

1 **TITLE: Assessment of Sociodemographic Factors Associated with Time to Self-reported**  
2 **COVID-19 Infection Among a Large Multi-Center Prospective Cohort Population in the**  
3 **Southeastern United States**

4 **Short title: Sociodemographic Factors and Self-Reported COVID-19 in a Southeastern**  
5 **U.S. Cohort Study**

6 Andrew J. Beron<sup>1</sup>, Joshua O. Yukich<sup>1</sup>, Andrea A. Berry<sup>2</sup>, Adolfo Correa<sup>3</sup>, Joseph Keating<sup>1</sup>,  
7 Matthew Bott<sup>4</sup>, Thomas F Wierzba<sup>5</sup>, William S. Weintraub<sup>6,7</sup>, DeAnna J. Friedman-Klabanoff<sup>2</sup>,  
8 Morgana Mongraw-Chaffin<sup>8</sup>, Michael A. Gibbs<sup>9</sup>, Yheneko J. Taylor<sup>10</sup>, Patricia J. Kissinger<sup>11</sup>,  
9 Devin Hayes<sup>12</sup>, John S. Schieffelin<sup>13</sup>, Brian K. Burke<sup>4</sup>, Richard A. Oberhelman<sup>1\*</sup>

10 1. Department of Tropical Medicine, Tulane School of Public Health and Tropical Medicine,  
11 New Orleans, LA, USA

12  
13 2. Center for Vaccine Development and Global Health and Department of Pediatrics,  
14 University of Maryland School of Medicine, Baltimore, MD, USA

15  
16 3. Department of Population Health Science, University of Mississippi Medical Center,  
17 Jackson, MS, USA

18  
19 4. Milken Institute School of Public Health; The George Washington University,  
20 Washington, DC, USA

21  
22 5. Section of Infectious Diseases Department of Internal Medicine, Wake Forest University  
23 School of Medicine, Winston-Salem, NC, USA

24  
25 6. MedStar Health Research Institute, Hyattsville, MD, USA

26  
27 7. Georgetown University, Washington, DC, USA

28  
29 8. Department of Epidemiology and Prevention, Wake Forest University School of  
30 Medicine, Winston-Salem, NC, USA

31  
32 9. Department of Emergency Medicine, Atrium Health, Charlotte, NC, USA

33  
34 10. Center for Health System Sciences, Atrium Health, Charlotte, NC, USA

35  
36 11. Department of Epidemiology, Tulane School of Public Health and Tropical Medicine,  
37 New Orleans, LA, USA

38  
39 12. Vysnova Partners, Landover, MD, USA

40  
41 13. Department of Pediatrics, Tulane School of Medicine, New Orleans, LA, USA

42  
43 \* Corresponding author Email: oberhel@tulane.edu

44 **Abstract**

45

46 **Objective:** We aimed to investigate sociodemographic factors associated with self-  
47 reported COVID-19 infection.

48 **Methods:** The study population is a multicenter prospective cohort of adult volunteers  
49 recruited from healthcare systems located in the mid-Atlantic and southern United  
50 States. Between April 2020 and October 2021 participants completed daily online  
51 questionnaires about symptoms, exposures, and risk behaviors related to COVID-19,  
52 including self-reports of positive SARS CoV-2 detection tests and COVID-19  
53 vaccination. Analysis of time from study enrollment to self-reported COVID-19 infection  
54 used a time-varying mixed effects Cox-proportional hazards framework.

55 **Results:** Overall, 1,603 of 27,214 study participants (5.9%) reported a positive COVID-  
56 19 test during the study period. The adjusted hazard ratio demonstrated lower risk for  
57 women, those with a graduate level degree, and smokers. A higher risk was observed  
58 for healthcare workers, those aged 18-34, those in rural areas, those from households  
59 where a member attends school or interacts with the public, and those who visited a  
60 health provider in the last year.

61 **Conclusions:** Increased risk of self-reported COVID-19 was associated with specific  
62 demographic characteristics, which may help to inform targeted interventions for future  
63 pandemics.

64

65

66

67 **Abbreviations and Acronyms**

68 CCRP: COVID-19 Community Research Partnership

69 CDC: Centers for Disease Control and Prevention

70 HER: Electronic Health Records

71 IRB: Institutional Review Board

72

73

74

## 75 INTRODUCTION

76 The COVID-19 pandemic has created and widened existing health disparities within US  
77 society. Epidemiologic data collected during COVID-19 surveillance provides  
78 important insights on at-risk populations including differential access to information and  
79 infection control measures, in particular social distancing, masking, and new vaccines to  
80 guide future pandemic preparedness. Regional studies carried out by local U.S.  
81 healthcare and public health organizations, as well as national cross-sectional  
82 studies<sup>1,2,3</sup> have found higher incidence in ethnic and race-based minority populations  
83 and certain age groups.<sup>4</sup> Nonetheless, there is a paucity of comprehensive risk factor  
84 analyses beyond demographic characteristics that examine other factors such as  
85 occupation, self-reported health conditions, health behaviors, and household  
86 characteristics.

87 The COVID-19 Community Research Partnership (CCRP) was a multi-state cohort  
88 study designed to monitor the evolution of the pandemic in a large population with both  
89 syndromic surveillance and periodic testing for serologic evidence of infection. The goal  
90 of the CCRP was to generate data to inform ongoing public health responses to COVID-  
91 19 as well as future pandemics by recruiting a diverse cohort of patients and community  
92 members. The methods and purpose have been described elsewhere.<sup>5</sup>

93

94 This large, multicenter study provided an opportunity to further examine population-  
95 based risk factors for COVID-19 to identify characteristics of subgroups most likely to  
96 become infected. As such, the purpose of this paper is to investigate factors associated  
97 with time from enrollment in follow-up to a self-reported COVID-19 infection. Herein we

98 report findings of our risk factor analysis comparing hazard rates in subgroups based on  
99 individual and household characteristics.

100

## 101 **MATERIALS AND METHODS**

102 Participants were drawn from the prospective, multi-site, CCRP COVID-19 surveillance  
103 cohort study, a convenience sample of patients and healthcare workers in ten  
104 healthcare systems from the Mid-Atlantic and southeastern US. We recruited adults  
105 through patient portals or email from the following health systems and institutions: Wake  
106 Forest Baptist Health, Atrium Health, Wake Med, New Hanover Regional Medical  
107 Center, Vidant Health, Campbell University, Tulane University affiliated partner systems,  
108 University of Mississippi, University of Maryland Medical System, and Medstar Health.  
109 Adults aged 18 years and older were eligible to participate if they were a patient or  
110 employee of a participating healthcare system. Only participants who reported residing  
111 in Maryland, Virginia, Washington D.C., North Carolina, South Carolina, Mississippi, and  
112 Louisiana were included. Participants who reported a prior positive test for COVID-19  
113 on enrollment were excluded.

114

115 Study procedures included daily online questionnaires for all participants, extracting  
116 electronic health records (EHR) for those who were patients at a participating health  
117 system, and periodic at-home serological testing for a subset of participants. The Wake  
118 Forest Baptist Health Institutional Review Board (IRB), which served as the central IRB  
119 for this study, approved the study protocol. Internet-based informed consent through  
120 secure patient portals was obtained from study participants prior to any study

121 procedures. The study is registered with ClinicalTrials.gov, NCT04342884. Enrollment  
122 began April 8, 2020, and ended on April 29, 2022.

123

#### 124 **Participant data**

125 Participants in the study completed daily online questionnaires about symptoms,  
126 exposures, and risk avoidance behaviors related to COVID-19, including self-reports of  
127 recent positive SARS CoV-2 detection tests (herein described as “self-reported COVID-  
128 19”). Respondents also reported history of COVID-19 vaccination (date of receipt,  
129 product, dose 1 or dose 2, participation in a clinical trial). Race was defined based on  
130 responses to the initial study enrollment questionnaire, with options listed as 1) Black or  
131 African American, 2) Asian, 3) Hispanic or Latino, 4) White (not Hispanic/Latino), 5)  
132 American Indian or Alaskan Native, and 6) Mixed Ethnicity. Participants were invited to  
133 complete two online supplemental questionnaires, one that was focused on the  
134 individual and another that was focused on the individual’s household, to provide more  
135 detailed information on demographic characteristics, occupations, self-reported health  
136 conditions, health behaviors, and household characteristics. Supplemental  
137 questionnaires were sent to all actively enrolled participants in May 2021, and  
138 subsequent newly enrolled study participants received the surveys within a month of  
139 starting the study.

140

141 The study sought to examine correlations between the occupational characteristics of  
142 participants and the primary outcome. The National Institute for Occupational Safety  
143 and Health (NIOSH), a United States federal agency responsible for conducting

144 research and making recommendations for the prevention of work-related injury and  
145 illness, has characterized workplace exposure to SARS-CoV-2 in hundreds of non-  
146 health care occupations using metrics from O\*NET, a national database with  
147 information on occupational characteristics, together with input from experts in  
148 occupational safety and health. Based on these data, occupations are categorized in  
149 three risk levels (i.e., high, medium, low) using the SARS-CoV-2 Occupational Exposure  
150 Matrix (SOEM) system.<sup>6</sup> SOEM exposure categories were based on three factors  
151 identified as contributing to increased risk of exposure in the workplace: whether an  
152 occupation involves routine in-person interaction with the public (*Public Facing*), working  
153 indoors (*Working Indoors*), and working in close physical proximity to others, either co-  
154 workers or the public (*Close Proximity*). Since health care workers were over-  
155 represented in the study population, the high exposure group was divided into two  
156 groups for data analysis, i.e., high-healthcare, and high-non healthcare.

157

## 158 **Data analysis**

159 Data from participants who completed both supplemental questionnaires were analyzed  
160 to determine demographic, occupational, health-related, and behavioral correlates of  
161 self-reported SARS CoV-2 infection. Descriptive statistics were produced by cross  
162 tabulation and covariates for model inclusion were first checked for collinearity using  
163 pairwise correlation. Analysis followed a time-varying mixed effects Cox-proportional  
164 hazards framework, with a shared frailty at the level of recruitment site to account for  
165 homogeneity within each site/health network and to account for intraclass correlation in  
166 the outcome of interest (self-reported COVID-19 infection) within health networks.

167 Hazard Ratios and 95% confidence intervals are reported as unadjusted estimates and  
168 as adjusted for all covariates included in the final model (Table 2). Survival-Curves are  
169 also presented using the Kaplan-Meier Product Limit estimator for selected covariates.

170

171 Two time-varying covariates were included in the categorical hazard analysis: 1) county  
172 level 7-day average COVID-19 incidence data updated daily and published by the New  
173 York Times in 2020 and 2021 <sup>7</sup> and 2) COVID-19 vaccination status of participants.  
174 Participants were considered vaccinated after they reported receiving their first  
175 vaccination dose of any COVID-19 vaccine.

176

## 177 **RESULTS**

178 A total of 69,714 participants were enrolled in the CCRP study. Of those, 42,701  
179 (61.3%) completed the individual adult supplemental survey and 31,642 (45.4%)  
180 completed the household supplemental survey. Just under thirty thousand (29,973  
181 (43.0%)) participants completed both the individual adult and household supplemental  
182 surveys. Participants that reported residing outside of Maryland, Virginia, Washington  
183 D.C., North Carolina, South Carolina, Mississippi, and Louisiana, as well as participants  
184 that were participating in clinical trials were excluded, leaving a total of 27,214  
185 participants (39% of the total number of CCRP study participants) in the analytic  
186 population.

187

188 Overall, 1,603 of a total of the 27,214 study participants (5.9%) reported that they had a  
189 positive test for COVID-19 between enrollment and the end of October 2021 (Table 1).

190 The study population was predominantly female (71.4%) and White/non-Hispanic



**Table 1. Characteristics at baseline of participants enrolled in the COVID-19 Community Research Panel included for analysis by Self-Reported COVID-19 Status**

Variable	Overall number of participants*	No*	Yes*	Percentage with Self-Reported COVID-19 Diagnosis
Total subjects	27,214	25,611	1,603	5.9
Female	19,443 (71.4)	18,266 (71.3)	1,177 (73.4)	6.1
Age group				
18-34	2,911 (10.7)	2,675 (10.4)	236 (14.7)	8.1
35-49	7,395 (27.2)	6,808 (26.6)	587 (36.6)	7.9
50-64	9,388 (34.5)	8,832 (34.5)	556 (34.7)	5.9
65+	7,520 (27.6)	7,296 (28.5)	224 (14.0)	3
Race/Ethnicity				
White NH	24,032 (88.3)	22,582 (88.2)	1,450 (90.5)	6
Black NH	1,439 (5.3)	1,377 (5.4)	62 (3.9)	4.3
Hispanic	604 (2.2)	563 (2.2)	41 (2.6)	6.8
Asian	488 (1.8)	470 (1.8)	18 (1.1)	3.7
Other	651 (2.4)	619 (2.4)	32 (2.0)	4.9
County Classification				
Rural	5,587 (20.5)	5,184 (20.2)	403 (25.1)	7.2
Suburban	6,355 (23.4)	5,950 (23.2)	405 (25.3)	6.4
Urban	15,271 (56.1)	14,476 (56.5)	795 (49.6)	5.2
Health Care Worker lives in Household	6,006 (22.1)	5,441 (21.2)	565 (35.2)	9.4
Education Level of Respondent				
Hs or less	845 (3.1)	782 (3.1)	63 (3.9)	7.5
Some college	2,756 (10.1)	2,564 (10.0)	192 (12.0)	7
Associate	2,090 (7.7)	1,882 (7.3)	208 (13.0)	10
Bachelor	8,483 (31.2)	7,927 (31.0)	556 (34.7)	6.6
Graduate	13,040 (47.9)	12,456 (48.6)	584 (36.4)	4.5
History of Influenza Vaccination	26,338 (96.8)	24,800 (96.8)	1,538 (95.9)	5.8
Time Since Last Primary Care Visit				
<1 Year	22,590 (83.0)	21,232 (82.9)	1,358 (84.7)	6
1-2 Years Ago	3,311 (12.2)	3,137 (12.2)	174 (10.9)	5.3
2-5 Years Ago	912 (3.4)	871 (3.4)	41 (2.6)	4.5
>5 Years Ago	401 (1.5)	371 (1.4)	30 (1.9)	7.5
Current Tobacco Use of Respondent	1,485 (5.5)	1,410 (5.5)	75 (4.7)	5.1
Number of Concurrent Medical Conditions Reported				
None	9,073 (33.3)	8,520 (33.3)	553 (34.5)	6.1
One	8,961 (32.9)	8,431 (32.9)	530 (33.1)	5.9
Two	5,603 (20.6)	5,294 (20.7)	309 (19.3)	5.5
3 or More	3,577 (13.1)	3,366 (13.1)	211 (13.2)	5.9
Household Annual Income				
<50k	2,749 (10.1)	2,595 (10.1)	154 (9.6)	5.6
50-100k	7,061 (25.9)	6,559 (25.6)	502 (31.3)	7.1
>100k	14,732 (54.1)	13,924 (54.4)	808 (50.4)	5.5
No answer	2,672 (9.8)	2,533 (9.9)	139 (8.7)	5.2
Someone in the Household Attends Class in Person	6,911 (25.4)	6,315 (24.7)	596 (37.2)	8.6
Social Contact Worker in the Household	10,279 (37.8)	9,471 (37.0)	808 (50.4)	7.9
Study Site				
Atrium	4,662 (17.1)	4,221 (16.5)	441 (27.5)	9.5
Campbell	247 (0.9)	238 (0.9)	9 (0.6)	3.6
Maryland	3,366 (12.4)	3,298 (12.9)	68 (4.2)	2
MedStar	7,265 (26.7)	7,100 (27.7)	165 (10.3)	2.3
Mississippi	182 (0.7)	172 (0.7)	10 (0.6)	5.5
New Hanover	438 (1.6)	408 (1.6)	30 (1.9)	6.8
Tulane	79 (0.3)	71 (0.3)	8 (0.5)	10.1

**Table 1. Characteristics at baseline of participants enrolled in the COVID-19 Community Research Panel included for analysis by Self-Reported COVID-19 Status**

Variable	Overall number of participants*	No*	Yes*	Percentage with Self-Reported COVID-19 Diagnosis
Vidant	619 (2.3)	585 (2.3)	34 (2.1)	5.5
Wake Forest	8,883 (32.6)	8,132 (31.8)	751 (46.8)	8.5
Wake Med	1,473 (5.4)	1,386 (5.4)	87 (5.4)	5.9
SOEM Exposure Risk Categorization of Subjects Occupation				
High - Healthcare	5,080 (20.1)	4,587 (19.4)	493 (31.4)	9.7
High - Other	5,179 (20.5)	4,849 (20.5)	330 (21.0)	6.4
Low	4,316 (17.1)	4,068 (17.2)	248 (15.8)	5.7
Medium	2,100 (8.3)	1,948 (8.2)	152 (9.7)	7.2
No Data	8,577 (34.0)	8,231 (34.8)	346 (22.1)	4

\* Values are all n (%) unless otherwise noted. No = Did not report COVID-19 infection during surveillance period; Yes = Did report COVID-19 infection during surveillance period

192  
193  
194 (88.3%; Table 1). Most participants lived in counties classified as urban (56.1%) and  
195 96.9% had at least some college education with a large proportion holding graduate  
196 level degrees (47.9%). The study population was affluent (54.1% had a household  
197 income over \$100,000). Data was not available for SOEM category designation for 32%  
198 of subjects, but among the other participants there was an even distribution between the  
199 four SOEM exposure groups (low, medium, high-non healthcare, and high-healthcare).  
200 The networks with the largest numbers of participants included Wake Forest (32.6%),  
201 MedStar (26.7%), and Atrium (17.1%).

202  
203 Adjusted and unadjusted hazard ratios for self-reported COVID-19 are shown in Table  
204 2. Despite the fact that a higher proportion of women in the study population reported  
205 COVID-19 infections as compared to men (6.1% vs. 5.5%), the adjusted hazard ratio  
206 demonstrated lower risk for women (aHR= 0.87 [0.77-0.98]). When compared to the  
207 risk of COVID-19 among participants ages 18-34, significantly lower risk for infection

**Table 2. Determinants of Self-Reported COVID-19 Diagnosis following a time-varying mixed-effects Cox-proportional hazards model with shared frailty for health system**

Term	Unadjusted Hazard Ratio (95% CI)	Adjusted Hazard Ratio (95% CI)
<b>Sex</b>		
Male	1.0	1.0
Female	1.07 (0.96 - 1.19)	0.88 (0.78 - 0.99)*
<b>Age</b>		
Age 18-34	1.0	1.0
Age 35-49	0.93 (0.79 - 1.09)	0.79 (0.67 - 0.92)*
Age 50-64	0.76 (0.62 - 0.94)*	0.61 (0.52 - 0.71)*
Age 65+	0.69 (0.56 - 0.85)*	0.44 (0.35 - 0.54)*
<b>Race-Ethnicity</b>		
Non-Hispanic White	1.0	1.0
Non-Hispanic Black	0.88 (0.68 - 1.14)	0.85 (0.66 - 1.10)
Hispanic	1.31 (0.96 - 1.79)	1.11 (0.81 - 1.51)
Asian	0.71 (0.45 - 1.13)	0.79 (0.49 - 1.25)
Other Race/Ethnicity	1.00 (0.71 - 1.42)	0.85 (0.59 - 1.21)
<b>Urban-Rural</b>		
County Rural	1.0	1.0
County Suburban	0.92 (0.80 - 1.06)	0.97 (0.84 - 1.12)
County Urban	0.66 (0.58 - 0.75)*	0.88 (0.78 - 1.01)
<b>Education</b>		
High School or Less	1.0	1.0
Some college	0.94 (0.71 - 1.25)	0.85 (0.64 - 1.14)
Associate Degree	1.21 (0.91 - 1.60)	0.95 (0.72 - 1.27)
Bachelor Degree	0.87 (0.67 - 1.13)	0.78 (0.59 - 1.02)
Graduate Degree	0.64 (0.49 - 0.83)	0.57 (0.44 - 0.76)*
<b>Influenza Vaccine History</b>		
Never Had Influenza Vaccine	1.0	1.0

**Table 2. Determinants of Self-Reported COVID-19 Diagnosis following a time-varying mixed-effects Cox-proportional hazards model with shared frailty for health system**

Term	Unadjusted Hazard Ratio (95% CI)	Adjusted Hazard Ratio (95% CI)
Ever Had Influenza Vaccine	0.78 (0.61 - 1.00)	0.93 (0.72 - 1.20)
<b>Time Since Last Primary Care Visit</b>		
Last Primary Care Visit < 1 Year Ago	1.0	1.0
Last Primary Care Visit 1-2 Years Ago	0.94 (0.80 - 1.10)	0.84 (0.71 - 0.98)*
Last Primary Care Visit 2-5 Years Ago	0.84 (0.62 - 1.15)	0.68 (0.49 - 0.94)*
Last Primary Care Visit >5 Years Ago	1.32 (0.92 - 1.89)	1.10 (0.77 - 1.59)
<b>Tobacco Use</b>		
Not A Current Tobacco User	1.0	1.0
Current Tobacco User	0.88 (0.70 - 1.11)	0.73 (0.58 - 0.92)*
<b>Co-morbidities</b>		
No Concurrent Medical Conditions	1.0	1.0
One Concurrent Medical Condition	0.98 (0.87 - 1.10)	1.01 (0.90 - 1.14)
Two Concurrent Medical Conditions	0.91 (0.79 - 1.04)	1.01 (0.88 - 1.17)
3 or More Concurrent Medical Conditions	0.99 (0.85 - 1.16)	1.18 (0.99 - 1.40)
<b>Household Income</b>		
Household Income <50k	1.0	1.0
Household income 50-100k	1.27 (1.06 - 1.52)*	1.22 (1.01 - 1.47)*
Household income >100k	1.11 (0.93 - 1.32)	1.14 (0.94 - 1.38)
Household income No answer	0.96 (0.76 - 1.21)	1.04 (0.82 - 1.33)
<b>In-person class</b>		
No one in Household attends in person class	1.0	1.0
Someone in Household attends in person class	1.64 (1.48 - 1.82)*	1.28 (1.14 - 1.43)*
<b>Social Contact Occupation</b>		
No one in Household works in a Social Contact Occupation	1.0	1.0
Someone in Household works in a Social Contact Occupation	1.60 (1.45 - 1.77)*	1.24 (1.12 - 1.38)*
<b>COVID-19 Vaccination</b>		

**Table 2. Determinants of Self-Reported COVID-19 Diagnosis following a time-varying mixed-effects Cox-proportional hazards model with shared frailty for health system**

Term	Unadjusted Hazard Ratio (95% CI)	Adjusted Hazard Ratio (95% CI)
Has not yet had any COVID-19 Vaccine Doses	1.0	1.0
Had At Least One Vaccine Dose	0.26 (0.23 - 0.30) *	0.37 (0.32 - 0.42) *
<b>SOEM Risk Class</b>		
SOEM - Class High Health Care Work	1.0	1.0
SOEM - Class High	0.85 (0.74 - 0.98) *	0.80 (0.69 - 0.93) *
SOEM Class Medium	0.94 (0.78 - 1.13) *	0.80 (0.66 - 0.97) *
SOEM - Class Low	0.81 (0.69 - 0.94) *	0.74 (0.63 - 0.87) *
SOEM Class No Data	0.47 (0.41 - 0.54) *	0.58 (0.49 - 0.68) *

\*  $p \leq 0.05$ ; Adjusted HR are adjusted for Sex, Age, Race-Ethnicity, Urban-Rural, Education, Influenza Vaccine History, Time Since Last Primary Care Visit, Tobacco Use, Co-morbidities, Household Income, In-person class, Social Contact Occupation, COVID-19 Vaccination and SOME Risk Class.

208

209 were seen in all other age groups (vs. ages 35-49 aHR=0.80; vs. ages 50-64  
210 aHR=0.63; vs. ages >65 aHR=0.45). No difference in risk of COVID-19 was seen  
211 based on race or ethnicity. A lower risk of self-reported infection for participants living  
212 in urban counties compared to rural counties seen in unadjusted analysis was  
213 attenuated and borderline significant in adjusted analysis (aHR=0.88 [0.77-1.00]).  
214 Educational level was strongly associated with risk, with significantly lower risk for  
215 participants with with a graduate level degree (aHR=0.57 [0.44-0.76]) compared to  
216 those with no college. As compared to participants who had seen their primary care  
217 provider within the past year, lower risk was observed for those whose last primary  
218 care visit was 1-2 years ago (aHR=0.84 [0.71-0.98]) or 2-5 years ago (aHR=0.68  
219 [0.49-0.94]). Smokers had significantly lower risk of infection as compared to non-  
220 smokers (aHR=0.73 [0.58 - 0.92]). Risk did not correlate with the number of self-

221 reported health conditions, but there was a borderline significant increase in risk for  
222 participants with 3 or more chronic health conditions as compared to those who  
223 reported no chronic conditions (aHR=1.18 [0.99 - 1.40]). Household income was not  
224 significantly associated with risk. Higher risk was observed in households where  
225 someone attended classes in-person (aHR=1.23 [1.11-1.37]) and in households  
226 where someone had occupational contact with the general public (aHR=1.24 [1.12 -  
227 1.38]; Figure 1a). Receipt of at least one dose of COVID-19 vaccine was strongly  
228 protective (aHR=0.37; Figure 1b). In adjusted analyses based on SOEM  
229 occupational risk level, healthcare workers in the high-risk group had higher risk as  
230 compared to the high risk/non healthcare group (aHR=0.79 [0.68-0.92]), the medium  
231 risk group (aHR=0.82 [0.68-1.00]), and the low risk group (aHR=0.75 [0.64-0.89];  
232 Figure 1c).

233

## 234 **DISCUSSION**

235 Our multicenter COVID-19 Community Research Partnership study provided a unique  
236 opportunity to examine the effect of demographic factors on risk of SARS CoV-2  
237 infections in populations across the southern United States. After adjustment, we found  
238 that multiple social and economic factors were strongly associated with self-reported  
239 SARS CoV-2 infection during the pandemic. Risk of infection was significantly higher in  
240 young adult participants ages 18-34 years as compared to older groups, and the hazard  
241 ratios indicated that risk of infection as compared to the youngest participant group  
242 decreased further with each sequentially older age stratum, which is consistent with  
243 findings from other studies.<sup>8,9</sup> This observation may be related to increasing concern

244 about disease outcome in older age groups leading to greater adoption of preventative  
245 behaviors among older age cohorts, based on data that were well publicized in the  
246 media. In contrast to findings from other studies<sup>10-12</sup>, we found no association between  
247 race or ethnicity and risk of self-reported infection, a finding that may reflect the  
248 underrepresentation of minorities and the relative affluence of our study population.  
249 Participants with graduate level college education also demonstrated lower risk as  
250 compared to participants without college education, again consistent with other studies.  
251 <sup>13</sup> Potential explanations for this observation include a lower likelihood of work-related  
252 contact with the general public and a greater awareness of risk and effective methods of  
253 protection among more educated subjects. The observation of lower risk among those  
254 whose last encounter with a primary care provider 1-5 years ago as compared to those  
255 who saw their provider within the past year could be due to greater awareness of  
256 COVID-19 from the health care provider, or to health care visits related to COVID-19  
257 illnesses.

258

259 Counterintuitively, smokers appeared to have a lower risk of infection as compared to  
260 non-smokers. This paradoxical relationship between smoking and risk of illness from  
261 COVID-19 has been observed in several other studies<sup>14,15</sup>, including some that  
262 speculate that nicotine could mitigate the effect of the cytokine storm in COVID-19  
263 patients.<sup>16</sup> However, these studies also point out that smokers are underrepresented in  
264 many COVID-19 population-based studies, as they appear to be in our study as well  
265 (5.4% of participants vs. the CDC estimated average of 12.4% for the southern United

266 States in 2021)<sup>17</sup>, and the harmful effects of smoking are often cited as offsetting any  
267 possible benefit from smoking for COVID-19 related morbidity.

268

269 As expected, participants with higher occupational risk of exposure to the general public  
270 demonstrated higher risk of infection, including those living in a household with a person  
271 who attends classes in person or who encounter the general public in their workplace.  
272 When risk was compared by occupational group using the NIOSH SOEM risk  
273 categories, the most significant increase in risk was from working in a healthcare  
274 setting. These results may be used to inform identification of high-risk groups for future  
275 respiratory disease outbreaks, allowing targeted programs to promote protective  
276 measures and behaviors that reduce the risk of infection.

277

278 Our study is subject to several limitations. COVID-19 cases were ascertained based on  
279 self-reported data from daily surveys, so the infections could not be verified.

280 Participants varied in the frequency with which they responded to the daily surveys,  
281 which limits our ability to determine the exact date of onset for the reported case of  
282 COVID-19. Generalizability is also limited due to the overrepresentation of lower risk  
283 socioeconomic groups in the study population and to recruitment that was limited to  
284 patients from healthcare networks and healthcare workers. As mentioned in the results  
285 section, this analysis was also limited to the 39% of participants who completed the  
286 supplemental questionnaires.

287



288 In contrast, the strengths of this study include a large sample size from a wide  
289 geographic area enrolled early in the pandemic and surveyed for a number of  
290 demographic, social, and economic characteristics. While our results illustrate the  
291 challenges inherent in understanding risk based on the complex interaction of age, race,  
292 and occupation and disentangling this from the risk due to local community infection  
293 rates, they also provide potential target groups for ongoing intervention. Under the  
294 assumption that those most at risk for first SARS-CoV-2 infection remain at higher risk  
295 for repeated infections even in the changing risk environment from vaccine uptake and  
296 shifting perceptions of the necessity of mitigations, this study may provide insight into  
297 ways to reduce ongoing disparities from the pandemic through risk stratification and  
298 targeted interventions.

299

300 **Acknowledgements.** The authors would like to thank the many dedicated CCRP  
301 participants who provided data for this study. All authors contributed to one or more  
302 areas of the methods (e.g. objectives, design, site selection and recruitment, form  
303 development, statistical approach) described in this article.

304

305 **References**

306

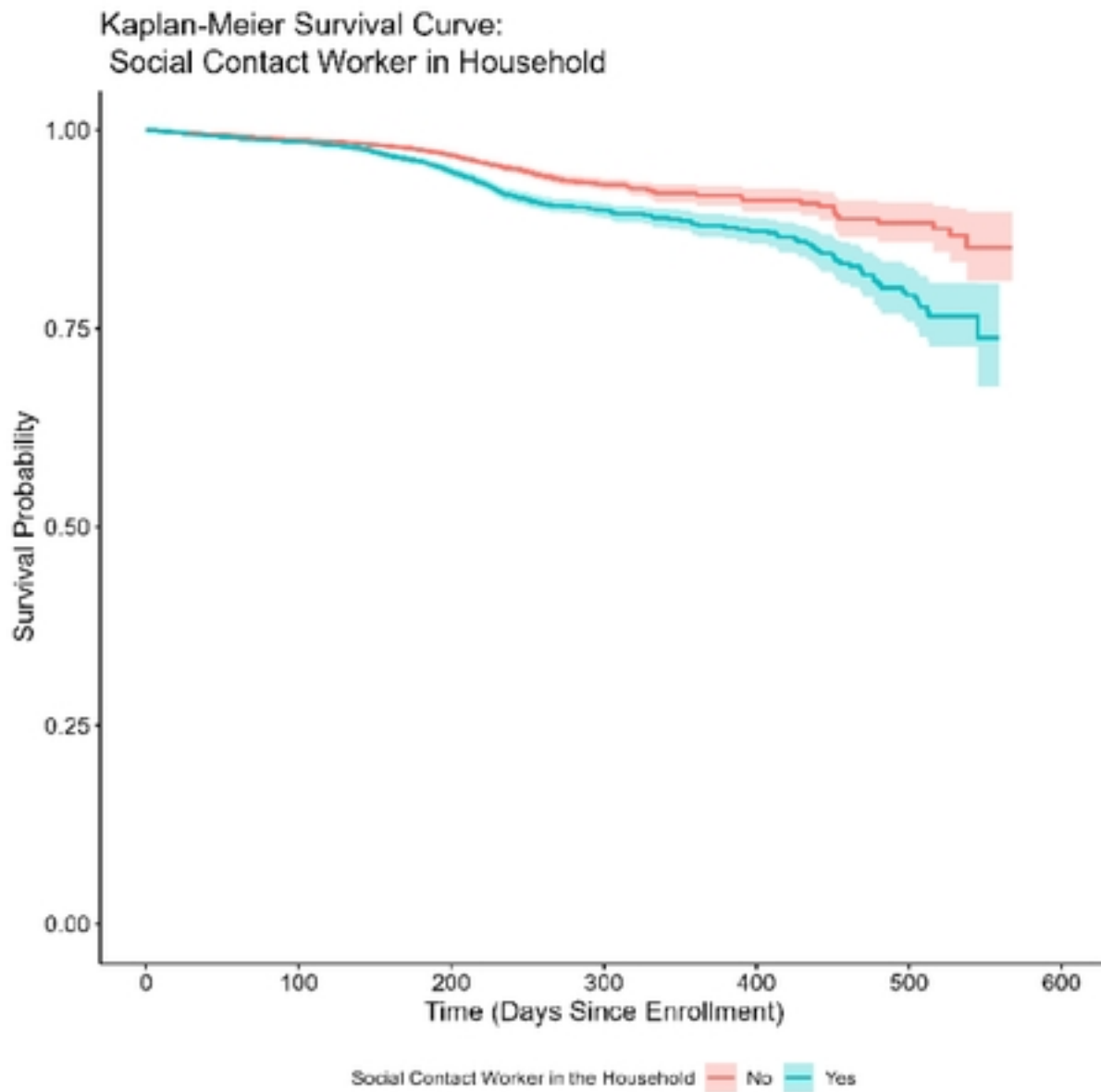
- 307 1. Ahmed SM, Shah RU, Fernandez V, et al. Robust Testing in Outpatient Settings to Explore COVID-  
 308 19 Epidemiology: Disparities in Race/Ethnicity and Age, Salt Lake County, Utah, 2020. *Public health*  
 309 *reports (Washington, DC : 1974)*. May 2021;136(3):345-353. doi:10.1177/0033354920988612
- 310 2. Guha A, Bonsu J, Dey A, Addison D. Community and Socioeconomic Factors Associated with  
 311 COVID-19 in the United States: Zip code level cross sectional analysis. *medRxiv*. Apr 22  
 312 2020;doi:10.1101/2020.04.19.20071944
- 313 3. Stokes EK, Zambrano LD, Anderson KN, et al. Coronavirus Disease 2019 Case Surveillance -  
 314 United States, January 22-May 30, 2020. *MMWR Morbidity and mortality weekly report*. Jun 19  
 315 2020;69(24):759-765. doi:10.15585/mmwr.mm6924e2
- 316 4. Athavale P, Kumar V, Clark J, Mondal S, Sur S. Differential Impact of COVID-19 Risk Factors on  
 317 Ethnicities in the United States. *Front Public Health*. 2021;9:743003. doi:10.3389/fpubh.2021.743003
- 318 5. Partnership C-CR. The COVID-19 Community Research Partnership: a multistate surveillance  
 319 platform for characterizing the epidemiology of the SARS-CoV-2 pandemic. *Biol Methods Protoc*.  
 320 2022;7(1):bpac033. doi:10.1093/biomethods/bpac033
- 321 6. van der Feltz S, Peters S, Pronk A, et al. Validation of a COVID-19 Job Exposure Matrix (COVID-  
 322 19-JEM) for Occupational Risk of a SARS-CoV-2 Infection at Work: Using Data of Dutch Workers. *Ann*  
 323 *Work Expo Health*. Jan 12 2023;67(1):9-20. doi:10.1093/annweh/wxac032
- 324 7. Covid-19-data. 2022. <https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html>
- 325 8. Fisman DN, Greer AL, Brankston G, et al. COVID-19 Case Age Distribution: Correction for  
 326 Differential Testing by Age. *Ann Intern Med*. Oct 2021;174(10):1430-1438. doi:10.7326/M20-7003
- 327 9. COVID-19 Stats: COVID-19 Incidence,\* by Age Group(dagger) - United States, March 1-  
 328 November 14, 2020( section sign). *MMWR Morbidity and mortality weekly report*. Jan 1  
 329 2021;69(5152):1664. doi:10.15585/mmwr.mm695152a8
- 330 10. Paremoer L, Nandi S, Serag H, Baum F. Covid-19 pandemic and the social determinants of  
 331 health. *BMJ*. Jan 28 2021;372:n129. doi:10.1136/bmj.n129
- 332 11. Yancy CW. COVID-19 and African Americans. *JAMA*. May 19 2020;323(19):1891-1892.  
 333 doi:10.1001/jama.2020.6548
- 334 12. Webb Hooper M, Napoles AM, Perez-Stable EJ. COVID-19 and Racial/Ethnic Disparities. *JAMA*.  
 335 Jun 23 2020;323(24):2466-2467. doi:10.1001/jama.2020.8598
- 336 13. Yoshikawa M, Asaba K. Educational Attainment Decreases the Risk of COVID-19 Severity in the  
 337 European Population: A Two-Sample Mendelian Randomization Study. *Front Public Health*.  
 338 2021;9:673451. doi:10.3389/fpubh.2021.673451
- 339 14. Usman MS, Siddiqi TJ, Khan MS, et al. Is there a smoker's paradox in COVID-19? *BMJ Evid Based*  
 340 *Med*. Dec 2021;26(6):279-284. doi:10.1136/bmjebm-2020-111492
- 341 15. Paleiron N, Mayet A, Marbac V, et al. Impact of Tobacco Smoking on the Risk of COVID-19: A  
 342 Large Scale Retrospective Cohort Study. *Nicotine Tob Res*. Aug 4 2021;23(8):1398-1404.  
 343 doi:10.1093/ntr/ntab004
- 344 16. Korzeniowska A, Reka G, Bilska M, Pieciewicz-Szczesna H. The smoker's paradox during the  
 345 COVID-19 pandemic? The influence of smoking and vaping on the incidence and course of SARS-CoV-2  
 346 virus infection as well as possibility of using nicotine in the treatment of COVID-19 - Review of the  
 347 literature. *Przegląd epidemiologiczny*. 2021;75(1):27-44. doi:10.32394/pe.75.03
- 348 17. Current Cigarette Smoking Among Adults in the United States. Centers for Disease Control and  
 349 Prevention; Office on Smoking and Health, National Center for Chronic Disease Prevention and Health

350 Promotion. Updated May 4, 2023. Accessed June 27, 2023,  
351 [https://www.cdc.gov/tobacco/data\\_statistics/fact\\_sheets/adult\\_data/cig\\_smoking/index.htm](https://www.cdc.gov/tobacco/data_statistics/fact_sheets/adult_data/cig_smoking/index.htm)

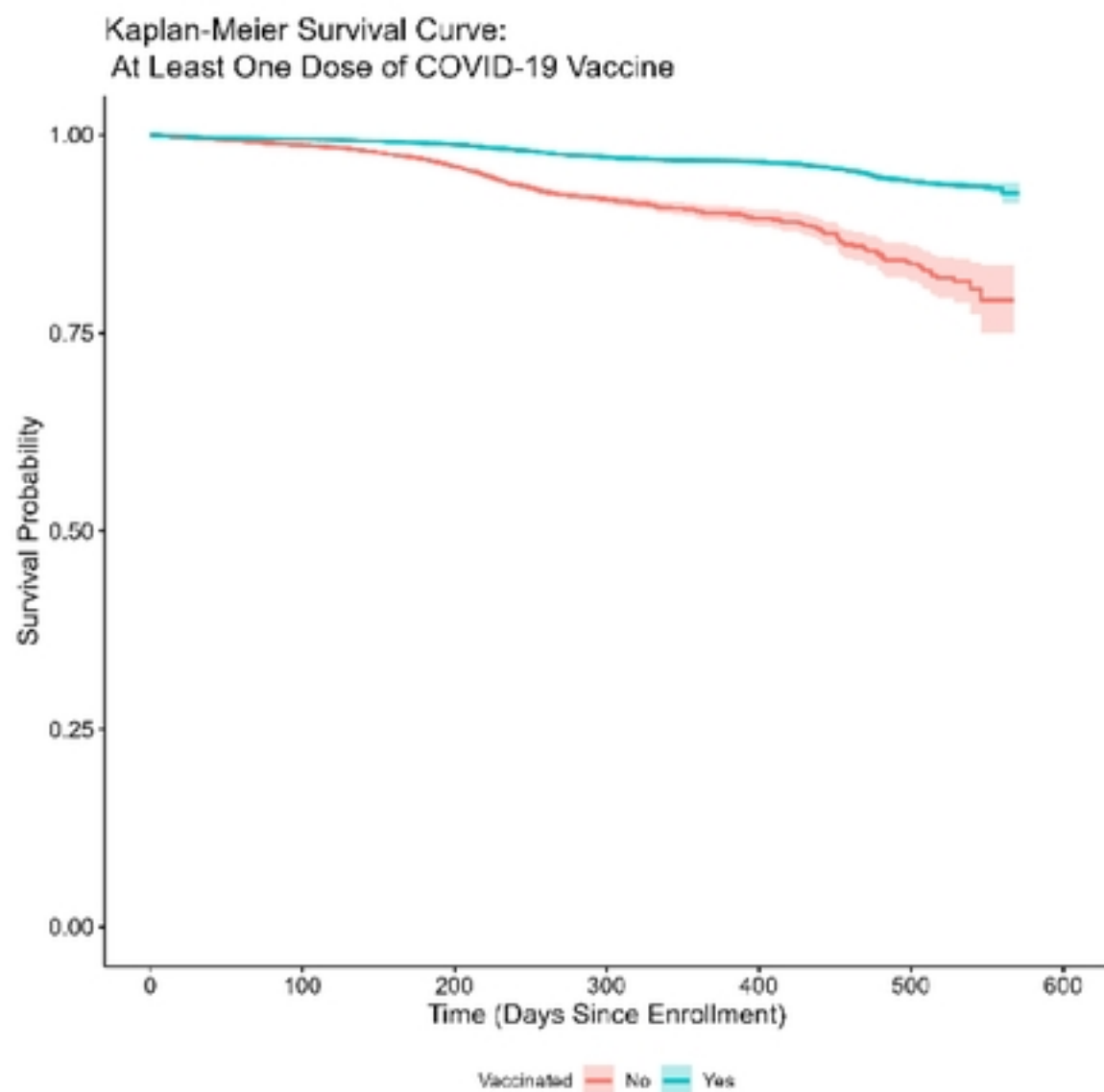
352

**Figure 1:** Kaplan Meyer survival analysis comparing cumulative risk of self-reported COVID-19 by specific risk factors.

**a.** Risk of self-reported COVID-19 and presence of a household member who works in a Social Contact Occupation



**b.** Risk of self-reported COVID-19 and receipt of at least one dose of COVID-19 vaccine



c. Risk of self-reported COVID-19 and occupational risk category based on the SARS-CoV-2 Occupational Exposure Matrix (SOEM) system

