BMJ Open

Characteristics and outcomes of hospitalised adults with COVID-19 in a Global Health Research Network: a cohort study

Julia Zhu ,[1] Zhongyuan Wei,[1] Manasi Suryavanshi,[1] Xiu Chen,[1] Qian Xia,[1] Jenny Jiang,[1] Olulade Ayodele,[2] Brian D Bradbury,[3] Corinne Brooks,[3] Carolyn A Brown,[3] Alvan Cheng,[3] Cathy W Critchlow,[4] Giovanna Devercelli,[2] Vivek Gandhi,[2] Kathleen Gondek,[2] Ajit A Londhe ,[3] Junjie Ma ,[3] Michele Jonsson-Funk,[5] Hillary A Keenan,[2] Sudhakar Manne,[2] Kaili Ren,[2] Lynn Sanders,[2] Peter Yu,[2] Jie Zhang,[3] Linyun Zhou,[2] Ying Bao[1]

To cite: Zhu J, Wei Z, Suryavanshi M, *et al*. Characteristics and outcomes of hospitalised adults with COVID-19 in a Global Health Research Network: a cohort study. *BMJ Open* 2021;**11**:e051588. doi:10.1136/bmjopen-2021-051588

► Prepublication history and additional online supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-051588).

Received 30 March 2021
Accepted 14 July 2021

For numbered affiliations see end of article.

Correspondence to
Dr Ying Bao; ying.bao@bms.com

ABSTRACT

Objective To examine age, gender, and temporal differences in baseline characteristics and clinical outcomes of adult patients hospitalised with COVID-19.

Design A cohort study using deidentified electronic medical records from a Global Research Network.

Setting/Participants 67 456 adult patients hospitalised with COVID-19 from the USA; 7306 from Europe, Latin America and Asia-Pacific between February 2020 and January 2021.

Results In the US cohort, compared with patients 18–34 years old, patients ≥65 had a greater risk of intensive care unit (ICU) admission (adjusted HR (aHR) 1.73, 95% CI 1.58 to 1.90), acute respiratory distress syndrome(ARDS)/respiratory failure (aHR 1.86, 95% CI 1.76 to 1.96), invasive mechanical ventilation (IMV, aHR 1.93, 95% CI, 1.73 to 2.15), and all-cause mortality (aHR 5.6, 95% CI 4.36 to 7.18). Men appeared to be at a greater risk for ICU admission (aHR 1.34, 95% CI 1.29 to 1.39), ARDS/respiratory failure (aHR 1.24, 95% CI1.21 to 1.27), IMV (aHR 1.38, 95% CI 1.32 to 1.45), and all-cause mortality (aHR 1.16, 95% CI 1.08 to 1.24) compared with women. Moreover, we observed a greater risk of adverse outcomes during the early pandemic (ie, February–April 2020) compared with later periods. In the ex-US cohort, the age and gender trends were similar; for the temporal trend, the highest proportion of patients with all-cause mortality were also in February–April 2020; however, the highest percentages of patients with IMV and ARDS/respiratory failure were in August–October 2020 followed by February–April 2020.

Conclusions This study provided valuable information on the temporal trends of characteristics and outcomes of hospitalised adult COVID-19 patients in both USA and ex-USA. It also described the population at a potentially greater risk for worse clinical outcomes by identifying the age and gender differences. Together, the information could inform the prevention and treatment strategies of COVID-19. Furthermore, it can be used to raise public awareness of COVID-19's impact on vulnerable populations.

Strengths and limitations of this study

- ► This study investigated characteristics and clinical outcomes of patients with COVID-19 using data from a Global Health Research Network, which includes COVID-19 test results and a wide range of clinical measures.
- ► The long follow-up (February 2020–January 2021) allowed a thorough examination of the temporal trend of the COVID-19 clinical outcomes.
- ► The large sample size allowed extensive sensitivity analyses to test the robustness of the study.
- ► As with other observational studies using electronic health records, random measurement error is inevitable and may bias the results towards the null.

INTRODUCTION

The COVID-19 is a newly discovered infectious viral disease caused by SARS-CoV-2. The virus not only affects the respiratory system but also causes damage to other systems, and may act as a precipitating factor to worsen existing conditions, potentially leading to death.[1 2] The first human case was reported in December 2019 in Wuhan City, China.[3] WHO declared the COVID-19 outbreak a public health emergency of international concern on 30 January 2020 and a pandemic on 11 March 2020.[4] As of 20 June 2021, the pandemic has resulted in over 178.4 million cases and 3.8 million deaths worldwide.[5] The first case of COVID-19 in the USA was reported on 20 January 2020 in Washington State, USA.[6 7] As of 20 June 2021, the USA had the highest number of reported infections, with more than 33.5 million confirmed cases and more than 0.6 million deaths.[5] A study conducted in China found higher

cases of COVID-19 among men compared with women with higher case fatality in men compared with women (2.8% vs 1.7%).[8] Similar gender disparity in mortality has been seen in reports from Italy,[9] England, and Wales.[10] The Centers for Disease Control and Prevention also reported higher fatality among older adults with 80% of all COVID-19 deaths reported in the US occurring in adults ≥65 years.[11]

As COVID-19 cases have risen exponentially around the world since the start of the pandemic, there is a necessity to document the temporal changes in patient characteristics, and the impact of real-world clinical practice on outcomes including intensive care unit (ICU) admissions, acute respiratory distress syndrome (ARDS)/respiratory failure, invasive mechanical ventilation (IMV), and all-cause mortality over time among patients with COVID-19. The objectives of this study were to provide real-world evidence on gender and age differences in COVID-19 outcomes and to understand temporal trends in outcomes among patients with COVID-19 in the USA and outside of the USA.

METHODS

Data source

This study used TriNetX (www.trinetx.com), a global federated research network. The TriNetX database has been described in detail in other published papers.[12–16] TriNetX network provides a dataset of electronic medical records (EMR, diagnoses, procedures, medications, laboratory values and genomic information) from different healthcare organisations (HCOs). The HCOs contributing EMR data to the TriNetX network are large academic medical institutions, specialty physician services and community hospitals providing on average 7 years of historical patient data from both inpatient and outpatient facilities. The US analysis was conducted using EMR data download from 44 different HCOs covering over 61 million patients that reside predominantly in the USA; the data are deidentified based on the standard defined in Section §164.514(a) of the HIPAA Privacy Rule. The process by which data sets are deidentified is attested to through a formal determination by a qualified expert as defined in Section §164.514(b)(1) of the HIPAA Privacy Rule. The ex-US analysis used the federated cloud-based TriNetX network, representing 6.6 million patients from 12 HCOs in Spain, the UK, Brazil, Australia, India, Malaysia and Taiwan, as of 31 January 2021. The ex-US TriNetX platform provides aggregated counts and statistical summaries of deidentified information. Protected Health Information or personal data is not available to the users of the platform. As TriNetX allows real-time access to the data, the platform was queried to generate results for this study. The network contains data that is provided by participating HCOs, each of which represents and warrants that it has all necessary rights, consents, approvals and authority to provide the data to TriNetX so long as their name remains anonymous as a data source and their data are used for research purposes. The data shared through the TriNetX platform are attenuated to ensure that they do not include sufficient information to facilitate reidentification nor allow for the determination of which HCO contributed specific information about a patient. Within HIPAA, TriNetX has a business associate agreement with each of the HCOs. Overall, the TriNetX EMR database includes COVID-19 patients from mixed payer types, geographies and demographic backgrounds, representing a geographically and socioeconomically diverse population both in and outside of the USA.

Study design, setting and participants

This was an observational cohort study that identified patients at their earliest episode of COVID-19 and subsequently followed up to describe their disease progression, treatment received, and outcomes using EMR. This study included adults (aged ≥18 years) who were hospitalised with COVID-19. COVID-19 diagnosis was defined as the first occurrence from 1 February 2020 to 31 January 2021 of any of the following: (1) positive SARS-CoV-2 RNA or antigen test; (2) ICD-10-CM diagnosis code U07.1, J12.81, J12.89, or J80; (3) ICD-10-CM code B97.29 or B34·2 occurring between 1 February 2020 and 30 April 2020.

Patients were excluded, if (1) first COVID-19 diagnosis occurring within the last 28 days of the available data; (2) missing data on age or gender; (3) continuous hospitalisation starting >10 days before COVID-19 diagnosis date; (4) with diagnosis or procedure codes for labor and delivery during the index hospitalisation or (5) with diagnosis codes for trauma, injury, fracture or poisoning during the first 2 days of the earliest hospitalisation. See figure 1A,B for the patient flow diagram for the US and ex-US analyses, respectively. If a patient had multiple hospitalisations that met the study criteria, only the earliest hospitalisation was included in this study.

The index date was defined as the first COVID-19 diagnosis date or hospital admission date, whichever comes earlier; the follow-up window was from the index date to the earliest of the following: the end of data availability, discharge date or death date. Health outcomes were assessed within the duration from the first COVID-19-related hospital admission date to the earliest of the following events: hospital discharge, death, end of data, or 28 days after hospital admission.

For temporal trend analysis, patients were stratified into four time periods based on their month of index date (February–April 2020, May–July 2020, August–October 2020, and November 2020–January 2021) for all analyses.

Variables

The primary outcomes of interest were ICU admission, ARDS/respiratory failure, IMV and all-cause mortality (please see medical codes in online supplemental eAppendix table 1). For ex-US, ICU admission was not reported given the consideration of potential misclassification. To comply with the data privacy agreement, only the patient's death month is available in the US

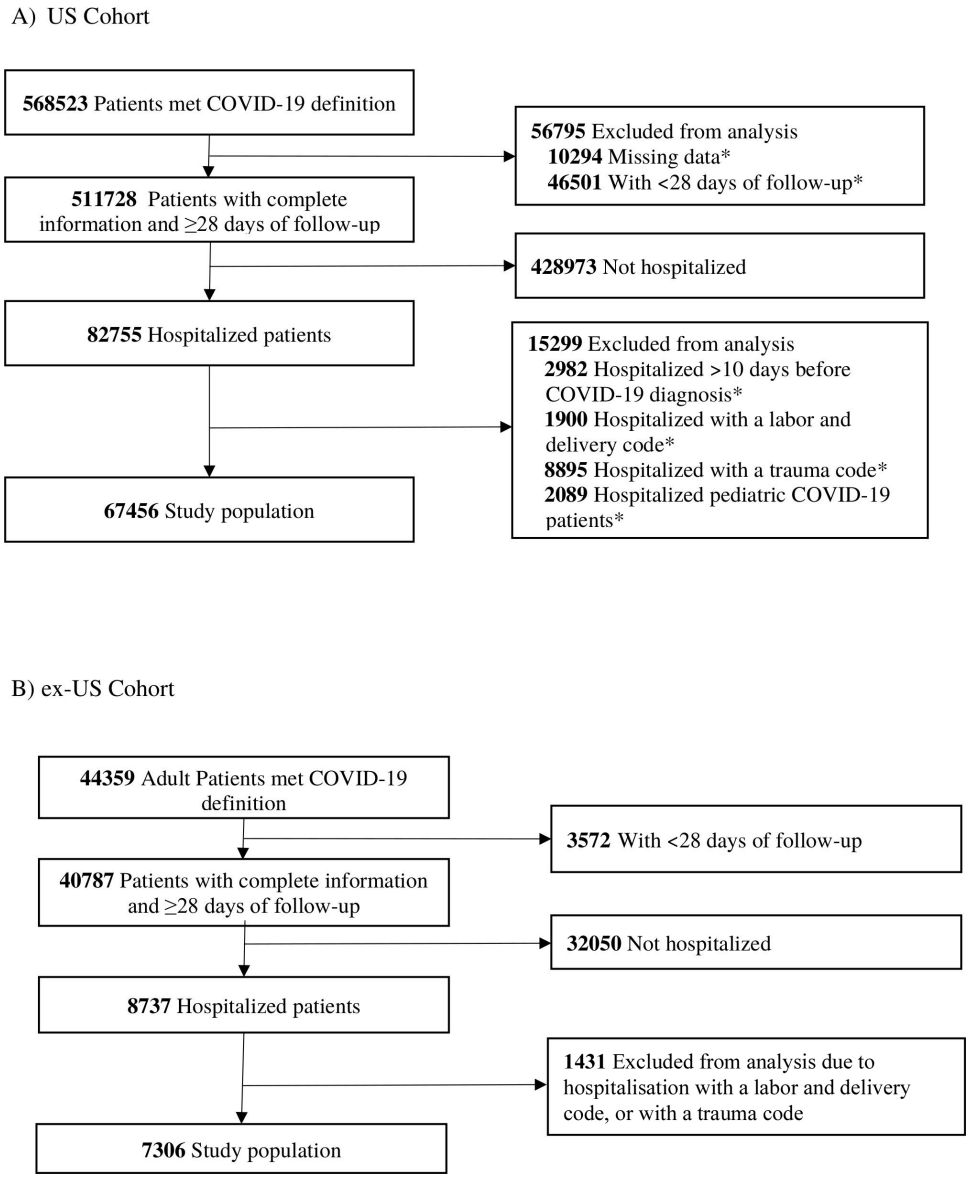


Figure 1 Flow diagram of US and ex-US study cohorts.*The counts in the exclusion criteria were not mutual exclusive.

data. Thus, the patient death date was inferred from a patient's last physically present/recorded date using the following: procedure (date), diagnosis (date), encounter (end date), vital signs (date) and medication (prescribing date).[17] In the ex-US analysis, the death date was used as reported in the platform.

Patient characteristics such as demographics (age, sex, race, geographical region, and ethnicity), and calendar month of index COVID-19 hospitalisation were reported. The baseline comorbid conditions were evaluated using 12-month data prior to the index date (excluding index date).

Statistical methods

The demographic and clinical characteristics of the primary cohort and subcohorts at the index date were summarised using descriptive statistics. Mean, SD, SE and 95% CI were reported for continuous variables, and counts and percentages were reported for categorical variables.

For the US cohort, the proportional subdistribution hazard model by Fine and Gray was used for estimating the HRs and cumulative incidence function (CIF) for all four outcomes; the hospital discharge date was treated as a competing risk in estimating HRs and CIF for mortality, whereas hospital discharge and death were treated as competing risks in estimating HRs and CIF for other outcomes.[18 19] Three separate Fine and Gray models were developed stratified by age group, gender, and time period; each model used the other two main categories of risk exposure (eg gender and time period in the age group model) as covariates. The other covariates in each model were race, ethnicity, US region, hypertension, Type I/II diabetes, active cancer, coronary artery disease, congestive heart failure, chronic kidney disease, chronic obstructive pulmonary disease, asthma

and obstructive sleep apnoea. The cumulative incidence of selected outcomes was plotted against time from the day of admission through 28 days assuming censoring at the study end date (ie, 31 January 2021) and stratified by gender, age and time period, respectively. All US analyses were performed using SAS V.9.4 (SAS Institute) with a two-tailed $p<0.05$ considered statistically significant. As for the ex-US cohort, the analysis was performed using a cloud-based TriNetX analytical platform. The platform provided aggregated counts and statistical summaries but did not allow the estimation of adjusted HRs (aHR) and model-adjusted cumulative incidences.

As some diagnosis codes used to identify US COVID-19 patients in the study were not specific to COVID-19, we performed sensitivity analyses that include only patients with 'confirmed' (at least one positive SARS-CoV-2 RNA or antigen test within 21 days of their index date) and 'probable' COVID-19 diagnosis (no documented positive lab test for COVID-19 within 21 days following the index date). In the sensitivity analyses, we also examined patients' baseline characteristics, treatment during hospitalisation and selected outcomes in 'confirmed' and 'probable' cohorts in the US, respectively. The sensitivity analysis was not available for the ex-US study due to the limitation of the TriNetX ex-US analytical platform.

Patient and public involvement

No patient involved.

RESULTS

A total of 67 456 patients hospitalised with COVID-19 were identified from February 2020 to January 2021 in the US (figure 1A), with a mean (SD) age of 58.3 (17.3) years old; 51.2% (n=34 518) were men (table 1). The highest proportion of patients in the US cohort had their index hospitalisation in November 2020 and December 2020 (15.9% and 17.7%, respectively). The most common comorbid condition was cardiovascular disease (n=25 970, 38.5%) including patients with hypertension (n=23 272, 34.5%), followed by gastrointestinal disorders (n=15 987, 23.7%) and type II diabetes (n=14 031, 20.8%) (table 1). The proportion of patients with chronic lung diseases was higher in women compared with men, driven by a higher proportion of asthma in women (online supplemental eAppendix table 2). The burden of comorbidities increased with age (online supplemental eAppendix table 3). In the USA, the proportion of White patients with index COVID-19 hospitalisation increased over subsequent months with the highest proportion seen during November 2020–January 2021; whereas the proportion of African American and Hispanic patients with index COVID-19 hospitalizations decreased over months, with the lowest seen during November 2020–January 2021. The index COVID-19 hospitalizations in the Northeast region were highest at the beginning of the pandemic (27.7%), followed by a subsequent drop in infection rates between May and October 2020 then a sharp rise

Table 1 Baseline characteristics of US and ex-US hospitalised adult patients with COVID-19

	Patient, %	
Characteristic	US cohort (n=67 456)	Ex-US cohort (n=7306)
Age, mean (SD), years	58.3 (17.3)	61.0 (16.9)
Male	51.2	55.1
Race		
White	56.1	19.4
Black or African American	23.7	0.8
Asian	3.2	0.8
American Indian or other Pacific Islander	0.9	NA
Unknown	16.1	79.0
Ethnicity		
Hispanic	24.8	NA
Month of index		
February 2020	1.2	0.3
March 2020	6.7	30.1
April 2020	11.1	18.4
May 2020	7.3	6.0
June 2020	7.8	3.2
July 2020	11.1	2.4
August 2020	6.5	8.5
September 2020	5.3	13.3
October 2020	8.5	7.7
November 2020	15.9	5.2
December 2020	17.7	4.6
January 2021	1.0	0.4
Baseline comorbidities		
Cardiovascular disease	38.5	15.9
Hypertension	34.5	13.1
Gastrointestinal disorders	23.7	8.6
Skin disorder	12.6	5.5
Cancer	7.6	5.8
Solid tumours	5.7	4.2
Haematologic malignancies	2.6	2.5
Chronic kidney disease	12.2	4.1
Chronic lung disease	12.8	6.8
Asthma	6	2.2
Chronic obstructive pulmonary disease	7.7	4.5
Pulmonary fibrosis	1	0.5
Diabetes mellitus		
Type I diabetes	1.6	0.4
Type II diabetes	20.8	6.7
Liver disease	2.5	5.6

in November 2020–January 2021 (14.7%). The Midwest region saw a gradual uptick in the infection rates over the course of the pandemic with the highest rates (21.9%) seen from November 2020–January 2021. The US patients hospitalised with COVID-19 during November 2020–January 2021 had more cardiovascular diseases compared with the patients in earlier months (online supplemental eAppendix Table 3).

The ex-US cohort included a total of 7306 patients hospitalised with COVID-19 (figure 1B), with 91.4% of them from Europe (online supplemental eAppendix table 4). The mean (SD) age, 61.0 (16.9), was greater than that in the US cohort, 55.1% of the cohort were men, and the 48.8% patients had their index hospitalisation in February–April 2020. A detailed description of the ex-US cohort was presented in table 1.

In the US cohort, the most commonly used medications post index were acetaminophen 45 269 (67.1%), followed by enoxaparin 33 370 (49.5%), dexamethasone (37.7%),insulin (32.7%), heparin (14.8%), aspirin (26.7%), azithromycin (23.7%) and methylprednisolone (19.8%). The most commonly used antiviral among the US cohort was remdesivir 13 667 (20.3%). Among the ex-US cohort, the most frequently used medication post index was also acetaminophen 4719 (64.6%), followed by azithromycin 2842 (38.9%), hydroxychloroquine 2119 (29.0%), methylprednisolone 1563 (21.4%), insulin 1476 (20.2%), dexamethasone 1454 (19.9%) and amoxicillin 1381 (18.9%). For additional details refer to online supplemental eAppendix tables 5 and 6.

Among US patients, this study showed that outcomes worsened with increased age, with the highest aHR in the age groups of 50–64 and 65+, compared with to those aged 18–39 (reference group). In particular, patients≥65 years had higher aHRs compared with the reference group for ICU admission (17.6% vs 7.2%; aHR 1.73, 95% CI 1.58 to 1.90, p<0.0001), ARDS/respiratory failure (43.1% vs 18.3%; aHR 1.86, 95% CI 1.76 to 1.96, p<0.0001), IMV (13.8% vs 5.1%; aHR 1.93, 95% CI, 1.73 to 2.15, p<0.0001), and all-cause mortality (9.6% vs 0.8%; aHR 5.6, 95% CI 4.36 to 7.18, p<0.0001). In contrast to women, men were more often admitted to the ICU (16.9% vs 12.3%; aHR 1.34, 95% CI 1.29 to 1.39, p<0.0001) and were at higher risk for ARDS/respiratory failure (41.0% vs 32.8%; aHR 1.24, 95% CI 1.21 to 1.27, p<0.0001), IMV (13.7% vs 9.4%; aHR 1.38, 95% CI 1.32 to 1.45, p<0.0001), and all-cause mortality (6.1% vs 4.6%; aHR 1.16, 95% CI 1.08 to 1.24, p<0.0001). Moreover, we observed the highest risk of worse outcomes during the early pandemic (ie, February–April 2020), compared with the later three time periods (May–July 2020, August–October 2020, and November 2020–January 2021), with significantly higher aHRs for all four outcomes. The risk of ICU admission and IMV decreased across the four time periods. The risk of ARDS/respiratory failure during the summer period of May–July 2020 were lowest among all four periods, while the all-cause mortality of August–October 2020 and November 2020–January 2021 were lower, compared with the periods of February–April 2020 or May–July 2020 (figure 2 and online supplemental eAppendix table 7).

In the US analysis, the model-adjusted cumulative incidence for all four outcomes measures at 7, 14, and 28 days were consistently higher among men compared with women, and among patients aged 50 and over compared with patients aged 18–49. With respect to time periods, the adjusted cumulative incidence for ICU admission, ARDS/respiratory failure, and IMV was markedly greater among patients with index date in February–April 2020 as compared with patients having their index COVID-19 hospital admission during the other three time periods (figure 3).

The ex-US analysis demonstrated similar age, and gender trends, although the data format and availability did not allow for similarly detailed analyses. Patients ≥65 years, compared with the age group of 18–39, had a higher incidence of ARDS/respiratory failure (31.5% vs 17.4%) and all-cause mortality (23.4% vs 1.0%). A higher percentage of patients aged between 50 and 64 years old had IMV compared with the other two age groups. The differences in adverse clinical events were also evident across gender; with higher proportions of men experienced ARDS/respiratory failure (29.0% vs 23.4%), IMV (5.0% vs 2.3%), and all-cause mortality (13.8% vs 10.4%) compared with women. The proportion of patients with ARDS/respiratory failure and IMV were highest in August–October 2020 and lowest in May–July 2020. As for the temporal trend, though the highest all-cause mortality was also observed in the period of February–April 2020, the highest and lowest percentages of patients with IMV and ARS/respiratory failure were found in August–October 2020 and May–July 2020, respectively (figure 4).

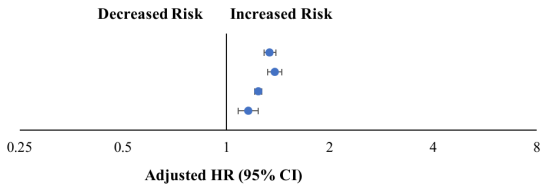
In the US analysis, the sensitivity analyses among subgroups of patients with confirmed COVID-19 diagnosis showed similar distributions of baseline characteristics, as well as comparable trends in adjusted cumulative incidence and aHRs for the primary outcomes of interest across age, gender, and time periods. Similar trends were also observed in the probable cohort. Details of sensitivity analysis results were presented in online supplemental eAppendix tables 8–12 and figures 1–6.

DISCUSSION

The COVID-19 pandemic has been rapidly evolving in the USA and other countries. It is critical to understand the demographic distribution and temporal trends in adverse clinical events associated with the disease such as ICU admissions, ARDS/respiratory failure, IMV, and all-cause mortality across the globe. The evidence generated from this study showed that patients ≥65 years seemed to have an approximately two fold increased risk of ICU admission, ARDS/respiratory failure and IMV; and about a 5.6-fold increase in all-cause mortality. The 7, 14 and 28 day (post-hospital admission date) adjusted cumulative incidence of these adverse clinical events was consistently higher in patients ≥65 years compared with those aged 18–49 years,

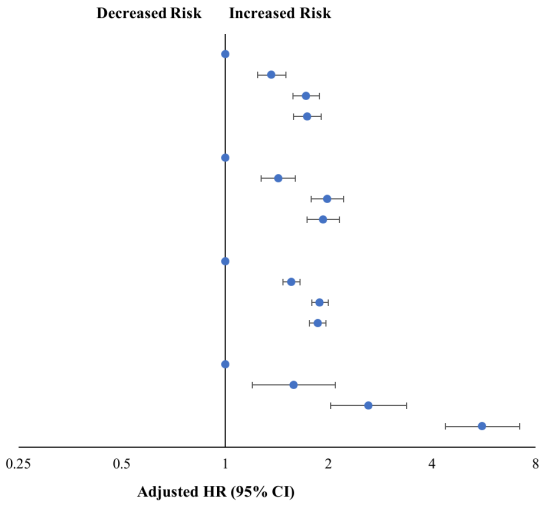
A) Gender

Characteristics	Adjusted HR (95%CI)	Pr > ChiSq
Male (Reference : Female)		
ICU Admission	1.34 (1.29-1.39)	<.0001
IMV	1.38 (1.32-1.45)	<.0001
ARDS/Respiratory Failure	1.24 (1.21-1.27)	<.0001
All-Cause Mortality	1.16 (1.08-1.24)	<.0001



B) Age

Characteristics	Adjusted HR (95%CI)	Pr > ChiSq
ICU Admission		
18-34	1 (ref)	
35-49	1.36 (1.24-1.50)	<.0001
50-64	1.72 (1.57-1.88)	<.0001
65+	1.73 (1.58-1.90)	<.0001
IMV		
18-34	1 (ref)	
35-49	1.43 (1.27-1.60)	<.0001
50-64	1.98 (1.78-2.21)	<.0001
65+	1.93 (1.73-2.15)	<.0001
ARDS/Respiratory Failure		
18-34	1 (ref)	
35-49	1.56 (1.47-1.65)	<.0001
50-64	1.88 (1.78-1.99)	<.0001
65+	1.86 (1.76-1.96)	<.0001
All-Cause Mortality		
18-34	1 (ref)	
35-49	1.58 (1.20-2.09)	0.0012
50-64	2.62 (2.03-3.37)	<.0001
65+	5.60 (4.36-7.18)	<.0001



C) Calendar Month

Characteristics	Adjusted HR (95%CI)	Pr > ChiSq
ICU Admission		
Feb, 2020 - Apr, 2020	1 (ref)	
May, 2020 - Jul, 2020	0.67 (0.64-0.71)	<.0001
Aug, 2020 - Oct, 2020	0.58 (0.55-0.62)	<.0001
Nov, 2020 - Jan, 2021	0.51 (0.49-0.54)	<.0001
IMV		
Feb, 2020 - Apr, 2020	1 (ref)	
May, 2020 - Jul, 2020	0.56 (0.53-0.60)	<.0001
Aug, 2020 - Oct, 2020	0.49 (0.45-0.52)	<.0001
Nov, 2020 - Jan, 2021	0.43 (0.40-0.45)	<.0001
ARDS/Respiratory Failure		
Feb, 2020 - Apr, 2020	1 (ref)	
May, 2020 - Jul, 2020	0.78 (0.76-0.81)	<.0001
Aug, 2020 - Oct, 2020	0.83 (0.80-0.87)	<.0001
Nov, 2020 - Jan, 2021	0.84 (0.81-0.87)	<.0001
All-Cause Mortality		
Feb, 2020 - Apr, 2020	1 (ref)	
May, 2020 - Jul, 2020	0.87 (0.80-0.96)	0.0056
Aug, 2020 - Oct, 2020	0.75 (0.67-0.83)	<.0001
Nov, 2020 - Jan, 2021	0.80 (0.73-0.88)	<.0001

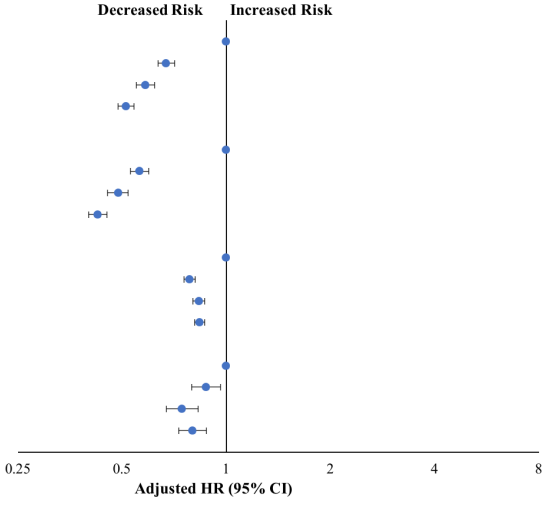


Figure 2 The forest plots of adjusted HRs among US hospitalised adult patients with COVID-19 by gender, by age group, and by time period. Three separate Fine and Gray models were developed stratified by age group, gender and time period; each model used the other two main categories of risk exposure (eg, gender and calendar month in the age group model) as covariates. The other covariates in each model were race, ethnicity, US region, hypertension, diabetes (type I and II), active cancer, coronary artery disease, congestive heart failure, chronic kidney disease, chronic obstructive pulmonary disease, asthma and obstructive sleep apnoea. ARDS, acute respiratory distress syndrome; ICU, intensive care unit; IMV, invasive mechanical ventilation.

suggesting elevated risk throughout the course of disease among older patients. Similar trends were observed in ex-US patients for ARDS/respiratory failure and all-cause mortality, while the risk of IMV was similar among patients aged 18–49 and aged 65 or over. These results were in accordance with prior reports from studies across the globe.[6 20–23] As shown in online supplemental eTable 2, elderly patients had a higher prevalence of pre-existing comorbidities, which could have undermined patients' ability to fight against infections.[24] In addition, Takahashi *et al*[25] found that declined T-cell response is associated with increased age, which could lead to lower efficacy in viral clearance and a higher likelihood of inflammatory cytokine storm, resulting in poor health outcomes.

Men had an approximate 20%–41% increased risk of all-cause mortality, ARDS/respiratory failure, IMV and

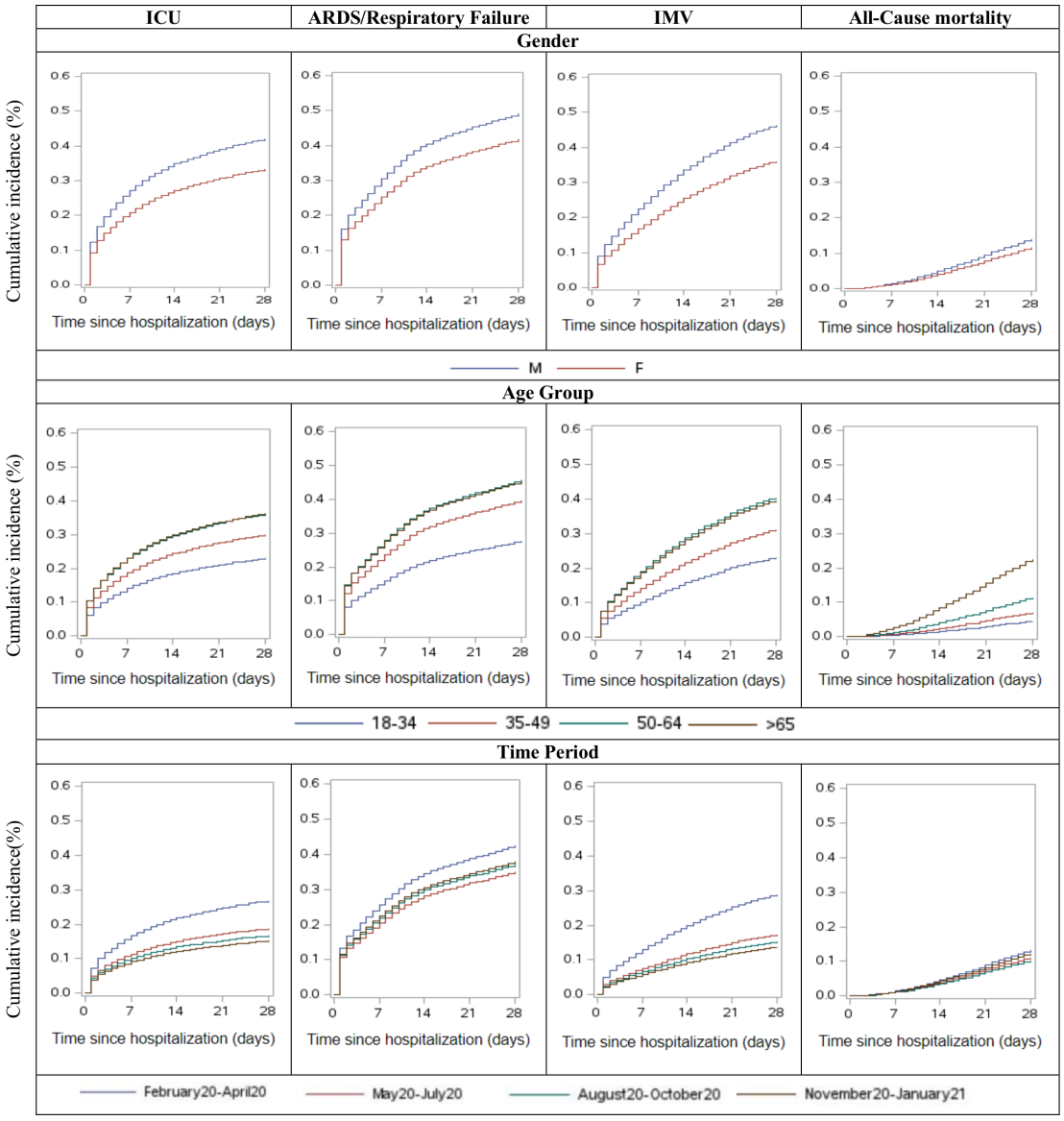


Figure 3 Regression-adjusted cumulative incidence plots of selected outcomes among US hospitalised adult patients with COVID-19 by gender, by age group and by time period. Three separate Fine and Gray models were developed stratified by age group, gender and time period; each model used the other two main categories of risk exposure (eg, gender and calendar month in the age group model) as covariates. The other covariates in each model were race, ethnicity, US region, hypertension, diabetes (type I and II), active cancer, coronary artery disease, congestive heart failure, chronic kidney disease, chronic obstructive pulmonary disease, asthma and obstructive sleep apnoea. ARDS, acute respiratory distress syndrome; ICU, intensive care unit; IMV, invasive mechanical ventilation.

ICU admissions compared with women. The increased risk of these clinical outcomes remained consistently higher in men over 7–28 days postindex suggesting worse disease prognosis in men. These results were also in line with prior reports.[6 21 26–31] For example, Fried *et al* showed increased mortality and morbidity associated with older age and male sex using medical claims data.[6] Palaiodimos *et al*, reported increasing age and male sex were independently associated with worse in-hospital outcomes.[21] However, Fried's study only covered the period from February to April 2020 while the study by Palaiodimos *et al* only included 200 patients from one medical centre in New York, USA. Our study included longitudinal data from thousands of patients lives in the USA and outside of the USA, thereby providing further support for the observation that gender seemed to be a differential factor in adverse clinical outcomes. Studies from national statistical agencies across England and Wales, France, Germany, Italy, Netherlands, Portugal, Korea, and Spain also observed that men had twice the risk of death from COVID-19 using data collated by the National Institute for Demographic Studies.[20] Many factors might contribute to the observed gender difference in adverse clinical outcomes. For example, Takahashi *et al* showed that immune responses to SARS-CoV-2 differed between men and women, possibly resulting from men having higher plasma levels of innate immune cytokines and poorer T cell activation.[25] Protein ACE2, a key protein

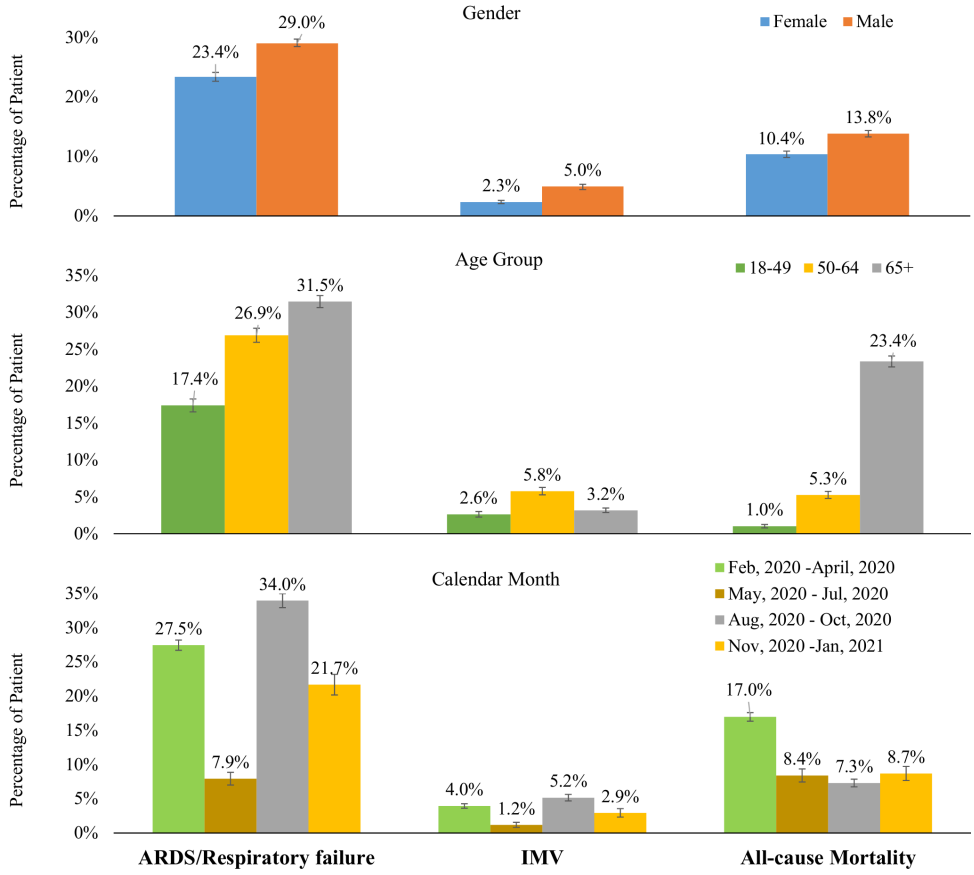


Figure 4 The selected outcomes among ex-US hospitalised adult patients with COVID-19 by gender, by age group and by time period. Error bars reflect SEs. ARDS, acute respiratory distress syndrome; IMV, invasive mechanical ventilation.

involved in the entry of SARS-CoV-2 into cell and the protection from lung injury, was also higher in women than men.[32] Moreover, women were more likely to follow healthy behaviours, such as better hygiene practices and lower level of smoking and alcohol use, which could also explain better outcomes in women.[33 34]

Results of this study found a significantly higher cumulative rate and aHR for ARDS/respiratory failure and IMV among patients ≥65 years and men, in line with prior reports of hospitalised COVID-19 patients showing greater risk among these demographic subgroups.[35 36] Analysis of the trends of ICU admissions in the USA and all-cause mortality over time periods in both USA and ex-USA revealed that patients with index date in February–April 2020 were at increased risk of death compared with patients diagnosed in the later time period. This could potentially be due to healthcare facilities being overwhelmed with the volume of COVID-19 admissions at the onset of the pandemic in some areas. In addition, limited information about COVID-19, and lack of experience in the clinical management of the disease leading to a trial and error approach using different pharmaceutical and non-pharmaceutical interventions to manage disease progression and spread, and inexperience in resource management of the healthcare staff in the early stage of COVID-19 pandemic[37–39] could have resulted in poorer outcomes among patients diagnosed at the beginning of the pandemic. Besides, the lack of SARS-CoV-2 diagnosis kits in the early pandemic could play an important role. Thus, the diagnosis kits were likely to have been reserved for the sickest patients, with the consequence that rates of ICU admission, ARDS/respiratory failure, IMV and death will appear higher at earlier time points. It was also possible that these trends were confounded by other unobserved patient-related or environmental risk factors.

This study was not meant to compare the US and ex-US cohorts; however, there were some differences between ARDS/respiratory failure and IMV in temporal trend. For example, ex-US patients seemed to have a much lower ARDS/respiratory failure and IMV in May–July 2020, compared with the US cohort. Additionally, ARDS/respiratory failure and IMV proportion in November 2020–January 2021 were much lower compared with August–October 2020 in the ex-USA cohort; while the proportions of these two outcomes during these two periods were closer in the US cohort. This was very likely due to different non-pharmaceutical interventions for the COVID-19 pandemic (eg, lockdown policies) implemented by each country within the ex-USA cohorts as well as different response strategies each country's HCO took. In addition, the baseline characteristics of the US and ex-US cohorts differed, which might reflect the differences in the underlying demographic distribution and healthcare systems across countries. Of note, the

prevalence of some comorbidities in the ex-US cohort was lower than the prevalence in previous studies conducted in non-US countries. For example, in the current study, the prevalence of hypertension was 13.1% in the ex-US cohort, whereas a recent meta-analysis of COVID-19 studies showed that the prevalence of hypertension was 39.5% in China, 35.9% in Italy and 27.8% in the UK.[40] The difference might be due to different compositions of the patient populations across studies. For example, the ex-US cohort in the current study had 55.1% men, while the studies included in the meta-analysis had a male to female ratio of 1.57 on average. Moreover, the current study had a longer follow-up than previous studies, which might also explain different patient characteristics across studies.

Our study has certain limitations. First, there is often a trade-off between sensitivity and specificity in defining a representative real-world cohort of patients with COVID-19. We performed a sensitivity analysis using only patients with COVID-19-specific clinical diagnosis, and patients with lab confirmation of disease. The results from our sensitivity analyses were similar to those from our primary analyses, suggesting that the potential effect of misclassification was minimal and did not impact the study conclusion. Second, compared with claims data, the EMR database may provide more timely, detailed, and accurate patient health information, but it only reflects the patient experience in the participating healthcare systems within the research network,[41 42] and the information from medical encounters before the COVID-19 hospitalisation or with other doctors/providers outside the research network is not captured.[42] Third, among deceased patients, as the death information was available at a monthly level, the patients' last physically present date in the database was used to infer the death date, but this measurement error should be non-differential and bias the results towards null. Fourth, for the ex-US analysis, the statistical analysis has been built into the platform and does not allow customisation, which has limited the ability to calculate statistics such as SE, CI and model the risk. Moreover, due to data privacy, the number of COVID-19 patients contributed by each country could not be specified. Finally, though COVID-19 patients included in this study were from multiple participating HCOs in the USA and outside of the USA, it may not fully represent the wider population across the globe.

CONCLUSIONS

Temporal trends in adverse clinical outcomes among patients with COVID-19 from this multinational EMR database comprising of patients from diverse demographic backgrounds suggest that older age, male gender and diagnosis in earlier months of the pandemic conferred greater risk for ICU admissions, ARDS/respiratory failure, IMV and all-cause mortality over 7–28 days postindex hospital admission. This evidence may be helpful in identifying patients at greater risk of these adverse clinical events, and in so doing, inform clinical interventions and increase public awareness.

Author affiliations
[1]Center for Observational Research and Data Science, Bristol-Myers Squibb, Princeton, New Jersey, USA
[2]Data Sciences Institute, Takeda Pharmaceutical Company Limited, Cambridge, Massachusetts, USA
[3]Center for Observational Research, Amgen, Inc, Thousand Oaks, California, USA
[4]Research & Development Strategy & Operations, Amgen, Inc, Thousand Oaks, California, USA
[5]Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina, USA

Acknowledgements We would like to acknowledge the support from David Anstatt and Marisa Co at Bristol Myers Squibb, as well as the COVID R&D Alliance members at Amgen Inc., Takeda Pharmaceutical Company Limited, and McKinsey & Company.

Contributors All authors (YB, JZ, ZW, MS, XC, QX, JJ, OA, BDB, CB, CAB, AC, CWC, GD, VG, KG, AAL, JM, MJ-F, HAK, SM, KR, LS, PY, JZ and LZ) contributed to the conception and design of the work; the acquisition, analysis or interpretation of data; drafting the manuscript or revising it critically. All authors approved the final manuscript as submitted and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding This study was funded by Bristol-Myers Squibb Co.; award/Grant number is not applicable.

Disclaimer This work used existing de-identified data which does not constitute research with human subjects because there is no interaction with any individual and no identifiable private information were used. No ethical approval was considered necessary.

Competing interests YB, XC, JJ, QX, MS, ZW and JZ are employees of Bristol Myers Squibb and hold stock or stock options at Bristol Myers Squibb. BDB, CB, CAB, AC, CWC, AAL, JM, JZ are employees and stockholders of Amgen. OA, GD, VG, KG, HAK, SM, KR, LS, PY and LZ are employees and stockholders in Takeda, Pharmaceutical Company Limited. GlaxoSmithKline (GSK), Takeda, AbbVie, Boehringer Ingelheim and UCB Bioscience (UCB) have collaborative agreements with the Center for Pharmacoepidemiology, Department of Epidemiology and University of North Carolina at Chapel Hill which provides salary support to MJ-F as Director. MJ-F is a member of the Scientific Steering Committee (SSC) for a postapproval safety study funded by GSK. All compensation for services provided on the SSC is invoiced by and paid to UNC Chapel Hill.

Patient consent for publication Not required.

Ethics approval This study was approved by Observational Protocol Review Committee at Bristol-Myers Squibb Co.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Data for this study are not publicly available; access to which is restricted on commercial terms. Please contact authors for information on data availability.

ORCID iDs
Julia Zhu http://orcid.org/0000-0001-8210-8389

Ajit A Londhe http://orcid.org/0000-0002-0338-2919
Junjie Ma http://orcid.org/0000-0002-6249-876X

REFERENCES

1 Tian J, Yuan X, Xiao J, *et al*. Clinical characteristics and risk factors associated with COVID-19 disease severity in patients with cancer in Wuhan, China: a multicentre, retrospective, cohort study. Lancet Oncol 2020;21:893–903.
2 Zhang Y, Geng X, Tan Y, *et al*. New understanding of the damage of SARS-CoV-2 infection outside the respiratory system. *Biomed* Pharmacother 2020;127:110195.
3 World Health Organization. Coronavirus disease (COVID-19), 2020.
4 World Health Organization. Archived: who timeline-covid-19. Viitattu, 2020.
5 Johns Hopkins Coronavirus Resource Center. COVID-19 case Tracker, 2020.
6 Fried MW, Crawford JM, Mospan AR, *et al*. Patient characteristics and outcomes of 11 721 patients with coronavirus disease 2019 (COVID-19) hospitalized across the United States. Clin Infect Dis 2021;72:e558–65.
7 Holshue ML, DeBolt C, Lindquist S, *et al*. First case of 2019 novel coronavirus in the United States. New England Journal of Medicine 2020;382:929–36.
8 Surveillances V. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020. China CDC Weekly 2020;2:113–22.
9 Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? The Lanc*et* 2020;395:1225–8.
10 White C, Nafilyan V. Coronavirus (COVID-19) related deaths by ethnic group, England and Wales: 2 March 2020 to 10 April 2020. Office for National Statistics, 2020.
11 Gold JAW, Rossen LM, Ahmad FB, *et al*. Race, ethnicity, and age trends in persons who died from COVID-19 — United States, May–August 2020. MMWR Morb Mortal Wkly Rep 2020;69:1517–21.
12 Stacey J, Mehta M. Using EHR data extraction to streamline the clinical trial process. Clinical Researcher 2017;4:2–7.
13 Singh S, Khan A. Clinical characteristics and outcomes of coronavirus disease 2019 among patients with preexisting liver disease in the United States: a multicenter research network study. Gastroenterology 2020;159:e763:768–71.
14 Taquet M, Luciano S, Geddes JR, *et al*. Bidirectional associations between COVID-19 and psychiatric disorder: retrospective cohort studies of 62 354 COVID-19 cases in the USA. Lancet Psychiatry 2021;8:130–40.
15 Topaloglu U, Palchuk MB. Using a Federated network of real-world data to optimize clinical trials operations. JCO Clin Cancer Inform 2018;2:1–10.
16 Jorge A, D'Silva KM, Cohen A, *et al*. Temporal trends in severe COVID-19 outcomes in patients with rheumatic disease: a cohort study. Lancet Rheumatol 2021;3:e137.
17 Harrison SL, Fazio-Eynullayeva E, Lane DA, *et al*. Comorbidities associated with mortality in 31,461 adults with COVID-19 in the United States: a federated electronic medical record analysis. PLoS Med 2020;17:e1003321.
18 Fine JP, Gray RJ. A proportional hazards model for the Subdistribution of a competing risk. J Am Stat Assoc 1999;94:496–509.
19 Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. Am J Epidemiol 2009;170:244–56.
20 Bhopal SS, Bhopal R. Sex differential in COVID-19 mortality varies markedly by age. The Lancet 2020;396:532–3.
21 Palaiodimos L, Kokkinidis DG, Li W, *et al*. Severe obesity, increasing age and male sex are independently associated with worse in-hospital outcomes, and higher in-hospital mortality, in a cohort of patients with COVID-19 in the Bronx, New York. Metabolism 2020;108:154262.
22 Wang L, He W, Yu X, *et al*. Coronavirus disease 2019 in elderly patients: characteristics and prognostic factors based on 4-week follow-up. J Infect 2020;80:639–45.
23 Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020;323:1239–42.
24 Bonanad C, García-Blas S, Tarazona-Santabalbina FJ, *et al*. Coronavirus: the geriatric emergency of 2020. joint document of the section on geriatric cardiology of the Spanish Society of cardiology and the Spanish Society of geriatrics and gerontology. Revista Española de Cardiología 2020;73:569–76.
25 Takahashi T, Ellingson MK, Wong P, *et al*. Sex differences in immune responses that underlie COVID-19 disease outcomes. Nature 2020;588:315–20.
26 Ahmed SB, Dumanski SM. Sex, gender and COVID-19: a call to action. Can J Public Health 2020;111:980–3.
27 The Lancet. The gendered dimensions of COVID-19. Lancet;395:11–17.
28 Dehingia N, Raj A. Sex differences in COVID-19 case fatality: do we know enough? Lancet Glob Health 2021;9:e14–15.
29 Alkhouli M, Nanjundappa A, Annie F, *et al*. Sex differences in case fatality rate of COVID-19: insights from a multinational registry. Mayo Clin Proc 2020;95:1613–20.
30 Voinsky I, Baristaite G, Gurwitz D. Effects of age and sex on recovery from COVID-19: analysis of 5769 Israeli patients. J Infect 2020;81:e102–3.
31 Richardson S, Hirsch JS, Narasimhan M, *et al*. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the new York City area. JAMA 2020;323:2052–9.
32 Ni W, Yang X, Yang D, *et al*. Role of angiotensin-converting enzyme 2 (ACE2) in COVID-19. Crit Care 2020;24:1–10.
33 Griffith DM, Sharma G, Holliday CS. A biopsychosocial approach to understanding sex differences in mortality and recommendations for practice and policy interventions, 2020.
34 Griffith DM, Sharma G, Holliday CS, *et al*. Men and COVID-19: a biopsychosocial approach to understanding sex differences in mortality and recommendations for practice and policy interventions. Prev Chronic Dis 2020;17:E63.
35 Grasselli G, Tonetti T, Protti A, *et al*. Pathophysiology of COVID-19-associated acute respiratory distress syndrome: a multicentre prospective observational study. Lancet Respir Med 2020;8:1201–8.
36 Ferrando C, Suarez-Sipmann F, Mellado-Artigas R, *et al*. Clinical features, ventilatory management, and outcome of ARDS caused by COVID-19 are similar to other causes of ARDS. Intensive Care Med 2020;46:2200–11.
37 Mannelli C. Whose life to save? scarce resources allocation in the COVID-19 outbreak. J Med Ethics 2020;46:364–6.
38 Ji Y, Ma Z, Peppelenbosch MP, *et al*. Potential association between COVID-19 mortality and health-care resource availability. Lancet Glob Health 2020;8:e480.
39 Lee CCM, Thampi S, Lewin B, *et al*. Battling COVID-19: critical care and peri-operative healthcare resource management strategies in a tertiary academic medical centre in Singapore. Anaesthesia 2020;75:861–71.
40 Bajgain KT, Badal S, Bajgain BB, *et al*. Prevalence of comorbidities among individuals with COVID-19: a rapid review of current literature. Am J Infect Control 2021;49:238–46.
41 Richesson RL, Horvath MM, Rusincovitch SA. Clinical research informatics and electronic health record data. Yearb Med Inform 2014;9:215–23.
42 Wilson J, Bock A. The benefit of using both claims data and electronic medical record data in health care analysis. Optum Insight 2012:1–4.