Association between HIV infection and clinical spectrum of COVID-19: a population level analysis based on US National COVID Cohort Collaborative (N3C) data

Xueying Yang*, Jing Sun*, Rena C Patel, Jiajia Zhang, Siyuan Guo, Qulu Zheng, Amy L Olex, Bankole Olatosi, Sharon B Weissman, Jessica Y Islam,

**South Carolina Smart State
Center for Healthcare Quality,**
(X Yang PhD, Prof J Zhang PhD,
S Guo MS, B Olatosi PhD,
Prof S B Weissman MD,
Prof X Li PhD), **Department of
Health Promotion, Education,
and Behavior,** (X Yang, Prof X Li),
**Department of Epidemiology
and Biostatistics** (Prof J Zhang,
S Guo), **and Department of
Health Services Policy and
Management** (B Olatosi),
**Arnold School of Public Health,
University of South Carolina,
Columbia, SC, USA; Department
of Internal Medicine, School of
Medicine** (Prof S B Weissman)
**and Big Data Health Science
Center** (X Yang, Prof J Zhang,
S Guo, B Olatosi,
Prof S B Weissman, Prof X Li),
**University of South Carolina,
Columbia, SC, USA; Department
of Epidemiology** (J Sun PhD,
Q Zheng PhD, Prof G D Kirk MD)
**and School of Medicine, Public
Health, and Nursing**
(C G Chute MD), **Johns Hopkins
Bloomberg School of Public
Health, Baltimore, MD, USA;
Department of Medicine**
(R C Patel MD) **and Department
of Global Health** (R C Patel),
**University of Washington,
Seattle, WA, USA; Wright
Center for Clinical and
Translational Research, Virginia
Commonwealth University,
Richmond, VA, USA**
(A L Olex MS); **Center for
Immunization and Infection in
Cancer, Cancer Epidemiology
Program, H Lee Moffitt Cancer
Center and Research Institute,
Tampa, FL, USA** (J Y Islam PhD);
**Center for Health AI, University
of Colorado Anschutz Medical
Campus, Aurora, CO, USA**
(M Haendel PhD)

Correspondence to:
Dr Xueying Yang PhD,
South Carolina SmartState
Center for Healthcare Quality,
Department of Health
Promotion, Education, and
Behavior, Arnold School of Public
Health, University of
South Carolina, Columbia,
SC 29208, USA
yang@mailbox.sc.edu

Research in context

Evidence before this study
We searched PubMed on June 3, 2021, for population-based epidemiological studies comparing risk of severe COVID-19 outcomes between people living with and without HIV. The search strategy "HIV" OR "AIDS" AND "COVID-19" OR "coronavirus" OR "SARS-CoV-2" was used, with results being filtered to articles from the past year and with abstracts available, no language restrictions were applied. We identified 140 papers for screening. One UK study that was included as the comparison of interest reported that people with HIV seemed to be at an increased risk of COVID-19 mortality than those without. Another relevant study identified on the medRxiv preprint server found a higher risk of COVID-19 mortality among people with HIV than in the general population in Western Cape, South Africa.

Added value of this study
We used the US National COVID Cohort Collaborative data from more than 5·8 million people from 54 clinical sites to compare risks of the full spectrum of COVID-19 outcomes (ie, disease severity, hospitalisation, death) between people with HIV and people without. Those with HIV were at elevated odds of COVID-19 hospitalisation and mortality compared with people

without HIV, even after accounting for demographic characteristics, lifestyle factors, and comorbidities. The association was more evident among people of an older age, males, and of Black or African American ethnicity. People with HIV with more pronounced immunodeficiency (low CD4 cell counts) might have more severe clinical course than people without.

Implications of all the available evidence
People with HIV in the USA, particularly those with pronounced immunodeficiency, seem to be at elevated risk of COVID-19 hospitalisation and mortality. The exacerbated COVID-19 outcomes in Black or African American and male people with HIV suggest profound health inequities faced during the COVID-19 pandemic. The robust risk assessment of this study could inform prioritisation of prevention messaging, disease monitoring and therapies, and vaccination for people with HIV, especially those with more pronounced immunodeficiency. Given the COVID-19 pandemic's exacerbating effects on health inequities, public health and clinical communities must strengthen services and support to prevent aggravated COVID-19 outcomes among people with HIV, particularly for those with pronounced immunodeficiency.

Africa suggested that people with HIV had a higher risk (more than double) of COVID-19 mortality than people without HIV, although different factors were adjusted in the different studies.[11,12] A prospective study of patients hospitalised with COVID-19 showed an increased 28-day mortality in people with HIV after adjusting for age.[13] One study in New York, USA, reported a standardised in-hospital mortality ratio of 1·23 for HIV patients.[14] Prognosis, according to HIV immune status, is also difficult to evaluate because most studies from Europe and the USA reported on individuals with overall high CD4 cell counts.[15] In the largest published cohorts, the potentially higher risk for poorer COVID-19 outcomes were observed in people with HIV with lower CD4 cell counts.[8,16] Investigating whether people with virologically controlled HIV who are clinically stable will have a greater risk for COVID-19 complications than people without HIV is of great clinical significance.

Nevertheless, the evidence linking HIV status and COVID-19 outcomes is still scarce and some knowledge gaps remain. Several studies were based on only a small number of cases; some either did not have direct comparative data for people without HIV and HIV markers,[5] or focused only on hospitalised patients.[7–9] A large, multicentre, representative clinical dataset is needed to provide timely and robust risk assessment and thereby inform prioritisation of critical therapies, vaccination, and targeted intervention. Using the US National COVID Cohort Collaborative (N3C) data, this study aims to understand the role of HIV infection and levels of immunity

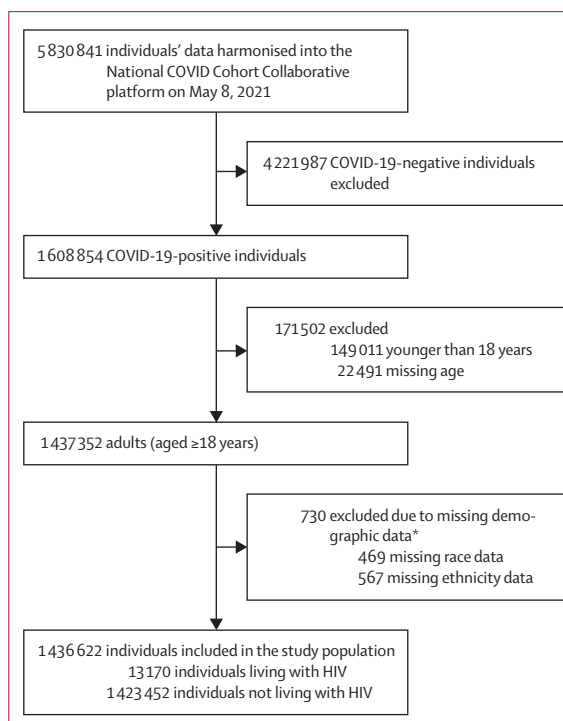affecting the COVID-19 clinical outcomes (ie, disease severity, hospitalisation, and mortality).

Figure 1: Study participant selection
*The numbers for missing race data and missing ethnicity data did not add up to 730 because some records missed both data.

Methods

Study design and population

The N3C Enclave, sponsored by multiple institutes of the US National Institutes of Health,[17] is the largest cohort of US COVID-19 cases and representative controls to date. The N3C is a large, multicentre dataset updated on an ongoing basis that harmonises electronic health records data for all individuals with laboratory confirmed, suspected, or possible COVID-19 during any encounter after Jan 1, 2020.[18] Control cases are those individuals who have tested negative for COVID-19, and are demographically matched on age group, sex, race, and ethnicity within the same submitting health-care system at a case

to control ratio of 1 to 2.[18] All patients in the N3C Enclave include historical data within the same health-care system as of Jan 1, 2018, which provides information about pre-existing health conditions (eg, comorbidities) and other medical history (look back data).[19] We included all adult (aged ≥18 years) COVID-19 cases from 54 clinical sites across the USA with data being deposited into the N3C and harmonised into a data release set from Jan 1, 2020, to May 8, 2021. The data ingestion and harmonisation process are described in the appendix (p 2). We excluded people with missing age, race, and ethnicity data because absence of data on these key variables probably indicated that poor data quality for these records.

The N3C data transfer to the National Center for Advancing Translational Sciences (NCATS) was done under a Johns Hopkins University Reliance Protocol (IRB00249128) or individual site agreements with the US National Institutes of Health (NIH). An institutional data use agreement was signed between the University of South Carolina and NCATS N3C Data Enclave. The N3C Data Enclave is managed under the authority of the NIH. The N3C Data Enclave is approved under the authority of the NIH Review Board. The analyses reported in this Article were approved separately by the institutional review board of University of South Carolina (Pro00107403) with data access. The NIH's N3C data access committee approved the data use request for this project (RP-E72986).

Procedures

The N3C phenotype[18] is designed to be inclusive of any diagnosis codes, procedure codes, laboratory tests, or combination thereof that might be indicative of COVID-19 (eg, Centers for Disease Control and Prevention coding guidance[20]). N3C includes patients with any encounter after Jan 1, 2020, who have either one or more of a set of a priori-defined SARS-CoV-2 laboratory tests with a positive result; or one or more strong positive diagnostic codes from the International Classification of Diseases (ICD) 10 or SNOMED tables; or two or more weak positive diagnostic codes from the ICD-10 or SNOMED tables. The cohort definition is publicly available on GitHub.[18]

N3C harmonises data across four clinical data models (ACT Network, PCORnet, Observational Health Data Sciences and Informatics, and TriNetX) and provides a unified analytical platform in which data are encoded by use of the Observational Medical Outcomes Partnership (OMOP) version 5.3.1.[21] The concept sets in OMOP[22,23] are a list of concepts from the standardised vocabulary that, taken together, describe a topic of interest for a study, were used to identify each clinical concept (eg, laboratory measure, conditions, or medication). Data domains extracted by N3C include demographics, encounter details, medications, diagnoses, procedures, vital signs, laboratory results, procedures, and social history. Specific variables included in each domain are listed in each model's

	Overall (n=1 436 622)	People with HIV (n=13 170)	People without HIV (n=1 423 452)	p value
Social demographics				
Age, years	47 (32–61)	49 (36–60)	47 (32–61)	..
18–49	770 099 (53·60%)	6703 (50·90%)	763 396 (53·63%)	<0·0001
50–64	371 489 (25·86%)	4533 (34·42%)	366 956 (25·78%)	..
≥65	295 034 (20·54%)	1934 (14·68%)	293 100 (20·59%)	..
Sex*				
Male	645 956 (44·96%)	9641 (73·20%)	636 315 (44·70%)	<0·0001
Female	789 148 (54·93%)	3521 (26·74%)	785 627 (55·19%)	..
Race				
Black or African American	202 947 (14·13%)	4092 (31·07%)	198 855 (13·97%)	<0·0001
White	853 997 (59·44%)	6013 (45·66%)	847 984 (59·57%)	..
Asian, other, or unknown	379 678 (26·43%)	3065 (23·27%)	376 613 (26·46%)	..
Ethnicity				
Hispanic or Latinx	213 205 (14·84%)	2227 (16·91%)	210 978 (14·82%)	<0·0001
Not Hispanic or Latinx	1 001 390 (69·70%)	9479 (71·97%)	991 911 (69·68%)	..
Other or unknown	222 027 (15·45%)	1464 (11·12%)	220 563 (15·49%)	..
Comorbidities				
Diabetes	233 838 (16·28%)	3066 (23·28%)	230 772 (16·21%)	<0·0001
Renal disease	86 803 (6·04%)	1758 (13·35%)	85 045 (5·97%)	<0·0001
Congestive heart failure	75 461 (5·25%)	1026 (7·79%)	74 435 (5·23%)	<0·0001
Chronic pulmonary disease	200 998 (13·99%)	2974 (22·58%)	198 024 (13·91%)	<0·0001
Peripheral vascular disease	72 808 (5·07%)	979 (7·43%)	71 829 (5·05%)	<0·0001
Stroke	68 171 (4·75%)	878 (6·67%)	67 293 (4·73%)	<0·0001
Cancer	78 909 (5·49%)	1205 (9·15%)	77 704 (5·46%)	<0·0001
Dementia	24 622 (1·71%)	251 (1·91%)	24 371 (1·71%)	0·095
Myocardial infarction	42 165 (2·94%)	693 (5·26%)	41 472 (2·91%)	<0·0001
Liver disease	72 701 (5·06%)	2152 (16·34%)	70 549 (4·96%)	<0·0001
Rheumatological disease	46 861 (3·26%)	501 (3·80%)	46 360 (3·26%)	0·0005
Hemiplegia or paraplegia	11 656 (0·81%)	222 (1·69%)	11 434 (0·80%)	<0·0001
Peptic ulcer disease	14 237 (0·99%)	230 (1·75%)	14 007 (0·98%)	<0·0001
Lifestyle factors				
Body-mass index, kg/m²				
>30	218 159 (15·19%)	2469 (18·75%)	215 690 (15·15%)	<0·0001
≤30	295 370 (20·56%)	4560 (34·62%)	290 810 (20·43%)	..
Unknown	923 093 (64·25%)	6141 (46·63%)	916 952 (64·42%)	..
Smoking status				
Non-smoker	1 194 746 (83·16%)	9318 (70·75%)	1 185 428 (83·28%)	<0·0001
Current or former smoker	241 876 (16·84%)	3852 (29·25%)	238 024 (16·72%)	..
			(Table 1 continues on next page)	

	Death, OR (95% CI)*	Hospitalisation, OR* (95% CI)	Mild or moderate COVID-19 vs unaffected‡, OR (95% CI)	Severe COVID-19 vs unaffected‡, OR (95% CI)
Unadjusted model	1·90 (1·73–2·09)	1·78 (1·71–1·84)	0·55 (0·53–0·57)	1·52 (1·38–1·67)
Adjusted models				
Adjusted for age + sex + race + ethnicity	1·85 (1·67–2·04)	1·62 (1·56–1·69)	0·53 (0·51–0·55)	1·34 (1·21–1·48)
Adjusted for age + sex + race + ethnicity + smoking + BMI	1·76 (1·59–1·94)	1·48 (1·42–1·54)	0·59(0·57–0·61)	1·34 (1·21–1·47)
Adjusted for age + sex + race + ethnicity + smoking + BMI + comorbidities‖ + month of diagnosis	1·29 (1·16–1·44)	1·20 (1·15–1·26)	0·61 (0·59–0·64)	1·04 (0·94–1·16)
Adjusted for age + sex + race + ethnicity + smoking + BMI + CCI + month of diagnosis	1·43 (1·29–1·59)	1·28 (1·23–1·34)	0·61 (0·58–0·63)	1·15 (1·04–1·27)
Interaction models**				
Age and HIV status				
With HIV and aged 50–64 years vs aged 18–49 years	7·86 (5·86–10·53)	2·17 (1·93–2·44)	0·48 (0·49–0·61)	3·08 (2·30–4·13)
With HIV and aged ≥65 years vs aged 18–49 years	22·20 (16·91–29·12)	3·42 (2·98–3·93)	0·57 (0·50–0·64)	7·57 (5·73–9·99)
Without HIV and aged 50–64 years vs aged 18–49 years	4·23 (4·00–4·48)	1·43 (1·41–1·45)	0·81 (0·81–0·82)	2·32 (2·21–2·43)
Without HIV and aged ≥65 years vs aged 18–49 years	12·46 (11·82–13·13)	2·39 (2·36–2·42)	0·71 (0·7–0·72)	4·66 (4·45–4·87)
Sex and HIV status				
With HIV and male vs female	3·13 (2·29–4·28)	1·32 (1·20–1·44)	0·50 (0·47–0·53)	1·89 (1·44–2·48)
Without HIV and male vs female	1·54 (1·50–1·59)	1·16 (1·15–1·17)	1·13 (1·12–1·14)	1·75 (1·70–1·8)
Race and HIV status				
With HIV and Black or African American vs White	3·64 (2·59–5·11)	3·16 (2·85–3·49)	1·10 (1·001–1·2)	3·25 (2·4–4·41)
With HIV and Asian, other, or unknown vs White	4·32 (2·88–6·46)	2·85 (2·48–3·27)	0·79 (0·70–0·89)	3·63 (2·52–5·23)
Without HIV and Black or African American vs White	1·16 (1·12–1·20)	1·66 (1·64–1·68)	0·92 (0·91–0·93)	1·25 (1·21–1·3)
Without HIV and Asian, other, or unknown vs White	1·16 (1·11–1·20)	1·35 (1·33–1·37)	1·01 (1–1·02)	1·18 (1·13–1·22)
Ethnicity and HIV status				
With HIV and Hispanic or Latinx vs not Hispanic or Latinx	3·48 (2·25–5·37)	1·88 (1·62–2·19)	0·86 (0·76–0·97)	2·33 (1·55–3·50)
With HIV and other or unknown vs not Hispanic or Latinx	3·50 (2·20–5·59)	1·70 (1·42–2·03)	0·49 (0·42–0·57)	1·98 (1·28–3·05)
Without HIV and Hispanic or Latino vs not Hispanic or Latinx	1·07 (1·02–1·12)	1·26 (1·24–1·28)	1·09 (1·08–1·10)	1·38 (1·32–1·44)
Without HIV and other or unknown vs not Hispanic or Latinx	0·87 (0·84–0·91)	0·62 (0·61–0·63)	0·72 (0·71–0·73)	0·77 (0·74–0·81)
Stratified models, people with HIV vs people without HIV at each subgroup				
Age, years				
18–49 (n=770 099)	1·24 (0·93–1·67)	1·25 (1·17–1·34)	0·62 (0·59–0·66)	0·96 (0·75–1·22)
50–64 (n=371 489)	1·05 (0·88–1·26)	1·04 (0·97–1·12)	0·59 (0·55–0·62)	0·81 (0·68–0·96)
≥65 (n=295 034)	1·26 (1·06–1·50)	1·27 (1·18–1·36)	0·67 (0·61–0·74)	1·23 (1·06–1·43)
Sex				
Female (n=645 956)	1·90 (1·57–2·30)	1·77 (1·64–1·92)	0·93 (0·86–0·997)	1·80 (1·49–2·18)
Male (n=789 148)	1·14 (0·98–1·26)	1·01 (0·96–1·06)	0·52 (0·50–0·54)	0·84 (0·75–0·95)
Race				
White (n=853 997)	1·31 (1·11–1·54)	1·11 (1·03–1·18)	0·50 (0·47–0·53)	0·90 (0·76–1·07)
Black or African American (n=202 947)	1·26 (1·06–1·50)	1·27 (1·18–1·36)	0·84 (0·79–0·90)	1·18 (0·995–1·40)
Asian, other, or unknown (n=379 678)	1·27 (1·02–1·57)	1·19 (1·09–1·30)	0·59 (0·55–0·64)	1·15 (0·94–1·40)

BMI=body-mass index. CCI= Charlson Comorbidity Index. OR=odds ratio. *Results from hierarchical logistic regression. †Includes both the mild (outpatient, WHO severity 1–3) and mild emergency department (outpatient with emergency department visit, WHO severity ~3) categories. ‡Unaffected refers to the individuals with no laboratory test, a negative laboratory test, or suspected COVID-19 with laboratory tests, but identified by other diagnosis codes or procedure codes. §Results from multinomial regression with log-transformation. ¶Includes both severe (hospitalised with invasive ventilation or extracorporeal membrane oxygenation, WHO severity 7–9) and mortality or hospice (hospital mortality or discharge to hospice, WHO severity 10) categories based on WHO criterion. ‖The model adjusted simultaneously for the following comorbidities: hemiplegia or paraplegia, dementia, liver disease, myocardial infarction, congestive heart failure, chronic pulmonary disease, cancer, diabetes, stroke, peripheral vascular disease, rheumatologic disease, renal disease, and peptic ulcer disease. **Results for interaction terms adjusted demographic characteristics (age, sex, race, ethnicity), smoking, BMI, and each individual comorbidities listed in the previous footnote.

Table 2: Association between HIV status and COVID-19 clinical spectrum outcomes based on hierarchical logistic regression models

disease severity), we used hierarchical logistic regression analyses to estimate the association of HIV infection and HIV markers (CD4 cell count, viral load) with these COVID-19 outcomes. A subsample of people with HIV (n=1544) who had data for both CD4 cell count and viral load were available for analysing the association between HIV markers and COVID-19 outcomes. In step one we first adjusted for age (18–49,

50–64, and ≥65 years), sex (male, female), race (Black or African American; White; Asian or other or unknown) and ethnicity (Hispanic or Latinx; not Hispanic or Latinx; others or unknown). In step two we subsequently adjusted for smoking (non-smoker, current, or former smoker) and obesity (BMI >30 kg/m², ≤30 kg/m², or unknown). In step three we adjusted for an array of comorbidities (hemiplegia or paraplegia, dementia, liver disease, myocardial infarction, congestive heart failure, chronic pulmonary disease, cancer, diabetes, stroke, peripheral vascular disease, rheumatological disease, renal disease, and peptic ulcer disease) or CCI score, and month of COVID-19 diagnosis. For COVID-19 severity, multinomial logistic regressions were applied with unaffected COVID-19 individuals as a reference group. Specifically, the same aforementioned factors were included in multiple multinominal regressions (mild or moderate vs unaffected, severe vs unaffected). Because information from at least one health-care encounter is required to generate index encounter and define COVID-19 severity, we first conducted multinominal regressions among the full sample in the main analysis, with those cases without health-care encounter information grouped into the first category of CPS (ie, unaffected). Then we did a sensitivity analysis to exclude those without health-care encounter information in the multinomial analysis (ie, 1396795 [97·2%] of total sample). We did additional sensitivity analyses with people with and without HIV matched on age group, sex, and number of comorbidities using 1:2 and 1:4 ratios. Collinearity was checked by calculating variance inflation factors for each covariate listed in the adjusted models.

To investigate whether age, sex, race, and ethnicity could be potential effect modifiers of HIV status, we fitted interaction terms between age (18–49, 50–64, and ≥65 years), sex, race (White vs Black or African American, and White vs Asian, other, or unknown), ethnicity (not Hispanic or Latinx vs Hispanic or Latinx) and HIV status in all analyses (not all the interaction analysis results are shown due to space limitation but are available from the authors upon request). We fitted stratified models with HIV infection, demographics, lifestyle, and comorbidities on each selected subgroup (ie, male; female; aged 18–49 years; aged 50–64 years; aged ≥65 years; White; Black or African American; and Asian, other, or unknown). We implemented all analyses with SQL and R (version 3.5) and created reproducible pipelines in the Code workbook on N3C Data Enclave.

Role of the funding source

Results

In this population level analysis, of the 5830841 COVID-19 cases and controls harmonised into the N3C data release set from Jan 1, 2020, to May 8, 2021, a total of 1436622 adult individuals who were positive for COVID-19 were included in this study (figure 1).

Compared with people without HIV, those with HIV had a narrower age distribution overall (lower proportion aged ≥65 years) but the median age was 2 years older (49 vs 47 years); a greater proportion were males and people of Black or African American race. People with HIV had higher prevalence of all comorbidities, including diabetes, chronic pulmonary disease, and liver disease (table 1).

Among the 1436622 COVID-19 cases, 262331 (18·26%) were hospitalised and 26130 (1·82%) died. People with HIV disproportionately required more COVID-19 related hospitalisation than those without (28·28% vs 18·17% [table 1]). Crude odds ratios (ORs) of COVID-19 hospitalisation and death were both higher in people with HIV (table 2). The associations were both attenuated, but remained significant, after sequentially adjusting for demographics, lifestyle factors, comorbidities, and month of COVID-19 diagnosis (hospitalisation adjusted OR [aOR] 1·20, 95% CI 1·15–1·26; mortality 1·29, 1·16–1·44; table 2; figure 2; appendix pp 3–6).

Compared with people without HIV, those with HIV had a higher proportion of severe illness (3·61% vs 1·73%), but a lower proportion of mild or moderate illness (47·15% vs 62·47%; table 1). Using unaffected COVID-19 individuals as a reference group in multinomial regression, people with HIV had lower odds of presenting with mild or moderate illness than people without HIV even after adjusting for all the covariates (aOR 0·61, 95% CI 0·59–0·64); by contrast, the odds of severe COVID-19 were comparable after sequential adjustments for all the covariates (1·04, 0·94–1·16; table 2; figure 2; appendix pp 7–8). In the sensitivity analysis (excluding individuals without health-care encounter information), the results were similar to the findings in the models with the full sample (appendix pp 9–10). The results from additional sensitivity analyses among the subsample of 1:2 and 1:4 matched people with and people without HIV were similar to the findings in the primary analyses (appendix p 11). Among 1544 people with HIV with both CD4 cell count and viral load data, a lower CD4 cell count (<200 cells per µL) was positively associated with all the adverse COVID-19 outcomes (ie, disease severity, hospitalisation, mortality) after adjusting for all the covariates, while viral suppression was only negatively associated with hospitalisation (table 3).

The interaction effect of age and HIV status suggested that the ageing process in people with HIV exacerbated all the adverse outcomes of COVID-19. Those with HIV in the older age groups had much higher odds of death and hospitalisation than those without HIV in the same age range. As another potential modifier, male sex could

Unadjusted

Adjusted for age + sex + race + ethnicity

Adjusted for age + sex + race + ethnicity + smoking + BMI

Adjusted for age + sex + race + ethnicity + smoking + BMI + comorbidities + month of diagnosis

Adjusted for age + sex + race + ethnicity + smoking + BMI + CCI + month of diagnosis

–

also interact with HIV infection in increasing the odds of severe clinical outcomes of COVID-19, yet with a smaller magnitude. Similar results were found in the interaction of race or ethnicity and HIV status, by which Black or African American race and Hispanic or Latinx ethnicity interacted with HIV infection in developing higher odds of adverse COVID-19 outcomes. Stratified models revealed that the elevated odds were higher in the similar subgroups (eg, older age, male sex, and Black or African American race; table 2; appendix pp 12–13).

To adjust the role of cumulative burden of comorbidities in the model development, CCI was considered in all adjusted models. However, a high collinearity was detected between CCI and the other covariates (variance inflation factor=7·97). Therefore, additional models were developed to include CCI (replacing individual comorbid conditions) in the analyses. Findings from the two sets of adjusted models (adjusting individual comorbid condition vs adjusting CCI) were similar (appendix pp 14–16).

Discussion

Our population-level analysis from N3C data found that people with HIV might not be disproportionately vulnerable to SARS-CoV-2 infection but are more likely to be hospitalised and die from COVID-19, although such risk might be attenuated when other confounding factors are taken into consideration. The associations between HIV and these outcomes seem particularly pronounced among older people, males, Black or African American adults, and Hispanic or Latinx adults. Among people with HIV, we find that the risks for poor COVID-19 outcomes are much higher among those with lower CD4 cell counts (<200 cells per μL) and an association between viral suppression and the COVID-19 outcome of hospitalisation.

To the best of our knowledge, this is the largest population-level analysis to investigate the role of HIV infection in COVID-19 clinical spectrum across the USA. Our results show a smaller but consistent effect of HIV infection on COVID-19 related mortality with large population-based cohort studies from South Africa[11] and the UK.[12] The differences of effect size between these

three studies could possibly be explained by the different sample characteristics. Results from the interaction effects illustrate that the adverse COVID-19 outcomes among people with HIV might be explained by the overlapping demographic (eg, male and African American) and co-morbidity characteristics (eg, a significant interaction effect of HIV and CCI, and data not shown but available upon request) that are highly prevalent in this population. Our study shows people with HIV require more COVID-19 hospitalisation, at a level of risk similar to a recent New York study[14] and other USA studies using TriNETX network data, which controlled for BMI and various comorbidities.[6,26]

Regarding the clinical severity of COVID-19, people with HIV are less likely to have mild illness, but more likely to have severe outcomes when only adjusting for demographics and lifestyle factors. The adjustment for comorbidities obviates the estimated risk of severe outcomes among people with HIV. This finding suggests

Figure 2: Estimates for the association between HIV status and COVID-19 clinical spectrum outcomes

(A) Mild or moderate disease. (B) Severe disease. (C) Hospitalisation. (D) Death. All stratified models (by age, sex, and race) were adjusted for age, sex, race, ethnicity, smoking, BMI, and comorbidities, including hemiplegia or paraplegia, dementia, liver disease, myocardial infarction, congestive heart failure, chronic pulmonary disease, cancer, diabetes, stroke, peripheral vascular disease, rheumatologic disease, renal disease, and peptic ulcer disease. Mild COVID-19 includes both the mild (outpatient, WHO severity 1–3) and mild emergency department (outpatient with emergency department visit, WHO severity ~3) categories. Moderate COVID-19 includes patients who were hospitalised but without invasive ventilation (WHO severity 4–6). Severe COVID-19 includes both severe (hospitalised with invasive ventilation or extracorporeal membrane oxygenation, WHO severity 7–9) and mortality or hospice (hospital mortality or discharge to hospice, WHO severity 10) categories based on WHO criterion. BMI=body-mass index.

	Death, OR (95% CI)*	Hospitalisation, OR (95% CI)*	Mild or moderate vs unaffected, OR (95% CI)*	Severe vs unaffected, OR (95% CI)*
HIV factors				
Most recent CD4 count§				
>500 cells per μL	1·00	1·00	1·00	1·00
200–500 cells per μL	1·49 (0·55–4·03)	1·28 (0·94–1·75)	1·15 (0·89–1·48)	1·62 (0·59–4·44)
<200 cells per μL	3·10 (1·06–9·13)	2·73 (1·80–4·14)	1·51 (1·04–2·21)	3·91 (1·31–11·62)
Most recent viral suppression, <200 copies per mL§	0·71 (0·27–1·89)	0·69 (0·49–0·97)	0·87 (0·64–1·17)	0·62 (0·24–1·57)
Social demographics				
Age, years				
18–49	1·00	1·00	1·00	1·00
50–64	1·51 (0·52–4·39)	0·88 (0·65–1·20)	0·60 (0·47–0·77)	0·62 (0·23–1·69)
≥65	3·39 (1·07–10·8)	0·79 (0·50–1·25)	0·37 (0·25–0·55)	0·58 (0·17–1·96)
Male vs female	1·12 (0·42–2·94)	0·88 (0·63–1·23)	0·69 (0·52–0·91)	1·29 (0·47–3·54)
Race				
White	1·00	1·00	1·00	1·00
Black or African American	2·33 (0·82–6·67)	1·77 (1·26–2·47)	1·56 (1·18–2·04)	2·08 (0·74–5·83)
Asian, other, or unknown	1·12 (0·30–4·18)	1·66 (1·09–2·54)	1·32 (0·93–1·86)	1·36 (0·36–5·06)
Ethnicity				
Not Hispanic or Latinx	1·00	1·00	1·00	1·00
Hispanic or Latinx	1·33 (0·33–5·35)	0·83 (0·55–1·25)	0·93 (0·68–1·27)	0·89 (0·23–3·50)
Other or unknown	2·67 (0·69–10·35)	0·82 (0·45–1·50)	0·48 (0·29–0·80)	1·16 (0·27–5·01)
Lifestyle factors				
Body-mass index, kg/m²				
≤30	1·00	1·00	1·00	1·00
>30	3·30 (1·14–9·53)	0·67 (0·47–0·96)	0·77 (0·55–1·08)	2·70 (1·00–7·29)
Unknown	1·71 (0·58–5·04)	0·22 (0·16–0·31)	0·45 (0·34–0·60)	0·52 (0·16–1·68)
Smoking status				
Non-smoker	1·00	1·00	1·00	1·00
Current or former smoker	2·57 (1·03–6·43)	1·09 (0·80–1·47)	0·46 (0·35–0·60)	1·41 (0·57–3·53)
				(Table 3 continues on next page)

that people with HIV might show less symptoms at the earlier stage of SARS-CoV-2 infection. Such protection from the most serious sequelae of COVID-19 might be attributable to the possible anti-SARS-CoV-2 activity of tenofovir disoproxil fumarate plus emtricitabine, as suggested in both observational and randomised closed trials studies.[27,28] Another hypothesis is that people with HIV with mild illness might be underrepresented (47·15% vs 62·33% in the overall group; table 1) because of higher stigma, increased fear of hospitalisation, higher social deprivation, and lower medical coverage when compared with people who do not have HIV. A consequence of such late linkage to care could be a higher risk of severe COVID-19.

As declining CD4 cell counts are associated with COVID-19 severity in general,[29] people with HIV and low CD4 cell counts might have a raised risk of severe COVID-19.[30] Our study supported this hypothesis and found that a lower CD4 cell count is associated with a higher risk of adverse COVID-19 outcomes, which is also in agreement with another multicentre study conducted by Dandachi and colleagues.[16] No association was observed between viral suppression and COVID-19 disease severity or mortality. Although our study observed the protective effect of viral suppression in reducing hospitalisation, the multicentre study did not.[16] Our larger sample size was possibly the reason for detecting such

a difference, because the multicentre study had a smaller sample and most of the study participants were receiving antiretroviral therapy.[16]

In this study, the age, sex, and race or ethnicity disparities in COVID-19 severe outcomes are pronounced among people with HIV. The study sample characteristics mirror the demographics of this population in the USA, with higher proportions of males, Hispanic or Latinx adults, and Black or African Americans adults. Hispanic or Latinx individuals, as well as those of an older age, have higher mortality and hospitalisation rates among people with HIV than in people without HIV, which were not reported in the UK study.[12] However, both our study and the UK study showed similar findings of the larger association between HIV and adverse COVID-19 outcomes among Black or African American adults. Understanding the reasons for the disproportionately large association between HIV and adverse COVID-19 outcomes in these subgroups will be a priority if effective policies are to be developed to mitigate any increased risks among these groups.

Our study had several limitations. First, although we included around 6 million individuals from the N3C dataset, the majority are based in the southeast, mid-Atlantic, and mid-west, and therefore might not be representative of the entire COVID-19 population or HIV population in the USA. Second, the algorithm for HIV case identification is not validated. Potential misclassification of HIV status might occur because of data availability and missing data on the N3C concept sets (HIV condition codes, antiretroviral therapy exposure, and laboratory results) that were available to the investigation team at the time of this study. Such identifications might change as the phenotype template of HIV patients changes as a result of continuous data updates from different contributing sites. Additionally, the release of other available concept sets might

might vary when stratifying by other vulnerable statuses of people with HIV, such as transgender individuals or injection drug users. However, codes for identifying these statuses were unavailable in this dataset.

In conclusion, using data from the largest COVID-19 population level analysis with a heterogeneous population in the USA, our study could identify people with HIV with mild or asymptomatic COVID-19 and examine the different risks for SARS-CoV-2 acquisition versus progression to severe disease or death once infected. In this large study, people with HIV have an elevated risk of adverse COVID-19 outcomes. The attenuated risk after controlling for comorbidities, which are more prevalent and typically occur at a younger age among people with HIV, indicates that certain underlying medical conditions had a greater influence on COVID-19 outcomes of this population. Our observation that people with lower CD4 cell counts are at a higher risk of poor outcomes suggests that people with a history of advanced immuno-suppression might warrant closer observation and monitoring. The robust risk assessment of this study could inform prioritisation of prevention messaging, disease monitoring and therapies, and vaccination for people with HIV, especially those with more pronounced immunodeficiency. Given the pandemic's exacerbating effects on health inequities, public health and clinical communities must strengthen services and support to prevent aggravated COVID-19 outcomes among people with HIV, particularly for those with pronounced immunodeficiency.