

Title: Body mass index changes and their association with SARS-CoV-2 infection: a real-world analysis

Author Names: Jithin Sam Varghese^{1,2}, Yi Guo³, Mohammed K. Ali^{1,2,4}, W. Troy Donahoo⁵, Rosette J. Chakkalakal^{1,2,6}

Author Affiliations:

¹Emory Global Diabetes Research Center of Woodruff Health Sciences Center and Emory University, Atlanta, USA.

²Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, USA

³Department of Health Outcomes and Biomedical Informatics, College of Medicine, University of Florida, Gainesville, Florida, USA.

⁴Department of Family and Preventive Medicine, School of Medicine, Emory University, Atlanta, USA

⁵Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine, College of Medicine University of Florida Gainesville FL USA.

⁶Department of Medicine, School of Medicine, Emory University, Atlanta, USA

Corresponding Author:

Dr. Rosette J. Chakkalakal, MD, MHS

Mailing address: 49 Jesse Hill Jr. Dr., Room 494, Atlanta, GA 30303

Email: rchakka@emory.edu

Office Phone: 404-778-1641

Sources of Support: National Institutes of Health (R01DK120814-05S1, P30DK111024)

Running title: BMI changes after SARS-CoV-2 infection

Abbreviations: BMI: Body mass index, CP: Computable phenotype, EHR: Electronic health record, HDL: High density lipoprotein, LDL: Low density lipoprotein, PASC: Post-acute sequelae of COVID-19, T2DM: Type 2 diabetes

Funding: This research was supported by the National Institute for Diabetes and Digestive and Kidney Diseases (NIDDK) of the National Institutes of Health, award number 3R01DK120814-05S1. MKA was partially supported by the Georgia Center for Diabetes Translation Research which is funded by the National Institutes of Health (P30DK111024).

Conflicts of Interest: The authors declare no competing financial interests.

ABSTRACT

Objective:

To study body mass index (BMI) changes among individuals aged 18-99 years with and without SARS-CoV-2 infection.

Subjects/Methods:

Using real-world data from the OneFlorida+ Clinical Research Network of the National Patient-Centered Clinical Research Network, we compared changes over time in BMI in an Exposed cohort (positive SARS-CoV-2 test between March 2020 - January 2022), to a contemporary Unexposed cohort (negative SARS-CoV-2 tests), and an age/sex-matched Historical control cohort (March 2018 - January 2020). Body mass index (kg/m^2) was retrieved from objective measures of height and weight in electronic health records. We used target trial approaches to estimate BMI at baseline and change per 100 days of follow-up for Unexposed and Historical cohorts relative to the Exposed cohort by categories of sex, race-ethnicity, age, and hospitalization status.

Results:

The study sample consisted of 44,436 (Exposed cohort), 164,118 (Unexposed cohort), and 41,189 (Historical cohort). Cumulatively, 62% were women, 21.5% Non-Hispanic Black, 21.4% Hispanic and 5.6% Non-Hispanic Other. Patients had an average age of 51.9 years (SD: 18.9). At baseline, relative to the Exposed cohort (mean BMI: 29.3 kg/m^2 [95% CI: 29.0, 29.7]), the Unexposed (-0.07 kg/m^2 [95% CI; -0.12, -0.01]) and Historical controls (-0.27 kg/m^2 [95% CI; -0.34, -0.20]) had lower BMI. Relative to no change in the Exposed over 100 days (0.00 kg/m^2 [95% CI; -0.03, 0.03]), the BMI of those Unexposed decreased (-0.04 kg/m^2 [95% CI; -0.06, -0.01]) while the Historical cohort's BMI increased ($+0.03 \text{ kg/m}^2$ [95% CI; 0.00, 0.06]). BMI changes were consistent between Exposed and Unexposed cohorts for most population groups, except at start of follow-up period among Males and those 65 years or older, and in changes over 100 days among Males and Hispanics.

Conclusions:

In a diverse real-world cohort of adults, mean BMI of those with and without SARS-CoV2 infection varied in their trajectories. The mechanisms and implications of weight retention following SARS-CoV-2 infection remain unclear.

1 **Introduction**

2 Poor cardiometabolic health, including obesity and diabetes, are known risk factors for
3 COVID-19 severity and mortality.¹ Recent studies of post-acute sequelae of COVID-19 (PASC)
4 suggest COVID-19 may, in turn, worsen cardiometabolic health as increased rates of type 2
5 diabetes, high blood pressure, and dyslipidemia have been reported for individuals infected with
6 SARS-CoV-2.²⁻⁸ Few studies have examined changes in body mass index (BMI) which may
7 contribute to observed associations of COVID-19 with incident cardiometabolic diseases,
8 particularly for younger, healthier adults. Racial and ethnic minorities in the US have also been
9 underrepresented in prior studies of PASC, making the findings of this research difficult to
10 generalize to communities that experience disproportionately higher rates of SARS-CoV-2
11 infection, COVID-19-related morbidity and mortality, and cardiometabolic disease in the US.⁹⁻¹¹
12 Understanding the magnitude and direction of changes in cardiometabolic risk factors during the
13 post-acute period of SARS-CoV-2 infection status in diverse, real-world populations can provide
14 insights into pathways through which COVID-19 may increase rates of type 2 diabetes and other
15 cardiometabolic conditions.

16 The purpose of this real-world evidence study was to characterize key changes in BMI
17 among individuals who survived the acute phase (defined as the first 30 days) of SARS-CoV-2
18 infection in a socio-demographically diverse cohort and compare them with a contemporary
19 cohort who always tested negative for SARS-CoV-2 as well as a historical cohort (prior to the
20 pandemic) matched on age and sex to those who tested positive. Using analytical approaches for
21 target trial emulation, we examined if BMI changes differed by socio-demographic (sex, race-
22 ethnicity, age) category and hospitalization status.

23 **Methods**

24 *Study Design*

25 We conducted a secondary analysis of patient-level electronic health record (EHR) data
26 from the OneFlorida+ clinical research network (CRN) in the National Patient-Centered Clinical
27 Research Network (PCORnet).¹² PCORnet CRNs have become a valuable resource to study
28 symptom profiles of PASC.^{13,14} For the current study, the OneFlorida+ Data Analytics team
29 independently queried the clinical data warehouse, built on the PCORnet Common Data Model,
30 to extract data on socio-demographic and clinical characteristics of eligible patients.¹⁵

31 We built three retrospective cohorts of adult individuals who sought care at least three
32 times in the OneFlorida+ clinical sites over the period 2016-22.¹² Individuals were included if
33 they were 18 years or older with at least two inpatient or outpatient encounters recorded in
34 OneFlorida+ in the two years before inclusion, and at least one inpatient or outpatient encounter
35 in the period after inclusion (30 days from inclusion until last visit or 28 February 2022).
36 Individuals were excluded if they had diabetes mellitus, as defined by previously validated
37 computable phenotypes for type 1 diabetes (T1DM) and type 2 diabetes (T2DM; at least two of
38 three criteria: ICD-9/10 diagnosis code for diabetes, antidiabetic medication, and/or HbA1c
39 $\geq 6.5\%$), in the 24 months preceding the date of first test or COVID-19 diagnosis code
40 **(Supplementary Table 1)**.¹⁶

41 The Institutional Review Board of Emory University determined the study met criteria
42 for exemption under 45 CFR 46.104(d). This study followed the REporting of studies Conducted
43 using Observational Routinely-collected health Data (RECORD) Statement.¹⁷

44 *Exposure Cohorts*

45 The Exposed cohort were individuals exposed to SARS-CoV-2, defined by meeting at
46 least one of two criteria between 1 March 2020 and 29 January 2022: any positive test (Nucleic
47 Acid Amplification Tests [NAAT] or antigen) or COVID-19 related ICD-10-CM codes (E08 to
48 E13).¹³ The index date of infection exposure (T_0) was defined as the date of the first positive test
49 or date of record for the ICD-10-CM code. The timeline for cohort selection and follow-up is
50 presented in **Supplementary Figure 1**.

51 The Unexposed cohort were individuals not measured to have had exposure to SARS-
52 CoV-2 and consisted of individuals who met all of three criteria between 1 March 2020 and 29
53 January 2022: at least one negative test (NAAT or antigen), no COVID-19 related International
54 Classification of Diseases-10 CM (ICD-10 CM) codes (E08.X to E13.X), and no positive test
55 during follow-up period. Since the Unexposed cohort tested negative throughout the follow-up
56 period, to maximize follow-up duration, the index date (T_0) was defined as the date of the first
57 negative test.

58 The Historical cohort was constructed to be representative of typical clinic visits among
59 patients who utilized the healthcare delivery systems in the pre-pandemic period, and ensure
60 differences observed in the Exposed and Unexposed cohorts were not due to the pandemic
61 period. Patients who were not subsequently part of Exposed or Unexposed cohorts were sampled
62 into the Historical cohort after matching on sex and 5-year age intervals. The index dates of
63 Historical cohort ranged from 2 March 2018 and 30 January 2020, and were selected such that
64 the last date of possible follow-up (29 February 2020) did not overlap with the start of the
65 calendar period of identification (1 March 2020) for the Exposed and Unexposed cohorts.

66 We restricted our study sample to individuals who had at least one measurement of
67 weight in the year preceding their observation period in our study (**Supplementary Figure 2**).

68 The final analytic sample consisted of 44,436 Exposed, 164,118 Unexposed controls and 41,189
69 Historical controls. All subsequent analysis accounted for the stratified sampling of Historical
70 controls.

71 ***Outcomes – Cardiometabolic Health Indicators***

72 We explored average longitudinal change in body mass index (BMI) based on objective
73 measures recorded in the EHR. Follow-up period (in days) was calculated as number of days
74 from 30 days after index date (i.e. Time $t_F = t - [T_0 + 30]$), defined as the post-acute phase of the
75 disease. To exclude weight changes associated with severe illness and weight loss prior to
76 mortality, we included only those participants who were alive at least 30 days after SARS-CoV2
77 infection was detected.^{2,13}

78 ***Effect Modifiers – Socio-demographic Characteristics and Hospitalization Status***

79 Socio-demographic characteristics included sex (male, female), race-ethnicity (Non-
80 Hispanic White, Non-Hispanic Black, Hispanic), and age at index date (18-39 years, 40-64 years,
81 65 years and older). We determined all-cause hospitalization status (not hospitalized,
82 hospitalized) based on the clinical encounter coded on the index date.

83 ***Confounders***

84 We used a combination of biologically plausible covariates and empirically identified
85 covariates to adjust for baseline differences between cohorts using inverse probability weighting
86 (IPW) (see **Supplementary Methods**). Biologically plausible covariates known to be associated
87 with higher cardiometabolic risk were derived from data collected before the index date. We
88 used data collected within 1 year of index date to define smoking status (yes/no), use of
89 medications relevant to cardiometabolic health (yes/no; antihypertensives, statins,

90 antidepressants, antipsychotics, steroids), and laboratory parameters (HbA1c, random glucose,
91 lipid panel). Given the lower frequency of reporting diagnosis codes, we used data collected
92 within 2 years before the index date to identify comorbidities using grouped ICD-10-CM codes
93 (yes/no; obesity, cardiovascular disease, cerebrovascular disease, hyperlipidemia, hypertension,
94 pulmonary disease). Information on medication, laboratory, and diagnosis codes used for
95 categorization are provided in **Supplementary Table 1**. We identified the health system and
96 primary insurance type (Medicare, Medicaid, Other Government, Private, No Insurance, No
97 Information) used on index date.

98 Empirically identified covariates were computed across six data dimensions available in
99 OneFlorida+ : inpatient diagnoses, outpatient diagnoses, inpatient procedures, outpatient
100 procedures, prescriptions, and laboratory tests.¹⁸ We used participant records from within 1 year
101 before index date to compute the covariates. Diagnostic codes and prescriptions were categorized
102 into 537 Clinical Classifications Software Refined (CCSR) categories and 89 Anatomical
103 Therapeutic Chemical (ATC) Level 3 classes respectively. The 1,951 procedure codes (Current
104 Procedural Terminology, ICD-10-CM) and abnormalities in 876 laboratory tests (Logical
105 Observation Identifiers Names and Codes [LOINC]) were included without additional
106 categorization. We identified the top 200 covariates in each data domain based on their
107 prevalence in the sample. For each covariate, we then computed three intensity dummy variables
108 (at least once, sporadic [more often than median participant], frequent [more often than the 75th
109 percentile]).

110 *Statistical Analysis*

111 We present an overview of the analytic strategy in **Supplementary Figure 3**.

112 *Inverse probability weighting for cohort membership and follow-up*

113 Since this is an observational study, baseline characteristics may differ between exposed,
114 unexposed, and historical control cohorts resulting in non-random cohort membership. To
115 minimize instrumental variable bias when constructing propensity scores, we performed variable
116 selection of empirically identified covariates longitudinally associated with BMI after adjusting
117 for multiple comparison correction using the Benjamini-Hochberg procedure. Next, random
118 forests models for probability of membership in each exposure cohort, relative to other cohorts,
119 were fit separately with 5-fold cross-validation. We tuned the parameters based on the area under
120 receiver operating characteristic curve (AUC) and identified 2000 trees and 10 observations per
121 node as the best combination of hyperparameters. We included both biologically-plausible and
122 algorithmically selected covariates to estimate the high-dimensional propensity score.^{19,20} We
123 assessed covariate balance between the Unexposed and Historical Control cohorts, relative to the
124 Exposed, after IPW using population standardized bias.²¹ Population standardized bias is the
125 maximum difference between IP weighted group mean and unweighted pooled mean.²¹ Bias less
126 than 0.1 after weighting was considered as indicative of covariate balance.

127 We additionally constructed IPW for availability of follow-up data for BMI to minimize
128 selection bias, since not all participants had data on all cardiometabolic health indicators during
129 the follow-up period.²² The final weight for each individual in the analytic sample is a product of
130 treatment weights and selection weights. We provide additional detail of the modeling strategy in
131 an extended methodological note (**Supplementary Methods**).

132 IPW were also constructed separately for each socio-demographic, clinical, and
133 community characteristic of interest with numerator reflecting probability of cohort membership

134 under levels of each characteristic (sex, race-ethnicity, age) when assessing differences between
135 levels.²³

136 *Changes in BMI*

137 First, we modeled BMI in the follow-up period for Exposed, Unexposed, and Historical
138 Control cohorts using marginal structural models. We used a difference-in-differences
139 framework to understand if changes in BMI per 100 days of follow-up differed for Unexposed
140 and Historical cohorts relative to the Exposed cohort. We adjusted for all imbalanced
141 characteristics except laboratory parameters in the regression model since the latter had high
142 rates of missingness. Second, we estimated average BMI at origin date and longitudinal change
143 in BMI per 100 days of follow-up across categories of sex (male vs female), race-ethnicity (NH
144 White vs NH Black vs Hispanic), age (18-39 vs 40-64 vs 65 and older) and hospitalization status
145 by exposure group. We did not adjust for multiple comparisons in these prespecified subgroup
146 analysis.

147 *Sensitivity Analysis*

148 First, to minimize the influence of patients who rarely used the health system, we
149 restricted the analytic sample to those patients with at least 100 days of follow-up and at least
150 two encounters where BMI was measured in the follow-up period. Second, to account for
151 differences in health outcomes due to differences in reasons for visit during follow-up period, we
152 adjusted for number and types of clinical encounters between measurements of the
153 cardiometabolic risk factors as a time-varying covariate. Third, to account for differences in
154 timing of follow-up visits, we adjusted for COVID-19 transmission in county of residence on
155 dates when cardiometabolic risk factors were measured. COVID-19 cases and test positivity per

156 100,000 persons in the 7 days prior to the date of measurement were obtained from Centers for
157 Disease Control and Prevention's COVID-19 County Level of Community Transmission data
158 archive. New cases and test positivity would be zero on all dates for the Historical Control
159 cohort. Finally, we used multiple imputation (5 datasets) to impute missing values of laboratory
160 parameters in lookback period and adjusted for imbalanced covariates in the regression model.

161 All analysis was carried out using R 4.2.0 and Python 3.11.4.

162 **Results**

163 The study sample consisted of 249,743 adults, of whom 44,436 (19.2%) belonged to the
164 Exposed cohort, 164,118 (61.5%) belonged to the Unexposed cohort, and 41,189 (19.3%)
165 belonged to the Historical cohort (**Table 1**). Most of the study sample were women (155,223
166 [62%]), with an average age of 51.9 years (SD: 18.9). Nearly half the sample belonged to
167 minority race-ethnicity groups (Non-Hispanic Black: 21.5%, Hispanic: 21.4%, Non-Hispanic
168 Other: 5.6%). Most patients used private insurance (46%) or Medicare (18%) on index date and
169 27% were hospitalized on index date. The average BMI on index date in the study sample was
170 29.5 kg/m² (SD: 7.4) with 30% and 40% having overweight and obesity respectively. The
171 excluded sample without weight available in the preceding year were healthier than the study
172 sample with weight available (**Supplementary Table 2**).

173 The Exposed cohort was younger (48.8 years [SD: 18.4]) and more likely to be female
174 (65%) relative to the Unexposed cohort (age: 53.6 years [SD: 19.0], 61% female). The Exposed
175 cohort was also more likely to be Non-Hispanic Black or Hispanic race-ethnicity (50.6% vs
176 40.6% Unexposed and 43.5% Historical), privately insured (53% vs 42% Unexposed and 42%
177 Historical) and had higher obesity (46% vs 39% Unexposed and 38% Historical). The

178 Unexposed cohort was more likely to be hospitalized due to any cause relative to other cohorts
179 (31% vs 28% Exposed and 7.4% Historical). The proportion with comorbidities and medications
180 prescribed were similar between the three cohorts. After inverse probability weighting
181 (**Supplementary Table 3**), some covariates remained imbalanced between the three groups,
182 namely age, race-ethnicity, smoking, health system at admission, hospitalization status, BMI,
183 and laboratory parameters (HbA1c, serum creatinine, HDL, LDL) in the previous year.

184 *Trajectories of body mass index in follow-up period*

185 Of the study sample, only 194,792 (77.9%) had BMI information in the follow-up period
186 (Exposed: 32,384, Unexposed: 127,142, Historical: 35,266). Socio-demographic and clinical
187 characteristics were similar between those with and without BMI information in the follow-up
188 period (**Supplementary Table 2**). The random forests model to predict loss to follow-up had
189 high model fit in the hold-out test set (AUROC: 0.97; **Supplementary Methods Note**). The time
190 (median [IQR]) to the first follow-up visit of BMI was shorter for Exposed (39 [12, 107]) and
191 Unexposed (33 [11, 98]) cohorts, relative to the Historical (74 [22, 180]) cohort (**Supplementary**
192 **Table 4**). The Historical cohort had more frequent (median: 4 vs 3 each for Exposed and
193 Unexposed) and longer duration of follow-up (median days [IQR]; 464 [271, 592] vs 178 [76,
194 357] for Exposed and 212 [85, 374] for Unexposed). BMI at the first follow-up was higher for
195 the Exposed (29.3 kg/m², 95% Confidence Interval [CI]: 24.9, 34.9), relative to Unexposed (27.9
196 kg/m², 95%CI: 23.9, 33.0) and Historical (27.8 kg/m², 95%CI: 23.8, 32.8) cohorts.

197 Adjusting for imbalanced covariates and after IPW, the average estimated BMI at origin
198 date (**Table 2**) was higher for the Exposed (29.3 kg/m², 95%CI: 29.0, 29.7) relative to the
199 Unexposed (difference: -0.07 kg/m², 95%CI: -0.12, -0.01) and Historical (difference: -0.27
200 kg/m², 95%CI: -0.34, -0.20) cohorts. Change in BMI per 100 days of follow-up was null for the

201 Exposed (0.00 kg/m^2 , 95%CI: -0.03, 0.03), negative for the Unexposed (-0.03 kg/m^2 , 95%CI: -
202 0.04, -0.02) and positive for the Historical (0.04 kg/m^2 , 95%CI: 0.03, 0.05) cohort. The change in
203 BMI per 100 days of follow-up for the Unexposed (-0.04 kg/m^2 , 95%CI: -0.06, -0.01) and
204 Historical (0.03 kg/m^2 , 95%CI: 0.00, 0.06) cohorts were different, relative to the Exposed cohort
205 (**Table 2**).

206 Adjusted differences in BMI at origin date for subgroups of sex, age, race-ethnicity, and
207 hospitalization status are presented in **Figure 1A**. Across most socio-demographic subgroups
208 (except among adult Males, those aged 65 and older, Hispanic adults, non-hospitalized adults)
209 BMI at origin date was higher among Exposed, relative to Unexposed and Historical cohorts
210 (**Supplementary Table 5**). Relative to Exposed cohort (**Figure 1B**), the change in BMI per 100
211 days of follow-up among the Unexposed cohort was lower among Males (difference-in-
212 differences [DiD]: -0.07 kg/m^2 , 95%CI: -0.12, -0.02) and non-Hospitalized adults (DiD: -0.04
213 kg/m^2 , 95%CI: -0.07, -0.01). Relative to the Exposed cohort, the change in BMI per 100 days of
214 follow-up among the Historical cohort was higher among Females (DiD: 0.07 kg/m^2 , 95%CI:
215 0.03, 0.11), Non-Hispanic White adults (DiD: 0.07 kg/m^2 , 95%CI: 0.03, 0.11) and non-
216 Hospitalized adults (DiD: 0.03 kg/m^2 , 95%CI: 0.00, 0.06)

217 Results were similar when restricting to those with at least 100 days of follow-up and at
218 least two encounters when BMI was measured (**Supplementary Table 6**), after adjusting for
219 number and types of encounters between BMI measurements (**Supplementary Table 7**) and
220 after adjusting for COVID-19 cases and testing (**Supplementary Table 8**). Results were also
221 similar after accounting for missing data in laboratory parameters using multiple imputation
222 (**Supplementary Table 9**).

223 **Discussion**

224 In a diverse real-world cohort of adults who sought medical care during the pandemic,
225 those who tested positive for COVID-19 (Exposed) were heavier at the start of the post-acute
226 phase of SARS-CoV-2 infection compared to contemporary controls who tested negative
227 (Unexposed), and age-sex matched historical controls. The BMI of those Exposed did not change
228 per 100 days of follow-up, while the BMI of those Unexposed decreased, and the BMI of
229 historical controls increased. The change in BMI over time differed between Exposed and
230 Unexposed cohorts at start of follow-up among Males and those 65 years or older, and in
231 changes per 100 days among Males and Hispanics.

232 The persistent higher BMI, including before infection, among those Exposed compared to
233 those who were Unexposed suggest that those infected by SARS-CoV-2 may have been at higher
234 risk of COVID-19 when seeking care.⁵ This finding is consistent with numerous reports of
235 elevated BMI being a risk factor for severe COVID-19.^{24,25} The current study extends the
236 findings of this prior research to examine BMI trajectories following SARS-CoV-2 infection at
237 the individual level. To date, most studies have examined the pandemic's impact on BMI at a
238 population level. For instance, a meta-analysis of observational studies showed that the
239 pandemic resulted in a small increase in BMI and prevalence of obesity among adults.^{26,27}
240 Potential explanations for BMI increases at the individual and population level include rise in
241 sedentary time and unhealthy dietary habits, downward social mobility from loss of income and
242 employment, and disruption of preventive care.²⁶ The bidirectional relationship of viral
243 infections and cardiometabolic disease may be mediated by excess adiposity, insulin resistance,
244 inflammation and/or delayed recovery after COVID-19 as a post-acute sequelae.^{2-4,28,29}

245 This analysis of real-world data has several strengths. We used a validated computable
246 phenotype for prevalent T2DM as part of our exclusion criteria to identify a cohort of adults free

247 of T2DM, who were users of the healthcare system before the pandemic, and who were tested for
248 SARS-CoV-2. We additionally included a historical control cohort, matched on age and sex, to
249 characterize differences in the patient population attributable to the pandemic. The OneFlorida+
250 Data Trust is a part of the National Patient-Centered Clinical Research Network (PCORnet) and
251 follows the common data model, facilitating reproducibility. Compared to other real-world
252 studies, the analytic sample was younger and more diverse across categories of race-ethnicity,
253 age, and sex.^{2,5} The duration of follow-up was longer than other studies reporting post-acute
254 sequelae of COVID-19.^{3,5,6,13} We used methods for target trial emulation to minimize
255 confounding, including minimizing the potential for instrumental variable bias in high
256 dimensional confounder selection.

257 There are several limitations for this study. First, the OneFlorida+ CRN includes data
258 from many but not all of Florida's healthcare providers; our analyses could not account for
259 patients testing positive at centers not participating in the OneFlorida+ CRN. However, the
260 criteria we used for the lookback period were meant to identify regular users of healthcare
261 systems participating in OneFlorida+.³⁰ Additionally, our sample selection was not biased by
262 home testing because it became widespread only after the last date of follow-up in February
263 2022. Second, although we used a high dimensional propensity score for confounding
264 adjustment, we cannot rule out all differences in clinical characteristics at the time of the index
265 encounter. For instance, those Exposed and Unexposed may have had different reasons for
266 seeking care. Finally, singly-robust machine learning based propensity scores are inferior to
267 doubly-robust methods for lower bias and confidence-interval coverage.³¹ However, to our
268 knowledge, there were no standard software packages to implement doubly-robust causal
269 inference methods for sparse, longitudinal datasets such as electronic health records.

270 This analysis of diverse, real-world data showed that among those who sought medical
271 care during the pandemic, patients testing positive were heavier at the start of the post-acute
272 phase of COVID-19 and retained weight during the post-acute phase of follow-up compared to
273 those testing negative throughout the observation period. Although the subgroup analysis
274 revealed some heterogeneity in these associations, the observed associations were present among
275 racial-ethnic minorities and young adults who were disproportionately affected by COVID-19.
276 The mechanisms influencing this observed increase in weight retention after COVID-19
277 infection warrant further investigation.

278

279

ACKNOWLEDGEMENT/SUPPORT

Acknowledgements: This research was supported by the National Institute for Diabetes and Digestive and Kidney Diseases (NIDDK) of the National Institutes of Health, award number 3R01DK120814-05S1. MKA was partially supported by the Georgia Center for Diabetes Translation Research which is funded by the National Institutes of Health (P30DK111024). The authors thank the OneFlorida+ Data Trust team (Kathryn Shaw, Meggen Kaufman, Jiang Bian, Elizabeth Shenkman) for support on query development and data extraction. The authors thank Shihab Chowdhury for administrative support.

Author contributions: RJC, WTD and JSV conceptualized the study with inputs from MKA and YG. JSV conducted the analysis. JSV wrote the first draft with inputs from RJC. All authors reviewed and edited subsequent drafts.

Competing Interest: The authors declare no competing financial interests.

Data availability statement: The code for the analysis is available on https://github.com/jvargh7/pasc_cardiometabolic_risk. Information of the OneFlorida+ CRN is provided at <https://onefloridaconsortium.org/>, and OneFlorida+ data are made available to researchers with an approved study protocol at <https://onefloridaconsortium.org/front-door/prepare-to-research-data-query/>. For questions regarding OneFlorida+, email: OneFloridaOperations@health.ufl.edu.

References

1. Singh AK, Khunti K. COVID-19 and Diabetes. *Annu Rev Med*. 2022;73(1):129-147. doi:10.1146/annurev-med-042220-011857
2. Xie Y, Al-Aly Z. Risks and burdens of incident diabetes in long COVID: a cohort study. *The Lancet Diabetes & Endocrinology*. 2022;10(5):311-321. doi:10.1016/S2213-8587(22)00044-4
3. Barrett CE, Koyama AK, Alvarez P, et al. Risk for Newly Diagnosed Diabetes >30 Days After SARS-CoV-2 Infection Among Persons Aged <18 Years — United States, March 1, 2020–June 28, 2021. *MMWR Morb Mortal Wkly Rep*. 2022;71(2):59-65. doi:10.15585/mmwr.mm7102e2
4. Rathmann W, Kuss O, Kostev K. Incidence of newly diagnosed diabetes after Covid-19. *Diabetologia*. Published online March 16, 2022. doi:10.1007/s00125-022-05670-0
5. Holman N, Barron E, Young B, et al. Comparative Incidence of Diabetes Following Hospital Admission for COVID-19 and Pneumonia: A Cohort Study. *Diabetes Care*. 2023;46(5):938-943. doi:10.2337/dc22-0670
6. Al-Aly Z, Xie Y, Bowe B. High-dimensional characterization of post-acute sequelae of COVID-19. *Nature*. 2021;594(7862):259-264. doi:10.1038/s41586-021-03553-9
7. Xu E, Xie Y, Al-Aly Z. Risks and burdens of incident dyslipidaemia in long COVID: a cohort study. *The Lancet Diabetes & Endocrinology*. 2023;11(2):120-128. doi:10.1016/S2213-8587(22)00355-2
8. Xie Y, Xu E, Bowe B, Al-Aly Z. Long-term cardiovascular outcomes of COVID-19. *Nat Med*. 2022;28(3):583-590. doi:10.1038/s41591-022-01689-3
9. Mude W, Oguoma VM, Nyanhanda T, Mwanri L, Njue C. Racial disparities in COVID-19 pandemic cases, hospitalisations, and deaths: A systematic review and meta-analysis. *J Glob Health*. 2021;11:05015. doi:10.7189/jogh.11.05015
10. Dalsania AK, Fastiggi MJ, Kahlam A, et al. The Relationship Between Social Determinants of Health and Racial Disparities in COVID-19 Mortality. *J Racial and Ethnic Health Disparities*. 2022;9(1):288-295. doi:10.1007/s40615-020-00952-y
11. Beckles GL, Chou CF. Disparities in the Prevalence of Diagnosed Diabetes — United States, 1999–2002 and 2011–2014. *MMWR Morb Mortal Wkly Rep*. 2016;65(45):1265-1269. doi:10.15585/mmwr.mm6545a4
12. Hogan WR, Shenkman EA, Robinson T, et al. The OneFlorida Data Trust: a centralized, translational research data infrastructure of statewide scope. *Journal of the American Medical Informatics Association*. 2022;29(4):686-693. doi:10.1093/jamia/ocab221

13. Zhang H, Zang C, Xu Z, et al. Data-driven identification of post-acute SARS-CoV-2 infection subphenotypes. *Nat Med*. 2023;29(1):226-235. doi:10.1038/s41591-022-02116-3
14. Hernandez-Romieu AC, Carton TW, Saydah S, et al. Prevalence of Select New Symptoms and Conditions Among Persons Aged Younger Than 20 Years and 20 Years or Older at 31 to 150 Days After Testing Positive or Negative for SARS-CoV-2. *JAMA Netw Open*. 2022;5(2):e2147053. doi:10.1001/jamanetworkopen.2021.47053
15. PCORnet. Common Data Model (CDM) Specification, Version 6.0. Published online January 2022. https://pcorntest.org/wp-content/uploads/2022/01/PCORnet-Common-Data-Model-v60-2020_10_221.pdf
16. Wiese AD, Roumie CL, Buse JB, et al. Performance of a computable phenotype for identification of patients with diabetes within PCORnet: The Patient-Centered Clinical Research Network. *Pharmacoepidemiol Drug Saf*. 2019;28(5):632-639. doi:10.1002/pds.4718
17. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Med*. 2015;12(10):e1001885. doi:10.1371/journal.pmed.1001885
18. Rassen JA, Blin P, Kloss S, et al. High-dimensional propensity scores for empirical covariate selection in secondary database studies: Planning, implementation, and reporting. *Pharmacoepidemiology and Drug*. 2023;32(2):93-106. doi:10.1002/pds.5566
19. Tian Y, Schuemie MJ, Suchard MA. Evaluating large-scale propensity score performance through real-world and synthetic data experiments. *International Journal of Epidemiology*. 2018;47(6):2005-2014. doi:10.1093/ije/dyy120
20. Simon N, Friedman J, Hastie T. A Blockwise Descent Algorithm for Group-penalized Multiresponse and Multinomial Regression. Published online 2013. doi:10.48550/ARXIV.1311.6529
21. McCaffrey DF, Griffin BA, Almirall D, Slaughter ME, Ramchand R, Burgette LF. A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statist Med*. 2013;32(19):3388-3414. doi:10.1002/sim.5753
22. Weuve J, Tchetgen Tchetgen EJ, Glymour MM, et al. Accounting for Bias Due to Selective Attrition: The Example of Smoking and Cognitive Decline. *Epidemiology*. 2012;23(1):119-128. doi:10.1097/EDE.0b013e318230e861
23. Robins JM, Hernán MÁ, Brumback B. Marginal Structural Models and Causal Inference in Epidemiology: *Epidemiology*. 2000;11(5):550-560. doi:10.1097/00001648-200009000-00011
24. Gao M, Piernas C, Astbury NM, et al. Associations between body-mass index and COVID-19 severity in 6.9 million people in England: a prospective, community-based, cohort study. *The Lancet Diabetes & Endocrinology*. 2021;9(6):350-359. doi:10.1016/S2213-8587(21)00089-9

25. Kompaniyets L, Goodman AB, Belay B, et al. Body Mass Index and Risk for COVID-19–Related Hospitalization, Intensive Care Unit Admission, Invasive Mechanical Ventilation, and Death — United States, March–December 2020. *MMWR Morb Mortal Wkly Rep*. 2021;70(10):355–361. doi:10.15585/mmwr.mm7010e4
26. Anderson LN, Yoshida □ Montezuma Y, Dewart N, et al. Obesity and weight change during the COVID □ 19 pandemic in children and adults: A systematic review and meta □ analysis. *Obesity Reviews*. 2023;24(5):e13550. doi:10.1111/obr.13550
27. Lin AL, Vittinghoff E, Olgin JE, Pletcher MJ, Marcus GM. Body Weight Changes During Pandemic-Related Shelter-in-Place in a Longitudinal Cohort Study. *JAMA Netw Open*. 2021;4(3):e212536. doi:10.1001/jamanetworkopen.2021.2536
28. Perakakis N, Harb H, Hale BG, et al. Mechanisms and clinical relevance of the bidirectional relationship of viral infections with metabolic diseases. *The Lancet Diabetes & Endocrinology*. 2023;11(9):675–693. doi:10.1016/S2213-8587(23)00154-7
29. Harding JL, Oviedo SA, Ali MK, et al. The bidirectional association between diabetes and long-COVID-19 - A systematic review. *Diabetes Res Clin Pract*. 2023;195:110202. doi:10.1016/j.diabres.2022.110202
30. Haneuse S, Arterburn D, Daniels MJ. Assessing Missing Data Assumptions in EHR-Based Studies: A Complex and Underappreciated Task. *JAMA Netw Open*. 2021;4(2):e210184. doi:10.1001/jamanetworkopen.2021.0184
31. Naimi AI, Mishler AE, Kennedy EH. Challenges in Obtaining Valid Causal Effect Estimates With Machine Learning Algorithms. *American Journal of Epidemiology*. 2023;192(9):1536–1544. doi:10.1093/aje/kwab201

Table 1. Socio-demographic, clinical and community characteristics of exposure cohorts before weighting and population standardized bias after weighting.

| | Overall (n = 249,743) | Unweighted | | | Population Standardized Bias |
|---|----------------------------------|---------------------------------|------------------------------------|------------------------------------|---|
| | | Exposed (n = 44,436) | Unexposed (n = 164,118) | Historical (n = 41,189) | |
| Female (%) | 155,223 (62%) | 29,028 (65%) | 99,949 (61%) | 26,246 (64%) | 0.051 |
| Age (years) | 51.9 (18.9) | 48.8 (18.4) | 53.6 (19.0) | 48.7 (18.4) | 0.186 |
| Race/ethnicity | | | | | |
| NH White | 128,613 (51.5%) | 19,554 (44.0%) | 88,129 (53.7%) | 20,930 (50.8%) | |
| NH Black | 53,702 (21.5%) | 11,919 (26.8%) | 32,591 (19.9%) | 9,192 (22.3%) | 0.108 |
| Hispanic | 53,354 (21.4%) | 10,581 (23.8%) | 34,038 (20.7%) | 8,735 (21.2%) | 0.059 |
| NH Other | 14,023 (5.6%) | 2,352 (5.3%) | 9,340 (5.7%) | 2,331 (5.7%) | 0.01 |
| Smoking | 117,593 (47%) | 18,818 (42%) | 78,470 (48%) | 20,305 (49%) | 0.162 |
| Health System | | | | | |
| Source 1 | 92,403 (42%) | 15,492 (36%) | 61,533 (46%) | 15,378 (37%) | 0.127 |
| Source 2 | 51,650 (24%) | 8,225 (19%) | 32,151 (24%) | 11,274 (27%) | 0.168 |
| Source 3 | 24,524 (11%) | 5,429 (12%) | 15,432 (12%) | 3,663 (8.9%) | 0.02 |
| Source 4 | 19,041 (8.7%) | 6,148 (14%) | 8,960 (6.7%) | 3,933 (9.5%) | 0.057 |
| Source 5 | 12,263 (5.6%) | 2,811 (6.5%) | 4,718 (3.6%) | 4,734 (11%) | 0.062 |
| Source 6 | 988 (0.5%) | 187 (0.4%) | 539 (0.4%) | 262 (0.6%) | 0.01 |
| Source 7 | 16,775 (7.7%) | 5,265 (12%) | 9,565 (7.2%) | 1,945 (4.7%) | 0.087 |
| Missing | 32,099 | 879 | 31,220 | 0 | |
| Primary Insurance | | | | | |
| Medicare | 38,768 (18%) | 5,809 (13%) | 25,135 (19%) | 7,824 (19%) | 0.014 |
| Medicaid | 18,154 (8.3%) | 3,365 (7.7%) | 11,196 (8.4%) | 3,593 (8.7%) | 0.007 |
| Other Government | 5,299 (2.4%) | 1,257 (2.9%) | 3,314 (2.5%) | 728 (1.8%) | 0.016 |
| Private | 99,648 (46%) | 23,248 (53%) | 54,879 (41%) | 21,521 (52%) | 0.02 |
| No Insurance | 4,210 (1.9%) | 568 (1.3%) | 1,895 (1.4%) | 1,747 (4.2%) | 0.022 |
| No Information | 51,565 (24%) | 9,310 (21%) | 36,479 (27%) | 5,776 (14%) | 0.025 |
| Missing | 32,099 | 879 | 31,220 | 0 | |
| Hospitalization within 30 days of index date | 66,386 (27%) | 12,506 (28%) | 50,830 (31%) | 3,050 (7.4%) | 0.36 |

| | | | | | |
|--|-------------------|-------------------|-------------------|-------------------|--------------|
| Diagnosis codes for comorbidities within last 2 years | | | | | |
| Obesity | 10,234 (4.1%) | 2,277 (5.1%) | 6,060 (3.7%) | 1,897 (4.6%) | 0.038 |
| Cardiovascular | 12,533 (5.0%) | 2,179 (4.9%) | 7,915 (4.8%) | 2,439 (5.9%) | 0.019 |
| Cerebrovascular | 549 (0.2%) | 106 (0.2%) | 306 (0.2%) | 137 (0.3%) | 0.007 |
| Hypertension | 23,882 (9.6%) | 4,259 (9.6%) | 14,545 (8.9%) | 5,078 (12%) | 0.027 |
| Pulmonary disease | 7,518 (3.0%) | 1,520 (3.4%) | 4,771 (2.9%) | 1,227 (3.0%) | 0.038 |
| Hyperlipidemia | 18,607 (7.5%) | 3,669 (8.3%) | 11,108 (6.8%) | 3,830 (9.3%) | 0.025 |
| Medication (any within last 1 year) | | | | | |
| Antidepressants | 32,214 (13%) | 5,988 (13%) | 21,003 (13%) | 5,223 (13%) | 0.027 |
| Antipsychotics | 11,011 (4.4%) | 2,240 (5.0%) | 7,199 (4.4%) | 1,572 (3.8%) | 0.035 |
| Antihypertensives | 68,912 (28%) | 12,560 (28%) | 46,792 (29%) | 9,560 (23%) | 0.093 |
| Statins | 32,296 (13%) | 5,776 (13%) | 22,422 (14%) | 4,098 (9.9%) | 0.098 |
| Immunosuppressants | 22,481 (9.0%) | 4,762 (11%) | 14,734 (9.0%) | 2,985 (7.2%) | 0.049 |
| Anthropometry within last 1 year | | | | | |
| Height | 66.2 (4.0) | 66.1 (4.0) | 66.2 (4.1) | 66.1 (4.0) | 0.015 |
| <i>Missing</i> | 1,442 | 362 | 971 | 109 | |
| BMI | 29.5 (7.4) | 30.5 (7.7) | 29.3 (7.3) | 29.0 (7.2) | 0.117 |
| Normal (18.5-24.9 kg/m ²) | 67,955 (27%) | 10,467 (24%) | 45,316 (28%) | 12,172 (30%) | |
| Overweight (25.0-29.9 kg/m ²) | 75,092 (30%) | 12,516 (28%) | 50,059 (31%) | 12,517 (30%) | |
| Obesity (≥ 30 kg/m ²) | 100,560 (40%) | 20,572 (46%) | 64,419 (39%) | 15,569 (38%) | |
| Systolic BP | 125.8 (18.6) | 125.1 (17.9) | 126.0 (19.0) | 125.7 (17.9) | 0.017 |
| <i>Missing</i> | 30,502 | 7,846 | 19,530 | 3,126 | |
| Labs within last 1 year | | | | | |
| HbA1c (%) | 6.3 (1.7) | 6.4 (1.8) | 6.2 (1.6) | 6.2 (1.6) | 0.141 |
| <i>Missing</i> | 201,296 | 33,909 | 131,239 | 36,148 | |
| Glucose (mg/dL) | 113.6 (47.3) | 114.7 (49.7) | 113.9 (47.2) | 110.4 (43.4) | 0.094 |
| <i>Missing</i> | 108,989 | 17,165 | 68,056 | 23,768 | |
| ALT | 17.0 (12.0, 26.0) | 17.0 (12.0, 26.0) | 17.0 (12.0, 26.0) | 17.0 (12.0, 26.0) | 0.032 |

| | | | | | |
|------------------|-------------------|-------------------|-------------------|-------------------|--------------|
| <i>Missing</i> | 130,240 | 20,512 | 82,511 | 27,217 | |
| AST | 20.0 (16.0, 26.0) | 19.0 (15.0, 25.0) | 20.0 (16.0, 26.0) | 19.0 (16.0, 25.0) | 0.057 |
| <i>Missing</i> | 128,790 | 20,166 | 81,843 | 26,781 | |
| Serum Creatinine | 0.8 (0.7, 1.1) | 0.8 (0.7, 1.0) | 0.9 (0.7, 1.1) | 0.8 (0.7, 1.0) | 0.109 |
| <i>Missing</i> | 111,019 | 17,398 | 69,416 | 24,205 | |
| HDL | 50.8 (16.9) | 50.1 (15.8) | 50.9 (16.8) | 51.7 (19.2) | 0.103 |
| <i>Missing</i> | 194,042 | 32,716 | 127,000 | 34,326 | |
| LDL | 83.1 (49.9) | 80.4 (53.5) | 81.6 (49.7) | 95.9 (42.0) | 0.393 |
| <i>Missing</i> | 191,750 | 31,939 | 125,748 | 34,063 | |

Medication and Comorbidities were imputed as ‘No’ if data were not found in PRESCRIBING and DIAGNOSIS datasets. Hospitalization within 30 days of index date was ‘Yes’ if any encounter of type emergency department to inpatient (ED), inpatient (IP), non-acute institutional stay (IS) and observation stay (OS). All others were coded as ‘No’.

Values were Mean (standard deviation) or median (25th percentile, 75th percentile) or frequency (percentage%)

a Population standardized bias was computed after inverse probability weighting. Descriptive statistics after weighting is provided in **Supplementary Table 2**. Population standardized bias less than 0.1 after weighting is considered as indicative of covariate balance. Values greater than or equal to 0.1 are bolded.

Table 2. Marginal estimates at origin date and change at 100 days of follow-up for body mass index

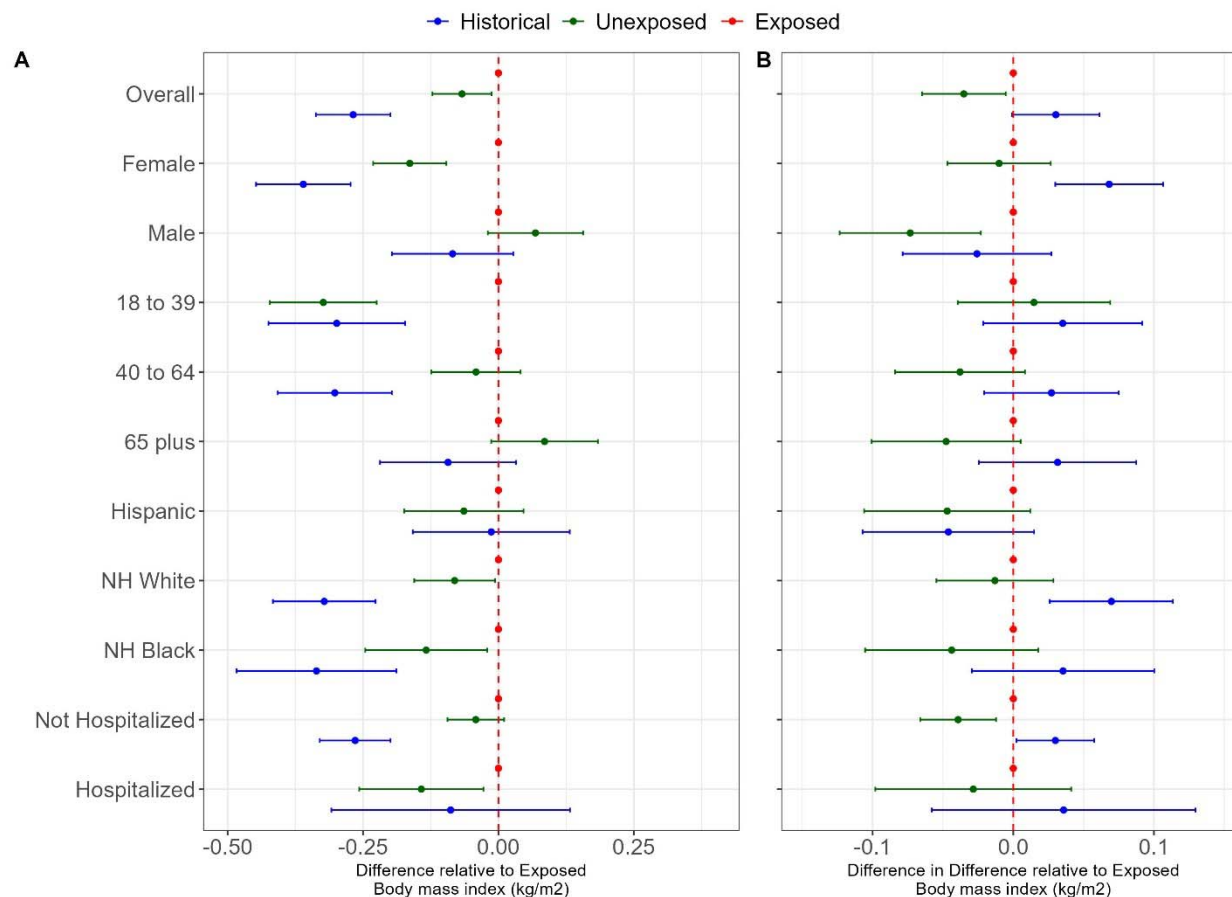
| | Exposed | Unexposed | Historical |
|--|-----------------|----------------------|----------------------|
| $\text{Time}_{0,\text{Cohort}} - \text{Time}_{0,\text{Exposed}}$ | Ref | -0.07 (-0.12, -0.01) | -0.27 (-0.34, -0.20) |
| $\Delta_{0 \rightarrow 100\text{days}, \text{Cohort}}$ | 0 (-0.03, 0.03) | -0.04 (-0.05, -0.02) | 0.03 (0.01, 0.04) |
| $\Delta_{0 \rightarrow 100\text{days}, \text{Cohort}} - \Delta_{0 \rightarrow 100\text{days}, \text{Exposed}}$ | Ref | -0.04 (-0.06, -0.01) | 0.03 (0, 0.06) |

All estimates are from the marginal structural model with statistical interaction of exposure group, socio-demographic characteristic (i.e., sex, age, race-ethnicity) and time, after inverse probability weighting for confounding and loss to follow-up. Associations are adjusted for BMI within last 1 year of lookback period and imbalanced covariates.

Figure 1 Legend

Our historical cohort corresponds with the blue lines. Our unexposed cohort corresponds with the green lines. Our exposed cohort corresponds with the red lines.

Figure 1. Difference from Exposed cohort in body mass index at start of post-acute period and change per 100 days of follow-up for Unexposed and Historical cohorts.



All estimates are from the marginal structural model with statistical interaction of exposure group, effect modifier (i.e., sex, age, race-ethnicity, hospitalization status) and time. Associations are adjusted for BMI closest to index date during lookback period and any imbalanced covariates. Associations reported above are also available in **Supplementary Table 5**.

Panel A: Difference relative to Exposed cohort at start of follow-up. Panel B: Difference relative to Exposed cohort per 100 days of follow-up.