Supplementary Material

Supplement to: Diet quality and risk and severity of COVID-19: a prospective cohort study

Jordi Merino,[1,2,3#] Amit D. Joshi,[4,5#] Long H. Nguyen,[4,5,6#] Emily R. Leeming,[7] Mohsen Mazidi,[7] David A Drew,[4,5] Rachel Gibson,[8] Mark S. Graham,[9] Chun-Han Lo,[4,5] Joan Capdevila,[10] Benjamin Murray,[9] Christina Hu,[10] Somesh Selvachandran,[10] Sohee Kwon,[4,5] Wenjie Ma,[4,5] Cristina Menni,[7] Alexander Hammers,[9,11] Shilpa N. Bhupathiraju,[3,12] Shreela V. Sharma,[14] Carole Sudre,[9] Christina M. Astley,[2,13] Walter C. Willet,[12,15,16] Jorge E. Chavarro,[12,15,16] Sebastien Ourselin,[9] Claire J. Steves,[7] Jonathan Wolf,[10] Paul W. Franks,[12,17] Tim D. Spector, MBBS,[8*] Sarah E. Berry,[8*] Andrew T. Chan,[4,5*]

Correspondence: Andrew T. Chan, M.D., M.P.H. Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, 100 Cambridge Street Boston, MA 02114
achan@mgh.harvard.edu

Table of Contents:

Supplementary Methods

Protocol

Supplementary Tables

Supplementary Figures

Supplementary Methods

Dietary intake assessment

Habitual dietary intake information was collected through an amended version of the Leeds Short Form Food Frequency Questionnaire (LSF-FFQ). [1] In brief, the LSF-FFQ includes 20 food items with reference to fruit, vegetables, fibre-rich foods, high fat and high-sugar foods, meat, meat products and fish. Seven additional food items were added to capture broader dietary intake information including refined carbohydrates (e.g. white rice, white pasta and white bread), eggs, fast food and live probiotic or fermented foods (e.g. live yogurt, kefir and kimchi). Participants were asked how often on average they had consumed one portion of each food in a typical week. The responses had eight frequency categories ranging from "rarely or never" to "five or more times per day. Further detail on the development, dissemination and procedures of the diet and lifestyle survey to UK and US participants is described elsewhere.[2]

Outcome ascertainment

Predicted COVID-19 definition: We used a symptom-based classifier developed by our group to predict COVID-19.[3] To build the prediction model, UK participants were randomly divided into a training set and a test set (ratio: 80:20). Based on the training set, a logistic model generated to predict symptomatic COVID-19 was: Log odds (Predicted COVID-19) = -1.32 - (0.01 x age) + (0.44 x male sex) + (1.75 x loss of smell or taste) + (0.31 x severe or significant persistent cough) + (0.49 x severe fatigue) + (0.39 x skipped meals). The prediction model achieved a sensitivity of 0.65 (95% CI 0.62-0.67) and specificity of 0.78 (95% CI 0.76-0.80) in the test set. In additional validation in the U.S. participants, the prediction model achieved a sensitivity of 0.66 (95% CI 0.62-0.69) and specificity of 0.83 (95% CI 0.82-0.85).

Severe COVID: To ascertain severe COVID-19, we used responses to the question "*What treatment did you receive while in the hospital / What treatment are you receiving right now?*" Participants had the option to respond a) None, b) Oxygen and fluids breathing support administered through an oxygen mask, no pressure applied, c) Non-invasive ventilation breathing support administered through an oxygen mask, which pushes oxygen into your lungs, d) Invasive ventilation breathing support administered through an inserted tube. People are usually asleep for this procedure, e) Other. COVID-19 severity was ascertained based on a report of the need for a hospital visit which required 1) non-invasive breathing support, 2) invasive breathing support, and 3) administration of antibiotics combined with oxygen support.

Covariate classification

Covariates were selected *a priori* based on putative confounders and risk factors for COVID-19 and included sex (male, female), race/ethnicity (White, Black, Asian, Other), index of multiple deprivation (most deprived <3, intermediate deprived 3 to 7, less deprived >7), population density (<500 individuals/km^2, 500 to 1,999 individuals/km^2, 2,000 to 4,999 individuals/km^2, and ≥ 5,000 individuals/km^2), healthcare worker status (yes with interaction with COVID-19 patients, yes without interaction with COVID-19 patients, no), presence of comorbidities [diabetes (yes, no), cardiovascular disease (yes, no), lung disease (yes, no), cancer (yes, no), kidney disease (yes, no)], body mass index (<18.5 kg/m2, 18.5 to 24.9 kg/m2, 25.0 to 29.9 kg/m2, and ≥30 kg/m2), smoking status (yes, no), and physical activity (<1 day/week, 1 to 2 days/week, 3 to 4 days/week, ≥5 days/week).

Interactions between diet quality and deprivation on COVID-19 risk

We tested for additive interactions by assessing the relative excess risk due to interaction, and further examined the risk proportions attributable to diet quality alone, to deprivation alone, and to their interaction. For these analyses, we considered diet quality and socioeconomic deprivation as continuous variables. We assessed the relative excess risk due to interaction as an index of additive interaction using the following formula ($RERI = RR_{11} - RR_{10} - RR_{01} + 1$)[4], and further examined the decomposition of the joint effect, which is the proportion attributable to genetic risk alone, to diet quality alone, and to their interaction (i.e., $AP = RERI / RR_{11}$).[4]

References

1 Cleghorn CL, Harrison RA, Ransley JK, et al. Can a dietary quality score derived from a short-form FFQ assess dietary quality in UK adult population surveys? Public Health Nutr 2016;19:2915–2923

2 Mazidi M, Leming E, Merino J, et al. Impact of COVID-19 on health behaviours and body weight: A prospective observational study in a cohort of 1.1 million UK and US individuals. Research Square. [Preprint]. Cited 2021 April 25.

3 Menni C, Valdes AM, Freidin MB, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020;26:1037–1040.

4 VanderWeele TJ, Tchetgen Tchetgen EJ. Attributing effects to interactions. Epidemiology 2014;25:711–22.

Pre-specified protocol

Purpose and meaning

The main objective of the present proposal is to use self-reported individual-level data from up to 1.1 million volunteers included in the COVID Symptom Study to evaluate the associations between diet quality and COVID-19 risk and severity. In addition we will investigate its intersection with deprivation.. Findings from this study have the potential to identify susceptible individuals to increased COVID-19 risk and severity and inform public health strategies to reduce the burden of the COVID-19 pandemic.

1. General methodological considerations

Dataset preparation: We will collect daily app responses to generate a dataset that contains follow-up data from March 24, 2020 and followed until December 2, 2020. We will obtain information on demographic factors, self-reported COVID-19 or any COVID-19 related symptoms and personal and medical history including lung disease, diabetes, cardiovascular disease, cancer, kidney disease, and use of medications.

Quality Control: For this study we will include individuals who responded to the diet and lifestyle survey for the pre-pandemic period. We will filter out multiple records for the same participant and time point, records without an indication of whether they were recorded pre- or peri- pandemic, and records not linked to UK or US participants.

Main exclusions: Prevalent COVID-19 prior to start of follow-up. Presence of symptoms that classify them as having predicted COVID-19 within 24 hours of first entry. Participants younger than 18 years old. Pregnant women. Participants who logged only one daily assessment during follow-up.

1.1 Primary outcome and exposures

Outcome: The main outcome will be COVID-19 risk defined using a validated symptom based-algorithm developed by our research team.

Secondary outcomes will include:

COVID-19 risk base on report of a positive COVID-19 test by RT-PCR.

COVID-19 severity will be defined based on the risk of hospitalization and the need of oxygen requirements based on responses to the following question "*What treatment did you receive while in the hospital / What treatment are you receiving right now?*"

Exposure definitions: Diet quality will be quantified using diet quality indices. We will generate the healthful plant-based diet index (hPDI) and the Diet Quality Score (DQS). To generate these scores we will use the items and weighting criteria used in previous studies.

1.2 Covariates

Covariates will be selected a priori based on putative confounders and risk factors for COVID-19. Modes will be adjusted for 10-year age group, country of origin (UK, US), sex (male, female), race/ethnicity (White, Black, Asian, Other), index of multiple deprivation (most deprived <3, intermediate deprived 3 to 7, less deprived >7), population density (<500 individuals/km2, 500 to 1,999 individuals/km2, 2,000 to 4,999 individuals/km2, and ≥ 5,000 individuals/km2), healthcare worker status (yes with interaction with COVID-19 patients, yes without interaction with COVID-19 patients, no), presence of comorbidities [diabetes (yes, no), cardiovascular disease (yes, no), lung disease (yes, no), cancer (yes, no), kidney disease (yes, no)], body mass index (<18.5 kg/m2, 18.5 to 24.9 kg/m2, 25.0 to 29.9 kg/m2, and ≥30 kg/m2), smoking status (yes, no), and physical activity (<1 day/week, 1 to 2 days/week, 3 to 4 days/week, ≥5 days/week).

1.3 Unit of analysis

Merino J, *et al. Gut* 2022; 70:2096–2104. doi: 10.1136/gutjnl-2021-325353

Estimated effect sizes will be reported per change in diet quality category (low diet quality would be the reference) or 1SD increase.

1.4 Subgroup analysis

We will assess the association between diet quality and COVID-19 risk according to comorbidities, demographic, and lifestyle characteristics. We will also classified participants according to categories of the diet quality score and socioeconomic deprivation and conducted joint analyses. We will also test for additive interactions by assessing the relative excess risk due to interaction, and further examined the COVID-19 risk proportions attributable to diet, deprivation, and to their interaction.

2. Statistical analyses

1. Multiple imputations by chained equations with five imputations will be used to handle missing data.

2. Follow-up time for each participant will start 24 hours after first log-in to the time of predicted COVID-19 (or to time of secondary outcomes) or date of last entry prior to December 2, 2020, whichever occurred first.

3. Cox regression models will be stratified by calendar date at study entry, country of origin, and 10-year age group (age-adjusted model 1). Model 2 will be further adjusted for sex, race/ethnicity, index of multiple deprivation, population density, and healthcare worker status. Model 3 will be further adjusted for presence of diabetes, cardiovascular disease, lung disease, cancer, kidney disease, body mass index, smoking status, and physical activity.

4. Absolute risk will be calculated as the percentage of COVID-19 cases occurring per 10,000 person-months.

5. We will use restricted cubic splines with four knots (at the 2.5th, 25th, 75th, and 97.5th percentiles) to assess for non-linear associations between diet quality and COVID-19 risk.

6. For COVID-19 risk defined by a positive test we will use inverse probability-weighted Cox models to account for predictors of btaining country-specific testing. Inverse probability-weighted analyses will include presence of COVID-19-related symptoms, interaction with a person with COVID-19, occupation as a healthcare worker, age group, and race. These models will be adjusted for the same confounders as before.

7. Subgroup analysis will be based according to comorbidities, demographic, and lifestyle characteristics; Age (<60, ≥60), Sex (Male, Female), Race (White, Non-white), Deprivation (Low, Intermediate, High), Population density (<2,000, ≥2,000), Healthcare worker (yes, no), BMI (<25, 25-30, ≥ 30), physical activity (<1d/wk, 1-4 d/wk, ≥5d/wk). Cox models will be adjusted for the same covariates as previous model 3. In a sensitivity analysis we will use the DQS score to investigate associations between diet quality and COVID-19 risk and severity. These models will be adjusted for the same confounders as before.

9. Sensitivity analysis to censor cases that occurred after completing the diet survey. These models will be adjusted for the same confounders as before.

10. Sensitivity analysis to account for regional differences in the effective reproductive number (Rt) or mask wearing. For Rt analyses, we will extract US state-level information from the COVID Tracking Project for the period between March 2020 and January 2021. For the UK we will calculate R_t time-series for Scotland, Wales, and each of the NHS regions in England, using a previously published methodology from our group. For these analyses, we will define community peak and nadir R_t time-windows as the period between one week before and two weeks after R_t was all-time high or low. Using censored time-windows, we will test the association between diet quality and COVID-19 risk after adjusting for the same confounders as included in model 3. For mask wearing analyses, we will use survey data launched on June/September 2020 on whether participants had worn a face mask when outside the house in the last week. Responses will be categorized into never, sometimes, most of the time, or always. Mask wearing analyses will include the same covariates as included in model 3.

Research team: Jordi Merino, Amit D. Joshi, Long H. Nguyen, Emily Leeming, Sarah E. Berry, Andrew T. Chan. This research proposal was approved by the research team on 11/02/2020.

Merino J, *et al. Gut* 2022; 70:2096–2104. doi: 10.1136/gutjnl-2021-325353

Supplementary table 1: Grouping and components of the hPDI

hPDI component	**FFQ items**
Wholegrains	Fibre-rich breakfast cereal, like Weetabix, Fruit ‘n Fibre, Porridge, Muesli; Wholemeal bread or chapattis
Fruits	Fruit (tinned / fresh)
Vegetables	Salad (not garnish added to sandwiches); Vegetables (tinned / frozen / fresh but not potatoes)
Nuts	N/A
Legumes	Beans or pulses like baked beans, chick peas, dahl
Vegetable oils	N/A
Tea and Coffee	N/A
Fruit Juice	Fruit juice (not cordial or squash)
Refined Grains	Crisps / savoury snacks; pasta; Refined breakfast cereals (e.g. rice krispies, cornflakes, coco pops); rice; white bread
Potatoes	Chips / fried potatoes
Sugar Sweetened Beverages	Nonalcoholic fizzy drinks/pop (not sugar free or diet)
Sweets and desserts	Sweet biscuits, cakes, chocolate, sweets
Animal fats	N/A
Dairy	Cheese / yoghurt; Ice cream / cream; live probiotic or fermented food products (e.g. yoghurt, kefir, kimchi)
Egg	Eggs - as boiled, fried, scrambled, etc
Fish and seafood	White fish in batter or breadcrumbs – like ‘fish ‘n chips’; White fish not in batter or breadcrumbs; Oily fish – like herrings, sardines, salmon, trout, mackerel, fresh tuna (not tinned tuna)
Meat	Beef, Lamb, Pork, Ham - steaks, roasts, joints, mince or chops; Chicken or Turkey – steaks, roasts, joints, mince or portions (not in batter or breadcrumbs); Sausages, bacon, corned beef, meat pies/pasties, burgers; Chicken/turkey nuggets/twizzlers, turkey burgers, chicken pies, or in batter or breadcrumbs
Miscellaneous	Fast food

Table legend: FFQ items constituting the 18 food groups originally used to generate the healthy plant-based diet index in Satija et al. JACC 2017. Four out of the 18 food groups originally considered were not included for the calculation of the healthy plant-based diet index in this study as they were not available (N/A). FFQ = food frequency questionnaire; hPDI = healthful plant-based diet index

Merino J, *et al. Gut* 2022; 70:2096–2104. doi: 10.1136/gutjnl-2021-325353

Supplementary table 2: Criteria for scoring each component of the hPDI

Component	Criteria for min score of 1	Criteria for max score of 5
Whole grain	Lowest quintile of intake	Highest quintile of intake
Fruits	Lowest quintile of intake	Highest quintile of intake
Vegetables	Lowest quintile of intake	Highest quintile of intake
Nuts	N/A	N/A
Legumes	Lowest quintile of intake	
Vegetable oils	N/A	N/A
Tea and coffee	N/A	N/A
Fruit juices	Highest quintile of intake	Lowest quintile of intake
Refined grains	Highest quintile of intake	Lowest quintile of intake
Potatoes	Highest quintile of intake	Lowest quintile of intake
Sugar sweetened beverages	Highest quintile of intake	Lowest quintile of intake
Sweets and desserts	Highest quintile of intake	Lowest quintile of intake
Animal fat	N/A	N/A
Dairy	Highest quintile of intake	Lowest quintile of intake
Egg	Highest quintile of intake	Lowest quintile of intake
Fish or seafood	Highest quintile of intake	Lowest quintile of intake
Meat	Highest quintile of intake	Lowest quintile of intake
Miscellaneous	Highest quintile of intake	Lowest quintile of intake

Table legend: Criteria for scoring the 18 food groups originally used to generate the healthy plant-based diet index in Satija et al. JACC 2017. Food groups were ranked into quintiles, and given positive (healthy plant food groups) or reverse scores (less healthy plant food groups and animal food groups). With positive scores, participants above the highest quintile of a food group received a score of 5, following on through to participants below the lowest quintile who received a score of 1. With reverse scores, this pattern of scoring was inverted. All component scores were summed to obtain a total score ranging from 0 (lowest diet quality) to 70 (highest diet quality) points.

Merino J, *et al. Gut* 2022; 70:2096–2104. doi: 10.1136/gutjnl-2021-325353

Supplementary table 3: Grouping and components of the DQS

DQS component	FFQ items
Fruits	Fruit (tinned / fresh)
Vegetables	Salad (not garnish added to sandwiches) Vegetables (tinned / frozen / fresh but not potatoes)
Oily fish	Oily fish – like herrings, sardines, salmon, trout, mackerel, fresh tuna (not tinned tuna)
Total fat	Fruit (tinned / fresh) Fruit juice (not cordial or squash) Salad (not garnish added to sandwiches) Vegetables (tinned / frozen / fresh but not potatoes) Chips / fried potatoes Beans or pulses like baked beans, chick peas, dahl Fiber-rich breakfast cereal, like Weetabix, Fruit ‘n Fiber, Porridge, Muesli Whole-meal bread or chapattis; Cheese / yoghurt; Crisps / savory snacks Sweet biscuits, cakes, chocolate, sweets Ice cream / cream Nonalcoholic fizzy drinks/pop (not sugar free or diet) Beef, Lamb, Pork, Ham - steaks, roasts, joints, mince or chops Chicken or Turkey – steaks, roasts, joints, mince or portions (not in batter or breadcrumbs) Processed meats/ meat products Sausages, bacon, corned beef, meat pies/pasties, burgers Chicken/turkey nuggets/twizzles, turkey burgers, chicken pies, or in batter or breadcrumbs White fish in batter or breadcrumbs – like ‘fish ‘n chips’ White fish not in batter or breadcrumbs
Non-milk extrinsic sugars	Fruit (tinned / fresh) Fruit juice (not cordial or squash) Salad (not garnish added to sandwiches) Vegetables (tinned / frozen / fresh but not potatoes) Chips / fried potatoes Beans or pulses like baked beans, chick peas, dahl Fiber-rich breakfast cereal, like Weetabix, Fruit ‘n Fiber, Porridge, Muesli Whole-meal bread or chapattis; Cheese / yoghurt; Crisps / savory snacks Sweet biscuits, cakes, chocolate, sweets Ice cream / cream Nonalcoholic fizzy drinks/pop (not sugar free or diet) Beef, Lamb, Pork, Ham - steaks, roasts, joints, mince or chops Chicken or Turkey – steaks, roasts, joints, mince or portions (not in batter or breadcrumbs) Processed meats/ meat products Sausages, bacon, corned beef, meat pies/pasties, burgers Chicken/turkey nuggets/twizzles, turkey burgers, chicken pies, or in batter or breadcrumbs White fish in batter or breadcrumbs – like ‘fish ‘n chips’ White fish not in batter or breadcrumbs

Table legend: FFQ items constituting the 5 food components originally used to generate the DQS score from Cleghorn et al., listed in the Nutritools library (nutritools.org). FFQ = food frequency questionnaire; DQS = diet quality score.

Merino J, *et al. Gut* 2022; 70:2096–2104. doi: 10.1136/gutjnl-2021-325353

Supplementary table 4: Criteria for scoring each component of the DQS

DQS Component	Criteria for score of 1	Criteria for score of 2	Criteria for score of 3
Fruit	≤ 2 servings/week	>2 servings/week and <2 servings/d	≥2 servings/d
Vegetables	≤ 1 servings/d	1-3 servings/d	≥ 3 servings/d
Oily Fish	No intake	0-200g/week	200g/week
Total Fat	≥1.5 x UK recommendations (≥127.5g/d)	1-1.5 x UK recommendations	≤ UK recommendations (≤ 85g/d)
Non-Milk-Extrinsic Sugars	≥1.5 x UK recommendations (≥ 90g/d)	1-1.5 x UK recommendations	≤ UK recommendations (≤ 60g/d)

Table legend: Criteria for scoring the 5 food groups originally used to generate the diet quality score from Cleghorn et al., listed in the Nutritools library (nutritools.org). Each component was scored from 1 (unhealthiest) to 3 (healthiest) points, with intermediate values scored proportionally. All component scores were summed to obtain a total score ranging from 5 (lowest diet quality) to 15 points(highest diet quality) points.

Supplementary Table 5: Demographic, lifestyle, and clinical characteristics according to diet and lifestyle survey participation

	Included participants (n=592,571)	**Participants who did not respond to the diet survey** (n= 3,289,680)
Age, years	56 (44-65)	43 (32-56)
≥18-24	14,397 (2.4)	312,927 (9.5)
25-34	52,922 (8.9)	715,438 (21.7)
35-44	86,251 (14.6)	721,725 (21.9)
45-54	125,802 (21.2)	635,749 (19.3)
55-64	158,637 (26.8)	482,356 (14.7)
≥65	153,810 (26.0)	421,485 (12.8)
Sex, No. (%)		
Male	187,450 (31.6)	1,311,439 (39.9)
Female	404,126 (68.2)	1,974,754 (60.1)
Race[ɛ], No. (%),		
White	568,770 (96.0)	2,214,416 (67.3)
Black	4,328 (0.7)	35,932 (1.1)
Asian	10,435 (1.8)	85,605 (2.6)
Other/Prefer not to say	9,038 (1.5)	953,727 (28.9)
Country, No. (%)		
UK	543,984 (91.8)	3,026,997 (92.0)
US	48,587 (8.2)	262,683 (8.0)
Index of deprivation, No. (%)[¶]		
Most deprived, decile 1	1,3416 (2.3)	128875 (3.9)
Least deprived, decile 10	103,608 (17.5)	408310 (12.4)
Population density, km^2, No. (%)[¶]		
<500	119,782 (20.2)	133,740 (4.1)
500-1,999	90,541 (15.3)	534,421 (16.2)
2,000 4,999	94,345 (15.9)	874,726 (26.6)
≥5,000	244,295 (41.2)	1,213,590 (36.9)
Healthcare worker, No. (%)		
Yes	41,141 (6.9)	274,052 (8.3)
Body mass index, Kg/m^2	25.1 (22.6-28.7)	25.7 (22.8-29.6)
<18.5	12,004 (2.0)	98,815 (3.1)
18.5-24.9	277,536 (46.8)	1,366,485 (41.5)
25-29.9	189,197 (31.9)	1,054,239 (32.0)
≥30	113,056 (19.1)	769,687 (23.4)
Diabetes	20,058 (3.4)	99,807 (3.0)
Heart disease	20,376 (3.4)	89,260 (2.7)
Cancer	6,559 (1.9)	28,545 (1.4)

Merino J, *et al. Gut* 2022; 70:2096–2104. doi: 10.1136/gutjnl-2021-325353

Lung disease	62,999 (10.6)	346,150 (10.5)
Kidney disease	5,134 (0.9)	24,251 (0.9)

Table Legend: Values are median (P_{25}-P_{75}) for continuous variables; numbers and (percentages) for categorical variables.

[ε] Race was self-reported by the participants.

[¶] Index of deprivation and population density were generated using zip code or postcode information linked with census track data.

Merino J, *et al. Gut* 2022; 70:2096–2104. doi: 10.1136/gutjnl-2021-325353

Supplementary table 6: Adjusted hazard ratios of COVID-19 risk and severity for diet quality in the COVID Symptom Study

	Low DQS	Intermediate DQS	High DQS	*P* for trend
Diet quality score, median (IQR)	9 (8-10)	11 (11-11)	13 (12 -13)	
COVID-19 risk				
No. of events/person-months	13,996 / 1,467,205	12,641 / 1,701,799	5,178 / 717,270	—
Incidence rate (10,000 person-months; 95% CI)	95.4 (93.8-97.0)	74.3 (73.0-75.6)	72.2 (70.3-74.2)	—
Age-adjusted model	1.00 (Ref)	0.90 (0.87-0.92)	0.92 (0.89-0.95)	<0.001
Multivariable model 2	1.00 (Ref)	0.90 (0.88-0.92)	0.92 (0.89-0.95)	0.019
Multivariable model 3	1.00 (Ref)	0.95 (0.93-0.98)	1.00 (0.97-1.03)	0.216
COVID-19 risk (positive test)				
No. of events/person-months	2,341 / 1,515,004	2,309 / 1,746,982	952 / 736,535	—
Incidence rate (10,000 person-months; 95% CI)	15.5 (14.8-16.1)	13.2 (12.7-13.8)	12.9 (12.1-13.7)	—
Age-adjusted model[$]	1.00 (Ref)	0.94 (0.91-0.98)	0.92 (0.87-0.96)	<0.001
Multivariable model 2[$]	1.00 (Ref)	0.95 (0.91-0.99)	0.93 (0.89-0.98)	0.006
Multivariable model 3[$]	1.00 (Ref)	0.96 (0.93-1.00)	0.95 (0.91-1.00)	0.047
COVID-19 severity				
No. of events/person-months	313 / 1,518,980	317 / 1,750,786	110 / 738,495	—
Incidence rate (10,000 person-months; 95% CI)	2.1 (1.8-2.3)	1.8 (1.6-2.0)	1.5 (1.2-1.8)	—
Age-adjusted model	1.00 (Ref)	0.82 (0.70-0.96)	0.67 (0.54-0.84)	<0.001
Multivariable model 2	1.00 (Ref)	0.82 (0.70-0.97)	0.68 (0.54-0.85)	<0.001
Multivariable model 3	1.00 (Ref)	0.93 (0.79-1.09)	0.83 (0.66-1.04)	0.141

Merino J, *et al. Gut* 2022; 70:2096–2104. doi: 10.1136/gutjnl-2021-325353

Table legend: Hazards ratios and 95% CI for COVID-19 risk and severity. Sensitivity analysis using the DQS to quantify diet quality. Cox proportional hazards models were stratified by calendar date at study entry, country of origin, and 10-year age group (Age-adjusted model). Multivariable model 2 was further adjusted for sex (male, female), race/ethnicity (White, Black, Asian, Other), index of multiple deprivation (most deprived <3, intermediate deprived 3 to 7, less deprived >7), population density (<500 individuals/km^2, 500 to 1,999 individuals/km^2, 2,000 to 4,999 individuals/km^2, and ≥ 5,000 individuals/km^2), and healthcare worker status (yes with interaction with COVID-19 patients, yes without interaction with COVID-19 patients, no). Model 3 was further adjusted for presence of comorbidities [diabetes (yes, no), cardiovascular disease (yes, no), lung disease (yes, no), cancer (yes, no), kidney disease (yes, no)], body mass index (<18.5 kg/m2, 18.5 to 24.9 kg/m2, 25.0 to 29.9 kg/m2, and ≥30 kg/m2), smoking status (yes, no), and physical activity (<1 day/week, 1 to 2 days/week, 3 to 4 days/week, ≥5 days/week).

[$] Inverse probability-weighted analyses were conducted to account for predictors of obtaining RT-PCR testing (presence of COVID-19-related symptoms, interaction with a COVID-19 case, healthcare worker, age group, and race). inverse probability-weighted Cox proportional hazards models were stratified by 10-year age group and date with additional adjustment for the covariates used in previous models.

Merino J, *et al. Gut* 2022; 70:2096–2104. doi: 10.1136/gutjnl-2021-325353

Supplementary table 7: Association between diet quality and COVID risk - censored to cases that occurred after completing the diet questionnaires

	Low hPDI	Intermediate hPDI	High hPDI	*P* for trend
COVID-19 risk				
Incidence rate (10,000 person-months; 95% CI)	116.2 (110.9-120.3)	84.4 (82.0-86.7)	74.1 (71.5-78.6)	—
Age-adjusted model	1.00 (Ref)	0.82 (0.78-0.86)	0.79 (0.75-0.83)	<0.001
Multivariable model 2	1.00 (Ref)	0.83 (0.79-0.87)	0.79 (0.75-0.84)	<0.001
Multivariable model 3	1.00 (Ref)	0.87 (0.83-0.92)	0.88 (0.83-0.93)	<0.001
COVID-19 risk (positive test)				
Incidence rate (10,000 person-months; 95% CI)	42.1 (40.2-44.3)	33.5 (32.1-35.0)	29.1 (27.2-31.0)	—
Age-adjusted model[$]	1.00 (Ref)	0.84 (0.77-0.92)	0.77 (0.70-0.86)	<0.001
Multivariable model 2[$]	1.00 (Ref)	0.85 (0.78-0.93)	0.79 (0.72-0.88)	<0.001
Multivariable model 3[$]	1.00 (Ref)	0.86 (0.79-0.94)	0.80 (0.72-0.89)	<0.001

Table legend: Hazards ratios and 95% CI for COVID-19 risk. Sensitivity analysis censored to cases that occurred after completing the diet questionnaires (September 21st, 2020). Cox proportional hazards models were stratified by calendar date at study entry, country of origin, and 10-year age group (Age-adjusted model). Multivariable model 2 was further adjusted for sex (male, female), race/ethnicity (White, Black, Asian, Other), index of multiple deprivation (most deprived <3, intermediate deprived 3 to 7, less deprived >7), population density (<500 individuals/km^2, 500 to 1,999 individuals/km^2, 2,000 to 4,999 individuals/km^2, and ≥ 5,000 individuals/km^2), and healthcare worker status (yes with interaction with COVID-19 patients, yes without interaction with COVID-19 patients, no). Model 3 was further adjusted for presence of comorbidities [diabetes (yes, no), cardiovascular disease (yes, no), lung disease (yes, no), cancer (yes, no), kidney disease (yes, no)], body mass index (<18.5 kg/m2, 18.5 to 24.9 kg/m2, 25.0 to 29.9 kg/m2, and ≥30 kg/m2), smoking status (yes, no), and physical activity (<1 day/week, 1 to 2 days/week, 3 to 4 days/week, ≥5 days/week).

[$] Inverse probability-weighted analyses were conducted to account for predictors of obtaining RT-PCR testing (presence of COVID-19-related symptoms, interaction with a COVID-19 case, healthcare worker, age group, and race). inverse probability-weighted Cox proportional hazards models were stratified by 10-year age group and date with additional adjustment for the covariates used in previous models.

Merino J, *et al. Gut* 2022; 70:2096–2104. doi: 10.1136/gutjnl-2021-325353

Supplementary table 8: Attributing associations to additive interaction between diet quality and socioeconomic deprivation on risk of COVID-19 infection

	Predicted COVID-19 infection
Main effects	
Diet quality, per 10 units decrease	1.05 (1.01-1.09)
Deprivation index, per category decrease	1.06 (1.01-1.12)
Joint effect	1.15 (1.09-1.21)
Relative excess risk due to interaction	
Relative excess risk due to interaction	0.05 (0.02-0.08)
P	0.005
Attributable proportion, %	
Diet quality	31.9 (18.2-45.6)
Deprivation index	38.4 (26.5-50.3)
Additive interaction	29.7 (2.1-57.3)

Table Legend: Multivariable-adjusted risk of predicted COVID-19 infection estimated from fully adjusted Cox models. The relative excess risk due to interaction was calculated using the following formula ($RERI_{RR} = RR_{11} - RR_{10} - RR_{01} + 1$). The decomposition of the joint effect, which is the proportions attributable to diet quality alone, to deprivation index alone, and to their interaction, was calculated using the following formula (i.e., $AP = RERI / RR_{11}$).

Merino J, *et al. Gut* 2022; 70:2096–2104. doi: 10.1136/gutjnl-2021-325353

Supplementary table 9: Association between diet quality and risk of COVID-19 infection after accounting for mask wearing

	Low hPDI	Intermediate hPDI	High hPDI	*P* for trend
COVID-19 risk				
No. of events/person-months	2,574 / 222,426	4,669 / 555,918	2,092 / 283,975	—
Incidence rate (10,000 person-months; 95% CI)	114.6 (110.2-119.0)	84.0 (81.6-86.4)	73.7 (70.6-76.9)	—
Multivariable Model 3 + Mask wearing	1.00 (Ref)	0.88 (0.83 to 0.92)	0.88 (0.83-0.94)	<0.001
COVID-19 risk (positive test)				
No. of events/person-months	989 / 233,564	1,907 / 576,267	874 / 293,760	
Incidence rate (10,000 person-months; 95% CI)	42.3 (39.8-45.1)	33.1 (31.6-34.6)	29.8 (27.8-31.8)	
Multivariable Model 3 + Mask wearing[$]	1.00 (Ref)	0.86 (0.79-0.94)	0.80 (0.72-0.89)	<0.001

Table legend: Hazards ratios and 95% CI for COVID-19 risk after accounting for mask wearing. These analyses were left censored to September 21st 2020. Mask wearing analyses included 524,825 participants. For confirmed COVID-19 analyses, inverse probability-weighted analyses were conducted to account for predictors of obtaining RT-PCR testing[$].

Supplementary table 10: Total, direct, and indirect effects of diet quality on COVID-19 risk

	COVID-19 risk	
	HR (95% CI)	**P value**
Total effect, 1SD increase in hPDI	0.96 (0.96-0.97)	<0.001
Direct effect	0.98 (0.97-0.98)	<0.001
Indirect effect	0.99 (0.98-0.99)	<0.001
Proportion mediated	37% (30-44)	<0.001

Table legend: Structural equation models were implemented to conduct a mediation analysis of BMI using the "lavaan" package in R. For this analysis, diet quality and BMI were used as continuous variables. We estimated the relative contribution of BMI to the association between diet quality and COVID-19 risk and computed the proportion of total effect that was explained by indirect effects of BMI. Indirect effects were estimated by taking the product of the effect of the exposure (diet quality) on the mediator (BMI) and the effect of the mediator (BMI) on the outcome (COVID-19 risk). The direct effect is defined as the association of diet quality on COVID-19 risk through mechanisms independent of mediation and was estimated from regressing COVID-19 on diet quality. To calculate the proportion of the mediated effect we divided the indirect effect by the total effect.

Supplementary figure 1: Diet and symptom data collection among participants of the COVID Symptom Study

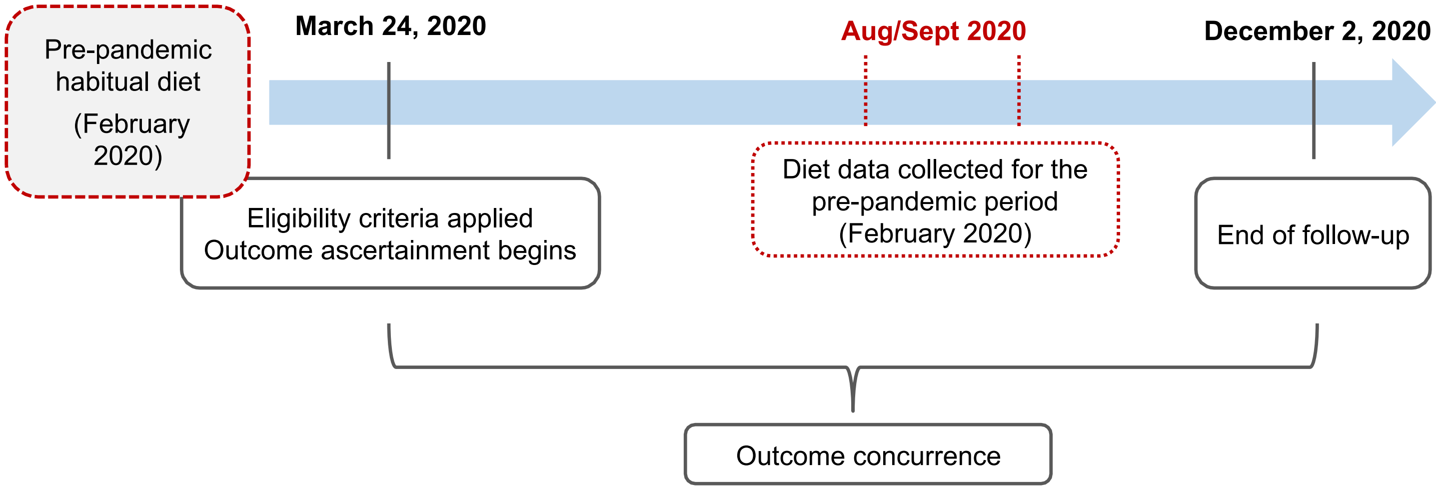

Figure legend: Schematic representation showing how and when diet and symptom data was collected. The diet survey was launched in August / September 2020 and queried about participant's habitual diet before (based on a time frame of February 2020) and during the pandemic (based on a time frame of July / August 2020). For the primary analysis, we used diet data deemed pre-pandemic. During follow-up, daily prompts queried for updates on interim symptoms, health care visits, and COVID-19 testing results.

Merino J, *et al. Gut* 2022; 70:2096–2104. doi: 10.1136/gutjnl-2021-325353

Supplementary figure 2: Pattern of missing data before multiple imputation

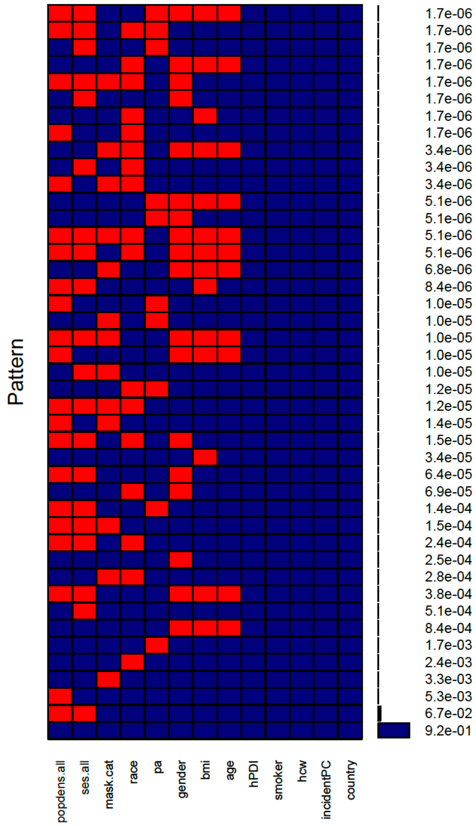

Figure legend: *The plot shows the pattern of missing data across all variables and individuals included in this study. The values on the left side indicate the % of participants with missing data for combinations of variables. About 92% of included participants had complete information.*

Supplementary figure 3: Directed acyclic graph depicting a possible scenario that could explain the association between diet quality and COVID-19 risk and severity.

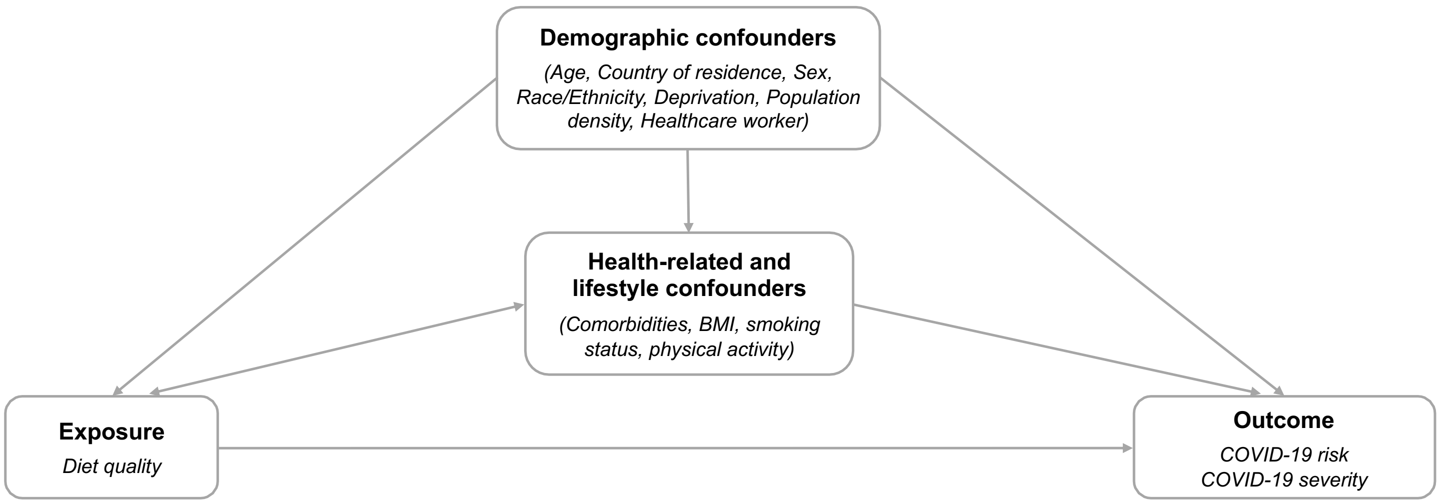

Figure legend: Directed acyclic graph showing the potential association between diet quality and COVID-19 risk and severity. Demographic confounders were included in the age-adjusted model and model 2. Model 3 was further adjusted for health-related and lifestyle confounders. In subgroup and sensitivity analyses, we investigated whether diet quality interacts with deprivation, and the extent to which BMI mediated the association between diet quality and COVID-19 risk.

Merino J, *et al. Gut* 2022; 70:2096–2104. doi: 10.1136/gutjnl-2021-325353

Supplementary figure 4: Flow diagram

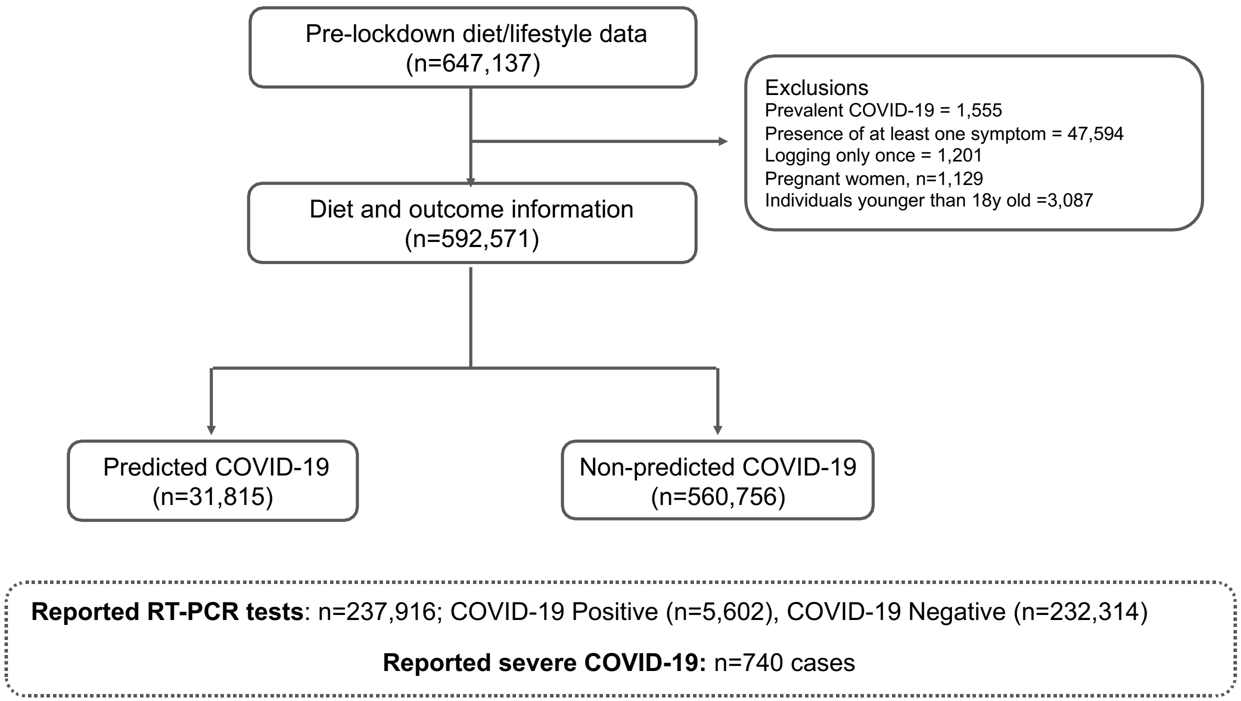

Figure legend: Identification of participants with diet and lifestyle data at baseline who met the eligibility criteria for this study. Number of cases and controls identified until the end of follow-up (December 2nd, 2020)

Supplementary figure 5: Distribution of the hPDI score

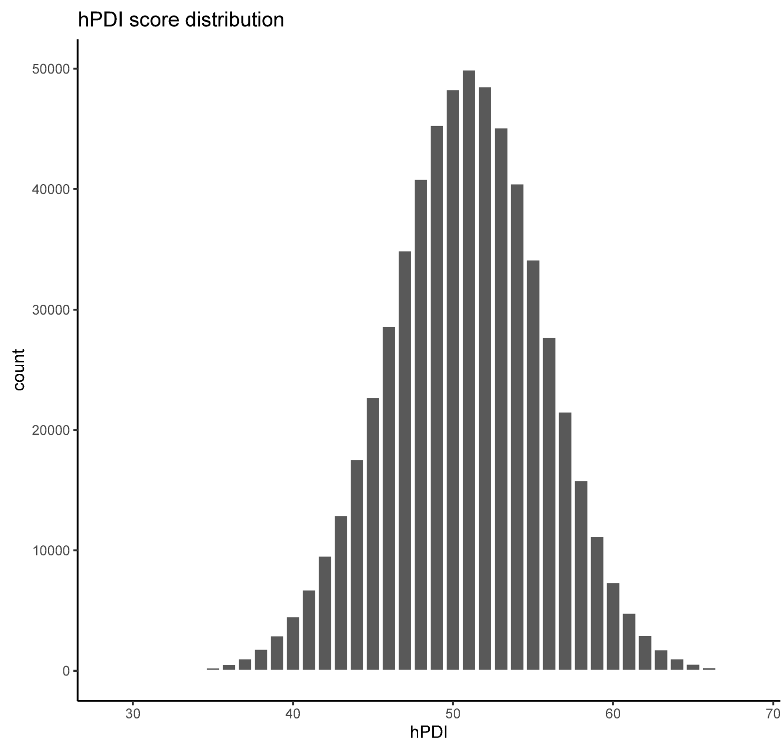

Figure legend: Distribution of the healthy plant-based diet index.

Merino J, *et al. Gut* 2022; 70:2096–2104. doi: 10.1136/gutjnl-2021-325353

Supplementary figure 6: Dose-response associations between diet quality and risk of COVID-19 infection.

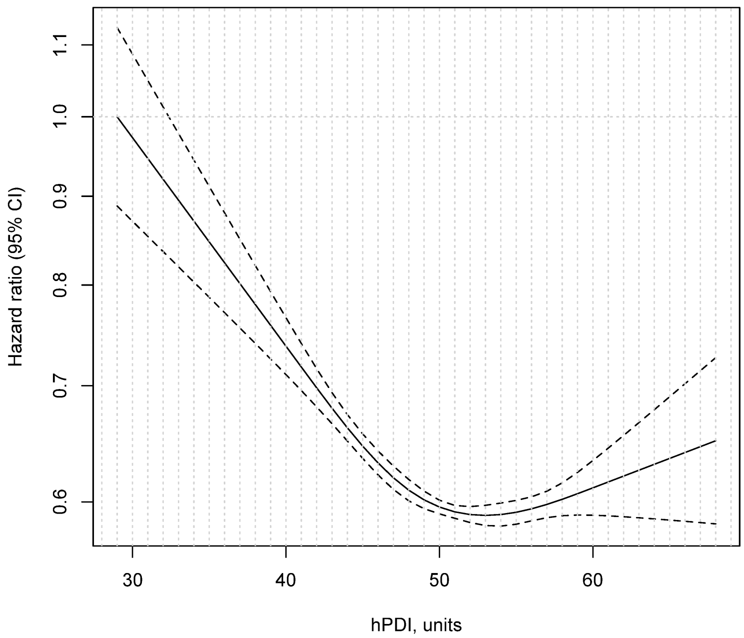

Figure legend: Dose-response associations between diet quality and risk of COVID-19 infection were calculated using restricted cubic splines with four knots (at the 2.5th, 25th, 75th, and 97.5th percentiles; methods). Cox models were adjusted for all confounders previously included in model 3. P for non-linearity <0.001.

Merino J, *et al. Gut* 2022; 70:2096–2104. doi: 10.1136/gutjnl-2021-325353

Supplementary figure 7: Absolute excess rate of COVID-19 per 10,000 person-months according to socioeconomic deprivation and diet quality

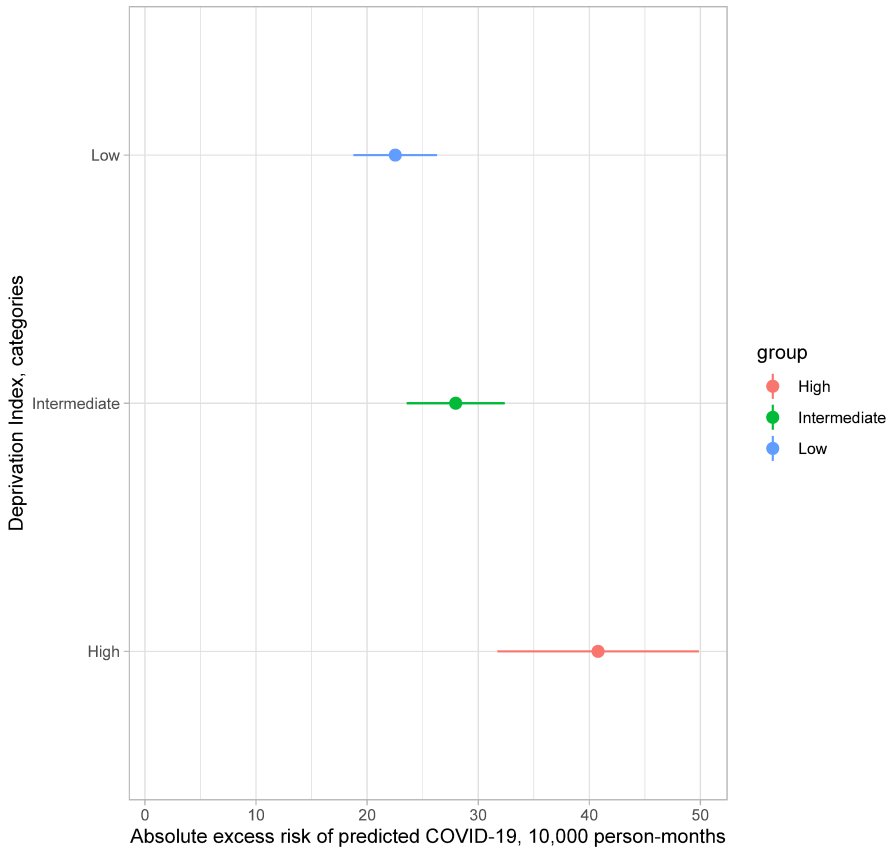

Figure legend: Absolute excess risk of COVID-19 per 10,000 person-months for lowest vs highest quartile of the diet score according to socioeconomic deprivation. Absolute excess risk was calculated based on the incidence rate per 1,000 person-months in each diet quality score and socioeconomic deprivation category using the “epiR” package in R.

Merino J, *et al. Gut* 2022; 70:2096–2104. doi: 10.1136/gutjnl-2021-325353