What factors increase the risk of complications in SARS-Cov-2 infected patients? A cohort study in a nationwide Israeli health organization

Chen Yanover, Barak Mizrahi, Nir Kalkstein, Karni Marcus, Pinchas Akiva, Yael Barer, Varda Shalev, Gabriel Chodick

Submitted to: JMIR Public Health and Surveillance
on: May 31, 2020

Table of Contents

What factors increase the risk of complications in SARS-Cov-2 infected patients? A cohort study in a nationwide Israeli health organization

Chen YanoverPhD, ; Barak MizrahiMSc, ; Nir KalksteinBSc, ; Karni MarcusMD, MBA, ; Pinchas AkivaPhD, ; Yael BarerMSc, ; Varda ShalevMD, ; Gabriel ChodickPhD,

Corresponding Author:
Chen YanoverPhD,
Phone: +972+972545493154
Email: chen@kinstitute.org.il

Abstract

Background: Reliably identifying patients at increased risk for coronavirus disease 2019 (COVID-19) complications could guide clinical decisions, public health policies, and preparedness efforts. Multiple studies have attempted to characterize at-risk patients, using various data sources and methodologies. Most of these studies, however, explored condition-specific patient cohorts (e.g., hospitalized patients) or had limited access to patients' medical history; thus, investigating related questions and, potentially, obtaining biased results.

Objective: To identify factors associated with COVID-19 complications from the complete medical records of a nationally representative SARS-CoV-2 infected cohort.

Methods: We studied the cohort of all SARS-CoV-2 positive individuals, confirmed by polymerase chain reaction testing of either nasopharyngeal or saliva samples, in a nationwide health organization (covering 2.3 million individuals) and identified those who suffered from serious complications (that is, experienced moderate or severe symptoms of COVID-19, admitted to intensive care unit or died). We then compared the prevalence of pre-existing conditions, extracted from electronic health records, between complicated and non-complicated COVID-19 patient cohorts to identify the conditions that significantly increase the risk of disease complications, in various age and sex strata.

Results: Of the 4,353 SARS-CoV-2 positive individuals, 173 (4%) patients suffered from COVID-19 complications (all above the age of 18 years). Our analysis suggests that cardiovascular and kidney diseases, obesity, and hypertension are significant risk factors for COVID-19 complications, as previously reported. Interestingly, it also indicates that depression (e.g., odds ratio, OR, for males 65 years or older: 2.94, 95% confidence intervals [1.55, 5.58]; P = .014) as well as cognitive and neurological disorder (e.g., OR for individuals ? 65 years old: 2.65 [1.69, 4.17]; P < .001) are significant risk factors; and that smoking and background of respiratory diseases do not significantly increase the risk of complications.

Conclusions: Adjusting existing risk definitions following these observations may improve their accuracy and impact the global pandemic containment and recovery efforts.

(JMIR Preprints 31/05/2020:20872)
DOI: https://doi.org/10.2196/preprints.20872

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
Please make my preprint PDF available to anyone at any time (recommended).
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
✓ **No, I do not wish to publish my submitted manuscript as a preprint.**
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

What factors increase the risk of complications in SARS-Cov-2 infected patients? A cohort study in a nationwide Israeli health organization

Authors. Chen Yanover, PhD[1,*,†], Barak Mizrahi, MSc[1,*], Nir Kalkstein, BSc[1], Karni Marcus, MD, MBA[1], Pinchas Akiva, PhD[1], Yael Barer, MSc[2], Varda Shalev, MD[2], Gabriel Chodick, PhD[2,3]
Affiliation. [1]KI Research Institute, Kfar Malal, Israel; [2]Maccabi Institute for Research and Innovation, Tel Aviv, Israel; [3]Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel
* Equally contributing authors
[†] Corresponding author: Chen Yanover, KI Research Institute, 11 Hazayit st, Kfar Malal, Israel. Email: chen@kinstitute.org.il
Word count**:** Abstract: 308 words; main text (excluding abstract, tables, figures, acknowledgments, references, and online-only material): 2163 words

Abstract

Background: Reliably identifying patients at increased risk for coronavirus disease 2019 (COVID-19) complications could guide clinical decisions, public health policies, and preparedness efforts. Multiple studies have attempted to characterize at-risk patients, using various data sources and methodologies. Most of these studies, however, explored condition-specific patient cohorts (e.g., hospitalized patients) or had limited access to patients' medical history; thus, investigating related questions and, potentially, obtaining biased results.

Objective: To identify factors associated with COVID-19 complications from the *complete* medical records of a *nationally representative* SARS-CoV-2 infected cohort

Methods: We studied a cohort of *all* SARS-CoV-2 positive individuals, confirmed by polymerase chain reaction testing of either nasopharyngeal or saliva samples, in a nationwide health organization (covering 2.3 million individuals) and identified those who suffered from serious complications (that is, experienced moderate or severe symptoms of COVID-19, admitted to intensive care unit or died). We then compared the prevalence of pre-existing conditions, extracted from electronic health records, between complicated and non-complicated COVID-19 patient cohorts to identify the conditions that significantly increase the risk of disease complications, in various age and sex strata.

Results: Of the 4,353 SARS-CoV-2 positive individuals, 173 (4%) patients suffered from COVID-19 complications (all above the age of 18 years). Our analysis suggests that cardiovascular and kidney diseases, obesity, and hypertension are significant risk factors for COVID-19 complications. It also indicates that depression (e.g., odds ratio, OR, for males 65 years or older: 2.94, 95% confidence intervals [1.55, 5.58]; $P = .014$) as well as cognitive and neurological disorder (e.g., OR for individuals ≥ 65 years old: 2.65 [1.69, 4.17]; $P < .001$) are significant risk factors; and that smoking and background of respiratory diseases do not significantly increase the risk of complications.

Conclusions: Our analysis agrees with previous studies on multiple risk factors, including hypertension and obesity. It also finds depression, cognitive and neurological disorders, but not smoking and respiratory diseases, to be significantly associated with COVID-19 complications. Adjusting existing risk definitions following these observations may improve their accuracy and impact the global pandemic containment and recovery efforts.

Introduction

As of April 30th, 2020, more than three million people worldwide contracted severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and close to 250,000 people died of corona virus disease 2019 (COVID-19) complications. In Israel, on that date, 16,004 individuals have been infected by the virus and 223 died from the disease. This pandemic poses grave challenges to patients, healthcare providers, and policy makers. Many of these challenges may be better addressed with timely stratification of patients to risk groups, based on their past and current medical characteristics. For example, reliably identifying patients at increased (or decreased) risk could guide clinical decisions (e.g., hospitalization vs home care), public health policies (e.g., risk-based quarantine), and preparedness efforts (e.g., expected medical equipment required).

Various algorithms for identifying patients at risk for COVID-19 (severe) complications have been proposed. The Centers for Disease Control and Prevention (CDC) identified individuals 65 years and older, living at nursing home or long-term care facility, or suffering from underlying medical conditions, particularly if not well controlled, as being at high risk for severe illness from COVID-19 [1]. Similarly, the European Centre for Disease Prevention and Control (ECDC) lists age above 70 years and some underlying conditions as risk factors for critical illness [2]. The United Kingdom National Health Service (NHS) included solid organ transplant recipients, patients with specific cancers or severe respiratory conditions, pregnant women with significant heart disease, and those with increased risk of infection (e.g., due to immunosuppressive therapies) in the highest clinical COVID-19 risk group [3]. In April 2020 approximately 1.3 million people in this group were asked to “shield” by staying at home for a period of at least 12 weeks. In addition, patients over 70 years and those suffering from some underlying health conditions (e.g., chronic respiratory diseases, BMI ≥ 40, and pregnant women) were considered in a wider vulnerable group (also referred to as the “flu group”). Finally, a more quantitative risk model (derived from [4]) has been adopted by the Israeli Ministry of Health (MoH), assigning a point for each underlying condition from a predefined list, then considering age group and point count to identify high risk patients.

Initially, these algorithms have been derived from a quickly growing number of epidemiological characterization studies (e.g., [5,6]), which report the prevalence of various conditions in a population of interest, typically severe hospitalized COVID-19 patients. These studies provide timely and important information; however, identifying risk factors calls for a comparative analysis, contrasting the prevalence of conditions in case and control populations. To date, only a handful of studies implemented such an approach, using, for example, the general population [7] or confirmed (and symptomatic) COVID-19 patient cohort [8]. Similar to these efforts, we analyze here the medical records of all SARS-CoV-2 positive patients in a nationwide health organization (covering 2.3 million individuals); compare the prevalence of existing conditions in complicated and non-complicated cohorts; and identify those conditions associated with COVID-19 complications in various age and sex strata. Our analysis highlights stratum-specific risk factors and may allow better identification of patients at risk, in different subpopulations.

Methods

Data source

Maccabi Health Services (MHS) is a nationwide health plan (payer-provider), representing a quarter of the Israeli population. The MHS database contains longitudinal data on a stable population of over 2.3 million people since 1993 (with annual attrition rate lower than 1%). Data are automatically collected and include comprehensive laboratory data from a single central lab, full pharmacy prescription and purchase data, and extensive demographic information on each patient.

Study design and setting

SARS-CoV-2 polymerase chain reaction testing in Israel uses both nasopharyngeal and saliva samples. Individuals with positive testing result (until April 22, 2020) are included in the *SARS-CoV-2 positive cohort*. Positive patients whose disease status, as updated by Israeli hospitals, deteriorated to moderate or severe (at any point in time), admitted to intensive care unit, or died, constitute the *complicated COVID-19 cohort*. Initially, the definition of disease status varied, to some extent, between hospitals, but was largely based on the severity of lower respiratory tract symptoms, including pneumonia, respiratory distress, and artificial respiration, as well as shock and system failure. The remaining SARS-CoV-2 positive patients (including asymptomatic, mild COVID-19 patients or those with unknown status) constitute the *non-complicated COVID-19 cohort*. Follow-up period ended on April 30th, 2020 (or upon patient's death).

Patient characteristics

Beside age and sex, we considered a set of existing conditions, comprising those included in the CDC, NHS, and Israeli MoH at-risk definitions, as well as a set of conditions showing significant association with flu and flu-like complications.

To identify each individual's existing conditions, we used, where available, registries created and maintained by MHS. These registries are based on validated inclusion and exclusion criteria (considering coded diagnoses, treatments, labs, and imaging, as applicable). The registries are continuously and retrospectively (since 1998) updated based on each patient's central medical record. Patients may be excluded from a registry when deemed misclassified by their primary physician. Linkage across registries and with other sources of information is performed via a unique national identification number. MHS registries used are: Cardiovascular diseases (specifically, ischemic heart disease, congestive heart failure, peripheral vascular disease, cerebrovascular disease, and other cardiovascular diseases) [9], diabetes [10,11], hypertension [12], osteoporosis [13], chronic kidney disease [14], cognitive disorders, mental illness [15], cancer, immunosuppression (including advanced kidney disease, immunosuppressive treatment, asplenia, and organ transplant), weight disorders (obesity, overweight, and underweight), smoking, hospitalization (in the last three years), nursing home, and home care (home visits, home respiratory care, respiratory and feeding equipment). For other conditions, we relied on previously grouped lists of diagnosis codes (Read codes or International Classification of Diseases, ICD, codes, ninth revision) [16–18]: Deficiency anemia, fluid and electrolyte disorders, respiratory diseases (specifically, COPD: chronic obstructive pulmonary disease, chronic pulmonary disease, pleural effusion, aspiration pneumonia, and bronchiectasis), neurological disorders, end stage renal disease, rheumatoid arthritis, paralysis, hip fracture, lymphoma, and alcohol consumption.

Statistical analysis

We extracted the prevalence of the studied conditions (excluding ones with less than 20 occurrences) in the non-complicated and complicated COVID-19 cohorts and measured the association between each condition and disease complication by computing the corresponding odds ratio and its estimated statistical significance (using Fisher's exact test). We conducted the analysis separately in three age groups: 18-50 years, 50-65 years, and 65 years and older; as well as four (age, sex) strata: male or female, younger or older than 65 years. Using different age groups (as sensitivity analysis), obtained similar results. Finally, to account for multiple testing, we controlled for the false discovery rate using Benjamini and Hochberg's method [19]. All analyses were performed using version 4.0.0 of the R programming language (R Project for Statistical Computing; R Foundation). We used the STROBE cohort checklist when writing our report [20].

Results

Maccabi Health Services (MHS) SARS-CoV-2 positive cohort included 4,353 individuals of whom 173 deteriorated to moderate (N=87, 50%) or severe condition (N=45, 26%), admitted to intensive care unit (N=66, 38%, partly overlapping with other conditions), or died (N=21, 12%), and make up the complicated COVID-19 cohort. Overall, patients in the complicated COVID-19 cohort were older, suffered from more comorbidities, and were more predominantly male (). Moreover, the prevalence of COVID-19 complications increased with age, and more steeply for men than for women (Figure 1.); and the risk of COVID-19 complications in men under 70 years was significantly higher than in women (Error: Reference source not found).

Figure 1. Age and complicated COVID-19. Prevalence of complicated COVID-19 (moderate or severe condition, y-axes) in different age groups (x-axis), shown separately for males (blue) and females (orange).

Comparing the prevalence of existing conditions in three age groups between the complicated and non-complicated COVID-19 cohorts, revealed multiple risk factors, including obesity for patients 18-50 years (OR: 11.09, 95% confidence intervals, CI: [4.15, 32.67]; $P < .001$), chronic kidney disease for patients 50-65 years (4.06 [1.89, 8.38]; $P = .005$); and neurological disorders (2.65 [1.69, 4.17]; $P < .001$) for patients 65 years or older (for a complete list, see Error: Reference source not found and Supplementary File 1.).

Stratifying over age (below and above 65 years) and sex (and Supplementary File 1.), kidney diseases appeared as a risk factor in all strata (e.g., OR 3.45 [1.57, 8.06]; $P =.015$ in women 65 years or more). Additional risk factors included hypertension in males under 65 years (4.56 [2.35, 8.55]; $P < .001$); neurological disorders in females 65 years or older (3.55 [1.68, 7.74]; $P = .008$); cognitive impairment (4.18 [1.81, 9.72]; $P = .009$) and depression (2.94 [1.55, 5.58]; $P =.014$) in males 65 years or older. Respiratory diseases and smoking, while typically more prevalent in complicated COVID-19 patients, were not identified as significant risk factors (e.g., chronic obstructive pulmonary disease in patients 65 years or more: OR 1.36 [0.75, 2.4]; $P = .634$); and see Supplementary File 1.

Table 1. Characteristics of the SARS-CoV-2 positive, complicated, and non-complicated COVID-19 patient cohorts.

	SARS-Cov-2 positive	Complicated COVID-19	Non-complicated COVID-19
Demographic information			
N	4353	173	4180
Age median [inter quantile range, IQR]	35 [22-54]	70 [58-80]	34 [22-52]
<18y	647 (15%)	0 (0%)	647 (15.6%)
18-50y	2354 (54.5%)	21 (12.1%)	2333 (56.3%)
50-60y	609 (14.1%)	29 (16.8%)	580 (14%)
60-70y	376 (8.7%)	35 (20.2%)	341 (8.2%)
70-80y	232 (5.4%)	42 (24.3%)	190 (4.6%)
80y≤	135 (3.1%)	46 (26.6%)	89 (2.1%)
Female	1939 (44.5%)	50 (28.9%)	1889 (45.2%)
Follow-up days [IQR]	30 [24-36]	28 [21-33]	30 [24-36]
Comorbidities			
Chronic respiratory diseases	481 (11%)	39 (22.5%)	442 (10.6%)
Chronic obstructive pulmonary disease	310 (7.1%)	24 (13.9%)	286 (6.8%)
Other chronic pulmonary disease	153 (3.5%)	10 (5.8%)	143 (3.4%)
Pleural effusion	41 (0.9%)	4 (2.3%)	37 (0.9%)
Cardiovascular diseases	310 (7.1%)	57 (32.9%)	253 (6.1%)
Ischemic heart disease	132 (3%)	27 (15.6%)	105 (2.5%)
Congestive heart failure	30 (0.7%)	11 (6.4%)	19 (0.5%)
Cerebrovascular disease	57 (1.3%)	15 (8.7%)	42 (1%)
Peripheral vascular disease	23 (0.5%)	7 (4%)	16 (0.4%)
Other cardiovascular diseases	199 (4.6%)	41 (23.7%)	158 (3.8%)
Hypertension	627 (14.4%)	102 (59%)	525 (12.6%)
Immunosuppression	164 (3.8%)	31 (17.9%)	133 (3.2%)
Cancer	205 (4.7%)	33 (19.1%)	172 (4.1%)
Deficiency anemia	423 (9.7%)	18 (10.4%)	405 (9.7%)
Liver and kidney diseases			
Liver disease	404 (9.3%)	28 (16.2%)	376 (9%)
Chronic kidney disease	384 (8.8%)	86 (49.7%)	298 (7.1%)
End stage renal disease	85 (2%)	26 (15%)	59 (1.4%)
Fluid and electrolyte disorders	394 (9.1%)	37 (21.4%)	357 (8.5%)
Metabolic diseases			
Diabetes	362 (8.3%)	58 (33.5%)	304 (7.3%)
Obesity (BMI≥30)	874 (20.1%)	73 (42.2%)	801 (19.2%)
Neurological and cognitive disorders			
Neurological disorders	294 (6.8%)	57 (32.9%)	237 (5.7%)
Paralysis	53 (1.2%)	12 (6.9%)	41 (1%)
Depression	578 (13.3%)	53 (30.6%)	525 (12.6%)
Cognitive impairment	87 (2%)	28 (16.2%)	59 (1.4%)
Other			
Hospitalization	931 (21.4%)	92 (53.2%)	839 (20.1%)
Smoking	643 (14.8%)	41 (23.7%)	602 (14.4%)
Current smoker	514 (11.8%)	30 (17.3%)	484 (11.6%)
Past smoker	129 (3%)	11 (6.4%)	118 (2.8%)
Nursing home	67 (1.5%)	23 (13.3%)	44 (1.1%)
Home care	44 (1%)	17 (9.8%)	27 (0.6%)

Table 2. Association of male sex and COVID-19 complications across age groups.

Age group	Patient counts[a]	OR[b]	*P-value*[c]
18-50y	(18/1300,3/1033)	4.77 [1.39, 25.32]	.013
50-60y	(25/314,4/266)	5.28 [1.79, 21.15]	.003
60-70y	(29/202,6/139)	3.32 [1.31, 10.03]	.013
70-80y	(27/108,15/82)	1.36 [0.65, 2.95]	.473
80y≤	(24/32,22/57)	1.93 [0.89, 4.26]	.145

[a] Number of males in the complicated/non-complicated COVID-19 cohorts, followed by the corresponding numbers in females

[b] Odds ratios (ORs) and 95% confidence intervals (in brackets). ORs greater than 1 suggest an increased risk for COVID-19 complications in males

[c] P-values adjusted for multiple testing using Benjamini and Hochberg method [19].

Table 3. Most statistically significant conditions associated with increased risk of COVID-19 complications in age stratified patient groups.

Condition	Age group	Patient counts[a]	OR[b]	*P-value*[b]
Obesity	18-50y	(14/356,7/1977)	11.09 [4.15, 32.67]	<.001
Depression	18-50y	(7/229,14/2104)	4.59 [1.55, 12.3]	.032
Hypertension	18-50y	(4/72,17/2261)	7.37 [1.76, 23.41]	.035
Liver disease	18-50y	(5/125,16/2208)	5.51 [1.55, 16.07]	.037
Chronic kidney disease	50-65y	(14/87,27/683)	4.06 [1.89, 8.38]	.005
End stage renal disease	50-65y	(5/8,36/762)	13.11 [3.21, 48.19]	.006
Neurological disorders	65y≤	(54/113,57/317)	2.65 [1.69, 4.17]	<.001
Chronic kidney disease	65y≤	(70/174,41/256)	2.51 [1.6, 3.97]	.001
Other cardiovascular diseases	65y≤	(36/70,75/360)	2.46 [1.49, 4.05]	.006
Cognitive impairment	65y≤	(28/52,83/378)	2.45 [1.4, 4.22]	.015
Home care	65y≤	(16/22,95/408)	3.12 [1.47, 6.48]	.022
Hypertension	65y≤	(82/249,29/181)	2.05 [1.27, 3.4]	.027
Cardiovascular diseases	65y≤	(50/129,61/301)	1.91 [1.22, 2.99]	.032
Nursing home	65y≤	(20/35,91/395)	2.48 [1.29, 4.65]	.036

[a] Number of patients with the condition (cases) in the complicated/non-complicated COVID-19 cohorts, followed by the corresponding numbers in patients without the studied condition (controls).

[b] Odds ratios (ORs) and 95% confidence intervals (in brackets). ORs greater than 1 suggest an increased risk for COVID-19 complications in patients with the noted condition. In each stratum, rows are sorted ascendingly by P-value.

Table 4. Most statistically significant conditions associated with increased risk of COVID-19 complications in age and sex stratified patient groups.

Condition	Age, sex group	Patient counts[a]	OR[b]	*P-value*[b]
End stage renal disease	age<65; Female	(2/5,7/1370)	75.7 [6.23, 570.01]	.014
Immunosuppression	age<65; Female	(3/46,6/1329)	14.35 [2.25, 69.89]	.032
Chronic kidney disease	age<65; Female	(3/58,6/1317)	11.3 [1.78, 54.41]	.041
Chronic kidney disease	age<65; Male	(13/66,40/1662)	8.16 [3.82, 16.5]	<.001
Hypertension	age<65; Male	(17/162,36/1566)	4.56 [2.35, 8.55]	<.001
Obesity	age<65; Male	(25/359,28/1369)	3.4 [1.88, 6.14]	.001
Hospitalization	age<65; Male	(21/285,32/1443)	3.32 [1.79, 6.04]	.004
End stage renal disease	age<65; Male	(3/7,50/1721)	14.67 [2.38, 66.53]	.029
Diabetes	age<65; Male	(9/105,44/1623)	3.16 [1.32, 6.79]	.040
Neurological disorders	65≤age; Female	(26/66,15/136)	3.55 [1.68, 7.74]	.008
Chronic kidney disease	65≤age; Female	(30/89,11/113)	3.45 [1.57, 8.06]	.015
Home care	65≤age; Female	(10/16,31/186)	3.72 [1.38, 9.69]	.036
Other cardiovascular diseases	65≤age; Female	(15/33,26/169)	2.94 [1.3, 6.51]	.038
Cardiovascular diseases	65≤age; Female	(19/48,22/154)	2.76 [1.29, 5.85]	.045
Cognitive impairment	65≤age; Male	(16/15,54/213)	4.18 [1.81, 9.72]	.009
Depression	65≤age; Male	(26/38,44/190)	2.94 [1.55, 5.58]	.014
Neurological disorders	65≤age; Male	(28/47,42/181)	2.56 [1.38, 4.73]	.022
End stage renal disease	65≤age; Male	(19/26,51/202)	2.88 [1.39, 5.9]	.027
Chronic kidney disease	65≤age; Male	(40/85,30/143)	2.24 [1.26, 4.02]	.034
Fluid and electrolyte disorders	65≤age; Male	(17/22,53/206)	2.99 [1.39, 6.38]	.034

[a] Number of patients with the condition (cases) in the complicated/non-complicated COVID-19 cohorts, followed by the corresponding numbers in patients without the studied condition (controls).

[b] Odds ratios (ORs) and 95% confidence intervals (in brackets).. ORs greater than 1 suggest an increased risk for COVID-19 complications in patients with the noted condition. In each stratum, rows are sorted ascendingly by P-value.

Discussion

We compared the prevalence of dozens of existing conditions in Israeli SARS-CoV-2 positive and complicated COVID-19 patient cohorts, to highlight those conditions associated with high risk of complications. A few other studies have employed a similar study design to identify risk factors for COVID-19 complications. For example, Ebinger et al [8] studied a cohort of symptomatic COVID-19 individuals (N=442) and examined the association of existing conditions with disease severity; and the OpenSAFELY Collaborative explored the risk of COVID-19-related hospital death among the general population (N>17M). We emphasize that cohort composition dictates the research question it can address: our analysis focuses on SARS-CoV-2 positive individuals, hence searches for risk factors of complications in patients who already contracted the virus (but are potentially asymptomatic), while studying the general population may combine risk factors for infection and COVID-19 severe outcome. Additionally, cohorts that consider only a subset of patients, defined based on disease outcome (e.g., symptomatic or hospitalized) or otherwise non-representative of the entire population (e.g., demographically skewed), may introduce biases to the analysis [21]; instead, we study here all SARS-CoV-2 infected patients in a large, nationwide health organization.

Multiple studies (e.g., [7,22]) have shown that COVID-19 complications are most strongly associated with age and sex. Stratifying by these factors provides readily interpretable insights on the supplemental associations (on top of older age and being male) between pre-existing conditions and disease complications.

Many conditions highlighted by our analysis have been previously reported [5,6,8] and are part of commonly used at-risk definitions [1,3], including hypertension, obesity, kidney and cardiovascular

diseases. We do, however, identify a few additional risk factors, notably depression in patients 18-50 years old and males 65 years or more; and cognitive and neurological disorders in patients 65 years or older. These additions may be, in part, associated with different age distribution in the 65+ years group (median 76y, IQR [70-83.5y] versus 72y [68-78y] in the complicated and non-complicated COVID-19 cohorts, respectively) and rely on small sample size (only seven 18-50y patients with depression in the complicated COVID-19 cohort; Table 3); nonetheless, with some preliminary support [7], they may deserve more consideration in future studies. Our analysis also points out to reduced importance of respiratory diseases and smoking. Both conditions appear as factors in most at-risk definitions [3,5]: Chronic obstructive pulmonary disease has been associated with severe COVID-19 in multiple studies [23] (though not all [6]), while the role of smoking has been somewhat controversial [23,24]. The discrepancies between our analysis and previous reports likely stem from the different cohorts analyzed: SARS-CoV-2 positive individuals, ranging from asymptomatic to severe COVID-19 versus hospitalized COVID-19 patients, respectively. Other study-related attributes, for example country-specific characteristics, may also contribute to the varying importance of the studied risk factors.

In parallel to the COVID-19 epidemiological characterization efforts, researchers have also attempted to use retrospective observational data to derive risk models for severe COVID-19 patients [25]. Such models require ample data of COVID-19 patients for both model training and performance assessment. As, currently, such data are scarce, some models compromised on using data for other diseases with, supposedly, similar clinical trajectory and complications. For example, DeCapprio et al [26] trained models on US Medicare claims data to predict inpatient visits with a primary diagnosis of either pneumonia, influenza, acute bronchitis, or other specified upper respiratory infections as proxy for COVID-19 complications. However, as previously reported (e.g., in [27]), and in agreement with our analysis, severe COVID-19 patient characteristics differ considerably from other diseases', thus limiting the generalizability of such models to COVID-19 and requiring adjustments of their parameters [4].

Our study has several limitations. First and foremost, it relies on routinely maintained electronic health records, which may be inaccurate and incomplete [28]. Second, the number of complicated COVID-19 patients in MHS data is below 200, limiting the statistical power of our analysis. Third, healthcare policies and, in particular, testing criteria, may systematically bias the composition of SARS-CoV-2 positive cohort. Fourth, asymptomatic and mild COVID-19 patients (currently in the non-complicated cohort) may deteriorate and eventually be part of the complicated cohort, potentially modifying the results of the analysis. Fifth, our analysis is univariate in nature, testing the association of individual conditions with COVID-19 complications; as such, it is unable to uncover more complex relations, e.g., interdependencies between existing conditions and COVID-19 complications, which may be discovered by multivariate analysis. Finally, we focused on data from Israel; characteristics in other geographies may differ [27]. We attempted to mitigate some of these limitations by age and sex stratification and robust estimations of statistical significance. We also note that, at the current point in time, many of these shortcomings are shared by all published COVID-19 research work.

Notwithstanding these limitations, our work adopts a novel vantage point to the problem of identifying patients at increased risk for COVID-19 complications. Importantly, as SARS-CoV-2 containment efforts focus on patients at risk for severe complications (for example, shielding vulnerable population in the UK [3]), changes in the list of considered conditions may have huge effect on a large number of individuals, thus calling for continuous fine-tuning of the corresponding definitions.

Supplementary Information

Supplementary File 1. Odds ratio analysis.

Ethical approval

The study was approved by Maccabi Health Services' institutional review board (0024-20-MHS).

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors

Data sharing statement

Data are available upon reasonable request. According to Israeli regulations, no patient-level secondary used medical data can be publicly shared.

Patient and Public Involvement

Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research

Acknowledgement

We thank Guy Amit, Maytal Bivas-Benita, and Irena Girshovitz, KI Research Institute, and Ran Nir-Paz, Hadassah Medical Center, for insightful discussions and comments.

Conflict of Interest

All authors have completed the ICMJE uniform disclosure form, with the following declarations made: PA reports personal fees and other from Medial Research, outside the submitted work.

Author Contributions

Conceptualization: NK; **Data Curation:** BM, KM, YB; **Investigation and methodology**: CY, BM, NK; **Project Administration**: PA, VS; **Writing – original draft preparation**: CY; **Writing – review and editing**: CY, BM, KM, PA, YB, GC.

References

1. Centers for Disease Control and Prevention. Coronavirus Disease 2019 (COVID-19) [Internet]. 2020 [cited 2020 Apr 21]. Available from: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-at-higher-risk.html

2. European Centre for Disease Prevention and Control. Rapid risk assessment: Coronavirus disease 2019 (COVID-19) pandemic: increased transmission in the EU/EEA and the UK – eighth update [Internet]. Eur Cent Dis Prev Control. 2020 [cited 2020 Apr 21]. Available from: https://www.ecdc.europa.eu/en/publications-data/rapid-risk-assessment-coronavirus-disease-2019-covid-19-pandemic-eighth-update

3. United Kingdom National Health Service. Guidance and updates for GPs: Highest clinical risk patients [Internet]. [cited 2020 Apr 27]. Available from: https://www.england.nhs.uk/coronavirus/publication/guidance-and-updates-for-gps-at-risk-patients/

4. Barda N, Riesel D, Akriv A, Levi J, Finkel U, Yona G, Greenfeld D, Sheiba S, Somer J, Bachmat E, Rothblum GN, Shalit U, Netzer D, Balicer R, Dagan N. Performing risk stratification for COVID-19

when individual level data is not available, the experience of a large healthcare organization [Internet]. 2020 [cited 2020 May 10]. [doi: 10.1101/2020.04.23.20076976]

5. Xu L, Yaqian M, Chen G. Risk factors for severe corona virus disease 2019 (COVID-19) patients : a systematic review and meta analysis | medRxiv [Internet]. [cited 2020 Apr 21]. Available from: https://www.medrxiv.org/content/10.1101/2020.03.30.20047415v1

6. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, Barnaby DP, Becker LB, Chelico JD, Cohen SL, Cookingham J, Coppa K, Diefenbach MA, Dominello AJ, Duer-Hefele J, Falzon L, Gitlin J, Hajizadeh N, Harvin TG, Hirschwerk DA, Kim EJ, Kozel ZM, Marrast LM, Mogavero JN, Osorio GA, Qiu M, Zanos TP. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA [Internet] 2020 Apr 22 [cited 2020 Apr 28]; [doi: 10.1001/jama.2020.6775]

7. The OpenSAFELY Collaborative, Williamson E, Walker AJ, Bhaskaran KJ, Bacon S, Bates C, Morton CE, Curtis HJ, Mehrkar A, Evans D, Inglesby P, Cockburn J, Mcdonald HI, MacKenna B, Tomlinson L, Douglas IJ, Rentsch CT, Mathur R, Wong A, Grieve R, Harrison D, Forbes H, Schultze A, Croker RT, Parry J, Hester F, Harper S, Perera R, Evans S, Smeeth L, Goldacre B. OpenSAFELY: factors associated with COVID-19-related hospital death in the linked electronic health records of 17 million adult NHS patients. [Internet]. 2020 [cited 2020 May 11]. [doi: 10.1101/2020.05.06.20092999]

8. Ebinger JE, Achamallah N, Ji H, Claggett BL, Sun N, Botting P, Nguyen T-T, Luong E, Kim EH, Park E, Liu Y, Rosenberry R, Matusov Y, Zhao S, Pedraza I, Zaman T, Thompson M, Raedschelders K, Berg AH, Grein JD, Noble PW, Chugh SS, Bairey Merz CN, Marbán E, Van Eyk JE, Solomon SD, Albert CM, Chen P, Cheng S. Pre-Existing Traits Associated with Covid-19 Illness Severity [Internet]. 2020 [cited 2020 May 11]. [doi: 10.1101/2020.04.29.20084533]

9. Shalev V, Chodick G, Goren I, Silber H, Kokia E, Heymann AD. The use of an automated patient registry to manage and monitor cardiovascular conditions and related outcomes in a large health organization. Int J Cardiol 2011 Nov;152(3):345–349. [doi: 10.1016/j.ijcard.2010.08.002]

10. Sella T, Shoshan A, Goren I, Shalev V, Blumenfeld O, Laron Z, Chodick G. A retrospective study of the incidence of diagnosed Type 1 diabetes among children and adolescents in a large health organization in Israel, 2000-2008. Diabet Med J Br Diabet Assoc 2011 Jan;28(1):48–53. PMID:21166845

11. Chodick G, Heymann AD, Shalev V, Kookia E. The epidemiology of diabetes in a large Israeli HMO. Eur J Epidemiol 2002 Dec;18(12):1143–1146. [doi: 10.1023/B:EJEP.0000006635.36802.c8]

12. Weitzman D, Chodick G, Shalev V, Grossman C, Grossman E. Prevalence and factors associated with resistant hypertension in a large health maintenance organization in Israel. Hypertens Dallas Tex 1979 2014 Sep;64(3):501–507. PMID:24958503

13. Goldshtein I, Nguyen AM, dePapp AE, Ish-Shalom S, Chandler JM, Chodick G, Shalev V. Epidemiology and correlates of osteoporotic fractures among type 2 diabetic patients. Arch Osteoporos 2018 03;13(1):15. PMID:29502187

14. Yu J, Goldshtein I, Shalev V, Chodick G, Ish-Shalom S, Sharon O, Modi A. Renal impairment among postmenopausal women with osteoporosis from a large health plan in Israel. Arch Osteoporos 2015;10:210. PMID:25753617

15. Kodesh A, Goldshtein I, Gelkopf M, Goren I, Chodick G, Shalev V. Epidemiology and comorbidity of severe mental illnesses in the community: findings from a computerized mental health registry in a large Israeli health organization. Soc Psychiatry Psychiatr Epidemiol 2012 Nov;47(11):1775–1782. PMID:22310700

16. Metcalfe D, Masters J, Delmestri A, Judge A, Perry D, Zogg C, Gabbe B, Costa M. Coding algorithms for defining Charlson and Elixhauser co-morbidities in Read-coded databases. BMC Med Res Methodol 2019 Jun 6;19(1):115. PMID:31170931

17. Kuan V, Denaxas S, Gonzalez-Izquierdo A, Direk K, Bhatti O, Husain S, Sutaria S, Hingorani M, Nitsch D, Parisinos CA, Lumbers RT, Mathur R, Sofat R, Casas JP, Wong ICK, Hemingway H, Hingorani AD. A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service. Lancet Digit Health 2019 Jun 1;1(2):e63–e77. [doi: 10.1016/S2589-7500(19)30012-3]

18. Clinical Classifications Software (CCS) 2015 [Internet]. Agency for Healthcare Research and Quality Healthcare Cost and Utilization Project (HCUP); Available from: https://www.hcup-us.ahrq.gov/toolssoftware/ccs/CCSUsersGuide.pdf

19. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B Methodol 1995;57(1):289–300. [doi: 10.2307/2346101]

20. Elm E von, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol Elsevier; 2008 Apr 1;61(4):344–349. PMID:18313558

21. Griffith G, Morris TT, Tudball M, Herbert A, Mancano G, Pike L, Sharp GC, Palmer TM, Smith GD, Tilling K, Zuccolo L, Davies NM, Hemani G. Collider bias undermines our understanding of COVID-19 disease risk and severity. medRxiv Cold Spring Harbor Laboratory Press; 2020 May 20;2020.05.04.20090506. [doi: 10.1101/2020.05.04.20090506]

22. Prieto-Alhambra D, Ballo E, Coma-Redon E, Mora N, Aragon M, Prats-Uribe A, Fina-Aviles F, Benitez M, Guiriguet C, Fabregas M, Medina-Peralta M, Duarte-Salles T. Hospitalization and 30-day fatality in 121,263 COVID-19 outpatient cases. medRxiv Cold Spring Harbor Laboratory Press; 2020 May 8;2020.05.04.20090050. [doi: 10.1101/2020.05.04.20090050]

23. Zhao Q, Meng M, Kumar R, Wu Y, Huang J, Lian N, Deng Y, Lin S. The impact of COPD and smoking history on the severity of Covid-19: A systemic review and meta-analysis. J Med Virol [Internet] [cited 2020 May 6];n/a(n/a). [doi: 10.1002/jmv.25889]

24. Lippi G, Henry BM. Active smoking is not associated with severity of coronavirus disease 2019 (COVID-19). Eur J Intern Med Elsevier; 2020 May 1;75:107–108. PMID:32192856

25. Wynants L, Van Calster B, Bonten MMJ, Collins GS, Debray TPA, De Vos M, Haller MC, Heinze G, Moons KGM, Riley RD, Schuit E, Smits LJM, Snell KIE, Steyerberg EW, Wallisch C, van Smeden M. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020 Apr 7;m1328. [doi: 10.1136/bmj.m1328]

26. DeCaprio D, Gartner J, Burgess T, Kothari S, Sayed S, McCall CJ. Building a COVID-19 Vulnerability Index. ArXiv200307347 Cs Stat [Internet] 2020 Mar 23 [cited 2020 Apr 21]; Available from: http://arxiv.org/abs/2003.07347

27. Burn E, You SC, Sena A, Kostka K, Abedtash H, Abrahao MTF, Alberga A, Alghoul H, Alser O, Alshammari TM, Areia C, Banda JM, Cho J, Culhane AC, Davydov A, DeFalco FJ, Duarte-Salles T, DuVall SL, Falconer T, Gao W, Golozar A, Hardin J, Hripcsak G, Huser V, Jeon H, Jing Y, Jung CY, Kaas-Hansen BS, Kaduk D, Kent S, Kim Y, Kolovos S, Lane J, Lee H, Lynch KE, Makadia R, Matheny ME, Mehta P, Morales DR, Natarajan K, Nyberg F, Ostropolets A, Park RW, Park J, Posada JD, Prats-Uribe A, Rao GA, Reich C, Rho Y, Rijnbeek P, Sathappan SMK, Schilling LM, Schuemie M, Shah NH, Shoaibi A, Song S, Spotnitz M, Suchard MA, Swerdel J, Vizcaya D, Volpe S, Wen H, Williams AE, Yimer BB, Zhang L, Zhuk O, Prieto-Alhambra D, Ryan P. An international characterisation of patients hospitalised with COVID-19 and a comparison with those previously hospitalised with influenza. medRxiv Cold Spring Harbor Laboratory Press; 2020 Apr 25;2020.04.22.20074336. [doi: 10.1101/2020.04.22.20074336]

28. Myers L, Stevens J. Using EHR to Conduct Outcome and Health Services Research. In: MIT Critical Data, editor. Second Anal Electron Health Rec [Internet] Cham: Springer International Publishing; 2016 [cited 2020 Jul 23]. p. 61–70. [doi: 10.1007/978-3-319-43742-2_7]

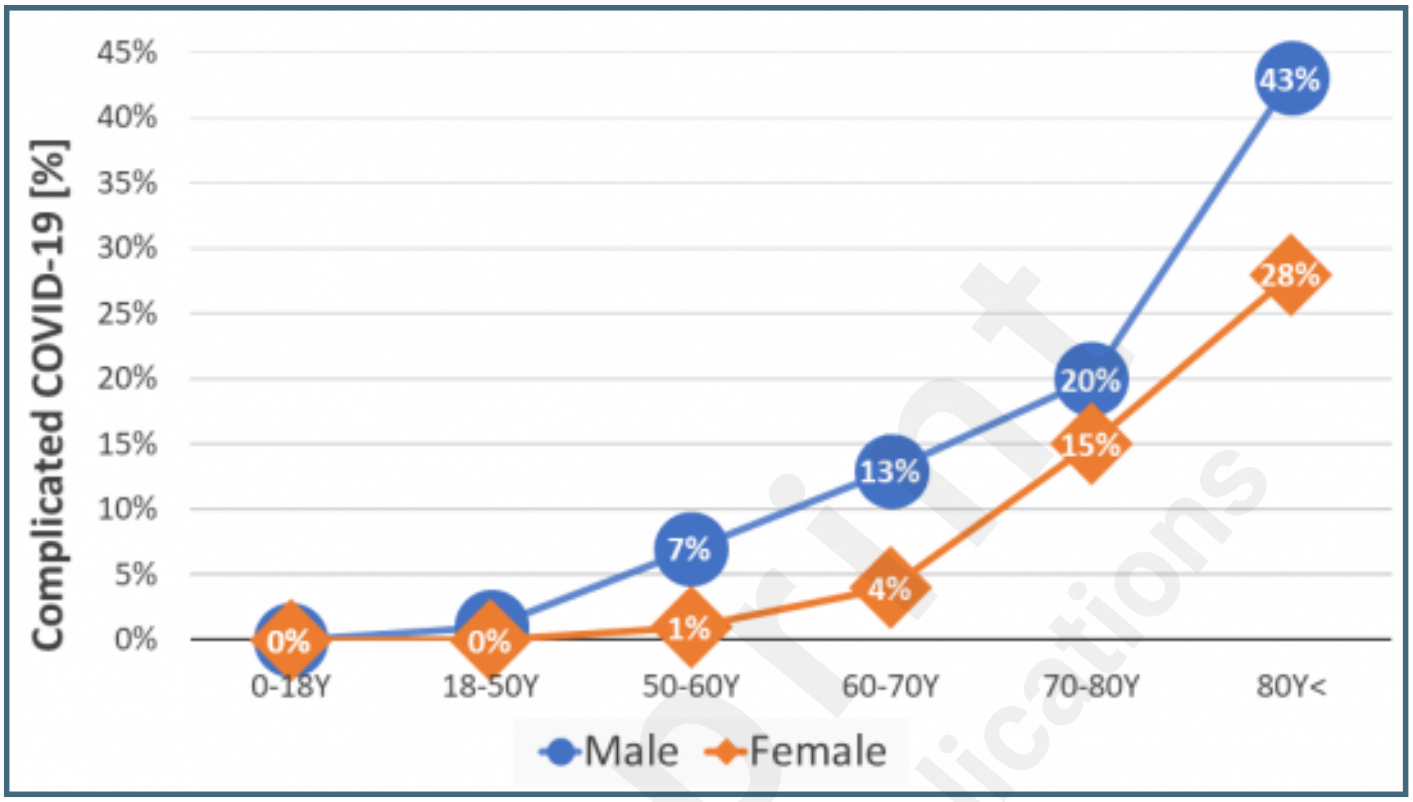
Supplementary Files

Multimedia Appendixes

Odds ratio analysis.
URL: https://asset.jmir.pub/assets/221a5b3fd1cc8ed4a5228d103695f08c.xlsx

Figures

Age and complicated COVID-19. Prevalence of complicated COVID-19 (moderate or severe condition, y-axes) in different age groups (x-axis), shown separately for males (blue) and females (orange).



CONSORT (or other) checklists

STROBE cohort checklist.
URL: https://asset.jmir.pub/assets/32e7538e42676b1fb56b72bac3507577.pdf

Related publication(s) - for reviewers eyes onlies

Cover letter and point-by-point response to editor/reviewer comments.
URL: https://asset.jmir.pub/assets/ca431aaff65305403302a6df8a4a1c96.pdf