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

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

1. Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA.
2. Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA.
3. Department of Medicine, Harvard Medical School, Boston, MA, USA.
4. Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.
5. Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.
6. Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA.
7. Department of Twin Research, King's College London, London UK
8. Department of Nutritional Sciences, King's College London, London UK
9. School of Biomedical Engineering & Imaging Sciences, King's College London, London, UK
10. Zoe Global Limited, London, UK
11. King's College London & Guy's and St Thomas' PET Centre, King's College London, London, UK
12. Department of Nutrition, Harvard TH Chan School of Public Health, Boston, MA, USA

13. Division of Endocrinology & Computational Epidemiology, Boston Children's Hospital, Harvard Medical School, Boston, M
14. Department of Epidemiology, Human Genetics, and Environmental Sciences, UTHealth, School of Public Health, Houston, TX, USA
15. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
16. Channing Division of Network Medicine, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA, USA
17. Department of Clinical Sciences, Lund University, Malmö, Sweden

Authors share co-first authorship

* Authors share co-last authorship

Correspondence:

Andrew T. Chan, M.D., M.P.H. Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, 100 Cambridge Street Boston, MA 02114; Phone: 617-726-7802; Fax: 617-643-0195; Email: achan@mgh.harvard.edu

Abstract

Objective: Poor metabolic health and certain lifestyle factors have been associated with risk and severity of coronavirus disease 2019 (COVID-19), but data for diet are lacking. We aimed to investigate the association of diet quality with risk and severity of COVID-19 and its intersection with socioeconomic deprivation.

Design: We used data from 592,571 participants of the smartphone-based COVID Symptom Study. Diet quality was assessed using a healthful plant-based diet score, which emphasizes healthy plant foods such as fruits or vegetables. Multivariable Cox models were fitted to calculate hazard ratios (HR) and 95% confidence intervals (95% CI) for COVID-19 risk and severity defined using a validated symptom-based algorithm or hospitalization with oxygen support, respectively.

Results: Over 3,886,274 person-months of follow-up, 31,815 COVID-19 cases were documented. Compared with individuals in the lowest quartile of the diet score, high diet quality was associated with lower risk of COVID-19 (HR, 0.91; 95% CI, 0.88-0.94) and severe COVID-19 (HR, 0.59; 95% CI, 0.47-0.74). The joint association of low diet quality and increased deprivation on COVID-19 risk was higher than the sum of the risk associated with each factor alone ($P_{interaction}$=0.005). The corresponding absolute excess rate for lowest vs highest quartile of diet score was 22.5 (95% CI, 18.8-26.3) and 40.8 (95% CI, 31.7-49.8; 10,000 person-months) among persons living in areas with low and high deprivation, respectively.

Conclusions: A dietary pattern characterized by healthy plant-based foods was associated with lower risk and severity of COVID-19. These association may be particularly evident among individuals living in areas with higher socioeconomic deprivation.

Introduction

Poor metabolic health[1,2] has been associated with increased risk and severity of coronavirus disease 2019 (COVID-19), and excess adiposity or preexisting liver disease might be causally associated with increased risk of death from COVID-19.[3,4] Underlying these conditions is the contribution of a diet, which may be independently associated with COVID-19 risk and severity.

On the basis of prior scientific evidence, diet quality scores have been developed to evaluate the healthfulness of dietary patterns.[5–7] Dietary patterns capture the complexity of food intakes better than any one individual food item and offer the advantage of describing usual consumption of foods in typical diets.[8] One such diet score is the healthful plant-based diet index (hPDI), which emphasizes intake of healthy plant foods, such as fruits, vegetables, and whole grains, and has been associated with lower risk of metabolic diseases.[5,9,10]

Adherence to healthful dietary patterns may also be a proximal manifestation of distal social determinants of health.[11–13] Addressing adverse social determinants of health, such as poor nutrition, has been shown to reduce the burden of certain infectious diseases in the past,[14] supporting calls for prioritizing social determinants of health in the public health response to COVID-19. However, evidence on the association between diet quality and the risk and severity of COVID-19 is lacking, especially in the context of upstream social determinants of health. To address this evidence gap, we analyzed data for 592,571 United Kingdom (UK) and United States (US) participants from the smartphone-based COVID Symptom Study,[15] to prospectively investigate the association of diet quality with risk and severity of COVID-19 and its intersection with socioeconomic deprivation.

Materials and Methods

Study design and participants

The COVID Symptom Study is a smartphone-based study conducted in the UK and US. Study design and sampling procedures have been published elsewhere.[15] This analysis included participants recruited from March 24, 2020 and followed until December 2, 2020. Participants who reported any symptoms related to COVID-19 prior to start of follow-up, or reported symptoms that classified them as having predicted COVID-19 within 24 hours of first entry, or who tested positive for COVID-19 at any time prior to start of follow-up or 24 hours after first entry were excluded. We also excluded participants younger than 18 years old, pregnant, and participants who logged only one daily assessment during follow-up. At enrollment, we obtained informed consent to the use of volunteered information for research purposes and shared relevant privacy policies and terms of use agreements. The study protocol was approved by the Mass General Brigham Human Research Committee (protocol 2020P000909) and King's College London Ethics Committee (REMAS ID 18210, LRS-19/20-18210).

Data collection procedures

Information on demographic factors was collected through standardized questionnaires at baseline,[15] including self-reported COVID-19 or any COVID-19 related symptoms and personal medical history including lung disease, diabetes, cardiovascular disease, cancer, kidney disease, and use of medications. During follow-up, daily prompts queried for updates on interim symptoms, health care visits, and COVID-19 testing results. Through software updates, a survey to examine self-reported diet and lifestyle habits during the pre/early-pandemic period was launched between August and September 2020. Details about this survey are available in the Supplementary Methods and published elsewhere.[16]

Assessment of diet quality

Diet quality was assessed using information obtained from an amended version of the Leeds Short Form Food Frequency Questionnaire[17] that included 27 food items (online supplementary methods). Participants were asked how often on average they had consumed one portion of each item in a typical week. The responses had eight frequency categories ranging from "rarely or never" to "five or more times per day".

Diet quality was quantified using the validated hPDI score.[5] To compute the hPDI, the 27 food items were combined into 14 food groups (online supplementary table 1). The original hPDI score included 18 food groups but nuts, vegetable oils, tea or coffee, and animal fat were not specifically queried. Food groups were ranked into quintiles and given positive (healthy plant food groups) or reverse scores (less healthy plant and animal food groups). With positive scores, participants within the highest quintile of a food group received a score of 5, following on through to participants within 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 14 (lowest diet quality) to 70 (highest) points. Criteria for generation of the hPDI are provided in online supplemental table 2. As an additional method to quantify diet quality based on available diet information, we used the Diet Quality Score (DQS).[17] The DQS is a score for adherence to UK dietary guidelines and was computed from five broad categories including fruits, vegetables, total fat, oily fish, and non-milk extrinsic sugars. Each component was scored from 1 (unhealthiest) to 3 (healthiest) points, with intermediate values scored proportionally (online supplementary table 3). All component scores were summed to obtain a total score ranging from 5 (lowest diet quality) to 15 (highest) points (online supplementary table 4).

Assessment of COVID-19 risk and severity

The primary outcome of this analysis was COVID-19 risk defined using a validated symptom-based algorithm,[18] which provides similar estimates of COVID-19 prevalence and incidence as those reported from the Office for National Statistics Community Infection Survey.[19] Details on the symptoms included in the predictive algorithm and corresponding weights are provided in the online supplementary methods. In brief, the symptom-based approach uses an algorithm to predict whether a participant has been infected with SARS-CoV-2 on the basis of their reported symptoms, age, and sex. The rationale for symptom-based classifier as a primary outcome was due to widespread difficulties obtaining testing during the early stages of the pandemic.[20] Secondary outcomes were confirmed COVID-19 based on a self-report of a reverse transcription polymerase chain reaction (RT-PCR) positive test and COVID-19 severity. 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 (online supplementary methods).

Statistical analysis

We summarized continuous measurements by using medians and interquartile ranges, and present categorical observations as frequency and percentages. Based on zip code (US) or post code (UK) of residence, participants were assigned to country-specific community-level socioeconomic measures including socioeconomic deprivation and population density (online supplementary methods). The methods for classifying socioeconomic deprivation, population density, and other *a priori* selected covariates are provided in online supplementary methods.

Multiple imputations by chained equations with five imputations were used to impute missing values. All covariates in the primary analysis were included in the multiple imputation procedure, and estimates generated from each imputed dataset were combined using Rubin's rules.[21]

Follow-up time for each participant started 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. We modeled the diet quality score as a continuous variable and generated categories of the score based on quartiles of the distribution (quartile 1, low diet quality; quartiles 2-3, intermediate diet quality; quartile 4, high diet quality). Cox regression models stratified by calendar date at study entry, country of origin, and 10-year age group were used to calculate hazard ratio (HR) and 95% confidence intervals (95% CI) for COVID-19 risk and severity (age-adjusted model 1). Model 2 was further adjusted for sex, race/ethnicity, index of multiple deprivation, population density, presence of diabetes, cardiovascular disease, lung disease, cancer, kidney disease, and healthcare worker status. Model 3 was further adjusted for body mass index, smoking status, and physical activity. We verified the proportional hazards assumption of the Cox model by using the Schoenfeld residuals technique.[22] Absolute risk was calculated as the percentage of COVID-19 cases occurring per 10,000 person-months in a given group. We used 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.

In secondary analyses, we used a self-report of a positive test to define COVID-19 risk. For these analyses, we used inverse probability-weighted Cox models to account for predictors of obtaining country-specific testing. Inverse probability-weighted analyses included presence of COVID-19-related symptoms, interaction with a person with COVID-19, occupation as a healthcare worker, age group, and race. Inverse probability-weighted Cox models were stratified by 10-year age group and date with additional adjustment for the covariates used in previous models. For severe COVID-19 analyses, we adjusted for the same covariates used in previous models. As an additional method to quantify diet quality we used the DQS and tested for associations between diet quality and COVID-19 risk and severity. In addition, we censored our analyses to cases that occurred after completing the diet survey to investigate potential bias due to time-varying confounding.

In subgroup analyses, we assessed the association between diet quality and COVID-19 risk according to comorbidities, demographic, and lifestyle characteristics. We also classified participants according to categories of the diet quality score and socioeconomic deprivation (nine categories based on thirds of diet quality score and deprivation index) and conducted joint analyses for COVID-19 risk. We tested 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 (online supplementary methods).[23]

We conducted sensitivity analyses to account for regional differences in the effective reproductive number (R_t) or other risk mitigating behaviors such as mask wearing. Details on how we obtained and classified individuals for these analyses are provided in online supplementary methods. Two-sided p-values of <0.05 were considered statistically significant for main analyses. All statistical analyses were performed using R software, version 4.0.3 (R Foundation).

Results

Self-reported diet quality was evaluated in 647,137 survey responders, of which 54,566 were excluded due to prevalent COVID (n=1,555), presence of any symptoms at baseline (n=47,594), logged only once (n=1,201), pregnancy (n=1,129), or age under 18 year (n=3,087; online supplementary figure 1). Baseline characteristics of the 592,571 participants included in this study according to categories of the hPDI score are shown in table 1. Participants in the highest quartile of the diet score (reflecting a healthier diet) were more likely than participants in the lowest quartile to be older, female, healthcare workers, of lower BMI, engage in physical activities ≥ 5 days/week, and less likely to reside in areas with higher socioeconomic deprivation. The hPDI score was normally distributed (online supplementary figure 2).

Over 3,886,274 person-months of follow-up, 31,815 COVID-19 cases were documented. Crude COVID-19 rates per 10,000 person-months were 72.0 (95% CI, 70.4-73.7) for participants in the highest quartile of the diet score and 104.1 (95% CI, 101.9-106.2) for those in the lowest quartile. The corresponding age-adjusted HR for COVID-19 risk was 0.80 (95% CI, 0.78-0.83, table 2). Differences in the risk of COVID-19 persisted after adjustment for potential confounders. In fully adjusted models, the multivariable-adjusted HR for COVID-19 risk was 0.91 (95% CI, 0.88-0.94) when we compared participants with high diet quality to those with low diet quality. We observed non-linear decreasing trends in the risk of COVID-19 with higher diet quality ($P < 0.001$ for non-linearity), in which COVID-19 risk plateau among individuals with a diet quality score > 50 (online supplementary figure 3). The association between diet quality and COVID-19 risk was consistent but attenuated in secondary analyses using the DQS score (HR, 0.92; 95% CI, 0.89-0.95; online supplementary table 5), and became non-significant in fully adjusted models (HR, 1.00; 95% CI, 0.97-1.03). We also investigated whether our primary findings were consistent in an analysis censored to cases that occurred after the completion of the diet survey. These analyses showed that high diet quality, compared to low diet quality, was associated with lower COVID-19 risk (multivariable-adjusted HR 0.88; 95% CI, 0.83-0.93; online supplementary table 6).

In secondary analyses for COVID-19 risk based on a positive test, we showed that crude COVID-19 incidence rates per 10,000 person-months were 12.9 (95% CI 12.2-13.6) for individuals with high diet quality and 16.4 (95% CI 15.5-17.2) for individuals with low diet quality. The corresponding multivariable-adjusted HR for risk of COVID-19 was 0.82 (95% CI, 0.78-0.86; table 2). For risk of severe COVID-19, crude incidence rates were lower for individuals reporting high diet quality compared to those with low diet quality (1.6 (95% CI, 1.3-1.8) vs. 2.1 (95% CI, 1.9-2.5; per 10,000 person-months) table 2). In the fully adjusted model, high diet quality, as compared to low diet quality, was associated lower risk of severe COVID-19 with an a HR of 0.59 (95% CI, 0.47-0.74; table 2).

In stratified analyses, the inverse association between diet quality and COVID-19 risk was more evident in participants living in areas of high socioeconomic deprivation and those reporting low physical activity levels ($P < 0.05$; table 3). We found no significant effect modification for other characteristics such age, BMI, race/ethnicity or population density. When diet quality and socioeconomic deprivation were combined, there was a risk gradient with low diet quality and high socioeconomic deprivation. Compared with individuals living in areas with low socioeconomic deprivation and high diet quality, the multivariable-adjusted HR for risk of COVID-19 for low diet quality was 1.08 (95% CI, 1.03-1.14) among those living in areas with low socioeconomic deprivation, 1.23 (95% CI, 1.17-1.29) for those living in areas with intermediate socioeconomic deprivation, and 1.47 (95% CI, 1.38-1.52) for those living in areas with high socioeconomic deprivation (figure 1). The joint associations of diet quality and socioeconomic deprivation was higher than the sum of the risk associated with each factor alone (relative excess risk due to interaction (RERI) = 0.05 (95% CI 0.02-0.08); $P_{interaction}$=0.005; online supplementary table 7). The proportion of contribution to excess COVID-19 risk was estimated to be 31.9% (95% CI, 18.2-45.6) to diet quality, 38.4% (95% CI, 26.5-50.3) to socioeconomic deprivation, and 29.7% (95% CI, 2.1-57.3) to their interaction. The absolute excess rate of COVID-19 per 10,000 person-months for lowest vs highest quartile of the diet score was 22.5 (95% CI, 18.8-26.3) among individuals living in areas with low socioeconomic deprivation and 40.8 (95% CI, 31.7-49.8) among individuals living in areas with high deprivation (online supplementary figure 4)

We conducted a series of sensitivity analyses to further account for variation in R_t or mask wearing. For peak R_t censored analyses, crude COVID-19 rates per 10,000 person-months were 148.1 (95% CI, 139.9-156.8) among participants with low diet quality and 92.9 (95% CI, 86.6-99.5) for participants with high diet quality. The corresponding multivariable-adjusted HR was 0.84 (95% CI, 0.76-0.92, figure 2). The same trend was observed for nadir R_t censored analyses, in which crude COVID-19 rates per 10,000 person-months were 67.1 (95% CI, 61.7-73.0) among participants with low diet quality and 45.8 (95% CI, 41.3-50.5) for participants with high diet quality (multivariable-adjusted HR, 0.89; 95% CI, 0.80-1.00, figure 2). We further adjusted our models for mask wearing. This analysis showed that high diet quality, as compared to low diet quality, was associated with lower risk of COVID-19 with an adjusted HR of 0.88 (95% CI, 0.83-0.94; online supplementary table 8).

Discussion

In this large survey among UK and US participants prospectively assessing risk and severity of COVID-19 infection, we found that a dietary patterns characterized by healthy plant foods was associated with lower risk and severity of COVID-19. We observed a risk gradient of poor diet quality and increased socioeconomic deprivation that departed from the additivity of the risks attributable to each factor separately, suggesting that the beneficial association of diet with COVID-19 may be particularly evident among individuals with higher socioeconomic deprivation.

Our findings are aligned with preliminary evidence showing that improving nutrition could help reduce the burden of infectious diseases.[12,14,24] Early studies have shown that the administration of arachidonic or linoleic acid partially suppresses SARS-CoV-1 and coronavirus 229E viral replication,[25] and that specific nutrients or dietary supplements associate with modest reductions in COVID-19 risk.[26] Results from this observational study could expand previous single nutrient observations and highlight the beneficial association of healthy dietary patterns, which was most pronounced for risk of severe COVID-19. Our findings also concur with a comparative risk assessment study suggesting that a 10% reduction in the prevalence of diet-related conditions such as obesity and type 2 diabetes would have prevented ~11% of the COVID-19 hospitalizations that have occurred among US adults since November 2020.[27]

The association of healthy diet with lower COVID-19 risk appears particularly evident among individuals living in areas of higher socioeconomic deprivation. Our models estimate that nearly a third of COVID-19 cases would have been prevented if one of two exposures (diet and deprivation) were not present. Although these estimated attributable risks should be interpreted in the context of the population-specific prevalence, and are likely to change over time with the prevailing SARS-CoV-2 infection rate, our observations are consistent with data from ecological studies showing that people living in regions with greater social inequalities are likely to have higher rates of COVID-19 incidence and deaths.[28] By generating a granular deprivation index based on zip code information our study adds to previous country-level ecological studies. In addition, recent studies on the impact of socioeconomic status on COVID-19 have shown that community-level deprivation indices are strongly associated with COVID-19 risk and mortality.[29,30] However, it is still possible that differences in deprivation exists within communities. Further studies including information about household characteristics, built environment, or access to healthy foods are needed to expand these initial associations.

Our study adds to knowledge by formally investigating how diet quality, in the context of distal social determinants of health, associates with risk and severity of COVID-19. While our study supports the beneficial association of diet quality with COVID-19 risk and severity, particularly among individuals with higher deprivation, we cannot completely rule out the potential for residual confounding. Individuals who eat healthier diets are likely to share other features that might be associated with lower risk of infection such as the adoption of other risk mitigation behaviors, better household conditions and hygiene, or access to care. However, it is reassuring that our findings were consistent despite controlling for additional surrogate markers of SARS-CoV-2 infection such as mask wearing or community transmission rate, two of the most relevant factors associated with virus transmission and COVID-19 risk.[31] These findings suggest that efforts to address disparities in COVID-19 risk and severity should consider specific attention to access to healthy foods as a social determinants of health.

We acknowledge several limitations. First, as an observational study, we are unable to confirm a direct causal association between diet and COVID-risk or infer specific mechanisms. Second, our study population was not a random sampling of the population. Although this limitation is inherent to any study requiring voluntary provision of information, we recognize our participants are mainly white participants and less likely to live in low deprived areas and are less ethnically diverse than the general population.[19] Thus, generalizability of our finding even to the wider British and American population is uncertain. Third, our results could be biased due to the time lapse between the dietary recalls, administered a few months after the relevant period of exposure (pre-pandemic). However, our sensitivity analyses in which we censored cases that had occurred before the administration of the diet survey showed consistent results. Fourth, the self-reported nature of the diet questionnaire is prone to measurement error and bias, and the use of a short food frequency survey could have further reduced the resolution of dietary data collected. More accurate dietary intake assessment methods such as the use of dietary intake biomarkers would be valuable in future studies,[32] but also difficult to implement in large-scale and time-sensitive investigations. Fifth, we defined risk of severe COVID-19 according to reports of hospitalization with oxygen support, which may not have captured more severe or fatal cases.

Conclusions

In conclusion, our data provide evidence that a healthy diet was associated with lower risk of COVID-19 and severe COVID-19 even after accounting for other healthy behaviors, social determinants of health, and virus transmission measures. The joint association of diet quality with socioeconomic deprivation was greater than the addition of the risks associated with each individual factor, suggesting that diet quality may play a direct influence in COVID-19 susceptibility and progression. Our findings suggest that public health interventions to improve nutrition and poor metabolic health and address social determinants of health may be important for reducing the burden of the pandemic.

Acknowledgments:

We express our sincere thanks to all of the participants who entered data into the app, including study volunteers enrolled in cohorts within the Coronavirus Pandemic Epidemiology (COPE) consortium. We thank the staff of Zoe Global, the Department of Twin Research at King's College London and the Clinical and Translational Epidemiology Unit at Massachusetts General Hospital for tireless work in contributing to the running of the study and data collection. This work was conducted using the Short Form FFQ tool developed by Cleghorn as reported in https://doi.org/10.1017/S1368980016001099 and listed in the Nutritools (www.nutritools.org) library.

Conflict of Interest:

JW, CH, SS, and JC are employees of Zoe Global Ltd. TDS, ERL, SEB, area consultant to Zoe Global Ltd. DAD, JM, and ATC previously served as investigators on a clinical trial of diet and lifestyle using a separate mobile application that was supported by Zoe Global Ltd. Other authors have no conflict of interest to declare.

Contributions:

JM, ADJ, LHN, ERL, TDS, SEB, and ATC conceived the study design. JM, ADJ, ERL, MSG, JC, BM, SS contributed to the statistical analysis. All authors were involved in acquisition, analysis, or interpretation of data. JM, LHN, and DAD wrote the first draft of the manuscript. DAD, WCW, SO, CJS, JW, PWF, TDS, SEB, ATC obtained funding. JM, ADJ, LHN provided administrative, technical, or material support. TDS, SEB, ATC jointly supervised this work. All authors contributed to the critical revision of the manuscript for important intellectual content and approved the final version of the manuscript. The corresponding authors attest that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Data availability statement:

Zoe Platform data used in this study is available to researchers through UK Health Data Research using the following link: https://web.www.healthdatagateway.org/dataset/fddcb382-3051-4394-8436-b92295f14259. The diet quality data used for this study are held by the department of Twin Research at Kings' College London. The data can be released to bona fide researchers using our normal procedures overseen by the Wellcome Trust and its guidelines as part of our core funding. We receive around 100 requests per year for our datasets and have a meeting three times a month with independent members to assess proposals. Application is via https://twinsuk.ac.uk/resources-for-researchers/access-our-data/. This means that the data needs to be anonymized and conform to GDPR standards.

Grant Support:

National Institutes of Health, American Gastroenterological Association, Massachusetts Consortium on Pathogen Readiness, National Institute for Health Research, Crohn's and Colitis Foundation, Chronic Disease Research Foundation, UK Medical Research Council, Wellcome Trust, UK Research and Innovation. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

References

1 Cummings MJ, Baldwin MR, Abrams D, *et al*. Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study. *Lancet* 2020;395:1763–70. doi:10.1016/S0140-6736(20)31189-2

2 The Lancet Diabetes & Endocrinology. Metabolic health: a priority for the post-pandemic era. *Lancet Diabetes Endocrinol* Published Online First: March 2021. doi:10.1016/S2213-8587(21)00058-9

3 Singh S, Khan A. Clinical Characteristics and Outcomes of Coronavirus Disease 2019 Among Patients With Preexisting Liver Disease in the United States: A Multicenter Research Network Study. *Gastroenterology* 2020;159:768-771.e3. doi:10.1053/j.gastro.2020.04.064

4 Leong A, Cole JB, Brenner LN, *et al*. Cardiometabolic risk factors for COVID-19 susceptibility and severity: A Mendelian randomization analysis. *PLOS Med* 2021;18:e1003553. doi:10.1371/journal.pmed.1003553

5 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;70:411–22. doi:10.1016/j.jacc.2017.05.047

6 Chiuve SE, Fung TT, Rimm EB, *et al*. Alternative dietary indices both strongly predict risk of chronic disease. *J Nutr* 2012;142:1009–18. doi:10.3945/jn.111.157222

7 Ley SH, Hamdy O, Mohan V, *et al*. Prevention and management of type 2 diabetes: dietary components and nutritional strategies. *Lancet* 2014;383:1999–2007. doi:10.1016/S0140-6736(14)60613-9

8 Willett WC, Stampfer MJ. Current evidence on healthy eating. *Annu Rev Public Health* 2013;34:77–95. doi:10.1146/annurev-publhealth-031811-124646

9 Satija A, Bhupathiraju SN, Rimm EB, *et al*. Plant-Based Dietary Patterns and Incidence of Type 2 Diabetes in US Men and Women: Results from Three Prospective Cohort Studies. *PLOS Med* 2016;13:e1002039. doi:10.1371/journal.pmed.1002039

10 Mazidi M, Kengne AP. Higher adherence to plant-based diets are associated with lower likelihood of fatty liver. *Clin Nutr* 2019;38:1672–7. doi:10.1016/j.clnu.2018.08.010

11 Marmot MG, Stansfeld S, Patel C, *et al*. Health inequalities among British civil servants: the Whitehall II study. *Lancet* 1991;337:1387–93. doi:10.1016/0140-6736(91)93068-K

12 Belanger MJ, Hill MA, Angelidi AM, *et al*. Covid-19 and Disparities in Nutrition and Obesity. *N Engl J Med* 2020;383:e69. doi:10.1056/nejmp2021264

13 Rehm CD, Peñalvo JL, Afshin A, *et al*. Dietary Intake Among US Adults, 1999-2012. *JAMA* 2016;315:2542. doi:10.1001/jama.2016.7491

14 Storm I, Den Hertog F, Van Oers H, *et al*. How to improve collaboration between the public health sector and other policy sectors to reduce health inequalities? - A study in sixteen municipalities in the Netherlands. *Int J Equity Health* 2016;15. doi:10.1186/s12939-016-0384-y

15 Drew DA, Nguyen LH, Steves CJ, *et al*. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science* 2020;368:1362–7. doi:10.1126/science.abc0473

16 Mazidii 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]. February 09, 2021 [cited 2021 March 25].

17 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;19:2915–23. doi:10.1017/S1368980016001099

18 Menni C, Valdes AM, Freidin MB, *et al*,. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med* 2020;26:1037–40. doi:10.1038/s41591-020-0916-2

19 Varsavsky T, Graham MS, Canas LS, *et al*. Detecting COVID-19 infection hotspots in England using large-scale self-reported data from a mobile application: a prospective, observational study. *Lancet Public Heal* 2021;6:e21–9. doi:10.1016/S2468-2667(20)30269-3

20 Botti-Lodovico Y, Rosenberg E, Sabeti PC. Testing in a Pandemic — Improving Access, Coordination, and Prioritization. *N Engl J Med* 2021;384:197–9. doi:10.1056/nejmp2025173

21 Toutenburg H. Rubin, D.B.: Multiple imputation for nonresponse in surveys. *Stat Pap* 1990;31:180–180. doi:10.1007/bf02924688

22 Schoenfeld D. Partial Residuals for The Proportional Hazards Regression Model. *Biometrika* 1982;69:239. doi:10.2307/2335876

23 VanderWeele TJ, Tchetgen Tchetgen EJ. Attributing effects to interactions. *Epidemiology* 2014;25:711–22. doi:10.1097/EDE.0000000000000096

24 Calder PC. Nutrition, immunity and COVID-19. *BMJ Nutr Prev Heal* 2020;**3**:74–92. doi:10.1136/bmjnph-2020-000085

25 Yan B, Chu H, Yang D, *et al*. Characterization of the lipidomic profile of human

coronavirus-infected cells: Implications for lipid metabolism remodeling upon coronavirus replication. *Viruses* 2019;11. doi:10.3390/v11010073

26 Louca P, Murray B AM, Klaser MK et al. Modest effects of dietary supplements during the COVID-19 pandemic: insights from 445,850 users of the COVID Symptom Study app. *BMJ Nutr Prev Health* 2021. In Pres

27 O'Hearn M, Liu J, Cudhea F, *et al*. Coronavirus Disease 2019 Hospitalizations Attributable to Cardiometabolic Conditions in the United States: A Comparative Risk Assessment Analysis. *J Am Heart Assoc* 2021;10:e019259. doi:10.1161/JAHA.120.019259

28 Wise J. Covid-19: Highest death rates seen in countries with most overweight populations. *BMJ* 2021;372:n623. doi:10.1136/bmj.n623

29 Mena GE, Martinez PP, Mahmud AS, *et al*. Socioeconomic status determines COVID-19 incidence and related mortality in Santiago, Chile. *Science* 2021;372:eabg5298. doi:10.1126/science.abg5298

30 Feldman JM, Bassett MT. The relationship between neighborhood poverty and COVID-19 mortality within racial/ethnic groups (Cook County, Illinois). *medRxiv* 2020;2020.10.04.20206318. doi:10.1101/2020.10.04.20206318

31 Brooks JT, Butler JC, Redfield RR. Universal Masking to Prevent SARS-CoV-2 Transmission - The Time Is Now. JAMA. 2020;324:635–7. doi:10.1001/jama.2020.13107

32 Savolainen O, Lind MV, Bergström G, *et al*. Biomarkers of food intake and nutrient status are associated with glucose tolerance status and development of type 2 diabetes in older Swedish women. *Am J Clin Nutr* 2017;106:1302–10. doi:10.3945/ajcn.117.152850

Figure legends

Figure 1.

Title: Risk of COVID-19 according to diet quality and socioeconomic deprivation.

Figure legend: Shown are adjusted hazard ratios and 95% confidence interval of the estimate for predicted COVID-19 according to categories of diet quality and socioeconomic deprivation. Cox model stratified by calendar date at study entry, country of origin, and 10-year age group, and adjusted for sex, race/ethnicity, index of multiple deprivation, population density, presence of diabetes, cardiovascular disease, lung disease, cancer, kidney disease, healthcare worker status, body mass index, smoking status, and physical activity. In these comparisons, participants with high-quality diet and low socioeconomic deprivation served as the reference group.

Figure 2.

Title: Risk of COVID-19 according to community transmission rate and diet quality

Figure legend: COVID-19 incidence rate per 10,000 person-month and 95% confidence interval of the estimate based on different community transmission rate and diet quality categories. Peak R_t and nadir R_t were defined using (methods). Adjusted hazard ratios and 95% confidence interval of the estimate for risk of COVID-19 were obtained from fully adjusted Cox models.

496 **TABLES and FIGURES**

497 **Table 1.** Baseline characteristics of study participants according to categories of the diet quality
498 score

	All participants (n=592,571)	Low hPDI (Q1; n=148,143)	Intermediate hPDI (Q2-Q3; n=296,286)	High hPDI (Q4; n=148,142)
hPDI score, median [IQR]	50 (47 to 54)	45 (43-47)	51 (49-52)	56 (55-58)
Demographic characteristics				
Age, years	56 (44-65)	52 (41-62)	57 (45-66)	57 (45-65)
≥18-24	14,397 (2.4)	5,146 (3.4)	5,846 (2.0)	3,405 (2.3)
25-34	52,922 (8.9)	16,535 (11.2)	23,150 (7.8)	13,237 (8.9)
35-44	86,251 (14.6)	26,907 (18.2)	40,145 (13.5)	19,199 (13.0)
45-54	125,802 (21.2)	34,890 (23.6)	62,491 (21.1)	28,421 (19.2)
55-64	158,637 (26.8)	34,279 (23.1)	81,837 (27.6)	42,521 (28.7)
≥65	153,810 (26.0)	30,215 (20.4)	82,413 (27.8)	41,182 (27.8)
Missing	752 (0.1)	171 (0.1)	404 (0.1)	177 (0.1)
Sex, No. (%)				
Male	187,450 (31.6)	58,199 (39.3)	93,162 (31.4)	36,089 (24.4)
Female	404,126 (68.2)	89,706 (60.5)	202,605 (68.4)	111,815 (75.5)
Prefer not to say	995 (0.2)	238 (0.2)	519 (0.2)	238 (0.2)
Race[ε], No. (%),				
White	568,770 (96.0)	141,365 (95.4)	284,804 (96.1)	142,601 (96.3)
Black	4,328 (0.7)	1,466 (1.0)	2,053 (0.7)	809 (0.5)
Asian	10,435 (1.8)	2,954 (1.9)	5,043 (1.7)	2,438 (1.6)
Other	7,228 (1.2)	1,925 (1.3)	3,463 (1.2)	1,840 (1.2)
Missing	1,810 (0.3)	433 (0.3)	923 (0.3)	454 (0.3)
Country, No. (%)				
UK	543,984 (91.8)	135,360 (91.4)	272,494 (92.0)	136,130 (91.9)
US	48,587 (8.2)	12,783 (8.6)	23,792 (8.0)	12,012 (8.1)
Index of deprivation, No. (%)[¶]				
Most deprived, decile 1	1,3416 (2.3)	4,696 (3.1)	6,163 (2.1)	2,557 (1.7)
Least deprived, decile 10	103,608 (17.5)	23,122 (15.6)	53,652 (18.1)	26,834 (18.1)
Missing	40,759 (6.9)	10,489 (7.1)	20,249 (6.8)	10,021 (6.8)
Population density, km^2, No. (%)[¶]				
<500	119,782 (20.2)	28,139 (19.0)	61,230 (20.7)	30,413 (20.5)
500-1,999	90,541 (15.3)	23,631 (16.0)	45,902 (15.5)	21,008 (14.2)
2,000 4,999	94,345 (15.9)	24,813 (16.7)	47,233 (15.9)	22,299 (15.1)
≥5,000	244,295 (41.2)	60,156 (40.6)	120,319 (40.6)	63,820 (43.1)
Missing	43,608 (7.4)	11,404 (7.7)	21,602 (7.3)	10,602 (7.2)
Healthcare worker, yes, No. (%)	41,141 (6.9)	10,633 (2.3)	20,183 (6.8)	10,325 (7.0)
Lifestyle characteristics				

Smoking status, No (%)				
Never	475,347 (81.9)	113,165 (79.0)	238,192 (81.9)	123,990 (84.6)
Former	87,901 (15.1)	22,683 (15.8)	44,771 (15.4)	20,447 (13.9)
Current	17,401 (3.0)	7,402 (5.2)	7,837 (2.6)	2,162 (1.5)
Physical activity				
< 1 day/week	106,294 (17.9)	37,258 (25.2)	50,713 (17.1)	18,323 (12.4)
1-2 days/week	224,606 (37.9)	59,325 (40.0)	113,749 (38.4)	51,532 (34.8)
3-4 days/week	143,548 (24.2)	30,009 (20.3)	73,601 (24.8)	39,938 (27.0)
≥ 5 days/week	117,007 (19.7)	21,164 (14.3)	57,701 (19.5)	38,142 (25.7)
Missing	1,116 (0.2)	387 (0.3)	522 (0.2)	207 (0.1)
Body mass index, Kg/m^2	25.1 (22.6-28.7)	26.6 (23.6-30.7)	25.2 (22.7-28.5	24.0 (21.8-26.9)
<18.5	12,004 (2.0)	2,680 (1.8)	5,540 (1.9)	3,784 (2.6)
18.5-24.9	277,536 (46.8)	52,109 (35.2)	138,503 (46.7)	86,924 (58.7)
25-29.9	189,197 (31.9)	51,517 (34.8)	97,919 (33.0)	39,761 (26.8)
≥30	113,056 (19.1)	41,655 (28.1)	53,909 (18.2)	17,492 (11.8)
Missing	778 (0.1)	182 (0.1)	415 (0.1)	181 (0.1)
Mask wearing, No (%)[†]				
Most of the time / always	437,782 (73.9)	113,202 (76.4)	218,402 (73.7)	106,178 (71.6)
Never / sometimes	152,551 (25.7)	34,240 (23.1)	76,809 (25.9)	41,502 (28.0)
Missing	2,238 (0.4)	701 (0.5)	1,075 (0.4)	462 (0.3)
Clinical history, yes, No. (%)				
Diabetes	20,058 (3.4)	6,079 (4.1)	10,158 (3.4)	3,821 (2.6)
Heart disease	20,376 (3.4)	5,200 (3.5)	10,660 (3.6)	4,516 (3.0)
Cancer	6,559 (1.9)	1,643 (1.8)	3,348 (1.9)	1,568 (1.8)
Lung disease	62,999 (10.6)	17,534 (11.8)	31,227 (10.5)	14,238 (9.6)
Kidney disease	5,134 (0.9)	1,492 (1.0)	2,594 (0.9)	1,048 (0.7)

Table Legend: Values are median (IQR) 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 zipcode or postcode information linked with census track data. Country-specific deprivation indices were generated (supplement).

[†] Mask wearing information was collapsed into two categories: participants who wore masks 'none of the time or sometimes', and those who reported wearing masks 'most of time/ always'.

[*] hPDI ranges from 0 to 70, with higher scores indicating higher adherence to a healthy plant-based diet.

Table 2. Adjusted hazard ratios of COVID-19 risk and severity according to healthful plant-based dietary index scores.

	Low hPDI (Q1; n=148,143)	**Intermediate hPDI** (Q2-Q3; n=296,286)	**High hPDI** (Q4; n=148,142)	**P for trend**
hPDI score, median (IQR)	45 (43-47)	51 (49-52)	56 (55-58)	
COVID-19 risk				
No. of events/person-months	8,739 / 839,747	15,733 / 2,026,824	7,359 / 1,022,078	—
Incidence rate (10,000 person-months; 95% CI)	104.1 (101.9-106.2)	77.6 (76.4-78.8)	72.0 (70.4-73.7)	—
Age-adjusted model	1.00 (Ref)	0.85 (0.82-0.87)	0.80 (0.78-0.83)	<0.001
Multivariable model 2	1.00 (Ref)	0.85 (0.83-0.87)	0.81 (0.78-0.83)	<0.001
Multivariable model 3	1.00 (Ref)	0.91 (0.89-0.93)	0.91 (0.88-0.94)	<0.001
COVID-19 risk (positive test)				
No. of events/person-months	1,423 / 869,664	2,829 / 2,081,970	1,350 / 1,046,887	—
Incidence rate (10,000 person-months; 95% CI)	16.4 (15.5-17.2)	13.6 (13.1-14.1)	12.9 (12.2-13.6)	—
Age-adjusted model[$]	1.00 (Ref)	0.86 (0.83-0.90)	0.79 (0.75-0.83)	<0.001
Multivariable model 2[$]	1.00 (Ref)	0.87 (0.84-0.91)	0.80 (0.76-0.84)	<0.001
Multivariable model 3[$]	1.00 (Ref)	0.88 (0.85-0.92)	0.82 (0.78-0.86)	<0.001
Severe COVID-19				
No. of events/person-months	187 / 871,995	390 / 2,086,790	163 / 1,049,476	—
Incidence rate (10,000 person-months; 95% CI)	2.1 (1.9-2.5)	1.9 (1.7-2.1)	1.6 (1.3-1.8)	—
Age-adjusted model	1.00 (Ref)	0.66 (0.56-0.77)	0.45 (0.36-0.57)	<0.001
Multivariable model 2	1.00 (Ref)	0.66 (0.57-0.78)	0.45 (0.36-0.57)	<0.001
Multivariable model 3	1.00 (Ref)	0.77 (0.66-0.91)	0.59 (0.47-0.74)	<0.001

Table Legend: Hazards ratios and 95% CI for COVID-19 risk and severity. COVID-19 risk defined using a validated symptom-based model. COVID-19 or a RT-PCR positive test report. COVID-19 severity was defined based on hospitalization with requirement of oxygen support (methods, supplement).

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.

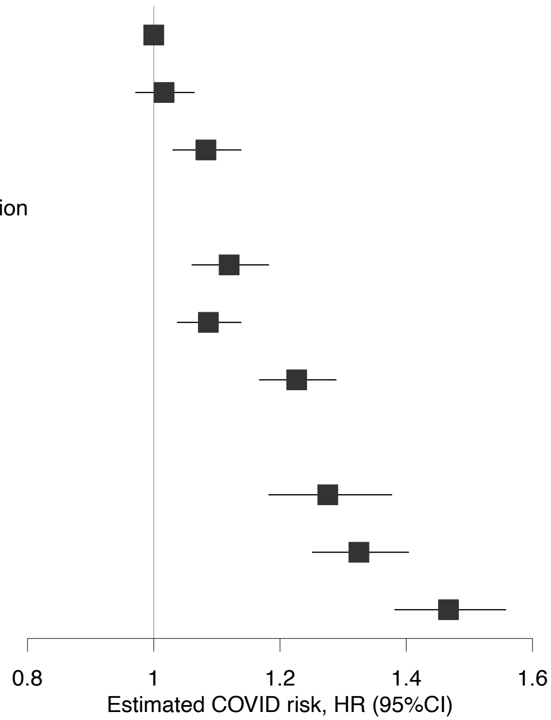
Table 3. Adjusted hazard ratios of COVID risk according to healthful plant-based dietary index scores stratified by sociodemographic and clinical characteristics.

Factor	No. of events/ person-months#	HR per 1-SD increase in diet quality score	P value
Age,			
<60	25,329 / 2,285,329	0.94 (0.93-0.95)	
≥60	6,486 / 1,600,945	0.94 (0.92-0.97)	0.63
Sex,			
Male	9,338 / 1,232,656	0 . 95 (0.93-0.97)	
Female	22,428 / 2,647,254	0 . 96 (0.95-0.98)	0.21
Race,			
White	30,335 / 3,736,972	0 . 96 (0.95-0.97)	
Non-white	1,480 / 149,303	0 . 96 (0.91-1.01)	0.95
Socioeconomic deprivation*			
High	5,244 / 45,6271	0.94 (0.91-0.96)	
Intermediate	13,172 / 1,567,516	0.96 (0.94-0.98)	
Low	13,399 / 1,862,489	0.97 (0.95-0.99)	0.04
Population density, km^2, No.			
<2000	10,581 / 1,490,084	0.96 (0.94-0.98)	
≥2,000	21,234 / 2,396,190	0.96 (0.94-0.97)	0.74
Healthcare worker			
Yes	2,908 / 140,087	0.95 (0.92-0.99)	
No	28,907 / 3,638,588	0.96 (0.95-0.97)	0.78
Body mass index, Kg/m^2			
<25	13,989 / 1,905,517	0.96 (0.94-0.97)	
25-30	9,854 / 1,252,222	0.96 (0.94-0.98)	
≥30	7,972 / 728,536	0.96 (0.94-0.98)	0.73
Physical activity			
< 1 day/week	6,751 / 683,562	0.94 (0.91-0.96)	
1-4 day/week	19,476 / 2,425,198	0.96 (0.94-0.97)	
≥ 5 day/week	5,588 / 777,515	0.99 (0.96-1.01)	0.01

Table Legend: Association between predicted COVID-19 and diet quality according to sociodemographic and clinical characteristics. *Socioeconomic deprivation categories were based on deciles of the deprivation index (methods). Cox models were adjusted for the same covariates as previous model 3. P-values obtained using the Q test for heterogeneity

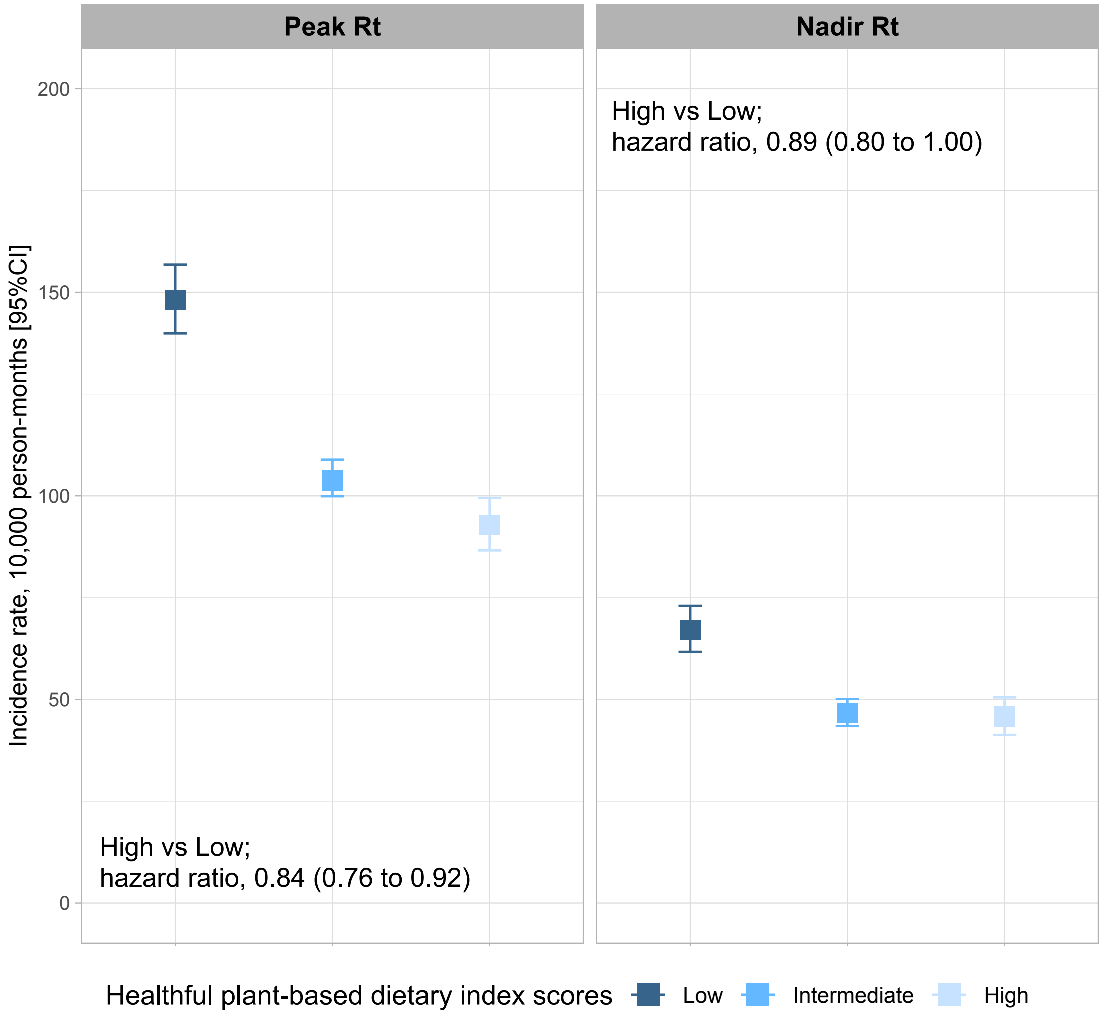
Number of observations varies among imputations.

Figure 1



Subgroup	Total No. of participants	COVID−19 cases Person−Month	Adjusted HR [95% CI]	P Value
Low Socioeconomic Deprivation				
High hPDI	66,714	2,916 / 446,623	1.00 [Reference]	
Intermediate hPDI	133,557	6,139 / 892,602	1.02 [0.97−1.06]	0.481
Low hPDI	60,759	3,499 / 398,407	1.08 [1.03−1.14]	0.002
Intermediate Socioeconomic Deprivation				
High hPDI	56,354	2,905 / 372,964	1.12 [1.06−1.18]	<0.001
Intermediate hPDI	110,890	5,660 / 731,895	1.09 [1.04−1.15]	<0.001
Low hPDI	56,143	3,810 / 359,896	1.23 [1.17−1.29]	<0.001
High Socioeconomic Deprivation				
High hPDI	15,053	979 / 97,708	1.28 [1.18−1.37]	<0.001
Intermediate hPDI	31,590	2,152 / 202,751	1.33 [1.25−1.40]	<0.001
Low hPDI	20,752	1,787 / 126,752	1.47 [1.38−1.56]	<0.001

Figure 2


Peak Rt
Nadir Rt
200
150
100
50
0
Incidence rate, 10,000 person-months [95%CI]
High vs Low;
hazard ratio, 0.84 (0.76 to 0.92)
High vs Low;
hazard ratio, 0.89 (0.80 to 1.00)
Healthful plant-based dietary index scores
Low
Intermediate
High