

1 **Title:** Disparities in SARS-CoV-2 seroprevalence among individuals presenting for care in central North Carolina
2 over a six-month period
3
4

5 **Authors:** Cesar A. Lopez^{1*}, Clark H. Cunningham^{2*}, Sierra Pugh³, Katerina Brandt⁴, Usaphea P. Vanna¹, Matthew J.
6 Delacruz¹, Quique Guerra¹, Samuel Jacob Goldstein⁵, Yixuan J. Hou⁶, Margaret Gearhart⁷, Christine Wiethorn⁸,
7 Candace Pope⁸, Carolyn Amditis⁹, Kathryn Pruitt¹⁰, Cinthia Newberry-Dillon¹⁰, John Schmitz¹¹, Lakshmanane
8 Premkumar¹, Adaora A. Adimora^{6,12}, Michael Emch⁴, Ross Boyce¹², Allison E. Aiello⁶, Bailey K. Fosdick³, Daniel
9 B. Larremore¹³, Aravinda M. de Silva¹, Jonathan J Juliano^{6,12}, Alena J. Markmann^{12#}

10
11 *co-first authors

12 #corresponding author

- 13
- 14 1. Department of Microbiology and Immunology, University of North Carolina School of Medicine, Chapel
15 Hill NC 27599, USA
- 16 2. Department of Genetics, University of North Carolina School of Medicine, University of North Carolina at
17 Chapel Hill, Chapel Hill NC 27599, USA
- 18 3. Department of Statistics, Colorado State University, Fort Collins, CO, 80523, USA
- 19 4. Department of Geography, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA;
20 Carolina Population Center, Chapel Hill, NC 27516, USA
- 21 5. Department of Environmental Sciences and Engineering, The University of North Carolina at Chapel Hill,
22 Chapel Hill, NC 27599, USA
- 23 6. Department of Epidemiology, University of North Carolina at Chapel Hill School of Public Health,
24 University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
- 25 7. McLendon Clinical Laboratories, UNC Healthcare, Chapel Hill, NC 27599, USA
- 26 8. Johnston Health Laboratories, Johnston Health, Smithfield, NC 27577
- 27 9. Rex Healthcare Laboratory, UNC Healthcare, Chapel Hill, NC 27607, USA
- 28 10. Chatham Clinical Laboratory, Chatham Hospital, Siler City, NC 27344, USA
- 29 11. Department of Pathology & Laboratory Medicine, University of North Carolina School of Medicine, Chapel
30 Hill NC 27599, USA
- 31 12. Department of Medicine, Division of Infectious Diseases, University of North Carolina School of Medicine,
32 Chapel Hill NC 27599, USA
- 33 13. Department of Computer Science & BioFrontiers Institute, University of Colorado Boulder, Boulder, CO,
34 80303, USA
- 35

36 Alena J Markmann, MD, PhD; 160 Dental Circle, Chapel Hill, NC 27514; 585-880-5812

37
38 **Email:** alena.markmann@unchealth.unc.edu

39 40 **ABSTRACT**

41 42 **Background**

43 Robust community-level SARS-CoV-2 prevalence estimates have been difficult to obtain in the American South and
44 outside of major metropolitan areas. Furthermore, though some previous studies have investigated the association of
45 demographic factors such as race with SARS-CoV-2 exposure risk, fewer have correlated exposure risk to surrogates
46 for socioeconomic status such as health insurance coverage.

47 48 **Methods**

49 We used a highly specific serological assay utilizing the receptor binding domain of the SARS-CoV-2 spike-protein
50 to identify SARS-CoV-2 antibodies in remnant blood samples collected by the University of North Carolina Health
51 system. We estimated the prevalence of SARS-CoV-2 in this cohort with Bayesian regression, as well as the
52 association of critical demographic factors with higher prevalence odds.

53 54 **Findings**

55 Between April 21st and October 3rd of 2020, a total of 9,624 unique samples were collected from clinical sites in central
56 NC and we observed a seroprevalence increase from 2·9 (1·7, 4·3) to 9·1 (7·2, 11·1) over the study period. Individuals

57 who identified as Latinx were associated with the highest odds ratio of SARS-CoV-2 exposure at 7.77 overall (5.20,
58 12.10). Increased odds were also observed among Black individuals and individuals without public or private health
59 insurance.

60
61 **Interpretation**

62 Our data suggests that for this care-accessing cohort, SARS-CoV-2 seroprevalence was significantly higher than
63 cumulative total cases reported for the study geographical area six months into the COVID-19 pandemic in North
64 Carolina. The increased odds of seropositivity by ethnoracial grouping as well as health insurance highlights the urgent
65 and ongoing need to address underlying health and social disparities in these populations.

66
67 **RESEARCH IN CONTEXT**

68
69 **Evidence before this study**

70 We searched PubMed for studies published through March 21st, 2021. We used search terms that included “COVID-
71 19”, “SARS-CoV-2”, “prevalence” and “seroprevalence”. Our search resulted in 399 papers, from which we identified
72 58 relevant studies describing SARS-CoV-2 seroprevalence at sites around the United States from March 1 to
73 December 9, 2020, 12 of which utilized remnant clinical samples and three of which overlapped with our study area.
74 Most notably, one study of 4,422 asymptomatic inpatients and outpatients in central NC from April 28-June 19, 2020
75 found an estimated seroprevalence of 0.7 - 0.8%, and another study of 177,919 inpatients and outpatients (3,817 from
76 NC) from July 27-September 24, 2020 found an estimated seroprevalence of 2.5 - 6.8%.

77
78 **Added value of this study**

79 This is the largest SARS-CoV-2 seroprevalence cohort published to date in NC. Importantly, we used a Bayesian
80 framework to account for uncertainty in antibody assay sensitivity and specificity and investigated seropositivity by
81 important demographic variables that have not yet been studied in this context in NC. This study corroborates other
82 reports that specific demographic factors including race, ethnicity and the lack of public or private insurance are
83 associated with elevated risk of SARS-CoV-2 infection. Furthermore, in a subset of serum samples, we identify other
84 SARS-CoV-2 antibodies elicited by these individuals, including functionally neutralizing antibodies.

85
86 **Implications of all the available evidence**

87 It is difficult to say the exact seroprevalence in the central North Carolina area, but a greater proportion of the
88 population accessing healthcare has been infected by SARS-CoV-2 than is reflected by infection cases confirmed by
89 molecular testing. Furthermore, local governments need to prioritize addressing the many forms of systemic racism
90 and socioeconomic disadvantage that drive SARS-CoV-2 exposure risk, such as residential and occupational risk, and
91 an urgent need to provide access to SARS-CoV-2 testing and vaccination to these groups.

92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112

113 INTRODUCTION

114 In December 2019, a cluster of pneumonia cases in China's Hubei province heralded the beginning of what would
115 become a global pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Despite
116 attempts to contain the virus, SARS-CoV-2 has spread around the world, causing over 100 million infections and over
117 2 million deaths due to the respiratory disease it causes, COVID-19.¹ Serological testing complements molecular
118 testing for evaluating the spread of SARS-CoV-2 and can be deployed efficiently at the population level.² Recently,
119 large prevalence studies around the United States using remnant samples from healthcare settings have reported
120 substantial geographic variation in prevalence by state: around 30% in New York but less than 2% in North Carolina
121 (NC), the focus of the present study.^{3,4} Notably, two other studies overlap with the present cohort both temporally and
122 geographically. One study of 4,422 asymptomatic inpatients and outpatients in central NC from April 28-June 19,
123 2020 found an estimated seroprevalence of 0.7 - 0.8%, and another study of 177,919 remnant clinical laboratory
124 samples from routine screening (3,817 from NC) from July 27-September 24, 2020 found an estimated seroprevalence
125 of 2.5 - 6.8%.^{5,6} While overall seroprevalence estimates of a given study depend on sampling method, assay
126 characteristics, geography, and temporal factors, seroprevalence studies can provide information on the spread of
127 COVID-19 that is missed by looking at the number of confirmed acute cases alone.

128
129 Seroprevalence studies are also useful for identifying demographic factors such as racial, ethnic and socioeconomic
130 disparities among those exposed to SARS-CoV-2.^{4,7,8} The COVID-19 pandemic has been shaped by the deep and
131 historic impacts of structural racism on disease disparities in US society as identified by serologic studies as well as
132 hospitalization and mortality rates.^{9,10} For example, COVID-19 case and hospitalization rates among Black, Hispanic
133 and Native American populations in the US, according to the Centers for Disease Control and Prevention, are 2.5-4.5
134 times higher than those in white populations¹¹. Structural and occupational factors previously identified as drivers of
135 race and ethnic disparities in health include unequal labor market opportunities and higher representation in essential
136 work positions that lack job security, access to infection prevention control, benefits, and sick leave.¹²⁻¹⁷ Here, we
137 confirm the findings of disparate SARS-CoV-2 exposure among racial and ethnic groups in the US by measuring
138 seroprevalence in a large southern US health-care seeking cohort using remnant blood samples.

139
140 The following results are from the first six months (April 21-October 3, 2020) of an ongoing seroprevalence study
141 using convenience remnant samples from clinical laboratories in central NC. The study catchment area covers Wake,
142 Orange, Chatham, Johnston, Durham and Alamance counties and includes the county of the first confirmed case in
143 NC,¹⁸ which occurred on March 3rd, 2020. On October 3rd, 2020, the cumulative total PCR and antigen-confirmed
144 SARS-CoV-2 cases in the study catchment area was 52,722 (2.7% of the population), with 1,266 confirmed deaths.^{19,20}
145 We used an in-house enzyme-linked immunoassay (ELISA) against the receptor-binding domain (RBD) of the spike
146 protein of SARS-CoV-2²¹ and applied Bayesian inference²² to estimate seroprevalence and demographic risk factors
147 of SARS-CoV-2 infection in a healthcare-seeking cohort over a six-month period.

148 149 150 METHODS

151 *Sampling Strategy and Data Collection*

152
153 Remnant plasma and serum samples were collected from four hospital-based clinical laboratories affiliated with the
154 University of North Carolina (UNC) Health system. These laboratories receive and process clinical samples from
155 inpatient units as well as outpatient clinics in NC. Each week, up to 300 remnant samples belonging to individuals 5-
156 99 years of age were arbitrarily selected by the clinical laboratory for testing from each location. Samples were
157 collected between April 21st, 2020 – October 3rd, 2020. Medical record numbers were recorded for each sample and
158 duplicates were discarded. We abstracted the following demographic and clinical data from electronic medical records
159 (EMR, Epic): age, sex, ethnicity, race, address including city, state and ZIP code, insurance coverage, insurance type,
160 inpatient or outpatient status, encounter diagnosis (ICD-10 code), inpatient date of discharge, and whether or not
161 COVID-19 testing was performed within a 30-day window prior to study sample collection. Written informed consent
162 was not required due to the use of routinely collected samples. All data for this study were collected under UNC IRB
163 #20-0791, which is conducted under Good Clinical Research Practices (GCP) and compliant with institutional IRB
164 oversight. De-identified samples used for assay validation were collected under UNC IRBs #20-0913 and #08-0895.

165 166 167 *Enzyme-Linked Immunosorbent Assays*

168

169 A total Ig and IgM SARS-CoV-2 RBD ELISA that does not react with common endemic human coronaviruses was
170 used in this study as previously described.²¹ The spike protein N-terminal domain (NTD) antigen (16–305 amino
171 acids, Accession: P0DTC2.1) was cloned into the pαH mammalian expression vector and purified using nickel-
172 nitrilotriacetic acid agarose in the same manner. Each measurement was conducted in duplicate and duplicate values
173 with variance > 25% and/or one value above assay cutoff were repeated. A correlation plot shows using 140 COVID-
174 19 PCR-confirmed cases between our RBD Ig P/N ratios and the neutralization assay described below (**Figure S1**).

175 176 *Nucleocapsid protein ELISA*

177
178 Detection of IgG antibody to SARS-CoV-2 N antigen was performed with the EUA approved Abbott SARS-CoV-2
179 IgG assay (Abbott Laboratories) on the Abbott Architect i2000SR immunoassay analyzer as previously described.²³

180 181 *SARS-CoV-2 Neutralization Assays*

182
183 To further characterize the SARS-CoV-2 antibody responses of this study, viral neutralization assays were obtained
184 for 110 ELISA-positive samples that were selected randomly using the `sample_n()` function of the `dplyr` R package.
185 Luciferase-expressing, full-length SARS-CoV-2 isolate WA1 strain (GenBank Accession#: MT020880) was
186 engineered and recovered via reverse genetics and used to titer serially diluted sera on Vero E6 USAMRID cell as
187 described previously.²⁴ The sample dilution at which a 50% reduction in RLU was observed relative to that of the
188 virus control wells was used as the 50% neutralization titer (NT₅₀) for that sample.

189 190 *Statistical Methods and Analyses*

191
192 To account for plate-to-plate variability, we used positive to negative (P/N) ratios defined as the average optical
193 density (OD) of the sample divided by the average OD of the negative control in the respective ELISA plate. Following
194 the CDC recommendation to set specificity to 99.5%, we chose the 0.995 quantile of the P/N ratio for the negative
195 validation samples as the P/N cutoff.²⁵

196
197 We fit two statistical models to estimate seroprevalence. First, we fit a Bayesian autoregressive logistic model to
198 estimate weekly prevalence across the six-month study period while accounting for uncertainty in the assay specificity
199 and sensitivity due to finite lab validation samples. Second, we fit a Bayesian logistic regression model to estimate
200 prevalence and conditional odds ratios by subpopulation with main effects for sex, race/ethnicity, age, in/out-patient
201 status, and health insurance payor, while again accounting for uncertainty in the assay test characteristics (**Table S1**).
202 Each group was compared to females, non-Latinx white, ages 5-17, outpatient, and private payor health insurance
203 status as respective baseline categories. Details are given in Supplementary Methods: Bayesian seroprevalence models
204 with unknown sensitivity and specificity. These Bayesian hierarchical models (BHM) simultaneously model study
205 data and validation data to produce prevalence estimates and credible intervals that reflect both uncertainty due to the
206 finite study sample as well as the uncertainty in the sensitivity and specificity of the ELISA, with statistical uncertainty
207 represented by 95% credible intervals.

208 209 210 **RESULTS**

211 212 *Cohort Characteristics*

213
214 From April 21, 2020 – October 3, 2020, after excluding duplicate samples, 9,624 remnant samples were analyzed
215 from four UNC Health hospitals in central North Carolina. The six counties most heavily sampled were Orange,
216 Johnson, Chatham, Wake, Durham and Alamance, with 6,946 (72.2%) of individuals residing in these counties
217 (**Figure 1**). The study consists of 5,417 females (56.3%) and 4,206 males (43.7%) which is similar to the
218 demographics of this region (**Table 1**). Less than 6% of individuals were in the youngest age group (5-17 years old),
219 though this age group represents over 18% of the study area's population. Approximately 90% of study individuals
220 were insured, with 8% falling into the self-pay category. The majority of sampled individuals were seen at UNC
221 Memorial Hospital, ~3% were acute or trauma cases and ~5% had a visit diagnosis of fever or respiratory symptoms
222 (**Table S2**). Overall, approximately 1% of patients had an associated COVID-19 visit diagnosis, with a significant
223 difference between inpatients (2.8%) and outpatients (0.3%) (Chi-squared test; $p < 0.0001$) (**Table S3**).

224

225 *Overall seroprevalence estimates*

226
227 The six-month period of the study was divided into three, two-month cohorts. The BHM-derived seroprevalence
228 estimates increased from around 3% in April/May to around 9% in August/September (**Table 2**). Raw seroprevalence
229 estimates also showed a similar increasing trend over the study period, but because they do not take into account assay
230 performance uncertainty, they are slightly higher at ~5% and ~11%. Furthermore, seroprevalence estimates peaked in
231 early August following a hospitalization peak in mid-July (**Figure 2A, 2C**). Cumulative PCR-positive COVID-19
232 cases reported by the state for these six counties increased over the study period (**Figure 2B**) with the most rapid
233 accumulation of cases occurring from June to August. Unexpectedly, seroprevalence peaks followed by a slight
234 decline, related to raw seroprevalence estimates at Johnston County hospital which surged from 7·81% in the first two
235 months to 18·00% in the second two months coinciding with a peak in PCR-confirmed cases in the region, followed
236 by a measured decline in raw seroprevalence to 14·80% in the final two-month period (**Table S6**). This peak and
237 decline was not affected by the removal of cases with ICD-10 visit codes for “COVID-19” or those we identify as
238 “respiratory disease” (data not shown).

239
240 *Clinical and demographic differences in seroprevalence estimates*

241
242 Latinx-identifying individuals have higher SARS-CoV-2 seroprevalence at 15-33% compared to non-Latinx
243 individuals which have only 1-11% seroprevalence over the study period (**Table 2**). Individuals with
244 “Other/Unknown” or “Self-pay” insurance status had a higher estimated seroprevalence (~20-40% or ~1-18%,
245 respectively) than those with private or public health insurance (~3-9%). Approximately 30% of Latinx individuals in
246 this study were either in the other/unknown or self-pay health categories, disproportionately comprising ~27% of these
247 two categories but only accounting for ~8% of our study population (**Table S5**).

248
249 To better compare the relative odds of SARS-CoV-2 seroprevalence for each clinical and/or demographic
250 characteristic, we calculated conditional odds ratios for each variable we collected using the BHM (**Table 3**). Latinx
251 individuals had the highest odds of SARS-CoV-2 exposure throughout the study period compared to non-Latinx white
252 individuals, OR 7·77 overall (5·20, 12·10), ranging from 14·53 (6·47, 36·72) in the first two months to 4·34 (2·61,
253 7·41) in the last two months of the study. Individuals with unknown insurance status also had an elevated odds ratio
254 of seropositivity at 3·81 (2·23, 6·54) compared to those with private insurance status. Over the entire period of the
255 study, non-Latinx Black individuals, individuals aged 50-64 years, and inpatients, also had increased odds ratios of
256 approximately two-fold compared to non-Latinx white individuals, individuals aged 0-17, and outpatients,
257 respectively. The overall difference in odds ratios by age appears to be driven primarily by increased odds ratios in
258 the first two months.

259
260 *SARS-CoV-2 RBD positive subset analysis*

261
262 To determine the SARS-CoV-2 antibody repertoire in a subset of RBD Ig seropositive individuals, we randomly
263 selected 110 participants and tested their sera for: RBD IgM, NTD IgG, and SARS-CoV-2 neutralizing antibodies.
264 About 75% of individuals were positive for RBD IgM, 60% had NTD IgG antibodies, and about 50% had detectable
265 neutralizing antibodies (**Figure 3A**). Of the participants with detectable functionally neutralizing antibodies, 23% had
266 a high titer > 1:1280, 47% had a moderate titer of 1:160-1:1279, and 30% had a lower titer of 1:10-1:159. Furthermore,
267 RBD Ig P/N antibody signal correlated more strongly with functionally neutralizing antibody levels (**Figure 3B**), than
268 NTD IgG signal (**Figure 3C**). We also found that 36% (29/80) of those in this subset with an ICD-10 code binned as
269 “Other” had detectable neutralizing antibodies, while 83% (25/30) of individuals with an ICD-10 code of “COVID-
270 19” or what we identify as “respiratory disease” had neutralizing antibodies (**Figure 3D**). There was substantial
271 agreement between the RBD Ig ELISA results reported here and 150 study individuals for which a clinical SARS-
272 CoV-2 nucleocapsid IgG (Abbott assay) was available (Cohen’s kappa=0·685) (**Table S4**).

273
274 **DISCUSSION**

275
276 Here we describe SARS-CoV-2 seroprevalence in a total of 9,624 unique healthcare-seeking individuals in central
277 North Carolina using clinical remnant samples from four regional hospitals between April and October 2020.
278 Employing a Bayesian framework²² to capture assay uncertainty in both field and lab validation data, we estimate a
279 significant increase in overall seroprevalence from 2·9% (95% CI 1·7% - 4·3%) at the start of the study period, to
280 9·1% (95% CI 7·2% - 11·1%) at the end of the study period, approximately six months after the first case in the state.

281 The end-of-study prevalence identified here is significantly higher than the cumulative number of cases identified by
282 PCR or antigen testing in the same county region at the same date, though determining the degree to which the
283 identified cases undercount true infections requires more representative sampling.
284

285 A previous study from central North Carolina that overlaps with the first two months of our study period found
286 seroprevalence in an asymptomatic healthcare-seeking cohort below 1% using the Abbott nucleocapsid IgG assay.⁵
287 This is much lower than the ~3% seropositive estimate in our cohort over this time period, and may be due to under-
288 sampling of Latinx individuals in that study and/or preferential sampling of asymptomatic individual. There is also
289 growing concern about the use and performance of nucleocapsid IgG assays in individuals with asymptomatic or mild
290 disease.²⁶ The nationwide CDC study that used remnant clinical samples from inpatients and outpatients found a
291 seroprevalence of 6.8% in NC in September 2020, which is closer to our estimate of 9.1% during the final two months
292 of this analysis.
293

294 The conditional odds ratios we calculated assume that all other variables are held constant while estimating the effect
295 of one demographic variable at a time. We found that Latinx individuals had the highest odds of SARS-CoV-2
296 seropositivity, and that non-Latinx Black individuals also had high odds of SARS-CoV-2 seropositivity, corroborating
297 previous observations.^{4,7,8} The high odds ratios by race and ethnicity decrease over time, consistent with the virus
298 spreading first among individuals with high exposure risk and later to the rest of the population. Residential
299 segregation, crowded households, socioeconomic disadvantage, mass incarceration, and inequities in access to
300 insurance, health care, and access to testing, vaccination, and treatments have all been cited as factors that have
301 contributed to the large and sustained racial and ethnic disparities in COVID-19 in the US.^{13,15,27-29} We also observed
302 that individuals that fell into the “self-pay” category for their healthcare or otherwise had unknown healthcare status
303 had higher SARS-CoV-2 seropositivity and odds ratios. The significant overlap in the Latinx population and these
304 insurance categories is concerning because the high odds ratios and seroprevalence in these categories can lead to
305 much higher exposure risk among the significant number of underinsured Latinx individuals³⁰.
306

307 Studies of PCR-positive symptomatic COVID-19 cases have reported good neutralizing antibody responses in these
308 individuals.³¹ Thus, it was surprising that we observed 51% of individuals in our RBD-positive subset analysis did not
309 have detectable neutralizing antibodies. Though we do not know what proportion of individuals in our study had
310 asymptomatic infections, low neutralizing antibody titers may be explained by short duration of viral replication in
311 respiratory compartments and low to no viral replication in the serum or blood of those with mild or asymptomatic
312 disease. Not surprisingly, when we looked at our neutralizing antibody results by ICD-10 code, the majority of all
313 individuals with a “respiratory disease” or “COVID-19” diagnosis had developed neutralizing antibodies. Reports of
314 mild disease COVID-19 cohorts support the idea that detectable neutralizing antibody titers are not necessarily
315 identified after mild COVID-19.^{23,32} In this subset analysis we also found that 75% had RBD IgM antibodies,
316 indicating that their infections likely occurred within the past three months.²³ Furthermore, a majority of individuals
317 in this subset had detectable NTD IgG antibodies; the NTD has recently been found to be an important target for the
318 B.1.1.7, B.1.351, and B.1.1.28.1 SARS-CoV-2 variants.³³
319

320 The primary limitation of this study is that the study population, composed of individuals accessing care at UNC area
321 hospitals and clinics may differ from the overall population in central North Carolina in ways that are not captured in
322 demographic data (e.g., overall health status). Accordingly, we have chosen to not weight our dataset to county
323 demographics and therefore do not provide overall estimates of seroprevalence in the six-county area as that would
324 require more representative sampling methodology.³⁴ Furthermore, many clinics and hospital elective procedures were
325 closed or only seeing patients virtually during the first few months of the study period.
326

327 The unexpected seroprevalence peak observed at the Johnston County hospital suggests that the population accessing
328 care at these clinical sites did not have consistent exposure risk over time. As expected, seroprevalence estimates in
329 this cohort track closely with COVID-19 hospitalizations in the four hospitals in this study with a two-week lag which
330 could be due to time to seroconvert. Declining antibody over this time period to undetectable levels is unlikely, as the
331 length of the study is shorter than it takes for significant antibody decline to undetectable levels, although little is
332 known about antibody levels over time in the asymptomatic population.³¹
333

334 Other limitations of the study include that we could not break down odds ratios by all races and/or by race and ethnicity
335 at the same time, or by multiracial categories because the number of individuals became too small to allow broad
336 interpretation. Finally, though the “self-pay” insurance category includes the uninsured, we cannot confidently state

337 that everyone in this category was uninsured because lack of insurance is not a specific category that is captured in
338 the EMR. Although SARS-CoV-2 seroprevalence of healthcare-seeking individuals is an imperfect comparison to the
339 general population, we maintain that it is a useful sentinel population to understand overall trends, especially when
340 attempting to surveil rural populations residing in areas without strong public health systems and spread over a large
341 geographic area.

342
343 Based on our estimates of seroprevalence in the population accessing healthcare, cumulative case numbers confirmed
344 by molecular diagnostics are likely under-representing the true number of cases. Public health distancing measures,
345 mask wearing, and vaccination should continue to be prioritized in order to lower the transmission of SARS-CoV-2
346 and subsequent loss of lives. Our findings of a significantly higher odds of SARS CoV-2 seropositivity among Latinx
347 and non-Latinx Black populations corroborate numerous studies describing large racial and ethnic disparities in
348 SARS-CoV-2 infection, morbidity and mortality in the US.^{4,7,8} Vaccination programs should address structural and
349 occupational factors that drive race and ethnic disparities in health outcomes in the US to ensure that individuals at
350 particularly high exposure risk of SARS-CoV-2 have timely access to SARS-CoV-2 vaccination.

351 **Acknowledgements**

352 We would like to thank Jennifer Dan, Alessandro Sette and Shane Crotty's Laboratories at La Jolla Institute of
353 Immunology for supplying some of the serum controls in our validation cohort.

354 **Data Sharing**

355 Deidentified individual data will be shared beginning 9 to 36 months following publication provided the investigator
356 who proposes to use the data has approval from an Institutional Review Board (IRB), Independent Ethics Committee
357 (IEC), or Research Ethics Board (REB), as applicable, and executes a data use/sharing agreement with UNC.

358 **Funding Sources**

359 This study was supported by funds from the NC Department of Health and Human Services, Division of Public Health
360 (Contract 00041877) and the SeroNet program of the National Cancer Institute (1U01CA261277-01). Funding sources
361 played no role in study design, in the collection, analysis, and interpretation of data, in the writing of the report, or in
362 the decision to submit the paper for publication.

363 **Declaration of interests**

364 The authors declare no conflicts of interest.

365 **Author contributions**

366 CL, CHC: Conceptualization, Data curation, Investigation, Project Administration, Writing – original draft. SP, KB:
367 Formal analysis, Software, Visualization. MD, QG, UPV: Data curation, Investigation, Project Administration. SJG:
368 Formal analysis, Software, Visualization. YJH: Investigation. PL: Methodology, Resources. RR, MG, CW, KP, CA:
369 Resources. JS: Conceptualization, Resources. ME: Conceptualization, Formal analysis, Resources, Supervision. RB,
370 AA: Conceptualization, Funding acquisition. BKF: Conceptualization, Formal analysis, Resources, Software,
371 Supervision. DBL: Conceptualization, Formal analysis, Software. ADS: Conceptualization, Funding acquisition,
372 Methodology, Project Administration, Resources, Supervision. JJJ: Conceptualization, Funding acquisition, Project
373 Administration, Supervision. AJM: Conceptualization, Funding acquisition, Investigation, Project Administration,
374 Supervision, Writing – original draft. CL, CHC, SP, KB, MD, QG, UPV, SJG, YJH, PL, JS, ME, RB, AA, BKF, DBL,
375 ADS, JJJ, AJM: Writing- reviewing & editing.

376 **REFERENCES**

- 377 1 Home - Johns Hopkins Coronavirus Resource Center. 2020. <https://coronavirus.jhu.edu/> (accessed Feb 26,
378 2021).
- 379 2 Population-based age-stratified seroepidemiological investigation protocol for coronavirus 2019 (COVID-
380 19) infection. <https://www.who.int/emergencies/diseases/novel-> (accessed June 21, 2020).
- 381 3 Havers FP, Reed C, Lim T, *et al.* Seroprevalence of Antibodies to SARS-CoV-2 in 10 Sites in the United
382 States, March 23-May 12, 2020. *JAMA Intern Med* 2020. DOI:10.1001/jamainternmed.2020.4130.
- 383 4 Anand S, Montez-Rath M, Han J, *et al.* Prevalence of SARS-CoV-2 antibodies in a large nationwide sample
384 of patients on dialysis in the USA: a cross-sectional study. *Lancet* 2020; **0**. DOI:10.1016/S0140-

- 393 6736(20)32009-2.
- 394 5 Barzin A, Schmitz JL, Rosin S, *et al.* SARS-CoV-2 seroprevalence among a southern U.S. population
395 indicates limited asymptomatic spread under physical distancing measures. *MBio* 2020; **11**: 1–8.
- 396 6 Bajema KL, Wiegand RE, Cuffe K, *et al.* Estimated SARS-CoV-2 Seroprevalence in the US as of
397 September 2020. *JAMA Intern Med* 2020. DOI:10.1001/jamainternmed.2020.7976.
- 398 7 Rogawski McQuade ET, Guertin KA, Becker L, *et al.* Assessment of Seroprevalence of SARS-CoV-2 and
399 Risk Factors Associated With COVID-19 Infection Among Outpatients in Virginia. *JAMA Netw open* 2021;
400 **4**: e2035234.
- 401 8 Feehan AK, Velasco C, Fort D, *et al.* Racial and workplace disparities in seroprevalence of SARS-CoV-2,
402 Baton Rouge, Louisiana, USA. *Emerg Infect Dis* 2021; **27**: 314–7.
- 403 9 Bailey ZD, Feldman JM, Bassett MT. How Structural Racism Works — Racist Policies as a Root Cause of
404 U.S. Racial Health Inequities. *N Engl J Med* 2020; **384**: 768–73.
- 405 10 Khazanchi R, Evans CT, Marcelin JR. Racism, Not Race, Drives Inequity Across the COVID-19
406 Continuum. *JAMA Netw. open.* 2020; **3**: e2019933.
- 407 11 Centers for Disease Control. Risk for COVID-19 Infection, Hospitalization, and Death By Race/Ethnicity.
408 2021; published online Feb 18. [https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-](https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-race-ethnicity.html)
409 [discovery/hospitalization-death-by-race-ethnicity.html](https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-race-ethnicity.html) (accessed Feb 23, 2021).
- 410 12 Kerwin D, Nicholson M, Alulema D, Warren R. US Foreign-Born Essential Workers by Status and State,
411 and the Global Pandemic. 2020 DOI:10.14240/cmsrpt0520.
- 412 13 Bailey ZD, Robin Moon J. Racism and the political economy of COVID-19: Will we continue to resurrect
413 the past? *J Health Polit Policy Law* 2020; **45**: 937–50.
- 414 14 Alberti PM, Lantz PM, Wilkins CH. Equitable Pandemic Preparedness and Rapid Response: Lessons from
415 COVID-19 for Pandemic Health Equity. *J Health Polit Policy Law* 2020; **45**: 921–35.
- 416 15 Laster Pirtle WN. Racial Capitalism: A Fundamental Cause of Novel Coronavirus (COVID-19) Pandemic
417 Inequities in the United States. *Heal Educ Behav* 2020; **47**: 504–8.
- 418 16 McClure ES, Vasudevan P, Bailey Z, Patel S, Robinson WR. Racial Capitalism Within Public Health—How
419 Occupational Settings Drive COVID-19 Disparities. *Am J Epidemiol* 2020; **189**: 1244–53.
- 420 17 Dyal JW, Grant MP, Broadwater K, *et al.* COVID-19 Among Workers in Meat and Poultry Processing
421 Facilities — 19 States, April 2020. *MMWR Morb Mortal Wkly Rep* 2020; **69**.
422 DOI:10.15585/mmwr.mm6918e3.
- 423 18 NCDHHS: North Carolina Identifies First Case of COVID-19.
424 [https://web.archive.org/web/20200407230739/https://www.ncdhhs.gov/news/press-releases/north-carolina-](https://web.archive.org/web/20200407230739/https://www.ncdhhs.gov/news/press-releases/north-carolina-identifies-first-case-covid-19)
425 [identifies-first-case-covid-19](https://web.archive.org/web/20200407230739/https://www.ncdhhs.gov/news/press-releases/north-carolina-identifies-first-case-covid-19) (accessed May 22, 2020).
- 426 19 US Census Bureau. American Community Survey 5-Year Data (2014-2018). 2019
427 <https://www.census.gov/data/developers/data-sets/acs-5year.html> (accessed Dec 3, 2020).
- 428 20 Data Behind the Dashboards: NC DHHS COVID-19. [https://covid19.ncdhhs.gov/dashboard/data-behind-](https://covid19.ncdhhs.gov/dashboard/data-behind-dashboards)
429 [dashboards](https://covid19.ncdhhs.gov/dashboard/data-behind-dashboards) (accessed Feb 13, 2021).
- 430 21 Premkumar L, Segovia-Chumbez B, Jadi R, *et al.* The receptor binding domain of the viral spike protein is
431 an immunodominant and highly specific target of antibodies in SARS-CoV-2 patients. *Sci Immunol* 2020; **5**:
432 eabc8413.
- 433 22 Larremore DB, Fosdick BK, Bubar KM, *et al.* Estimating SARS-CoV-2 seroprevalence and epidemiological
434 parameters with uncertainty from serological surveys. *Elife* 2021; **10**. DOI:10.7554/eLife.64206.
- 435 23 Markmann AJ, Giallourou N, Ryan Bhowmik D, *et al.* Sex disparities and neutralizing antibody durability to
436 SARS-CoV-2 infection in convalescent individuals. *medRxiv* 2021; : 2021.02.01.21250493.
- 437 24 Hou YJ, Okuda K, Edwards CE, *et al.* SARS-CoV-2 Reverse Genetics Reveals a Variable Infection
438 Gradient in the Respiratory Tract. *Cell* 2020; published online May 27. DOI:10.1016/j.cell.2020.05.042.
- 439 25 Centers for Disease Control and Prevention. Interim Guidelines for COVID-19 Antibody Testing | CDC.
440 2020; published online Aug 1. [https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/antibody-tests-](https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/antibody-tests-guidelines.html)
441 [guidelines.html](https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/antibody-tests-guidelines.html) (accessed Feb 26, 2021).
- 442 26 Ripperger TJ, Uhrlaub JL, Watanabe M, *et al.* Orthogonal SARS-CoV-2 Serological Assays Enable
443 Surveillance of Low-Prevalence Communities and Reveal Durable Humoral Immunity. *Immunity* 2020; **53**:
444 925-933.e4.
- 445 27 Hagan LM, Williams SP, Spaulding AC, *et al.* Mass Testing for SARS-CoV-2 in 16 Prisons and Jails — Six
446 Jurisdictions, United States, April–May 2020. *MMWR Morb Mortal Wkly Rep* 2020; **69**: 1139–43.
- 447 28 Webb Hooper M, Nápoles AM, Pérez-Stable EJ. COVID-19 and Racial/Ethnic Disparities. *JAMA - J. Am.*
448 *Med. Assoc.* 2020; **323**: 2466–7.

- 449 29 Figueroa JF, Wadhera RK, Lee D, Yeh RW, Sommers BD. Community-Level Factors Associated With
450 Racial And Ethnic Disparities In COVID-19 Rates In Massachusetts. *Health Aff* 2020; **39**: 1984–92.
451 30 Berchick ER, Barnett JC, Upton RD. Health Insurance Coverage in the United States: 2018 Current
452 Population Reports. 2019.
453 31 Long QX, Tang XJ, Shi QL, *et al.* Clinical and immunological assessment of asymptomatic SARS-CoV-2
454 infections. *Nat Med* 2020; **26**: 1200–4.
455 32 Dan JM, Mateus J, Kato Y, *et al.* Immunological memory to SARS-CoV-2 assessed for up to 8 months after
456 infection. *Science (80-)* 2021; **371**. DOI:10.1126/science.abf4063.
457 33 Diamond M, Chen R, Xie X, *et al.* SARS-CoV-2 variants show resistance to neutralization by many
458 monoclonal and serum-derived polyclonal antibodies. *Res Sq* 2021; published online Feb 10.
459 DOI:10.21203/rs.3.rs-228079/v1.
460 34 Shook-Sa BE, Boyce RM, Aiello AE. Estimation Without Representation: Early Severe Acute Respiratory
461 Syndrome Coronavirus 2 Seroprevalence Studies and the Path Forward. *J Infect Dis* 2020; **222**: 1086–9.
462

463 FIGURE LEGENDS

464 **Figure 1. Catchment area for hospital remnant sample collection for UNC Health hospitals.** Remnant samples
465 were collected from hospital clinical laboratories from each of the four sites indicated by the red dots. (A) Number of
466 samples collected by count as well as (B) the rate of sampling.¹⁹
467

468 **Figure 2. Trends in seroprevalence estimates.** (A) Weekly posterior mean seroprevalence estimates and 95%
469 credible intervals for the study period of 4/21-10/3 of the hospital samples by ELISA plotted over time over the course
470 of the study period. (B) Cumulative daily COVID-19 PCR+ cases from the six-county area 4/19-10/3, and (C) weekly
471 COVID-19 hospitalizations in the six-county area 4/19-10/3 from NC Department of Health and Human Services.
472

473 **Figure 3. Antibody repertoires in an RBD Ig positive subset.** 110 RBD Ig positive samples were chosen at random
474 to undergo SARS-2 antibody repertoire analysis. (A) Percent of individuals with RBD IgM, NTD IgG and functionally
475 neutralizing antibodies (NT50). (B) Correlation plot of NT50 and RBD Ig. (C) Correlation plot of NTD IgG and RBD
476 Ig, r_s = Spearman correlation coefficient displayed in the top left of panels (B) and (C). (D) NT50 values for each
477 diagnosis binning category based on ICD-10 codes. Medians shown in blue. Two-tailed Mann-Whitney,
478 *** $p < 0.0001$, ** $p = 0.0078$.
479

480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504

505 TABLES
506

Table 1. Study participants by demographic factors of interest.							
	4/19-6/13		6/14-8/08		8/09-10/03		6-county
	N	(%)	N	(%)	N	(%)	Demographics (%)
Sex							
Female	1947	56.2	2020	57.3	1450	55.1	51.8
Male	1515	43.7	1508	42.7	1183	44.9	48.2
Unreported	1	0.0	0	0.0	0	0.0	—
Age							
5-17	259	7.5	163	4.6	150	5.7	18.4
18-49	1311	37.9	1052	29.8	830	31.5	48.7
50-64	926	26.7	1030	29.2	725	27.5	19.7
65-99	967	27.9	1283	36.4	928	35.2	13.1
Race/Ethnicity							
NL White	2113	61.0	2267	64.3	1628	61.8	59.7
NL Black	845	24.4	803	22.8	603	22.9	21.0
NL Other	210	6.1	195	5.5	194	7.4	8.2
Latinx	295	8.5	263	7.5	208	7.9	11.1
In/Out patient							
Inpatient	1057	30.5	961	27.2	839	31.9	—
Outpatient	2394	69.1	2562	72.6	1792	68.1	—
Unknown	12	0.3	5	0.1	2	0.1	—
Payor							
Public	1825	52.7	2050	58.1	1509	57.3	—
Private	1249	36.1	1172	33.2	920	34.9	—
Self-Pay	326	9.4	254	7.2	181	6.9	—
Other/Unknown	63	1.8	52	1.4	23	0.8	—

507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531

Table 1. Study participants by demographic factors of interest. Note, because of how the NC census reports data, the sex and age breakdowns of the 6-county demographics includes only individuals over the age of 4 (including those over age 99), but the race/ethnicity breakdown includes individuals of all ages. Additionally, the 65-99 age category is actually age 65+ for the 6-county demographics.

Table 2. Cohort prevalence estimates									
	Positivity			BHM prevalence estimates					
	4/19-6/13	6/14-8/08	8/09-10/03	4/19-6/13		6/14-8/08		8/09-10/03	
				Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Overall	5.3	10.5	10.8	2.9	(1.7, 4.3)	8.8	(7.1, 10.6)	9.1	(7.2, 11.1)
Age									
5-17	3.1	9.8	9.3	1.4	(0.3, 3.3)	8.1	(3.9, 13.4)	7.6	(3.5, 13.0)
18-49	6.0	12.6	10.5	3.6	(2.2, 5.4)	11.1	(8.6, 13.8)	8.7	(6.2, 11.5)
50-64	5.9	10.4	13.0	3.7	(1.9, 5.8)	8.7	(6.3, 11.3)	11.5	(8.5, 14.7)
65-99	4.3	9.0	9.6	1.5	(0.2, 3.4)	7.1	(5.0, 9.4)	7.7	(5.2, 10.4)
Sex									
Female	4.5	10.3	10.7	2.1	(1.0, 3.5)	8.5	(6.6, 10.6)	8.9	(6.8, 11.3)
Male	6.3	10.7	10.9	3.9	(2.3, 5.8)	9.2	(7.1, 11.3)	9.2	(6.9, 11.8)
Race/Ethnicity									
NL White	3.7	7.5	8.3	1.4	(0.5, 2.7)	5.4	(3.7, 7.3)	6.3	(4.3, 8.4)
NL Black	5.6	12.0	12.8	2.6	(0.6, 5.0)	10.4	(7.5, 13.4)	11.4	(8.2, 14.8)
NL Other	5.7	10.3	11.3	2.0	(0.1, 5.9)	8.5	(3.9, 13.9)	9.3	(4.5, 14.9)
Latinx	15.9	31.9	24.0	14.8	(10.4, 19.6)	33.2	(26.8, 40.0)	23.9	(17.5, 31.1)
In/out patient									
Outpatient	4.3	9.0	9.1	2.0	(1.0, 3.3)	7.1	(5.4, 9.0)	7.1	(5.1, 9.2)
Inpatient	7.7	14.6	14.4	5.0	(2.9, 7.4)	13.3	(10.5, 16.2)	13.3	(10.3, 16.4)
Payor									
Private	5.2	9.0	8.9	2.9	(1.5, 4.6)	7.3	(5.3, 9.6)	7.1	(4.7, 9.6)
Public	5.0	9.8	10.7	2.5	(1.2, 4.2)	7.9	(5.9, 9.9)	8.9	(6.8, 11.2)
Self-Pay	4.0	18.9	17.1	1.3	(0.2, 3.5)	18.3	(13.1, 23.8)	16.3	(10.4, 23.1)
Other/Unknown	22.2	30.8	43.5	21.1	(11.8, 31.7)	31.2	(19.4, 44.5)	40.4	(22.4, 60.6)

532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555

Table 2. Cohort prevalence estimates. Raw seropositivity (%) and posterior mean seroprevalence estimates (%) from BHM with 95% credible intervals (lower bound, upper bound). NL, Non-Latinx.

Table 3. Conditional odds ratios of being SARS-CoV-2 seropositive over the study period.								
	4/19-6/13		6/14-8/08		8/09-10/03		4/19-10/03 (overall)	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Sex								
Female	—	—	—	—	—	—	—	—
Male	2·05	(1·08, 4·25)	1·10	(0·80, 1·51)	0·91	(0·64, 1·29)	1·27	(0·98, 1·69)
Race/Ethnicity								
NL White	—	—	—	—	—	—	—	—
NL Black	1·66	(0·53, 4·28)	1·94	(1·31, 2·92)	1·82	(1·20, 2·79)	1·80	(1·19, 2·65)
NL Other	1·26	(0·12, 5·74)	1·58	(0·74, 3·19)	1·81	(0·87, 3·57)	1·54	(0·66, 2·84)
Latinx	14·53	(6·47, 36·72)	7·43	(4·70, 11·97)	4·34	(2·61, 7·41)	7·77	(5·20, 12·10)
Age								
5-17	—	—	—	—	—	—	—	—
18-49	3·09	(0·99, 11·43)	1·38	(0·68, 3·05)	0·89	(0·42, 2·03)	1·56	(0·92, 2·77)
50-64	3·62	(1·13, 13·56)	1·34	(0·64, 2·99)	1·56	(0·76, 3·54)	1·96	(1·15, 3·55)
65-99	1·62	(0·28, 6·90)	1·49	(0·71, 3·34)	1·13	(0·52, 2·64)	1·40	(0·71, 2·61)
In/out patient								
Outpatient	—	—	—	—	—	—	—	—
Inpatient	2·50	(1·31, 5·10)	1·91	(1·38, 2·68)	1·92	(1·34, 2·80)	2·09	(1·59, 2·85)
Payor								
Private	—	—	—	—	—	—	—	—
Public	0·85	(0·41, 1·73)	0·89	(0·58, 1·34)	1·16	(0·74, 1·85)	0·96	(0·70, 1·30)
Self-Pay	0·18	(0·03, 0·64)	1·78	(1·07, 2·93)	1·94	(1·03, 3·63)	0·85	(0·45, 1·41)
Other/ Unknown	3·08	(1·15, 8·23)	2·73	(1·25, 5·98)	6·60	(2·29, 18·71)	3·81	(2·23, 6·54)

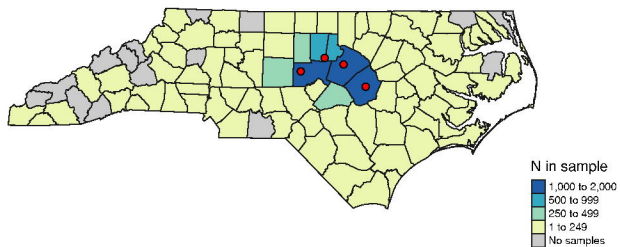
556

557

558 **Table 3. Conditional odds ratios of being SARS-CoV-2 seropositive over the study period.** Data is broken down
 559 into three two-month long periods in central North Carolina. Odds ratios of seropositivity calculated from the BHM
 560 with 95% credible intervals (lower bound, upper bound) are reported where the baseline groups for comparison are
 561 female, Non-Latinx white, age 5-17, outpatient, and private insurance. Odds ratios that do not overlap a value of one
 562 are bolded.

563

A. Samples by county



B. Rate of sampling (samples by county population)

