A phenome-wide association study (PheWAS) of COVID-19 outcomes by race using the electronic health records data in Michigan Medicine

Maxwell Salvatore[a], Tian Gu[a], Jasmine A. Mack[a], Swaraaj Prabhu Sankar[b,c], Snehal Patil[a,d], Thomas S. Valley[e,f], Karandeep Singh[f,g], Brahmajee K. Nallamothu[h], Sachin Kheterpal[f,i], Lynda Lisabeth[j], Lars G. Fritsche[a,b,k], Bhramar Mukherjee[a,b,d]*

[a] Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI 48109, United States
[b] Rogel Cancer Center, University of Michigan Medicine, Ann Arbor, MI 48109, United States
[c] Data Office for Clinical and Translational Research, University of Michigan, Ann Arbor, MI 41809, United States
[d] Precision Health, University of Michigan, Ann Arbor, MI 48109, United States
[e] Division of Pulmonary and Critical Care Medicine and Department of Internal Medicine, University of Michigan Medicine, Ann Arbor, MI 48109, United States
[f] Institute for Healthcare Policy and Innovation, University of Michigan, Ann Arbor, MI 48109, United States
[g] Department of Learning Health Sciences, University of Michigan, Ann Arbor, MI 48109, United States
[h] Division of Cardiovascular Medicine and Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI 48109, United States
[i] Department of Anesthesiology, University of Michigan Medical School, Ann Arbor, MI 48109, United States
[j] Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI 48109, United States
[k] Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI 48109, United States

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

* Corresponding author
Bhramar Mukherjee
University of Michigan School of Public Health Department of Biostatistics
1415 Washington Heights
Ann Arbor, MI 48109
United States
bhramar@umich.edu
+1 (734)-764-6544

Abstract

Background: We perform a phenome-wide scan to identify pre-existing conditions related to COVID-19 susceptibility and prognosis across the medical phenome and how they vary by race.

Methods: The study is comprised of 53,853 patients who were tested/positive for COVID-19 between March 10 and September 2, 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 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.

1. Introduction

The emergence of electronic health records (EHR) and rise of EHR-linked biobanks has 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]. Due to computational advances and development of widely available analytic frameworks[2–6], PheWAS are 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.

As of January 15, 2021, there are 23,759,743 confirmed COVID-19 cases in the US,[7] representing approximately 25% of all global cases. Because COVID-19 is a respiratory disease and produces flu-like symptoms, testing strategies in the US initially focused on those with symptoms, the elderly, and those with pre-existing conditions [8] - populations who are at risk of severe disease and complications. However, because COVID-19 is a novel disease, only a handful of pre-existing phenotypes are known to be associated with developing symptoms or experiencing adverse outcomes. These include liver, kidney, heart, and respiratory disease.

There has been a remarkable surge within the academic and medical communities to conduct rapid research on COVID-19[9]. However, only recently have there been studies examining disparities across an ensemble of COVID-19 associated conditions and outcomes in US patient cohorts[10–17]. Instead of a hypothesis driven approach based on the literature, this study applies an agnostic *disease-disease* PheWAS framework to COVID-19 outcomes in a cohort of 53,853 patients who were tested for or treated at a large academic medical center. We look at 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 is to agnostically identify conditions present in an individual's medical record that may be associated with hospitalization, ICU admission, and mortality. We also present the results from race stratified susceptibility PheWAS 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 it hard to find a suitable comparison group. A naïve comparison of the positive versus negative test results is highly biased.[17] However, conditional on testing positive, downstream outcomes are less prone to such selection biases and we focus on these outcomes.

2. Materials and Methods

2.1. Study design

2.1.1. COVID-19 cohort

We extracted the EHR for patients tested for COVID-19 at the University of Michigan Health System, also known as Michigan Medicine (MM), from March 10, 2020 to September 2, 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 2,582 tested positive. Since the testing protocol in MM[18] focused on prioritized testing, this is a non-random 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

We categorized 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 (17 had at least one ICU state 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) code from EHR and forming them up to 1,813 traits using the PheWAS R package (as described in detail elsewhere).[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. 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 1,363 traits that appeared in the EHR 14-day-prior of at least ten COVID-19 positive patients. Further, we realize that the aggregation of ICD codes into phecodes may result in clinically unusable or unclear phenotypes. 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.

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 did not (no untested controls were used, only considers tested positive cohort)

We also present results from the susceptibility PheWAS (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. All logistic regression models were of the following form:

$$\mathrm{logit}\, P(Y_{COVID} = 1 | \mathrm{Covariates}, \mathrm{PheCode}_k) = \beta_0 + \beta_{Cov}^{T}\mathrm{Covariates} + \beta_k I[\mathrm{Phecode}_k = 1] \quad (1)$$

$k = 1, \ldots, 1051$. 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.[19–21] 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).[19] PheWAS adjusting for an additional comorbidity score covariate (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 on our accompanying website: https://cphds.sph.umich.edu/covidphewas/.

2.3.1. Race-stratified analysis

Since the prognostic factors could potentially be different across races, we carried out 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).

For all models, we report the Firth corrected estimate of the odds ratio, 95% Wald-type confidence interval and *P*-value. A conservative Bonferroni multiple testing correction was implemented to conclude statistically significant results (*P*=0.05/number of phecodes in analysis), and P < .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 eligible for inclusion in this study. Of those eligible for inclusion, our study population comprised 47,862 individuals (n_{tested}=47,862 [n_{positive}=2,133]) who had available International Classification of Disease (ICD; ninth and tenth editions) code data after applying the 14-day-prior to testing restriction to the EHR. Furthermore, a total of 1,813 qualified ICD-code-based phenotypes, referred to as PheWAS codes, were initially screened of which 1,363 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 (5,763). Out of the study cohort, 4.8% (2,582) were tested positive (**Table 1**). Among the 2,582 positive patients, 54.6% (1,411) were White, 25.0% (646) were Black, 27.8% (719) were hospitalized, 14.6% (377) were admitted to ICU and 5.0% (129) died.

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

	Individuals, no. (%)[a]					
	Tested for COVID-19					
			Positive Results			
Variable	Overall (n = 53853)	Negative Results (n = 51271)	Overall (n = 2582)	Hospitalized (n = 719)	ICU (n = 377)	Deceased (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)	12652 (23.5)	12017 (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)	12116 (22.5)	11440 (22.3)	676 (26.2)	224 (31.2)	120 (31.8)	33 (25.6)
[65,80)	10257 (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)
Gender	23814 (44.2)	22651 (44.2)	1163 (45)	403 (56.1)	233 (61.8)	80 (62)
Primary Care in MM	31357 (58.2)	29969 (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)	12857 (29.7)	12357 (30)	500 (22.9)	102 (14.9)	61 (16.9)	17 (13.7)
[25,30)	13371 (30.8)	12723 (30.9)	648 (29.7)	211 (30.9)	110 (30.5)	45 (36.3)
>=30	16291 (37.6)	15281 (37.1)	1010 (46.3)	361 (52.9)	186 (51.5)	61 (49.2)
Smoking Status						
Never	31041 (63.2)	29549 (63)	1492 (68.7)	368 (60.2)	159 (54.6)	30 (39)
Past	13725 (28)	13145 (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	18039 (36.8)	17360 (37)	679 (31.3)	243 (39.8)	132 (45.4)	47 (61)
Alcohol consumption	25894 (68.4)	24768 (68.6)	1126 (66.2)	261 (63.2)	128 (63.7)	35 (61.4)
Race/ethnicity						
White	38977 (72.4)	37566 (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/square mile	2375.8 (2422.1)	2343.2 (2412.8)	2997.3 (2512.8)	3658.7 (2635)	3826.4 (2675.2)	4128.4 (2770.3)
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.

3.1. Phenome-wide comorbidity association analysis

The association results for 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/. This 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 additional 263 suggestive traits under threshold of P<.05 (**Figure 1A**)— respiratory failure, insufficiency, arrest (P=3.98x10^{-20}), acute renal failure (P=6.31x10^{-13}), viral pneumonia (P=2.51x10^{-11}), and acid-base balance disorder (P=2.40x10^{-10}). Moreover, 58 phenome-wide significant hits (e.g., respiratory failure, insufficiency, arrest [P=1.58x10^{-15}], acid-base balance disorder [P=3.98x10^{-14}], and hypotension [P=1.58x10^{-11}]) as well as 286 suggestive hits were noted for association with ICU admission/mortality (**Figure 1D**), 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 1G**), along with additional 227 suggestive traits under 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*=9.33x10^{-7}), memory loss (*P*=3.98x10^{-4}) and aphasia (*P*=5.37x10^{-4}).

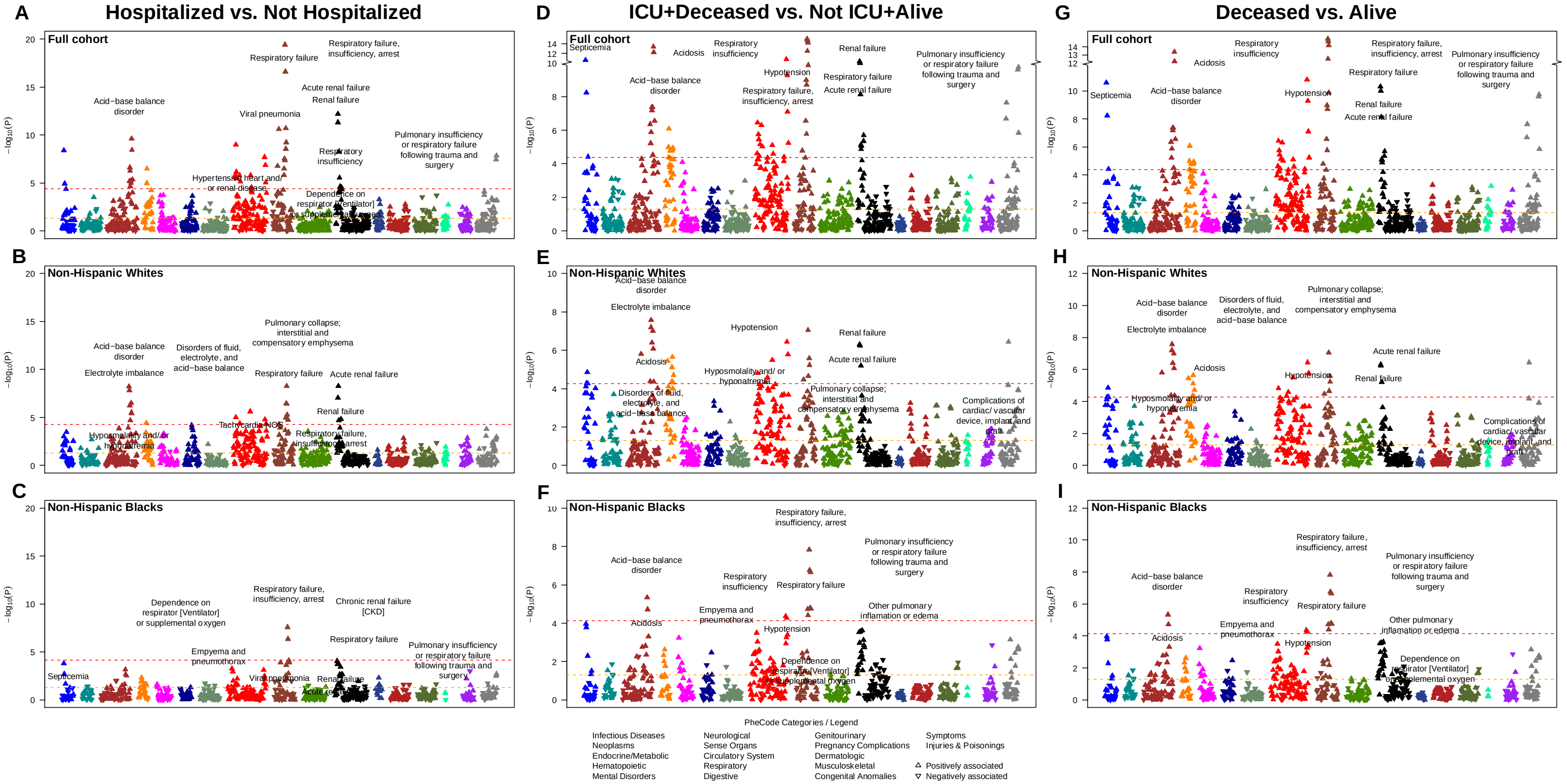

Figure 1. Manhattan plot showing the phenome-wide association between disease conditions and prognostic outcomes for COVID-19. *Models are adjusted for age, sex, race (full cohort only), and three census tract-level socioeconomic indicators: proportion with less than high school education, proportion 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.*

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*=5.25x10^{-9}], acute renal failure [*P*=5.25x10^{-9}], and electrolyte imbalance [*P*=1.51x10^{-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*=2.63x10^{-8}], pulmonary collapse, interstitial and compensatory emphysema [*P*=8.91x10^{-8}] and hypotension [*P*=3.63x10^{-7}]) and 239 suggestive traits were associated with ICU admission/mortality (**Figure 1E**), including 60 circulatory system, 27 respiratory, 27 digestive, and 23 hematopoietic diseases. One phenome-wide significant trait (elevated white blood cell count [*P*=3.55x10^{-5}]) and 130 suggestive traits were associated with COVID-19 mortality (**Figure 1H**), including 18 circulatory system, 17 endocrine/metabolic, 16 mental disorders, 14 genitourinary diseases such as osteomyelitis (*P*=1.74x10^{-4}), neurological disorder (*P*=4.37x10^{-4}) and aphasia (*P*=4.57x10^{-4}).

Among Black patients, two phenome-wide significant traits were detected (respiratory failure, insufficiency, arrest [*P*=2.63x10^{-8}], respiratory failure [*P*=4.37x10^{-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*=1.48x10^{-8}], acid-base balance disorder [*P*=4.57x10^{-6}], hypotension [*P*=4.37x10^{-5}]) and 119 suggestive traits were associated with ICU admission/mortality, including 33 circulator, 26 genitourinary, and 17 endocrine/metabolic diseases. Six phenome-wide significant traits (e.g., empyema and pneumothorax [P=3.98x10^{-5}], hyperosmolality and/or hypernatremia [P=5.37x10^{-5}], atrial fibrillation [P=5.62x10^{-5}]) and 105 suggestive traits were associated with mortality, including 34 circulator, 24 endocrine/metabolic, and 12 genitourinary diseases. As shown in **Figure 1H** and **Figure 1I**, the strength of association between circulatory system disorders and COVID-19 mortality was higher in Black patients compared with 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=2.34x10^{-4}) compared with Whites.

When comparing the effect sizes of phenome-wide significant associations (in the full cohort) across Whites and Blacks, we found no significant differences in the effect sizes (though there are numerical differences) but these traits exhibited consistent risks among races for hospitalization (**Figure 1A**) and for ICU admission/mortality (**Figure 1B**).

3.1.3. 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 2A**); 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 (**Figure 2B**) and ICU admission/mortality (**Figure 2C**) in both races, while acute renal consistently stood out as well as in mortality (**Figure 2D**). Effect estimates for parent phecodes and corresponding confidence intervals for phenome-wide significant traits by outcome by cohort are present in forest plots in **Figure 3**. A similar forest plot consisting of child phecodes are in **Figure S1**. A description of the susceptibility outcome results and corresponding PheWAS plots (**Figure S2**) are in the **supplementary materials**.

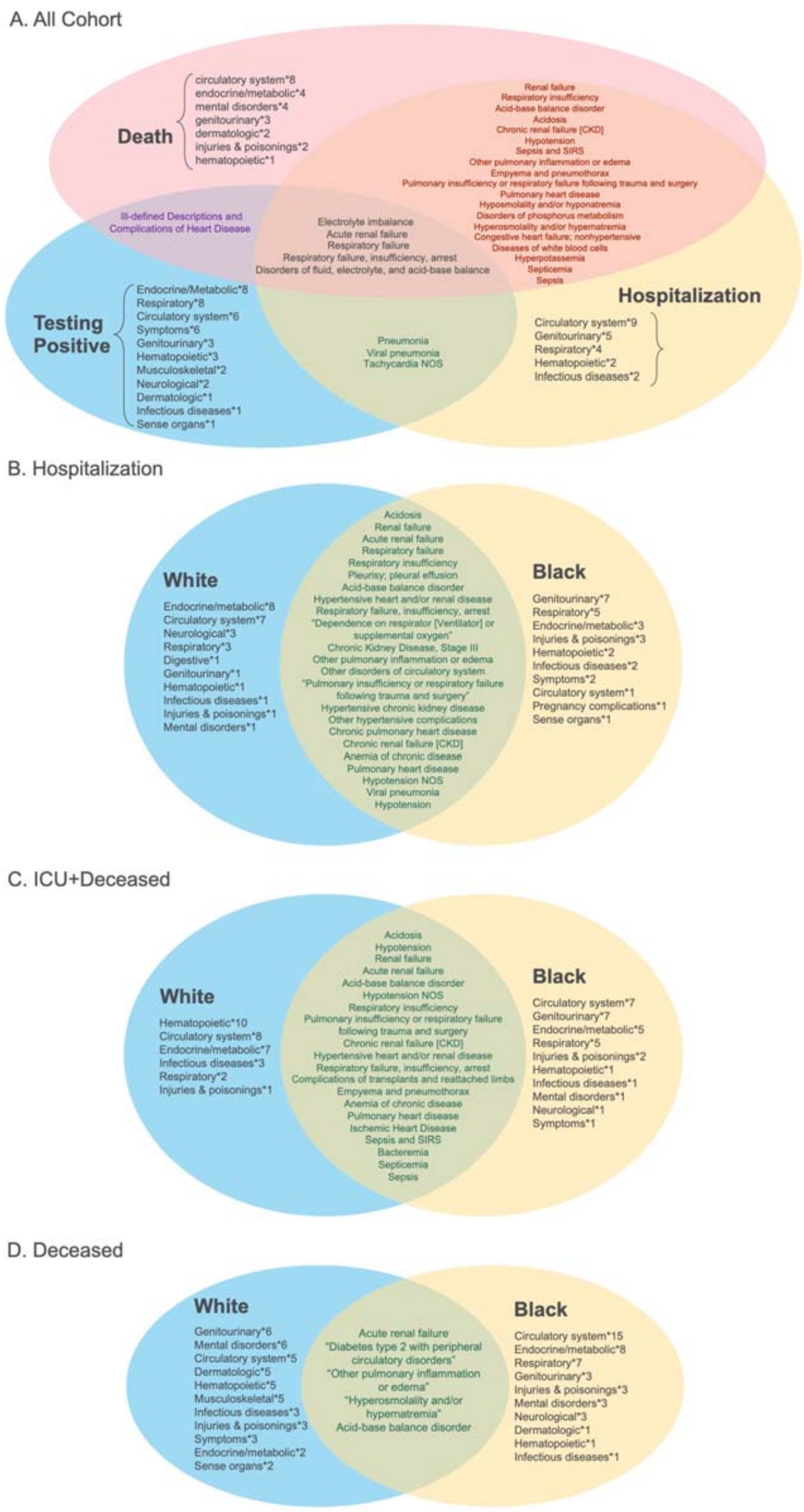

Figure 2. Venn diagrams of the top 50 traits. *Each circle represents the top 50 hits from the full cohort PheWAS (panel A) and the racial PheWAS (panels B, C and D), respectively. Traits shared across PheWAS are stated, while the corresponding number of traits within a given disease category that are unique to that PheWAS are also provided.*

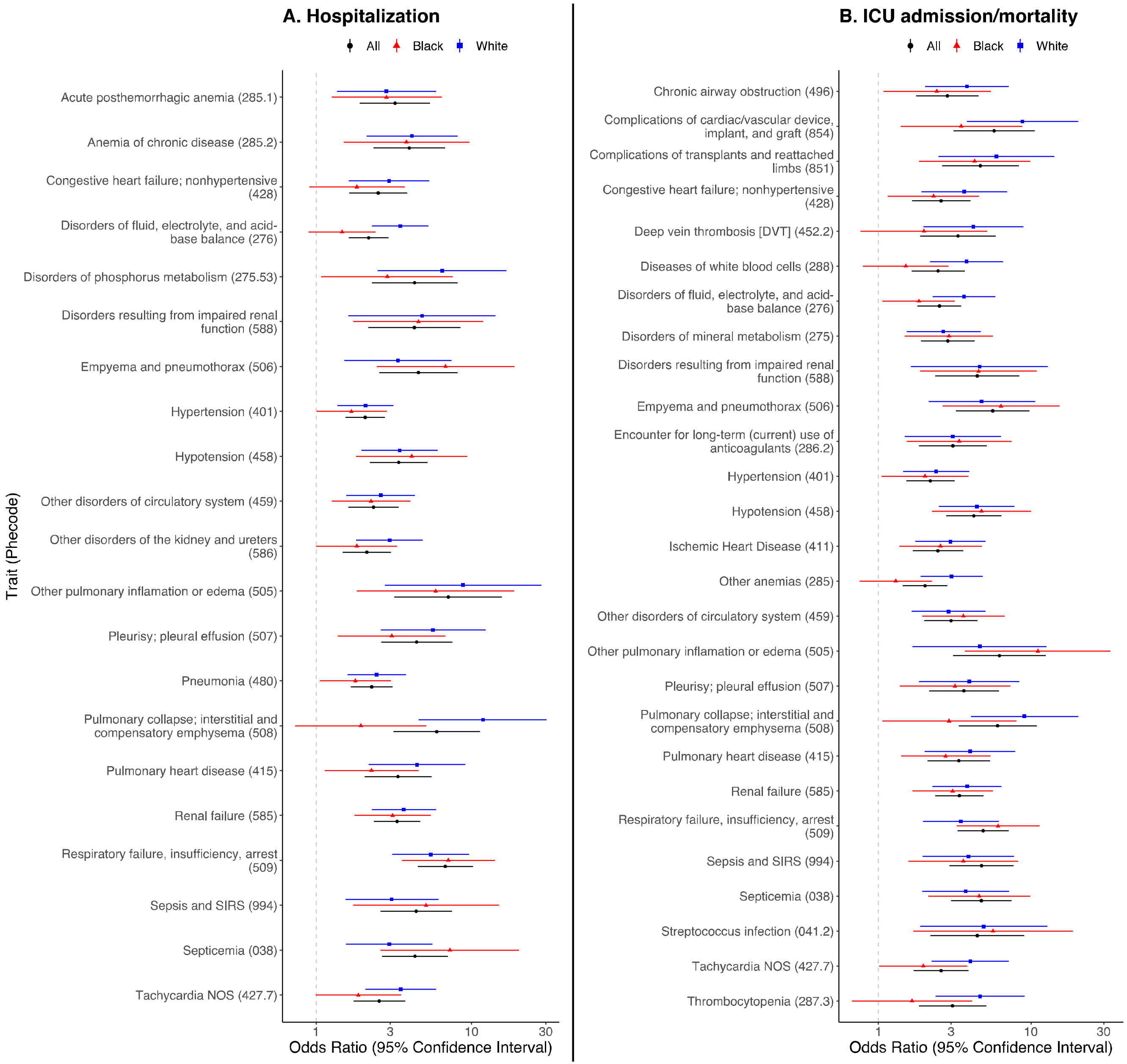

Figure 3. Forest plots of traits associated with poorer prognosis in overall cohort. *A. Hospitalization. B. ICU admission/mortality. Odds ratios and 95% confidence intervals are shown for each trait whose PheWAS code is given in parentheses. Plots show parent codes only. Child codes found in Figure S*

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. Oetjens et al. recently published a PheWAS looking at phenotypes associated with COVID-19-related hospitalization, which is consistent with our results.[23] 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 for increasingly severe prognosis such as pulmonary heart disease, respiratory failure and type 2 diabetes. Our stratified analysis showed that respiratory conditions and mental disorders appear to be associated with more severe outcomes among Whites while coagulation renal disease and heart disease are more strongly associated with severe outcomes among Blacks. **Figures 2B-2D** show that the disease categories that comprise the top 50 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[24] 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 applies here including inaccurate data from billing codes[20] and failure of physicians to report/record problems.[21] 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 primary care records of 17,728,392 adults in England (covering 40% of all patients),[27] 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.

4.1. Conclusion

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 SES covariate, which are important to consider when comparing races [22]. 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 [23]. 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/. We hope the summary data and the phenomic landscape for COVID-19 will help future replication and meta-analysis efforts.

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

Running head

PheWAS of COVID-19 outcomes by race

Acknowledgments

Equal contribution of work

Maxwell Salvatore and Tian Gu contributed equally to this work

Financial support

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).

Conflicts of interest

The authors have no conflicts of interests to declare.

References

1. Denny JC, Ritchie MD, Basford MA, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics **2010**; 26:1205–1210.

2. Gagliano Taliun SA, VandeHaar P, Boughton AP, et al. Exploring and visualizing large-scale genetic associations by using PheWeb. Nat Genet **2020**; 52:550–552.

3. Verma A, Lucas A, Verma SS, 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.

4. Zhao X, Geng X, Srinivasasainagendra V, et al. A PheWAS study of a large observational epidemiological cohort of African Americans from the REGARDS study. BMC Med Genomics **2019**; 12:26.

5. Cai T, Zhang Y, Ho Y-L, et al. Association of Interleukin 6 Receptor Variant With Cardiovascular Disease Effects of Interleukin 6 Receptor Blocking Therapy. JAMA Cardiol **2018**; 3:849.

6. Leppert B, Millard LAC, Riglin L, et al. A cross-disorder PRS-pheWAS of 5 major psychiatric disorders in UK Biobank. PLOS Genet **2020**; 16:e1008185.

7. Microsoft bing COVID-19 Tracker. Available at: https://www.bing.com/covid. Accessed 29 June 2020.

8. Centers for Disease Control and Prevention. Overview of testing for SARS-CoV-2. 2020. Available at: 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 15 June 2020.

9. Brainard J. Scientists are drowning in COVID-19 papers. Can new tools keep them afloat? Sci. Mag. 2020;

10. Price-Haywood EG, Burton J, Fort D, Seoane L. Hospitalization and Mortality among Black Patients and White Patients with Covid-19. N Engl J Med **2020**; :NEJMsa2011686.

11. Brandt EB, Beck AF, Mersha TB. Air pollution, racial disparities, and COVID-19 mortality. J Allergy Clin Immunol **2020**;

12. Laurencin CT, 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.

13. Vahidy FS, Pan AP, Ahnstedt H, et al. 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.

14. Pflugeisen BM, Mou J. Empiric evidence of ethnic disparities in coronavirus positivity in Washington State. Ethn Health **2021**; :1–13.

15. Chang M-H, Moonesinghe R, Truman BI. 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**;

16. Wiley Z, Kubes JN, Cobb J, et al. Age, Comorbid Conditions, and Racial Disparities in COVID-19 Outcomes. J Racial Ethn Health Disparities **2021**;

17. Gu T, Mack JA, Salvatore M, et al. Characteristics Associated With Racial/Ethnic Disparities in COVID-19 Outcomes in an Academic Health Care System. JAMA Netw Open **2020**; 3:e2025197.

18. Michigan Medicine. Indications for COVID-19 diagnostics testing for adult patients in all clinical settings. 2020. Available at: http://www.med.umich.edu/asp/pdf/adult_guidelines/COVID-19-testing.pdf. Accessed 9 June 2020.

19. NaNDA | Social Environment and Health Program. Available at: https://seh.isr.umich.edu/signature-projects/nanda/. Accessed 10 June 2020.

20. Rhodes ET, Laffel LMB, Gonzalez T V., Ludwig DS. Accuracy of Administrative Coding for Type 2 Diabetes in Children, Adolescents, and Young Adults. Diabetes Care **2007**; 30:141–143.

21. Williams C, Mosley-Williams A, C M. Accuracy of provider generated computerized problem lists in the Veterans Administration. AMIA Annu Symp Proc **2007**; :1155.

22. Williams DR, Priest N, Anderson NB. Understanding associations among race, socioeconomic status, and health: Patterns and prospects. Health Psychol **2016**; 35:407–411.

23. Yao H, Chen J-H, Xu Y-F. Patients with mental health disorders in the COVID-19 epidemic. Lancet Psychiatry **2020**; 7:e21.