280 (0, 840)
Poultry	175 (0, 607)	228 (0, 607)	182 (0, 607)	152 (0, 607)	0 (0, 607)
Fish	175 (0, 595)	175 (0, 607)	175 (0, 560)	175 (0, 578)	127 (0, 621)
Refined cereals	851 (257, 1,675)	849 (255, 1,677)	873 (288, 1,647)	869 (273, 1,677)	816 (238, 1,686)
Whole grain cereals	475 (0, 1,425)	431 (0, 1,384)	490 (0, 1,410)	518 (0, 1,468)	588 (0, 1,617)
Mixed dishes	210 (0, 1,167)	175 (0, 1,143)	329 (0, 1,177)	263 (0, 1,208)	303 (0, 1,342)
Dairy	1,943 (718, 3,456)	1,986 (840, 3,483)	1,948 (783, 3,404)	1,993 (759, 3,505)	1,604 (292, 3,290)
Fats	80 (8, 181)	81 (8, 182)	75 (8, 172)	81 (8, 183)	77 (6, 182)
Fruits	1,341 (280, 2,905)	1,292 (245, 2,814)	1,365 (350, 2,901)	1,386 (327, 2,996)	1,512 (350, 3,210)
Nuts	7 (0, 151)	0 (0, 140)	13 (0, 154)	11 (0, 164)	11 (0, 191)
Vegetables	1,190 (331, 2,571)	1,127 (298, 2,423)	1,244 (451, 2,538)	1,271 (403, 2,753)	1,349 (379, 3,049)

It is made available under a CC-BY 4.0 International license .

20

Potatoes	626 (0, 1,260)	630 (0, 1,260)	613 (0, 1,248)	624 (0, 1,260)	630 (0, 1,324)
Eggs and egg dishes	0 (0, 420)	0 (0, 420)	70 (0, 420)	88 (0, 438)	70 (0, 513)
Non alashalia havanasaa	10,733 (7,158,	10,605 (7,070,	10,798 (7,271,	10,815 (7,280,	11,078 (7,350,
Non-alconolic beverages	15,295)	15,155)	15,260)	15,49)	15,820)
Alcoholic beverages	938 (0, 4,676)	987 (0, 4,760)	1,050 (0, 4,718)	940 (0, 4,782.4)	613 (0, 4,237)
Snacks and sweets	518 (112, 1,195)	527 (116, 1,212)	495 (117, 1,120)	512 (117, 1,180)	495 (98, 1,204)
Sauces and condiments	117 (0, 385)	117 (0, 385)	128 (0, 385)	117 (0, 373)	105 (0, 390)
Total weight of consumed	22,023 (16,611,	21,689 (16,331,	22,134 (16,764,	22,453 (17,067,	22,993 (17,386,
foods	28,879)	28,384)	28,633)	29,534)	30,367)

Continuous variables are presented as median (10%, 90%) and categorical values as number of participants (%). ^aNAFLD, non-alcoholic fatty liver disease. ^bEducational level was defined as low (Certificate of Secondary Education (CSE), National Vocational Qualifications, Higher National Diploma, Higher National Certificates, other professional qualifications, or equivalent), intermediate (A levels, O levels, General Certificate of Secondary Education, or equivalent), and high (College or University degree). ^cDeprivation was assessed with the Townsend Deprivation Index based on four indicators of material deprivation: non-home ownership, non-car ownership, unemployment, and overcrowding. Positive values indicate that individuals live in areas with high material deprivation and negative values indicate relative affluence.⁽³⁰⁾ ^dPhysical activity was based on total metabolic equivalent task (MET) minutes per week for all activity including walking, moderate, and vigorous activity, and defined as low (0-9.9 METs/week), moderate (10-49.9 METs/week), high (≥50 METs/week), and unknown. ^eHistory of NAFLD-related diseases was defined as participant's previous diagnosis of diabetes, alcoholic liver disease, angina pectoris, hypertension, myocardial infarction, stroke, elevated cholesterol, gallbladder disease, or cancer, ^fFamily history of related diseases was defined as diagnosis of diabetes, heart disease, stroke, hypertension, or elevated cholesterol among the mother, father, and/or biological sibling(s) of the participant. ^gThe definition and inclusion of foods in each food group can be found in Supplementary Table 1.

It is made available under a CC-BY 4.0 International license

Table 2. Hazard ratios and 95 % confidence intervals for non-alcoholic fatty liver disease in the UK Biobank when consuming 80 g/week of legumes in specific and non-specific substitutions, and without food substitution.

Statistical model	Consuming 80 g/week of legumes instead of 80 g/week of			
Full sample	Red and processed	Poultry ^b	Fish ^c	
(N = 124, 194, events = 1201)	meat ^a			
Model 1	0.96 (0.94; 0.97)	0.97 (0.95; 0.99)	0.99 (0.97; 1.01)	
Model 2	0.96 (0.94; 0.98)	0.97 (0.95; 0.99)	0.98 (0.96; 1.01)	
Model 3	0.97 (0.95; 0.99)	0.98 (0.96; 1.00)	0.99 (0.97; 1.01)	
Consumers of legumes ^d				
(N = 50,483, events = 465)				
Model 2	0.97 (0.94; 0.99)	0.98 (0.95; 1.02)	0.99 (0.95; 1.02)	

Model 1 was adjusted for age at recruitment, geographical region of recruitment, sex, g/week intake of all other dietary components (red and processed meat, poultry, fish, refined cereal, whole grain cereal, fruits, vegetables, potatoes, nuts, dairy, fats, eggs and egg-dishes, mixed dishes, snacks and sweets, sauce and condiments, non-alcoholic beverages, and alcoholic beverages) apart from the food to be substituted, and total intake of all dietary components in g/week. Model 2 was further adjusted for alcohol intake (ethanol g/week), ethnicity, Townsend Deprivation Index, educational level, yearly income, cohabitation, physical activity, smoking status, history of NAFLD-related diseases, family history of NAFLD-related diseases, and cancer. Model 3 was further adjusted for BMI \geq 30 kg/m². ^aRed and processed meat included beef, pork, lamb, and other meats including offal, and sausages, bacon, ham, and liver pâté. ^bPoultry included poultry with or without skin and fried poultry with batter or breadcrumbs. ^cFish included oily fish, white fish, tinned tuna, fried fish with batter or breadcrumbs, and shellfish. ^dStudy sample after excluding participants who reported no consumption of legumes at either of the completed dietary assessments.

It is made available under a CC-BY 4.0 International license

It is made available under a CC-BY 4.0 International license

Table 3. Hazard ratio and 95% confidence intervals for NAFLD when substituting 80 g/week of meat, poultry, or fish with 80 g/week of legumes across altered exposure level and exclusion criteria.

	Hazard ratios and 95 % confidence intervals		
Altered exposure level	Red and	Poultry ^b	Fish ^c
	processed meat ^a		
Replacing 80 g/week of legumes, including peas, for animal-	0.95 (0.92; 0.99)	0.97 (0.93; 1.01)	0.98 (0.94; 1.02)
based foods, $N = 124,194$ (events = 1201)			
Replacing 80 g/week of legumes, excluding soymilk, for	0.96 (0.93; 0.98)	0.97 (0.95; 1.00)	0.98 (0.96; 1.01)
animal-based foods, N =124,194 (events = 1201)			
Altered exclusion criteria			
Individuals with alcohol intake below 90 th percentile, N=	0.95 (0.94; 0.97)	0.97 (0.95; 0.99)	0.98 (0.96; 1.00)
111,774 (events = 1043)			
Individuals with normal alanine aminotransferase level ^d , N =	0.96 (0.94; 0.98)	0.98 (0.95; 1.00)	0.98 (0.96; 1.01)
108,754 (events = 823)			
Individuals with \geq 3 24-hour assessments ^e , N = 76,440 (events	0.96 (0.93; 0.98)	0.98 (0.95; 1.01)	0.98 (0.95; 1.01)
= 696)			

Adjustment followed the main analyses Model 2 and adjusted for age at recruitment, geographical region of recruitment, sex, g/week intake of all other dietary components (meat, poultry, fish, refined cereal, whole grain cereal, fruits, vegetables, potatoes, nuts, dairy, fats, eggs and egg-dishes, mixed dishes, snacks and sweets, sauce and condiments, non-alcoholic beverages, and alcoholic beverages) apart from the food to be substituted, total intake of all dietary components in g/week, alcohol intake (ethanol g/week), ethnicity, Townsend Deprivation Index, educational level, yearly income, cohabitation, physical activity, smoking status, history of NAFLD-related diseases, family history of NAFLD-related diseases, and cancer. ^aRed and processed meat included beef, pork, lamb, and other meats including offal, and sausages, bacon, ham, and liver pâté. ^bPoultry included poultry with or without skin and fried poultry with batter or breadcrumbs. ^cFish included oily fish, white fish, tinned tuna, fried fish with batter or breadcrumbs, and shellfish. ^dAnalyses including individuals with alanine aminotransferase levels below 40 U/L. ^eAnalyses restricted to individuals with three or more completed 24-hour dietary assessments.

It is made available under a CC-BY 4.0 International license .



Figure 1. Flowchart of participants in the UK Biobank eligible for inclusion