

Higher mortality following SARS-CoV-2 infection in rural versus urban dwellers persists for two years post-infection

Received: 17 October 2024

Accepted: 29 August 2025

Published online: 08 October 2025

 Check for updates

A. Jerrod Anzalone¹✉, Michael T. Vest², Makayla E. Schissel¹, Bradley Price³, William B. Hillegass⁴, Ronald Horswell⁵, San Chu⁵, Clifford J. Rosen^{6,7}, Lucio Miele⁸, Susan L. Santangelo^{6,7}, Gordon S. Smith³ & Sally L. Hodder³ On behalf of the National Clinical Cohort Collaborative (N3C) Consortium*

Previous studies demonstrated higher short-term mortality among rural compared with urban residents infected with SARS-CoV-2. However, whether this difference persists remains uncertain. This retrospective cohort study analyzed two-year post-COVID-19 mortality by rurality using the National Clinical Cohort Collaborative COVID-19 Enclave, a United States-based longitudinal electronic health record repository. We analyzed mortality among patients infected with SARS-CoV-2 between April 2020 and December 2022, with follow-up until December 2024. Patients were categorized into urban, urban-adjacent rural (UAR), and nonurban-adjacent rural (NAR) groups based on residential ZIP Code. Mortality differences were assessed using Kaplan-Meier analysis and weighted multivariable Cox regression, with weights derived from demographic factors and models adjusted for background clinical risk and social vulnerability. Among 3,082,978 SARS-CoV-2-infected patients, we found a significant association between rurality and increased two-year all-cause mortality post-infection. Adjusted hazards for two-year mortality for UAR and NAR were 1.19 (95% CI 1.18-1.21) and 1.26 (1.22-1.29). A reference cohort of 4,153,216 COVID-19-negative patients showed a modest yet consistent rural mortality penalty, with a similar relative hazard across cohorts, an observed rurality-COVID-19 interaction, and a greater absolute number of deaths following SARS-CoV-2 infection. Our findings emphasize ongoing rural mortality disparities and the importance of public health efforts in rural communities.

Rural regions in the U.S. have unique health challenges, including higher socioeconomic disparities¹, greater comorbid burden², and limited healthcare access³. As highlighted in Rural Healthy People 2030⁴, rural residents face distinct health disparities in many areas,

such as mental health⁵, substance use disorders⁶, obesity⁷, diabetes⁸, and preventive care⁹. These and other disparities contribute to the growing divide in rural-urban health outcomes and what has collectively been described as the rural mortality penalty¹⁰. This term refers

¹University of Nebraska Medical Center, Omaha, NE, USA. ²Christiana Care Health System, Newark, DE, USA. ³West Virginia University, Morgantown, WV, USA. ⁴University of Mississippi Medical Center, Jackson, MS, USA. ⁵Pennington Biomedical Research Center, Baton Rouge, LA, USA. ⁶MaineHealth Institute for Research, Scarborough, ME, USA. ⁷Tufts University School of Medicine, Boston, MA, USA. ⁸Louisiana State University Health Sciences Center, New Orleans, LA, USA. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: alfred.anzalone@unmc.edu

to the higher mortality rates experienced in the U.S. by rural communities relative to urban communities, including shorter life expectancies¹¹. This phenomenon contrasts with the longer life expectancy in rural areas observed before the mid-twentieth century¹².

The COVID-19 pandemic has exacerbated the existing rural mortality penalty, resulting in even greater loss of life in rural areas since 2020¹³. Higher rates of acute COVID-19 mortality have been observed in rural areas during the Delta and Omicron periods¹³, driven by factors such as limited access to healthcare, higher prevalence of comorbidities, and social determinants of health (SDOH). COVID-19 has increased rates of social isolation, substance use, and mental health disorders^{6,14}, placing more strain on rural communities as rural hospital closures continue to increase¹⁵. These issues, combined with limited access to specialized and critical care resources in rural facilities, underscore the heightened risk of COVID-19-associated deaths in rural areas during the acute phase of infection, even after adjustment for background risk factors¹⁶.

While we know from prior work that COVID-19 has a profound impact on outcomes in the acute period after infection¹⁶, including among vaccinated persons¹⁷ and after accounting for therapeutic interventions¹⁸, the long-term impact of COVID-19 in rural communities and its impact on the rural mortality penalty has yet to be studied. This study seeks to understand differences in short- and long-term mortality among rural and urban dwellers with SARS-CoV-2 infection, controlling for demographic differences and comorbid burden, and to quantify differences in outcomes based on region and degree of rurality. We hypothesized that rural dwellers would have worse outcomes than their urban counterparts, even after adjusting for demographic differences, comorbid conditions, vaccination status, and infection time period. However, we hypothesized that these differences may attenuate with time. This study aims to systematically understand the rural-urban mortality gap during and after SARS-CoV-2 infection to inform targeted public health interventions addressing these disparities.

Results

After applying inclusion/exclusion criteria, our cohort included 3,082,978 SARS-CoV-2-infected patients (Table 1, Fig. 1). This included 2,466,473 urban (80%), 505,382 UAR (16.4%), and 111,123 NAR dwellers (3.6%), which is a similar distribution to the US distribution of urban-rural population density as of the 2020 Census (~20% rural)¹⁹. The sample includes patients from all US Census Divisions, with the highest representation from the East North Central and South Atlantic regions across all rurality groups (Supplementary Fig. 1). Median age was higher in nonurban-adjacent rural (NAR; 51 years) and urban-adjacent rural (UAR; 49 years) compared to urban areas (46 years). Most patients were white non-Hispanic across all areas, but this proportion was higher in rural (UAR 85%, NAR 91%) than in urban areas (67%). Urban areas had higher proportions of Black or African American non-Hispanic (14%) and Hispanic or Latino (10%) compared to UAR and NAR areas.

The distribution of infections across different SARS-CoV-2 variant periods was consistent across all areas. Urban areas had a higher percentage of patients with primary (15%) and additional vaccination doses before infection (12%) compared to UAR (13% and 8.0%, respectively) and NAR (12% and 7.2%, respectively). Myocardial infarction, heart failure, vascular, cerebrovascular, rheumatic, metabolic, neurodegenerative, and renal diseases were more prevalent in rural areas, while pulmonary disease and HIV were more common in urban areas. Obesity prevalence was slightly higher among urban (39%) than rural dwellers (37%), but greater BMI missingness in rural residents (44% vs 33%) may contribute to underestimation. Tobacco usage and substance use disorder were higher among urban residents, while rural residents had lower Social Vulnerability Index (SVI) scores (UAR 0.43 SVI and NAR 0.37 SVI) relative to urban residents (0.48 SVI).

Kaplan-meier cumulative incidence

Kaplan-Meier cumulative incidence curves (Fig. 2) revealed a significant association between rurality and two-year cumulative mortality post-infection ($p < 0.001$), with a gradient risk by rurality. In month 1, the mortality difference between urban and UAR was 0.62% (95% confidence interval [CI]: 0.58%–0.68%), and between urban and NAR areas was 0.80% (0.71%–0.90%). These gaps widened over time, reaching 0.81% (0.74%–0.87%) and 1.00% (0.89%–1.12%) at month 3; 1.17% (1.10%–1.25%) and 1.46% (1.33%–1.60%) at one year; and 1.62% (1.54%–1.71%) and 1.86% (1.70%–2.02%) at two years for UAR and NAR relative to urban residents, respectively. This demonstrates a growing mortality gap, with rural areas experiencing higher absolute and relative mortality rates than urban areas.

Cumulative and excess mortality

The cumulative mortality and associated excess mortality revealed a consistent trend of increasing rural-versus-urban mortality over time (Fig. 3, and Supplementary Data 1). Within the first month, the age-adjusted overall mortality was 1005.67 per 100,000 persons (95% CI: 995.99–1015.36). Urban areas had a lower mortality of 943.44 (932.75–954.13). Rural areas experienced significantly higher mortality, with UAR at 1213.09 (1188.35–1237.83) and NAR at 1249.57 (1197.36–1301.78), showing excess mortality of 269.65 (216.82–322.47) and 306.13 (201.67–410.59), respectively, per 100,000 persons relative to urban.

This difference increased gradually at three months, one year, and two years after infection. By three months, age-adjusted overall mortality increased to 1387.50 (1376.12–1398.87); at one year, it rose to 2158.71 (2144.52–2172.91); and by two years, the overall mortality reached 2970.37 (2953.71–2987.04) per 100,000 persons. Two years post-infection, urban mortality stood at 2823.77 (2805.27–2842.27), while rural areas were much higher, with UAR at 3484.71 (3442.60–3526.82) and NAR at 3,436.15 (3348.90–3523.40) per 100,000 persons. The excess mortality at two years was 660.94 (570.78–751.09) for UAR and 612.38 (437.57–787.19) for NAR per 100,000 persons compared to the urban mortality rate. Crude and age-adjusted risk and rates per 100,000 person-months are provided in Supplementary Data 1.

Cox proportional hazards for post-COVID-19 death

To understand the progressive risk of death after SARS-CoV-2 infection over time (Fig. 4, Supplementary Data 2), we assessed one-month, three-month, one-year, and two-year mortality risk by rurality after adjusting for background risk at the time of infection. After IPW, we observed a stable adjusted mortality risk across all periods in unweighted-unadjusted and weighted-adjusted models. At one-month post-COVID-19, UAR was associated with an adjusted hazard ratio (aHR) of 1.19 (95% CI 1.16–1.22) for all-cause mortality, while NAR was associated with an aHR of 1.36 (1.30–1.42) relative to urban dwellers. We found a similar increased risk for rural dwellers at three months (UAR aHR 1.18 [1.15–1.20] and NAR aHR 1.31 [1.26–1.37]), one year (UAR aHR 1.18 [1.16–1.20] and NAR aHR 1.29 [1.25–1.34]), and two years (UAR aHR 1.19 [1.18–1.21] and NAR aHR 1.26 [1.22–1.29]) after infection relative to urban dwellers.

Older age, male sex, non-white race/ethnicity, and a history of cardiovascular, pulmonary, neurological, metabolic, rheumatic, hepatic, renal, or malignant disease, as well as HIV, tobacco use, and substance use disorder, were consistently associated with increased post-COVID-19 all-cause mortality (Supplementary Data 2). Patients who received SARS-CoV-2 vaccination exhibited a significantly reduced mortality risk at all observed time points post-COVID-19, with the greatest protective effect observed in the immediate post-COVID-19 period (e.g., month one versus two years). The Delta and ancestral variant dominant periods had a higher risk of death than the Alpha, Beta, and Gamma variant periods. In contrast, both Omicron periods

Table 1 | Descriptive Statistics for Patients with a SARS-CoV-2 Infection between April 2020 and December 2022 by Rurality

Characteristic ¹	Overall, N = 3,082,978	Urban, N = 2,466,473	Urban-Adjacent Rural, N = 505,382	Nonurban-Adjacent Rural, N = 111,123
Sex				
Female	1,805,923 (59%)	1,449,910 (59%)	292,918 (58%)	63,095 (57%)
Male	1,277,055 (41%)	1,016,563 (41%)	212,464 (42%)	48,028 (43%)
Age at COVID-19, Median (IQR)	47 (32, 62)	46 (32, 61)	49 (33, 64)	51 (35, 65)
Age Group				
<30	607,957 (20%)	496,926 (20%)	92,560 (18%)	18,471 (17%)
30-49	1,078,944 (35%)	880,778 (36%)	164,373 (33%)	33,793 (30%)
50-64	756,768 (25%)	598,101 (24%)	128,822 (25%)	29,845 (27%)
65-74	363,341 (12%)	281,534 (11%)	65,535 (13%)	16,272 (15%)
75+	275,968 (9.0%)	209,134 (8.5%)	54,092 (11%)	12,742 (11%)
Race/Ethnicity				
White Non-Hispanic	2,185,101 (71%)	1,655,102 (67%)	429,223 (85%)	100,776 (91%)
Black or African American Non-Hispanic	384,750 (12%)	355,568 (14%)	25,584 (5.1%)	3598 (3.2%)
Hispanic or Latino Any Race	289,221 (9.4%)	256,294 (10%)	29,377 (5.8%)	3550 (3.2%)
Other	121,679 (3.9%)	106,337 (4.3%)	13,432 (2.7%)	1910 (1.7%)
Unknown	102,227 (3.3%)	93,172 (3.8%)	7766 (1.5%)	1289 (1.2%)
SARS-CoV-2 Variant Dominant Period				
Ancestral COVID-19	778,618 (25%)	626,929 (25%)	124,738 (25%)	26,951 (24%)
Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1)	358,222 (12%)	289,436 (12%)	56,239 (11%)	12,547 (11%)
Delta (B.1.617.2)	582,071 (19%)	434,812 (18%)	119,093 (24%)	28,166 (25%)
Omicron (B.1.1.529, BA.2, BA.2.12.1)	918,020 (30%)	753,989 (31%)	135,835 (27%)	28,196 (25%)
Omicron (BA.5, BQ.1.1, XBB.1.5)	446,047 (14%)	361,307 (15%)	69,477 (14%)	15,263 (14%)
Vaccination Status				
No Documented Vaccination Before Acute COVID-19	2,287,372 (74%)	1,800,465 (73%)	397,525 (79%)	89,382 (80%)
Primary Vaccination Series Documented Before COVID-19	456,654 (15%)	375,461 (15%)	67,459 (13%)	13,734 (12%)
Primary and Additional Dose(s) Documented Before Acute COVID-19	338,952 (11%)	290,547 (12%)	40,398 (8.0%)	8007 (7.2%)
Comorbid Conditions				
Obesity ²	1,201,038 (39%)	972,503 (39%)	187,370 (37%)	41,165 (37%)
Hypertension	852,422 (28%)	680,984 (28%)	139,965 (28%)	31,473 (28%)
Myocardial Infarction	82,117 (2.7%)	62,275 (2.5%)	16,145 (3.2%)	3697 (3.3%)
Congestive Heart Failure	104,139 (3.4%)	81,125 (3.3%)	18,716 (3.7%)	4298 (3.9%)
Peripheral Vascular Disease	94,972 (3.1%)	75,142 (3.0%)	16,314 (3.2%)	3516 (3.2%)
Cerebrovascular Disease	100,470 (3.3%)	79,994 (3.2%)	16,709 (3.3%)	3767 (3.4%)
Rheumatic Disease	152,698 (5.0%)	122,322 (5.0%)	24,702 (4.9%)	5674 (5.1%)
Pulmonary Disease	622,794 (20%)	506,387 (21%)	95,014 (19%)	21,393 (19%)
Peptic Ulcer Disease	32,999 (1.1%)	26,149 (1.1%)	5656 (1.1%)	1194 (1.1%)
Hemiplegia/Paraplegia	24,042 (0.8%)	18,940 (0.8%)	4190 (0.8%)	912 (0.8%)
Diabetes	409,204 (13%)	323,796 (13%)	70,046 (14%)	15,362 (14%)
Dementia	43,851 (1.4%)	34,353 (1.4%)	7,884 (1.6%)	1614 (1.5%)
Liver Disease	137,060 (4.4%)	110,162 (4.5%)	22,198 (4.4%)	4700 (4.2%)
Renal Disease	263,038 (8.5%)	208,258 (8.4%)	44,799 (8.9%)	9981 (9.0%)
Cancer	192,043 (6.2%)	154,484 (6.3%)	30,011 (5.9%)	7548 (6.8%)
HIV	12,571 (0.4%)	11,355 (0.5%)	1046 (0.2%)	170 (0.2%)
Current or Former Tobacco Usage	296,361 (9.6%)	239,828 (9.7%)	46,337 (9.2%)	10,196 (9.2%)
Substance Use Disorder	84,525 (2.7%)	69,277 (2.8%)	12,829 (2.5%)	2419 (2.2%)
Social Vulnerability Index Composite Score	0.47 (0.25, 0.68)	0.48 (0.25, 0.69)	0.43 (0.27, 0.64)	0.37 (0.20, 0.63)
US Census Division				
East North Central	1,529,587 (50%)	1,185,094 (48%)	282,638 (56%)	61,855 (56%)
East South Central	187,720 (6.1%)	146,319 (5.9%)	35,677 (7.1%)	5724 (5.2%)
Middle Atlantic	37,862 (1.2%)	34,119 (1.4%)	3070 (0.6%)	673 (0.6%)
Mountain	246,596 (8.0%)	229,598 (9.3%)	14,436 (2.9%)	2562 (2.3%)
New England	55,737 (1.8%)	47,149 (1.9%)	4732 (0.9%)	3856 (3.5%)

Table 1 (continued) | Descriptive Statistics for Patients with a SARS-CoV-2 Infection between April 2020 and December 2022 by Rurality

Characteristic ¹	Overall, N = 3,082,978	Urban, N = 2,466,473	Urban-Adjacent Rural, N = 505,382	Nonurban-Adjacent Rural, N = 111,123
Pacific	130,202 (4.2%)	117,807 (4.8%)	9312 (1.8%)	3083 (2.8%)
South Atlantic	544,959 (18%)	429,288 (17%)	95,197 (19%)	20,474 (18%)
West North Central	195,606 (6.3%)	129,227 (5.2%)	54,692 (11%)	11,687 (11%)
West South Central	154,709 (5.0%)	147,872 (6.0%)	5628 (1.1%)	1209 (1.1%)

1. Median (IQR) for continuous variables; *n* (%) for categorical variables.
 2. Missing BMI values were more common among rural residents, including both Urban-Adjacent Rural (44%) and Nonurban-Adjacent Rural (44%), compared to Urban residents (33%). This may contribute to underestimation of obesity prevalence in rural settings. Additionally, normal weight and overweight BMI classifications were more frequently identified among urban dwellers (11% and 18%, respectively) than among urban-adjacent (7.0%, 14%) and nonurban-adjacent rural residents (6.8%, 14%).
 Table 1 provides descriptive statistics for patients with a documented SARS-CoV-2 infection between April 1, 2020, and December 31, 2022.

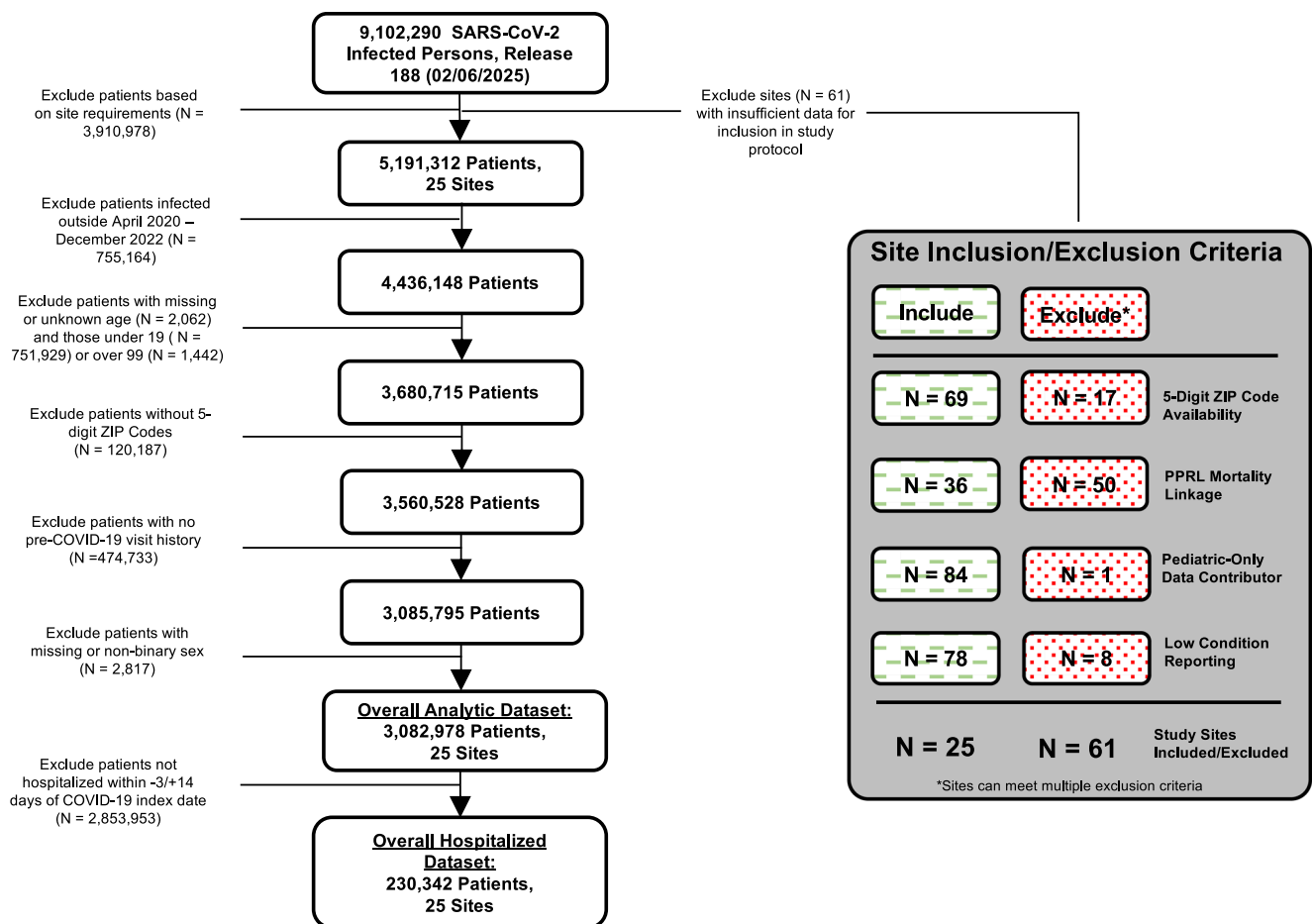


Fig. 1 | Cohort Selection Flow Diagram. Figure 1 shows the inclusion and exclusion criteria applied to N3C release v188 (February 6, 2025) to reach the analytic dataset used in this study.

had a lower risk. Patients residing in communities with higher SVI values had an increased risk of adverse outcomes up to two years post-infection (aHR 1.23 [1.21–1.26]). This means that for every one-unit increase in SVI, there is a 26% increase in the risk of the event occurring, even after adjusting for other factors. Since SVI ranges from 0 to 1, smaller increments in SVI (e.g., from 0.5 to 0.6) would correspond to a proportionally smaller increase in risk.

To evaluate the relative risk of death among patients who survived various periods post-COVID-19 (Fig. 4, Supplementary Data 2), we examined two-year survival among those who survived to landmarks at one month, three months, and one year post-infection. The association between rurality and two-year all-cause mortality

was consistent among patients who survived the acute period of COVID-19. We observed a similar increased risk for rural dwellers who were alive after month one (UAR aHR 1.20 [1.18–1.22] and NAR aHR 1.20 [1.16–1.24]), month three (UAR aHR 1.21 [1.18–1.23] and NAR aHR 1.20 [1.16–1.25]), and one year (UAR aHR 1.22 [1.19–1.25] and NAR aHR 1.16 [1.10–1.22]) after initial infection relative to urban dwellers.

New-Onset Diagnoses Among Decedents Within Two Years of COVID-19

Among patients who died within two years of infection (Fig. 5, and Supplementary Data 3), rural dwellers exhibited lower documented

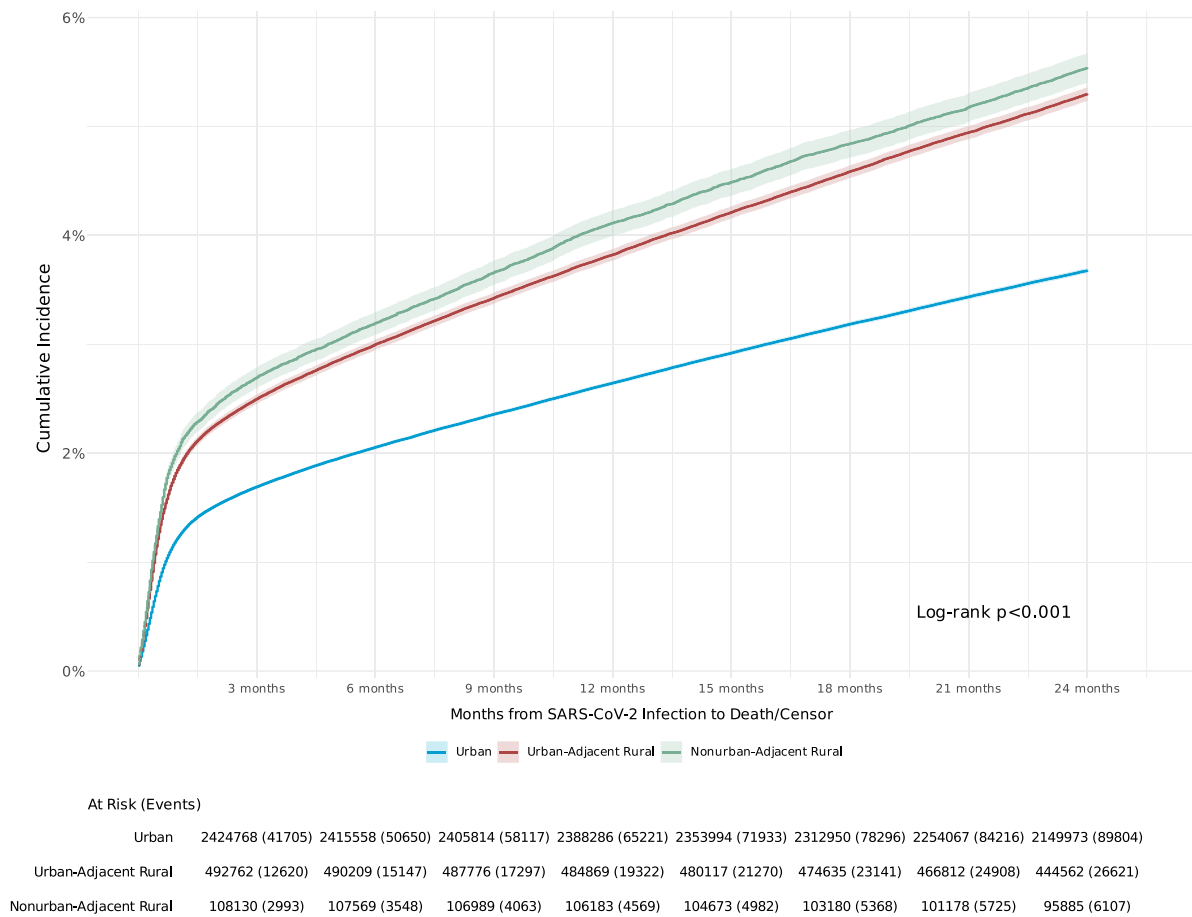


Fig. 2 | Kaplan-Meier cumulative incidence curves for two-year mortality following SARS-CoV-2 infection from April 2020 through December 2022 by rurality. Figure 2 shows Kaplan-Meier cumulative incidence from day 0 through day 730 in the period after diagnosis of COVID-19. Patients were included if they had a documented SARS-CoV-2 infection between April 1, 2020, and December 31, 2022, with follow-up through December 31, 2024. Patients are followed from their initial SARS-CoV-2 infection for up to 730 days, ending at 1) death, 2) end of

observation window, or 3) censoring (e.g., end of available data). This analysis uses a three-level rural categorization: urban, urban-adjacent rural, and nonurban-adjacent rural, to assess long-term manifestations of COVID-19 based on proximity and commuting patterns to urban centers. Data are presented as Kaplan-Meier cumulative incidence estimates, with shaded bands representing 95% confidence intervals around the mean cumulative incidence function.

rates of new-onset diagnoses across multiple systems—including respiratory, cardiovascular, immune, nervous, digestive, endocrine, musculoskeletal, integumentary, and genitourinary—as well as reduced diagnoses of neoplasms, constitutional symptoms, blood and blood-forming tissue abnormalities, metabolic homeostasis disturbances, and a combined variable indicating any post-COVID condition ($p < 0.001$). Rural residents had similar rates of SARS-CoV-2 reinfection and long COVID among decedents compared with urban residents ($p = 0.11$ and $p = 0.15$, respectively).

Post-COVID-19 mortality among patients hospitalized with acute COVID-19

Among the 230,342 patients hospitalized with COVID-19 (-3 to +14 days of infection), 178,542 (78%) were urban, 41,857 (18%) were UAR, and 9,943 (4.3%) were NAR dwellers. Compared to the overall cohort, hospitalized rural patients were more likely to have several comorbid conditions (Supplementary Data 4) but were still less likely to be vaccinated. Those hospitalized had an age-adjusted overall two-year mortality risk of 11,416.18 (95% CI 11,291.81–11,540.55) per 100,000 persons, with an urban risk of 11,097.53 (10,959.82–11,235.23), UAR risk of 12,490.37 (12,174.17–12,806.56), and NAR risk of 12,905.96 (12,167.04–13,644.87). The excess mortality at two years among hospitalized patients was 1,392.84 (716.88–2,068.79) for UAR and 1,808.43 (335.24–3,281.61) for NAR per 100,000 persons compared to the urban

mortality rate. Hospitalized rural dwellers had a higher mortality across all periods.

After IPW within the hospitalized cohort with good balance (Supplementary Fig. 3), we observed similar increased risk across all post-infection periods as in the overall cohort (Supplementary Data 6, Fig. 6). At one-month post-COVID-19, UAR was associated with an aHR of 1.18 (95% CI 1.15–1.22) for all-cause mortality, while NAR was associated with an aHR of 1.22 (1.15–1.29) relative to urban dwellers. We found a similar increased risk for rural dwellers at three months (UAR aHR 1.16 [1.13–1.19] and NAR aHR 1.22 [1.16–1.28]), one year (UAR aHR 1.16 [1.13–1.19] and NAR aHR 1.21 [1.16–1.26]), and two years (UAR aHR 1.16 [1.13–1.18] and NAR aHR 1.19 [1.15–1.24]) after infection relative to urban dwellers. We observed similar associations between rurality and two-year mortality in patients who survived to landmarks at one month, three months, and one year after infection. We also observed similar post-COVID conditions among hospitalized decedents, with rural residents having lower new-onset conditions observed (Supplementary Data 7).

Post-COVID-19 Mortality Differences Stratified by Rurality

In models stratified by rural residency (Fig. 7, and Supplementary Data 8), we observed significant differences in risk across key demographic and clinical factors. Males in rural and urban settings had a higher two-year mortality risk than females, with rural males having an

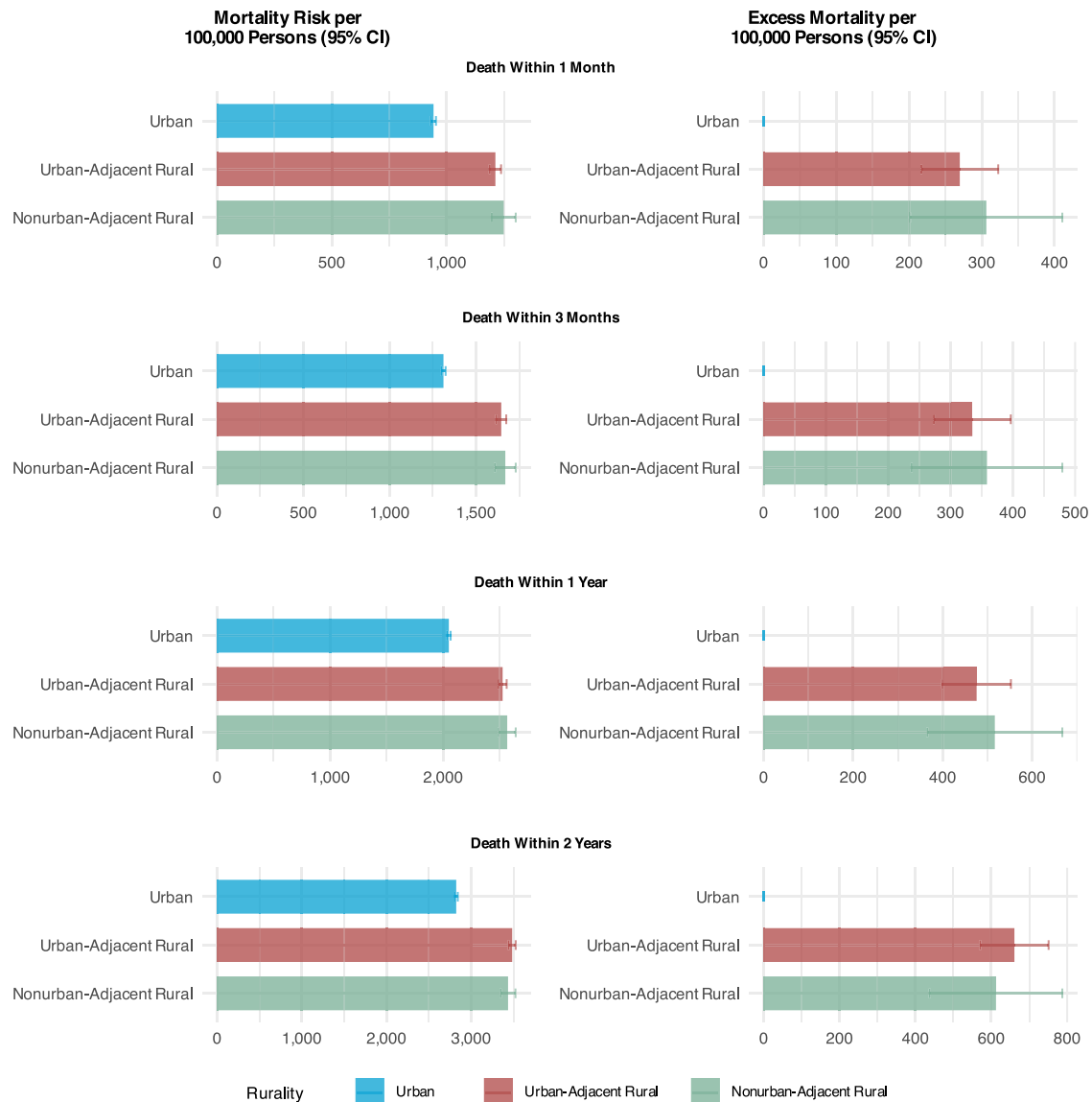


Fig. 3 | Age-Standardized Cumulative Risk and Excess Mortality Following SARS-CoV-2 Infection by Rurality. Figure 3 shows age-adjusted differences in cumulative mortality risk (per 100,000) after SARS-CoV-2 infection, stratified by rurality. Risks were estimated at 1, 3, 12, and 24 months post-infection, with risk differences compared to the urban cohort. Point estimates are displayed with 95% confidence intervals, calculated assuming a Poisson distribution. Data are

presented as age-standardized mean risk, with whiskers indicating 95% confidence intervals. Because these are population-level standardized estimates, individual data distributions are not shown. Estimates were standardized to the 2020 U.S. Standard Population to account for age differences. Full risks, rates, and confidence intervals are available in Supplementary Data 1.

aHR of 1.21 (95% CI: 1.18–1.24) and urban males showing a slightly higher aHR of 1.25 (1.23–1.27). Age remained a strong determinant of mortality, with older populations showing steep increases in risk. In rural areas, individuals aged 75 and older had an aHR of 21.98 (20.88–23.14), compared to 18.27 (17.74–18.81) in urban areas.

Social vulnerability was associated with higher mortality in rural and urban settings, with SVI having an aHR of 1.29 (1.22–1.36) for rural residents and 1.31 (1.27–1.34) for urban residents. These trends persisted for patients with a COVID-19-associated hospitalization (Fig. 8, and Supplementary Data 9), except that social vulnerability was associated with higher all-cause mortality in rural than urban settings.

Post-COVID-19 mortality differences by dichotomous rurality

We conducted a secondary analysis using a binary rural indicator instead of the three-level rural classification scheme and found that the

main findings were consistent with the original analysis. Detailed results of this analysis are available in Supplementary Data 10–11, which further support the robustness of the findings across different rural classification approaches.

COVID-19-Negative Comparison Cohort

To assess whether rural mortality disparities were specific to SARS-CoV-2 infection or reflected broader rural mortality differences, we examined a COVID-19-negative comparison cohort. This cohort included 4,153,216 individuals without documented SARS-CoV-2 infection, comprising 3,480,053 urban, 535,415 UAR, and 137,748 NAR residents (Supplementary Data 12). In models restricted to this cohort, two-year all-cause mortality was higher among rural dwellers (Supplementary Data 13), with UAR and NAR residents having aHRs of 1.17 (95% CI: 1.15–1.19) and 1.21 (1.17–1.25), respectively, compared to urban dwellers.

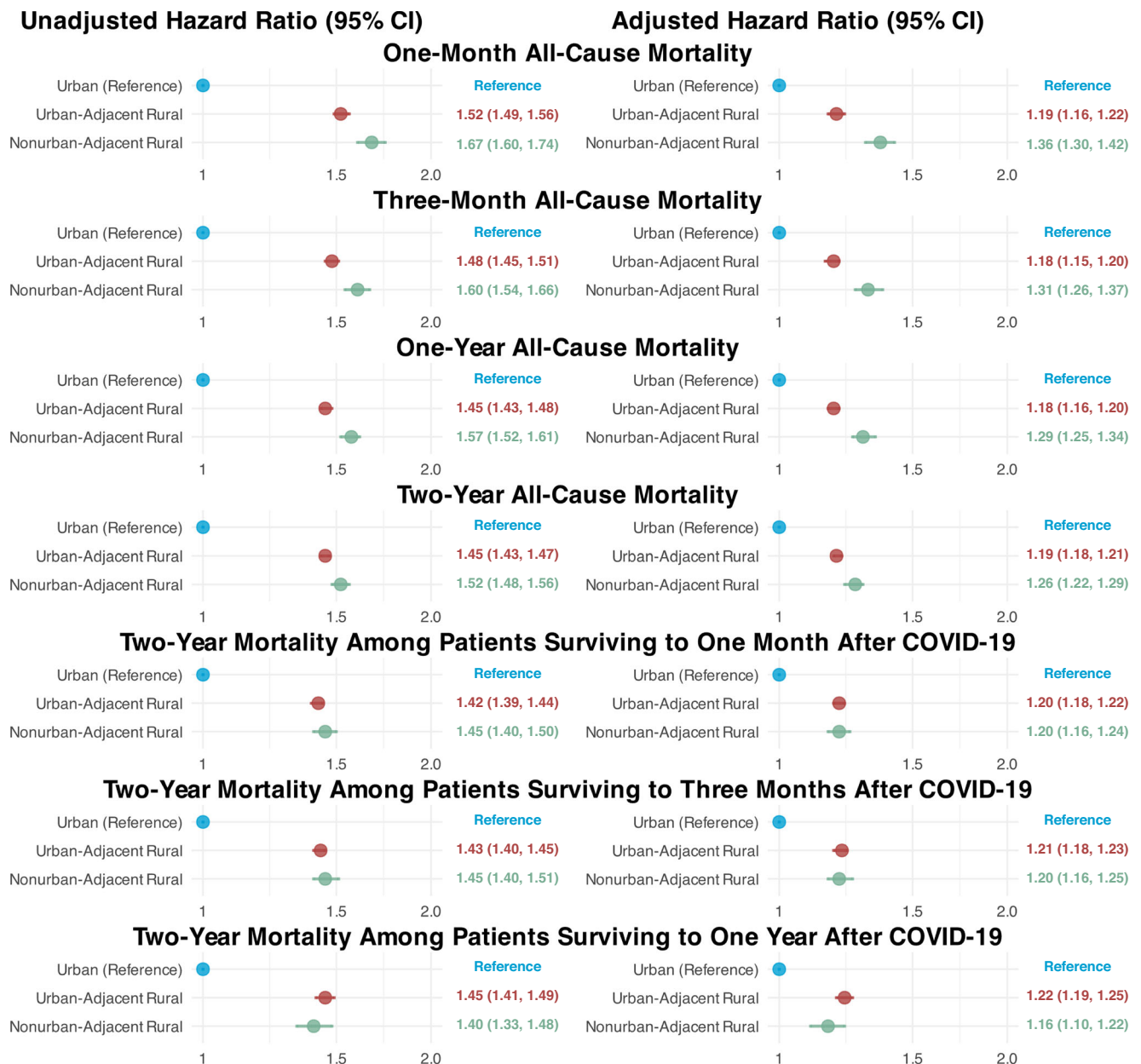


Fig. 4 | Unadjusted and adjusted hazard ratio for all-cause mortality among patients with an acute SARS-CoV-2 infection from April 2020 through December 2022 by rurality. Figure 4 contains Cox Proportional Hazards models for all-cause mortality after an acute SARS-CoV-2 infection by rurality. Deaths are captured from month 0 (day 0) through month 24 (day 730) in the period after diagnosis of COVID-19 at intervals of A) one month, B) three months, C) one year, D) two years, E) two years among patients surviving the initial month after infection, F) two years among patients surviving the initial 3 months after infection, and G) two years among patients surviving the initial one year after infection. Adjustments were made for sex, age, race/ethnicity, SARS-CoV-2 variant-dominant period at the

time of infection, vaccination status, Social Vulnerability Index, U.S. Census Division, and a history of obesity, hypertension, myocardial infarction, congestive heart failure, peripheral vascular disease, rheumatic disease cerebrovascular disease, dementia, chronic pulmonary disease, peptic ulcer disease, hemiplegia or paraplegia, diabetes, liver disease, renal disease, cancer, HIV, tobacco usage, and substance use disorder documented before COVID-19. Forest plots display hazard ratios (points) on a log scale, with horizontal lines representing 95% confidence intervals (error bars). The vertical line indicates the null value of 1.0. Full model specifications are available in Supplementary Data 2.

Interaction Between Rurality and COVID-19

To assess whether COVID-19 modified the rural mortality penalty, we included a rurality-by-COVID-19 interaction term in weighted Cox models that included both SARS-CoV-2-positive and SARS-CoV-2-negative patients (Supplementary Data 14). At two years, we observed an interaction aHR of 1.06 (95% CI, 1.04–1.08) for UAR and 1.06 (1.04–1.08) for NAR between rural residency and COVID-19. These findings suggest an interaction, with COVID-19 slightly increasing associations between rurality and death.

To quantify the mortality burden of SARS-CoV-2 infection, we calculated the age-adjusted excess mortality risk per 100,000 persons by rurality (Supplementary Data 15) among infected versus uninfected patients. Overall, COVID-19 was associated with an excess risk of 1668.4 (1668.4–1668.4) deaths per 100,000 persons. This risk was higher in UAR (1794.9 [1794.9–1794.9]) and NAR (1755.7 [1755.7–1755.7]) than among urban residents (1603.6 [1603.6–1603.6]).

In a combined model that included both SARS-CoV-2-positive and -negative individuals (Supplementary Data 16), but with no interaction

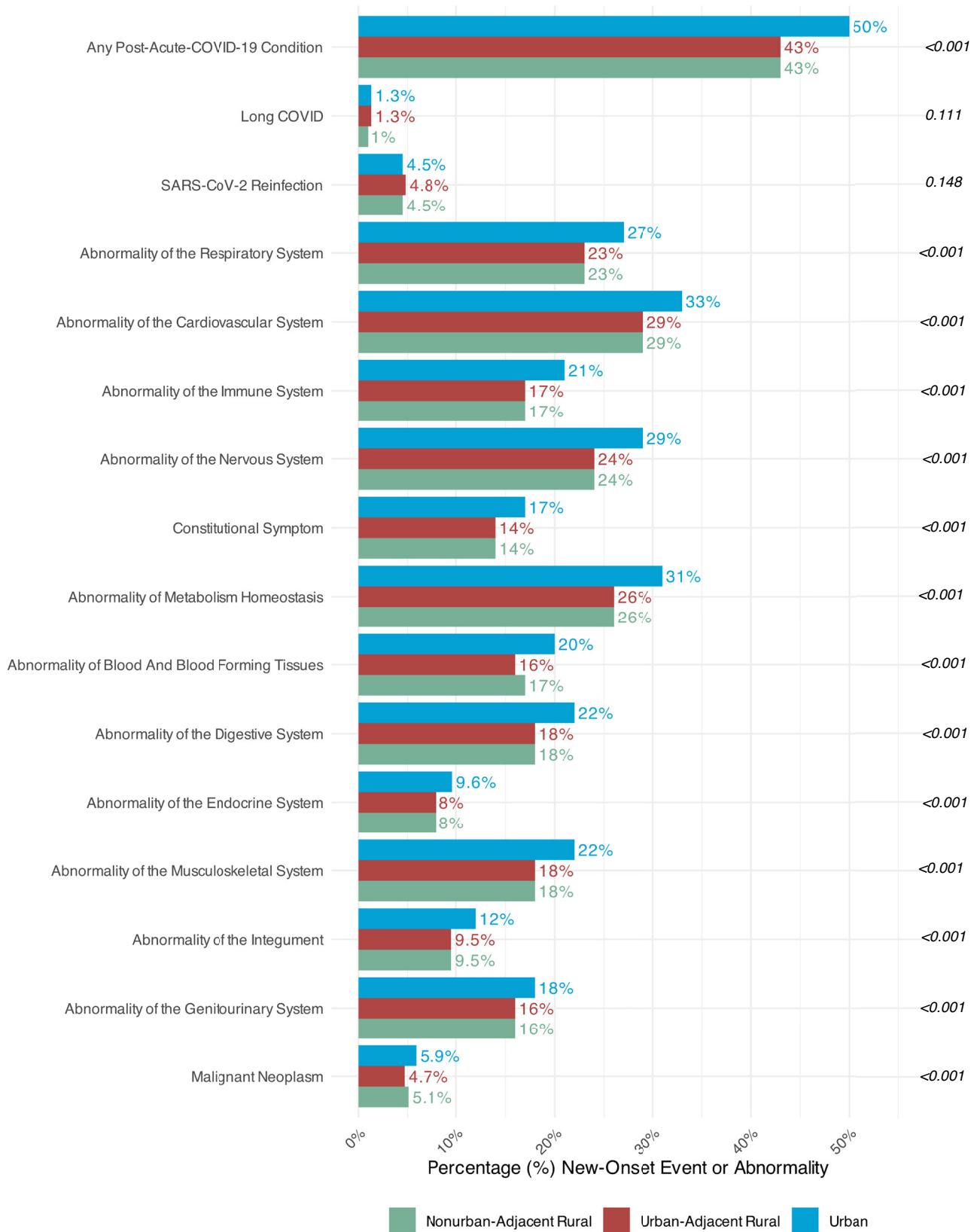


Fig. 5 | New-Onset Post-COVID-19 Conditions and Events Among Patients Dying Within 2 Years by Rurality. Figure 5 provides new-onset conditions and events for patients with a documented SARS-CoV-2 infection between April 1, 2020, and December 31, 2022, who died within two years. Outcomes and conditions are

documented following acute COVID-19 but before death by rurality. P values are calculated using two-sided Pearson's chi-squared tests. Full results are available in Supplementary Data 3.



Fig. 6 | Unadjusted and adjusted hazard ratio for all-cause mortality among patients hospitalized after an acute SARS-CoV-2 infection from April 2020 through December 2022 by rurality. Figure 6 contains Cox Proportional Hazards models for all-cause mortality among patients hospitalized within two weeks (-3/+14 days) after an acute SARS-CoV-2 infection by rurality. Deaths are captured from month 0 (day 0) through month 24 (day 730) in the period after diagnosis of COVID-19 at intervals of A) one month, B) three months, C) one year, D) two years, E) two years among patients surviving the initial month after infection, F) two years among patients surviving the initial three months after infection, and G) two years among patients surviving the initial one year after infection. Adjustments were

made for sex, age, race/ethnicity, SARS-CoV-2 variant-dominant period at the time of infection, vaccination status, Social Vulnerability Index, U.S. Census Division, and a history of obesity, hypertension, myocardial infarction, congestive heart failure, peripheral vascular disease, rheumatic disease cerebrovascular disease, dementia, chronic pulmonary disease, peptic ulcer disease, hemiplegia or paraplegia, diabetes, liver disease, renal disease, cancer, HIV, tobacco usage, and substance use disorder documented before COVID-19. Forest plots display hazard ratios (points) on a log scale, with horizontal lines representing 95% confidence intervals (error bars). The vertical line indicates the null value of 1.0. Full model specifications are available in Supplementary Data 6.

term, rurality remained independently associated with higher two-year mortality (UAR aHR 1.21 [1.19–1.22]; NAR aHR 1.26 [1.23–1.29]). SARS-CoV-2 infection was a strong independent predictor of mortality, with an aHR of 1.63 (1.62–1.65) relative to those without documented infection, with the risk reducing with greater time from infection.

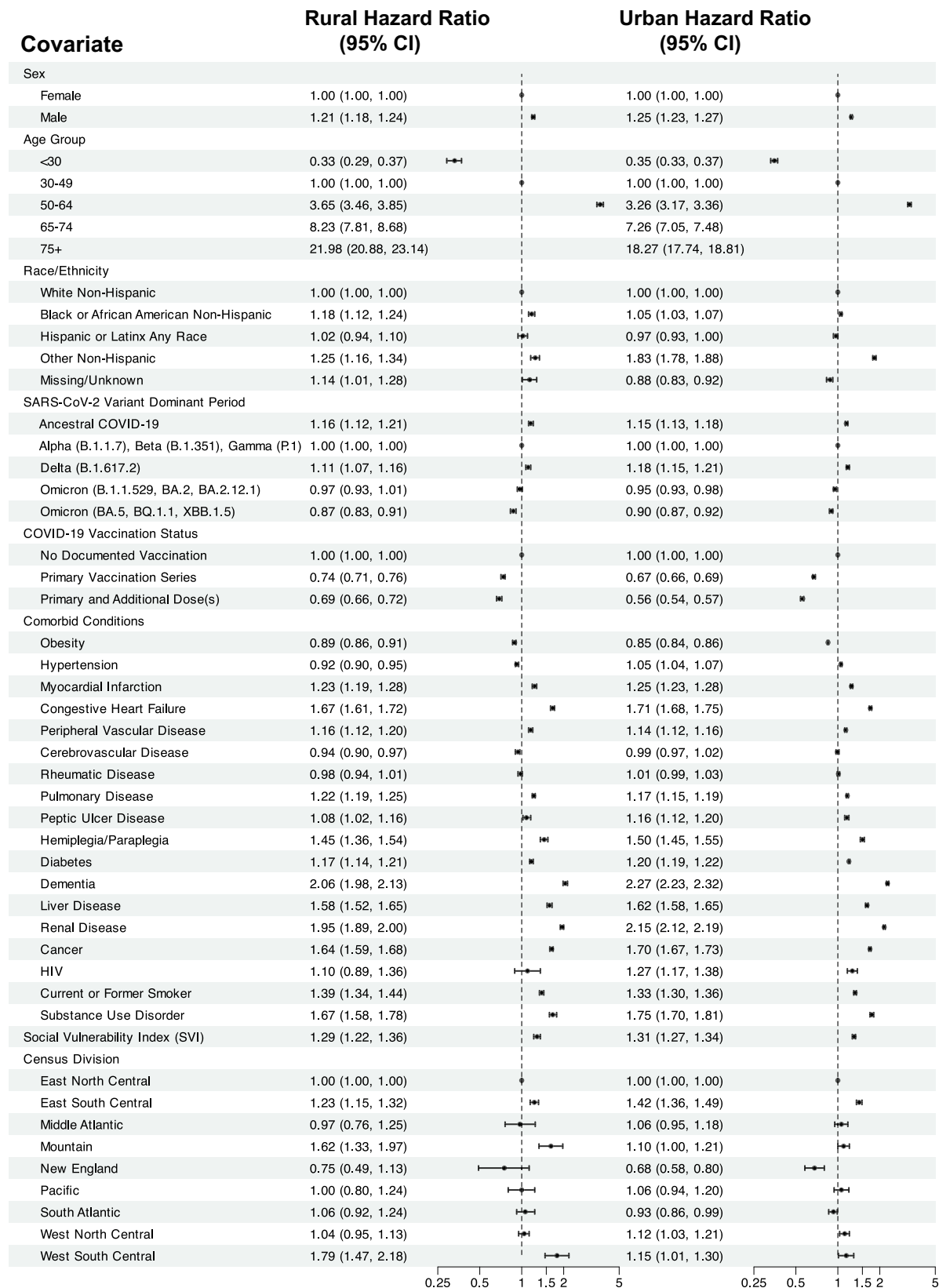
Reinfection as a Time-Varying Covariate

We also conducted a sensitivity analysis incorporating SARS-CoV-2 reinfection as a time-varying covariate to account for potential changes in risk over time (Supplementary Data 17). We observed 234,397

reinfections in our cohort with consistent findings to our main models. Rural dwellers had a similarly elevated mortality risk (UAR aHR 1.19 [1.17–1.21]; NAR aHR 1.26 [1.22–1.29]) relative to urban residents. SARS-CoV-2 reinfection was strongly associated with two-year mortality, with an aHR of 1.95 (1.89–2.00).

Discussion

Our study identified a rural-urban gradient in two-year all-cause mortality following SARS-CoV-2 infection, with adjusted risks 19% higher for urban-adjacent rural (UAR) and 26% higher for nonurban-adjacent



rural (NAR) dwellers than urban residents, even after accounting for confounding factors and community vulnerability. This increased risk of death was present for both patients requiring hospitalization for COVID-19 and those managed solely as outpatients. Previous research focusing on acute outcomes at the individual¹⁶ and population levels²⁰ has shown higher COVID-associated mortality in rural versus urban

areas. Our data expand upon this work by showing that the COVID-exacerbated rural mortality penalty persists for at least two years after adjusting for background risk factors.

The rural mortality penalty in the US is a significant public health issue marked by disparities in mortality rates between rural and urban populations. This phenomenon, exacerbated by healthcare access,

Fig. 7 | Adjusted Hazard Ratio for 2-Year Death Among Patients with an Acute SARS-CoV-2 Infection from April 2020 through December 2022 Stratified by Rural-Urban Residency. Figure 7 contains Cox Proportional Hazards models for all-cause mortality among patients with an acute SARS-CoV-2 infection from April 2020 through December 2022, stratified by rurality (rural residents on the left and urban residents on the right). Adjustments were made for sex, age, race/ethnicity, SARS-CoV-2 variant-dominant period at the time of infection, vaccination status, Social Vulnerability Index, U.S. Census Division, and a history of obesity, hypertension,

myocardial infarction, congestive heart failure, peripheral vascular disease, rheumatic disease cerebrovascular disease, dementia, chronic pulmonary disease, peptic ulcer disease, hemiplegia or paraplegia, diabetes, liver disease, renal disease, cancer, HIV, tobacco usage, and substance use disorder documented before COVID-19. Forest plots display hazard ratios (points) on a log scale, with horizontal lines representing 95% confidence intervals (error bars). The vertical line indicates the null value of 1.0. Full model specifications are available in Supplementary Data 8.

socioeconomic status, educational inequality, and lifestyle choices, has been the subject of much recent scholarship. Cosby et al. first described the evolution of the rural mortality penalty in 2008²¹, exploring the underlying causes and identifying factors such as healthcare access, economic challenges, and educational disparities as key contributors to this penalty³.

The rural mortality penalty varies significantly across different regions in the US, demonstrating diverse health outcomes², differences in life expectancy¹¹, and marked regional²² and socioeconomic disparities¹ across the rural spectrum. Additionally, several studies have analyzed lifespan variation across racial and ethnic groups and documented race-specific penalties, notably among rural Black and white populations²³.

These persistent disparities underscore the urgent need for targeted public health strategies in rural communities. The significant rural-urban divide regarding health risks emphasizes addressing the underlying factors contributing to these disparities, ensuring equitable healthcare access, and improving health outcomes for all populations, irrespective of geographical location. While studies have assessed the role of place-based mortality¹² and the characteristics of rural areas that contribute to higher mortality rates from preventable causes²⁴, quantifying a COVID-specific rural mortality penalty that persists beyond the immediate acute phase has not, to our knowledge, been established. Our study examines this and highlights a COVID-associated rural mortality penalty extending two years after initial infection.

We examined a COVID-19-negative comparison cohort of over 4.1 million individuals to further contextualize the observed rural mortality disparities. Within this cohort, rural dwellers had modestly higher two-year all-cause mortality compared to urban residents, with adjusted risks 17% higher for UAR and 21% higher for NAR. These findings align with known patterns of elevated mortality in rural populations. The rural mortality gradient was slightly more pronounced in the COVID-19 cohort (19% UAR and 26% NAR).

In a combined model that included COVID-19-positive and -negative individuals, we observed a modest interaction between rurality and COVID-19. In subsequent analysis, SARS-CoV-2 infection was associated with a 63% increased risk of two-year mortality, and rural-urban differences in mortality persisted. Although the relative hazard associated with rurality remained stable, the absolute number of deaths attributable to COVID-19 was higher among rural residents. We also found a similar risk for rurality in a sensitivity analysis with SARS-CoV-2 reinfection as a time-varying covariate. These findings indicate that while rural-urban mortality differences exist independently of COVID-19, infection with SARS-CoV-2 exacerbates this gap.

Prior findings have demonstrated that COVID-19 treatments are effective in both urban and rural dwellers, yet mortality at 90 days is 36% greater risk among rural residents¹⁸. This fact and the persistence of the mortality penalty for rural dwellers so long after surviving an episode of COVID-19 begs the question as to why this disparity persists. Behavioral factors such as vaccine hesitancy may also have contributed, as prior research^{25–27} has shown greater skepticism toward COVID-19 vaccination among rural residents. Understanding and addressing the reasons for this penalty may ultimately improve health outcomes in rural areas beyond the COVID-19 pandemic.

Rural health outcome disparities have been reported for multiple conditions in addition to COVID-19, including colorectal cancer²⁸ and

diabetes⁸. Rural residency presents unique challenges in care coordination²⁹. The lack of health care resources in rural areas is well documented. Even if rural dwellers can access primary care physicians and rehabilitation services, those professionals may lack access to information about care received at the referral center, leading to potentially worse outcomes.

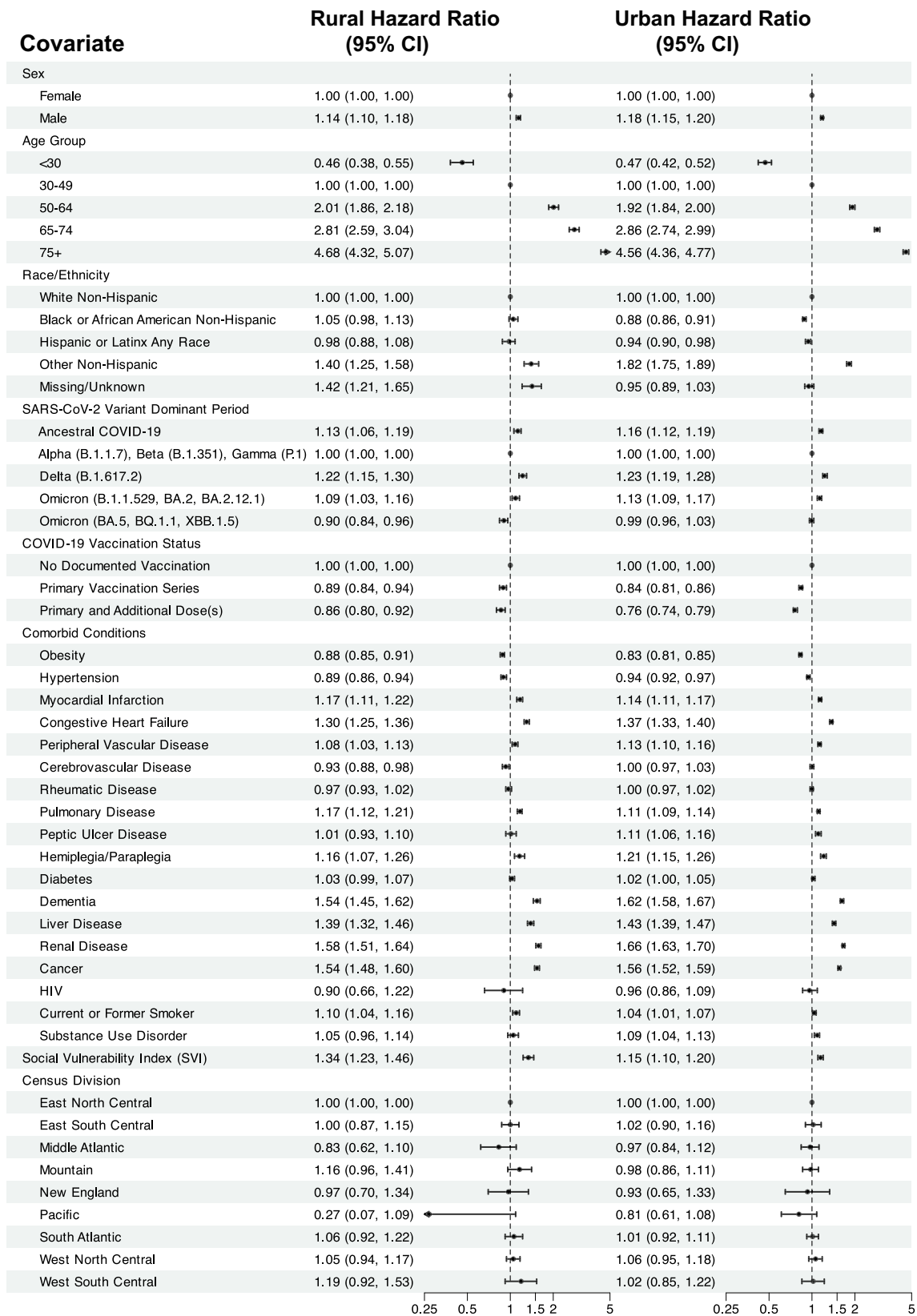
We found that outcomes are worse for rural versus urban dwellers across all time points post-COVID. Still, paradoxically, rural dwellers have lower rates of new diagnoses of other post-COVID conditions after their SARS-CoV-2 infection. The combination of lower rates of new diagnoses and higher mortality suggests that under-diagnosis may occur in the rural population. This under-diagnosis hypothesis is consistent with rural communities' well-known lack of healthcare resources compared to urban communities. While further work is needed to determine the factors contributing to the persistent mortality penalty suffered by survivors of acute COVID-19 in rural areas, the existence of this disparity highlights a broader crisis in rural healthcare.

Another important consideration in interpreting our findings is the role of social vulnerability. SVI is a composite of 16 variables in four broad areas (i.e., socioeconomic status, household characteristics, race/ethnicity minority status, and housing type and transportation)³⁰. Our models adjusted for social vulnerability, however, social vulnerability may not be entirely distinct from rurality. Social vulnerability may be one of the deeper factors driving the rural effect on mortality, meaning that our inclusion of social vulnerability likely led to underestimating the impact of rural residency on post-infection mortality. High social vulnerability, which reflects cumulative disadvantage across socioeconomic and demographic factors, was associated with significantly higher mortality risks, particularly among rural dwellers. Recent research demonstrates that counties with high social vulnerability have higher rates of chronic diseases³¹, reduced access to healthcare³², and more barriers to preventive care³³. The study underscores the compounded effects of living in a geographically isolated and socioeconomically disadvantaged area by adjusting for rurality and social vulnerability.

We also identified an important interaction between social vulnerability and rurality, further highlighting these compounded effects. However, this interaction did not substantially attenuate the rural effect when social vulnerability was excluded from the original models. As a result, we opted to present stratified models to better illustrate the differences in outcomes between rural and urban settings. The stratified analysis provides valuable insights into the differential impact of patient- and community-level factors on long-term mortality. Specifically, social vulnerability was associated with higher mortality in rural and urban dwellers, but higher in rural dwellers hospitalized with COVID-19.

Limitations

This study has several limitations. First, not all sites in N3C opt-in to PPRL, which means they did not have externally validated death records. Given that this study seeks to understand long-term differences in all-cause mortality, this resulted in the exclusion of approximately half of N3C-contributing sites that do not participate in PPRL for externally-validated mortality data³⁴. At the patient level, data on key variables – such as 5-digit ZIP codes, comorbidities, and



vaccination status – were missing for many patients in the N3C Enclave and were systematically missing from multiple sites. The reliance on EHR data, with its varying reporting practices across sites, introduces the potential for missingness bias, which is a well-known concern when dealing with real-world data³⁵. We mitigated this by excluding patients with insufficient pre-COVID visit histories and employing robust

statistical methods, including weighting and adjusting for site-specific random effects. However, unmeasured confounders, such as undocumented health behaviors and environmental exposures, may still influence our findings.

Second, referral-in bias poses another limitation, particularly for rural patients. This referral-in bias may take two forms. First, among

Fig. 8 | Adjusted Hazard Ratio for 2-Year Death Among Patients Hospitalized After an Acute SARS-CoV-2 Infection from April 2020 through December 2022 Stratified by Rural-Urban Residency. Figure 8 contains Cox Proportional Hazards models for all-cause mortality among patients hospitalized within two weeks (-3/+14 days) after an acute SARS-CoV-2 infection, stratified by rurality (rural residents on the left and urban residents on the right). Adjustments were made for sex, age, race/ethnicity, SARS-CoV-2 variant-dominant period at the time of infection, vaccination status, Social Vulnerability Index, U.S. Census Division, and a history of

obesity, hypertension, myocardial infarction, congestive heart failure, peripheral vascular disease, rheumatic disease cerebrovascular disease, dementia, chronic pulmonary disease, peptic ulcer disease, hemiplegia or paraplegia, diabetes, liver disease, renal disease, cancer, HIV, tobacco usage, and substance use disorder documented before COVID-19. Forest plots display hazard ratios (points) on a log scale, with horizontal lines representing 95% confidence intervals (error bars). The vertical line indicates the null value of 1.0. Full model specifications are available in Supplementary Data 9.

those rural patients with COVID-19 requiring hospitalization, the more severely ill may have a greater tendency to be transferred to referral tertiary centers with a greater likelihood of contributing data to N3C than local community hospitals, potentially caring for less severely ill or comorbid COVID-19 patients. N3C data-contributing sites potentially have an enrichment of more severe COVID-19 cases among rural-dwelling patients. Despite adjusting for multiple comorbidities and factors associated with the severity of illness, residual confounding may persist from this potential referral-in bias. Rural residents requiring specialized care are often transferred to urban centers, which may result in a second form of post-referral-in bias or effect. Specifically, being discharged from an urban care setting into a rural residence at discharge may affect healthcare access, care coordination, and follow-up, thus influencing long-term mortality outcomes. These factors may translate into a real disadvantage in rural dweller outcomes, which are explanatory rather than a bias from confounding or a study limitation. The paradoxically lower incidence of new diagnoses in follow-up among rural dwellers is consistent with this mechanistic supposition. Nonetheless, we attempted to address this by including adjustments for rurality and degree of rurality (UAR vs. NAR) and using site-specific random effects. Still, residual long-term mortality differences may remain due to explanatory disparities in healthcare access and delivery, adversely affecting rural compared to urban dwellers.

Third, although the N3C includes a diverse population from various U.S. regions, including equivalent numbers of rural versus urban dwellers as the U.S. population, it may not fully capture the experiences of all rural populations, particularly those in underrepresented areas or healthcare systems that did not participate in the N3C. This limits the generalizability of our findings, especially in regions with limited healthcare access, where mortality rates could be higher.

Lastly, while we adjusted for a wide range of demographic, clinical, and variant-period factors, residual confounding by socioeconomic status, health behaviors, and environmental exposures—factors often not captured in EHR data—may still contribute to the observed rural-urban disparities. These limitations highlight the complexity of understanding the full scope of rural health disparities and emphasize the need for future research that addresses these unmeasured factors.

Despite these limitations, this study had several strengths—a large sample size, adjustments for background risk, and an extensive follow-up period, including all-cause mortality linkage—that contribute to our understanding of the persistent disparities in long-term mortality between rural and urban populations following SARS-CoV-2 infection.

In conclusion, rural dwellers have significantly higher mortality than urban dwellers after SARS-CoV-2 infection, and this poorer prognosis persists for at least two years. The increased long-term mortality for rural dwellers, coupled with a paradoxically lower incidence of new diagnoses during the same period, underscores the critical need for improved healthcare access and follow-up care in rural communities. Greater social vulnerability was also associated with higher mortality, further highlighting the importance of tailored public health strategies to address ongoing health disparities in rural populations.

Methods

This retrospective cohort study utilized data from the National Clinical Cohort Collaborative (N3C) COVID-19 Enclave³⁶, a longitudinal

electronic health record (EHR) repository of SARS-CoV-2-infected individuals. N3C is a next-generation registry³⁷ containing EHR data contributed from sites across the U.S., including externally linked data through a process called privacy-preserving record linkage (PPRL)³⁴. This includes de-duplication of patients across sites, externally validated death records, and linkage with external patient records and registries, including GenBank-linked SARS-CoV-2 viral variants, Medicaid/Medicare claims data, and SEER registry data³⁸. The Regenstrief Institute serves as the linkage honest broker³⁹. Over half of N3C sites (N = 50) participate in PPRL, with 36 supporting external mortality linkage.

The University of Nebraska Medical Center's IRB (0176-21-EP) and the N3C Data Access Committee (RP-924490) approved the study, which adheres to RECORD reporting guidelines (<https://www.equator-network.org/reporting-guidelines/record/>). Additionally, each participating site's IRB also approved this study. N3C operates under the authority of the National Institutes of Health IRB, with Johns Hopkins University serving as the central IRB (IRB00309495). No informed consent was obtained from individual patients because the study used a limited data set among sites opting into PPRL with enhanced mortality linkage, stripping direct identifiers in compliance with the HIPAA Privacy Rule.

Cohort definition

Persons were classified based on rurality, identified by mapping 5-digit ZIP Codes to the 2010 Rural-Urban Commuting Area (RUCA) codes⁴⁰. We grouped RUCA codes into a binary urban-rural distinction (RUCA codes 1–3 are urban and 4–10 are rural). Further, we subclassified their degree of rurality into urban-adjacent rural (UAR) and nonurban-adjacent rural (NAR), using a previously described methodology based on commuting flows to urban clusters^{16,17}, which aligns with the Federal Office for Rural Health Policy guidance on rural-qualifying ZIP Codes⁴¹.

We included weighting or adjustments for age at COVID-19 diagnosis, sex, race/ethnicity, receipt of any COVID-19 vaccination before incident COVID-19 (categorized as no documented vaccination before infection or completion of a primary series (≥ 2 mRNA doses or 1 viral vector dose, with or without additional doses), tobacco usage, obesity (both diagnosed and identified by body mass index >30), and comorbid conditions diagnosed by a provider, including substance use disorder, obesity, hypertension, and individual categories from the Charlson Comorbidity Index (CCI)⁴². We also adjusted for community risk using the Social Vulnerability Index (SVI)³⁰, a composite measure of 15 social and environmental factors developed by the U.S. Centers for Disease Control and Prevention (CDC) and Agency for Toxic Substances and Disease Registry (ATSDR) mapped to U.S. residency at the county level, the nine U.S. Census divisions to understand the differential impact of rural residence by region⁴³, and differences across SARS-CoV-2 variant-dominant periods based on predominant strain from CDC reporting (see **Supplementary Methods** for CDC variant-dominant period ascertainment process and timing), which included ancestral COVID-19; Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1); Delta (B.1.617.2); early Omicron (B.1.1.529, BA.2, BA.2.12.1); and later Omicron (BA.5, BQ.1.1, XBB.1.5)⁴⁴.

These variables were selected a priori based on established associations with COVID-19 outcomes and their relevance to post-acute mortality risk. Individual-level variables such as age,

comorbidities, and vaccination status were modeled at the patient level, while ecological variables such as SVI and Census Division were assigned based on ZIP Code or county of residence.

Our primary cohort was all patients infected during the study period (April 2020 – December 2022). To understand differences based on initial COVID-19 severity (i.e., patients hospitalized for COVID-19), we also assessed mortality differences among patients hospitalized within -3 to +14 days of their initial SARS-CoV-2 infection. This timeframe was selected to account for possible delays in post-admission testing and to include the typical incubation period of the virus⁴⁵, thereby capturing hospitalizations directly associated with the acute phase of COVID-19. We excluded patients with no pre-COVID-19 visit history (required to understand pre-COVID risk factors), without a 5-digit ZIP Code, pediatric patients (<19 years at infection date), centenarians (100+ years at infection date), and those from sites not participating in PPRL (required to assess deaths occurring outside the contributing hospital system) or with limited data reporting in the condition domain. These exclusions were primarily due to limitations in site-level reporting, as many sites did not provide key structured data, such as ZIP Code or externally verified mortality data, necessary for cohort construction. We include detailed inclusion and exclusion criteria and site characteristics in the Supplementary Methods.

Outcome definition

We analyzed two-year all-cause mortality among patients infected with SARS-CoV-2 between April 1, 2020, and December 31, 2022, with follow-up until December 31, 2024. The data were extracted from N3C on February 2, 2025 (N3C Release 188). Mortality was assessed at intervals after initial SARS-CoV-2 infection at one month (30 days), three months (91 days), one year (365 days), and two years (730 days), using hospital-documented deaths and enhanced mortality records using PPRL³⁴ that enabled linkage with the Social Security Administration's Death Master File.

We employed a piecewise landmark analysis to explore long-term mortality trends further, focusing on patients who survived to the end of each critical time point: month one (>30 days), month three (>91 days), and year one (>365 days) post-infection. This method was chosen to provide a more detailed understanding of mortality risk among those who survived the immediate acute phase of COVID-19. By segmenting the analysis into these distinct intervals, we aimed to capture potential changes in mortality risk over time, particularly concerning rural residency, which may influence outcomes due to varying healthcare access and social determinants of health.

This piecewise approach complements the primary Cox proportional hazards models by allowing us to focus on specific post-acute phases, where the risk profile may differ significantly from the acute phase. We applied this granular analysis to the overall cohort and specifically to those hospitalized during the acute infection phase, providing insights into the persistence of risk in different subgroups over time.

Secondary outcomes

We explored new-onset intermediate events that occurred after SARS-CoV-2 infection but before death among those dying in the two years after COVID-19. New onset post-COVID-19 conditions were classified as those not documented using the same condition code in the pre-COVID period and occurring between SARS-CoV-2 infection and death. We assessed for SARS-CoV-2 reinfection (2 PCR or Ag tests at least 90 days apart), long COVID (ICD-10 U09.9), and the following phenotypic abnormalities using the Human Phenotype Ontology⁴⁶ mapped to OMOP: respiratory system, cardiovascular system, immune system, nervous system, metabolism homeostasis, blood and blood-forming tissues, digestive system, endocrine system, musculoskeletal system, integument, and genitourinary system as well as constitutional symptoms and neoplasms⁴⁷.

Statistical analysis

Mortality differences by rurality were assessed using Kaplan-Meier survival analysis with a log-rank test. We calculated mortality risks per 100,000 persons at 1-, 3-, 12-, and 24-months post-infection. Risks were calculated as cumulative incidence proportions, and 95% confidence intervals (95% CI) were derived under a Poisson assumption for the event count. Excess mortality was expressed as the risk difference between each rural-urban group and the overall cohort, standardized using the 2020 U.S. Standard Population. We report crude and age-adjusted rates using the 2020 U.S. Standard Population to account for differences in age distribution across rural-urban groups⁴⁸. We also report event rates per person-months in the **Supplementary Data**.

We assessed associations between rurality and two-year mortality using multivariable Cox Proportional Hazards with inverse probability treatment weighting (IPW). Stabilized weights were calculated as the marginal probability of each rural group divided by the predicted probability from a multinomial logistic regression model. Weighting included age group, sex, race/ethnicity, SARS-CoV-2 variant-dominant period, and SARS-CoV-2 vaccination status. We chose these variables to represent universal confounders of rural residence and mortality risk, while avoiding covariates that might reflect downstream effects or inherent components of rurality itself (e.g., comorbid burden, social vulnerability). These downstream variables are potential mediators, not colliders, as they may be on the pathway between rurality and mortality. We excluded them from the propensity score model to avoid blocking these pathways in total-effect estimation, but included them in the outcome model when estimating the direct effect. This approach also maintains the temporal order between exposure and confounders, preventing the adjustment of meaningful rural health disparities.

In the weighted regression model, we adjusted for additional variables excluded from the propensity score model, including comorbid conditions, SVI, and Census Division, to estimate the direct association of rurality with mortality conditional on these pathways. This specification avoids overlap in covariates between the weighting and regression stages and maintains the doubly robust property⁴⁹, whereby the association is consistently estimated if either the propensity score or the outcome model is correctly specified. We included a random effect for the data-contributing site to account for data-reporting differences.

We re-analyzed all tests in a cohort using a dichotomous rural-urban variable to assess the robustness of the findings with a more stable sample size across all periods. Additionally, we performed a sensitivity analysis that accounted for multiple infections by treating reinfections as a time-varying covariate in a case-level approach instead of a person-level approach based on the initial infection.

To assess whether the effect of rurality on mortality is specific to COVID-19, consistent across COVID-positive and COVID-negative cohorts, or exacerbated by COVID-19, we conducted additional analyses using a cohort of COVID-negative individuals from the same data-contributing sites. COVID-negative individuals were defined as those without a documented SARS-CoV-2 infection during the study period, with index dates assigned to align with the distribution of first positive dates in the primary cohort. We analyzed two-year mortality within the COVID-negative cohort alone and in a combined model including both cohorts, incorporating COVID-19 status as a covariate. To formally assess effect modification, we included an interaction term between rurality and COVID-19 status in the combined model. All methods and adjustments were consistent with the main analysis (excluding COVID-19 vaccinations and variant-dominant period), allowing for a comparison of rurality's impact across cohorts.

The statistical tests performed were 2-sided at a significance level of $p < .05$ and presented with 95% confidence intervals. The **Supplementary Methods** describe methodological details, including all computable phenotype definitions and an overview of the N3C environment and sampling approach. Supplementary Data 1-15 and Figs. 1-7

present full model specifications and sensitivity analyses. Statistical analyses were performed in R v4.1.3, Python, and SQL within the N3C Enclave, which is available on GitHub⁵⁰.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The N3C Enclave is available for public research use. More than 4000 researchers have access to data in N3C, representing more than 300 US research institutions. Institutions must have a signed Data Use Agreement executed with the U.S. National Center for Advancing Translational Sciences (NCATS) to access data. Investigators must complete mandatory training and submit a Data Use Request (DUR) to N3C. To request N3C data access, follow the instructions at (<https://covid.cd2h.org/onboarding>). This project utilized data from N3C release 188, which can be replicated within the Enclave by qualified N3C users. All concepts and definitions are provided in the **eMethods**.

Code availability

All code used for statistical analyses is available in a public GitHub repository: (<https://github.com/National-Clinical-Cohort-Collaborative/CS-Rural-Health/tree/main/short-and-long-term-mortality>). This includes a minimally reproducible example using synthetic data and all concept sets used for clinical definitions throughout the study.

References

- Gong, G., Phillips, S. G., Hudson, C., Curti, D. & Philips, B. U. Higher US rural mortality rates linked to socioeconomic status, physician shortages, and lack of health insurance. *Health Aff. (Millwood)* **38**, 2003–2010 (2019).
- James, W. L. All rural places are not created equal: revisiting the rural mortality penalty in the United States. *Am. J. Public Health* **104**, 2122–2129 (2014).
- Cossmann, J. S., James, W. L., Cosby, A. G. & Cossmann, R. E. Underlying causes of the emerging nonmetropolitan mortality penalty. *Am. J. Public Health* **100**, 1417–1419 (2010).
- Callaghan, T. et al. Rural healthy people 2030: New decade, new challenges. *Prev. Med Rep.* **33**, 102176 (2023).
- Lawrence-Bourne, J. et al. What is rural adversity, how does it affect wellbeing and what are the implications for action? *Int. J. Environ. Res. Public Health* **17**, 7205 (2020).
- Stack, E. et al. The impacts of COVID-19 on mental health, substance use, and overdose concerns of people who use drugs in rural communities. *J. Addict. Med.* **15**, 383–389 (2021).
- Cohen, S. A., Cook, S. K., Kelley, L., Foutz, J. D. & Sando, T. A. A Closer look at rural-urban health disparities: associations between obesity and rurality vary by geospatial and sociodemographic factors. *J. Rural Health* **33**, 167–179 (2017).
- Dugani, S. B., Mielke, M. M. & Vella, A. Burden and management of type 2 diabetes in rural United States. *Diab. Metab. Res Rev.* **37**, e3410 (2021).
- Probst, J. C., Moore, C. G., Baxley, E. G. & Lammie, J. J. Rural-urban differences in visits to primary care physicians. *Fam. Med.* **34**, 609–615 (2002).
- Cross, S. H., Califf, R. M. & Warraich, H. J. Rural-urban disparity in mortality in the US from 1999 to 2019. *Jama* **325**, 2312–2314 (2021).
- Walker, B. H. & Brown, D. C. Trends in lifespan variation across the spectrum of rural and urban places in the United States, 1990–2017. *SSM Popul Health* **19**, 101213 (2022).
- Cosby, A. G. et al. Growth and persistence of place-based mortality in the United States: the rural mortality penalty. *Am. J. Public Health* **109**, 155–162 (2019).
- Jones, M., Bhattar, M., Henning, E. & Monnat, S. M. Explaining the U.S. rural disadvantage in COVID-19 case and death rates during the Delta-Omicron surge: The role of politics, vaccinations, population health, and social determinants. *Soc. Sci. Med.* **335**, 116180 (2023).
- Jeffers, A. et al. Impact of social isolation during the COVID-19 pandemic on mental health, substance use, and homelessness: qualitative interviews with behavioral health providers. *Int. J. Environ. Res. Public Health* **19**, 12120 (2022).
- Slonim, A. D., See, H. & Slonim, S. Challenges confronting rural hospitals accentuated during COVID-19. *J. Biomed. Res* **34**, 397–409 (2020).
- Anzalone, A. J. et al. Higher hospitalization and mortality rates among SARS-CoV-2-infected persons in rural America. *J. Rural Health* **39**, 39–54 (2022).
- Anzalone, A. J. et al. Community risks for SARS-CoV-2 infection among fully vaccinated US adults by rurality: a retrospective cohort study from the National COVID Cohort Collaborative. *PLoS One* **18**, e0279968 (2023).
- Anzalone, A. J. et al. Associations between COVID-19 therapies and outcomes in rural and urban America: A multisite, temporal analysis from the alpha to omicron SARS-CoV-2 variants. *J. Rural Health*, **41**, e12857 (2024).
- Bureau, U. C. et al. Nation's urban and rural populations shift following 2020 census. CB22-CN.25. <https://www.census.gov/newsroom/pressreleases/2022/urban-rural-populations.html> (2022).
- Lundberg, D. J. et al. COVID-19 mortality by race and ethnicity in US metropolitan and nonmetropolitan areas, march 2020 to february 2022. *JAMA Netw. Open* **6**, e2311098 (2023).
- Cosby, A. G. et al. Preliminary evidence for an emerging non-metropolitan mortality penalty in the United States. *Am. J. Public Health* **98**, 1470–1472 (2008).
- Monnat, S. M. Trends in U.S. Working-age non-hispanic white mortality: rural-urban and within-rural differences. *Popul Res Policy Rev.* **39**, 805–834 (2020).
- Cossmann, J., James, W. & Wolf, J. K. The differential effects of rural health care access on race-specific mortality. *SSM Popul Health* **3**, 618–623 (2017).
- Roth, A. R., Denney, J. T., Amiri, S. & Amram, O. Characteristics of place and the rural disadvantage in deaths from highly preventable causes. *Soc. Sci. Med.* **245**, 112689 (2020).
- Saelee, R. et al. Disparities in COVID-19 vaccination coverage between urban and rural counties - United States, December 14, 2020–January 31, 2022. *MMWR Morb. Mortal. Wkly Rep.* **71**, 335–340 (2022).
- Callaghan, T. et al. Correlates and disparities of COVID-19 vaccine hesitancy. Available at SSRN 3667971 (2020).
- Fridman, A., Gershon, R. & Gneezy, A. COVID-19 and vaccine hesitancy: a longitudinal study. *PLoS One* **16**, e0250123 (2021).
- Zahnd, W. E. et al. Rural-urban and racial/ethnic trends and disparities in early-onset and average-onset colorectal cancer. *Cancer* **127**, 239–248 (2021).
- Garvin, L. A., Pugatch, M., Gurewich, D., Pendergast, J. N. & Miller, C. J. Interorganizational care coordination of rural veterans by veterans affairs and community care programs: a systematic review. *Med Care* **59**, S259–s269 (2021).
- Flanagan, B. E., Gregory, E. W., Hallisey, E. J., Heitgerd, J. L. & Lewis, B. A social vulnerability index for disaster management. *J. Homeland Security And Emergency Management* **8** (2011).
- Ibrahim, R. et al. Impact of social vulnerability and demographics on ischemic heart disease mortality in the United States. *JACC Adv.* **2**, 100577 (2023).
- Mayhew, M. et al. Social deprivation, the area deprivation Index, and emergency department utilization within a community-based

- primary and preventive care program at a Florida medical school. *J. Public Health* **32**, 827–835 (2024).
33. Bauer, C. et al. County-level social vulnerability and breast, cervical, and colorectal cancer screening rates in the US, 2018. *JAMA Netw. Open* **5**, e2233429 (2022).
 34. Tachinardi, U. et al. Privacy-preserving record linkage across disparate institutions and datasets to enable a learning health system: the national COVID cohort collaborative (N3C) experience. *Learning Health Syst.* **8**, e10404 (2024).
 35. Groenwold, R. H. H. Informative missingness in electronic health record systems: the curse of knowing. *Diagn. Progn. Res* **4**, 8 (2020).
 36. Haendel, M. A. et al. The National COVID Cohort Collaborative (N3C): rationale, design, infrastructure, and deployment. *J. Am. Med. Inf. Assoc.* **28**, 427–443 (2021).
 37. Labkoff, S. E., Quintana, Y. & Rozenblit, L. Identifying the capabilities for creating next-generation registries: a guide for data leaders and a case for “registry science”. *J. Am. Med. Inform. Assoc.* **31**, 1001–1008 (2024).
 38. N3C Privacy-Preserving Record Linkage: National COVID Cohort Collaborative (N3C); 2024 [April 1, 2024]. Available from: <https://covid.cd2h.org/PPRL/>.
 39. PPRL Linkage Honest Broker: Regenstrief Institute; 2024 [updated May 17, 2024/May 18, 2024]. Available from: <https://www.regenstrief.org/n3c-lhb/>.
 40. Rural Urban Commuting Area Codes: US Department of Agriculture – Economic Research Service; 2010. Available from: <https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/>.
 41. Defining Rural Population Guidance Portal: U.S. Department of Health and Human Services; (2020). [updated 06/25/2020]. Available from: <https://www.hhs.gov/guidance/document/defining-rural-population>.
 42. Charlson, M. E., Pompei, P., Ales, K. L. & MacKenzie, C. R. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J. chronic Dis.* **40**, 373–383 (1987).
 43. Racine, R. et al. Profound racial disparities in COVID-19 associated hospitalizations in rural Southwest Georgia. *Am J Med Sci.* **364**, 1–6 (2022).
 44. COVID data tracker: monitoring variant proportions [Internet]. Centers for Disease Control and Prevention. 2024 [cited October 4, 2024]. Available from: <https://covid.cdc.gov/covid-data-tracker/#variantproportions>.
 45. Elias, C., Sekri, A., Leblanc, P., Cucherat, M. & Vanhems, P. The incubation period of COVID-19: A meta-analysis. *Int J. Infect. Dis.* **104**, 708–710 (2021).
 46. Gargano, M. A. et al. The human phenotype ontology in 2024: phenotypes around the world. *Nucleic Acids Res* **52**, D1333–d1346 (2024).
 47. Callahan, T., Hunter, L. & Kahn, M. Leveraging a neural-symbolic representation of biomedical knowledge to improve pediatric subphenotyping. Preprint at <https://doi.org/10.5281/zenodo.5746187> (2021).
 48. Anderson, R. N. & Rosenberg, H. M. Age standardization of death rates; implementation of the year 2000 standard. *Natl. Vital. Stat. Rep.* **47**, 1–16 (1998).
 49. Bang, H. & Robins, J. M. Doubly robust estimation in missing data and causal inference models. *Biometrics* **61**, 962–973 (2005).
 50. A. Jerrod Anzalone, et al. *Higher Mortality Following SARS-CoV-2 Infection in Rural Versus Urban Dwellers Persists for Two Years Post-Infection GitHub Repository*. (Rural Health Domain Team, 2024).
- (All). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.
- N3C Attribution:** The analyses described in this publication were conducted with data or tools accessed through the NCATS N3C Data Enclave (<https://covid.cd2h.org>) and N3C Attribution & Publication Policy v 1.2-2020-08-25b supported by NCATS U24 TR002306, Axle Informatics Subcontract: NCATS-P00438-B. This research was possible because of the patients whose information is included within the data and the organizations (<https://ncats.nih.gov/n3c/resources/data-contribution/data-transfer-agreement-signatories>) and scientists who have contributed to the on-going development of this community resource (<https://doi.org/10.1093/jamia/ocaa196>).
- Disclaimer** The N3C Publication Committee confirmed that this manuscript MSID:2128.383 is in accordance with N3C data use and attribution policies; however, this content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the N3C program.
- Individual Acknowledgements for Core Contributors:** We gratefully acknowledge the following core contributors to N3C: Adam B. Wilcox, Adam M. Lee, Alexis Graves, Alfred (Jerrod) Anzalone, Amin Manna, Amit Saha, Amy Olex, Andrea Zhou, Andrew E. Williams, Andrew Southerland, Andrew T. Girvin, Anita Walden, Anjali A. Sharathkumar, Benjamin Amor, Benjamin Bates, Brian Hendricks, Brijesh Patel, Caleb Alexander, Carolyn Bramante, Cavin Ward-Caviness, Charisse Madlock-Brown, Christine Suver, Christopher Chute, Christopher Dillon, Chunlei Wu, Clare Schmitt, Cliff Takemoto, Dan Housman, Davera Gabriel, David A. Eichmann, Diego Mazzotti, Don Brown, Eilis Boudreau, Elaine Hill, Elizabeth Zampino, Emily Carlson Marti, Emily R. Pfaff, Evan French, Farrukh M Korashy, Federico Mariona, Fred Prior, George Sokos, Greg Martin, Harold Lehmann, Heidi Spratt, Hemalkumar Mehta, Hongfang Liu, Hythem Sidky, J.W. Awori Hayanga, Jami Pincavitch, Jaylyn Clark, Jeremy Richard Harper, Jessica Islam, Jin Ge, Joel Gagnier, Joel H. Saltz, Joel Saltz, Johanna Loomba, John Buse, Jomol Mathew, Joni L. Rutter, Julie A. McMurry, Justin Guinney, Justin Starren, Karen Crowley, Katie Rebecca Bradwell, Kellie M. Walters, Ken Wilkins, Kenneth R. Gersing, Kenrick Dwain Cato, Kimberly Murray, Kristin Kostka, Lavance Northington, Lee Allan Pyles, Leonie Misquitta, Lesley Cottrell, Lili Portilla, Mariam Deacy, Mark M. Bissell, Marshall Clark, Mary Emmett, Mary Morrison Saltz, Matvey B. Palchuk, Melissa A. Haendel, Meredith Adams, Meredith Temple-O’Connor, Michael G. Kurilla, Michele Morris, Nabeel Qureshi, Nasia Safdar, Nicole Garbarini, Noha Sharafeldin, Ofer Sadan, Patricia A. Francis, Penny Wung Burgoon, Peter Robinson, Philip R.O. Payne, Rafael Fuentes, Randeep Jawa, Rebecca Erwin-Cohen, Rena Patel, Richard A. Moffitt, Richard L. Zhu, Rishi Kamaleswaran, Robert Hurley, Robert T. Miller, Saiju Pyarajan, Sam G. Michael, Samuel Bozette, Sandeep Mallipattu, Satyanarayana Vedula, Scott Chapman, Shawn T. O’Neil, Soko Setoguchi, Stephanie S. Hong, Steve Johnson, Tellen D. Bennett, Tiffany Callahan, Umit Topaloglu, Usman Sheikh, Valery Gordon, Vignesh Subbian, Warren A. Kibbe, Wendy Hernandez, Will Beasley, Will Cooper, William Hillegass, Xiaohan Tanner Zhang. Details of contributions available at covid.cd2h.org/core-contributors.
- Data Partners with Released Data:** The following institutions whose data is released or pending: Available: Advocate Health Care Network—UL1TR002389: The Institute for Translational Medicine (ITM) • Aurora Health Care Inc—UL1TR002373: Wisconsin Network For Health Research • Boston University Medical Campus—UL1TR001430: Boston University Clinical and Translational Science Institute • Brown University—U54GM115677: Advance Clinical Translational Research (Advance-CTR) • Carilion Clinic—UL1TR003015: iTHRIV Integrated Translational health Research Institute of Virginia • Case Western Reserve University—UL1TR002548: The Clinical & Translational Science Collaborative of Cleveland (CTSC) • Charleston Area Medical Center—U54GM104942: West Virginia Clinical and Translational Science Institute (WVCTSI) • Children’s Hospital Colorado—UL1TR002535: Colorado Clinical and Translational Sciences Institute • Columbia University Irving Medical

Acknowledgements

The project described was supported by the National Institute of General Medical Sciences, U54GM104942 (BP, GSM, SLH), U54GM115458 (AJA, MS), U54GM104940 (RH, SC, LM), U54GM104938 (All), U54GM115516 (CJS, SLS), U54GM115677 (All), U54GM115428 (WBH), and U54GM104941

Center—UL1TR001873: Irving Institute for Clinical and Translational Research • Dartmouth College — None (Voluntary) Duke University—UL1TR002553: Duke Clinical and Translational Science Institute • George Washington Children’s Research Institute—UL1TR001876: Clinical and Translational Science Institute at Children’s National (CTSA-CN) • George Washington University—UL1TR001876: Clinical and Translational Science Institute at Children’s National (CTSA-CN) • Harvard Medical School—UL1TR002541: Harvard Catalyst • Indiana University School of Medicine—UL1TR002529: Indiana Clinical and Translational Science Institute • Johns Hopkins University—UL1TR003098: Johns Hopkins Institute for Clinical and Translational Research • Louisiana Public Health Institute—None (Voluntary) • Loyola Medicine—Loyola University Medical Center • Loyola University Medical Center—UL1TR002389: The Institute for Translational Medicine (ITM) • Maine Medical Center—U54GM115516: Northern New England Clinical & Translational Research (NNE-CTR) Network • Mary Hitchcock Memorial Hospital & Dartmouth Hitchcock Clinic—None (Voluntary) • Massachusetts General Brigham—UL1TR002541: Harvard Catalyst • Mayo Clinic Rochester—UL1TR002377: Mayo Clinic Center for Clinical and Translational Science (CCaTS) • Medical University of South Carolina—UL1TR001450: South Carolina Clinical & Translational Research Institute (SCTR) • MITRE Corporation—None (Voluntary) • Montefiore Medical Center—UL1TR002556: Institute for Clinical and Translational Research at Einstein and Montefiore • Nemours—U54GM104941: Delaware CTR ACCEL Program • NorthShore University HealthSystem—UL1TR002389: The Institute for Translational Medicine (ITM) • Northwestern University at Chicago—UL1TR001422: Northwestern University Clinical and Translational Science Institute (NUCATS) • OCHIN—INV-018455: Bill and Melinda Gates Foundation grant to Sage Bionetworks • Oregon Health & Science University—UL1TR002369: Oregon Clinical and Translational Research Institute • Penn State Health Milton S. Hershey Medical Center—UL1TR002014: Penn State Clinical and Translational Science Institute • Rush University Medical Center—UL1TR002389: The Institute for Translational Medicine (ITM) • Rutgers, The State University of New Jersey—UL1TR003017: New Jersey Alliance for Clinical and Translational Science • Stony Brook University—U24TR002306 • The Alliance at the University of Puerto Rico, Medical Sciences Campus—U54GM133807: Hispanic Alliance for Clinical and Translational Research (The Alliance) • The Ohio State University—UL1TR002733: Center for Clinical and Translational Science • The State University of New York at Buffalo—UL1TR001412: Clinical and Translational Science Institute • The University of Chicago—UL1TR002389: The Institute for Translational Medicine (ITM) • The University of Iowa—UL1TR002537: Institute for Clinical and Translational Science • The University of Miami Leonard M. Miller School of Medicine—UL1TR002736: University of Miami Clinical and Translational Science Institute • The University of Michigan at Ann Arbor—UL1TR002240: Michigan Institute for Clinical and Health Research • The University of Texas Health Science Center at Houston—UL1TR003167: Center for Clinical and Translational Sciences (CCTS) • The University of Texas Medical Branch at Galveston—UL1TR001439: The Institute for Translational Sciences • The University of Utah—UL1TR002538: Uhealth Center for Clinical and Translational Science • Tufts Medical Center—UL1TR002544: Tufts Clinical and Translational Science Institute • Tulane University—UL1TR003096: Center for Clinical and Translational Science • The Queens Medical Center—None (Voluntary) • University Medical Center New Orleans—U54GM104940: Louisiana Clinical and Translational Science (LA CaTS) Center • University of Alabama at Birmingham—UL1TR003096: Center for Clinical and Translational Science • University of Arkansas for Medical Sciences—UL1TR003107: UAMS Translational Research Institute • University of Cincinnati—UL1TR001425: Center for Clinical and Translational Science and Training • University of Colorado Denver, Anschutz Medical Campus—UL1TR002535: Colorado Clinical and Translational Sciences Institute • University of Illinois at Chicago—UL1TR002003: UIC Center for Clinical and Translational Science • University of Kansas Medical Center—UL1TR002366: Frontiers: University of Kansas Clinical and Translational Science Institute • University of Kentucky—UL1TR001998: UK Center for Clinical and Translational Science • University of Massachusetts Medical School Worcester—UL1TR001453: The UMass Center for Clinical and Translational Science (UMCCTS) • University Medical Center of Southern Nevada—None (voluntary) • University of Minnesota—UL1TR002494: Clinical and Translational Science Institute • University of Mississippi Medical Center—U54GM115428: Mississippi Center for Clinical and Translational Research (CCTR) • University of Nebraska Medical Center—U54GM115458: Great Plains IDEa-Clinical & Translational Research • University of North Carolina at Chapel Hill—UL1TR002489: North Carolina Translational and Clinical Science Institute • University of Oklahoma Health Sciences Center—U54GM104938: Oklahoma Clinical and Translational Science Institute (OCTSI) • University of Pittsburgh—UL1TR001857: The Clinical and Translational Science Institute (CTSI) • University of Pennsylvania—UL1TR001878: Institute for Translational Medicine and Therapeutics • University of Rochester—UL1TR002001: UR Clinical & Translational Science Institute • University of Southern California—UL1TR001855: The Southern California Clinical and Translational Science Institute (SC CTSI) • University of Vermont—U54GM115516: Northern New England Clinical & Translational Research (NNE-CTR) Network • University of Virginia—UL1TR003015: iTHRIV Integrated Translational health Research Institute of Virginia • University of Washington—UL1TR002319: Institute of Translational Health Sciences • University of Wisconsin-Madison—UL1TR002373: UW Institute for Clinical and Translational Research • Vanderbilt University Medical Center—UL1TR002243: Vanderbilt Institute for Clinical and Translational Research • Virginia Commonwealth University—UL1TR002649: C. Kenneth and Dianne Wright Center for Clinical and Translational Research • Wake Forest University Health Sciences—UL1TR001420: Wake Forest Clinical and Translational Science Institute • Washington University in St. Louis—UL1TR002345: Institute of Clinical and Translational Sciences • Weill Medical College of Cornell University—UL1TR002384: Weill Cornell Medicine Clinical and Translational Science Center • West Virginia University—U54GM104942: West Virginia Clinical and Translational Science Institute (WVCTSI) Submitted: Icahn School of Medicine at Mount Sinai—UL1TR001433: ConduITS Institute for Translational Sciences • The University of Texas Health Science Center at Tyler—UL1TR003167: Center for Clinical and Translational Sciences (CCTS) • University of California, Davis—UL1TR001860: UC Davis Health Clinical and Translational Science Center • University of California, Irvine—UL1TR001414: The UC Irvine Institute for Clinical and Translational Science (ICTS) • University of California, Los Angeles—UL1TR001881: UCLA Clinical Translational Science Institute • University of California, San Diego—UL1TR001442: Altman Clinical and Translational Research Institute • University of California, San Francisco—UL1TR001872: UCSF Clinical and Translational Science Institute Pending: Arkansas Children’s Hospital—UL1TR003107: UAMS Translational Research Institute • Baylor College of Medicine—None (Voluntary) • Children’s Hospital of Philadelphia—UL1TR001878: Institute for Translational Medicine and Therapeutics • Cincinnati Children’s Hospital Medical Center—UL1TR001425: Center for Clinical and Translational Science and Training • Emory University—UL1TR002378: Georgia Clinical and Translational Science Alliance • HonorHealth—None (Voluntary) • Loyola University Chicago—UL1TR002389: The Institute for Translational Medicine (ITM) • Medical College of Wisconsin—UL1TR001436: Clinical and Translational Science Institute of Southeast Wisconsin • MedStar Health Research Institute—None (Voluntary) • Georgetown University—UL1TR001409: The Georgetown-Howard Universities Center for Clinical and Translational Science (GHUCCTS) • MetroHealth—None (Voluntary) • Montana State University—U54GM115371: American Indian/Alaska Native CTR • NYU Langone Medical Center—UL1TR001445: Langone Health’s Clinical and Translational Science Institute • Ochsner Medical Center—U54GM104940: Louisiana Clinical and Translational Science (LA CaTS) Center • Regenstrief Institute—UL1TR002529: Indiana Clinical and

Translational Science Institute • Sanford Research—None (Voluntary) • Stanford University—UL1TR003142: Spectrum: The Stanford Center for Clinical and Translational Research and Education • The Rockefeller University—UL1TR001866: Center for Clinical and Translational Science • The Scripps Research Institute—UL1TR002550: Scripps Research Translational Institute • University of Florida—UL1TR001427: UF Clinical and Translational Science Institute • University of New Mexico Health Sciences Center—UL1TR001449: University of New Mexico Clinical and Translational Science Center • University of Texas Health Science Center at San Antonio—UL1TR002645: Institute for Integration of Medicine and Science • Yale New Haven Hospital—UL1TR001863: Yale Center for Clinical Investigation.

Author contributions

Concept and design: A.A. Acquisition, analysis, or interpretation of data: A.A., M.E.S., B.P., R.H., and S.C. Drafting of the manuscript: A.A., M.V., and S.H. Interpretation of Clinical Findings: S.H., M.V., L.M., C.R., W.H., S.S., G.S. Critical revision of the manuscript for important intellectual content: A.A., M.V., M.E.S., B.P., W.H., R.H., S.C., C.R., L.M., S.S., G.S., and S.H. Obtained funding: S.H. Administrative, technical, editorial, or material support: A. Anzalone and the N3C Rural Health Domain Team Supervision: S. Hodder. All authors approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-63944-8>.

Correspondence and requests for materials should be addressed to A. Jerrod Anzalone.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

the National Clinical Cohort Collaborative (N3C) Consortium

Christopher G. Chute⁹

⁹Johns Hopkins University, Baltimore, USA.