THE LANCET
Rheumatology

Supplementary appendix

This appendix formed part of the original submission. We post it as supplied by the authors.

Supplement to: Bovijn J, Lindgren CM, Holmes MV. Genetic variants mimicking therapeutic inhibition of IL-6 receptor signaling and risk of COVID-19.
Lancet Rheumatol 2020; published online Sept 25. https://doi.org/10.1016/S2665-9913(20)30345-3.

Supplementary Appendix for "Genetic variants mimicking therapeutic inhibition of interleukin-6 receptor signaling and risk of COVID-19"

Supplementary Methods 2

Supplementary Figure 1. Direction of effects on circulating biomarkers for pharmacological inhibition of IL-6R and *IL6R* variants. 4

Supplementary Figure 2. Association of CRP-lowering *IL6R* variants with SARS-CoV-2-related outcomes. 6

Supplementary Figure 3. Scaled association of *IL6R* instruments with rheumatoid arthritis and risk of hospitalization for confirmed SARS-CoV-2 infection, using different instruments. 7

Supplementary Table 1. Pairwise linkage disequilibrium (in r^2) between seven *IL6R* variants. 8

Supplementary Table 2. Pairwise linkage disequilibrium (in D') between seven *IL6R* variants. 9

Supplementary Table 3. Definitions of outcomes used by The Covid-19 Host Genetics Initiative. 10

Supplementary References 11

Supplementary Methods

Data-sources

Clinical trials of interleukin-6 receptor (IL-6R) blockers. We identified the number of actively recruiting trials for IL-6R blockers by searching *ClinicalTrials.gov* with the following search terms; Condition:"covid", Other:("sarilumab OR tocilizumab"), with filter on "recruiting". This yielded 43 results on 06 July 2020.

IL6R instrument. We used an instrument consisting of 7 variants located in, and up to 300 kilobases (kb) on either side of, *IL6R*. This instrument was recently shown to be associated with lower C-reactive protein (CRP) and fibrinogen, and higher IL-6 and soluble IL-6R, recapitulating the effects of pharmacological IL-6R blockade (Supplementary Figure 1).[1] We also examined a range of alternative instruments to evaluate the robustness of our findings (see *Sensitivity Analyses* below).

CRP. We used the largest available genome-wide association study (GWAS) dataset for CRP concentration, derived in UK Biobank (UKBB).[2] We used the summary statistics derived in the combined male and female analysis and using the inverse-rank normalized transformed trait values.

Rheumatoid arthritis. We used an independent GWAS dataset including up to 14,361 rheumatoid arthritis cases and 43,923 controls of European ancestry.[3] Summary statistics were downloaded from: http://plaza.umin.ac.jp/~yokada/datasource/software.htm.

Coronary heart disease. We used an independent GWAS dataset including up to 60,801 coronary heart disease cases and 123,504 controls from the CARDIoGRAMplusC4D consortium.[4] Summary statistics were downloaded from: http://www.cardiogramplusc4d.org/data-downloads/.

SARS-CoV-2 outcomes. We examined the effect of the *IL6R* variants on SARS-CoV-2 outcomes using data from The COVID-19 Host Genetics Initiative GWAS meta-analyses round 3, the largest available GWAS datasets for COVID-19.[5] Data were available for various SARS-CoV-2 outcomes (definitions given in Supplementary Table 3). Details pertaining to the analysis performed by the COVID-19 Host Genetics Initiative are available at https://www.covid19hg.org/about/.

Analyses

We first confirmed the validity of the instrument by examining each variant's association with CRP (a biomarker of IL-6R inhibition) in UKBB. We subsequently scaled all estimates relating to further outcomes of interest to a 0.1 SD lower CRP (approximately equivalent to the effect size of the missense variant, rs2228145). To do this, we derived the allelic effect estimate for the association of each variant with CRP in UKBB. We then derived a scaling factor for each variant by dividing 0.1 by the variant's allelic effect estimate for CRP (in SD units) and then multiplying all per-allele effect estimates [log(OR) and the standard error of log(OR)] pertaining to outcomes by this scaling factor. We meta-analysed scaled estimates for each outcome using fixed-effect meta-analysis, implemented in the *metafor* R library.[6]

We examined the instrument's association with two positive controls: rheumatoid arthritis (given the established indication of IL-6R blockers) and coronary heart disease (which has been previously linked to genetic variation in *IL6R*[7]). Our primary outcome of interest was risk of COVID-19 hospitalization, using both confirmed-negative controls and population-based controls (see Supplementary Table 3). We then investigated further outcomes related to SARS-CoV-2 infection.

Sensitivity Analyses

We repeated our main analysis (association with rheumatoid arthritis and risk of hospitalization for COVID-19) using different sets of *IL6R* variants, to evaluate the robustness of our findings.

Excluding variants outside IL6R. We excluded all variants located outside the *IL6R* transcript boundary (chr1:154,377,669-154,441,926) from the instrument.

Excluding variants in moderate D'. We checked pairwise r^2 and D' between all variants using the *LDLinkeR* R package,[8] using a 1000 Genomes CEU reference panel. Pairwise r^2 was low (≤ 0.1) for all pairs (Supplementary Table 1). Some pairs showed moderately high D' (Supplementary Table 2). For each pair with D' > 0.5, we

therefore also excluded the variant with the weakest effect on CRP (based on the *P*-value for association with CRP in UKBB). This led to the exclusion of three variants from the instrument (rs12059682, rs4556348, rs73026617).

Identifying an alternative IL6R instrument in UKBB. We extracted all variants in *IL6R* ± a 500 kb region on either side of *IL6R* and identified the variant with the strongest association with CRP in UKBB (rs12133641, an intronic *IL6R* variant). We then ran locus-specific conditional analysis (using the GCTA-COJO "--cojo-cond" function[9], with genotype data from UKBB as an LD reference panel), conditioning on this lead variant. We repeated an iterative process of conditioning until no further genome-wide significant ($P < 5 \times 10^{-8}$ after conditioning on other independent variants) variants could be identified in this window. We selected variants as instruments if their conditional *P*-value and their *P*-value in the original (i.e. unconditioned) dataset were both less than 5×10^{-8}. This identified 2 independent variants (rs12133641 and rs12083537, with pairwise r^2 and D' of 0.0001 and 0.02, respectively).

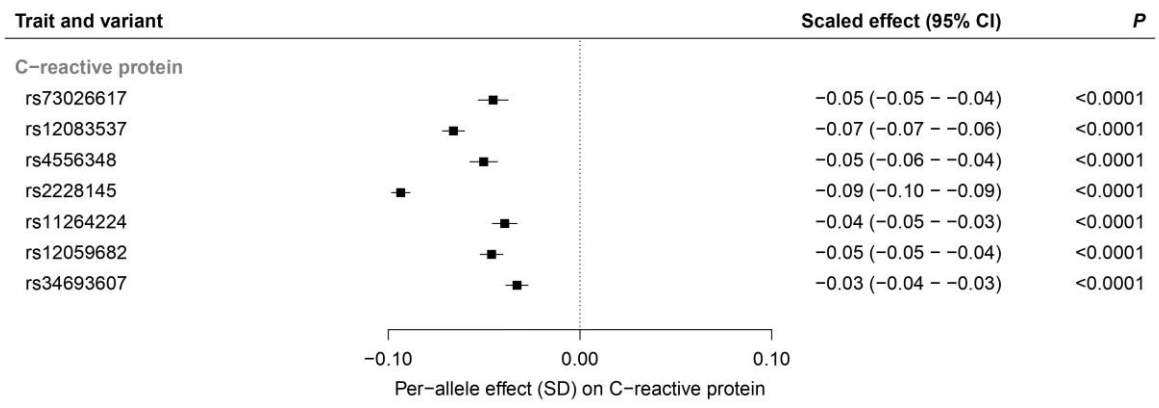
The effect of the missense variant rs2228145. We studied the effect of just the rs2228145 *IL6R* missense variant. This variant has previously been found to lead to impaired classical IL-6R signaling,[1,7,10] mimicking the effect of pharmacological IL-6R inhibition.

Acknowledgements
We express our gratitude to the studies and consortia (including the Neale lab, the COVID-19 Host Genetics Initiative, the CARDIoGRAMplusC4D consortium and Okada et al.), in particular their participants and their investigators, for providing access to summary statistics.

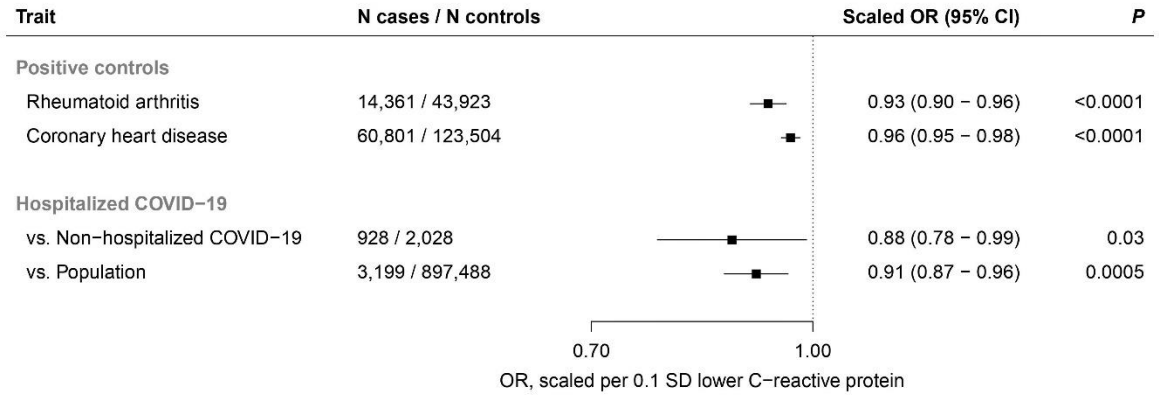
Circulating biomarker	Pharmacological inhibition of IL-6R	*IL6R* variants
IL-6	↑	↑
sIL-6R	↑	↑
CRP	↓	↓
Fibrinogen	↓	↓

Supplementary Figure 1. Direction of effects on circulating biomarkers for pharmacological inhibition of IL-6R and *IL6R* variants.
Figure adapted from Swerdlow *et al*[7] and Georgakis *et al.*[1] CRP, C-reactive protein; IL-6, interleukin-6; sIL-6R, soluble interleukin-6 receptor.

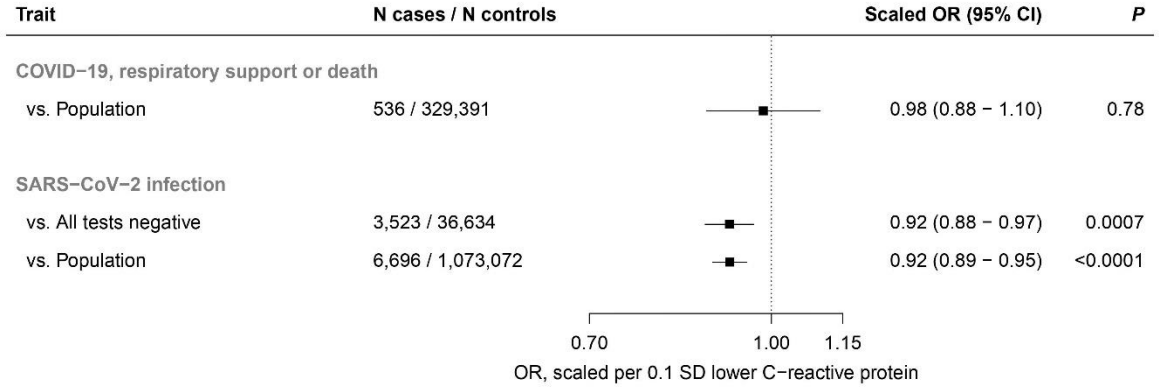
A


Trait and variant
Scaled effect (95% CI)
P
C−reactive protein
rs73026617 −0.05 (−0.05 − −0.04) <0.0001
rs12083537 −0.07 (−0.07 − −0.06) <0.0001
rs4556348 −0.05 (−0.06 − −0.04) <0.0001
rs2228145 −0.09 (−0.10 − −0.09) <0.0001
rs11264224 −0.04 (−0.05 − −0.03) <0.0001
rs12059682 −0.05 (−0.05 − −0.04) <0.0001
rs34693607 −0.03 (−0.04 − −0.03) <0.0001
−0.10
0.00
0.10
Per−allele effect (SD) on C−reactive protein

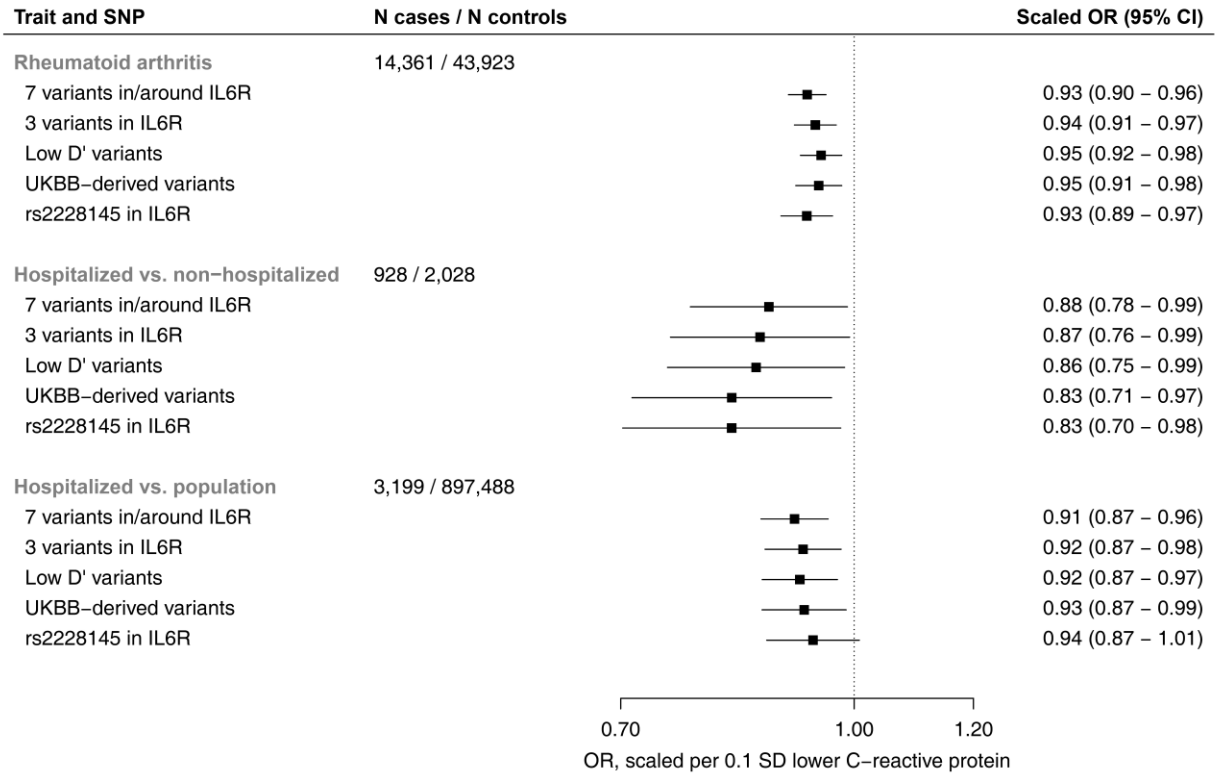
B


Trait
N cases / N controls
Scaled OR (95% CI)
P
Positive controls
Rheumatoid arthritis 14,361 / 43,923 0.93 (0.90 − 0.96) <0.0001
Coronary heart disease 60,801 / 123,504 0.96 (0.95 − 0.98) <0.0001
Hospitalized COVID−19
vs. Non−hospitalized COVID−19 928 / 2,028 0.88 (0.78 − 0.99) 0.03
vs. Population 3,199 / 897,488 0.91 (0.87 − 0.96) 0.0005
0.70
1.00
OR, scaled per 0.1 SD lower C−reactive protein

C


Trait
N cases / N controls
Scaled OR (95% CI)
P
COVID−19, respiratory support or death
vs. Population 536 / 329,391 0.98 (0.88 − 1.10) 0.78
SARS−CoV−2 infection
vs. All tests negative 3,523 / 36,634 0.92 (0.88 − 0.97) 0.0007
vs. Population 6,696 / 1,073,072 0.92 (0.89 − 0.95) <0.0001
0.70
1.00
1.15
OR, scaled per 0.1 SD lower C−reactive protein

Supplementary Figure 2. Association of CRP-lowering *IL6R* variants with SARS-CoV-2-related outcomes. Panel A shows the per-allele association of 7 independent CRP-lowering variants in or near *IL6R* with CRP in UKBB. Panel B shows the scaled association of the *IL6R* instrument with two positive controls (rheumatoid arthritis, given the established indication of IL-6R blockers, and coronary heart disease, which has been previously linked to variants in *IL6R*) and risk of hospitalization for confirmed SARS-CoV-2 infection. Risk of hospitalization was assessed using a control group of individuals with confirmed SARS-CoV-2 infection and no hospitalization at 21 days post-test, and a population-based control group of all non-cases. Panel C shows the scaled associations of the *IL6R* instrument with secondary outcomes (COVID-19 complicated by death or requiring respiratory support, and risk of SARS-CoV-2 infection, as defined by laboratory confirmed SARS-CoV-2 infection and/or a confirmed diagnosis of COVID-19). See appendix p 10 for definitions of outcomes used in The COVID-19 Host Genetics Initiative. OR, odds ratio; CI, confidence interval.



Supplementary Figure 3. Scaled association of *IL6R* instruments with rheumatoid arthritis and risk of hospitalization for confirmed SARS-CoV-2 infection, using different instruments.
"7 variants in/around *IL6R*" refers to instrument used in the main analysis, consisting of seven variants located in, and up to 300 kb around, the *IL6R* gene; "3 variants in *IL6R*" refers to subset of three variants (from the 7 variants) located within the *IL6R* gene; "Low D' variants" refers to instrument with exclusion of variants with high pairwise D' (D' > 0.5) ; "UKBB-derived variants" refers to an *IL6R* instrument (consisting of 2 independent variants) derived using UKBB CRP data; "rs2228145 in *IL6R*" refers to an *IL6R* missense variant previously shown to impede classical IL-6R signaling.[11]

	rs73026617	rs12083537	rs4556348	rs2228145	rs11264224	rs12059682	rs34693607
rs73026617	1	0.017	0.1	0.011	0	0.018	0.006
rs12083537	0.017	1	0.014	0	0.008	0.007	0.017
rs4556348	0.1	0.014	1	0.066	0.006	0.007	0.011
rs2228145	0.011	0	0.066	1	0.08	0.076	0.033
rs11264224	0	0.008	0.006	0.08	1	0.052	0.057
rs12059682	0.018	0.007	0.007	0.076	0.052	1	0.001
rs34693607	0.006	0.017	0.011	0.033	0.057	0.001	1

Supplementary Table 1. Pairwise linkage disequilibrium (in r^2) between seven *IL6R* variants.

	rs73026617	rs12083537	rs4556348	rs2228145	rs11264224	rs12059682	rs34693607
rs73026617	1	0.759	0.407	0.426	0.072	0.764	0.479
rs12083537	0.759	1	0.533	0.02	0.396	0.31	0.137
rs4556348	0.407	0.533	1	0.815	0.401	0.099	0.496
rs2228145	0.426	0.02	0.815	1	0.472	0.727	0.273
rs11264224	0.072	0.396	0.401	0.472	1	1	0.265
rs12059682	0.764	0.31	0.099	0.727	1	1	0.132
rs34693607	0.479	0.137	0.496	0.273	0.265	0.132	1

Supplementary Table 2. Pairwise linkage disequilibrium (in D$^{'}$) between seven *IL6R* variants.

Name	Case definition	Control definition	Figure presented
Hospitalized COVID-19 vs. Non-hospitalized COVID-19	Hospitalized, laboratory confirmed SARS-CoV-2 infection (RNA and/or serology based), AND Hospitalization due to Covid-19-related symptoms.	Laboratory confirmed SARS-CoV-2 infection (RNA and/or serology based), AND Not hospitalized 21 days after the test.	Supplementary Figure 2; Supplementary Figure 3
Hospitalized COVID-19 vs. Population		Everybody that is not a case, e.g. population.	Supplementary Figure 2; Supplementary Figure 3
Covid-19, respiratory support or death vs. Population	Hospitalized, laboratory confirmed SARS-CoV-2 infection (RNA and/or serology based), AND Death OR respiratory support, including intubation, CPAP, BiPAP, CNP, HFNO/very high flow PEEP oxygen (but excluding simple supplementary oxygen), AND hospitalization with Covid-19 as primary reason for admission.	Everybody that is not a case, e.g. population	Supplementary Figure 2
SARS-CoV-2 infection vs. All tests negative	Laboratory confirmed SARS-CoV-2 infection (RNA and/or serology based) OR EHR/ICD coded/physician confirmed Covid-19 OR self-reported Covid-19 positive (e.g. by questionnaire)	Laboratory tested for SARS-CoV-2 infection (RNA and/or serology based) AND All tests negative, OR self-reported tested negative for SARS-CoV-2 infection (e.g. by questionnaire)	Supplementary Figure 2
SARS-CoV-2 infection vs. Population		Everybody that is not a case, e.g. population	Supplementary Figure 2
BiPAP, Bilevel Positive Airway Pressure; CNP, Continuous external negative pressure; CPAP, Continuous positive airway pressure; EHR, Electronic health record; HFNO, High-flow nasal oxygen therapy; ICD, International Classification of Diseases; PEEP, Positive-end expiratory pressure.			

Supplementary Table 3. Definitions of outcomes used by The Covid-19 Host Genetics Initiative.
Definitions are given as per the following URL: Phenotype definitions for analyses v2.0 (https://docs.google.com/document/d/1okamrqYmJfa35ClLvCt_vEe4PkvrTwggHq7T3jbeyCI/edit).

Supplementary References

1 Georgakis MK, Malik R, Gill D, *et al*. Interleukin-6 Signaling Effects on Ischemic Stroke and Other Cardiovascular Outcomes: A Mendelian Randomization Study. *Circ Genom Precis Med.* 2020; **13**: e002872. https://doi.org/10.1161/CIRCGEN.119.002872

2 Neale lab. UK Biobank GWAS results. http://www.nealelab.is/uk-biobank. Data accessed: June 05, 2020.

3 Okada Y, Wu D, Trynka G, *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature.* 2014; **506**: 376–81.

4 The CARDIoGRAMplusC4D Consortium. A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nat Genet.* 2015; **47**: 1121–30.

5 The COVID-19 Host Genetics Initiative. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur J Hum Genet.* 2020; **28**: 715–8.

6 Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw.* 2010; **36**. https://doi.org/10.18637/jss.v036.i03.

7 Interleukin-6 Receptor Mendelian Randomisation Analysis (IL6R MR) Consortium, Swerdlow DI, Holmes MV, *et al.* The interleukin-6 receptor as a target for prevention of coronary heart disease: a Mendelian randomisation analysis. *Lancet.* 2012; **379**: 1214–24.

8 Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics.* 2015; **31**: 3555–7.

9 Yang J, Ferreira T, Morris AP, *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet.* 2012; **44**: 369–75.

10 Haycock PC, Burgess S, Wade KH, Bowden J, Relton C, Davey Smith G. Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. *Am J Clin Nutr.* 2016; **103**: 965–78.

11 Ferreira RC, Freitag DF, Cutler AJ, *et al.* Functional IL6R 358Ala allele impairs classical IL-6 receptor signaling and influences risk of diverse inflammatory diseases. *PLoS Genet.* 2013; **9**: e1003444. https://doi.org/10.1371/journal.pgen.1003444.