Journal of the American Heart Association

ORIGINAL RESEARCH

Enriching the American Heart Association COVID-19 Cardiovascular Disease Registry Through Linkage With External Data Sources: Rationale and Design

Andrew S. Oseran, MD, MBA*; Tianyu Sun, PhD*; Rishi K. Wadhera, MD, MPP, MPhil; Issa J. Dahabreh, MD, ScD; James A. de Lemos, MD; Sandeep R. Das, MD, MPH, MBA; Christine Rutan, BS; Aarti H. Asnani, MD; Robert W. Yeh, MD, MSc†; Dhruv S. Kazi, MD, MSc, MS†

BACKGROUND: The AHA Registry (American Heart Association COVID-19 Cardiovascular Disease Registry) captures detailed information on hospitalized patients with COVID-19. The registry, however, does not capture information on social determinants of health or long-term outcomes. Here we describe the linkage of the AHA Registry with external data sources, including fee-for-service (FFS) Medicare claims, to fill these gaps and assess the representativeness of linked registry patients to the broader Medicare FFS population hospitalized with COVID-19.

METHODS AND RESULTS: We linked AHA Registry records of adults ≥65 years from March 2020 to September 2021 with Medicare FFS claims using a deterministic linkage algorithm and with the American Hospital Association Annual Survey, Rural Urban Commuting Area codes, and the Social Vulnerability Index using hospital and geographic identifiers. We compared linked individuals with unlinked FFS beneficiaries hospitalized with COVID-19 to assess the representativeness of the AHA Registry. A total of 10 010 (47.0%) records in the AHA Registry were successfully linked to FFS Medicare claims. Linked and unlinked FFS beneficiaries were similar with respect to mean age (78.1 versus 77.9, absolute standardized difference [ASD] 0.03); female sex (48.3% versus 50.2%, ASD 0.04); Black race (15.1% versus 12.0%, ASD 0.09); dual-eligibility status (26.1% versus 23.2%, ASD 0.07); and comorbidity burden. Linked patients were more likely to live in the northeastern United States (35.7% versus 18.2%, ASD 0.40) and urban/metropolitan areas (83.9% versus 76.8%, ASD 0.18). There were also differences in hospital-level characteristics between cohorts. However, in-hospital outcomes were similar (mortality, 23.3% versus 20.1%, ASD 0.08; home discharge, 45.5% versus 50.7%, ASD 0.10; skilled nursing facility discharge, 24.4% versus 22.2%, ASD 0.05).

CONCLUSIONS: Linkage of the AHA Registry with external data sources such as Medicare FFS claims creates a unique and generalizable resource to evaluate long-term health outcomes after COVID-19 hospitalization.

Key Words: cardiovascular diseases ■ COVID-19 ■ medicare ■ fee-for-service ■ mortality ■ readmissions ■ postdischarge outcomes

Patients hospitalized with COVID-19 are at increased risk of adverse cardiovascular events, including myocardial injury, arrhythmia, and thromboembolic phenomena.[1,2] Data on the longer-term cardiovascular sequelae of COVID-19 are limited,[3] and high rates of postacute COVID-19 syndromes (eg, "long COVID") present a potentially significant public health burden.[4] Additionally, throughout the course of the pandemic, existing health disparities associated with race/ethnicity and socioeconomic status have

Correspondence to: Dhruv S. Kazi, MD, MSc, MS and Robert W. Yeh, MD, MSc, Richard A. and Susan F. Smith Center for Outcomes Research in Cardiology, Beth Israel Deaconess Medical Center, 375 Longwood Ave, 4th Floor, Boston, MA 02215. Email: dkazi@bidmc.harvard.edu; rwyeh@bidmc.harvard.edu

*A. S. Oseran and T. Sun contributed equally.

†R. W. Yeh and D. S. Kazi contributed equally.

Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.122.027094

For Sources of Funding and Disclosures, see page 10.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- We successfully linked the American Heart Association COVID-19 Cardiovascular Disease Registry with Medicare fee-for-service claims, socioeconomic data, and hospital-level characteristics.
- We found that Medicare fee-for-service patients in the American Heart Association COVID-19 Cardiovascular Disease Registry are broadly representative of the general population of fee-for-service patients with a COVID-19 hospitalization with respect to sociodemographic characteristics and comorbidity burden.

What Are the Clinical Implications?

- This novel data source provides the first opportunity to comprehensively assess both in-hospital and long-term cardiovascular sequelae of COVID-19 among a representative cohort of Medicare fee-for-service patients, and to examine the socioeconomic and hospital factors influencing these outcomes.
- Future studies using data from this multidimensional linked registry have the potential to efficiently inform public health decision making, clinical care, and strategies to reduce inequities among individuals hospitalized with COVID-19.

Nonstandard Abbreviations and Acronyms

FFS	fee-for-service

been exacerbated, with vulnerable populations experiencing higher rates of infection and worse short-term outcomes, which may place them at increased risk of long-term complications.[5,6] Therefore, there is a critical need to better establish the association between COVID-19 and long-term cardiovascular outcomes in order to inform clinical care and public health decision making and to reduce systemic inequities.

The AHA Registry (American Heart Association COVID-19 Cardiovascular Disease Registry) was created with the goals of better understanding the effect of COVID-19 on in-hospital cardiovascular outcomes, accelerating the development of effective therapeutic strategies, facilitating quality improvement efforts, and informing preparations for future waves of the pandemic.[7] Since its creation, the registry has supported research on the short-term (inpatient) cardiovascular outcomes associated with COVID-19 hospitalizations.[8–10] The registry captures information such as prior cardiac history and major medical comorbidities, vital signs, laboratory values, medications, and treatments; however, it does not capture complete historical information or incorporate postdischarge follow-up, limiting its utility in examining the long-term impact of COVID-19. Prior studies have demonstrated that linking short-term registry or trial data with claims data can enable investigators to combine the granular phenotyping of a registry with the efficient long-term follow-up offered by claims.[11,12] Additional linkages with hospital-level data sets and community-level socioeconomic data can help answer questions about the role of structural and social determinants of health on long-term outcomes (Figure 1).[13–15]

Here, we describe the linkage of AHA Registry data with Medicare fee-for-service (FFS) claims, the American Hospital Association Annual Survey, the Social Vulnerability Index, and Rural–Urban Commuting Area codes to create a unique data resource that enriches the AHA Registry. We then use this new data resource to assess the representativeness of the AHA Registry for adults ≥65 years enrolled in Medicare FFS to the broader Medicare FFS population with a COVID-19 hospitalization.

METHODS

Data Management

This work is a National Heart, Lung, and Blood Institute–funded study (NIH/NHLBI R01HL157530) that will use linked data from the AHA Registry and external data sources, including Medicare claims, to better understand the impact of COVID-19 on cardiovascular disease. This represents a collaboration between the Richard A. and Susan F. Smith Center for Outcomes Research in Cardiology at Beth Israel Deaconess Medical Center and the American Heart Association. Data use agreements were established between the Smith Center and Centers for Medicare and Medicaid Services (CMS). All data management and analyses were performed through the CMS Virtual Research Data Center to allow timelier access to CMS data and secure handling of American Heart Association data.[16] Data linkage and all analyses were conducted using SAS version 9.4 (SAS Institute, Cary, NC). The data are not publicly available, but analytic methods and materials have been made available to other researchers within the article and its online supplementary files for the purposes of replicating the procedure.

Data Sources

Registry Data

The AHA Registry is a quality improvement program built on the Get With The Guidelines platform.[17] Site-level participation is voluntary, and any hospital or

Figure 1. American Hospital Association COVID-19 Registry linkage data sets.
The American Hospital Association (AHA) COVID-19 Registry was linked to: (1) a 100% sample of Medicare fee-for-service claims, which includes demographic information (Medicare Beneficiary Summary Files), hospitalization information (inpatient fee-for-service claims), and historical comorbidity burden (Chronic Conditions Segment); (2) the Social Vulnerability Index, which provides a county-level rank/score of a community's social vulnerability based on socioeconomic status, household composition and disability, minority status and language, housing type, and transportation; (3) Rural–Urban Commuting Area Codes, which categorize an individual's residence as urban (metropolitan and micropolitan) or rural based on ZIP code; and (4) American Hospital Association Survey, which provides hospital-level characteristics. FFS indicates fee-for-service.

health system in the United States treating patients with COVID-19 is eligible to enroll. Hospitalization details are abstracted from consecutive patients ≥18 years of age admitted with SARS CoV-2 infection. Individual patient informed consent is not required because data are abstracted retrospectively and anonymously.[7]

More than 200 data elements are abstracted from patient charts including demographic information, medical history, medications taken before hospitalization, admission vital signs, admission and serial laboratory values, COVID-19-directed medical therapy, cardiovascular medical therapy, inpatient interventions, and in-hospital outcomes. Additional information on the data elements, including the full case record form, is available at www.heart.org/COVIDRegistry. Data abstraction occurs onsite with active educational and technical support, quality control, and auditing from the American Heart Association.

Claims Data

Payer claims data for linkage were derived from a 100% sample of Medicare FFS beneficiary claims. Specifically, we used data from 3 sources: (1) Medicare Beneficiary Summary Files (demographic characteristics, monthly enrollment status, and mortality information); (2) inpatient FFS claims (admission and discharge dates, and *International Classification of Diseases, Tenth Revision* (*ICD-10*) diagnosis and procedure codes); and (3) the Chronic Conditions Warehouse, which identifies a history of 27 specific comorbidities for each beneficiary, including both cardiac (eg, heart failure, ischemic heart

disease, and atrial fibrillation) and noncardiac (eg, chronic obstructive pulmonary disease, asthma, and chronic kidney disease) conditions.[18–20] In addition to the supplemental covariate information provided by the Chronic Conditions Warehouse, long-term post-discharge outcomes of AHA Registry participants (eg, number of outpatient visits, readmission rates and primary readmission reason, and mortality) will be monitored.

Other Data Sets

Hospital characteristics were obtained from the 2020 American Hospital Association Annual Survey.[15] Community-level data were obtained from the 2018 Social Vulnerability Index, which utilizes US census data to create a county-level rank of a community's social vulnerability (eg, socioeconomic status, household composition and disability, minority status and language, housing type, and transportation).[14] Finally, 2010 Rural–Urban Commuting Area codes were used to identify the rural/urban status of beneficiaries.[13]

Linkage Method

For this analysis, we identified AHA registry hospitalizations admitted on or after March 1, 2020 and discharged on or before September 30, 2021. We restricted our sample to records of individuals ≥65 years and excluded records with missing information from any of the 7 variables (n=645) required for linkage and from Veterans Affairs hospitals (n=71), (Figure 2).

We then used a deterministic matching algorithm to link hospitalization records in the AHA Registry with Medicare FFS claims. Our approach was similar to those used in prior efforts to link registry data (including Get With The Guidelines) with CMS claims.[21–23] First, we identified inpatient claims from hospitalizations with a diagnosis of COVID-19 based on *ICD-10* codes (U07.1 [COVID-19], B97.29 [other coronavirus as the cause of disease classified elsewhere], J12.82 [pneumonia due to coronavirus disease 2019]).[24,25] We then selected 7 deterministic matching factors: date of birth, sex, race, 5-digit ZIP code, hospital ID, admission date, and discharge date. AHA registry hospitalizations were linked to CMS claims in an iterative fashion (Table 1). As a first step, we required a perfect match across all 7 factors. Next, for unmatched hospitalizations, we gradually relaxed our matching criteria by allowing a 6-factor match (excluding either race or ZIP code), a window around admission and discharge dates (±1, 2, or 7 days), and finally only requiring 2 of 3 date of birth elements (ie, day, month, year). AHA Registry Hospital ID numbers were cross walked to Medicare Provider ID numbers using American Hospital Association ID numbers.

We excluded beneficiaries enrolled in Medicare FFS for <12 months (n=502) to allow accurate assessment of baseline comorbidity burden using Chronic Conditions Segment files. To ensure a cohort of unique beneficiaries, we excluded records where multiple hospitalizations in the AHA Registry matched to a single Medicare beneficiary (n=376) or where one hospitalization was matched to multiple Medicare beneficiaries (n=2). Inpatient claims from different hospitals that were separated by 1 day or less were assumed to represent a hospital-to-hospital transfer and treated as a single episode of care.

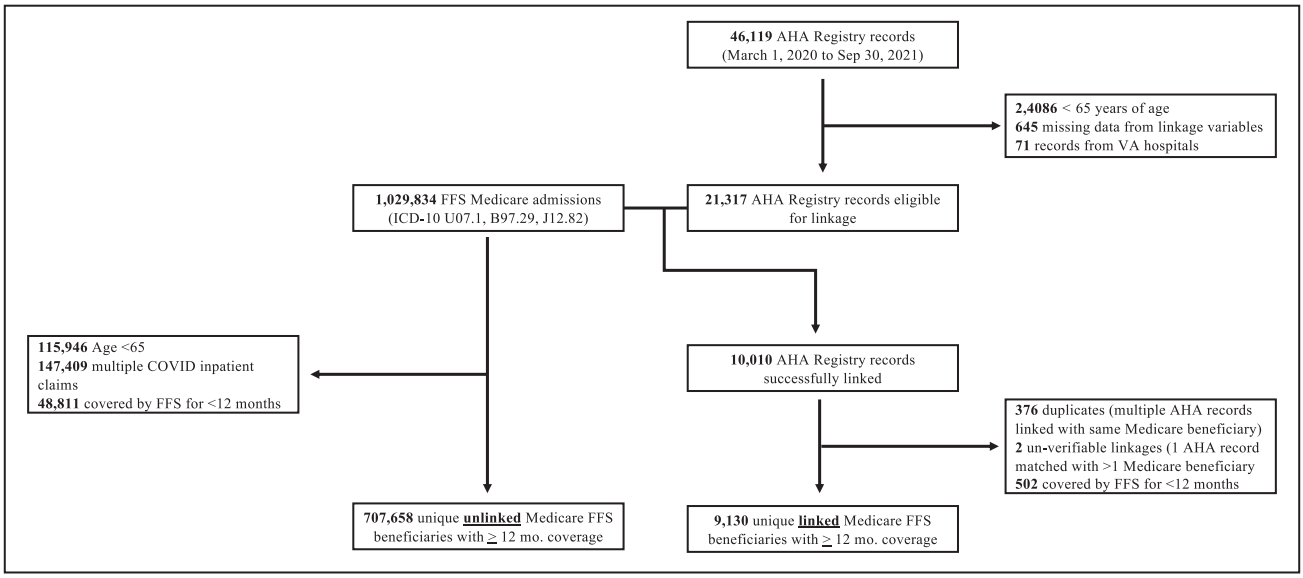

Figure 2. Flow chart: linkage of American Hospital Association COVID-19 Registry and fee-for-service Medicare COVID-19 admissions.
AHA indicates American Heart Association; FFS, fee-for-service; *ICD-10*, *International Classification of Diseases, Tenth Revision*; and VA, Veterans Administration.

Table 1. Stepwise Matching Algorithm Used to Link American Heart Association COVID-19 Registry with Medicare Fee-For-Service Claims*,†

	Date of birth	Sex	Hospital identification	ZIP code	Race	Admission date	Discharge date	Number linked
Step 1	✓	✓	✓	✓	✓	Exact	Exact	5766
Step 2	✓	✓	✓	✓		Exact	Exact	639
Step 3	✓	✓	✓		✓	Exact	Exact	1453
Step 4	✓	✓	✓	✓	✓	±1 d	±1 d	839
Step 5	✓	✓	✓	✓	✓	±2 d	±2 d	213
Step 6	✓	✓	✓	✓	✓	±7 d	±7 d	210
Step 7	✓	✓	✓		✓	±1 d	±1 d	248
Step 8	✓	✓	✓	✓		±1 d	±1 d	123
Step 9	✓	✓	✓			±1 d	±1 d	200
Step 10	✓	✓	✓			Exact		42
Step 11	✓	✓	✓				Exact	140
Step 12	2 of 3‡	✓	✓			Exact	Exact	137
Step 13	2 of 3‡	✓	✓			±1 d	±1 d	0
Total								10 010

*Each subsequent step is incremental to the prior (ie, identifies new matches that were not included in the prior step[s]).

†Steps 1–9 required a COVID-19 *International Classification of Diseases, Tenth Revision* code (U07.1, B97.29, J12.82) to be present in the inpatient fee-for-service claim; these steps were subsequently repeated relaxing this requirement (ie, allowing linkage with any fee-for-service inpatient claim) to match additional records.

‡Step 12, 13 required 2 of 3 elements (ie, day, month, year) of birthday for the matching.

American Hospital Association Annual Survey data were linked to the AHA Registry using American Hospital Association ID and to FFS claims using the Medicare Provider Number. Rural–Urban Commuting Area codes were linked using beneficiary 5-digit ZIP code and Social Vulnerability Index data were linked using 5-digit county-level Federal Information Processing System code.

Example SAS code implementing this linkage algorithm, along with details for data set preparation, are provided in Figure S1. In the future, linkage of these data sets will be repeated at 6-month intervals as data from both the AHA Registry and Medicare claims data continue to accrue.

Statistical Analysis

To assess whether linkable registry patients were representative of hospitalized patients with COVID-19, we compared unique AHA Registry patients who could be linked to Medicare claims ('linked') with those of Medicare FFS beneficiaries hospitalized with COVID-19 identified only in claims ('unlinked'). We first compared patient-level sociodemographic characteristics (age, sex, race, geographic region, urban/rural status, dual-eligible status, and Social Vulnerability Index quartile) and clinical comorbidities between linked and unlinked patients. Next, to understand differences in COVID-19 disease severity and care quality, we compared in-hospital outcomes (mortality, discharge status, and length of stay) between groups. Finally, we compared the hospital characteristics where each group received care (bed size [small, medium, and large], teaching status [major, minor, and none], ownership type, and presence of specific resources [hemodialysis, intensive care unit, cardiac catheterization laboratory, and cardiac surgery]). We used absolute standardized differences to compare groups, as has been done in other similar analyses (Data S1; Supplemental Methods).[26,27] Multivariate Mahalanobis distance was used for multinomial variables.[28] An absolute standard difference ≤0.1 is suggestive of balance between groups.[27,29] For this analysis, we relied on data elements available in CMS files (rather than those elements only available in the AHA Registry) to facilitate a comparison between the linked and unlinked cohorts. Hospital characteristics were based on the discharging hospital.

As a supplementary analysis, we also compared linked and unlinked patients from the AHA Registry to assess whether any patient characteristics were associated with linkage using data elements from the AHA Registry.

RESULTS

There were 46 119 COVID-19 hospitalizations in the AHA Registry during our study period. Of those, 21 317 records were eligible for linkage (eg, ≥65 years old, complete information across all linkage variables, admitted at non-VA hospitals). In the inpatient Medicare claims files during this same period, there

were 1 029 834 inpatient claims with a COVID-19 diagnosis. Using the described linkage algorithm, 10 010 (47.0%) hospitalizations from the AHA registry were successfully linked with Medicare claims. After excluding beneficiaries enrolled in FFS for <12 months and those with duplicate records, there were 9130 unique Medicare FFS beneficiaries in the linked cohort and 707 658 unique Medicare FFS beneficiaries in the unlinked cohort.

Linked and unlinked beneficiaries were similar with respect to mean age (78.1 years in the linked cohort versus 77.9 years in the unlinked cohort, absolute standardized difference [ASD] 0.03), sex (48.3% female versus 50.2% female, ASD 0.04), race (75.3% versus 78.6% White, ASD 0.08; 15.1% versus 12.0% Black, ASD 0.09), and proportion of beneficiaries with dual-eligible status (26.1% versus 23.2%, ASD 0.07). The linked and unlinked cohorts also had a similar burden of all chronic comorbidities evaluated, including similar prevalence of hypertension (89.2% versus 87.6%, ASD 0.05), diabetes (55.6% versus 54.1%, ASD 0.03), ischemic heart disease (64.3% versus 61.9%, ASD 0.05), heart failure (47.9% versus 44.2%, ASD 0.08), and stroke (26.3% versus 24.2%, ASD 0.05) (Table 2, Figure 3).

Linked patients were more likely to live in the northeastern United States (35.7% versus 18.2%, ASD 0.40) and in urban/metropolitan areas (83.9% versus 76.8%, ASD 0.18). Additionally, linked patients were more likely to be treated at large (56.7% versus 32.9%, ASD 0.49), major teaching (49.5% versus 14.8%, ASD 0.80), and publicly owned (21.1% versus 11.5%, ASD 0.26) hospitals. These hospitals also had more resources including a higher proportion with an intensive care unit (98.0% versus 93.6%, ASD 0.22), cardiac catheterization laboratory (92.6% versus 79.4%, ASD 0.39), and cardiac surgery capability (79.4% versus 61.7%, ASD 0.40).

With respect to in-hospital outcomes, linked and unlinked patient had similar rates of in-hospital mortality (23.3% versus 20.1%, ASD 0.08), mean length of stay (11.1 versus 9.9 days, ASD 0.09), as well as home (45.5% versus 50.7%, ASD 0.10) and skilled nursing facility (24.4% versus 22.2%, ASD 0.05) discharges, Table 3.

Linked and unlinked patients from the AHA Registry had slightly different mean age (78.0 versus 76.5 years, ASD 0.17) and race distribution (75.4% versus 64.0% White, ASD 0.25; 14.3% versus 21.2% Black, ASD 0.18) but similar comorbidity burdens, Table S1.

DISCUSSION

In this study, we describe the creation of a unique, longitudinal, and representative data resource that includes the AHA Registry linked with external data, including Medicare FFS claims. This linked registry will be used to characterize long-term health outcomes and resource utilization among older adults following acute COVID-19 hospitalization and to identify clinical, biological, health system, and social risk factors associated with their occurrence.

In our initial effort, we successfully linked hospitalizations from the AHA Registry with a 100% sample of Medicare FFS claims using a deterministic matching algorithm. Among individuals ≥65 years of age with an acute COVID-19 hospitalization in the AHA registry, 47.0% matched with Medicare FFS claims. Our observed matching rate is slightly lower than those seen in other studies linking registry data to CMS claims, demonstrating limitations in the matching process (ie, the use of indirect identifiers) and also likely because of the significant and growing proportion of elderly individuals covered by Medicare Advantage plans, and therefore potentially without inpatient FFS claims.[22,23,30] We were then able to leverage this linkage strategy to show that Medicare FFS patients in the AHA Registry are broadly representative of the general population of FFS patients with a COVID-19 hospitalization with respect to sociodemographic characteristics and comorbidity burden. The case for representativeness is also supported by similar in-hospital outcomes. On the other hand, linkable registry patients tended to be from larger, major teaching hospitals with more resources, suggesting differences in the types of hospitals that opted to participate in the registry.

While adverse cardiac events are common during acute COVID-19 infection and are associated with worse short-term outcomes, how they relate to rates of long-term major adverse cardiac events and other postacute sequelae of COVID-19 remain incompletely understood. With ongoing waves of COVID-19 cases and hospitalizations, and with some studies estimating that over half of COVID-19 survivors experience long-term sequelae of infection,[31] the ability to quickly leverage "real world" data to clarify the complex nature of cardiovascular disease in COVID-19 survivors is critical. The demonstration of the representativeness of the AHA Registry is a basic step in that process and, overall, our findings support the use of the linked AHA Registry-Medicare data set to examine long-term outcomes after COVID-19 hospitalization in Medicare FFS beneficiaries. Future work using this linked data set can harness both the granularity of registry data and the efficiency provided by claims in collecting long-term outcomes, with the ultimate goal of providing insights that may inform public health interventions, clinical management, and system resource allocation.

The COVID-19 pandemic has exposed unacceptable inequities in our health system. Ample data suggest that low-income individuals and those from racial/ethnic minority groups are at substantially increased risk of both direct and indirect effects of the pandemic.[32,33] Many of these disparities are upstream

Table 2. Baseline Characteristics of Linked American Heart Association Registry-Medicare Fee-For-Service Patients and Unlinked Medicare Fee-For-Service Patients With COVID-19 Hospitalization

	Linked patients n=9130	Unlinked patients n=707 658	Absolute Standardized Difference*
Patient-level sociodemographic characteristics			
Age (y), mean (SD)	78.1 (8.5)	77.9 (8.4)	0.03
Female, N (%)	4411 (48.3)	354 914 (50.2)	0.04
Race, N (%)†			0.10
White	6879 (75.3)	555 909 (78.6)	0.08
Black	1379 (15.1)	85 071 (12.0)	0.09
Other‡	872 (9.6)	66 678 (9.4)	<0.01
Region, N (%)†,§			0.41
Midwest	1674 (18.3)	159 026 (22.5)	0.10
Northeast	3262 (35.7)	128 418 (18.2)	0.40
South	3200 (35.1)	307 852 (43.6)	0.17
West	990 (10.8)	111 430 (15.8)	0.15
Medicaid-Medicare dual enrollment, N (%)	2379 (26.1)	163 840 (23.2)	0.07
Chronic comorbidities, N (%)			
Cardiovascular comorbidities			
Atrial fibrillation	2333 (25.6)	168 287 (23.8)	0.04
Diabetes	5079 (55.6)	382 768 (54.1)	0.03
Heart failure	4377 (47.9)	312 807 (44.2)	0.08
Hyperlipidemia	7765 (85.0)	589 510 (83.3)	0.05
Ischemic heart disease	5868 (64.3)	438 288 (61.9)	0.05
Hypertension	8141 (89.2)	619 773 (87.6)	0.05
Obesity	3835 (42.0)	300 700 (42.5)	0.01
Peripheral vascular disease	3923 (43.0)	286 486 (40.5)	0.05
Stroke or transient ischemic attack	2404 (26.3)	171 114 (24.2)	0.05
Tobacco use disorders	1508 (16.5)	118 012 (16.7)	<0.01
Oncologic comorbidities			
Breast cancer	525 (5.8)	39 026 (5.5)	0.01
Colorectal cancer	366 (4.0)	25 757 (3.6)	0.02
Endometrial cancer	115 (1.3)	8337 (1.2)	0.01
Lung cancer	234 (2.6)	17 304 (2.4)	0.01
Prostate cancer	723 (7.9)	50 735 (7.2)	0.03
Other comorbidities			
Alzheimer disease and related disorders	3519 (38.5)	228 090 (32.2)	0.13
Anemia	6599 (72.3)	478 192 (67.6)	0.10
Asthma	1756 (19.2)	137 906 (19.5)	0.01
Benign prostatic hyperplasia	2581 (28.3)	191 421 (27.0)	0.03
Cataract	6302 (69.0)	478 270 (67.6)	0.03
Chronic kidney disease	5648 (61.9)	410 766 (58.0)	0.08
Chronic obstructive pulmonary disease	3570 (39.1)	278 992 (39.4)	0.01
Depression	4615 (50.5)	327 464 (46.3)	0.09
Glaucoma	2473 (27.1)	176 153 (24.9)	0.05
Hip/pelvic fracture	755 (8.3)	48 551 (6.9)	0.05
Hypothyroidism	3061 (33.5)	236 580 (33.4)	<0.01
Mobility impairments	1158 (12.7)	77 844 (11.0)	0.05
Osteoporosis	2259 (24.7)	165 829 (23.4)	0.03
Rheumatoid arthritis	6398 (70.1)	481 604 (68.1)	0.04

(*Continued*)

Table 2. Continued

	Linked patients n=9130	Unlinked patients n=707 658	Absolute Standardized Difference*
Community-level sociodemographic characteristics‖,¶			
RUCA, N (%)†			0.20
Metro	7655 (83.9)	543 137 (76.8)	0.18
Micro	662 (7.3)	84 552 (12.0)	0.16
Small-town/rural	810 (8.9)	79 499 (11.2)	0.08
SVI, N (%)†			0.03
Least vulnerable quantile	1729 (19.0)	123 937 (17.6)	0.04
Third vulnerable quantile	2138 (23.5)	163 730 (23.3)	<0.01
Second vulnerable quantile	2966 (32.6)	233 936 (33.3)	0.014
Most vulnerable quantile	2260 (24.9)	181 797 (25.8)	0.023
Hospital-level characteristics¶			
Hospital bed size, N (%)†			0.54
Small (<100)	466 (5.2)	107 804 (15.5)	0.34
Medium (100–399)	3441 (38.2)	359 805 (51.6)	0.27
Large (≥400)	5108 (56.7)	229 075 (32.9)	0.49
Hospital teaching status, N (%)†			0.85
Major	4466 (49.5)	102 983 (14.8)	0.80
Minor	3686 (40.9)	407 347 (58.5)	0.36
None	863 (9.6)	186 354 (26.7)	0.46
Hospital ownership, N (%)†			0.45
Federal/military	0 (0.0)	1263 (0.2)	0.06
Private, for-profit	365 (4.0)	113 505 (16.3)	0.41
Private, not-for-profit	6744 (74.8)	501 803 (72.0)	0.063
Public/municipal	1906 (21.1)	80 113 (11.5)	0.26
Hemodialysis capability, N (%)	5517 (65.9)	353 481 (61.6)	0.09
ICU presence, N (%)	8197 (98.0)	537 262 (93.6)	0.22
Cardiac catheterization laboratory presence, N (%)	7745 (92.6)	455 815 (79.4)	0.39
Cardiac surgery presence, N (%)	6646 (79.4)	353 938 (61.7)	0.40
Number of intensivists, mean (SD)	43.7 (85.3)	13.2 (33.2)	0.47

ICU indicates intensive care unit; RUCA, Rural–Urban Commuting Area codes; and SVI, social vulnerability index.

*Absolute standardized difference (ASD) is calculated by taking the difference in means of a covariate across treatment groups, divided by the combined SD of both groups. Multivariate Mahalanobis distance was used for multinomial variables.

†Categorical variables may not sum to 1 because of rounding.

‡In the AHA COVID-19 registry, race is categorized as American Indian/Alaska Native, Asian (with subcategories for Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, Other Asian), Black or African American, Native Hawaiian or Pacific Islander, White, or Unable to Determine. Hispanic ethnicity is recorded as a separate variable, and includes subcategories for Mexican/Mexican American/Chicano(a), Puerto Rican, Cuban, and Other Hispanic, Latino, or Spanish Origin.

§Puerto Rico and Guam (0.13%) were excluded for the region variable.

‖Based on beneficiary ZIP/FIPS code (not hospital location).

¶Missing values were excluded from analysis. For hospital bed size, hospital teaching status, and hospital ownership the missing rate was 1.55%. For hemodialysis capability, ICU presence, cardiac catheterization laboratory presence, and cardiac surgery presence the missing rate was 18.79%.

of the health system and related to community-level factors that increase exposure and susceptibility, or limit access to health care.[34–36] Therefore, the ability to examine the association between community-level factors and COVID-19 outcomes may help develop targeted interventions to reduce some of these stark disparities. However, because social determinants of health are not captured well in registries and health care claims, linking these data with established data sets of social vulnerability, as in this case, creates a powerful resource to examine the association of community characteristics on long-term health outcomes.

Insights derived from the AHA COVID-19 linked registry will need to be interpreted with the following considerations. Findings are limited to individuals enrolled in Medicare FFS and may not generalize to hospitalized COVID-19 patients younger than 65 years of age or those covered by Medicare Advantage plans. However, older adults are disproportionately likely to be hospitalized for COVID-19 and face in-hospital

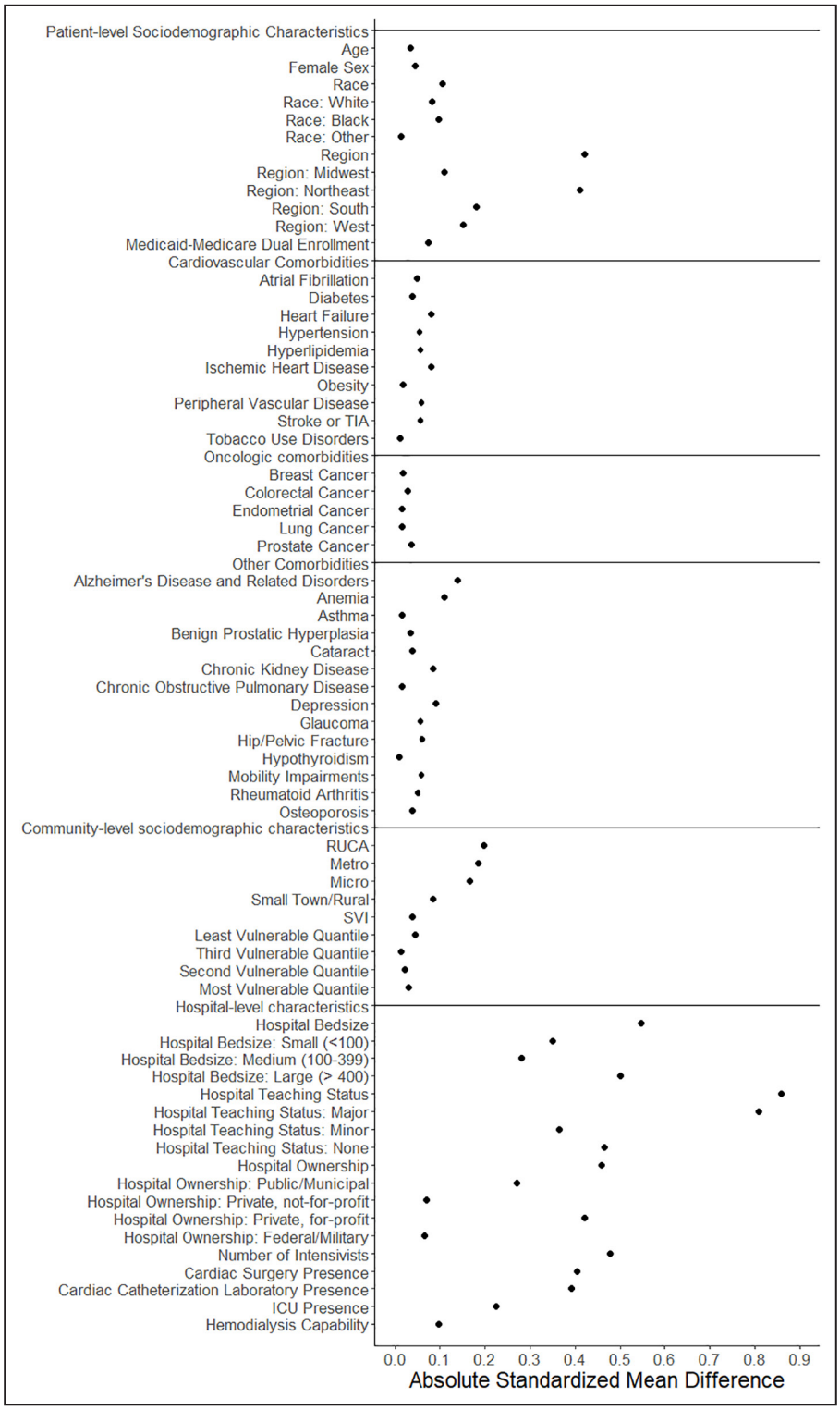

Figure 3. Balance of patient, community, and hospital characteristics and in-hospital outcomes among linked American Hospital Association Registry-Medicare fee-for-service patients and unlinked Medicare fee-for-service patients with COVID-19 hospitalization.

Absolute standardized difference is calculated by taking the difference in means of a covariate across treatment groups, divided by the combined SD of both groups. Multivariate Mahalanobis distance was used for multinomial variables. ICU indicates intensive care unit; RUCA, Rural–Urban Commuting Area codes; SVI, Social Vulnerability Index; and TIA, transient ischemic attack.

Table 3. In-Hospital Outcomes Among Linked American Hospital Association Registry-Medicare Fee-For-Service Patients and Unlinked Medicare Fee-For-Service Patients With COVID-19 Hospitalization

	Linked patients n=9130	Unlinked patients n=707 658	Absolute standardized difference*
Discharge status, N (%)†			0.10
Died	2131 (23.3)	142 187 (20.1)	0.08
Home	4155 (45.5)	358 710 (50.7)	0.10
Hospice	467 (5.1)	33 928 (4.8)	0.01
Skilled nursing facility or rehabilitation center	2230 (24.4)	156 770 (22.2)	0.05
Others	147 (1.6)	16 063 (2.3)	0.05
Length of stay (d), mean (SD)	11.1 (13.3)	9.9 (12.0)	0.09

*Absolute standardized difference (ASD) is calculated by taking the difference in means of a covariate across treatment groups, divided by the combined SD of both groups. Multivariate Mahalanobis distance was used for multinomial variables.
†Sum of categorical variable percent may not equal 1 because of rounding.

complications, so the study focuses on a particularly vulnerable population. Second, we were unable to directly compare COVID-19 disease severity between linked and unlinked patients because of the absence of granular clinical information in health care claims. However, because the linked and unlinked cohorts had similar short-term outcomes, it is likely that disease severity was similar between groups. Third, this initial description only examines data before September 2021. Therefore, the assessment of representativeness may need to be updated to evaluate comparability in subsequent waves of the pandemic.

The linkage between the AHA registry, Medicare claims, socioeconomic data, and hospital-level characteristics provides the first opportunity to comprehensively assess both in-hospital and long-term cardiovascular sequelae of COVID-19 among a representative cohort of Medicare FFS patients, and to examine the socioeconomic and hospital factors influencing these outcomes. This multidimensional linked registry will provide "real world" data to efficiently inform future public health decision making, clinical care, and strategies to reduce inequities.

ARTICLE INFORMATION

Received June 21, 2022; accepted July 12, 2022.

Affiliations

Richard A. and Susan F. Smith Center for Outcomes Research, Beth Israel Deaconess Medical Center, Boston, MA (A.S.O., T.S., R.K.W., I.J.D., R.W.Y., D.S.K.); Division of Cardiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA (A.S.O.); Division of Cardiology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA (R.K.W., A.H.A., R.W.Y., D.S.K.); CAUSALab (I.J.D.); and Departments of Epidemiology and Biostatistics (I.J.D.), Harvard T.H. Chan School of Public Health, Boston, MA; Division of Cardiology, University of Texas Southwestern Medical Center, Dallas, TX (J.A.d.L., S.R.D.); and Quality, Outcomes Research and Analytics, American Heart Association, Dallas, TX (C.R.).

Acknowledgments

We acknowledge Joanne Laffan, MBS, the Senior Administrative Manager at the Richard A. and Susan F. Smith Center for Outcomes Research in Cardiology. The content of this work is solely the responsibility of the authors and does not necessarily represent the official views of PCORI, PCORI's Board of Governors, or the PCORI Methodology Committee.

Sources of Funding

This work is supported in part by grants from the National Heart, Lung, and Blood Institute (R01HL157530) and Patient-Centered Outcomes Research Institute (PCORI) (ME-1502-27794).

Disclosures

None.

Supplemental Material

Data S1
Table S1
Figure S1

REFERENCES

1. Kazi DS, Martin LM, Litmanovich D, Pinto DS, Clerkin KJ, Zimetbaum PJ, Dudzinski DM. Case 18-2020: a 73-year-old man with hypoxemic respiratory failure and cardiac dysfunction. *N Engl J Med*. 2020;382(24):2354–2364. doi: 10.1056/NEJMcpc2002417
2. Giustino G, Pinney SP, Lala A, Reddy VY, Johnston-Cox HA, Mechanick JI, Halperin JL, Fuster V. Coronavirus and cardiovascular disease, myocardial injury, and arrhythmia. *J Am Coll Cardiol*. 2020;76:2011–2023. doi: 10.1016/j.jacc.2020.08.059
3. Xie Y, Xu E, Bowe B, Al-Aly Z. Long-term cardiovascular outcomes of COVID-19. *Nat Med*. 2022;28:583–590. doi: 10.1038/s41591-022-01689-3
4. Satterfield BA, Bhatt DL, Gersh BJ. Cardiac involvement in the long-term implications of COVID-19. *Nat Rev Cardiol*. 2021;19:332–341. doi: 10.1038/s41569-021-00631-3
5. Lopez L III, Hart LH III, Katz MH. Racial and ethnic health disparities related to COVID-19. *JAMA*. 2021;325:719–720. doi: 10.1001/jama.2020.26443
6. Garcia S, Dehghani P, Grines C, Davidson L, Nayak KR, Saw J, Waksman R, Blair J, Akshay B, Garberich R, et al. Initial findings from the north American COVID-19 myocardial infarction registry. *J Am Coll Cardiol*. 2021;77:1994–2003. doi: 10.1016/j.jacc.2021.02.055
7. Alger HM, Rutan C, Williams JHt, Walchok JG, Bolles M, Hall JL, Bradley SM, Elkind MSV, Rodriguez F, Wang TY, et al. American Heart Association COVID-19 CVD registry powered by get with the guidelines. *Circ Cardiovasc Qual Outcomes* 2020;13:e006967. doi: 10.1161/circoutcomes.120.006967
8. Bradley SM, Emmons-Bell S, Mutharasan RK, Rodriguez F, Gupta D, Roth G, Gluckman TJ, Shah RU, Wang TY, Khera R, et al. Repeated cross-sectional analysis of hydroxychloroquine deimplementation in the AHA COVID-19 CVD registry. *Sci Rep*. 2021;11:15097. doi: 10.1038/s41598-021-94203-7
9. Roth GA, Emmons-Bell S, Alger HM, Bradley SM, Das SR, de Lemos JA, Gakidou E, Elkind MSV, Hay S, Hall JL, et al. Trends in patient

characteristics and COVID-19 in-hospital mortality in the United States during the COVID-19 pandemic. *JAMA Netw Open*. 2021;4:e218828. doi: 10.1001/jamanetworkopen.2021.8828
10. Acharya P, Ranka S, Sethi P, Bharati R, Hu J, Noheria A, Nallamothu BK, Hayek SS, Gupta K. Incidence, predictors, and outcomes of in-hospital cardiac arrest in COVID-19 patients admitted to intensive and non-intensive care units: insights from the AHA COVID-19 CVD registry. *J Am Heart Assoc*. 2021;10:e021204. doi: 10.1161/jaha.120.021204
11. Butala NM, Faridi KF, Tamez H, Strom JB, Song Y, Shen C, Secemsky EA, Mauri L, Kereiakes DJ, Curtis JP, et al. Estimation of DAPT study treatment effects in contemporary clinical practice: findings from the EXTEND-DAPT study. *Circulation*. 2022;145:97–106. doi: 10.1161/CIRCULATIONAHA.121.056878
12. Butala NM, Secemsky E, Kazi DS, Song Y, Strom JB, Faridi KF, Brennan JM, Elmariah S, Shen C, Yeh RW. Applicability of transcatheter aortic valve replacement trials to real-world clinical practice. *J Am Coll Cardiol Interv*. 2021;14:2112–2123. doi: 10.1016/j.jcin.2021.08.006
13. Department of Agriculture, Economic Research Service. Rural-urban commuting area codes. Available at: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx. Accessed January 15, 2022.
14. CDC/ATSDR Social Vulnerability Index. Agency for Toxic Substances and Disease Registry. Available at: https://www.atsdr.cdc.gov/placeandhealth/svi/index.html. Accessed January 15, 2022.
15. AHA Annual Survey Database. American Hospital Association. Available at: https://www.ahadata.com/aha-annual-survey-database. Accessed January 15, 2022.
16. CMS Virtual Research Data Center (VRDC). Research Data Assistance Center. Available at: https://resdac.org/cms-virtual-research-data-center-vrdc. Accessed January 15, 2022.
17. Hong Y, LaBresh KA. Overview of the American Heart Association "get with the guidelines" programs: coronary heart disease, stroke, and heart failure. *Crit Pathw Cardiol*. 2006;5:179–186. doi: 10.1097/01.hpc.0000243588.00012.79
18. Inpatient (Fee-for-Service). Research Data Asssitance Center. Available at: https://resdac.org/cms-data/files/ip-ffs. Accessed January 15, 2022.
19. Master Beneficiary Summary File (MBSF) Base. Research Data Assistance Center. Available at: https://resdac.org/cms-data/files/mbsf-base. Accessed January 15, 2022.
20. Master Beneficiary Summary File (MBSF): Chronic Conditions Segment. Research Data Assistance Center. https://resdac.org/cms-data/files/mbsf-cc. Accessed January 15, 2022.
21. Strom JB, Tamez H, Zhao Y, Valsdottir LR, Curtis J, Brennan JM, Shen C, Popma JJ, Mauri L, Yeh RW. Validating the use of registries and claims data to support randomized trials: rationale and design of the extending trial-based evaluations of medical therapies using novel sources of data (EXTEND) study. *Am Heart J*. 2019;212:64–71. doi: 10.1016/j.ahj.2019.02.007
22. Hammill BG, Hernandez AF, Peterson ED, Fonarow GC, Schulman KA, Curtis LH. Linking inpatient clinical registry data to Medicare claims data using indirect identifiers. *Am Heart J*. 2009;157:995–1000. doi: 10.1016/j.ahj.2009.04.002
23. Reeves MJ, Fonarow GC, Smith EE, Pan W, Olson D, Hernandez AF, Peterson ED, Schwamm LH. Representativeness of the Get With the Guidelines–Stroke Registry: comparison of patient and hospital characteristics among Medicare beneficiaries hospitalized with ischemic stroke. *Stroke*. 2012;43:44–49. doi: 10.1161/STROKEAHA.111.626978
24. Kadri SS, Gundrum J, Warner S, Cao Z, Babiker A, Klompas M, Rosenthal N. Uptake and accuracy of the diagnosis code for COVID-19 among US hospitalizations. *JAMA*. 2020;324:2553–2554. doi: 10.1001/jama.2020.20323
25. Santostefano CM, White EM, Feifer RA, Mor V. Accuracy of ICD-10 codes for identifying skilled nursing facility residents with lab-confirmed COVID-19. *J Am Geriatr Soc*. 2021;69:3397–3399. doi: 10.1111/jgs.17412
26. Gehi AK, Doros G, Glorioso TJ, Grunwald GK, Hsu J, Song Y, Turakhia MP, Turchin A, Virani SS, Maddox TM. Factors associated with rhythm control treatment decisions in patients with atrial fibrillation-insights from the NCDR PINNACLE registry. *Am Heart J*. 2017;187:88–97. doi: 10.1016/j.ahj.2017.02.006
27. Sullivan GM, Feinn R. Using effect size-or why the P value is not enough. *J Grad Med Educ*. 2012;4:279–282. doi: 10.4300/jgme-d-12-00156.1
28. Yang D, Dalton JE. A unified approach to measuring the effect size between two groups using SAS®. Paper/Poster presented at: SAS global forum. 2012.
29. Zhang XD. Strictly standardized mean difference, standardized mean difference and classical t-test for the comparison of two groups. *Stat Biopharm Res*. 2010;2:292–299.
30. Centers for Medicare and Medicaid Services. Report to the Congress: Medicare Payment Policy. Medicare Payment Advisory Commission. Available at: https://www.medpac.gov/wp-content/uploads/import_data/scrape_files/docs/default-source/reports/jun21_medpac_report_to_congress_sec.pdf. 2021. Accessed September 30, 2021.
31. Groff D, Sun A, Ssentongo AE, Ba DM, Parsons N, Poudel GR, Lekoubou A, Oh JS, Ericson JE, Ssentongo P, et al. Short-term and long-term rates of Postacute sequelae of SARS-CoV-2 infection: a systematic review. *JAMA Netw Open*. 2021;4:e2128568. doi: 10.1001/jamanetworkopen.2021.28568
32. Wadhera RK, Figueroa JF, Rodriguez F, Liu M, Tian W, Kazi DS, Song Y, Yeh RW, Joynt Maddox KE. Racial and ethnic disparities in heart and cerebrovascular disease deaths during the COVID-19 pandemic in the United States. *Circulation*. 2021;143:2346–2354. doi: 10.1161/circulationaha.121.054378
33. Wadhera RK, Wadhera P, Gaba P, Figueroa JF, Joynt Maddox KE, Yeh RW, Shen C. Variation in COVID-19 hospitalizations and deaths across New York City boroughs. *JAMA*. 2020;323:2192–2195. doi: 10.1001/jama.2020.7197
34. Figueroa JF, Wadhera RK, Lee D, Yeh RW, Sommers BD. Community-level factors associated with racial and ethnic disparities in COVID-19 rates in Massachusetts. *Health Aff (Millwood)*. 2020;39:1984–1992. doi: 10.1377/hlthaff.2020.01040
35. Figueroa JF, Wadhera RK, Mehtsun WT, Riley K, Phelan J, Jha AK. Association of race, ethnicity, and community-level factors with COVID-19 cases and deaths across U.S. counties. *Healthcare*. 2021;9:100495. doi: 10.1016/j.hjdsi.2020.100495
36. Rodriguez F, Solomon N, Lemos JA, Das SR, Morrow DA, Bradley SM, Elkind MSV, Williams JH, Holmes D, Matsouaka RA, et al. Racial and ethnic differences in presentation and outcomes for patients hospitalized with COVID-19: Findings from the American Heart Association's COVID-19 Cardiovascular Disease Registry. *Circulation*. 2021;143:2332–2342. doi: 10.1161/CIRCULATIONAHA.120.052278

SUPPLEMENTAL MATERIAL

Data S1. Supplemental Methods

Calculating and Interpreting Absolute Standardized Differences

The absolute standardized difference (ASD) is calculated by taking the difference in means of a covariate across treatment groups and then dividing that difference by the combined standard deviation of both groups ([mean group 1] – [mean group 2] / standard deviation). In general, absolute standardized differences of ≤ 0.1 are considered to represent differences of small magnitude.

Table S1. Baseline characteristics of linked versus unlinked American Heart Association Registry participants

	Linked patients n = 10,009	Unlinked patients n = 11,308	Absolute Standardized Difference*
Patient-level sociodemographic characteristics			
Age (year), Mean (SD)	78.0 (8.5)	76.5 (8.1)	0.17
Female, N (%)	4,837 (48.3)	5450 (48.2)	<0.01
Race, N (%)†			0.26
White	7,546 (75.4)	7,242 (64.0)	0.25
Black	1,436 (14.3)	2,400 (21.2)	0.18
Other	1,027 (10.3)	1,666 (14.7)	0.14
Region, N (%)†,‡			0.14
Midwest	1,822 (18.2)	1,752 (15.5)	0.0.7
Northeast	3,591 (35.9)	4,830 (42.7)	0.14
South	3,492 (34.9)	3,614 (32.0)	0.06
West	1103 (11.0)	1,111 (9.8)	0.04
Chronic comorbidities, N (%)			
Hypertension	7,725 (77.2)	8,791 (77.7)	0.01
Diabetes	3,925 (3.2)	4,933 (43.6)	0.01
Heart failure	2,079 (20.8)	2,118 (18.7)	0.05
Myocardial infarction	1,028 (10.3)	1,072 (9.5)	0.03
Peripheral arterial disease	490 (4.9)	529 (4.7)	0.01
Stroke/TIA	1,668 (16.7)	1,610 (14.2)	0.07
Smoking	614 (6.1)	760 (6.7)	0.02
Chronic kidney disease	2,071 (20.7)	2,172 (19.2)	0.04
COPD	1,689 (16.9)	1,858 (16.4)	0.01
Asthma	761 (7.6)	901 (8.0)	0.01
Community-level sociodemographic characteristics§			
RUCA, N (%)†			0.14
Metro	8,442 (84.5)	10,018 (88.8)	0.13
Micro	683 (6.8)	560 (5.0)	0.08
Small-town/rural	868 (8.7)	699 (6.2)	0.09

*Absolute standardized difference (ASD) is calculated by taking the difference in means of a covariate across treatment groups, divided by the combined standard deviation of both groups. Multivariate Mahalanobis distance was used for multinomial variables.

† Categorical variables may not sum to 1 due to rounding

‡ Puerto Rico and Guam (0.13%) were excluded for the region variable

§ Based on beneficiary ZIP code

Figure S1. Sample SAS Code for Linkage of AHA COVID-19 Cardiovascular Disease Registry with External Data

```
/*******************************************************/
/** Link dataset X and Y**/
/** Dataset X is the smaller one **/
/** *_ID: the ID variable in data X/Y**/
/** adate_*: the admission date in X/Y **/
/** ddate_*: the discharge date in X/Y **/
/** bdate_*: the birth date in X/Y **/
/** bday_*,bmonth_*,byear_*: 3 elements (day, month, year) of the
birth date in X/Y **/
/** prov_num: provider number (hospital ID)**/
/** No missing of these variables in X or Y**/
/*******************************************************/

/*-------------------------------------------------------
Step 1 all 7 factors
-------------------------------------------------------*/;
proc sql noprint;
create table temp as select a.*,b.*
from  x as a
left join Y as b
on compress(a.prov_num) = compress(b.prov_num) and
   a.gender = b.gender and
   a.adate_x= b.adate_y and
   a.bdate_x= b.bdate_y and
   a.ddate_x= b.ddate_y and
   a.zip_x= b.zip_y  and
   a.race_x = b.race_y
;
quit;

data matched1;
set TEMP;
where missing(Y_ID) = 0 ;
length Matched $120.;
matched = '1 all factors' ;
drop adate_x bdate_x ddate_x zip_x race_x
     bday_x bday_y
      bmonth_x bmonth_y
      byear_x byear_y ;
run;

data matched;
set matched1 ;
run;
```

```
/*-----------------------------------------
 Step #2 without race
-----------------------------------------*/

proc sql noprint;
create table finder_x as select distinct * from x where X_ID not in
(select X_ID from  matched);
create table master as select distinct * from Y where Y_ID not in
(select Y_ID from  matched);

create table TEMP as select a.*, b.*
from finder_x as a
left join master as b
on a.prov_num = b.prov_num and
   a.gender = b.gender and
   a.adate_x= b.adate_y and
   a.bdate_x= b.bdate_y and
   a.ddate_x= b.ddate_y and
   a.zip_x= b.zip_y
;
quit;

data matched2;
set TEMP;
where missing(Y_ID) = 0 ;
length Matched $120.;
matched = '2 wo race' ;
drop adate_x bdate_x ddate_x zip_x race_x
     bday_x bday_y
      bmonth_x bmonth_y
      byear_x byear_y ;
run;

data matched;
set matched1 matched2 ;
run;


/*-----------------------------------------
 Step #3 without zipcode
-----------------------------------------*/

proc sql noprint;
create table finder_x as select distinct * from x where X_ID not in
(select X_ID from  matched);
create table master as select distinct * from Y where Y_ID not in
(select Y_ID from  matched);

create table TEMP as select a.*, b.*
from finder_x as a
left join master as b
on a.prov_num = b.prov_num and
```

```
   a.gender = b.gender and
   a.adate_x= b.adate_y and
   a.bdate_x= b.bdate_y and
   a.ddate_x= b.ddate_y and
   a.race_x = b.race_y
;
quit;

data matched3;
set TEMP;
where missing(Y_ID) = 0 ;
length Matched $120.;
matched = '3 wo zip' ;
drop adate_x bdate_x ddate_x zip_x race_x
     bday_x bday_y
      bmonth_x bmonth_y
      byear_x byear_y ;
run;

data matched;
set matched1 matched2 matched3 ;
run;



/*-------------------------------------------------------------
 Step #4 open an 1-day window for dates of admission and discharge
-------------------------------------------------------------*/
proc sql noprint;
create table finder_x as select distinct * from x where X_ID not in
(select X_ID from  matched);
create table master as select distinct * from Y where Y_ID not in
(select Y_ID from  matched);

create table TEMP as select a.*, b.*
from finder_x as a
left join master as b
on a.prov_num = b.prov_num and
   a.gender = b.gender and
   abs(a.adate_x - b.adate_y) < 2 and
   abs(a.ddate_x - b.ddate_y) < 2 and
   a.bdate_x= b.bdate_y and
   a.race_x = b.race_y  and
   a.zip_x= b.zip_y
;
quit;

data matched4;
set TEMP;
where missing(Y_ID) = 0 ;
length Matched $120.;
matched = '4 7 factor with 1d win for a/d date' ;
```

```
drop adate_x bdate_x ddate_x zip_x race_x
     bday_x bday_y
      bmonth_x bmonth_y
      byear_x byear_y ;
run;

data matched;
set matched1 - matched4 ;
run;


/*-----------------------------------------------------------
 Step #5 open a 2-day window for dates of admission and discharge
-----------------------------------------------------------*/
proc sql noprint;
create table finder_x as select distinct * from x where X_ID not in
(select X_ID from  matched);
create table master as select distinct * from Y where Y_ID not in
(select Y_ID from  matched);

create table TEMP as select a.*, b.*
from finder_x as a
left join master as b
on a.prov_num = b.prov_num and
   a.gender = b.gender and
   abs(a.adate_x - b.adate_y) < 3 and
   abs(a.ddate_x - b.ddate_y) < 3 and
   a.bdate_x= b.bdate_y and
   a.race_x = b.race_y  and
   a.zip_x= b.zip_y
;
quit;

data matched5;
set TEMP;
where missing(Y_ID) = 0 ;
length Matched $120.;
matched = '5 7 factor with 2d win for a/d date' ;
drop adate_x bdate_x ddate_x zip_x race_x
     bday_x bday_y
      bmonth_x bmonth_y
      byear_x byear_y ;
run;

data matched;
set matched1 - matched5 ;
run;

/*-----------------------------------------------------------
 Step #6 open a 7-day window for dates of admission and discharge
-----------------------------------------------------------*/
proc sql noprint;
```

```
create table finder_x as select distinct * from x where X_ID not in
(select X_ID from  matched);
create table master as select distinct * from Y where Y_ID not in
(select Y_ID from  matched);

create table TEMP as select a.*, b.*
from finder_x as a
left join master as b
on a.prov_num = b.prov_num and
   a.gender = b.gender and
   abs(a.adate_x - b.adate_y) < 8 and
   abs(a.ddate_x - b.ddate_y) < 8 and
   a.bdate_x= b.bdate_y and
   a.race_x = b.race_y  and
   a.zip_x= b.zip_y
;
quit;

data matched6;
set TEMP;
where missing(Y_ID) = 0 ;
length Matched $120.;
matched = '6 7 factor with 7d win for a/d date' ;
drop adate_x bdate_x ddate_x zip_x race_x
     bday_x bday_y
      bmonth_x bmonth_y
      byear_x byear_y ;
run;

data matched;
set matched1 - matched6 ;
run;

/*-------------------------------------------------------------
 Step #7 without zipcode and 1-day window for dates of admission and
discharge
-------------------------------------------------------------*/
proc sql noprint;
create table finder_x as select distinct * from x where X_ID not in
(select X_ID from  matched);
create table master as select distinct * from Y where Y_ID not in
(select Y_ID from  matched);

create table TEMP as select a.*, b.*
from finder_x as a
left join master as b
on a.prov_num = b.prov_num and
   a.gender = b.gender and
   abs(a.adate_x - b.adate_y) < 2 and
   abs(a.ddate_x - b.ddate_y) < 2 and
   a.bdate_x= b.bdate_y and
   a.race_x = b.race_y
```

```
;
quit;

data matched7;
set TEMP;
where missing(Y_ID) = 0 ;
length Matched $120.;
matched = '7 wo zip and 1d win for a/d date' ;
drop adate_x bdate_x ddate_x zip_x race_x
     bday_x bday_y
      bmonth_x bmonth_y
      byear_x byear_y ;
run;

data matched;
set matched1 - matched7 ;
run;

/*-------------------------------------------------------------
 Step #8 without race and 1-day window for dates of admission and
discharge
-------------------------------------------------------------*/
proc sql noprint;
create table finder_x as select distinct * from x where X_ID not in
(select X_ID from  matched);
create table master as select distinct * from Y where Y_ID not in
(select Y_ID from  matched);

create table TEMP as select a.*, b.*
from finder_x as a
left join master as b
on a.prov_num = b.prov_num and
   a.gender = b.gender and
   abs(a.adate_x - b.adate_y) < 2 and
   abs(a.ddate_x - b.ddate_y) < 2 and
   a.bdate_x= b.bdate_y and
   a.zip_x= b.zip_y
;
quit;

data matched8;
set TEMP;
where missing(Y_ID) = 0 ;
length Matched $120.;
matched = '8 wo race and 1d win for a/d date' ;
drop adate_x bdate_x ddate_x zip_x race_x
     bday_x bday_y
      bmonth_x bmonth_y
      byear_x byear_y ;
run;

data matched;
```

```
set matched1 - matched8 ;
run;

/*-------------------------------------------------------
 Step #9 without race, zip, and 1-day window for dates of admission
and discharge
-------------------------------------------------------*/
proc sql noprint;
create table finder_x as select distinct * from x where X_ID not in
(select X_ID from  matched);
create table master as select distinct * from Y where Y_ID not in
(select Y_ID from  matched);

create table TEMP as select a.*, b.*
from finder_x as a
left join master as b
on a.prov_num = b.prov_num and
   a.gender = b.gender and
   abs(a.adate_x - b.adate_y) < 2 and
   abs(a.ddate_x - b.ddate_y) < 2 and
   a.bdate_x= b.bdate_y
;
quit;

data matched9;
set TEMP;
where missing(Y_ID) = 0 ;
length Matched $120.;
matched = '9 wo race/zip and allow 1d win for a/d date' ;
drop adate_x bdate_x ddate_x zip_x race_x
     bday_x bday_y
      bmonth_x bmonth_y
      byear_x byear_y ;
run;

data matched;
set matched1 - matched9 ;
run;


/*-------------------------------------------------------
 Step #10 without race, zip, discharge
-------------------------------------------------------*/
proc sql noprint;
create table finder_x as select distinct * from x where X_ID not in
(select X_ID from  matched);
create table master as select distinct * from Y where Y_ID not in
(select Y_ID from  matched);

create table TEMP as select a.*, b.*
from finder_x as a
left join master as b
```

```
on a.prov_num = b.prov_num and
   a.gender = b.gender and
   a.adate_x = b.adate_y and
   a.bdate_x= b.bdate_y
;
quit;


data matched10;
set TEMP;
where missing(Y_ID) = 0 ;
length Matched $120.;
matched = '10 wo race zip discharge date' ;
drop adate_x bdate_x ddate_x zip_x race_x
     bday_x bday_y
      bmonth_x bmonth_y
      byear_x byear_y ;
run;

data matched;
set matched1 - matched10 ;
run;

/*-------------------------------------------------------------
 Step #11 without race, zip, admission
-------------------------------------------------------------*/
proc sql noprint;
create table finder_x as select distinct * from x where X_ID not in
(select X_ID from  matched);
create table master as select distinct * from Y where Y_ID not in
(select Y_ID from  matched);

create table TEMP as select a.*, b.*
from finder_x as a
left join master as b
on a.prov_num = b.prov_num and
   a.gender = b.gender and
   a.ddate_x = b.ddate_y and
   a.bdate_x= b.bdate_y
;
quit;

data matched11;
set TEMP;
where missing(Y_ID) = 0 ;
length Matched $120.;
matched = '11 wo race zip admission date' ;
drop adate_x bdate_x ddate_x zip_x race_x
     bday_x bday_y
      bmonth_x bmonth_y
      byear_x byear_y ;
run;
```

```
data matched;
set matched1 - matched11 ;
run;

/*-----------------------------------------------------------
 Step #12 without race, zip, 2 out of 3 dob
-----------------------------------------------------------*/
proc sql noprint;
create table finder_x as select distinct * from x where X_ID not in
(select X_ID from  matched);
create table master as select distinct * from Y where Y_ID not in
(select Y_ID from  matched);

create table TEMP as select a.*, b.*
from finder_x as a
left join master as b
on a.prov_num = b.prov_num and
   a.gender = b.gender and
   a.ddate_x = b.ddate_y and
   a.adate_x = b.adate_y and
   ((a.bday_x =  b.bday_y and a.bmonth_x = b.bmonth_y) or (a.bday_x =
b.bday_y and a.byear_x = b.byear_y) or ( a.bmonth_x = b.bmonth_y and
a.byear_x = b.byear_y))
;
quit;

data matched12;
set TEMP;
where missing(Y_ID) = 0 ;
length Matched $120.;
matched = '12 wo race zip and 2/3 element of dob' ;
drop adate_x bdate_x ddate_x zip_x race_x
     bday_x bday_y
      bmonth_x bmonth_y
      byear_x byear_y ;
run;

data matched;
set matched1 - matched12 ;
run;

/*-----------------------------------------------------------
 Step #13 without race, zip, 2 out of 3 dob, 1d win for a/d date
-----------------------------------------------------------*/
proc sql noprint;
create table finder_x as select distinct * from x where X_ID not in
(select X_ID from  matched);
create table master as select distinct * from Y where Y_ID not in
(select Y_ID from  matched);

create table TEMP as select a.*, b.*
```

```
from finder_x as a
left join master as b
on a.prov_num = b.prov_num and
   a.gender = b.gender and
   abs(a.ddate_x - b.ddate_y) <1 and
   abs(a.adate_x - b.adate_y) <1 and
   ((a.bday_x =  b.bday_y and a.bmonth_x = b.bmonth_y) or (a.bday_x =
b.bday_y and a.byear_x = b.byear_y) or ( a.bmonth_x = b.bmonth_y and
a.byear_x = b.byear_y))
;
quit;

data matched13;
set TEMP;
where missing(Y_ID) = 0 ;
length Matched $120.;
matched = '13 wo race zip and 2/3 element of dob, 1d win' ;
drop adate_x bdate_x ddate_x zip_x race_x
     bday_x bday_y
      bmonth_x bmonth_y
      byear_x byear_y ;
run;

data final_matched;
set matched1 - matched13 ;
run;
```