RESEARCH ◼ SEPSIS

Diagnosis-wide analysis of COVID-19 complications: an exposure-crossover study

William Murk PhD MPH, Monica Gierada MPH, Michael Fralick MD PhD SM, Andrew Weckstein BA, Reyna Klesh MS, Jeremy A. Rassen ScD

ABSTRACT

BACKGROUND: Many studies reporting coronavirus disease 2019 (COVID-19) complications have involved case series or small cohorts that could not establish a causal association with COVID-19 or provide risk estimates in different care settings. We sought to study all possible complications of COVID-19 to confirm previously reported complications and to identify potential complications not yet known.

METHODS: Using United States health claims data, we compared the frequency of all *International Classification of Diseases, 10th Revision, Clinical Modification* (ICD-10-CM) diagnosis codes occurring before and after the onset of

the COVID-19 pandemic in an exposure-crossover design. We included patients who received a diagnosis of COVID-19 between Mar. 1, 2020, and Apr. 30, 2020, and computed risk estimates and odds ratios (ORs) of association with COVID-19 for every ICD-10-CM diagnosis code.

RESULTS: Among 70 288 patients with COVID-19, 69 of 1724 analyzed ICD-10-CM diagnosis codes were significantly associated with COVID-19. Disorders showing both strong association with COVID-19 and high absolute risk included viral pneumonia (OR 177.63, 95% confidence interval [CI] 147.19–214.37, absolute risk 27.6%), respiratory failure (OR 11.36, 95% CI 10.74–12.02,

absolute risk 22.6%), acute kidney failure (OR 3.50, 95% CI 3.34–3.68, absolute risk 11.8%) and sepsis (OR 4.23, 95% CI 4.01–4.46, absolute risk 10.4%). Disorders showing strong associations with COVID-19 but low absolute risk included myocarditis (OR 8.17, 95% CI 3.58–18.62, absolute risk 0.1%), disseminated intravascular coagulation (OR 11.83, 95% CI 5.26–26.62, absolute risk 0.1%) and pneumothorax (OR 3.38, 95% CI 2.68–4.26, absolute risk 0.4%).

INTERPRETATION: We confirmed and provided risk estimates for numerous complications of COVID-19. These results may guide prognosis, treatment decisions and patient counselling.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a novel strain of coronavirus that has been identified as the cause of the coronavirus disease 2019 (COVID-19) pandemic. As of Nov. 20, 2020, more than 50 million people have received a diagnosis of COVID-19 globally.[1] The clinical spectrum of disease is wide and can range from symptoms typical of the common cold to respiratory failure and death.[2] Most patients have mild symptoms and can be managed as outpatients, but as many as 20% have a severe form of the disease requiring admission to hospital, commonly presenting with hypoxia secondary to pneumonia.[3]

Studies also show that COVID-19 is associated with a wide variety of nonrespiratory sequelae, including endothelial, thrombotic, cardiac, inflammatory, neurologic and other complications.[4–9] Whether these associations are causal is not well established, as many of these findings originate from case reports, which are prone to publication bias and cannot provide risk

estimates, or from cohort studies that often do not provide relative risk estimates.

An alternative strategy for identifying potential complications of COVID-19 is studying all possible complications as captured in *International Classification of Diseases, 10th Revision, Clinical Modification* (ICD-10 CM) diagnosis codes, which allows for the discovery of unreported complications and can confirm previously identified ones. The objective of our study was to analyze all diagnoses associated with COVID-19, to identify those that could be complications of the disease and to present both the absolute risk and relative odds of any complications identified.

Methods

Data source

This study used de-identified United States medical claims from HealthVerity's Marketplace data set. These data are

claims from nationally representative health plans that encompass all major payer types (commercial, Medicaid and Medicare). The data include open claims, which are sourced in near real time from practice management systems, billing systems and claims clearinghouses, as well as closed claims, which are sourced from insurance providers and payers. Data from open claims can be captured within days of a health care encounter, while data from closed claims have longer lag times. Thus, open claims provide more recent data, and closed claims encompass a more complete view of a patient's interactions with the health care system. These data sources are further described in Appendix 1, Methods, available at www.cmaj.ca/lookup/doi/10.1503/cmaj.201686/tab-related-content. Claims were available with dates of service from Nov. 1, 2019, through May 30, 2020.

Study population

We included all patients who had had at least 1 medical encounter related to COVID-19, defined as any claim with an ICD-10-CM diagnosis code of U07.1 (COVID-19, virus identified) or B97.29 (other coronavirus as the cause of diseases classified elsewhere) that occurred within the period from Mar. 1, 2020, to Apr. 30, 2020. Apr. 30 was the last eligibility date to allow for a month of possible follow-up for all patients. Codes were selected based on US Centers for Disease Control and Prevention (CDC) coding guidance for confirmed COVID-19 cases.[10,11] The first eligible claim within the study period was the index date. Patients with a code for U07.1, B97.29 or B97.21 (SARS-associated coronavirus) occurring before index, and those with a missing year of birth, were excluded. Patients with a code for Z03.818 occurring between 7 days before index and 30 days after index were also excluded, because this code indicates that SARS-CoV-2 exposure has been ruled out.[10] To ensure that all analyzed patients were observable before their COVID-19 diagnosis, patients were excluded if they did not have any medical claim between 120 days before index and 30 days before index. We also assessed subgroups of patients treated only on an outpatient basis, patients treated as inpatients, and patients who were admitted to the intensive care unit (ICU) (Appendix 1, Methods).

Study design

We used an exposure-crossover design[12] to identify diagnoses whose odds increased after the onset of COVID-19. This design minimizes unmeasured confounding by comparing a patient's condition upon COVID-19 diagnosis (the "hazard period") with their own condition at a previous time (the "baseline period"). Any increased odds of a condition are likely attributable to COVID-19 sequelae or treatment.

To capture activity related to COVID-19 that occurred just before the patient received a diagnosis of the disease (Appendix 1, eFigure 2), we defined the hazard period as the 7 days before the index date through the 30 days after index. We defined the baseline period as the 120 days before the index date through 30 days before index. We chose this 90-day window as it was long enough to capture most chronic conditions while ensuring equivalent capture of baseline status for all patients. We considered the 23 days between the baseline and hazard period as a period of

unknown COVID-19 disease state (Appendix 1, eFigure 2) and excluded that person-time from the analysis. We also performed sensitivity analyses to assess the effect of the duration of the baseline period by conducting an analysis where the baseline period was identical in length to the hazard period (days 68 through 30 before index; 38 days total). The study design is illustrated in Appendix 1, eFigure 3.

Statistical analysis

To identify ICD-10-CM codes associated with COVID-19, we paired observations of each patient in the baseline and hazard periods and computed matched odds ratios (ORs) to estimate the strength of association of the code with COVID-19, as well as McNemar's test p values.[13] We used exact tests if the number of discordant pairs was less than 25. ICD-10-CM codes are constructed as a 3-character header indicating system classification and disease and can be followed by up to 4 digits offering greater specificity.[14,15] For our primary analysis, we analyzed all 1724 ICD-10-CM diagnosis codes present in the data set, aggregated at the header level (e.g., "R43 – Disturbances of smell and taste"). For any header code from the primary analysis found to be significantly associated, we also analyzed it at the individual code level (e.g., "R43.0 – Anosmia"). For our secondary analysis, we analyzed all 64 931 individual ICD-10-CM codes. We applied Bonferroni-corrected significance thresholds, defined as a nominal type 1 error of 0.05 divided by the number of analyzed codes (2.9E-05 and 7.7E-07 for primary and secondary analyses, respectively).[16] We excluded codes used to identify cases (B97 and U07). A candidate COVID-19 complication was any code that increased in odds with COVID-19 at a Bonferroni-corrected level of significance.

We also calculated the absolute risk of becoming newly diagnosed with each ICD-10-CM code upon having a COVID-19 diagnosis. This was the percentage of patients with the code in the hazard period, calculated among patients who did not have the code in the baseline period (Appendix 1, eFigure 3).

Ethics approval

This study was approved under exemption by the New England Institutional Review Board (#1–9757–1).

Results

We identified 70 288 patients with a diagnosis of COVID-19 (Table 1). Excluded patients are described in Appendix 1, eFigure 4. Of these patients, 53.4% were admitted to hospital and 4.7% to the ICU. The median age was 65 years, and 55.8% were female. The 5 most common states of residence of patients were New York (19.2%), New Jersey (10.2%), Michigan (9.3%), Pennsylvania (7.5%) and Illinois (7.2%). The median numbers of diagnosis codes in the baseline and hazard period are shown in Appendix 1, eTable 1. Baseline prevalence estimates of selected conditions that may affect the risk of COVID-19 or its sequelae are listed in Table 1.

Identification of candidate COVID-19 complications

Among all 1724 diagnosis codes considered at the code header level in our primary analysis, we identified 69 codes that

Table 1: Demographic characteristics and baseline conditions of patients with COVID-19*

Characteristic	No. (%)† of all patients n = 70 288	No. (%)† of outpatients‡ n = 32 745	No. (%)† of inpatients§ n = 37 543	No. (%)† of patients in ICU¶ n = 3269
Percent of total, n = 70 288	100	46.6	53.4	4.7
Age, yr; median (IQR)	65 (51–77)	56 (41–70)	71 (59–81)	69 (59–78)
Age, yr; category				
0 to < 20	1722 (2.4)	1387 (4.2)	335 (0.9)	81 (2.5)
20 to < 30	3212 (4.6)	2582 (7.9)	630 (1.7)	53 (1.6)
30 to < 40	4859 (6.9)	3605 (11.0)	1254 (3.3)	99 (3.0)
40 to < 50	6769 (9.6)	4524 (13.8)	2245 (6.0)	200 (6.1)
50 to < 60	11 220 (16.0)	6264 (19.1)	4956 (13.2)	435 (13.3)
60 to < 70	14 457 (20.6)	6075 (18.6)	8382 (22.3)	773 (23.6)
70 to < 80	13 345 (19.0)	4324 (13.2)	9021 (24.0)	946 (28.9)
≥ 80	14 704 (20.9)	3984 (12.2)	10 720 (28.6)	682 (20.9)
Sex				
Female	39 213 (55.8)	19 586 (59.8)	19 627 (52.3)	1584 (48.4)
Male	30 042 (42.7)	12 625 (38.6)	17 417 (46.4)	1643 (50.3)
Unknown	1033 (1.5)	534 (1.6)	499 (1.3)	42 (1.3)
Baseline (pre-COVID-19) conditions				
Acute kidney failure	3798 (5.4)	831 (2.5)	2967 (7.9)	323 (9.9)
Asthma	4274 (6.1)	2042 (6.2)	2232 (5.9)	238 (7.3)
Atrial fibrillation or flutter	5667 (8.1)	1528 (4.7)	4139 (11.0)	381 (11.7)
Bipolar disorder or depression	11 035 (15.7)	4139 (12.6)	6896 (18.4)	481 (14.7)
Cerebral infarction	2726 (3.9)	664 (2.0)	2062 (5.5)	138 (4.2)
Chronic kidney disease	9856 (14.0)	2363 (7.2)	7493 (20.0)	778 (23.8)
Chronic liver disease	662 (0.9)	195 (0.6)	467 (1.2)	42 (1.3)
Chronic obstructive pulmonary disease	7424 (10.6)	2056 (6.3)	5368 (14.3)	534 (16.3)
Dementia	8477 (12.1)	2335 (7.1)	6142 (16.4)	323 (9.9)
Diabetes mellitus, type 1	1088 (1.5)	361 (1.1)	727 (1.9)	79 (2.4)
Diabetes mellitus, type 2	19 452 (27.7)	6446 (19.7)	13 006 (34.6)	1341 (41.0)
Dyslipidemia	19 479 (27.7)	7378 (22.5)	12 101 (32.2)	1323 (40.5)
Heart failure	8072 (11.5)	1989 (6.1)	6083 (16.2)	593 (18.1)
Hypertension, essential	28 746 (40.9)	10 333 (31.6)	18 413 (49.0)	1789 (54.7)
Inflammatory bowel disease	413 (0.6)	188 (0.6)	225 (0.6)	28 (0.9)
Ischemic heart disease	8959 (12.7)	2445 (7.5)	6514 (17.4)	679 (20.8)
Malignant disease	4184 (6.0)	1310 (4.0)	2874 (7.7)	270 (8.3)
Peripheral vascular disease	6468 (9.2)	1680 (5.1)	4788 (12.8)	429 (13.1)
Psychotic disorders	3221 (4.6)	926 (2.8)	2295 (6.1)	135 (4.1)
Rheumatic disease	1918 (2.7)	789 (2.4)	1129 (3.0)	131 (4.0)
Venous thromboembolism	1960 (2.8)	529 (1.6)	1431 (3.8)	129 (3.9)

Note: COVID-19 = coronavirus disease 2019, ICU = intensive care unit, IQR = interquartile range.
*Some baseline conditions listed here are composites of more than 1 header code. Definitions are provided in the Appendix 1, Methods.
†Unless otherwise specified.
‡Outpatients are those who never had an inpatient claim in the COVID-19 hazard period.
§Inpatients were those who had an inpatient claim at any time in the hazard period. Inpatients include patients in the ICU.
¶Patients in the ICU were those who had an ICU claim at any time in the hazard period.

increased in odds with COVID-19 at a Bonferroni-corrected level of statistical significance. The respiratory and circulatory systems were the most broadly affected physiologic systems (Figure 1). Odds ratios for these candidate COVID-19 complications are shown in Figure 2.

For the respiratory system, codes having the strongest association with COVID-19 included pneumonia (e.g., viral pneumonia: OR 177.63, 95% CI 147.19–214.37), acute respiratory distress syndrome (ARDS) (OR 41.60, 95% CI 33.00–52.45), acute lower respiratory infection (OR 28.77, 95% CI 23.75–34.84), respiratory failure (OR 11.36, 95% CI 10.74–12.02) and pneumothorax (OR 3.38, 95% CI 2.68–4.26). For the circulatory system, cardiac arrest (OR 10.43, 95% CI 8.23–13.21) and acute myocarditis (OR 8.17, 95% CI 3.58–18.62) had the strongest association. Among hematologic disorders, disseminated intravascular coagulation (OR 11.83, 95% CI 5.26–26.62) had the strongest association, and among renal disorders, acute kidney failure (OR 3.50, 95% CI 3.34–3.66) was associated with COVID-19. Disturbances of smell and taste (OR 14.62, 95% CI 10.06–21.25) were also strongly associated. Details for all candidate COVID-19 complications identified in the primary analysis, as well as codes that increased in frequency with nominal statistical significance, are provided in Appendix 1, eTable 2. Codes that significantly decreased in frequency with COVID-19 are provided in Appendix 1, eTable 3. Details of the

69 candidate COVID-19 complications examined at the individual code level are provided in Appendix 1, eTable 4.

Secondary analysis of all 64 931 ICD-10-CM diagnoses at the individual code level revealed an additional 23 codes that significantly increased in frequency with COVID-19 (Appendix 1, eTable 5), including codes for viral enteritis (OR 18.00, 95% CI 8.96–36.16), ketoacidosis (from type 2 diabetes: OR 3.12, 95% CI 2.50–3.89) and rhabdomyolysis (OR 1.96, 95% CI 1.65–2.33).

Risk of candidate COVID-19 complications

Among identified respiratory system codes, those with the highest risks included pneumonia (e.g., viral pneumonia: 27.6% and 81.0% among all patients and patients admitted to the ICU, respectively), respiratory failure (22.6% and 75.3%, respectively) and ARDS (4.3% and 26.0%, respectively). Among identified circulatory system codes, the highest risks included atrial fibrillation or flutter (4.6% and 16.1% among all patients and patients in the ICU, respectively), hypotension (3.8% and 17.3%, respectively) and acute myocardial infarction (1.8% and 8.1%, respectively). The risk for acute kidney failure was 11.8% among all patients and 50.7% among patients in the ICU, and the risk for other sepsis was 10.4% and 54.1%, respectively. Other disorders of the brain (including encephalopathy and other conditions; Appendix 1, eTable 4) showed risks of 4.9% and 24.9%, respectively.
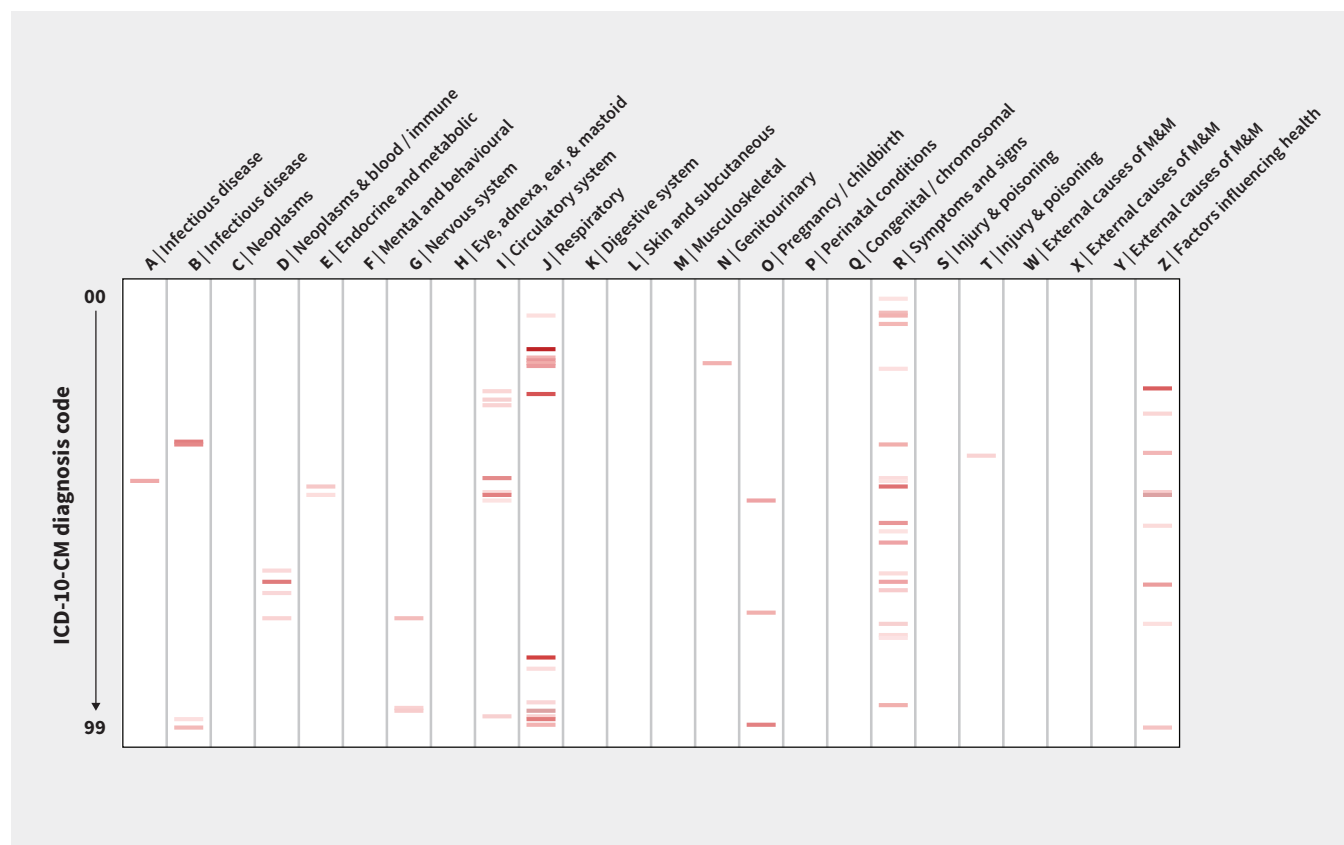
Figure 1: Heat map of candidate coronavirus disease 2019 (COVID-19) complications, by *International Classification of Diseases, 10th Revision, Clinical Modification* (ICD-10-CM) chapter. Odds ratios for association with COVID-19 are shown. Bands show ICD-10-CM diagnosis codes (aggregated at the code header) that were significantly associated with COVID-19 at a Bonferroni-corrected level of significance and for which the odds ratio was > 1. Darker colours show higher odds ratios. Codes highlighted here are described in Figure 2. Note: M&M = morbidity and mortality.

Code type	ICD-10-CM diagnosis code	Prev. (%)	Risk (%)	OR (95% CI)	
Infectious	A41 - Other sepsis	3.4	10.4	4.23 (4.01–4.46)	
	B33 - Other viral diseases–NEC	0.0	0.2	10.88 (6.55–18.05)	
	B34 - Viral infection of unspecified site	1.3	9.5	8.47 (7.86–9.12)	
	B96 - Other bacterial agents–CE	1.9	2.2	1.25 (1.16–1.35)	
	B99 - Other infectious diseases	0.2	0.6	2.94 (2.42–3.57)	
Blood/ immune	D61 - Other aplastic anemias / bone marrow	0.5	0.5	1.50 (1.28–1.75)	
	D65 - Disseminated intravascular coagulation	0.0	0.1	11.83 (5.26–26.62)	
	D69 - Purpura/other hemorrhagic conditions	1.7	2.0	1.52 (1.40–1.65)	
	D72 - Other disorders of white blood cells	2.7	4.0	1.67 (1.57–1.78)	
Endocrine/ metabolism	E43 - Unspecified severe malnutrition	0.8	1.1	2.11 (1.86–2.38)	
	E46 - Unspecified malnutrition	1.1	1.0	1.32 (1.18–1.48)	
Nervous	G72 - Other & unspecified myopathies	0.2	0.4	2.70 (2.17–3.36)	
	G92 - Toxic encephalopathy	0.5	0.8	1.73 (1.51–1.98)	
	G93 - Other disorders of brain	3.6	4.9	1.93 (1.82–2.04)	
Circulatory	I21 - Acute myocardial infarction	1.5	1.8	1.59 (1.45–1.74)	
	I24 - Other acute ischemic heart diseases	0.4	0.6	1.79 (1.53–2.09)	
	I26 - Pulmonary embolism	0.9	1.0	1.66 (1.47–1.87)	
	I40 - Acute myocarditis	0.0	0.1	8.17 (3.58–18.62)	
	I45 - Other conduction disorders	1.1	1.3	1.39 (1.26–1.54)	
	I46 - Cardiac arrest	0.1	1.1	10.43 (8.23–13.21)	
	I48 - Atrial fibrillation & flutter	8.1	4.6	1.19 (1.13–1.26)	
	I95 - Hypotension	2.7	3.8	1.80 (1.68–1.92)	
Respiratory	J06 - Acute upper respiratory infections	5.0	5.9	1.29 (1.23–1.36)	
	J12 - Viral pneumonia–NEC	0.3	27.6	177.63 (147.19–214.37)	
	J15 - Bacterial pneumonia–NEC	0.7	2.0	3.46 (3.09–3.87)	
	J16 - Pneumonia due to other organisms	0.1	0.4	6.26 (4.65–8.43)	
	J17 - Pneumonia in diseases–CE	0.0	0.1	3.83 (2.05–7.16)	
	J18 - Pneumonia–unspecified organism	5.4	20.4	5.68 (5.44–5.94)	
	J22 - Unspecified acute lower resp. infection	0.2	4.4	28.77 (23.75–34.84)	
	J80 - Acute respiratory distress syndrome	0.1	4.3	41.60 (33.00–52.45)	
	J84 - Other interstitial pulmonary diseases	0.8	0.8	1.36 (1.20–1.55)	
	J90 - Pleural effusion–NEC	2.4	2.8	1.56 (1.45–1.68)	
	J93 - Pneumothorax & air leak	0.1	0.4	3.38 (2.68–4.26)	
	J95 - Procedural complications of resp. system	0.2	0.3	1.72 (1.39–2.12)	
	J96 - Respiratory failure–NEC	4.6	22.6	11.36 (10.74–12.02)	
	J98 - Other respiratory disorders	3.8	10.6	3.43 (3.27–3.60)	
Genitourinary	N17 - Acute kidney failure	5.4	11.8	3.50 (3.34–3.66)	
Pregnancy	O48 - Late pregnancy	0.0	0.1	4.75 (2.25–10.01)	
	O70 - Perineal laceration during delivery	0.0	0.1	3.71 (2.08–6.65)	
	O98 - Maternal infectious diseases	0.1	0.4	9.84 (6.86–14.13)	
Symptoms/ signs	R00 - Abnormalities of heart beat	5.1	5.4	1.21 (1.15–1.27)	
	R05 - Cough	10.4	22.0	2.79 (2.70–2.88)	
	R06 - Abnormalities of breathing	11.9	26.2	3.42 (3.31–3.53)	
	R09 - Other s/s circulatory/respiratory system	5.8	14.0	3.10 (2.97–3.23)	
	R19 - Other s/s digestive system & abdomen	3.9	4.4	1.26 (1.19–1.33)	
	R34 - Anuria & oliguria	0.1	0.3	3.81 (2.77–5.24)	
	R40 - Somnolence–stupor & coma	1.6	1.9	1.37 (1.26–1.48)	
	R41 - Other s/s involving cognitive functions	8.4	7.3	1.21 (1.16–1.26)	
	R43 - Disturbances of smell & taste	0.0	0.6	14.62 (10.06–21.25)	
	R50 - Fever of other & unknown origin	4.8	22.2	6.52 (6.24–6.82)	
	R53 - Malaise & fatigue	13.1	12.7	1.21 (1.17–1.25)	
	R57 - Shock–NEC	0.3	1.4	4.76 (4.09–5.54)	
	R62 - Lack of expected normal development	1.2	1.1	1.44 (1.29–1.60)	
	R65 - S/s of systemic inflammation/infection	1.2	4.9	4.83 (4.45–5.23)	
	R68 - Other general s/s	3.8	6.0	2.00 (1.89–2.11)	
	R74 - Abnormal serum enzyme levels	1.5	2.3	1.79 (1.65–1.94)	
	R78 - Findings of drugs & other substances	1.0	1.2	1.41 (1.27–1.56)	
	R79 - Other abnormal findings of blood	4.0	4.2	1.15 (1.09–1.21)	
	R91 - Abnormal findings on imaging of lung	4.7	12.8	3.70 (3.53–3.87)	
External	T38 - Adverse effects hormones–NEC	0.2	0.3	1.66 (1.35–2.04)	
Health factors	Z20 - Exposure to communicable diseases	1.0	17.3	22.10 (20.28–24.08)	
	Z29 - Other prophylactic measures	0.3	0.4	1.59 (1.31–1.95)	
	Z37 - Outcome of delivery	0.1	0.3	3.19 (2.45–4.16)	
	Z45 - Management of implanted device	1.7	2.8	1.89 (1.75–2.04)	
	Z46 - Fitting/adjustment of other devices	1.3	3.7	3.38 (3.12–3.66)	
	Z51 - Encounter for other medical care	3.2	3.7	1.43 (1.35–1.52)	
	Z66 - Do not resuscitate	0.7	2.8	5.51 (4.92–6.17)	
	Z74 - Problems of provider dependency	7.8	6.7	1.28 (1.22–1.34)	
	Z99 - Dependence on devices–NEC	6.0	7.3	2.38 (2.26–2.50)	

1.0   3.0   30.0

Figure 2: Diagnosis codes identified as candidate complications related to coronavirus disease 2019 (COVID-19) for all patients. Baseline pr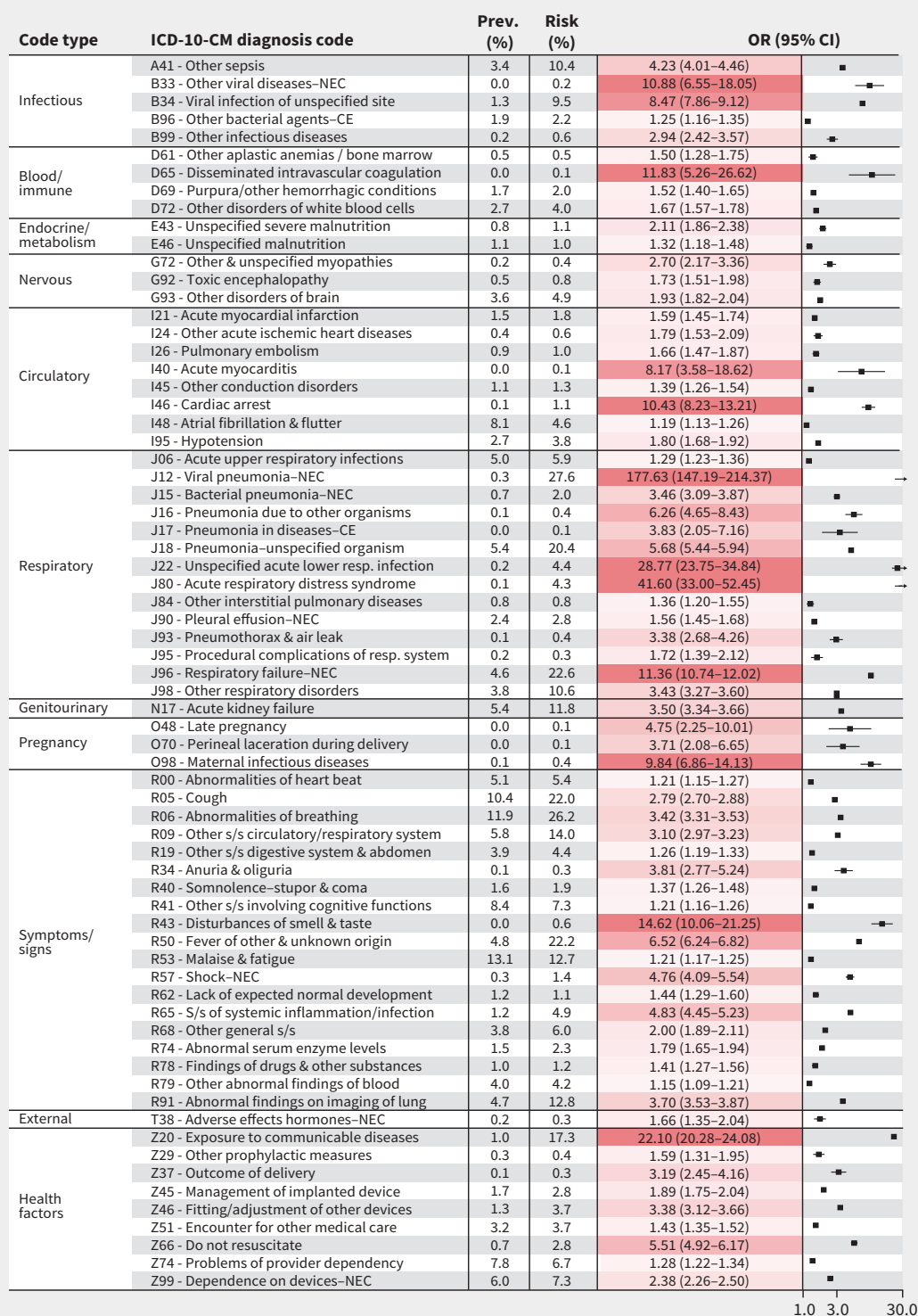evalences ("Prev."), overall risks ("Risk"), and odds ratios (ORs) with 95% confidence intervals (CIs) of association with COVID-19 are shown for *International Classification of Diseases, 10th Revision, Clinical Modification* (ICD-10-CM) diagnosis codes, aggregated at the code header level. Baseline prevalence is the proportion of patients with the condition in the baseline period, among all patients. Overall risk is the proportion of patients with the condition after acquiring COVID-19, among patients who did not previously have the condition. Only significantly associated codes with OR > 1 are listed. The significance threshold for this analysis was 0.05/(total number of codes analyzed) = 0.05/1724 = 2.90E-05. Complete data are provided in Appendix 1, eTable 2 (available at www.cmaj.ca/lookup/doi/10.1503/cmaj.201686/tab-related-content). Calculations are described in Appendix 1, eFigure 3. The X-axis shows ORs represented on a log scale. Darker shading shows higher values. Note: CE = as the cause of diseases classified elsewhere, NEC = not elsewhere classified, s/s = symptoms/signs.
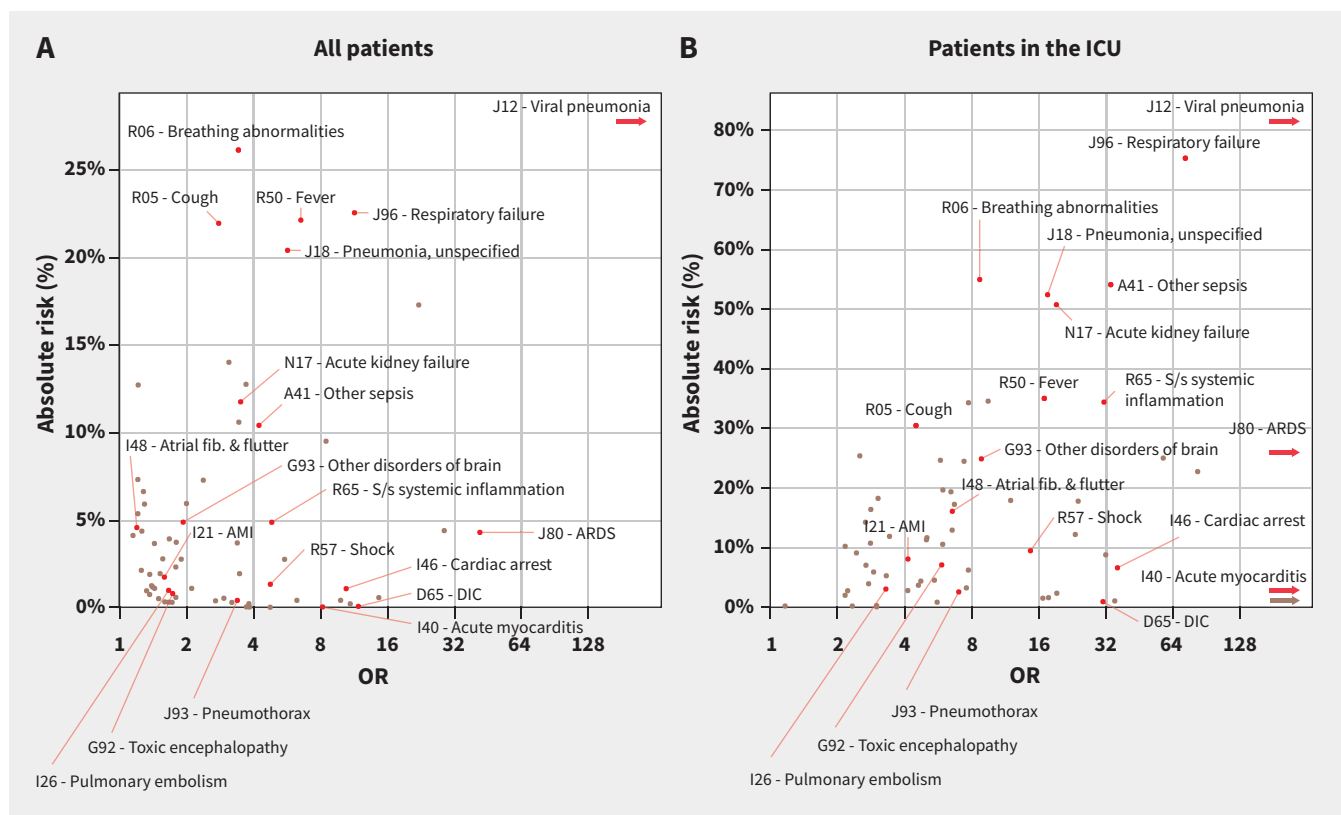
Figure 3: Risks and odds ratios (ORs) of identified complications of coronavirus disease 2019 (COVID-19). Candidate COVID-19 complications identified in the primary analysis are plotted here. The X-axis shows ORs on a log scale, representing the strength of association between COVID-19 and the indicated complication. The Y-axis shows overall risk for the complication (i.e., how common the condition is after acquiring COVID-19, among patients who previously did not have the condition). Odds ratios and risk estimates were calculated in the respective populations (either [A] all patients or [B] patients in the intensive care unit [ICU]). Arrowheads show *International Classification of Diseases, 10th Revision, Clinical Modification* (ICD-10-CM) diagnosis codes with ORs that are higher than the scale limit. Because of space constraints, only a selection of complications representing different ranges of ORs and risks are highlighted; full data are provided in Figure 2. Note: AMI = acute myocardial infarction, ARDS = acute respiratory distress syndrome, DIC = disseminated intravascular coagulation, fib = fibrillation, NEC = not elsewhere classified, s/s = signs and symptoms.

Figure 3 shows risk estimates plotted against OR estimates, where it is evident that the most high-risk, highly associated disorders included viral pneumonia, respiratory failure, sepsis or systemic inflammation, acute kidney failure and ARDS. Risks for these conditions by age and hospital admission status are shown in Figure 4 and Appendix 1, eTable 6. Most other complications had relatively low overall risk estimates even if they were highly associated with COVID-19, such as pneumothorax (0.4% and 2.6% absolute risk among all patients and patients in the ICU, respectively), disseminated intravascular coagulation (0.1% and 0.9%, respectively) and acute myocarditis (0.1% and 0.5%, respectively). Appendix 1 contains risk information for all candidate COVID-19 complications identified in the primary analysis, calculated among all patients (eTable 2), age-stratified patients (eTable 7), sex-stratified patients (eTable 8), and by hospital admission status (eTable 9).

Sensitivity analysis

Our sensitivity analysis to evaluate the effect of using a shorter, 30-day baseline period resulted in many chronic conditions apparently increasing in frequency after COVID-19, confirming that the longer, 90-day baseline period of the primary analysis

was necessary to detect chronic conditions that were already present before COVD-19 (Appendix 1, eTable 10). It also found that codes that may represent acute conditions, such as "other venous embolism and thrombosis," became moderately increased in frequency with COVID-19 (OR 1.74, 95% CI 1.58–1.93), suggesting that the primary analysis could have missed some acute conditions with comparatively modest associations with COVID-19.

Interpretation

In this study of more than 70 000 individuals who received a diagnosis of COVID-19, we found that the disease was associated with a broad range of complications. The more common complications that we identified — including viral pneumonia, respiratory failure, acute kidney failure and sepsis — were expected, as they have been well described in the literature.[17–20] We also identified less common complications, previously described in case series or small studies, such as disseminated intravascular coagulation,[21] pneumothorax,[22] myocarditis[23] and rhabdomyolysis.[24] This study provides estimates of absolute risk and relative odds for all identified diagnoses related to COVID-19, which are needed to

Figure 4: Risk estimates of the most common complications of coronavirus disease 2019 (COVID-19), by age and hospital admission status. Some of these conditions are composites of more than 1 *International Classification of Diseases, 10th Revision, Clinical Modification* (ICD-10-CM) header code. Pneumonia: defined as J12 (viral pneumonia, not elsewhere classified) or J18 (pneumonia, unspecified organism). Respiratory failure or acute respiratory distress syndrome (ARDS): defined as J96 (respiratory failure, not elsewhere classified) or J80 (acute respiratory distress syndrome). Acute kidney failure: defined as N17 (acute kidney failure). Systemic inflammatory response syndrome (SIRS), sepsis or septic shock: defined as A41 (other sepsis) or R65 (symptoms and signs specifically associated with systemic inflammation and infection). Complete data for these findings are provided in Appendix 1, eTable 6 (available at www.cmaj.ca/lookup/doi/10.1503/cmaj.201686/tab-related-content).

help providers, patients and policy-makers understand the likelihood of complications. For example, acute myocarditis was found to have an OR of 8.17 but an overall risk of 0.1%, illustrating how a very strong association of a condition with COVID-19 does not necessarily translate into a high overall risk.

Since the first case of COVID-19 described in 2019, a sharp rise in the number of studies related to COVID-19 has made it challenging for clinicians to keep up with the literature and distinguish spurious findings from causal effects. Reporting bias and publication bias can result in an over-representation of some findings, and published findings beget confirmation of the same. A key strength of our analytic approach is that it considered all possible ICD-10-CM diagnoses codes and simply quantified diagnoses that occurred after the onset of COVID-19.

Although COVID-19 has been widely reported to increase the risk of stroke,[25] this was not seen in our study. For I63 ("cerebral infarction") in our overall population, we observed an OR of 0.58 and an overall risk of 1.5% (Appendix 1, eTable 3), suggesting that while patients with COVID-19 do experience stroke at significant frequencies, a causal association with COVID-19 was not supported in this population. Multisystem inflammatory syndrome in children could not be directly assessed because it has no specific ICD-10-CM code, although no association was noted for the similar condition Kawasaki disease, as evaluated under the code M30.3.

We also observed many diagnosis codes that appeared to decrease in odds after the onset of COVID-19. We propose 2 possible explanations. During the initial stages of COVID-19 treatment, many chronic conditions and less severe conditions may not have been considered priorities for care and were therefore less likely to be captured in a claim. In addition, for acute conditions, the baseline period (90 d) could be accruing more events

than the hazard period (38 d) because it is longer in duration. This could explain why, for example, cerebral infarction appeared to decrease in odds after the onset of COVID-19 in our primary analysis, while it slightly increased in odds in our sensitivity analysis using a shorter baseline period.

In interpreting our risk estimates, it is important to note that mild or less clinically overt conditions are not as commonly coded in claims data. This likely explains the fact that, although we observed strong associations with COVID-19 for complications such as cough and disturbances of smell or taste, their overall risk estimates (22.0% and 0.6%, respectively, in the overall population) were substantially lower than has been reported (79% and 65%, respectively).[9,26] Thus, the findings for codes related to relatively mild conditions may be more valuable in representing strengths of association than absolute risk estimates. In contrast, risk estimates for severe, overt disorders are more likely to reflect actual risk, as they are more likely to be treated or brought to the attention of a care provider, and thus are more consistently captured in a medical claim.[27] For example, our risk estimates among inpatients for respiratory failure (40.0%) and acute kidney failure (21.2%) are similar to previously reported estimates (54% and 15%, respectively).[20] An additional caveat to the interpretation of our risk estimates is that they estimate only the risk of newly diagnosed disease and do not estimate the risk of events where a pre-existing condition may have been exacerbated by COVID-19. Moreover, our estimates reflect only the risk among patients who seek medical care for COVID-19.

Strengths of this study include its large sample size and data-driven approach. An additional strength is the exposure-crossover design, which controls for confounding by matching observations pre- and post-COVID-19 diagnosis.[12] By using inpatient and outpatient claim records, we were able to leverage a relatively complete medical history for each patient to account for conditions that were already present before COVID-19 diagnosis.

Limitations

A limitation of the claims data is that we identified COVID-19 cases using diagnosis codes. Although the codes we used to identify patients are intended for confirmed cases of COVID-19,[10,11] it is possible that some patients were misclassified. Furthermore, the recency of the COVID-19 pandemic necessitated the use of open medical claims that may not be fully adjudicated at the time of capture and may be subject to change if disputed by a payer. As a result, some misclassification may be present in the data, which could reduce the precision of estimates. In addition, the ICD-10-CM codes are not necessarily validated disease definitions and are assigned by medical professionals working in different places of care throughout the US; thus, these codes do not capture their intended disease concepts with complete consistency or fidelity. However, given the size of the data set, we believe that these issues did not have a substantial impact on our results. Moreover, because older data in the baseline period are likely more complete than those of the hazard period, there may have been an underestimation of ORs, resulting in a gener-

ally conservative bias in the identification of COVID-19 complications. Given that the results of this study are broadly in line with known complications of COVID-19, we believe that this did not meaningfully affect our results.

The care settings reflected in the database were not exclusively academic medical centres, where more severe cases of COVID-19 may be treated; as such, our results may have missed some of the more severe cases. Another limitation on generalizability is that patients were required to have at least 1 medical claim and thus may be more ill in the baseline period than the general population. As a result, risk estimates reported here may be greater than the risks in the general population, as the studied patients were more likely to have comorbidities. Finally, some of our findings may not be direct consequences of infection with SARS-CoV-2 but instead may be iatrogenic effects of treatment.

Conclusion

Overall, the most common complications associated with COVID-19 among patients seeking medical care include pneumonia, respiratory failure, kidney failure, and sepsis or systemic inflammation. After analyzing all possible diagnosis codes, we confirm that COVID-19 is also associated with a diverse array of additional cardiac, thrombotic and other conditions, although the overall risks for most of these complications are comparatively low. Understanding the full range of associated conditions can aid in prognosis, guide treatment decisions and better inform patients as to their actual risks for the variety of COVID-19 complications reported in the literature and media.

References

1. WHO coronavirus disease (COVID-19) dashboard. Geneva: World Health Organization; 2020. Available: https://covid19.who.int/ (accessed 2020 Nov. 20).
2. Huang C, Wang Y, Xingwang L, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020;395:497-506.
3. Interim clinical guidance for management of patients with confirmed coronavirus disease (COVID-19). Atlanta: Centers for Disease Control and Prevention; 2020. Available: www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html (accessed 2020 July 1).
4. Varga Z, Flammer AJ, Steiger P, et al. Endothelial cell infection and endotheliitis in COVID-19. *Lancet* 2020;395:1417-8.
5. Klok FA, Kruip MJHA, van der Meer NJM, et al. Incidence of thrombotic complications in critically ill ICU patients with COVID-19. *Thromb Res* 2020;191:145-7.
6. Long B, Brady WJ, Koyfman A, et al. Cardiovascular complications in COVID-19. *Am J Emerg Med* 2020;38:1504-07.
7. Chiotos K, Bassiri H, Behrens EM, et al. Multisystem inflammatory syndrome in children during the COVID-19 pandemic: a case series. *J Pediatric Infect Dis Soc* 2020;9:393-8.
8. Mao L, Jin H, Wang M, et al. Neurologic manifestations of hospitalized patients with coronavirus disease 2019 in Wuhan, China. *JAMA Neurol* 2020;77:683-90.
9. Spinato G, Fabbris C, Polesel J. Alterations in smell or taste in mildly symptomatic outpatients with SARS-CoV-2 infection. *JAMA* 2020;323: 2089-2090.
10. ICD-10-CM official coding guidelines — supplement: coding encounters related to COVID-19 coronavirus outbreak: effective February 20, 2020. Atlanta: Centers for Disease Control and Prevention; 2020. Available: www.cdc.gov/nchs/data/icd/ICD-10-CM-Official-Coding-Gudance-Interim-Advice-coronavirus-feb-20-2020.pdf (accessed 2020 July 1).
11. ICD-10-CM official coding and reporting guidelines April 1, 2020 through September 30, 2020. Atlanta: Centers for Disease Control and Prevention; 2020. Available: www.cdc.gov/nchs/data/icd/COVID-19-guidelines-final.pdf (accessed 2020 July 1).
12. Redelmeier DA. The exposure-crossover design is a new method for studying sustained changes in recurrent events. *J Clin Epidemiol* 2013;66:955-63.

13. Fay MP. Exact McNemar's Test and matching confidence intervals. 2016. Available: https://cran.r-project.org/web/packages/exact2x2/vignettes/exactMcNemar.pdf (accessed 2016 July 1).
14. Hirsch JA, Nicola G, McGinty G, et al. ICD-10: history and context. *AJNR Am J Neuroradiol* 2016;37:596-9.
15. ICD-10-CM official guidelines for coding and reporting. Baltimore: Centers for Medicare and Medicaid Services; 2020. Available: www.cms.gov/Medicare/Coding/ICD10/Downloads/2020-Coding-Guidelines.pdf (accessed 2020 July 1).
16. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ* 1995;310:170.
17. Li X, Ma X. Acute respiratory failure in COVID-19: Is it "typical" ARDS? *Crit Care* 2020;24:198.
18. Richardson S, Hirsch JS, Narasimhan M. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. *JAMA* 2020;323:2052-9.
19. Goyal P, Choi JJ, Pinheiro LC, et al. Clinical characteristics of COVID-19 in New York City. *N Engl J Med* 2020;382:2372-4.
20. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020;395:1054-62.
21. Tang N, Li D, Wang X, et al. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. *J Thromb Haemost* 2020;18:844-7.
22. Rohailla S, Ahmed N, Gough K. SARS-CoV-2 infection associated with spontaneous pneumothorax. *CMAJ* 2020;192:E510.
23. Ruan Q, Yang K, Wang W, et al. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive Care Med* 2020;46:846-8.
24. Jin M, Tong Q. Rhabdomyolysis as potential late complication associated with COVID-19. *Emerg Infect Dis* 2020;26:1618-20.
25. Fridman S, Bullrich MB, Jimenez-Ruiz A, et al. Stroke risk, phenotypes, and death in COVID-19: Systematic review and newly reported cases. *Neurology* 2020;Sep 15 [Epub ahead of print]. doi: 10.1212/WNL.0000000000010851.
26. Chen HJ, Qiu J, Wu B, et al. Early chest CT features of patients with 2019 novel coronavirus (COVID-19) pneumonia: relationship to diagnosis and prognosis. *Eur Radiol* 2020;30:6178-85.
27. *Guidance for industry and FDA staff: best practices for conducting and reporting pharmacoepidemiologic safety studies using electronic healthcare data*. Silver Spring (MD): US Food and Drug Administration, Center for Drug Evaluation and Research; 2013.