Identifying Long-Term Effects of SARS-CoV-2 and Their Association with Social Determinants of Health in a Cohort of Over One Million COVID-19 Survivors

Sumit Mukherjee
Microsoft Corporation
Meghana Kshirsagar (Meghana.Kshirsagar@microsoft.com)
Microsoft Corporation
Nicholas Becker
Microsoft Corporation
Yixi Xu
Microsoft Corporation
William B Weeks
Microsoft Corporation
Shwetak Patel
University of Washington
Juan Lavista Ferres
Microsoft Corporation
Michael L. Jackson
Kaiser Permanente Washington

Research Article

Keywords: SARS-CoV-2, COVID-19, long-term effects, social determinants of health, medical claims, health disparities, infectious diseases, observational study.

Posted Date: November 11th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1032897/v1

Abstract

Background: Despite an abundance of information on the risk factors of SARS-CoV-2, large scale studies of long-term effects are lacking. In this paper we analyzed a large medical claims database of US based individuals to identify common long-term effects as well as their associations with various social and medical risk factors.

Methods: The medical claims database was obtained from a prominent US based claims data processing company, namely Change Healthcare. In addition to the claims data, the dataset also consisted of various social determinants of health such as race, income, education level and veteran status of the individuals. A self-controlled cohort design (SCCD) observational study was performed to identify ICD-10 codes whose proportion was significantly increased in the outcome period compared to the control period to identify significant long-term effects. A logistic regression-based association analysis was then performed between identified long-term effects and social determinants of health.

Results: Among the over 1.37 million COVID patients in our datasets we found 36 out of 1,724 3-digit ICD-10 codes to be statistically significantly increased in the post-COVID period (p-value <0.05). We also found one combination of ICD-10 codes, corresponding to 'other anemias' and 'hypertension', that was statistically significantly increased in the post-COVID period (p-value <0.05). Our logistic regression-based association analysis with social determinants of health variables, after adjusting for comorbidities and prior conditions, showed that age and gender were significantly associated with the multiple long-term effects. Race was only associated with 'other sepsis', income was only associated with 'Alopecia areata', while education level was only associated with 'Maternal infectious and parasitic diseases' (p-value <0.05).

Conclusion: We identified several long-term effects of SARS-CoV-2 through a self-controlled study on a cohort of over one million patients. Furthermore, we found that while age and gender are commonly associated with the long-term effects, other social determinants of health such as race, income and education levels have rare or no significant associations.

Background

Since emerging in late 2019, the SARS-CoV-2 virus is known to have infected over 200 million persons globally and caused over 4.5 million deaths.[1] SARS-CoV-2 infection can lead to severe primary illness, including pneumonia and acute respiratory distress syndrome.[2] Infection can also lead to numerous immune-mediated pathologies such as lymphopenia during the acute illness phase.[3] Beyond the initial infection, evidence is accumulating that SARS-CoV-2 infection may cause long-term health complications for some individuals.

While early studies suggest that SARS-CoV-2 infection can cause multiple long-term complications, much remains unknown about the clinical course following SARS-CoV-2. There are few systematic studies of conditions that may be triggered by infection, and risk factors for long-term SARS-CoV-2 complications[8] on small sample sized cohorts. Furthermore, there are no studies exploring the effects of various social and economic factors, that are known to be powerful determinants of population health, on long term effects of SARS-CoV-2.

In this study we utilized claims data from a large sample sized cohort of patients diagnosed with SARS-CoV-2 to study the long-term complications arising due to SARS-CoV-2. Our primary contributions are: i) identification of conditions that are significantly more likely to occur after exposure to SARS-CoV-2, ii) identification of the relative timing of when such conditions become significant, iii) identification of the association of significant long-term effects with various social determinants of health (SDOH) such as race, education level, income etc.

Methods

Data source

Our study uses de-identified United States medical claims records from Change Healthcare collected over a period from April 1, 2018, to Dec 31, 2020, encompassing over 50 million records from over 2 million patients. Every claims record contains information about the diagnoses (in the form of ICD-10 codes), the procedures performed and prescribed drugs. The claims in our dataset include primarily open claims, and a subset of closed payer claims which are normalized for analytics purposes providing sound directional insight for this study. The open claims are derived from broad based healthcare sources and consists of all the medical claims that Change Healthcare processes and for which they have the rights to use. The closed claims are derived from the payer and capture nearly all events that occur during the patient's enrollment period. Roughly 95% of the claims used for this study are commercial and 5% are Medicare Advantage/other types of plans.

In addition to medical claims, we use patient-level social determinants of health (SDOH) data from Change Healthcare. The SDOH attributes included in this study are: i) race, ii) gender, iii) age, iii) income, iv) education level, v) veteran status. Of these attributes, gender and age are obtained from patient claims. SDOH data (other than gender and age) are available for 43.91% of the individuals in the data.

Study population

Our dataset includes all COVID-19 positive patients, identified by the ICD-10 diagnosis codes of U07.1 (COVID-19, virus identified, lab confirmed) or U07.2 (COVID-19, virus not identified but clinically diagnosed) as the principal diagnosis. We defined a subject's *index date* as the date of the SARS-CoV-2 diagnosis and only included patients whose index date was between March 1, 2020, and September 30, 2020. For these patients, we had claims data available between April 1, 2018, to December 31, 2020. The total size of our study population was 2.7 million, reduced to 1.37 million after discarding records with missing fields. Of this group, we possess supplementary SDOH data for 602,025 patients. Henceforth, we shall refer to the cohort of patients for whom we possess the SDOH data as the 'SDOH cohort' and the other patients as the 'non-SDOH cohort'. The non-SDOH cohort is used to first define the long-term effects of interest, as described in the statistical analysis section. We then test the association of certain long-term effect outcomes with the SDOH variables using the SDOH cohort. The descriptive statistics of both cohorts can be found in Table 1. We can see that the populations are qualitatively similar in terms of age and gender.

Table 1
Descriptive statistics of the study cohort

Variable	Category	SDOH fraction	non-SDOH fraction	All fraction
Age	0-20	0.009333996	0.15870638	0.09311697
	21-30	0.080847434	0.118384518	0.10190198
	31-40	0.120001865	0.116660258	0.11812756
	41-50	0.149404386	0.12315018	0.1346784
	51-60	0.208195612	0.153282147	0.17739465
	61-70	0.199574833	0.147831305	0.17055189
	71-80	0.14482287	0.109365417	0.12493478
	80+	0.087819005	0.072619797	0.07929377
Gender	Female	0.611106506	0.5813309	0.59440537
	Male	0.388893494	0.41866777	0.40559389
Veteran status	Non-veteran	0.798827767	x	x
	Veteran	0.201172233	x	x
Race	Asian	0.027929459	x	x
	Black	0.119549412	x	x
	Hispanic	0.189645049	x	x
	White	0.66287608	x	x
Income	Less than $15,000	0.101988374	x	x
	$15,000 - $19,999	0.071895086	x	x
	$20,000 - $29,999	0.105519923	x	x
	$30,000 - $39,999	0.107134593	x	x
	$40,000 - $49,999	0.103200671	x	x
	$50,000 - $74,999	0.201743843	x	x
	$75,000 - $99,999	0.115691476	x	x
	$100,000 - $124,999	0.061549115	x	x
	Greater than $124,999	0.131276918	x	x
Education	Completed High School	0.607909937	x	x
	Completed College	0.269532168	x	x
	Completed Graduate School	0.115415918	x	x
	Attended Vocational/Technical	0.007104643	x	x

Study design

We utilize a self-controlled cohort design (SCCD)[6] in this study. In this design, event rates during a time window after SARS-CoV-2 diagnosis are compared to event rates during a time window prior to diagnosis, where the study population is restricted to patients diagnosed with SARS-CoV-2. The outcome period is defined as beginning 2 months after the index date and continuing through January 31, 2021, the last date for which reliable claims data are present (see Supplementary figure 1). The control period is defined as the three-month period from 10 months to 7 months prior to the index date. This control period begins during the same calendar month as the outcome period, and so should reduce possible confounding by seasonal variations in incidence of events of interest.

Pre-existing comorbidities were defined based on ICD-10 codes assigned to medical encounters during the six-month period from 16 months to 10 months prior to the index date (see Supplementary figure 1). This period does not overlap with the control period, so events during the control period will not also be counted as comorbidities. The Elixhauser comorbidity index[7] was used to define comorbid conditions and their corresponding ICD-10 codes[9].

Statistical Analysis

Identification of statistically significant ICD10 codes that define long term effects – Following common practice, we grouped the ICD10 codes by their first three digits which approximately represents high level health conditions. Relative abundances for each condition (represented by a three-digit ICD10 code) were calculated for both control and post-covid periods. Conditions that occurred in less than 0.01% of the post-covid population were discarded to limit the analysis to conditions that were present in a large enough population. A 2-proportion one-way z-test was performed to identify conditions that were significantly higher in the post-covid period, compared to the control period. The significance level was set to 0.05 with multiple testing correction using the Bonferroni method, for this and all subsequent analyses unless mentioned otherwise. This analysis was done on the non-SDOH cohort. However, for the purposes of validation, we also replicated the same analysis on the SDOH cohort.

Identification of month-wise long-term effects – To study the month-wise prevalence of the long-term effects that we identified, we perform the same analysis as described in the previous section, on one month long post-covid and matched control periods shown in Supplementary figure 1. The analysis was done for months 3, 4 and 5 post-covid. Since we had used the non-SDOH cohort to identify the long-term effects, to prevent 'double dipping', we performed this analysis on the SDOH cohort.

Identification of co-occurring long-term effects – Identification of frequently co-occurring conditions was done using a data mining technique known as market-basket analysis or affinity analysis[7]. Briefly, affinity analysis identifies co-occurring items (long-term effects in our case) in the data by comparing the observed co-occurrence frequency with the expected co-occurrence frequency (assuming that the co-occurrence was purely random). We first performed market affinity analysis with (support$\geq$0.01, lift$\geq$1) on the post-covid period to identify co-occurring conditions. We then identified the relative proportion of patients who experienced each 'basket' of conditions in the post-covid and control periods. Finally, we performed a 2-proportion one-way z-test to identify which buckets were significantly overrepresented in the post-covid period compared to the control period. This analysis was performed on the non-SDOH cohort.

Studying associations of SDOH variables with long-term effects – Association testing of SDOH variables with each significant long-term effect was done using a logistic regression model, which adjusted for comorbidities and presence of the same long-term conditions in the control period (prior events). The mathematical model can be expressed as:

$$\log\left(\frac{p_m}{1-p_m}\right) = \beta_0 + \sum_{i \in SDOH} \beta_i^{SDOH} X_i + \sum_{j \in Comorb} \beta_j^{Comorb} X_j + \sum_{k \in PriorEvents} \beta_k^{PriorEvents} X_k$$

Where, $p_m = \Pr(Y_m = 1)$ is the probability of long-term effect *m* occurring. Prior to performing the logistic regression, we performed feature selection using a chi-squared test of independence between each outcome and independent variable. Only variables that met a significance level of 0.05 were used in the logistic regression. However, a Bonferroni corrected p-value (correcting for m outcomes) was used to determine significant associations in the logistic regression model. The selected baseline categories were: Race-White, Education – Completed college, Income-greater than $124,999, Gender-male, Non-veteran, Age-31-40.

Results

Long term effects of COVID-19 and cooccurring conditions

The study population consisted of 1,371,110 patients with an ICD-10 diagnosis code for SARS-CoV-2 infection. This population was predominantly older (mean age 55.36 years) and female (59.44%). Among the 43.91% of the cohort with SDOH data available, 66% were non-Hispanic White.

Of the 1,724 3-digit ICD10 codes considered, 36 met the significance threshold after the Bonferroni correction (Table 2). These ICD10 codes are reported in Table 1, along with the rate of occurrence in the control and post-COVID period for the non-SDOH and SDOH cohorts. The identification of significant ICD10 codes was done using the p-values from the non-SDOH cohort only and the SDOH cohort numbers are only reported for validation purposes. We find that 25 of 36 ICD10 codes are significant in both cohorts, thereby indicating consistency of our findings.

Several broad categories of associations are notable. Unsurprisingly, multiple codes suggest ongoing pulmonary complications, such as J12 (viral pneumonia) and J80 (acute respiratory distress syndrome). Cardiac and thrombotic events comprise a second category (e.g. I40 [acute myocarditis], I82 (other venous embolism and thrombosis). A third category is apparent complications of treatment during acute SARS-CoV-2 infection, such as codes J95 (intraoperative and postprocedural complications), K94 (complications of artificial openings of the digestive system), and L89 (pressure ulcer). A fourth is malnutrition or wasting, such as codes E43, E44, and E46 (protein-calorie malnutrition) or R34 (cachexia).

Next, we investigated two of the significant 3-digit digit codes: D84 (other immunodeficiencies) and G93 (other disorders of brain). We evaluated whether the significant associations with these codes were driven by specific sub-codes. For these two ICD10 code families, we identified the constituent ICD10 codes that were significantly increased in the post-COVID period compared to the control period using the same method as above. However, unlike the previous analysis, here we only look at the SDOH cohort, since the significant 3-digit codes were identified on the non-SDOH cohort. Of the 19 constituent codes, we find 5 that meet our Bonferroni corrected significance threshold (see Supplementary table 1). The only significant sub-code to D89 was D89.9 (immunodeficiency, unspecified). Four sub-codes were significant from G93. Of these, the most significant was G93.3 (postviral fatigue syndrome), which was 4.4 times more common in the post-COVID period than the pre-COVID period.

Using association analysis on the post-COVID period, we identified several co-occurring long-term conditions. We then looked at the co-occurring conditions that are significantly overrepresented in the post-COVID period compared to the control period (see Supplementary table 2) after excluding those that contained Z codes. Only one co-occurring condition was found to meet our significance threshold: D64 (other anemias) and I10 (essential hypertension).

Table 2

ICD10 codes that were observed in a significantly higher proportion in the post-covid window compared to the control window. ICD10 codes that are significant for both non-SDOH and SDOH cohorts are in bold.

ICD10	Description	Non-SDOH cohort			SDOH cohort		
		Control%	Post%	p-value	Control%	Post%	p-value
A41	Other sepsis	0.667	0.813	2.0E-26	0.684	0.761	3.9E-07
B49	Unspecified mycosis	0.012	0.019	5.0E-05	0.012	0.021	1.3E-04
B94	Sequelae of infectious and parasitic diseases	0.002	0.041	3.6E-63	0.001	0.040	1.3E-49
D84	Other immunodeficiencies	0.036	0.052	1.1E-06	0.043	0.058	8.4E-05
E43	Severe protein-calorie malnutrition	0.126	0.214	4.8E-40	0.131	0.204	3.5E-23
E44	Medium/Mild protein-calorie malnutrition	0.158	0.200	1.8E-10	0.167	0.201	6.7E-06
E46	Unspecified protein-calorie malnutrition	0.148	0.232	2.4E-33	0.144	0.224	3.9E-25
G72	Unspecified myopathies	0.032	0.107	5.4E-71	0.042	0.144	5.0E-75
G92	Toxic encephalopathy	0.090	0.112	5.0E-06	0.093	0.111	6.9E-04
G93	Other disorders of brain	0.702	0.847	5.6E-25	0.702	0.846	8.7E-20
I26	Pulmonary embolism	0.206	0.309	1.7E-36	0.293	0.392	4.5E-21
I40	Acute myocarditis	0.002	0.010	2.9E-10	0.002	0.006	5.5E-04
I46	Cardiac arrest	0.030	0.099	3.6E-63	0.030	0.095	2.1E-46
I82	Other venous embolism/thrombosis	0.419	0.484	7.2E-10	0.539	0.574	4.9E-03
J12	Viral pneumonia	0.055	0.938	0.0E+00	0.055	1.110	0.0E+00
J69	Pneumonitis due to solids and liquids	0.151	0.190	2.3E-09	0.128	0.160	1.4E-06
J80	Acute respiratory distress syndrome	0.018	0.127	2.6E-137	0.019	0.137	1.3E-118
J84	Other interstitial pulmonary diseases	0.218	0.276	1.5E-13	0.326	0.362	4.1E-04
J91	Pleural effusion	0.036	0.049	4.6E-05	0.041	0.053	1.3E-03
J93	Pneumothorax and air leak	0.036	0.067	7.9E-18	0.040	0.070	5.4E-13
J95	Intraoperative/postprocedural complications	0.043	0.072	2.8E-14	0.040	0.072	5.0E-14
J96	Respiratory failure	1.065	1.822	0.0E+00	1.178	1.910	4.5E-233
K94	Complications of artificial openings of the digestive system	0.085	0.114	9.2E-09	0.054	0.093	4.0E-15
L63	Alopecia areata	0.029	0.041	4.0E-05	0.038	0.050	7.7E-04
L64	Androgenic alopecia	0.013	0.026	9.1E-09	0.018	0.033	3.4E-07

ICD10	Description	Non-SDOH cohort			SDOH cohort		
		Control%	Post%	p-value	Control%	Post%	p-value
L65	Telogen effluvium	0.116	0.354	3.9E-204	0.142	0.439	5.5E-202
L89	Pressure ulcer	0.335	0.605	3.9E-132	0.368	0.674	2.3E-120
M30	Polyarteritis nodosa and related conditions	0.005	0.013	1.1E-07	0.004	0.003	9.0E-01
O98	Maternal infectious and parasitic diseases	0.057	0.079	2.7E-07	0.041	0.045	1.4E-01
R13	Aphagia and dysphagia	1.077	1.249	1.8E-23	1.069	1.168	1.1E-07
R43	Disturbances of smell and taste	0.026	0.147	1.2E-143	0.040	0.153	2.1E-89
R57	Shock	0.052	0.095	1.1E-23	0.054	0.089	1.2E-13
R64	Cachexia	0.032	0.047	3.3E-06	0.033	0.045	4.7E-04
R65	Systemic inflammation and infection	0.238	0.320	3.2E-22	0.256	0.304	3.2E-07
R77	Other abnormalities of plasma proteins	0.042	0.077	2.2E-19	0.048	0.084	1.3E-14
R78	Findings of drugs and other substances, not normally found in blood	0.166	0.207	2.0E-09	0.173	0.212	3.8E-07

Persistent and fleeting long term effects

We explored the timing of significant associations at the 3-digit code level, within one-month windows post-COVID (Table 3). Several code groups were significantly elevated early on but appeared to have resolved by month 5 post-onset, while others were elevated through the full follow-up period. Of the categories of codes identified earlier, no category saw resolution of all codes by 5 months post-onset, and no category saw persistence of all codes through 5 months. Only L64 (androgenic alopecia) became significantly elevated at month 5 after not being elevated previously.

Table 3

ICD10 codes that are significantly over-present in the one month long post-covid periods compared to corresponding month-long control period.

ICD10		Month 3	Month 4	Month 5
D84	Other immunodeficiencies	No	No	No
L63	Alopecia areata	No	No	No
B49	Unspecified mycosis	Yes	No	No
G92	Toxic encephalopathy	Yes	No	No
I82	Other venous embolism/thrombosis	Yes	No	No
J69	Pneumonitis due to solids and liquids	Yes	No	No
J91	Pleural effusion	Yes	No	No
K94	Complications of artificial openings of the digestive system	Yes	No	No
M30	Polyarteritis nodosa and related conditions	Yes	No	No
O98	Maternal infectious and parasitic diseases	Yes	No	No
R78	Findings of drugs and other substances, not normally found in blood	Yes	No	No
L64	Androgenic alopecia	No	No	Yes
A41	Other sepsis	Yes	Yes	No
E44	Medium/Mild protein-calorie malnutrition	Yes	Yes	No
G93	Other disorders of brain	Yes	Yes	No
I40	Acute myocarditis	Yes	Yes	No
J84	Other interstitial pulmonary diseases	Yes	Yes	No
J93	Pneumothorax and air leak	Yes	Yes	No
J95	Intraoperative/postprocedural complications	Yes	Yes	No
R13	Aphagia and dysphagia	Yes	Yes	No
R57	Shock	Yes	Yes	No
R64	Cachexia	Yes	Yes	No
R65	Systemic inflammation and infection	Yes	Yes	No
R77	Other abnormalities of plasma proteins	Yes	No	Yes
B94	Sequelae of infectious and parasitic diseases	Yes	Yes	Yes
E43	Severe protein-calorie malnutrition	Yes	Yes	Yes
E46	Unspecified protein-calorie malnutrition	Yes	Yes	Yes
G72	Unspecified myopathies	Yes	Yes	Yes
I26	Pulmonary embolism	Yes	Yes	Yes

ICD10		Month 3	Month 4	Month 5
I46	Cardiac arrest	Yes	Yes	Yes
J12	Viral pneumonia	Yes	Yes	Yes
J80	Acute respiratory distress syndrome	Yes	Yes	Yes
J96	Respiratory failure	Yes	Yes	Yes
L65	Telogen effluvium	Yes	Yes	Yes
L89	Pressure ulcer	Yes	Yes	Yes
R43	Disturbances of smell and taste	Yes	Yes	Yes

Associations with SDOH variables

As described in the Methods section, we estimated the association of every significant 3-digit code group with the SDOH variables, adjusting for comorbidities and the presence of the same conditions in the control period (Figure 1). Interestingly, for this population, race was only significantly associated with A41 (other sepsis). Similarly, only L63 (alopecia areata) was significantly associated with some income categories. Female gender showed negative association with A41, I40 and positive association with L63. In contrast, age was significantly associated with most of the long-term effect conditions. A complete table of association results for SDOH, as well as comorbidity and prior conditions can be found in Supplementary table 2.

Discussion

Long-term sequelae of SARS-CoV-2 infection has received substantial attention in scientific literature, legacy media, and social media. Numerous studies have explored the long-term health concerns among people previously infected with SARS-CoV-2. However, prior research has generally been limited by small sample sizes or lack of comparison groups. This study addresses these limitations by comparing the frequency of ICD-10-coded diagnoses during pre- and post-COVID-19 periods among more than 1.37 million study subjects.

Our first notable finding was that SARS-CoV-2 infection was associated with subsequent codes for malnutrition/wasting. This accords with a prior cohort study that found 31% of patients hospitalized for COVID-19 lost ≥5% of their body weight at roughly three weeks post-discharge compared to admission body weight[10]. Malnutrition is known to be a risk factor for pneumonia in diverse populations, including community-dwelling seniors[11], seniors in long-term care settings[12] and children in resource-poor settings[13]. COVID-19-induced malnutrition/wasting thus may pre-dispose patients to future episodes of pneumonia or other respiratory disease.

A second notable finding is the frequency of codes likely related to complications of COVID-19 hospitalization or treatment. These include conditions such as pressure ulcers and enterostomy. Even among those successfully treated for COVID-19, hospitalization and treatment can have long-term impacts on health independent of physiologic damage caused by infection.

Third, the ICD-10 codes associated with SARS-CoV-2 infection in this study support many self-reported complications from surveys of COVID-19 patients. Post-viral fatigue syndrome (G93.3) was significantly associated with SARS-CoV-2 infection in this study. This matches patient-reported data, where fatigue is commonly reported[14]. Other codes that

match common patient-reported outcomes include persistent respiratory symptoms, myalgia, and ongoing disturbances to taste/smell.

Interestingly, some post-COVID-19 complications frequently reported by patients did not show up in the ICD-10 data. Examples include headache, anxiety, and sleep disturbances. It could be that these complications are frequently experienced by patients but do not result in medical encounters; or that other, more severe symptoms ended up in the ICD-10 coded data instead; or that these symptoms are actually not elevated among persons infected with SARS-CoV-2 relative to pre-infection periods. Further research will be needed to distinguish between these possibilities.

One recent study by Murk et al applied a similar design to medical claims data to identify short-term (<31 days) complications of SARS-CoV-2 infection[15]. Like our study, that study found elevated risks of codes associated with respiratory infection and respiratory complications, disturbance of taste/smell, and cardiovascular conditions such as cardiac arrest. The most notable difference is that Murk and colleagues found associations with acute kidney failure, which was not observed in the present study.

Several limitations of this study are important to highlight. First, the self-controlled cohort design assumes that differences in event frequencies after vs. before SARS-CoV-2 diagnosis are causally related to infection. Other temporal trends in these diagnoses unrelated to infection could bias effect estimates either upward or downward. Second, this study relies on ICD-10 codes assigned to medical encounters. ICD-10 codes are imperfect proxies for actual disease and do not allow evaluation of complications that are not severe enough to warrant medical attention. These codes also do not include indicators of disease severity. Finally, SDOH data were only available for 43.91% of our population. This data may not be missing at random, and the group with SDOH data may not be representative of the underlying population. SDOH associations should thus be interpreted with caution.

Conclusions

In this study, we have identified potential complications of SARS-CoV-2 infection that require ongoing medical evaluation and care. This builds on and supplements patient-reported outcomes and illustrates the potential for long-term complications of SARS-CoV-2 infection. Furthermore, we find that after controlling for prior health conditions, only age and gender consistently show significant associations with the identified long-term effects.

Declarations

Ethics approval and consent to participate

This study does not constitute as human subjects research due to the usage and reporting of only deidentified observational data as determined by the ethics committee of the University of Washington School of Medicine. An ethics approval waiver was received from the ethics committee of the University of Washington School of Medicine. A consent waiver was received from the ethics committee of the University of Washington School of Medicine. All methods were carried out in accordance with the relevant guidelines and regulations.

Consent for publication

Not applicable.

Availability of data and materials

The study is conducted on proprietary data from a medical claims processing company (Change Healthcare) and hence cannot be shared publicly. Requests for data access should be directed to Change Healthcare. All programming code will be made available upon reasonable request.

Competing interests

The authors declare no competing interests.

Funding

Not applicable.

Authors' contributions

MLJ designed the study. SM, MK performed the analyses. YX, WBW, NB, JLF, MLJ, SP, SM and MK wrote the paper.

Acknowledgements

The authors acknowledge Mohammed Nasir of Microsoft for thoughtful discussions about the paper. The authors also acknowledge Change Healthcare for providing them access to the dataset.

References

1. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Johns Hopkins University, 2020. (Accessed 27 April 2020, 2020, at https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6.)
2. Kakodkar P, Kaka N, Baig MN. A Comprehensive Literature Review on the Clinical Presentation, and Management of the Pandemic Coronavirus Disease 2019 (COVID-19). Cureus 2020;12:e7560.
3. Yuki K, Fujiogi M, Koutsogiannaki S. COVID-19 pathophysiology: A review. Clin Immunol 2020;215:108427.
4. Wang Y, Dong C, Hu Y, et al. Temporal Changes of CT Findings in 90 Patients with COVID-19 Pneumonia: A Longitudinal Study. Radiology 2020;296:E55-E64.
5. Galeotti C, Bayry J. Autoimmune and inflammatory diseases following COVID-19. Nat Rev Rheumatol 2020;16:413-4.
6. Ryan PB, Schuemie MJ, Madigan D. Empirical performance of a self-controlled cohort method: lessons for developing a risk identification and analysis system. Drug Saf 2013;36 Suppl 1:S95-106.
7. Elixhauser, A., Steiner, C., Harris, D. R., & Coffey, R. M. (1998). Comorbidity measures for use with administrative data. Medical care, 8-27.
8. Murk, W., Gierada, M., Fralick, M., Weckstein, A., Klesh, R., & Rassen, J. A. (2021). Diagnosis-wide analysis of COVID-19 complications: an exposure-crossover study. Cmaj, 193(1), E10-E18.
9. Elixhauser Comorbidity Software Refined for ICD-10-CM. URL: https://www.hcup-us.ahrq.gov/toolssoftware/comorbidityicd10/comorbidity_icd10.jsp [accessed 2021-10-01].
10. Di Filippo, L., De Lorenzo, R., D'Amico, M., Sofia, V., Roveri, L., Mele, R., ... & Conte, C. (2021). COVID-19 is associated with clinically significant weight loss and risk of malnutrition, independent of hospitalisation: A post-hoc analysis of a prospective cohort study. Clinical Nutrition, 40(4), 2420-2426.
11. Uematsu, H., Yamashita, K., Kunisawa, S., & Imanaka, Y. (2021). Prediction model for prolonged length of stay in patients with community-acquired pneumonia based on Japanese administrative data. Respiratory Investigation,

59(2), 194-203.

12. Graversen, S. B., Pedersen, H. S., Sandbaek, A., Foss, C. H., Palmer, V. J., & Ribe, A. R. (2021). Dementia and the risk of short-term readmission and mortality after a pneumonia admission. PloS one, 16(1), e0246153.
13. Chowdhury, F., Shahid, A. S. M. S. B., Ghosh, P. K., Rahman, M., Hassan, M. Z., Akhtar, Z., ... & Chisti, M. J. (2020). Viral etiology of pneumonia among severely malnourished under-five children in an urban hospital, Bangladesh. PloS one, 15(2), e0228329.
14. Moreno-Pérez, O., Merino, E., Leon-Ramirez, J. M., Andres, M., Ramos, J. M., Arenas-Jiménez, J., ... & COVID19-ALC research group. (2021). Post-acute COVID-19 syndrome. Incidence and risk factors: A Mediterranean cohort study. Journal of Infection, 82(3), 378-383.
15. Murk, W., Gierada, M., Fralick, M., Weckstein, A., Klesh, R., & Rassen, J. A. (2021). Diagnosis-wide analysis of COVID-19 complications: an exposure-crossover study. Cmaj, 193(1), E10-E18.

Figures

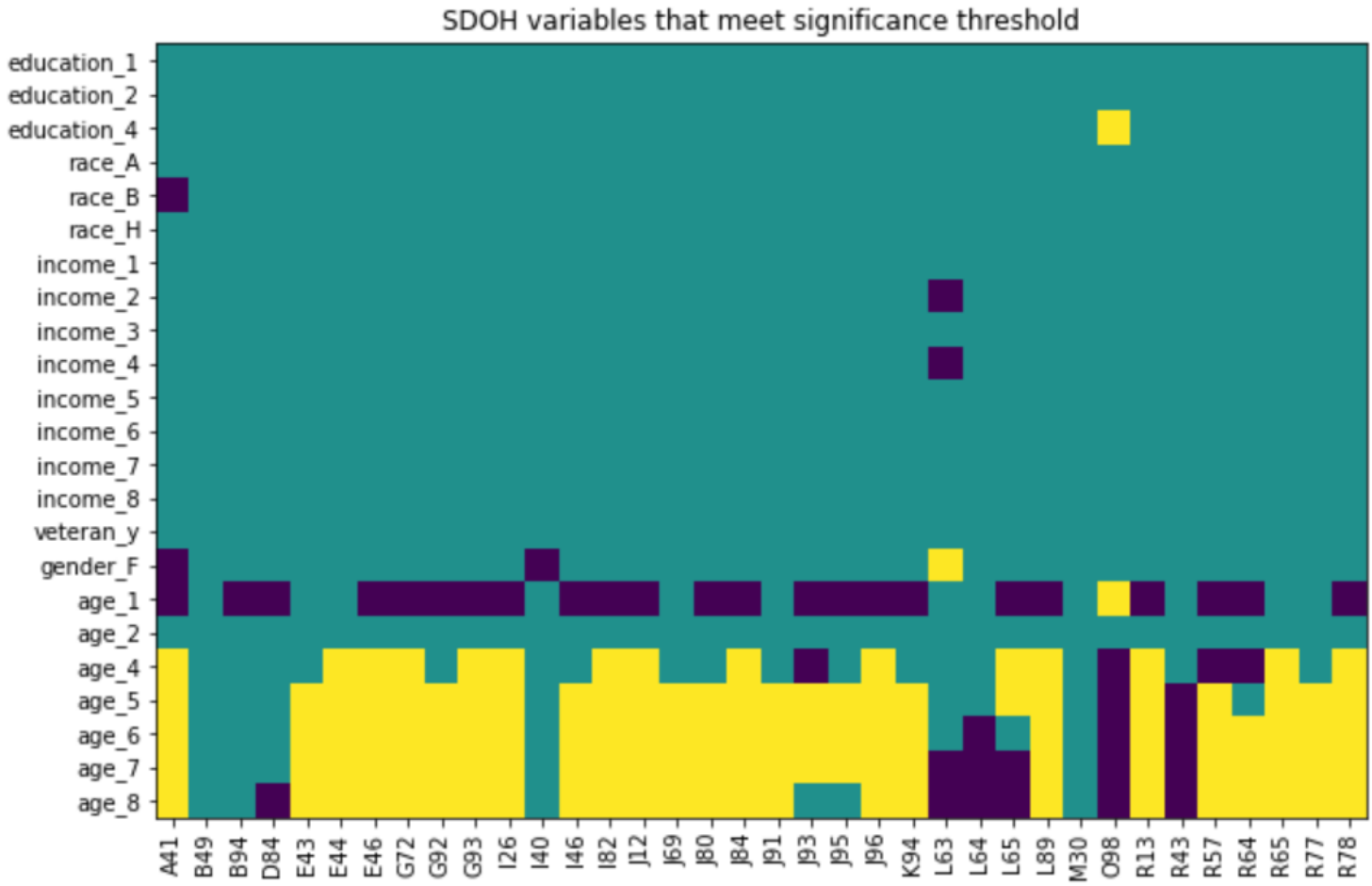

Figure 1

Association analysis of SDOH variables with ICD10 codes corresponding to long term effects of COVID-19. Green indicates no association, blue indicates a negative association, yellow indicates a positive association.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalInfoBMC.docx