Journal of
Clinical Medicine

MDPI

Article

A Phenome-Wide Association Study (PheWAS) of COVID-19 Outcomes by Race Using the Electronic Health Records Data in Michigan Medicine

Maxwell Salvatore [1,2,3], Tian Gu [1,2], Jasmine A. Mack [1], Swaraaj Prabhu Sankar [2,4,5], Snehal Patil [1,2], Thomas S. Valley [6,7,8], Karandeep Singh [8,9], Brahmajee K. Nallamothu [7,10], Sachin Kheterpal [8,11], Lynda Lisabeth [3], Lars G. Fritsche [1,2,4,12] and Bhramar Mukherjee [1,2,3,*]

1 Department of Biostatistics, University of Michigan School of Public Health, 1415 Washington Heights, Ann Arbor, MI 48109, USA; mmsalva@umich.edu (M.S.); gutian@umich.edu (T.G.); jasamack@umich.edu (J.A.M.); snehal@umich.edu (S.P.); larsf@umich.edu (L.G.F.)
2 Center for Precision Health Data Science, University of Michigan, Ann Arbor, MI 48109, USA; swarsank@umich.edu
3 Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI 48109, USA; llisabet@umich.edu
4 Rogel Cancer Center, Michigan Medicine, Ann Arbor, MI 48109, USA
5 Data Office for Clinical and Translational Research, University of Michigan, Ann Arbor, MI 41809, USA
6 Division of Pulmonary and Critical Care Medicine, University of Michigan Medicine, Ann Arbor, MI 48109, USA; valleyt@med.umich.edu
7 Department of Internal Medicine, Michigan Medicine, Ann Arbor, MI 48109, USA; bnallamo@med.umich.edu
8 Institute for Healthcare Policy and Innovation, University of Michigan, Ann Arbor, MI 48109, USA; kdpsingh@med.umich.edu (K.S.); sachinkh@med.umich.edu (S.K.)
9 Department of Learning Health Sciences, University of Michigan, Ann Arbor, MI 48109, USA
10 Division of Cardiovascular Medicine, Michigan Medicine, Ann Arbor, MI 48109, USA
11 Department of Anesthesiology, Michigan Medicine, Ann Arbor, MI 48109, USA
12 Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI 48109, USA
* Correspondence: bhramar@umich.edu; Tel.: +1-(734)-764-6544

Abstract: Background: We performed a phenome-wide association study to identify pre-existing conditions related to Coronavirus disease 2019 (COVID-19) prognosis across the medical phenome and how they vary by race. Methods: The study is comprised of 53,853 patients who were tested/diagnosed for COVID-19 between 10 March and 2 September 2020 at a large academic medical center. Results: Pre-existing conditions strongly associated with hospitalization were renal failure, pulmonary heart disease, and respiratory failure. Hematopoietic conditions were associated with intensive care unit (ICU) admission/mortality and mental disorders were associated with mortality in non-Hispanic Whites. Circulatory system and genitourinary conditions were associated with ICU admission/mortality in non-Hispanic Blacks. Conclusions: Understanding pre-existing clinical diagnoses related to COVID-19 outcomes informs the need for targeted screening to support specific vulnerable populations to improve disease prevention and healthcare delivery.

Keywords: biobank; health disparities; EHR; phenome; odds ratio; risk profile

check for updates

Citation: Salvatore, M.; Gu, T.; Mack, J.A.; Prabhu Sankar, S.; Patil, S.; Valley, T.S.; Singh, K.; Nallamothu, B.K.; Kheterpal, S.; Lisabeth, L.; et al. A Phenome-Wide Association Study (PheWAS) of COVID-19 Outcomes by Race Using the Electronic Health Records Data in Michigan Medicine. *J. Clin. Med.* **2021**, *10*, 1351. https://doi.org/10.3390/jcm10071351

Academic Editor: Alessandra Falchi

Received: 10 February 2021
Accepted: 17 March 2021
Published: 25 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The emergence of electronic health records (EHR) and rise of EHR-linked biobanks have made it possible for researchers to explore omics-based relationships agnostically on a large scale instead of targeted hypothesis testing. Introduced by Denny et al. in 2010, a phenome-wide association study (PheWAS) is an omnibus scan to identify gene–disease associations across the medical phenome [1]. PheWAS typically associates a genetic variant (G) with hundreds of disease codes (D_j, $j = 1, \ldots, J$) with association models of the structure $logit\left(P\left(D_j \middle| G, Confounders\right)\right) = \beta_{0j} + \beta_{Gj}G + \beta_C Confounders$. Due to

computational advances and development of widely available analytic frameworks [2–6], PheWAS is now relatively easy to implement. The main goal of a PheWAS is to replicate known gene–disease relationships and to search for hidden and unanticipated associations (for example, Li et al. found that there is a strong negative association between the tag single nucleotide polymorphism (SNP) for blood group O antigen and arm impedance) [2,7–10].

As of 15 January 2021, there were 23,759,743 confirmed COVID-19 cases in the US [11], representing approximately 25% of all global cases. Because COVID-19 is a respiratory disease and produces flu-like symptoms, the testing strategies in the US initially focused on those with symptoms, the elderly, and those with pre-existing conditions [12]—i.e., populations who are at risk of severe disease and complications. Only a handful of pre-existing comorbidities are known to be associated with experiencing adverse COVID-19-related outcomes. These pre-existing conditions include liver, kidney, heart, and respiratory disease [2,13–17].

There has been a remarkable surge within the academic and medical communities to conduct rapid research on COVID-19 [18]. There have been a number of studies examining differences across racial groups for an ensemble of COVID-19-associated conditions and outcomes in US patient cohorts [13,19–25]. Instead of a hypothesis-driven approach a priori restricted to certain disease categories, this study applied an agnostic *disease–disease* PheWAS framework to COVID-19 outcomes in a cohort of 53,853 patients who were tested or diagnosed with COVID-19 at a large academic medical center. We looked at correlates of disease prognosis among all COVID-19 patients as well as separately among non-Hispanic White (White) and non-Hispanic Black/African American (Black) patients. The primary objective of this study was to agnostically identify pre-existing conditions present in an individual's medical record that may be associated with hospitalization, intensive care unit (ICU) admission, and mortality. We also present the results from race-stratified susceptibility PheWAS to predict who tests positive for COVID-19 in the Supplementary Materials. Our reason to downplay the outcome of who gets COVID-19 or who tests positive for COVID-19 is due to the prioritized testing strategy that makes this tested sample highly non-representative of the population. A naïve comparison of the positive versus negative test results is highly biased [13]. However, conditional on testing or being diagnosed positive, downstream prognostic outcomes are less prone to such selection biases and we primarily focus on these outcomes.

2. Materials and Methods

2.1. Study Design

COVID-19 Cohort

We extracted the EHR for patients who were tested for COVID-19 at the University of Michigan Health System, also known as Michigan Medicine (MM), from 10 March 2020 to 2 September 2020. A total of 53,260 patients (98.9%) who were tested at MM and 593 patients (1.1%) who were treated for COVID-19 in MM, but tested elsewhere, constituted our initial study cohort of 53,853 patients, of whom 2582 tested positive. Our analytic cohort was restricted to those individuals on whom we have EHR data for at least 14 days prior to the first COVID-19 test. This restriction is used to eliminate symptoms which may indicate manifestation of underlying COVID-19 disease or symptoms, whereas our goal was to search for truly "pre-existing" conditions prior to COVID-19 testing/diagnosis. Our resulting analytic cohort comprised 47,862 tested/diagnosed patients, of whom 2133 tested positive. Since the testing protocol in MM [26] focused on prioritized testing based on symptoms, exposure, occupation and other patient level factors, this is a non-representative sample of the population. Study protocols were reviewed and approved by the University of Michigan Medical School Institutional Review Board (IRB ID HUM00180294).

2.2. Data Processing

2.2.1. Classifying Patients Who Were Still in Hospital and ICU to Define COVID-19 Outcomes

We categorized COVID-19-positive patients into non-hospitalized, hospitalized (includes ICU stays), and hospitalized with ICU stay based on the admission and discharge

data. A total of 22 patients were still admitted in the hospital at the time of data extraction (17 had at least one ICU stay and five had no ICU stay).

2.2.2. Generation of the Medical Phenome

We constructed the medical phenome by extracting available International Classification of Diseases (ICD; ninth and tenth editions) codes from EHR and grouping them into 1813 traits using the PheWAS R package (as described in [1]). Each of these traits (PheWAS codes) was coded as a binary risk factor (present/absent) and used as a predictor in the association models with COVID-19 outcomes. As mentioned before, to differentiate *pre-existing* conditions from phenotypes related to COVID-19 testing/treatment, we applied a 14-day-prior restriction on the tested cohort by removing diagnoses that first appeared within the 14 days before the first test or diagnosis date, whichever was earlier. The analyses in this study were restricted to 1363 traits that appeared in the EHR 14-day-prior of at least ten COVID-19-positive patients. While the PheWAS is performed on PheWAS codes, one can view the mapping of ICD-to-PheWAS code relationships on this website: https://prsweb.sph.umich.edu:8443/phecodeData/searchPhecode (accessed on 22 September 2020).

2.2.3. Description of Variables

A summary data dictionary is available with the source and definition of each variable used in our analysis (Table S1A in Supplement).

2.3. Statistical Analysis

We performed PheWAS to identify predictors of three COVID-19 prognostic outcomes in this study (detailed definition in Table S1B in the Supplement), among those who were diagnosed/tested positive, comparing:

(i) those who were hospitalized with those who were not;
(ii) those who were admitted to ICU or died with those who were not;
(iii) those who died with those who were alive at the time of data extraction.

We also present results from the susceptibility PheWAS (comparing those who were diagnosed with COVID-19 with those who were not tested at all [matched controls]) in the Supplementary Materials. All COVID-19 outcomes of interest are binary; thus, logistic regression was our primary tool for association analysis. All logistic regression models were of the following form:

$$\text{logit}\, p(Y_{COVID} = 1 | \text{Covariates}, \text{PheCode}_j) = \beta_0 + \beta_{Cov}^{\top}\text{Covariates} + \beta_j I[\text{Phecode}_j = 1] \quad (1)$$

$j = 1, \ldots, 1363$. Here Y_{COVID} is various COVID-19-related outcomes under consideration (e.g., COVID-19 hospitalization, ICU admission, and mortality). The Firth correction was used to address potential separation issues in logistic regression models. For all models, adjusted odds ratios (OR), 95% Wald-type confidence interval and *p*-values were presented [27–29]. Full models were adjusted for age, sex, race, and the neighborhood deprivation index (NDI). The NDI is defined by US census tract (corresponding to the residential address available in each patient's EHR) for the year 2010 and are from the National Neighborhood Data Archive (NaNDA) [30]. PheWAS adjusting for an additional comorbidity score (indicating whether the patient was diagnosed with conditions across seven disease categories associated with COVID-19 susceptibility and adverse outcomes: respiratory, circulatory, any cancer, type II diabetes, kidney, liver, and autoimmune; ranges from 0 to 7) is included as a sensitivity analysis on our accompanying website: https://cphds.sph.umich.edu/covidphewas/ (accessed on 10 March 2021). Results for PheWAS analysis are easier to visualize when $-\log_{10}$(*p*-values) corresponding to each of the 1363 tests are plotted against the disease codes grouped into disease categories. We use this visualization tool to present our analysis while all detailed summary results are available at the website above and in the online supplement.

Race-Stratified Analysis

Since the prognostic factors could potentially be different across races, we repeated the entire analysis stratified by race. We restricted our attention to Whites and Blacks due to limitations of sample size for other racial groups. Table S2 contains descriptive statistics stratified by race. We checked for the equality of the log(OR) corresponding to Whites and Blacks through a Wald test for the difference of the log(OR). A conservative Bonferroni multiple testing correction was implemented to conclude statistically significant results ($p = 0.05$/number of tests in analysis), and $p < 0.05$ was used as a threshold for suggestive traits.

3. Results

There were 53,853 patients who were either tested for or diagnosed with COVID-19 who were eligible for inclusion in this study. Of those eligible for inclusion, our study population comprised 47,862 individuals (n_{tested} = 47,862 [$n_{positive}$ = 2133]) who had available ICD code data after applying the 14-day-prior to testing restriction to the EHR. Furthermore, a total of 1813 qualified ICD-code-based phenotypes, referred to as PheWAS codes, were initially screened, of which 1363 had at least 10 occurrences in our COVID-19-positive cohort and were included in the analysis.

Of those 53,853 who were tested for COVID-19, 44.2% (23,814) were males and the median age was 47 years. The majority were White (72.4% (38,977)), while 10.7% were Black (5763). We note that the Black cohort is both younger and more female than the White cohort, a trend that also appears in the tested and hospitalized cohorts (Table S2). Similarly, Blacks tend to have more autoimmune disease, kidney disease, type 2 diabetes, and circulatory disease diagnoses, while Whites tend to have more cancer diagnoses (note that our definition of cancer includes skin cancer) in our tested cohort (Table S2). Out of the study cohort, 4.8% (2582) were tested positive (Table 1). Among the 2582 positive patients, 54.6% (1411) were White, 25.0% (646) were Black, 27.8% (719) were hospitalized, 14.6% (377) were admitted to ICU and 5.0% (129) died. A flowchart describing the sample sizes of the overall cohort and race-specific cohorts by COVID-19 outcome is included in Figure S1.

3.1. Phenome-Wide Comorbidity Association Analysis

The association results for the top 50 traits from the comorbidity PheWAS can be found in Tables S3–S6 for the full cohort, Whites, and Blacks, side-by-side. Interactive versions of the PheWAS plots are online at https://cphds.sph.umich.edu/covidphewas/ (accessed on 10 March 2021). This online resource also provides tables with the adjusted odds ratios, 95% confidence intervals, *p*-values, and counts of occurrence in cases and controls for all traits included in the PheWAS performed.

3.1.1. Full Cohort Prognostic Associations

As the disease outcome progresses (from hospitalized to ICU, and to deceased), stronger associations with circulatory system, genitourinary (renal diseases in particular) and respiratory diseases were observed. Forty-four traits, including 12 circulatory system and 11 respiratory diseases, were phenome-wide significantly associated with hospitalization, as well as an additional 263 suggestive traits under the threshold of $p < 0.05$ (Figure 1A)—respiratory failure, insufficiency, arrest ($p = 3.98 \times 10^{-20}$), acute renal failure ($p = 6.31 \times 10^{-13}$), viral pneumonia ($p = 2.51 \times 10^{-11}$), and acid-base balance disorder ($p = 2.40 \times 10^{-10}$). Moreover, 58 phenome-wide significant hits (e.g., respiratory failure, insufficiency, arrest [$p = 1.58 \times 10^{-15}$], acid-base balance disorder [$p = 3.98 \times 10^{-14}$], and hypotension [$p = 1.58 \times 10^{-11}$]) as well as 286 suggestive hits were noted for association with ICU admission/mortality (Figure 2A), including 77 circulatory system, 36 endocrine/metabolic, 35 genitourinary, and 31 respiratory diseases. There were 22 phenome-wide significant traits associated with COVID-19 mortality (Figure 3A), along with an additional 227 suggestive traits under the threshold $p < 0.05$. In addition to 64 circulatory system and 31 endocrine/metabolic diseases, 23 mental disorders stood out as the

third largest disease group associated with mortality, including delirium due to conditions classified elsewhere ($p = 9.33 \times 10^{-7}$), memory loss ($p = 3.98 \times 10^{-4}$) and aphasia ($p = 5.37 \times 10^{-4}$).

Table 1. Descriptive Characteristics of the COVID-19 Tested/Diagnosed cohort at Michigan Medicine (10 March–2 September 2020).

	Individuals, No. (%) [a]					
	Tested for COVID-19					
			Positive Results			
	Overall	Negative Results	Overall	Hospitalized	ICU	Deceased
Variable	**(*n* = 53,853)**	**(*n* = 51,271)**	**(*n* = 2582)**	**(*n* = 719)**	**(*n* = 377)**	**(*n* = 129)**
Age, y						
Mean (SD)	44.8 (23.1)	44.7 (23.2)	47.4 (20)	58.5 (17.6)	58.6 (17.5)	69 (14.3)
Median (IQR)	47 (38)	46 (38)	49 (31)	61 (23)	61 (22)	71 (22)
<18	6895 (12.8)	6768 (13.2)	127 (4.9)	14 (1.9)	10 (2.7)	0 (0)
[18,35)	12,652 (23.5)	12,017 (23.4)	635 (24.6)	65 (9)	33 (8.8)	3 (2.3)
[35,50)	9273 (17.2)	8697 (17)	576 (22.3)	125 (17.4)	56 (14.9)	11 (8.5)
[50,65)	12,116 (22.5)	11,440 (22.3)	676 (26.2)	224 (31.2)	120 (31.8)	33 (25.6)
[65,80)	10,257 (19)	9825 (19.2)	432 (16.7)	209 (29.1)	124 (32.9)	43 (33.3)
≥80	2660 (4.9)	2524 (4.9)	136 (5.3)	82 (11.4)	34 (9)	39 (30.2)
Male Gender	23,814 (44.2)	22,651 (44.2)	1163 (45)	403 (56.1)	233 (61.8)	80 (62)
Primary Care in MM	31,357 (58.2)	29,969 (58.5)	1388 (53.8)	253 (35.2)	128 (34)	35 (27.1)
BMI						
Mean (SD)	29.1 (7.6)	29.1 (7.6)	30.9 (8.4)	32.6 (10.1)	32.9 (11.5)	31.3 (6.9)
<18.5	826 (1.9)	804 (2)	22 (1)	9 (1.3)	4 (1.1)	1 (0.8)
[18.5,25)	12,857 (29.7)	12,357 (30)	500 (22.9)	102 (14.9)	61 (16.9)	17 (13.7)
[25,30)	13,371 (30.8)	12,723 (30.9)	648 (29.7)	211 (30.9)	110 (30.5)	45 (36.3)
≥30	16,291 (37.6)	15,281 (37.1)	1010 (46.3)	361 (52.9)	186 (51.5)	61 (49.2)
Smoking Status						
Never	31,041 (63.2)	29,549 (63)	1492 (68.7)	368 (60.2)	159 (54.6)	30 (39)
Past	13,725 (28)	13,145 (28)	580 (26.7)	219 (35.8)	120 (41.2)	44 (57.1)
Current	4314 (8.8)	4215 (9)	99 (4.6)	24 (3.9)	12 (4.1)	3 (3.9)
Ever	18,039 (36.8)	17,360 (37)	679 (31.3)	243 (39.8)	132 (45.4)	47 (61)
Alcohol consumption	25,894 (68.4)	24,768 (68.6)	1126 (66.2)	261 (63.2)	128 (63.7)	35 (61.4)
Race/ethnicity						
White	38,977 (72.4)	37,566 (73.3)	1411 (54.6)	326 (45.3)	172 (45.6)	56 (43.4)
Black	5763 (10.7)	5117 (10)	646 (25)	265 (36.9)	139 (36.9)	42 (32.6)
Other [b]	4869 (9)	4616 (9)	253 (9.8)	63 (8.8)	21 (5.6)	6 (4.7)
Unknown [c]	4244 (7.9)	3972 (7.7)	272 (10.5)	65 (9)	45 (11.9)	25 (19.4)
NDI, mean (SD)	0.1 (0.07)	0.1 (0.07)	0.12 (0.09)	0.15 (0.1)	0.16 (0.11)	0.16 (0.11)
Population density persons/mile^2	2375.8 (2422.1)	2343.2 (2412.8)	2997.3 (2512.8)	3658.7 (2635)	3826.4 (2675.2)	4128.4 (2770.3)
Respiratory Diseases	34,471 (72)	32,850 (71.8)	1621 (76)	399 (79.6)	205 (81.7)	82 (90.1)
Circulatory Diseases	32,419 (67.7)	30,940 (67.7)	1479 (69.3)	428 (85.4)	218 (86.9)	87 (95.6)
Any Cancer	13,831 (28.9)	13,344 (29.2)	487 (22.8)	164 (32.7)	88 (35.1)	42 (46.2)
Type 2 Diabetes	7841 (16.4)	7409 (16.2)	432 (20.3)	191 (38.1)	107 (42.6)	57 (62.6)
Kidney Diseases	7206 (15.1)	6867 (15)	339 (15.9)	194 (38.7)	119 (47.4)	56 (61.5)
Liver Diseases	4406 (9.2)	4234 (9.3)	172 (8.1)	58 (11.6)	32 (12.7)	14 (15.4)
Autoimmune Diseases	7544 (15.8)	7163 (15.7)	381 (17.9)	109 (21.8)	61 (24.3)	19 (20.9)
Comorbidity score mean (SD)	2.3 (1.5)	2.2 (1.5)	2.3 (1.5)	3.1 (1.6)	3.3 (1.6)	3.9 (1.5)

Abbreviations: BMI, body mass index (calculated as weight in kilograms divided by height in meters squared); COVID-19, coronavirus disease 2019; ICU, intensive care unit; IQR, interquartile range; NDI, 2010 Neighborhood Socioeconomic Disadvantage Index; MM, Michigan Medicine. [a] Percentages are reported as fraction of column totals excluding missing entries. [b] Includes White Hispanic or unknown; Black Hispanic or unknown; Asian Hispanic, non-Hispanic, or unknown; Native American Hispanic, non-Hispanic, or unknown; Pacific Islander Hispanic, non-Hispanic, or unknown; and other Hispanic, non-Hispanic, or unknown. [c] Includes missing race and/or ethnicity.

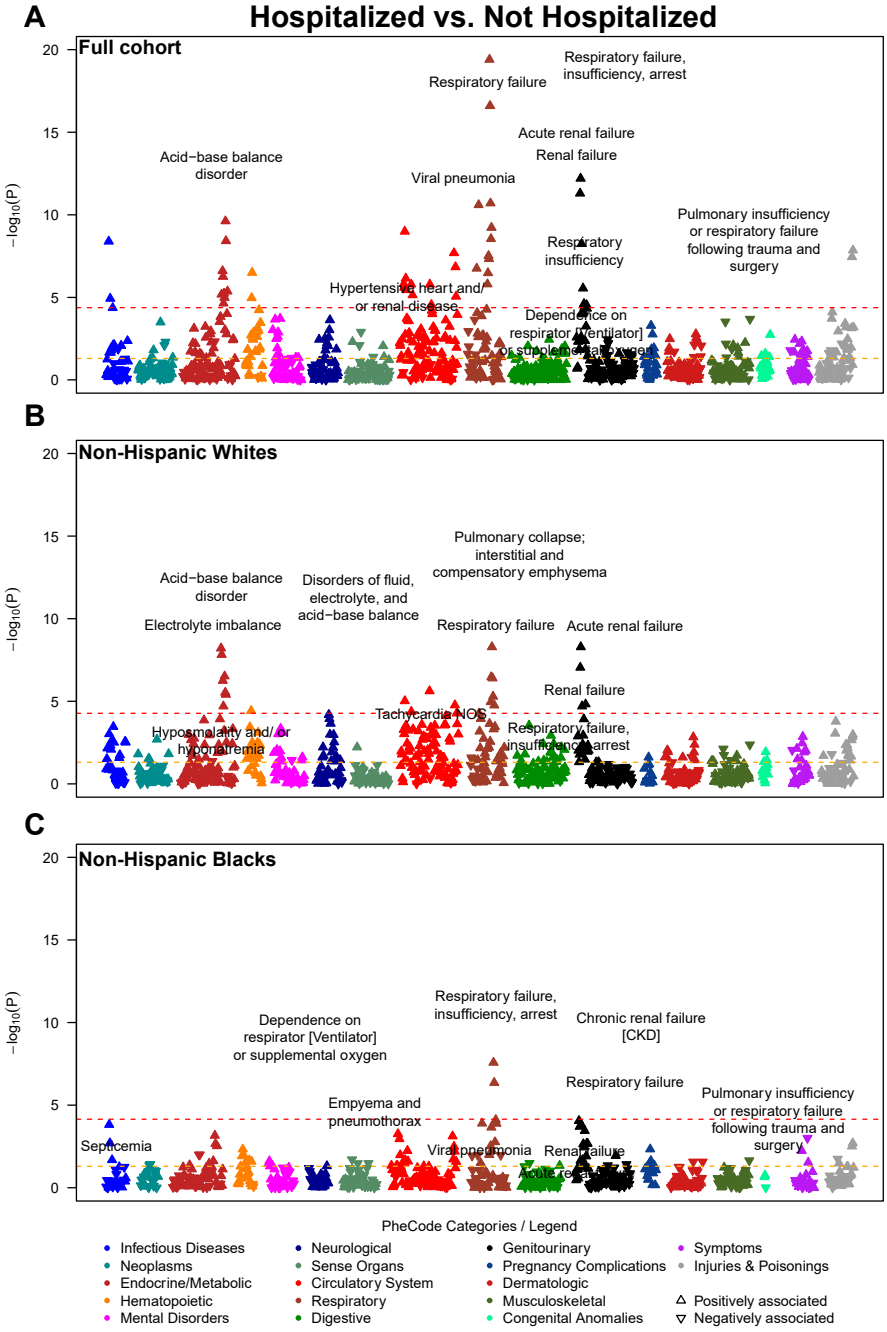

Figure 1. Manhattan plot showing the phenome-wide association between disease conditions and hospitalization for COVID-19. Models are adjusted for age, sex, race (full cohort only), and three census tract-level socioeconomic indicators: proportion with less than high school education, proportion unemployed, and proportion with annual income below the federal poverty level. The *x*-axis are individual disease codes, color-coded by their corresponding disease category as described in the shared legend. The *y*-axis represents the $-\log_{10}$ transformed *p*-value of the association. The dashed, horizontal lines represent the $p = 0.05$ (in orange) and the Bonferroni corrected *p*-value (0.05/number of tests; in red). Each point is represented by either an upward triangle indicating a positive association or a downward triangle indicating a negative association. (**A**): Full cohort, (**B**): Restricted to Whites, (**C**): Restricted to Blacks

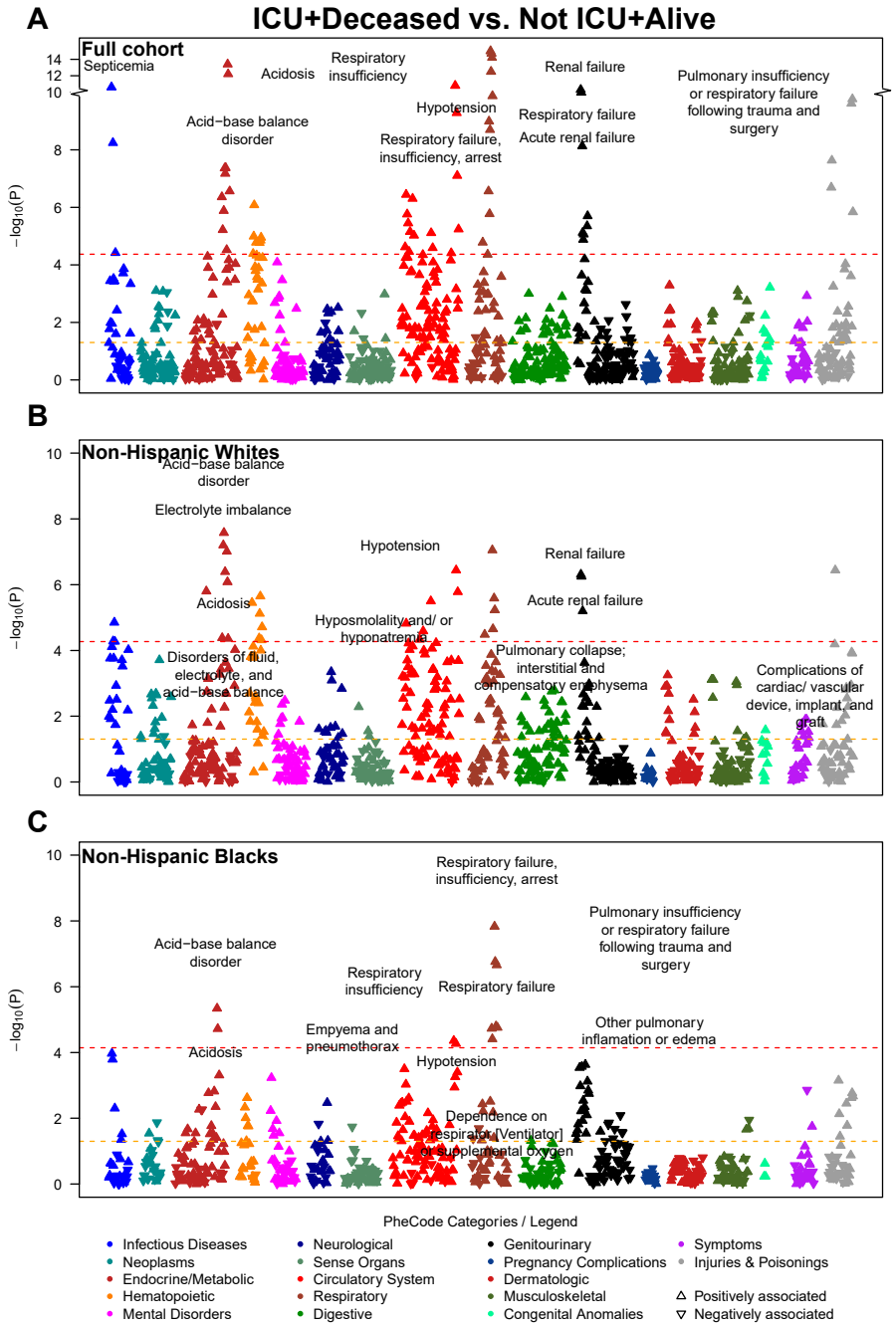

Figure 2. Manhattan plot showing the phenome-wide association between disease conditions and ICU admission for COVID-19. Models are adjusted for age, sex, race (full cohort only), and three census tract-level socioeconomic indicators: proportion with less than high school education, proportion unemployed, and proportion with annual income below the federal poverty level. The *x*-axis are individual disease codes, color-coded by their corresponding disease category as described in the shared legend. The *y*-axis represents the $-\log_{10}$ transformed *p*-value of the association. The dashed, horizontal lines represent the $p = 0.05$ (in orange) and the Bonferroni corrected *p*-value (0.05/number of tests; in red). Each point is represented by either an upward triangle indicating a positive association or a downward triangle indicating a negative association. (**A**): Full cohort, (**B**): Restricted to Whites, (**C**): Restricted to Blacks

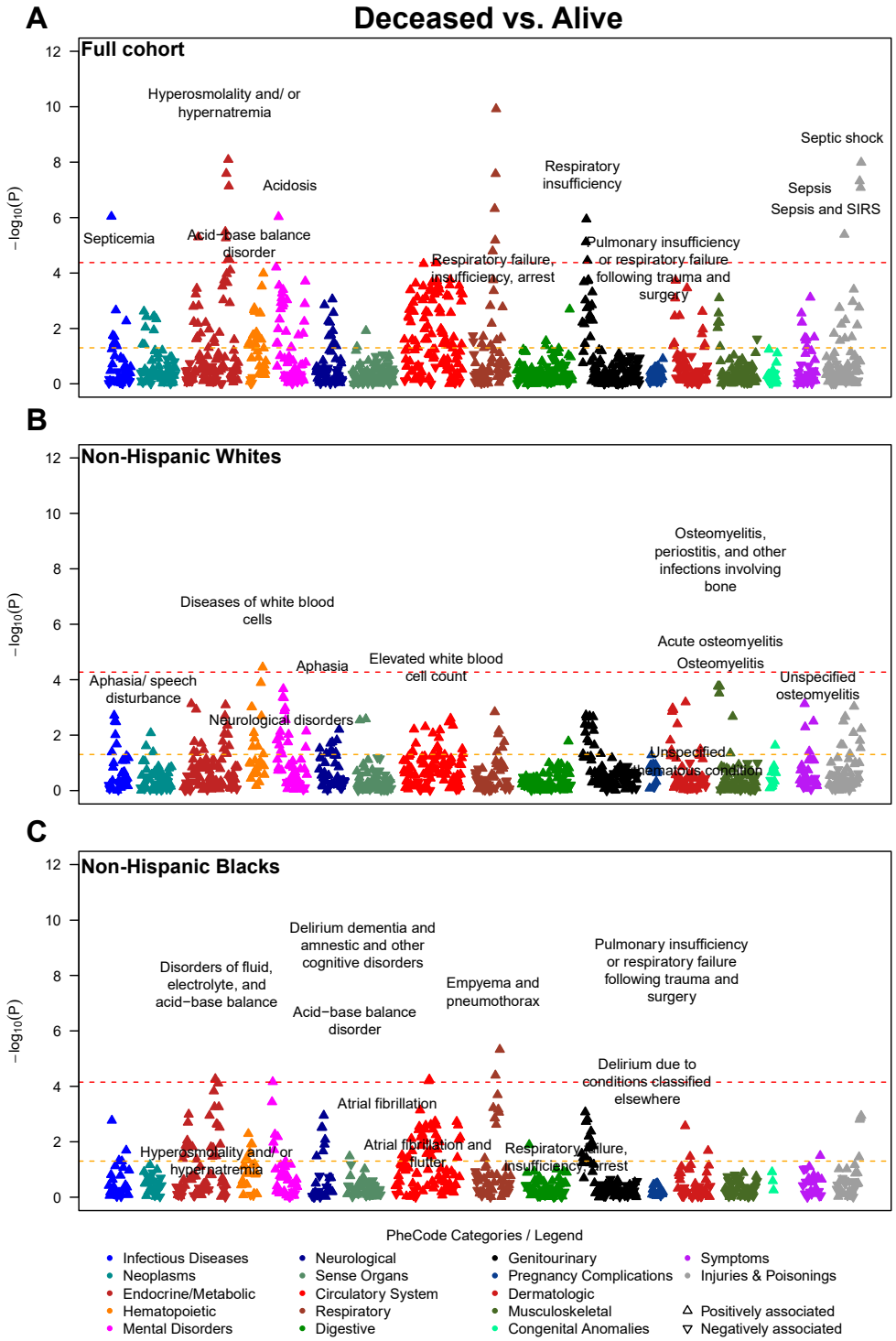

Figure 3. Manhattan plot showing the phenome-wide association between disease conditions and prognostic outcomes for COVID-19. Models are adjusted for age, sex, race (full cohort only), and three census tract-level socioeconomic indicators: proportion with less than high school education, proportion unemployed, and proportion with annual income below the federal poverty level. The *x*-axis are individual disease codes, color-coded by their corresponding disease category as described in the shared legend. The *y*-axis represents the $-\log_{10}$ transformed *p*-value of the association. The dashed, horizontal lines represent the $p = 0.05$ (in orange) and the Bonferroni corrected *p*-value (0.05/number of tests; in red). Each point is represented by either an upward triangle indicating a positive association or a downward triangle indicating a negative association. (**A**): Full cohort, (**B**): Restricted to Whites, (**C**): Restricted to Blacks

3.1.2. Race-Stratified Prognostic Associations

Among White patients, we identified 23 traits phenome-wide significantly associated with hospitalization (e.g., respiratory failure, insufficiency, arrest [$p = 5.25 \times 10^{-9}$], acute renal failure [$p = 5.25 \times 10^{-9}$], and electrolyte imbalance [$p = 1.51 \times 10^{-8}$] Figure 1B), as well as 239 suggestive traits, including 54 circulatory system, 30 respiratory, 29 endocrine/metabolic, and 21 genitourinary diseases. Thirty-two phenome-wide significant traits (e.g., electrolyte imbalance [$p = 2.63 \times 10^{-8}$], pulmonary collapse, interstitial and compensatory emphysema [$p = 8.91 \times 10^{-8}$] and hypotension [$p = 3.63 \times 10^{-7}$]) and 239 suggestive traits were associated with ICU admission/mortality (Figure 2B), including 60 circulatory system, 27 respiratory, 27 digestive, and 23 hematopoietic diseases. One phenome-wide significant trait (elevated white blood cell count [$p = 3.55 \times 10^{-5}$]) and 130 suggestive traits were associated with COVID-19 mortality (Figure 3B), including 18 circulatory system, 17 endocrine/metabolic, 16 mental disorders, 14 genitourinary diseases such as osteomyelitis ($p = 1.74 \times 10^{-4}$), neurological disorder ($p = 4.37 \times 10^{-4}$) and aphasia ($p = 4.57 \times 10^{-4}$).

Among Black patients, two phenome-wide significant traits were detected (respiratory failure, insufficiency, arrest [$p = 2.63 \times 10^{-8}$], respiratory failure [$p = 4.37 \times 10^{-7}$]) along with 89 traits nominally associated with hospitalization (Figure 1C), including 17 circulatory, 15 genitourinary, and 14 respiratory diseases. Eleven phenome-wide significant traits (e.g., respiratory failure, insufficiency, arrest [$p = 1.48 \times 10^{-8}$], acid-base balance disorder [$p = 4.57 \times 10^{-6}$], hypotension [$p = 4.37 \times 10^{-5}$]) and 119 suggestive traits were associated with ICU admission/mortality, including 33 circulatory, 26 genitourinary, and 17 endocrine/metabolic diseases (Figure 2C). Six phenome-wide significant traits (e.g., empyema and pneumothorax [$p = 3.98 \times 10^{-5}$], hyperosmolality and/or hypernatremia [$p = 5.37 \times 10^{-5}$], atrial fibrillation [$p = 5.62 \times 10^{-5}$]) and 105 suggestive traits were associated with mortality, including 34 circulatory, 24 endocrine/metabolic, and 12 genitourinary diseases. As shown in Figure 3B,C, the strength of association between circulatory system disorders and COVID-19 mortality was higher in Black patients compared to White. Similarly, we observe a higher prevalence of genitourinary diseases in Blacks associated with COVID-19 mortality such as stage I or II chronic kidney disease ($p = 2.34 \times 10^{-4}$) compared to Whites.

When comparing the effect sizes of the top 50 hits significant in the full cohort, stratified by racial groups, we found—as expected—no significant differences in the effect sizes (though there are numerical differences). These traits exhibited consistent risks among races for hospitalization (Figures S2A and S3A) and for ICU admission/mortality (Figures S2B and S3B).

3.1.3. Differences between Blacks and Whites across the Phenome

When comparing the estimated effect sizes for the hospitalization outcome across the phenome between Blacks and Whites, we found 35 pre-existing traits (including 10 digestive, 7 endocrine/metabolic, 6 circulatory system, and 5 respiratory conditions) where the effect estimates differed across Blacks and Whites at a p-value less than 0.05 (Figure S4). In all cases, there was a stronger, positive association with the trait and COVID-19-related hospitalization among Whites compared to Blacks. For the ICU admission outcome, we found 31 such pre-existing traits (including 6 neoplasm, 4 hematopoietic, 4 respiratory, and 4 digestive conditions) where the effect estimates differed (p for difference < 0.05) (Figure S5). Similarly, there was a stronger, positive association with the trait and COVID-19-related ICU admission among Whites compared to Blacks, except for delirium due to conditions classified elsewhere (phecode 290.2), chronic pain syndrome (phecode 355.1), disorders of lacrimal system (phecode 375), and degeneration of intervertebral disc (phecode 722.6), where the opposite was true. For the COVID-19-related death outcome, we found 8 such conditions, including 3 circulatory system and 2 respiratory conditions (Figure S6). There was a strong, positive association with the trait and COVID-19-related death among Blacks compared to Whites for 6 of the traits, while the opposite was true for decreased white blood cell count (phecode 288.1) and cervicalgia (phecode 761).

3.1.4. Summary Takeaways

In all cohorts, as the disease progressed to increasingly severe prognosis, the associated phenotypes concentrated in circulatory heart diseases and renal diseases (Figure 4); pre-existing *cardiovascular system problems*, and chronic diseases such as *chronic pulmonary heart disease* and *chronic renal failure* appeared to be associated with poor prognosis, while mental disorders constituted the third largest category associated with COVID-19 mortality behind circulatory system and endocrine/metabolic diseases. When comparing the top 50 traits between Whites and Blacks, acidosis, pulmonary, acute/chronic renal diseases showed an association with hospitalization and ICU admission/mortality in both races, while acute renal consistently stood out as well as in mortality (Table 2). Phenome-wide significant (for the overall cohort) effect estimates for parent phecodes and corresponding confidence intervals for phenome-wide significant traits by outcome by cohort are present in forest plots of parent phecodes (Figure S2) and child phecodes (Figure S3). While no effect differences were significant at Bonferroni-corrected p-value thresholds, we observed 35, 31, and 8 nominally significant ($p < 0.05$) effect differences for Blacks compared to Whites for hospitalization (Figure S4), ICU/admission (Figure S5), and death (Figure S6), respectively. A description of the susceptibility outcome results and corresponding PheWAS plots (Figure S7) are in the Supplementary Materials. Respiratory, endocrine/metabolic, and circulatory system conditions were associated with COVID-19 susceptibility overall, while mental disorders appear overrepresented among Whites.

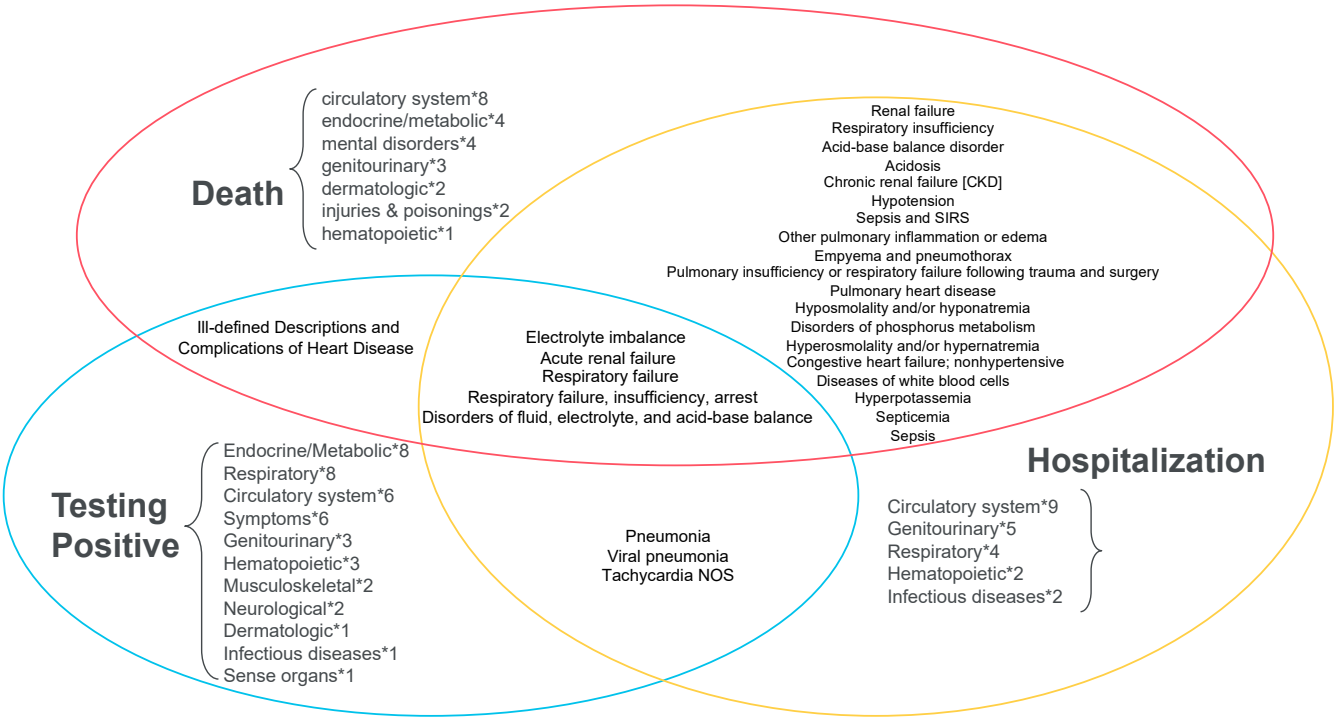

Figure 4. Venn diagrams of the top 50 traits. Each circle represents the top 50 hits from the full cohort PheWAS. Traits shared across PheWAS are stated, while the corresponding number of traits within a given disease category that are unique to that PheWAS are also provided. Abbreviations: NOS, not otherwise specified; SIRS, systemic inflammatory response syndrome

Table 2. Comparison of the top 20 traits from White and Black cohorts across COVID-19 outcome PheWAS.

		Outcome in Top 20 Traits		
Phecode	**Description**	**Hospitalization**	**ICU Admission**	**Death**
276	Disorders of fluid, electrolyte, and acid-base balance	White	White	Black
276.1	Electrolyte imbalance	White	White	Black
276.12	Hyposmolality and/or hyponatremia	White	White	White
276.13	Hyperpotassemia	White		
276.4	Acid-base balance disorder	Both	Both	Black
276.41	Acidosis	White	Both	Black
276.5	Hypovolemia	White		
401.2	Hypertensive heart and/or renal disease	Both		
427.7	Tachycardia, not otherwise specified (NOS)	White	White	
458	Hypotension	Both	Both	
507	Pleurisy; pleural effusion	White		
508	Pulmonary collapse; interstitial and compensatory emphysema	White	White	
509	Respiratory failure, insufficiency, arrest	Both	Black	Black
509.1	Respiratory failure	Both	Black	Black
509.2	Respiratory insufficiency	Both	Both	Black
509.8	Dependence on respirator [Ventilator] or supplemental oxygen	Both	Black	
585	Renal failure	Both	Both	Black
585.1	Acute renal failure	Both	White	
585.3	Chronic renal failure [CKD]	Both	Both	
586	Other disorders of the kidney and ureters	White		
38	Septicemia	Black	Black	
38.3	Bacteremia	Black	Black	
401.22	Hypertensive chronic kidney disease	Black		
480.2	Viral pneumonia	Black		
506	Empyema and pneumothorax	Black	Black	Black
509.3	Pulmonary insufficiency or respiratory failure following trauma and surgery	Black	Both	Black
585.4	Chronic kidney disease, Stage I or II	Black	Black	
588	Disorders resulting from impaired renal function	Black		
785	Abdominal pain	Black		
994.2	Sepsis	Black		
41.1	Staphylococcus infections		White	
260	Protein-calorie malnutrition		White	
285	Other anemias		White	
287.3	Thrombocytopenia		White	
288	Diseases of white blood cells		White	White
458.9	Hypotension, not otherwise specified (NOS)		Both	
854	Complications of cardiac/vascular device, implant, and graft		White	
276.6	Fluid overload		Black	
411.4	Coronary atherosclerosis		Black	
459	Other disorders of circulatory system		Black	
459.9	Circulatory disease, not elsewhere classifiable (NEC)		Black	
505	Other pulmonary inflammation or edema		Black	
249	Secondary diabetes mellitus			White
250.25	Diabetes type 2 with peripheral circulatory disorders			Both
284	Aplastic anemia			White
284.1	Pancytopenia			White
288.2	Elevated white blood cell count			White
292	Neurological disorders			White
292.1	Aphasia/speech disturbance			White

Table 2. *Cont.*

		Outcome in Top 20 Traits		
Phecode	**Description**	**Hospitalization**	**ICU Admission**	**Death**
292.11	Aphasia			White
292.3	Memory loss			White
681.5	Cellulitis and abscess of leg, except foot			White
681.6	Cellulitis and abscess of foot, toe			White
695.9	Unspecified erythematous condition			White
710	Osteomyelitis, periostitis, and other infections involving bone			White
710.1	Osteomyelitis			White
710.11	Acute osteomyelitis			White
710.19	Unspecified osteomyelitis			White
771	Musculoskeletal symptoms referable to limbs			White
962.3	Hormones and synthetic substitutes causing adverse effects in therapeutic use			White
275	Disorders of mineral metabolism			Black
276.11	Hyperosmolality and/or hypernatremia			Black
290	Delirium dementia and amnestic and other cognitive disorders			Black
290.2	Delirium due to conditions classified elsewhere			Black
348	Other conditions of brain			Black
426.21	First degree atrioventricular (AV) block			Black
427.2	Atrial fibrillation and flutter			Black
427.21	Atrial fibrillation			Black
503	Pulmonary congestion and hypostasis			Black

These results enable us to understand the risk profiles that are associated with poor COVID-19 prognosis. It will be interesting to study the association of these pre-existing conditions with post-covid acute complications or "long COVID" syndrome [31,32].

4. Discussion

Using data from a cohort of tested/diagnosed COVID-19 patients at MM, we performed what we believe is the first PheWAS looking at multiple COVID-19 outcomes stratified by race. A recently published PheWAS by Oetjens et al. found 21 phenome-wide significant traits—including six kidney (e.g., end stage renal disease or stage 5 CKD [OR = 11.07, $p = 1.96 \times 10^{-8}$]), six cardiovascular (e.g., congestive heart failure [OR = 3.8, $p = 3.24 \times 10^{-5}$]), five respiratory (e.g., chronic airway obstruction [OR = 2.54, $p = 3.71 \times 10^{-5}$]), and three metabolic (e.g., type 2 diabetes [OR = 1.80, $p = 7.51 \times 10^{-5}$]) phenotypes—associated with COVID-19-related hospitalization, which is consistent with our results [14]. This technique allowed us to explore and identify potentially associated conditions across the medical phenome that are associated with susceptibility, hospitalization, ICU admission or mortality. Our results yield many previously known or plausibly associated phenotypes with increasingly severe prognosis such as pulmonary heart disease, respiratory failure and type 2 diabetes. Our stratified analysis showed that kidney disease (e.g., renal failure OR = 1.34, $p = 4.90 \times 10^{-7}$ among Whites vs. OR = 1.12, $p = 2.88 \times 10^{-4}$ among Blacks for ICU admission) and hematopoietic conditions (e.g., anemia of chronic disease OR = 1.39, $p = 7.59 \times 10^{-5}$ among Whites vs. OR = 1.19, $p = 4.68 \times 10^{-3}$ among Blacks for ICU admission) appear to be associated with more severe outcomes among Whites, while respiratory (e.g., pulmonary insufficiency or respiratory failure following trauma OR = 2.75, $p = 4.68 \times 10^{-6}$ among Blacks vs. OR = 1.70, $p = 6.92 \times 10^{-3}$ among Whites for death) and circulatory system (e.g., atrial fibrillation OR = 2.26, $p = 5.62 \times 10^{-5}$ among Blacks vs. OR = 1.90, $p = 2.15 \times 10^{-1}$ among Whites for death) conditions are more strongly associated with severe outcomes among Blacks. Table 2 shows that the disease categories that comprise the top 20 hits by prognostic outcome by race are different (with the caveat that for ICU admission and mortality these hits are largely suggestive due to limited power). Our results can inform targeted prevention across racial groups, which includes

increased testing and encouraging self-isolation from household members with specific disease profiles along with education of enhanced public health prevention guidelines.

There are several limitations to this analysis. First, there is the agnostic nature of PheWAS, which can identify potentially spurious associations. While we feel that many of the top traits have been highlighted elsewhere and are biologically plausible, there is currently no process in place for rapidly discerning potentially novel from spurious associations [33] beyond extensive manual review and follow-up research, particularly for a novel disease. Second, many of the issues with utilizing EHR data for research purposes also apply here, including inaccurate data from billing codes [34] and failure of physicians to report/record problems [35]. Third, the sample size for a PheWAS is still rather small to be able to identify statistically significant associations—particularly for mortality. Moreover, we did not distinguish between transfer patients (i.e., those who were diagnosed elsewhere and transferred to MM for treatment), who may have been sicker patients than the cohort diagnosed at MM. However, given that this is an emerging and novel disease, we feel it is important to identify suggestive associations so that future research and clinicians can potentially consider other conditions outside those that have been previously identified—namely, pulmonary and cardiovascular conditions—and to inspire follow-up studies in larger cohorts. For example, OpenSAFELY, a platform including the primary care records of 17,728,392 adults in England (covering 40% of all patients) [16], shares many of our full-cohort conclusions (with consistent effect sizes), but has not published race-stratified PheWAS results. Finally, our analysis is scanning through each phenotype one at a time, though they occur in a correlated and interactive manner. A richer multivariate model needs to be constructed with more complex features.

While potentially relevant, we did not explore past medication data in our analyses because the available EHR data did not provide comprehensive medication coverage but predominantly medication orders and administrations for hospitalized patients. Other population-based cohorts with prescription registries, as exemplified by Butt and colleagues [36], might be better suited for such an exploration.

5. Conclusions

This work contributes to a new area of COVID-19 research that rigorously examines racial differences in disease prognosis with pre-existing conditions captured across the medical phenome. Moreover, we incorporated a census tract-level socioeconomic status (SES) covariate, which is important to consider when comparing races [37]. We found several potentially novel diseases unexpectedly associated with different outcomes in the course of COVID-19 progression and that some disease profiles differ by race. For example, we provide additional evidence on the previously reported concern that patients with mental health disorders are at higher risk of infection and experience barriers in seeking treatment leading to poor prognosis [38]. We hope this exploratory effort will inspire hypothesis generation for future research that might result in targeted prevention and care as we are still combatting this pandemic. In this spirit, we have made all PheWAS results available for exploration here: https://cphds.sph.umich.edu/covidphewas/ (accessed on 10 March 2021). We hope the summary data and the phenomic landscape for COVID-19 will help future replication and meta-analysis efforts.

Supplementary Materials: The following are available online at https://www.mdpi.com/2077-0383/10/7/1351/s1, Figure S1: Flowchart describing the sample sizes of the overall and race-specific cohorts by progression of COVID-19 outcomes, Figure S2: Forest plots of traits associated with poorer prognosis reaching phenome-wide significance in the overall cohort. A. Hospitalization. B. ICU admission/mortality. Odds ratios and 95% confidence intervals are shown for each trait whose PheWAS code is given in parentheses. Plots show parent codes only. Child codes found in Figure S3, Figure S3: Forest plots of traits associated with poorer prognosis in overall cohort. A. Hospitalization. B. ICU admission/mortality. Odds ratios and 95% confidence intervals are shown for each trait whose PheWAS code is given in parentheses. Plots show child codes only. Parent codes found in Figure S2, Figure S4: Forest plot showing nominally significant effect differences comparing Whites and Blacks with respect to COVID-19-related

hospitalization, Figure S5: Forest plot showing nominally significant effect differences comparing Whites and Blacks with respect to COVID-19-related ICU admission, Figure S6: Forest plot showing nominally significant effect differences comparing Whites and Blacks with respect to COVID-19-related death, Figure S7: Manhattan plot showing the phenome-wide association between disease codes and testing positive for COVID-19. Models are adjusted for age, sex, race (full cohort only), and four census tract-level socioeconomic indicators: proportion with less than high school education, proportion unemployed, proportion with annual income below the federal poverty level, and population density (persons per mile^2). The x-axis are individual disease codes, color-coded by their corresponding disease category as described in the legend. The y-axis represents the -$\log_{10}$ transformed p-value of the association. The dashed, horizontal lines represent the $p = 0.05$ (in orange) and the Bonferroni corrected p-value (0.05/number of tests; in red). Each point is represented by either an upward triangle indicating a positive association or a downward triangle indicating a negative association, Table S1A: Definition and Sources of All Variables, Table S1B: Analysis Setup, Table S2: Descriptive cohort characteristics stratified by race, Table S3: "COVID-19 Positive vs. Controls" PheWAS Merged Top 50 Traits Using Adjustment 2, Table S4: "Hospitalized vs. Alive & Not hospitalized" PheWAS Merged Top 50 Traits Using Adjustment 2, Table S5: "ICU & Dead vs. Alive & no ICU Stay" Merged PheWAS Top 50 Traits Using Adjustment 2, Table S6: "Deceased vs. Alive" Merged PheWAS Top 50 Traits Using Adjustment 2.

Author Contributions: Conceptualization, L.G.F. and B.M.; methodology, L.G.F.; software, S.P.; formal analysis, L.G.F..; data curation, S.P.S. and L.G.F.; writing—original draft preparation, M.S., T.G., J.A.M., L.G.F., T.S.V., K.S., B.K.N., S.K., L.L. and B.M.; writing—review and editing, M.S., L.G.F. and B.M.; visualization, T.G., S.P. and L.G.F.; supervision, B.M.; project administra-tion, B.M.; funding acquisition, B.M. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the University of Michigan Precision Health Initiative, University of Michigan Rogel Cancer Center, Michigan Institute of Data Science, National Science Foundation (grant number DMS 1712933) and the National Institutes of Health (grant number P30 CA 046592-30-S3).

Institutional Review Board Statement: Ethical review and approval were waived for this study, due to its qualification for a federal exemption as secondary research for which consent is not required. Determination for exemption made by the University of Michigan Medical School Institutional Review Board (IRBMED) (study ID: HUM00180294; determination made: 4/13/2020).

Informed Consent Statement: Data was deidentified by honest broker office prior to analysis; not applicable.

Data Availability Statement: Data cannot be shared publicly due to patient confidentiality.

Acknowledgments: Equal contribution of work. Maxwell Salvatore and Tian Gu contributed equally to this work.

Conflicts of Interest: The authors have no conflict of interests to declare.

Abbreviations

Black	non-Hispanic Black/African American
EHR	electronic health record
ICD	International Classification of Disease
ICU	intensive care unit
MM	Michigan Medicine
NaNDA	National Neighborhood Data Archive
OR	odds ratio
PheWAS	phenome-wide association study
SES	socioeconomic status
White	non-Hispanic White

References

1. Denny, J.C.; Ritchie, M.D.; Basford, M.A.; Pulley, J.M.; Bastarache, L.; Brown-Gentry, K.; Wang, D.; Masys, D.R.; Roden, D.M.; Crawford, D.C. PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* **2010**, *26*, 1205–1210. [CrossRef] [PubMed]
2. Taliun, S.A.G.; VandeHaar, P.; Boughton, A.P.; Welch, R.P.; Taliun, D.; Schmidt, E.M.; Zhou, W.; Nielsen, J.B.; Willer, C.J.; Lee, S.; et al. Exploring and visualizing large-scale genetic associations by using PheWeb. *Nat. Genet.* **2020**, *52*, 550–552. [CrossRef]

3. Verma, A.; Lucas, A.; Verma, S.S.; Zhang, Y.; Josyula, N.; Khan, A.; Hartzel, D.N.; Lavage, D.R.; Leader, J.; Ritchie, M.D.; et al. PheWAS and Beyond: The Landscape of Associations with Medical Diagnoses and Clinical Measures across 38,662 Individuals from Geisinger. *Am. J. Hum. Genet.* **2018**, *102*, 592–608. [CrossRef] [PubMed]
4. Zhao, X.; Geng, X.; Srinivasasainagendra, V.; Chaudhary, N.; Judd, S.; Wadley, V.; Gutiérrez, O.M.; Wang, H.; Lange, E.M.; Lange, L.A.; et al. A PheWAS study of a large observational epidemiological cohort of African Americans from the REGARDS study. *BMC Med Genom.* **2019**, *12*, 167–177. [CrossRef]
5. Cai, T.; Zhang, Y.; Ho, Y.-L.; Link, N.; Sun, J.; Huang, J.; Cai, T.A.; Damrauer, S.; Ahuja, Y.; Honerlaw, J.; et al. Association of Interleukin 6 Receptor Variant With Cardiovascular Disease Effects of Interleukin 6 Receptor Blocking Therapy. *JAMA Cardiol.* **2018**, *3*, 849–857. [CrossRef] [PubMed]
6. Leppert, B.; Millard, L.A.; Riglin, L.; Smith, G.D.; Thapar, A.; Tilling, K.; Walton, E.; Stergiakouli, E. A cross-disorder PRS-pheWAS of 5 major psychiatric disorders in UK Biobank. *PLoS Genet.* **2020**, *16*, e1008185. [CrossRef] [PubMed]
7. Li, S.; Schooling, C.M. A phenome-wide association study of ABO blood groups. *BMC Med.* **2020**, *18*, 334. [CrossRef] [PubMed]
8. A Pendergrass, S.; Ritchie, M.D. Phenome-Wide Association Studies: Leveraging Comprehensive Phenotypic and Genotypic Data for Discovery. *Curr. Genet. Med. Rep.* **2015**, *3*, 92–100. [CrossRef] [PubMed]
9. Abdellaoui, A.; Sanchez-Roige, S.; Sealock, J.; Treur, J.L.; Dennis, J.; Fontanillas, P.; Elson, S.; Nivard, M.G.; Ip, H.F.; Van Der Zee, M.; et al. Phenome-wide investigation of health outcomes associated with genetic predisposition to loneliness. *Hum. Mol. Genet.* **2019**, *28*, 3853–3865. [CrossRef] [PubMed]
10. Fritsche, L.G.; Gruber, S.B.; Wu, Z.; Schmidt, E.M.; Zawistowski, M.; Moser, S.E.; Blanc, V.M.; Brummett, C.M.; Kheterpal, S.; Abecasis, G.R.; et al. Association of Polygenic Risk Scores for Multiple Cancers in a Phenome-wide Study: Results from The Michigan Genomics Initiative. *Am. J. Hum. Genet.* **2018**, *102*, 1048–1061. [CrossRef]
11. Microsoft bing COVID-19 Tracker. Available online: https://www.bing.com/covid (accessed on 29 June 2020).
12. Centers for Disease Control and Prevention. Overview of testing for SARS-CoV-2. 2020. Available online: https://www.cdc.gov/coronavirus/2019-ncov/hcp/testing-overview.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fhcp%2Fclinical-criteria.html (accessed on 15 June 2020).
13. Gu, T.; Mack, J.A.; Salvatore, M.; Sankar, S.P.; Valley, T.S.; Singh, K.; Nallamothu, B.K.; Kheterpal, S.; Lisabeth, L.; Fritsche, L.G.; et al. Characteristics Associated With Racial/Ethnic Disparities in COVID-19 Outcomes in an Academic Health Care System. *JAMA Netw. Open* **2020**, *3*, e2025197. [CrossRef]
14. Oetjens, M.T.; Luo, J.Z.; Chang, A.; Leader, J.B.; Hartzel, D.N.; Moore, B.S.; Strande, N.T.; Kirchner, H.L.; Ledbetter, D.H.; Justice, A.E.; et al. Electronic health record analysis identifies kidney disease as the leading risk factor for hospitalization in confirmed COVID-19 patients. *PLoS ONE* **2020**, *15*, e0242182. [CrossRef]
15. Rosenthal, N.; Cao, Z.; Gundrum, J.; Sianis, J.; Safo, S. Risk Factors Associated With In-Hospital Mortality in a US National Sample of Patients With COVID-19. *JAMA Netw. Open* **2020**, *3*, e2029058. [CrossRef] [PubMed]
16. Williamson, E.J.; Walker, A.J.; Bhaskaran, K.; Bacon, S.; Bates, C.; E Morton, C.; Curtis, H.J.; Mehrkar, A.; Evans, D.; Inglesby, P.; et al. OpenSAFELY: Factors associated with COVID-19 death in 17 million patients. *Nature* **2020**, *584*, 430–436. [CrossRef]
17. Sim, B.L.H.; Chidambaram, S.K.; Wong, X.C.; Pathmanathan, M.D.; Peariasamy, K.M.; Hor, C.P.; Chua, H.J.; Goh, P.P. Clinical characteristics and risk factors for severe COVID-19 infections in Malaysia: A nationwide observational study. *Lancet Reg. Health West. Pac.* **2020**, *4*, 100055. [CrossRef]
18. Brainard, J. Scientists are drowning in COVID-19 papers. Can new tools keep them afloat? *Science* **2020**. [CrossRef]
19. Price-Haywood, E.G.; Burton, J.; Fort, D.; Seoane, L. Hospitalization and Mortality among Black Patients and White Patients with Covid-19. *N. Engl. J. Med.* **2020**, *382*, 2534–2543. [CrossRef] [PubMed]
20. Brandt, E.B.; Beck, A.F.; Mersha, T.B. Air pollution, racial disparities, and COVID-19 mortality. *J. Allergy Clin. Immunol.* **2020**, *146*, 61–63. [CrossRef]
21. Laurencin, C.T.; McClinton, A. The COVID-19 Pandemic: A Call to Action to Identify and Address Racial and Ethnic Disparities. *J. Racial Ethn. Health Disparities* **2020**, *7*, 398–402. [CrossRef] [PubMed]
22. Vahidy, F.S.; Pan, A.P.; Ahnstedt, H.; Munshi, Y.; Choi, H.A.; Tiruneh, Y.; Nasir, K.; Kash, B.A.; Andrieni, J.D.; McCullough, L.D. Sex differences in susceptibility, severity, and outcomes of coronavirus disease 2019: Cross-sectional analysis from a diverse US metropolitan area. *PLoS ONE* **2021**, *16*, e0245556. [CrossRef]
23. Pflugeisen, B.M.; Mou, J. Empiric evidence of ethnic disparities in coronavirus positivity in Washington State. *Ethn. Health* **2021**, *26*, 36–48. [CrossRef]
24. Chang, M.-H.; Moonesinghe, R.; Truman, B.I. COVID-19 Hospitalization by Race and Ethnicity: Association with Chronic Conditions Among Medicare Beneficiaries, January 1–September 30, 2020. *J. Racial Ethn. Health Disparities* **2021**, 1–10, ahead of print. [CrossRef]
25. Wiley, Z.; Kubes, J.N.; Cobb, J.; Jacob, J.T.; Franks, N.; Plantinga, L.; Lea, J. Age, Comorbid Conditions, and Racial Disparities in COVID-19 Outcomes. *J. Racial Ethn. Health Disparities* **2021**, 1–7, ahead of print. [CrossRef]
26. Michigan Medicine. Indications for COVID-19 Diagnostics Testing for Adult Patients in All Clinical Settings. 2020. Available online: http://www.med.umich.edu/asp/pdf/adult_guidelines/COVID-19-testing.pdf (accessed on 9 June 2020).
27. Firth, D. Bias reduction of maximum likelihood estimates. *Biometrika* **1993**, *80*, 27–38. [CrossRef]
28. Kosmidis, I.; Pagui, E.C.K.; Sartori, N. Mean and median bias reduction in generalized linear models. *Stat. Comput.* **2019**, *30*, 43–59. [CrossRef]

29. brglm2: Bias Reduction in Generalized Linear Models. R package version 0.7.1. Available online: https://CRAN.R-project.org/package=brglm2 (accessed on 10 March 2021).
30. NaNDA | Social Environment and Health Program. Available online: https://seh.isr.umich.edu/signature-projects/nanda/ (accessed on 10 June 2020).
31. Huang, C.; Huang, L.; Wang, Y.; Li, X.; Ren, L.; Gu, X.; Kang, L.; Guo, L.; Liu, M.; Zhou, X.; et al. 6-month consequences of COVID-19 in patients discharged from hospital: A cohort study. *Lancet* **2021**, *397*, 220–232. [CrossRef]
32. Marshall, M. The lasting misery of coronavirus long-haulers. *Nature* **2020**, *585*, 339–341. [CrossRef]
33. A Hanauer, D.; Saeed, M.; Zheng, K.; Mei, Q.; Shedden, K.; Aronson, A.R.; Ramakrishnan, N. Applying MetaMap to Medline for identifying novel associations in a large clinical dataset: A feasibility analysis. *J. Am. Med. Inform. Assoc.* **2014**, *21*, 925–937. [CrossRef] [PubMed]
34. Rhodes, E.T.; Laffel, L.M.; Gonzalez, T.V.; Ludwig, D.S. Accuracy of Administrative Coding for Type 2 Diabetes in Children, Adolescents, and Young Adults. *Diabetes Care* **2006**, *30*, 141–143. [CrossRef] [PubMed]
35. Williams, C.; Mosley-Williams, A.; McDonald, C. Accuracy of provider generated computerized problem lists in the Veterans Administration. *AMIA Annu. Symp. Proc.* **2007**, *1155*.
36. Butt, J.H.; Gerds, T.A.; Schou, M.; Kragholm, K.; Phelps, M.; Havers-Borgersen, E.; Yafasova, A.; Gislason, G.H.; Torp-Pedersen, C.; Køber, L.; et al. Association between statin use and outcomes in patients with coronavirus disease 2019 (COVID-19): A nationwide cohort study. *BMJ Open* **2020**, *10*, e044421. [CrossRef] [PubMed]
37. Williams, D.R.; Priest, N.; Anderson, N.B. Understanding associations among race, socioeconomic status, and health: Patterns and prospects. *Health Psychol.* **2016**, *35*, 407–411. [CrossRef] [PubMed]
38. Yao, H.; Chen, J.-H.; Xu, Y.-F. Patients with mental health disorders in the COVID-19 epidemic. *Lancet Psychiatry* **2020**, *7*, e21. [CrossRef]