

1 **Genetically predicted serum vitamin D and COVID-19: a Mendelian randomization study**

2 Bonnie K Patchen¹, Andrew G Clark², Dana B Hancock³, Nathan Gaddis³,
3 Patricia A Cassano^{1,4}

4 ¹ Division of Nutritional Sciences, Cornell University, Ithaca NY, United States 14853

5 ² Department of Molecular Biology and Genetics, Cornell University, Ithaca NY, United States
6 14853

7 ³ GenOmics, Bioinformatics and Translational Research Center, Research Triangle Institute
8 International, Research Triangle Park, NC, United States 27709-2194

9 ⁴ Weill Cornell Medicine, Department of Population Health Sciences, Division of Epidemiology,
10 NY, NY, United States 10065

11 Correspondence to: P.A. Cassano pac6@cornell.edu

12 January 2021

13

14 **ABSTRACT**

15 **Objectives:** To investigate causality of the association of serum vitamin D with the risk and
16 severity of COVID-19 infection.

17
18 **Design:** Two-sample Mendelian randomization study.

19
20 **Setting:** Summary data from genome-wide analyses in the population-based UK Biobank and
21 SUNLIGHT Consortium, applied to meta-analyzed results of genome-wide analyses in the
22 COVID-19 Host Genetics Initiative.

23
24 **Participants:** 17,965 COVID-19 cases including 11,085 laboratory or physician confirmed
25 cases, 7,885 hospitalized cases, and 4,336 severe respiratory cases, and 1,370,547 controls,
26 primarily of European ancestry.

27
28 **Exposures:** Genetically predicted variation in serum vitamin D status, based on genome-wide
29 significant single nucleotide polymorphisms (SNPs) associated with serum vitamin D or risk of
30 vitamin D deficiency/insufficiency.

31
32 **Main outcome measures:** Susceptibility to and severity of COVID-19 infection, including
33 severe respiratory infection and hospitalization.

34
35 **Results:** Mendelian randomization analysis, powered to detect moderate effects comparable to
36 those seen in observational studies, provided little to no evidence for an effect of genetically
37 predicted serum vitamin D on susceptibility to or severity of COVID-19 infection. Using SNPs
38 in loci related to vitamin D metabolism as proxies for serum vitamin D concentration, the odds
39 ratio for a standard deviation increase in serum vitamin D was 1.04 (95% confidence interval
40 0.92 to 1.18) for any COVID-19 infection versus population controls, 1.05 (0.84-1.31) for
41 hospitalized COVID-19 versus population controls, 0.96 (0.64 to 1.43) for severe respiratory
42 COVID-19 versus population controls, 1.15 (0.99 to 1.35) for COVID-19 positive versus
43 COVID-19 negative, and 1.44 (0.75 to 2.78) for hospitalized COVID-19 versus non-hospitalized
44 COVID-19. Results were similar in analyses that used all SNPs with genome-wide significant
45 associations with serum vitamin D (i.e., including SNPs in loci with no known relationship to
46 vitamin D metabolism) and in analyses using SNPs with genome-wide significant associations
47 with risk of vitamin D deficiency or insufficiency.

48
49 **Conclusions:** These findings suggest that genetically predicted differences in long-term vitamin
50 D nutritional status do not causally affect susceptibility to and severity of COVID-19 infection,
51 and that associations observed in previous studies may have been driven by confounding. These
52 results do not exclude the possibility of low-magnitude causal effects, nor do they preclude
53 potential causal effects of acute responses to therapeutic doses of vitamin D. Future directions
54 include extension of this work to non-European ancestry populations, and high-risk populations,
55 for example persons with comorbid disease.

56

57 INTRODUCTION

58 The COVID-19 pandemic has reached every corner of the globe and continues to spread.
59 With more than 100 million cases and 2.0 million deaths globally at time of writing,¹
60 identification of risk factors for susceptibility to SARS-CoV-2 infection and severity of COVID-
61 19 is critical. Vitamin D nutritional status is a promising modifiable risk factor, and higher
62 vitamin D is posited to reduce the risk of SARS-CoV-2 infection and the severity of the clinical
63 course of COVID-19. Hypothesized vitamin D effects are biologically plausible given prior
64 evidence that vitamin D upregulates innate and adaptive immunity to fight infection and reduce
65 inflammation,² is associated with a reduced risk of respiratory disease mortality,³ and enhances
66 expression of ACE2, which is hypothesized to modulate the immune system response to SARS-
67 CoV-2 infection.^{4,5} A recent *in vitro* study showed that vitamin D reduces viral load of nasal
68 epithelial cells infected with SARS-CoV-2,⁶ and preliminary evidence from two small human
69 trials suggested that vitamin D supplementation may help improve the prognosis of COVID-19
70 in infected individuals.^{7,8}

71 Ecological and observational studies lend further support to the hypothesis that lower
72 serum vitamin D concentrations are associated with both an increased risk of COVID-19 and an
73 increased severity of the infection. Early in the pandemic, the prevalence of COVID-19 in
74 European countries was negatively correlated with national averages for serum vitamin D
75 concentration.⁹ COVID-19 mortality rates were positively correlated with latitude, a proxy for
76 UV B exposure (higher latitude associates with lower UV B), which is required for synthesis of
77 vitamin D in the body.^{10,11} More recent studies report associations of lower pre-pandemic serum
78 vitamin D concentration or higher risk of vitamin D insufficiency with susceptibility to COVID-
79 19 infection¹²⁻¹⁴ and with severity of COVID-19 infection in hospitalized patients as indicated by
80 elevated biological markers of the cytokine storm, increased risk of intubation or use of

81 supplemental oxygen, ICU admission, and death.¹⁵⁻¹⁸ The mounting evidence from observational
82 studies and the known effects of vitamin D on the immune system contribute to speculation
83 about whether a simple and immediate intervention like vitamin D supplementation might be
84 effective in reducing risk of COVID-19 infection or severity, but sources of guidance for
85 clinicians and the public cite a lack of evidence for a causal association.^{19,20} Further evidence is
86 urgently needed.

87 The uncertainty about the causal association of vitamin D and COVID-19 arises because
88 of the well-known association of serum vitamin D concentration with many known risk factors
89 for COVID-19, including age, sex, body mass index (BMI), and race/ethnicity. The largest
90 observational study of the UK Biobank data (N = 348,598 participants) reported little to no
91 association of serum vitamin D with risk of COVID-19 infection in multivariate models adjusted
92 for covariates.²¹ In the absence of large, rigorous randomized trials, it is challenging to determine
93 whether other reported inverse associations of vitamin D and risk of COVID-19 are causal.
94 Given the high global prevalence of vitamin D insufficiency, estimating the true effect of vitamin
95 D on risk of COVID-19 infection and severity is important.

96 Mendelian randomization (MR), a study design to address causality, uses a form of
97 instrumental variable analysis to improve causal inference and to address the biases inherent in
98 observational studies. MR uses genetic variants that are associated with the exposure (i.e., serum
99 vitamin D) as instrumental variables, or genetic instruments, that represent the long-term usual
100 exposure. Following the laws of independent assortment, genetic variants are inherited from
101 parent to offspring in a random and independent manner. An individual's genotype therefore
102 mimics the lifelong randomization of individuals into groups with different long-term serum
103 vitamin D levels. Valid MR analysis depends on three key assumptions: (1) adequate strength

104 and validity of the genetic instruments (i.e., genetic variants that underlie vitamin D
105 metabolism), (2) independence of the genetic instruments from any confounders, and (3) absence
106 of direct effects of the genetic instruments on the outcome of interest (i.e., COVID-19 status).
107 MR addresses limitations in observational data including confounding, reverse causality, and
108 measurement error, and supports triangulation on the evidence for causality when randomized
109 trials are either impossible to conduct or currently unavailable. Genome-wide association (GWA)
110 studies of serum vitamin D levels identified genetic variants that are robustly associated with
111 serum vitamin D status in different populations and ancestries.^{22–28} Capitalizing on this finding,
112 researchers successfully applied MR to study associations of vitamin D with many clinical
113 outcomes, including diseases of immune system dysregulation and inflammation.^{29–45} We used
114 MR to investigate causality of the relationship of serum vitamin D status with the risk and
115 severity of COVID-19 infection.

116 **METHODS**

117 **Two-sample Mendelian Randomization with Multiple Instruments**

118 We estimated the effect of serum vitamin D levels on risk of COVID-19 infection and
119 severity with two-sample MR.⁴⁶ The associations of the genetic instruments with the exposure(s)
120 and outcome(s) were estimated in separate, independent samples, then used to calculate the MR
121 estimate of the effect of serum vitamin D status on COVID-19.⁴⁶ We used summary data from a
122 GWA study of serum vitamin D in the UK Biobank (sample 1)²² and genome-wide data for
123 COVID-19 patients versus comparison groups from the COVID-19 Host Genetics Initiative
124 (sample 2) (Figure 1).⁴⁷ Because of overlap between the UK Biobank and the COVID-19 Host
125 Genetics Initiative samples, which may bias MR results toward effects estimated from traditional

126 observational studies,⁴⁸ we replicated the analyses using summary GWAS data for serum vitamin
127 D from the SUNLIGHT Consortium.²³ We also performed sensitivity analyses to evaluate prior
128 hypotheses about the direct effect of vitamin D deficiency or insufficiency on COVID-19
129 outcomes using summary data from a meta-analysis of associations of vitamin D SNPs with the
130 dichotomous outcome of vitamin D deficiency versus sufficiency.²⁴

131 **Vitamin D Genetic Instrument(s)**

132 We constructed all genetic instruments for serum vitamin D levels based on independent
133 genetic loci associated with serum vitamin D levels at genome-wide significance ($p < 5 \times 10^{-8}$).
134 For the primary analysis, we used a biologically plausible genetic instrument (instrument A)
135 consisting of SNPs in genetic loci that encode proteins in vitamin D transport and metabolism
136 (Supplemental Figure 1), including the vitamin D binding protein (*GC*), 25 OH hydroxylase
137 (*CYP2R1*), 7-dehydrocholesterol reductase (*DHCR7*), and 24-hydroxylase (*CYP24A1*). All SNPs
138 included in Instrument A were associated with serum vitamin D at genome-wide significance in
139 the UK Biobank and replicated in the SUNLIGHT Consortium.^{22,23}

140 In secondary analyses, we expanded the genetic instrument to include additional genome-
141 wide significant SNPs in loci with no known relation to vitamin D metabolism (Figure 1). The
142 first expanded instrument (instrument B) included SNPs in two additional loci, Sec23 homolog A
143 (*SEC23A*) and amidohydrolase domain containing 1 (*AMDHDI*), that were associated with
144 serum vitamin D at genome-wide significance in both the UK Biobank and the SUNLIGHT
145 Consortium. The second expanded instrument (instrument C) included SNPs in 63 additional loci
146 associated with serum vitamin D at genome-wide significance in the UK Biobank (Supplemental
147 Table 1).

148 Finally, we constructed separate genetic instruments for risk of vitamin D deficiency and
149 insufficiency (defined as serum vitamin D < 50 nmol/L and <75 nmol/L, respectively) based on
150 meta-analysis results of vitamin D SNP—risk of vitamin D deficiency/insufficiency associations
151 in the SUNLIGHT Consortium.²⁴ The genome-wide significant SNPs contributing to the vitamin
152 D deficiency and insufficiency instruments were all in loci related to vitamin D metabolism,
153 including *GC*, *DHCR7*, and *CYP2R1*.

154 **SNP—Vitamin D Associations**

155 We extracted summary data, including beta-coefficients, standard errors, and effect
156 alleles and their frequencies for each SNP contributing to the genetic instruments. We primarily
157 used summary data from a UK Biobank GWAS of serum vitamin D in participants of “white
158 British” ancestry as defined by genotype principal component analysis (N = 401,460)²² and
159 replicated the analyses using summary data from a SUNLIGHT Consortium meta-analysis
160 GWAS of serum vitamin D (31 European ancestry cohorts, N = 79,366).²³ For sensitivity
161 analyses, we used data from a SUNLIGHT Consortium meta-analysis of associations of vitamin
162 D SNPs with risk of vitamin D deficiency/insufficiency (four European ancestry cohorts, N =
163 16,905).²⁴

164 Details of the UK Biobank and SUNLIGHT Consortia GWA studies of serum vitamin D
165 and the candidate SNP—vitamin D deficiency/insufficiency meta-analysis are described
166 elsewhere.^{22–24} Briefly, in the UK Biobank GWAS, serum vitamin D data were log-transformed
167 and standardized to a mean of 0 and standard deviation of 1, and SNP–vitamin D association
168 models were adjusted for age, sex, season of vitamin D measurement, vitamin D
169 supplementation, genotype batch, genotype array, and assessment center (as a proxy for latitude).
170 In the SUNLIGHT Consortium meta-analysis GWAS of serum vitamin D, serum vitamin D data

171 were log-transformed and SNP–vitamin D association models were adjusted for age, sex, BMI,
172 month of sample collection, cohort-specific variables such as geographical location and assay
173 batch, and genotype principal components.²² In the SUNLIGHT Consortium meta-analysis of
174 SNP—risk of vitamin D deficiency/insufficiency associations, models were adjusted for age, sex,
175 BMI, and season, and genomic control was applied to control for population stratification.²⁴

176 **Vitamin D Instrument Strength**

177 We assessed the strength and validity of the genetic instruments for vitamin D by
178 calculating the F-statistic,⁴⁹ according to the method described by Burgess and colleagues for
179 two-sample MR.⁴⁸ First, we calculated the variance in vitamin D explained by each SNP (R^2_{snp})
180 using the equation $R^2_{\text{snp}}=2\alpha^2 \text{MAF}(1-\text{MAF})$, where α = SNP–serum vitamin D association and
181 MAF = minor allele frequency. Next, we calculated the F-statistic for each of the instruments,
182 using the equation $F=(N-K-1)/K \times (R^2_{\text{instrument}})/(1-R^2_{\text{instrument}})$, where N = sample size, K =
183 number of SNPs contributing to the genetic instrument, and $R^2_{\text{instrument}}$ = sum of R^2_{snp} across
184 SNPs contributing to the instrument. We based all calculations on GWAS data from the UK
185 Biobank²² and calculated the F-statistic for sample sizes of 10,000 and 1.3 million individuals,
186 the approximate minimum and maximum sample sizes for the COVID-19 outcomes. We
187 considered the MR standard of F-statistic >10 as an indicator of instrument strength.⁴⁹

188 **SNP—COVID Associations**

189 We used genome-wide data for COVID-19 cases and comparison groups from the
190 COVID-19 Host Genetics Initiative, which is described elsewhere.⁴⁷ Briefly, the initiative
191 comprises a global effort to study genetic contributions to variability in COVID-19 infection and
192 severity. Contributing studies performed genome-wide analyses for COVID-19 case and

193 comparison groups following an analysis plan developed by the initiative. Studies with multiple
194 ancestries performed stratified analyses, and all analyses were adjusted for sex, age, age,² sex ×
195 age interaction, and genotype principal components. Results were summarized across studies and
196 ancestries by meta-analysis.

197 The publicly available summary results were downloaded from the COVID-19 Host
198 Genetics Initiative’s fourth data freeze and meta-analysis (<https://www.covid19hg.org/results/>),
199 which were released on October 20, 2020. We considered the following five COVID-19 case
200 versus control comparisons, using definitions provided by the initiative: (1) any COVID-19
201 versus population controls, (2) hospitalized COVID-19 versus population controls, (3) very
202 severe respiratory confirmed COVID-19 versus population controls, (4) any COVID-19 versus
203 confirmed COVID-19 negative, and (5) hospitalized versus non-hospitalized COVID-19. For
204 each COVID-19 “case” versus comparison group, we extracted summary data, including beta-
205 coefficients, standard errors, effect alleles, and effect allele frequencies corresponding to the
206 SNPs included in the vitamin D genetic instruments. For SNPs that were not available in the
207 public data, we identified proxy SNPs based on linkage disequilibrium (LD), using the standard
208 MR threshold for LD proxies of $r^2 > 0.8$.⁵⁰ All LD proxy SNPs were defined using 1000
209 Genomes European (CEU and GBR) sample data and identified through LDlink.⁵¹ Case and
210 comparison group definitions, sample sizes, and ancestry distributions for the different
211 comparisons are summarized in Table 1. Studies contributing to these analyses are listed in
212 Supplemental Table 1.

213 Summary data were available for mixed ancestry analyses for all outcomes. For the
214 outcomes of COVID-19 versus population controls and hospitalized COVID-19 versus
215 population controls, summary data from analyses limited to European ancestry participants were

216 also available. For these outcomes, we extracted summary data from both the mixed ancestry and
217 the European ancestry analyses for comparison and to address differences in ancestry
218 composition between vitamin D GWAS and COVID-19 Host Genetics Initiative sample
219 populations.

220 **Stratification of SNP—Vitamin D Associations by COVID-19 Risk Factors**

221 Our study design assumes that the potential effect of vitamin D nutritional status on
222 susceptibility to and severity of COVID-19 infection is the same across different COVID-19 risk
223 categories, including ancestry, age, sex, BMI, and smoking status. MR analysis under this
224 assumption requires the genetic instruments for vitamin D to be robust across the COVID-19 risk
225 subgroups. To address this assumption, we conducted stratified analyses of the SNP-vitamin D
226 associations for all SNPs contributing to the genetic instruments using data from the UK
227 Biobank. UK Biobank Resource data—including serum vitamin D levels, vitamin D
228 supplementation, covariates and stratification variables, and imputed genotype dosages with
229 reference to Haplotype Reference Consortium and UK10K haplotype resource—were obtained
230 under Application Number 24603. The Institutional Review Board at RTI International approved
231 the UK Biobank data use and analysis.

232 The SNP–vitamin D associations were estimated stratified by sex (male vs. female), age
233 group (40-50 vs. 50-60 vs. 60-70), BMI category (18.5-25 vs. 25-30 vs. 30-40 vs. >40), and
234 smoking status (never vs. former vs. current smokers). Analyses were performed using the
235 GENESIS R/ Bioconductor package,⁵² separately for European ancestry (N = 421,407) and
236 African ancestry (N = 7,859) participants, as classified via analysis with STRUCTURE⁵³ in
237 comparison to the CEU, YRI, and CHB 1000 Genomes populations.^{54,55} Statistical models

238 resembled those described for SNP–vitamin D associations in UK Biobank,²² except that
239 genotype batch was not included as a covariate and vitamin D levels were not log-transformed.

240 **Statistical Analysis for Two-Sample MR**

241 We performed two-sample MR using the inverse variance weighted (IVW) method,
242 which assumes that pleiotropy is either non-existent or balanced (i.e., any associations of genetic
243 instrument SNPs with phenotypes other than vitamin D are randomly positive and negative such
244 that the mean pleiotropic effect is zero).⁵⁶ We used the MR-Egger intercept test to check for
245 evidence of directional pleiotropy, and performed sensitivity analyses using the MR-Egger,
246 weighted-median and weighted-mode methods.^{57–59} The MR-Egger method relaxes the
247 assumption of no directional pleiotropy and corrects for pleiotropy-induced bias by allowing for
248 a nonzero intercept.^{56,57} The weighted-median method assumes that at least 50% of the weight of
249 the genetic instrument comes from nonpleiotropic SNPs and produces an unbiased estimate so
250 long as less than 50% of SNPs have pleiotropic effects.⁵⁸ The weighted-mode method assumes
251 that the most common (modal) SNP–vitamin D effect is the true effect and therefore allows for
252 the inclusion of pleiotropic SNPs without biasing the MR estimate.⁵⁹ Where evidence of
253 directional pleiotropy existed (MR-Egger intercept test P-value <0.05) we prioritized the MR
254 estimates from the MR-Egger, mode-weighted, and median-weighted analyses, otherwise we
255 prioritized the estimates from the IVW method. Results for our primary analysis are presented as
256 odds ratios and 95% confidence intervals (OR, 95% CI) for COVID-19 case versus comparator
257 per standard deviation increase in log-transformed serum vitamin D. The threshold for statistical
258 significance was set at a P-value < 0.05. All analyses were performed using R Studio version
259 1.2.1335 and the “TwoSampleMR” R package (<https://github.com/MRCIEU/TwoSampleMR>)⁵⁰.

260 We estimated potential bias under the null due to sample overlap between the UK
261 Biobank vitamin D GWAS and the COVID-19 Host Genetics Initiative using estimates of the
262 vitamin D–COVID-19 associations from observational studies, the percentage of sample overlap,
263 and the relative bias (reciprocal of the F-statistic), as previously described.⁴⁸ We compared SNP-
264 vitamin D associations stratified by sex, age group, BMI category, and smoking status with
265 Pearson’s correlation coefficients.

266 We calculated the minimum odds ratio detectable at 80% power for each COVID-19
267 outcome using the “mRnd” tool (<https://shiny.cnsgenomics.com/mRnd/>).⁶⁰ Based on heritability
268 estimates from the UK Biobank and SUNLIGHT Consortium vitamin D GWA studies,^{23,61} we
269 assumed the genetic instruments accounted for 3% of the variance in serum vitamin D for all
270 power calculations.

271 **Patient and Public Involvement**

272 We used publicly available data from several consortia including the UK Biobank, the
273 SUNLIGHT Consortium and the COVID-19 Host Genetics Initiative in this study. These
274 consortia were not involved in any stage of the design and conduct of this research, nor were
275 they asked to advise on interpretation or writing up of results. No patients were involved in any
276 stage of the research process. There are no plans to disseminate the results of the research to
277 study participants or the relevant patient community.

278 **RESULTS**

279 **Genetic Instruments for Vitamin D**

280 Characteristics of the genetic instruments for serum vitamin D, including contributing
281 SNPs, SNP–vitamin D associations in the UK Biobank, and F-statistics, indicate that instruments

282 A and B clearly exceed the MR standard for instrument strength of F-statistic > 10 across the
283 range of COVID-19 outcome sample sizes (Table 2, Supplemental Table 2). The F-statistics for
284 instruments A, B, and C were respectively 53, 37, and 5 for the minimum sample size of 10,000
285 participants and 6,915, 4,746, and 600 for the maximum sample size of 1.3 million participants.
286 Instrument A, which was limited to SNPs in loci related to vitamin D metabolism
287 (Supplementary Figure 1), had the largest F-statistics across all sample sizes, reflecting the
288 strength of the SNP–vitamin D associations for the SNPs in loci related to the vitamin D
289 metabolic pathway.

290 **COVID-19 Case-Comparison Groups and Power Calculations**

291 We considered multiple case definitions, sample sizes, and ancestral compositions for the
292 COVID-19 case-control comparisons in our analysis (Table 1). Sample sizes across the analyses
293 ranged from just over 10,000 (for hospitalized COVID-19 vs. non-hospitalized COVID-19) to
294 over 1.3 million (for COVID-19 vs. population). The majority ($>75\%$) of participants were of
295 European ancestry. Non-European ancestry participants made up a much smaller fraction of the
296 total sample.

297 Based on the sample sizes across the COVID-19 case-control comparisons, and assuming
298 our genetic instruments explained 3% of the variance in serum vitamin D, we calculated that this
299 study had 80% power to detect moderate effect sizes comparable to those seen in observational
300 studies. The minimum detectable ORs for a standard deviation increase in serum vitamin D
301 ranged from 0.66 (hospitalized COVID-19 vs. non-hospitalized COVID-19) to 0.86 (COVID-19
302 vs. population; Supplemental Table 3).

303 **MR Estimates for Serum Vitamin D Effect on COVID-19 Outcomes**

304 MR estimates (odds ratios and 95% confidence intervals) for the effect of genetically
305 predicted serum vitamin D on risk of COVID-19 outcomes, calculated separately for instruments
306 A, B, and C, are shown in Figure 2 and Table 3. MR scatter plots and forest plots further
307 illustrating these results are shown in Supplementary Figure 2. MR-Egger intercept tests for
308 pleiotropy were not statistically significant (MR-Egger intercept p-values > 0.1) for all
309 instruments and all outcomes (Table 3). Thus, we prioritized the MR estimates from the IVW
310 analyses and found little to no evidence for an effect of vitamin D on risk of any COVID-19
311 outcome considered. For instrument A, using SNPs in loci related to vitamin D metabolism, the
312 odds ratio for a standard deviation increase in serum vitamin D was 1.04 (95% confidence
313 interval 0.92 to 1.18) for any COVID-19 infection versus population controls, 1.05 (0.84-1.31)
314 for hospitalized COVID-19 versus population controls, 0.96 (0.64 to 1.43) for severe respiratory
315 COVID-19 versus population controls, 1.15 (0.99 to 1.35) for COVID-19 positive versus
316 COVID-19 negative, and 1.44 (0.75 to 2.78) for hospitalized COVID-19 versus non-hospitalized
317 COVID-19 (p-values >0.1 for all comparisons). Results were similar for instruments B and C.
318 Results of MR analyses limited to European ancestry participants, which we were able to
319 perform for the outcomes of COVID-19 versus population and hospitalized COVID-19 versus
320 population, were also non-significant (Supplementary Figure 3).

321 For all outcomes, the MR-Egger intercept test showed little to no evidence of pleiotropy
322 (Table 3). Nevertheless, we performed sensitivity analysis to explore possible pleiotropy using
323 the MR-Egger, weighted-median, and weighted-mode MR methods, and this analysis showed a
324 *direct* association of serum vitamin D with risk of hospitalized COVID-19 versus non-
325 hospitalized COVID-19 (Supplemental Figure 4). For this outcome, MR estimates from the MR-

326 Egger and mode-weighted analyses found that chronically *higher* vitamin D levels increase the
327 odds of hospitalization in COVID-19-infected participants (OR [95% CI] for instrument A: 1.83
328 [1.12, 2.97] for mode-weighted MR and 3.69 [1.53, 8.93] for MR-Egger). MR estimates from the
329 MR-Egger, mode- and median-weighted MR analyses were non-significant for all other
330 outcomes, consistent with results from the IVW analysis.

331 Estimations of bias because of participant overlap between the UK Biobank and the
332 COVID-19 Host Genetics Initiative were negligible. For the outcome with maximum participant
333 overlap (58% for severe respiratory COVID-19 vs. population), we estimated the bias for
334 instruments A, B, and C to be 0.1%, 0.2%, and 1.0% respectively. We confirmed the negligible
335 effect of participant overlap in our study by repeating the analyses using summary data for
336 GWAS significant SNP–vitamin D associations from the SUNLIGHT consortium in place of UK
337 Biobank data (Supplementary Table 4). Results using the SUNLIGHT Consortium data were
338 consistent with results using the UK Biobank data; there was no evidence of pleiotropy for any
339 outcome (MR-Egger intercept p-value >0.1), and IVW MR estimates for the effect of genetically
340 predicted serum vitamin D on risk of COVID-19 outcomes were similar in magnitude and
341 direction and were not statistically significant (Supplementary Table 4).

342 **MR Estimates for Vitamin D Deficiency/Insufficiency and COVID-19 Outcomes**

343 A further sensitivity analysis evaluated the effect of genetically predicted vitamin D
344 deficiency and insufficiency on COVID-19 outcomes (Table 3, Figure 4). MR-Egger intercept
345 tests for pleiotropy were all non-significant (p-value > 0.1), and the IVW MR estimates did not
346 support an effect of long-term vitamin D deficiency or insufficiency on the risk of any of the
347 COVID-19 outcomes evaluated (Table 3, Figure 4).

348 **Covariate Effects on SNP-Vitamin D Associations**

349 Stratified analyses of SNP-vitamin D associations evaluated the validity of the genetic
350 instruments across known COVID-19 risk factors, including ancestry, age, sex, BMI, and
351 smoking status. Correlations of age-, sex-, BMI-, and smoking status-specific SNP–vitamin D
352 associations are shown separately for European (N = 421,407) and African (N = 7,859) ancestry
353 participants in Supplemental Figure 5. For European ancestry participants, subgroup-specific
354 SNP–vitamin D associations were strongly correlated across age, sex, BMI category, and
355 smoking status (Pearson’s correlation coefficients ranging from 0.90 to 0.99), supporting the use
356 of the instrument in the general European ancestry population. For African ancestry participants,
357 the correlations among subgroup-specific SNP–vitamin D associations were less clear (Pearson’s
358 correlation coefficients ranged from -0.95 to 0.72), suggesting that optimization of the genetic
359 instruments for the African ancestry population may be necessary. Given the small sample sizes
360 in the African ancestry subgroups, further research is needed to validate the instruments and their
361 use in African and other non-European Ancestry populations.

362 **DISCUSSION**

363 We used summary data from genome-wide analysis in the UK Biobank, SUNLIGHT
364 Consortium, and COVID-19 Host Genetics Initiative to conduct the first MR study of serum
365 vitamin D and COVID-19 outcomes. Although our study was powered to detect effects
366 comparable to those seen in observational studies of vitamin D and COVID-19, we found little to
367 no evidence of an effect of genetically predicted serum vitamin D levels on COVID-19
368 outcomes, including risk of COVID-19 infection, severe respiratory infection, and
369 hospitalization. Our results were robust to multiple genetic instruments for vitamin D and were
370 replicated using SNP–vitamin D association data from two independent samples. Results from

371 sensitivity analyses evaluating the effect of genetically predicted vitamin D deficiency or
372 insufficiency on risk of COVID-19 outcomes were similarly null. In summary, our results
373 suggest that long-term usual vitamin D nutritional status does not have a causal effect on
374 susceptibility to COVID-19 infection and its severity.

375 Our results are consistent with findings from the largest observational study to date of
376 serum vitamin D and COVID-19 infection,²¹ which used UK Biobank data on approximately
377 350,000 White, Black, and South Asian participants, and found associations of pre-pandemic
378 vitamin D levels with COVID-19 infection in univariate models, but not in models adjusting for
379 confounding by known COVID-19 risk factors, including age, sex, race/ethnicity, BMI,
380 socioeconomic status, smoking status, diabetes, blood pressure, chronic illness, and disability.
381 We used the MR approach as an alternative way to address potential confounding. We
382 instrumented long-term serum vitamin D status with genetic variants that are unlikely to have
383 direct associations with known risk factors for COVID-19 infection. Specifically, the genetic
384 instrument for our primary analysis (instrument A) was limited to SNPs in loci with functional
385 links to serum vitamin D, including *GC*, *DHCR7*, *CYP2R1*, and *CYP24A1*. With this approach,
386 we saw no evidence for an effect of genetically predicted (i.e., unconfounded) variation in serum
387 vitamin D on risk and severity of COVID-19 infection.

388 These results are contrary to findings from several smaller observational studies showing
389 an association of lower vitamin D levels with higher risk of testing positive for COVID-19 in the
390 general population¹²⁻¹⁴ and higher risk of COVID-19 related hospitalization, ICU admission,
391 intubation and death among infected individuals.¹⁵⁻¹⁸ However, these studies were inconsistent in
392 the covariates included in statistical models, and in several studies estimates of vitamin D
393 associations with susceptibility to and severity of COVID-19 infection were attenuated with

394 covariate adjustment, suggesting the observed effects could be the result of residual confounding.
395 Additionally, many of these studies were case-control or cross-sectional studies, making it
396 difficult to rule out the possibility of reverse causality (i.e., that COVID-19 infection or its
397 symptoms could lead to lowering of serum vitamin D). The MR approach addresses key
398 limitations of these studies, including confounding and reverse causality, and is recognized as a
399 method for improving causal inference in epidemiology.

400 This study has many strengths. We used summary data from the largest GWA studies of
401 serum vitamin D to maximize the strength and validity of the genetic instruments for vitamin D
402 and leveraged publicly available data on nearly 1.4 million participants, including 17,965
403 COVID-19 cases, from 38 study cohorts in the COVID-19 Host Genetics Initiative. We
404 addressed potential pleiotropy of the genetic instruments, which if present would limit the
405 interpretation of the MR results, by comparing results across the biologically plausible and
406 expanded instruments, testing for evidence of pleiotropy with the MR-Egger intercept test, and
407 performing sensitivity analysis with alternative MR methods that address pleiotropy. We
408 explored potential threshold effects using genetic instruments for vitamin D deficiency and
409 insufficiency. Our findings were robust across all exploration, and, in triangulation with results
410 of the largest to date observational study of vitamin D and COVID-19, the findings help to
411 clarify the nature of the association of vitamin D with COVID-19 outcomes.

412 As in any MR study, the valid interpretation of our results rests on the MR assumptions
413 of adequate strength of the genetic instruments for vitamin D and absence of direct effects of the
414 instruments on COVID-19 outcomes and potential confounders (i.e., pleiotropy). Our
415 instruments easily met the MR standard for instrument strength (F-statistic >10) and we found no
416 evidence of pleiotropy from the MR-Egger intercept test. Our results were consistent across the

417 biologically plausible and expanded genetic instruments and largely similar in sensitivity
418 analysis using MR methods designed to address pleiotropy. While the findings in our study are
419 unlikely to be biased by pleiotropy, exhaustive tests for pleiotropy are challenging, making the
420 MR assumptions difficult to fully verify.

421 We used a two-sample MR approach, using estimates of associations of the genetic
422 instruments with vitamin D and with COVID-19 outcomes from two separate samples. This
423 approach maximizes sample size and reduces the likelihood of bias toward estimates from
424 observational studies⁴⁶ but requires that the two samples are drawn from similar populations. In
425 our study, the summary data for associations of the genetic instruments with vitamin D were
426 from analyses in European ancestry participants, while the data for associations of the genetic
427 instruments with COVID-19 outcomes were from analyses including up to 25% non-European
428 ancestry participants. Repeating the analysis with a subgroup of European-only participants for
429 two outcomes, COVID-19 versus population controls and hospitalized COVID-19 versus
430 population controls, indicated minimal impact of ancestral differences on MR estimates.
431 However, ancestry-specific data were not available for COVID-19 outcomes with larger
432 proportions of non-European ancestry participants, and the impact of ancestral differences could
433 not be fully explored. In sensitivity analyses using alternative MR methods that assume some
434 pleiotropy, findings indicated an association of *higher* vitamin D levels with *higher* risk of
435 hospitalization among COVID-19-infected individuals, the outcome with the highest proportion
436 of non-European ancestry participants. This finding could be driven by racial/ethnic disparities in
437 access to health care (i.e., hospitalized COVID-19 individuals may be more likely of European
438 ancestry and therefore have higher vitamin D status). We also found evidence for racial
439 differences in risk-factor stratified SNP–vitamin D associations. Our results suggest that SNP–

440 vitamin D associations are strongly correlated across sex, age, BMI, and smoking subgroups for
441 individuals of European ancestry, but not for individuals of African ancestry, although because
442 of small sample sizes for the African ancestry subgroups we are unable to draw firm conclusions.
443 Further research using data from ancestry-specific GWA studies of vitamin D and COVID-19
444 outcomes is warranted to further investigate the impact of race/ethnicity on the vitamin D–
445 COVID-19 relationship.

446 This study was not designed to evaluate the effect of acute changes in vitamin D status
447 (i.e., from supplementation) on prevention or treatment of COVID-19. One randomized trial and
448 one quasi-experimental trial have provided evidence for a positive effect of vitamin D
449 supplementation around time of diagnosis on COVID-19 prognosis in infected individuals. The
450 doses of vitamin D administered in these studies could result in an acute change in the
451 availability of vitamin D, which may support the immune system’s response to the virus,
452 mitigate acute lung injury, and contribute to improved prognosis. Our results, which pertain to
453 long-term vitamin D status, do not preclude the possibility that therapeutic doses of vitamin D
454 may be effective in preventing or treating COVID-19 infection. Larger randomized trials using
455 diverse sample populations are needed to investigate the potential use of therapeutic doses of
456 vitamin D supplementation for COVID-19 prevention and treatment.

457 In conclusion, we used two-sample MR to study the associations of vitamin D with
458 COVID-19 outcomes to address limitations of existing observational studies, including
459 confounding and reverse causality. We found no evidence for an effect of genetically predicted
460 variation in serum vitamin D on risk or severity of COVID-19 infection. Our findings suggest
461 that chronic differences in serum vitamin D do not have a causal effect on susceptibility to
462 COVID-19 infection or severity of COVID-19 among those infected, and that associations

463 observed in previous studies may have been driven by confounding. Future directions of this
464 work include extension of the MR approach to non-European ancestry populations, and
465 investigation of potential modification of genetically predicted vitamin D effects on COVID-19
466 risk and severity by COVID-19 risk factors, including race/ethnicity, age, sex, and BMI.
467 Randomized trials are needed to inform the question of whether acute changes in vitamin D
468 levels (i.e., from supplementation) are efficacious in COVID-19 prevention and treatment across
469 diverse populations.

470

471 **SUMMARY BOX**

472 **What is already known on this topic**

- 473 • Observational studies report that lower vitamin D levels are associated with increased risk
474 and severity of COVID-19
- 475 • Known risk factors for COVID-19 are associated with lower vitamin D levels in blood
- 476 • It is unknown whether observed associations of vitamin D and COVID-19 are causal

477 **What this study adds**

- 478 • This study found no evidence for associations of genetically predicted long-term vitamin D
479 levels with risk and severity of COVID-19 infection
- 480 • The findings do not support a causal effect of long-term usual vitamin D nutritional status on
481 COVID-19 infection and its severity
- 482 • These findings have implications for the potential effectiveness of low level vitamin D
483 supplementation as a strategy for COVID-19 prevention and treatment, but do not address the
484 therapeutic use of vitamin D in acute disease

485

486 **FOOTNOTES**

487 **Contributor and Guarantor Confirmation:** The study was designed and conceived by BKP,
488 PAC and AGC. The MR analyses were performed by BKP in close consultation with PAC. The
489 follow-up stratified analyses of SNP—vitamin D associations were performed by NG. The
490 manuscript was drafted by BKP and PAC, and edited with input from AGC, DBH and NG. All
491 authors reviewed and contributed to the discussion of findings and the crafting of the manuscript
492 and gave final approval to the version submitted for publication. PAC is the guarantor.

493 **Funding:** BKP was supported by the National Institutes of Health under award T32-DK007158.
494 ACG was supported by the National Institutes of Health under award R01-HG006849. NG, DBH
495 and PAC were partially supported by the National Institutes of Health under award R01-
496 HL149352. The study content is solely the responsibility of the authors and does not necessarily
497 represent the official views of the funders. The funders had no role in the study design, data
498 collection, analysis, interpretation, or writing, nor in the decision to submit the article for
499 publication.

500 **Competing interests:** All authors have completed the ICMJE uniform disclosure form at
501 www.icmje.org/coi_disclosure.pdf and declare: no support from any organization for the
502 submitted work; BKP was supported by the National Institutes of Health during the conduct of
503 the study, AGC, DBH and PAC report receiving grants from the National Institutes of Health
504 during the conduct of the study; no other relationships or activities that could appear to have
505 influenced the submitted work.

506 **Ethics approval:** The main analysis conducted in this study used publicly available summary
507 data and did not require ethical approval. The use of UK Biobank data for the stratified analysis

508 of SNP—vitamin D associations by COVID-19 risk factors was approved by the Institutional
509 Review Board at RTI International.

510 **Data sharing:** All data used for this analysis are publicly available. Code implementing the
511 MR analysis is available upon request from the corresponding author.

512 **Transparency:** PAC is the corresponding author and guarantor for this paper and affirms that
513 the manuscript is an honest, accurate, and transparent account of the study being reported; that
514 no important aspects of the study have been omitted; and that any discrepancies from the study
515 as originally planned have been explained.

516 **Copyright:** The Corresponding Author has the right to grant on behalf of all authors and does
517 grant on behalf of all authors, a worldwide licence to the Publishers and its licensees in
518 perpetuity, in all forms, formats and media (whether known now or created in the future), to i)
519 publish, reproduce, distribute, display and store the Contribution, ii) translate the Contribution
520 into other languages, create adaptations, reprints, include within collections and create
521 summaries, extracts and/or, abstracts of the Contribution, iii) create any other derivative work(s)
522 based on the Contribution, iv) to exploit all subsidiary rights in the Contribution, v) the inclusion
523 of electronic links from the Contribution to third party material where-ever it may be located;
524 and, vi) licence any third party to do any or all of the above.

525

526

527

528

529 **REFERENCES**

- 530 1. Home. *Johns Hopkins Coronavirus Resource Center* <https://coronavirus.jhu.edu/>.
- 531 2. Guillot, X., Semerano, L., Saidenberg-Kermanac'h, N., Falgarone, G. & Boissier, M.-C.
532 Vitamin D and inflammation. *Joint Bone Spine* **77**, 552–557 (2010).
- 533 3. Brenner, H., Holleczeck, B. & Schöttker, B. Vitamin D Insufficiency and Deficiency and
534 Mortality from Respiratory Diseases in a Cohort of Older Adults: Potential for Limiting the
535 Death Toll during and beyond the COVID-19 Pandemic? *Nutrients* **12**, (2020).
- 536 4. Cui, C. *et al.* Vitamin D receptor activation regulates microglia polarization and oxidative
537 stress in spontaneously hypertensive rats and angiotensin II-exposed microglial cells: Role
538 of renin-angiotensin system. *Redox Biol.* **26**, 101295 (2019).
- 539 5. Kai, H. & Kai, M. Interactions of coronaviruses with ACE2, angiotensin II, and RAS
540 inhibitors—lessons from available evidence and insights into COVID-19. *Hypertens. Res.*
541 1–7 (2020) doi:10.1038/s41440-020-0455-8.
- 542 6. Mok, C. K. *et al.* Calcitriol, the active form of vitamin D, is a promising candidate for
543 COVID-19 prophylaxis. *bioRxiv* 2020.06.21.162396 (2020)
544 doi:10.1101/2020.06.21.162396.
- 545 7. Entrenas Castillo, M. *et al.* Effect of calcifediol treatment and best available therapy versus
546 best available therapy on intensive care unit admission and mortality among patients
547 hospitalized for COVID-19: A pilot randomized clinical study. *J. Steroid Biochem. Mol.*
548 *Biol.* **203**, 105751 (2020).
- 549 8. Annweiler, C. *et al.* Vitamin D and survival in COVID-19 patients: A quasi-experimental
550 study. *J. Steroid Biochem. Mol. Biol.* **204**, 105771 (2020).
- 551 9. Ilie, P. C., Stefanescu, S. & Smith, L. The role of vitamin D in the prevention of coronavirus
552 disease 2019 infection and mortality. *Aging Clin. Exp. Res.* **32**, 1195–1198 (2020).
- 553 10. Whittemore, P. B. COVID-19 fatalities, latitude, sunlight, and vitamin D. *Am. J. Infect.*
554 *Control* **48**, 1042–1044 (2020).
- 555 11. Kohlmeier, M. Avoidance of vitamin D deficiency to slow the COVID-19 pandemic. *BMJ*
556 *Nutr. Prev. Health* bmjnph-2020-000096 (2020) doi:10.1136/bmjnph-2020-000096.
- 557 12. Kaufman, H. W., Niles, J. K., Kroll, M. H., Bi, C. & Holick, M. F. SARS-CoV-2 positivity
558 rates associated with circulating 25-hydroxyvitamin D levels. *PLoS ONE* **15**, (2020).
- 559 13. Meltzer, D. O. *et al.* Association of Vitamin D Status and Other Clinical Characteristics
560 With COVID-19 Test Results. *JAMA Netw. Open* **3**, e2019722–e2019722 (2020).
- 561 14. D'Avolio, A. *et al.* 25-Hydroxyvitamin D Concentrations Are Lower in Patients with
562 Positive PCR for SARS-CoV-2. *Nutrients* **12**, 1359 (2020).

- 563 15. Ye, K. *et al.* Does Serum Vitamin D Level Affect COVID-19 Infection and Its Severity?-A
564 Case-Control Study. *J. Am. Coll. Nutr.* (2020) doi:10.1080/07315724.2020.1826005.
- 565 16. Maghbooli, Z. *et al.* Vitamin D sufficiency, a serum 25-hydroxyvitamin D at least 30 ng/mL
566 reduced risk for adverse clinical outcomes in patients with COVID-19 infection. *PLoS One*
567 **15**, e0239799 (2020).
- 568 17. Macaya, F. *et al.* Interaction between age and vitamin D deficiency in severe COVID-19
569 infection. *Nutr. Hosp.* **37**, 1039–1042 (2020).
- 570 18. Baktash, V. *et al.* Vitamin D status and outcomes for hospitalised older patients with
571 COVID-19. *Postgrad. Med. J.* (2020) doi:10.1136/postgradmedj-2020-138712.
- 572 19. COVID-19 Treatment Guidelines Panel. Coronavirus Disease 2019 (COVID-19) Treatment
573 Guidelines. *COVID-19 Treatment Guidelines*
574 <https://www.covid19treatmentguidelines.nih.gov/adjunctive-therapy/vitamin-d/>.
- 575 20. Williams, C. & Gingery, J. G. Joint guidance on vitamin D in the era of COVID-19.
576 *Endocrine Society* [https://www.endocrine.org/news-and-advocacy/news-room/2020/joint-](https://www.endocrine.org/news-and-advocacy/news-room/2020/joint-guidance-on-vitamin-d)
577 [guidance-on-vitamin-d](https://www.endocrine.org/news-and-advocacy/news-room/2020/joint-guidance-on-vitamin-d) (2020).
- 578 21. Hastie, C. E. *et al.* Vitamin D concentrations and COVID-19 infection in UK Biobank.
579 *Diabetes Metab. Syndr. Clin. Res. Rev.* **14**, 561–565 (2020).
- 580 22. Manousaki, D. *et al.* Genome-wide Association Study for Vitamin D Levels Reveals 69
581 Independent Loci. *Am. J. Hum. Genet.* **106**, 327–337 (2020).
- 582 23. Jiang, X. *et al.* Genome-wide association study in 79,366 European-ancestry individuals
583 informs the genetic architecture of 25-hydroxyvitamin D levels. *Nat. Commun.* **9**, 260
584 (2018).
- 585 24. Wang, T. J. *et al.* Common genetic determinants of vitamin D insufficiency: a genome-wide
586 association study. *Lancet* **376**, 180–188 (2010).
- 587 25. Ahn, J. *et al.* Genome-wide association study of circulating vitamin D levels. *Hum. Mol.*
588 *Genet.* **19**, 2739–2745 (2010).
- 589 26. Hong, J. *et al.* Transethnic Evaluation Identifies Low-Frequency Loci Associated With 25-
590 Hydroxyvitamin D Concentrations. *J. Clin. Endocrinol. Metab.* **103**, 1380–1392 (2018).
- 591 27. Arabi, A. *et al.* CYP2R1 polymorphisms are important modulators of circulating 25-
592 hydroxyvitamin D levels in elderly females with vitamin insufficiency, but not of the
593 response to vitamin D supplementation. *Osteoporos. Int. J. Establ. Result Coop. Eur.*
594 *Found. Osteoporos. Natl. Osteoporos. Found. USA* **28**, 279–290 (2017).
- 595 28. Jones, P. *et al.* Distribution of variants in multiple vitamin D-related loci
596 (DHCR7/NADSYN1, GC, CYP2R1, CYP11A1, CYP24A1, VDR, RXR α and RXR γ) vary

- 597 between European, East-Asian and Sub-Saharan African-ancestry populations. *Genes Nutr.*
598 **15**, 5 (2020).
- 599 29. Liefwaard, M. C. *et al.* Vitamin D and C-Reactive Protein: A Mendelian Randomization
600 Study. *PloS One* **10**, e0131740 (2015).
- 601 30. Afzal, S., Brondum-Jacobsen, P., Bojesen, S. E. & Nordestgaard, B. G. Genetically low
602 vitamin D concentrations and increased mortality: Mendelian randomisation analysis in
603 three large cohorts. *BMJ* **349**, g6330 (2014).
- 604 31. Brondum-Jacobsen, P., Benn, M., Afzal, S. & Nordestgaard, B. G. No evidence that
605 genetically reduced 25-hydroxyvitamin D is associated with increased risk of ischaemic
606 heart disease or myocardial infarction: a Mendelian randomization study. *Int. J. Epidemiol.*
607 **44**, 651–661 (2015).
- 608 32. Bae, S.-C. & Lee, Y. H. Vitamin D level and risk of systemic lupus erythematosus and
609 rheumatoid arthritis: a Mendelian randomization. *Clin. Rheumatol.* **37**, 2415–2421 (2018).
- 610 33. Hysinger, E. B. *et al.* Mendelian randomization analysis demonstrates that low vitamin D is
611 unlikely causative for pediatric asthma. *J. Allergy Clin. Immunol.* **138**, 1747-1749.e4 (2016).
- 612 34. Manousaki, D., Mokry, L. E., Ross, S., Goltzman, D. & Richards, J. B. Mendelian
613 Randomization Studies Do Not Support a Role for Vitamin D in Coronary Artery Disease.
614 *Circ. Cardiovasc. Genet.* **9**, 349–356 (2016).
- 615 35. Manousaki, D. *et al.* Vitamin D levels and susceptibility to asthma, elevated
616 immunoglobulin E levels, and atopic dermatitis: A Mendelian randomization study. *PLoS*
617 *Med.* **14**, e1002294 (2017).
- 618 36. Mao, Y., Zhan, Y. & Huang, Y. Vitamin D and asthma: A Mendelian randomization study.
619 *Ann. Allergy Asthma Immunol. Off. Publ. Am. Coll. Allergy Asthma Immunol.* **119**, 95-97.e1
620 (2017).
- 621 37. Mokry, L. E. *et al.* Genetically decreased vitamin D and risk of Alzheimer disease.
622 *Neurology* **87**, 2567–2574 (2016).
- 623 38. Ong, J.-S. *et al.* Association of vitamin D levels and risk of ovarian cancer: a Mendelian
624 randomization study. *Int. J. Epidemiol.* **45**, 1619–1630 (2016).
- 625 39. Skaaby, T. *et al.* Vitamin D status, filaggrin genotype, and cardiovascular risk factors: a
626 Mendelian randomization approach. *PloS One* **8**, e57647 (2013).
- 627 40. Taylor, A. E. *et al.* Investigating causality in the association between 25(OH)D and
628 schizophrenia. *Sci. Rep.* **6**, 26496 (2016).
- 629 41. Trummer, O. *et al.* Vitamin D and mortality: a Mendelian randomization study. *Clin. Chem.*
630 **59**, 793–797 (2013).

- 631 42. Viatte, S. *et al.* The role of genetic polymorphisms regulating vitamin D levels in
632 rheumatoid arthritis outcome: a Mendelian randomisation approach. *Ann. Rheum. Dis.* **73**,
633 1430–1433 (2014).
- 634 43. Vimalleswaran, K. S. *et al.* Causal relationship between obesity and vitamin D status: bi-
635 directional Mendelian randomization analysis of multiple cohorts. *PLoS Med.* **10**, e1001383
636 (2013).
- 637 44. Vimalleswaran, K. S. *et al.* Association of vitamin D status with arterial blood pressure and
638 hypertension risk: a mendelian randomisation study. *Lancet Diabetes Endocrinol.* **2**, 719–
639 729 (2014).
- 640 45. Ye, Z. *et al.* Association between circulating 25-hydroxyvitamin D and incident type 2
641 diabetes: a mendelian randomisation study. *Lancet Diabetes Endocrinol.* **3**, 35–42 (2015).
- 642 46. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian Randomization Analysis With
643 Multiple Genetic Variants Using Summarized Data. *Genet. Epidemiol.* **37**, 658–665 (2013).
- 644 47. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host
645 genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur. J.*
646 *Hum. Genet.* 1–4 (2020) doi:10.1038/s41431-020-0636-6.
- 647 48. Burgess, S., Davies, N. M. & Thompson, S. G. Bias due to participant overlap in
648 two-sample Mendelian randomization. *Genet. Epidemiol.* **40**, 597–608 (2016).
- 649 49. Burgess, S., Thompson, S. G. & Collaboration, C. C. G. Avoiding bias from weak
650 instruments in Mendelian randomization studies. *Int. J. Epidemiol.* **40**, 755–764 (2011).
- 651 50. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the
652 human phenome. *eLife* **7**, e34408 (2018).
- 653 51. Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-
654 specific haplotype structure and linking correlated alleles of possible functional variants.
655 *Bioinforma. Oxf. Engl.* **31**, 3555–3557 (2015).
- 656 52. Gogarten, S. M. *et al.* Genetic association testing using the GENESIS R/Bioconductor
657 package. *Bioinformatics* **35**, 5346–5348 (2019).
- 658 53. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using
659 multilocus genotype data. *Genetics* **155**, 945–959 (2000).
- 660 54. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
- 661 55. An integrated map of structural variation in 2,504 human genomes | Nature.
662 <https://www.nature.com/articles/nature15394>.
- 663 56. Hemani, G., Bowden, J. & Davey Smith, G. Evaluating the potential role of pleiotropy in
664 Mendelian randomization studies. *Hum. Mol. Genet.* **27**, R195–R208 (2018).

- 665 57. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid
666 instruments: effect estimation and bias detection through Egger regression. *Int. J.*
667 *Epidemiol.* **44**, 512–525 (2015).
- 668 58. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent Estimation in
669 Mendelian Randomization with Some Invalid Instruments Using a Weighted Median
670 Estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
- 671 59. Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data Mendelian
672 randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46**, 1985–1998
673 (2017).
- 674 60. Brion, M.-J. A., Shakhbazov, K. & Visscher, P. M. Calculating statistical power in
675 Mendelian randomization studies. *Int. J. Epidemiol.* **42**, 1497–1501 (2013).
- 676 61. Manousaki, D. *et al.* Genome-wide Association Study for Vitamin D Levels Reveals 69
677 Independent Loci. *Am. J. Hum. Genet.* (2020) doi:10.1016/j.ajhg.2020.01.017.
- 678

Table 1. Sample Description for COVID-19 case and comparison groups

Comparison	Case definition	Comparison group definition	Sample size	Ancestry
Any COVID-19 case vs. population controls	laboratory/physician confirmed OR self-reported positive	Everyone that is not a case (e.g., the population)	17,965 cases, 1,370,547 controls	93.6% EUR 3.1% EAS/SAS 1.8% AFR 1% ARAB <1% AMR <1% HIS
Hospitalized COVID-19 cases vs. population controls	laboratory/physician confirmed AND hospitalized for COVID-19	Everyone that is not a case (e.g., the population)	7,885 cases, 961,804 controls	93.7% EUR 3.5% EAS/SAS 1.1% AFR 1.4% ARAB <1% AMR <1% HIS
Severe respiratory COVID-19 cases vs. population controls	laboratory confirmed AND hospitalized AND respiratory support or death	Everyone that is not a case (e.g., the population)	4,336 cases, 623,902 controls	99.6% EUR <1% AMR
COVID-19 positive cases vs. COVID-19 negative controls	laboratory/physician confirmed OR self-reported positive	laboratory confirmed OR self-reported negative	11,085 cases, 116,794 controls	86.5% EUR <1% EAS/SAS 9.6% AFR 3.2% HIS
Hospitalized vs. nonhospitalized COVID-19	laboratory/physician confirmed AND hospitalized for COVID-19	Laboratory or physician confirmed COVID-19	2,430 hospitalized 8,478 nonhospitalized	75.3% EUR <1% EAS/SAS 12.6% AFR 6.4% ARAB 4.7% HIS

AFR = African; ARAB = Arabic; AMR = Ad Mixed American; EAS/SAS = East Asian/South Asian; EUR = European; HIS = Hispanic

Table 2. Characteristics of vitamin D instruments and association with measured vitamin D status in UK Biobank (N=401,460, Source: Manousaki et al., 2020)

Loci	SNP	EA	NEA	EAF	Beta-vitD	P-vitD	F-statistic	
							N=10,000	N=1,300,000
Instrument A							53	6,915
GC	rs11723621	G	A	0.29	-0.19	2.9E-1689	147	19,072
CYP2R1	rs10832289	T	A	0.41	-0.07	2.03E-266	23	4,597
DHCR7	rs12803256	G	A	0.78	0.10	1.3E-378	35	715
CYP24A1	rs6127099	T	A	0.28	-0.04	9.30E-62	5	6,915
Instrument B							37	4,746
SEC23A	rs8018720	C	G	0.82	-0.03	4.04E-36	3	391
AMDHD1	rs10859995	C	T	0.58	-0.04	7.03E-89	8	983
Instrument C*							5	600

*Characteristics of all SNPs in instrument C are shown in Supplemental Table 1

Beta-vitD = beta coefficient for association of an additional effect allele with a SD change in serum vitamin D in the log scale; EA = effect allele; EAF = effect allele frequency; F-statistic = measure of instrument strength, calculated as described in the methods, shown for approximate minimum and maximum sample sizes across the COVID-19 outcomes; NEA = non-effect allele; P-vitD = P-value for the beta coefficient; SNP = single nucleotide polymorphism

Table 3. Mendelian Randomization (MR) estimates of effect of vitamin D on COVID-19 outcomes

MR method	COVID vs. population		Hosp COVID vs. population		Severe Resp COVID vs. population		COVID pos vs. COVID neg		Hosp COVID vs. non-hosp COVID	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Instrument A										
IVW	1.04	0.92-1.18	1.05	0.84-1.31	0.96	0.64-1.43	1.15	0.99-1.35	1.44	0.75-2.78
MR-Egger intercept test	p-value = 0.57		p-value = 0.19		p-value = 0.14		p-value = 0.88		p-value = 0.14	
Instrument B										
IVW	1.04	0.92-1.18	1.15	0.92-1.44	1.01	0.71-1.43	1.15	0.98-1.34	1.42	0.86-2.35
MR-Egger intercept test	p-value = 0.62		p-value = 0.15		p-value = 0.06		p-value = 0.61		p-value = 0.14	
Instrument C										
IVW	1.02	0.90-1.15	1.12	0.92-1.37	1.05	0.82-1.34	1.05	0.92-1.20	1.23	0.85-1.77
MR-Egger intercept test	p-value = 0.88		p-value = 0.85		p-value = 0.23		p-value = 0.12		p-value = 0.14	
Instrument of risk of vitamin D deficiency										
IVW	1.00	0.99-1.01	1.00	0.98-1.02	1.02	0.99-1.05	0.99	0.97-1.00	0.97	0.90-1.05
MR-Egger intercept test	p-value = 0.95		p-value = 0.61		p-value = 0.92		p-value = 0.46		p-value = 0.21	
Instrument of risk of vitamin D insufficiency										
IVW	1.00	0.99-1.01	1.00	0.98-1.02	1.06	1.00-1.05	0.99	0.98-1.00	0.97	0.90-1.05
MR-Egger intercept test	p-value = 0.87		p-value = 0.64		p-value = 0.76		p-value = 0.59		p-value = 0.22	

OR = odds ratio representing change in risk of outcome per SD increase in log-transformed serum vitamin D (Instruments A, B and C) or difference in risk of outcome for individuals at risk of vitamin D deficiency/insufficiency vs. individuals with sufficient vitamin D; 95% CI, 95% confidence interval

Figure 1) Schematic of two-sample MR study design

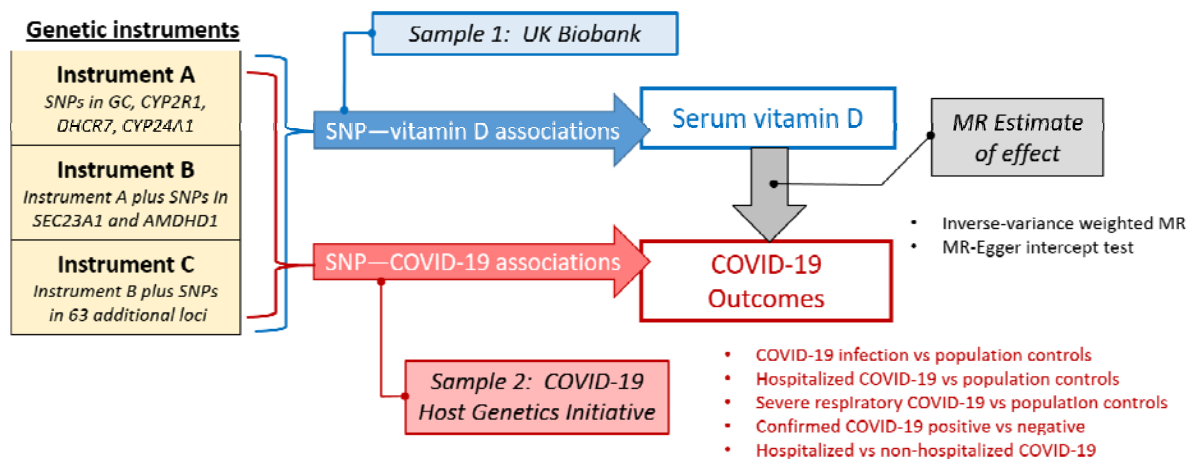


Figure 2) Inverse variance weighted MR estimates for effect of serum vitamin D on risk of COVID-19 outcomes

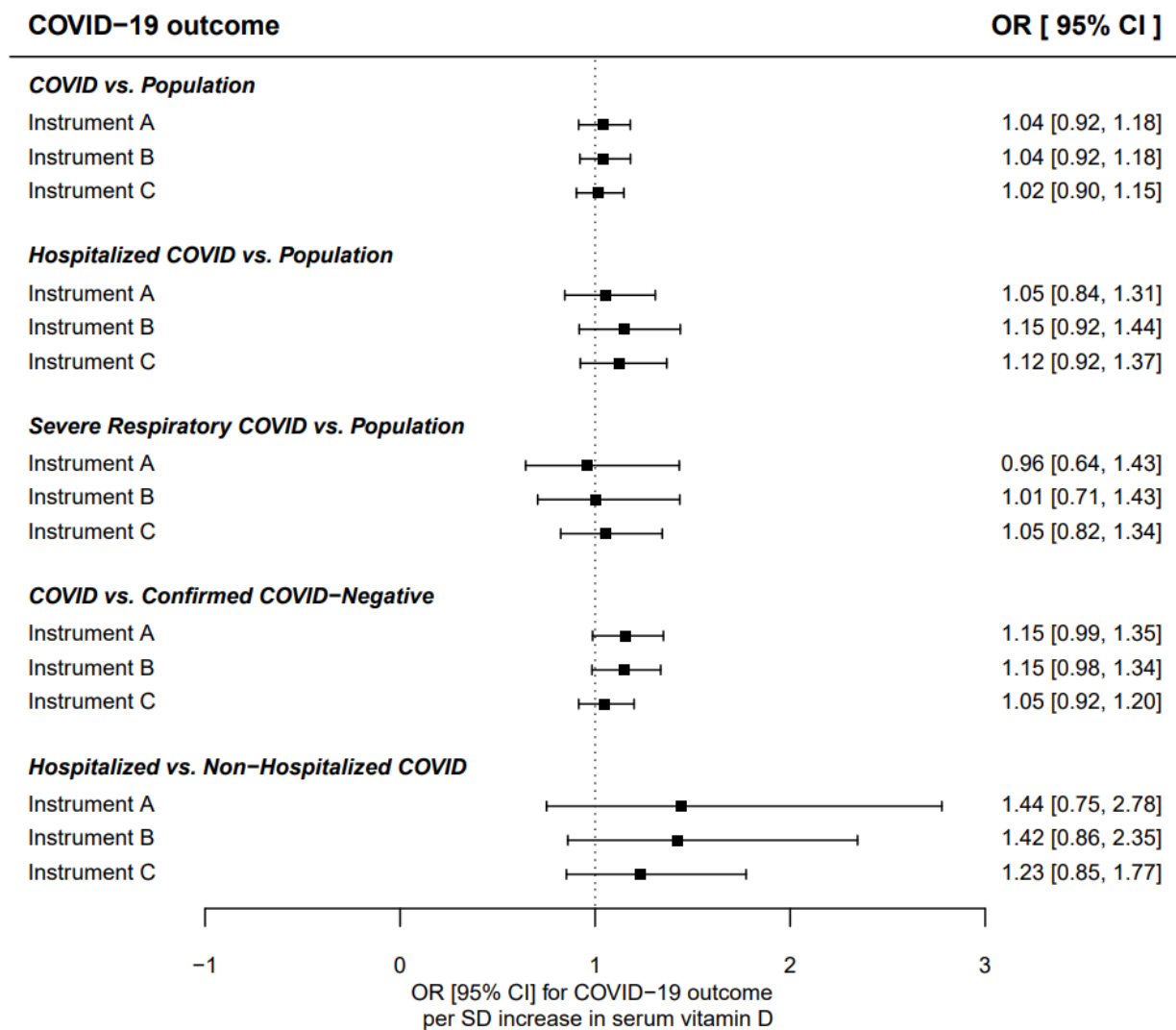


Figure 3) Inverse variance weighted MR estimates for effect of vitamin D insufficiency or deficiency on risk of COVID-19 outcomes

