

Design and Analysis of Outcomes following SARS-CoV-2 Infection in Veterans

Valerie A. Smith, DrPH^{1,2,3}, Theodore S. Z. Berkowitz, MS¹, Paul Hebert, PhD^{4,5}, Edwin S. Wong, PhD^{4,5},
Meike Niederhausen, PhD^{6,7,8}, John A. Pura, PhD¹, Kristin Berry, PhD^{4,12}, Pamela Green, MPH PhD⁴,
Anna Korpak, PhD¹², Alexandra Fox, MS¹², Aaron Baraff, MS¹², Alex Hickok, MS⁶,
Troy A Shahoumian, MPH PhD¹⁹, Amy S.B. Bohnert, PhD^{9,10}, Denise Hynes, PhD^{6,11},
Edward J. Boyko, MD MPH¹², George N. Ioannou, MD^{4,13}, Theodore J. Iwashyna, MD^{9,14,15},
C. Barrett Bowling, MD^{1,16,17}, Ann M. O'Hare, MD^{4,18}, Matthew L. Maciejewski, Ph.D.^{1,2,3}

¹ Center of Innovation to Accelerate Discovery and Practice Transformation, Durham VA Medical Center, Durham, NC

² Department of Population Health Sciences, Duke University, Durham, NC

³ Division of General Internal Medicine, Department of Medicine, Duke University, Durham, NC

⁴ Health Services Research & Development Center of Innovation for Veteran-Centered and Value-Driven Care, and Gastroenterology section, Veterans Affairs Puget Sound Health Care System, Seattle, WA

⁵ Department of Health Systems and Population Health, University of Washington, Seattle, WA

⁶ Center to Improve Veteran Involvement in Care, VA Portland Health Care System, Portland, OR

⁷ Oregon Health & Science University (OHSU), Portland, OR

⁸ Portland State University School of Public Health, Portland, OR

⁹ VA Center for Clinical Management Research, Ann Arbor VA, Ann Arbor, MI

¹⁰ Departments of Anesthesiology and Psychiatry, University of Michigan Medical School, Ann Arbor, MI

¹¹ College of Public Health and Human Sciences and Center for Quantitative Life Sciences, Oregon State University, Corvallis, OR

¹² Seattle Epidemiologic Research and Information Center, VA Puget Sound, Seattle, WA

¹³ Division of Gastroenterology, University of Washington, Seattle, WA

¹⁴ National Clinical Scholars Program, University of Michigan Medical School, Ann Arbor, MI

¹⁵ Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI

¹⁶ Geriatric Research Education and Clinical Center, Durham VA Medical Center, Durham, NC

¹⁷ Department of Medicine, Duke University, Durham, NC

¹⁸ Division of Nephrology, University of Washington, Seattle, WA

¹⁹ Population Health: Health Solutions, Veterans Health Administration, Washington, DC

Corresponding Author:

Matthew L. Maciejewski, Ph.D.

Department of Population Health Sciences

Duke University Medical Center

Durham, NC 27705

Phone: (919) 286-0411 ext. 5198

e-mail: mlm34@duke.edu

ABSTRACT: 327 WORDS

WORD COUNT: 2,896 WORDS

TABLES: 1

FIGURES: 1

APPENDICES: 1

MAY 31, 2022

51 **ABSTRACT**

52

53 **Background:** Understanding how SARS-CoV-2 infection impacts long-term patient outcomes
54 requires identification of comparable persons with and without infection. We report the design
55 and implementation of a matching strategy employed by the Department of Veterans Affairs'
56 (VA) COVID-19 Observational Research Collaboratory (CORC) to develop comparable cohorts
57 of SARS-CoV-2 infected and uninfected persons for the purpose of inferring potential causative
58 long-term adverse effects of SARS-CoV-2 infection in the Veteran population.

59 **Methods:** In a retrospective cohort study, we identified VA health care system patients who
60 were and were not infected with SARS-CoV-2 on a rolling monthly basis. We generated
61 matched cohorts utilizing a combination of exact and time-varying propensity score matching
62 based on electronic health record (EHR)-derived covariates that can be confounders or risk
63 factors across a range of outcomes.

64 **Results:** From an initial pool of 126,689,864 person-months of observation, we generated final
65 matched cohorts of 208,536 Veterans infected between March 2020-April 2021 and 3,014,091
66 uninfected Veterans. Matched cohorts were well-balanced on all 38 covariates used in matching
67 after excluding patients for: no VA health care utilization; implausible age, weight, or height;
68 living outside of the 50 states or Washington, D.C.; prior SARS-CoV-2 diagnosis per Medicare
69 claims; or lack of a suitable match. Most Veterans in the matched cohort were male (88.3%),
70 non-Hispanic (87.1%), white (67.2%), and living in urban areas (71.5%), with a mean age of
71 60.6, BMI of 31.3, Gagne comorbidity score of 1.4 and a mean of 2.3 CDC high-risk conditions.
72 The most common diagnoses were hypertension (61.4%), diabetes (34.3%), major depression

73 (32.2%), coronary heart disease (28.5%), PTSD (25.5%), anxiety (22.5%), and chronic kidney
74 disease (22.5%).

75 **Conclusions:** This successful creation of matched SARS-CoV-2 infected and uninfected patient
76 cohorts from the largest integrated health system in the United States will support cohort studies
77 of outcomes derived from EHRs and sample selection for qualitative interviews and patient
78 surveys. These studies will increase our understanding of the long-term outcomes of Veterans
79 who were infected with SARS-CoV-2.

80 **Key words:** COVID-19, Veterans, matching, study design, retrospective, causal inference,
81 longitudinal

82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109

110 **Background**

111 The U.S. faces continued infections with Severe Acute Respiratory Syndrome
112 Coronavirus 2 (SARS-CoV-2) following earlier waves in 2020 and 2021. Numerous studies
113 have examined short-term symptoms, hospitalization, and death(1, 2) among patients infected
114 with SARS-CoV-2 both before and after vaccine availability. The coronavirus disease 2019
115 (COVID-19) pandemic caused unprecedented disruptions across a wide range of domains
116 relevant to health and health care including to health systems and care processes, informal family
117 support, long-term care, safety net programs, social organizations, and economic stability,
118 making it challenging to isolate the direct impact of infection with SARS-CoV-2 on individual
119 health. Thus, more work is needed to characterize the direct long-term health effects of SARS-
120 CoV-2 infection distinct from these systemic disruptions.

121 To research the long-term health consequences of COVID-19, the U.S. Department of
122 Veterans Affairs (VA) provided resources in May 2021 to create the COVID-19 Outcomes
123 Research Collaboratory (CORC). The main purposes of this program are investigating long-term
124 health outcomes associated with SARS-CoV-2 infection using national electronic health record
125 (EHR) data and survey research to assess factors not well captured by the EHR. The focus of
126 this paper is on the study design utilized to facilitate EHR research through cohort construction
127 of Veterans with SARS-CoV-2 infection and well-matched controls, with matching for a wide
128 range of covariates potentially associated with the exposure (SARS-CoV-2 infection) and
129 outcomes. The creation of these populations will enable study of outcomes associated with this
130 infection and potential mediating factors using observational research methods applied to both
131 retrospective EHR and prospective survey-based data.

132 To best identify potential causal associations, we selected the target trial emulation
133 approach to evaluate the causal effect of SARS-CoV-2 infection on long-term health-related
134 outcomes. This approach is intended to minimize sources of bias, including observed and
135 unobserved confounding and immortal time bias in estimating the effect of SARS-CoV-2
136 infection.(3, 4) To address unobserved confounding and selection bias, we matched Veterans
137 infected with SARS-CoV-2 to similar contemporaneous uninfected Veterans before comparing a
138 broad array of outcomes available in the VA's comprehensive longitudinal EHR.

139 Analyses based on the EHR will be supplemented with a prospective longitudinal survey
140 in which a subset of matched cohorts of infected and uninfected persons will be invited to
141 participate. Longitudinal survey responses will provide more detailed information on patient-
142 reported outcomes that will complement outcomes ascertained from the EHR. Selection and
143 matching of survey samples will be conducted in such a way to ensure covariate balance is
144 maintained in the survey subsample. In addition, source data from the EHR and survey will be
145 used to guide purposive sampling for a prospective qualitative study to understand the diversity
146 of experiences of Veterans infected with SARS-CoV-2.

147 Designing a matched cohort study to address a wide array of EHR-based outcomes and
148 embedded survey subsets requires a more inclusive consideration of confounding than when
149 estimating the effect of SARS-CoV-2 infection on a single outcome. This protocol paper
150 describes the design and methodological approach to identify a matched cohort of comparable
151 patients infected and not infected with SARS-CoV-2. This matched cohort will be used in future
152 research to analyze the effect of SARS-CoV-2 infection on clinical, functional, and economic
153 outcomes among Veterans. This work could also inform and support efforts by other groups

154 interested in creating matched cohorts to address a wide range of unanswered questions related to
155 SARS-CoV-2 infection.

156

157 **Methods**

158 *Study Design and Data*

159 We designed a retrospective cohort study of EHR-based outcomes with a non-equivalent
160 comparator of uninfected Veterans. To facilitate measurement of patient-reported outcomes, this
161 retrospective cohort is paired with an embedded smaller post-only survey-based prospective
162 cohort study. In both components, comparator non-equivalence was reduced by generating
163 matched cohorts.

164 As described previously,⁽⁵⁾ we assembled a cohort of VA enrollees who tested positive
165 for SARS-CoV-2 RNA in a respiratory specimen within the VA system based on polymerase
166 chain reaction (PCR) tests as well as those with such tests performed outside the VA but
167 documented in VA records as identified by the VA National Surveillance Tool between March
168 1, 2020 and April 30, 2021. The earliest date of a documented positive test was taken as each
169 patient's date of infection. We included only those Veterans who had an assigned VA primary
170 care team (e.g., Patient Aligned Care Team) or at least one VA primary care clinic visit in the
171 two-year period prior to infection to minimize missingness in EHR-based covariates that are
172 generated from health system interaction. Cohorts were identified sequentially on a monthly
173 basis, with assignment to a particular month for cases based on the date of the positive test. VA-
174 enrolled Veterans without a positive test prior to or during the month who met the same inclusion
175 criteria were considered uninfected potential comparators for that month. The uninfected control
176 group members were eligible for repeated sampling and matching with replacement until they

177 had a positive test. To avoid misclassification of first infection date based on a positive test,
178 infected Veterans with COVID-19-related diagnostic codes (ICD-10: B97.29, U07.1, J12.82,
179 Z86.16) listed in fee-for-service Medicare claims 15 or more days before their VA test were
180 excluded. In addition, Veterans from the uninfected comparator group with any such diagnostic
181 codes were excluded from sampling for matching in the month the COVID-19-related code arose
182 and any months thereafter.

183 We developed 14 separate monthly patient cohorts—one for each month (March 2020-
184 April 2021) —for the purpose of defining index dates and matching covariates. For example, the
185 March 2020 cohort included all VA enrollees with an initial positive test during March 2020 and
186 all VA enrollees who were alive as of March 1, 2020 and had not been infected prior to April 1,
187 2020. SARS-CoV-2-infected patients were included as potential comparator patients in months
188 before infection. In a given month, uninfected Veterans could be matched to multiple infected
189 Veterans in that same month and uninfected Veterans could be included in multiple month-
190 specific cohorts as long as they remained uninfected and continued to meet other eligibility
191 criteria. To minimize immortal time bias, the index date was defined as the date of the earliest
192 positive test for SARS-CoV-2- infected Veterans and as the 1st day of the relevant month for
193 uninfected Veterans.(6) Each patient’s index date served as the anchor for defining matching
194 covariates (with covariate construction starting 14 days prior to the positive test date for infected
195 patients), based on EHR data from the prior two years.

196

197 *Matching Specification*

198 Our goal was to conduct many-to-one matching that would maximize retention of
199 infected patients for external validity and covariate balance for internal validity. A priori, we

200 defined a suitable matching strategy as one that would result in <5% attrition of the infected
201 cohort and achieve covariate balance among the selected covariates for matching based on
202 standardized differences <0.1.(7)

203 Coarsened exact matching (CEM) was initially attempted. Covariates used for matching
204 were derived iteratively at a single point in time (summer 2021) with the understanding that the
205 evidence base about causes and consequences of COVID-19 was (and is) evolving rapidly. In
206 collaboration with clinician-investigators (see left column, Appendix 1), we identified a broad
207 list of demographic, clinical, and health care utilization measures hypothesized to be either risk
208 factors for pre-specified outcomes alone (e.g., survival, depression, total VA costs, disability,
209 financial toxicity) or confounders associated with both infection and outcomes.(8)

210 To minimize sample loss when attempting to match on many covariates in CEM,(9) the
211 five physician principal investigators then worked together to prioritize covariates for the final
212 matching specification (see right column, Appendix 1). Modified coarsened exact matching was
213 then implemented using this prioritized set of covariates. However, a suitable exact match could
214 not be identified for 53.7% of infected Veterans, so we reverted to a form of combined exact and
215 calendar time-specific propensity score matching,(10) with cohorts identified by index month.

216 In a two-step process, infected patients were exact matched to uninfected controls based
217 on index month, sex, immunosuppressive medication use (binary), state of residence, and
218 COVID-19 vaccination status (effective in January-April 2021 cohorts only) because these
219 covariates were strong potential confounders. In the second step, a total of 38 binary,
220 categorical, and continuous covariates were included in the propensity score model, including
221 immunosuppressive medication use (binary), nursing home residence any time, and diagnosed
222 CDC high-risk conditions:(11) coronary heart disease, cancer (excluding non-metastatic skin

223 cancers), chronic kidney disease, congestive heart failure, pulmonary-associated conditions
224 (including asthma, COPD, interstitial lung disease, and cystic fibrosis), dementia, diabetes,
225 hypertension, liver disease, sickle cell/thalassemia, solid organ or blood stem cell transplant,
226 stroke/cerebrovascular disorders, substance use disorder, anxiety disorder, bipolar disorder,
227 major depression, PTSD, and schizophrenia.

228 Other categorical variables in the propensity score model included sex, race, ethnicity,
229 rurality of the Veteran's home ZIP code, state of residence, smoking status, and categorization of
230 two comorbidity scores (CAN(12), Nosos(13)). Continuous covariates included age, body mass
231 index (BMI), comorbidity score via Gagne index, distance from a Veteran's home to nearest VA
232 hospital, count of CDC high-risk conditions, count of mental health conditions, and four VA
233 utilization measures (inpatient admissions, primary care visits, specialty care visits, mental health
234 visits in the prior 2 years).

235 A caliper of 0.2 times the pooled estimate of the standard deviation of the logit of the
236 propensity score was used to bound which uninfected patients could be matched to each infected
237 patient.(14) To provide the survey team a sufficiently deep pool of matched controls to account
238 for survey non-participation, the 25 matched uninfected patients closest in propensity score were
239 retained for each infected patient. Infected patients with fewer than 25 matched uninfected
240 patients had all their comparator patients selected as eligible matches. Matching was performed
241 by the PSMATCH procedure from SAS/STAT 15.1 in SAS® 9.4M6 via the VA Informatics and
242 Computing Infrastructure (VINCI) platform.

243

244

245

246 *Outcomes Comparisons to be Conducted*

247 The EHR-based clinical outcomes that we intend to compare between matched cohorts
248 are mortality, depression, suicide, onset of new clinical diagnoses, exacerbation of prevalent
249 conditions, development of COVID-19 sequelae, health care use, and costs. The survey-based
250 outcomes to be compared between matched cohorts include disability, financial toxicity, and
251 health-related quality of life. For all analyses, we take a “per-protocol” approach such that
252 uninfected patients who cross over to become infected will be censored at the time of infection.
253 The study team discussed inclusion of negative control outcomes, but an outcome expected to be
254 null between comparators could not be identified due to the ubiquitous effects of SARS-CoV-2
255 infection and the conditioning of negative control outcomes on health care utilization that might
256 be differential between comparators.

257

258 **Results**

259 From a sampling frame of 231,160 Veterans who had documentation of at least one
260 SARS-CoV-2 infection between March 2020 and April 2021, and 9,291,822 Veterans without
261 evidence of infection over the same time period, we excluded patients for missing comorbidity
262 (CAN) score, no primary care use 24 months prior to index, missing or implausible height,
263 weight, or age (Figure 1). We also excluded Veterans with missing ZIP codes or ZIP codes
264 outside of Washington, D.C. or the 50 states, patients who were uninfected on the 1st of each
265 month but became infected later in the same month (for the uninfected group), or had a prior
266 infection documented in Medicare. Lastly, we excluded 776 (0.4%) of 209,312 infected patients
267 who did not have a suitable match, which generated final matched cohorts of 208,536 infected

268 and 3,014,091 uninfected Veterans (comprising 5,173,400 total person-months of follow-up
269 because of matching with replacement).

270 The matched cohorts were well-balanced on all covariates, based on standardized mean
271 differences (SMD) <0.1 (Table 1). The cohorts included Veterans from all 50 states and
272 Washington, D.C. Most Veterans in the matched cohorts were male (88.3%), non-Hispanic
273 (87.1%), white (67.2%), and living in urban areas (71.5%), with a mean (standard deviation, SD)
274 age of 60.6 (16.4), BMI of 31.3 (6.6), and mean (SD) straight-line distance to the closest VA
275 medical center of 35.8 (35.2) miles. A minority were current smokers (12.6%), 39.3% had never
276 smoked and 42.5% were former smokers. Comorbidity was assessed several ways, including
277 Gagne score (mean=1.4, SD=2.2), count of CDC high-risk conditions (mean=2.3, SD=1.9) and
278 count of 5 mental health conditions prevalent in Veterans (mean=0.9, SD=1.0). The most
279 common diagnoses were hypertension (61.4%), diabetes (34.3%), major depression (32.2%),
280 coronary heart disease (28.5%), PTSD (25.5%), anxiety (22.5%), and chronic kidney disease
281 (22.5%). Approximately 10% of matched cohort members had been prescribed one or more
282 immunosuppressive medications within 24 months before the index date. Of the 34.0% of the
283 cohort with index dates between January-April 2021 when vaccinations became available,
284 70,985 Veterans (1.5% of the entire infected cohort) received at least one dose of a vaccine
285 before their first positive COVID-19 test result.

286 In the 24 months prior to the index date, cohort members had a mean (SD) of 8.3 (10.2)
287 primary care visits, 13.4 (14.9) specialty care visits, and 7.8 (21.9) mental health visits in VA.
288 Over one-half (53.3%) of infected patients were drawn from three of the 14 months in the
289 observation period (November 2020, December 2020, and January 2021).

290

291 **Discussion**

292 Despite the very large sample size available for this research with 231,160 infected and
293 just over 9 million uninfected Veterans, a strategy of bias reduction based on coarsened exact
294 matching on resulted in lack of an exact match for 53.7% of cases. The large sample loss reduced
295 statistical power and generalizability since the exposure-disease associations may have differed
296 between successfully matched and unmatched populations. The combined exact matching and
297 propensity score approach, on the other hand, resulted in a much lower failure to match
298 frequency at only 0.4%, with a high rate of success as assessed by the SMDs < 0.1 for all
299 matching covariates. The work performed by the CORC to identify important covariates on
300 which to and match cases and uninfected controls using propensity score methodology should
301 facilitate the performance of causal research on long COVID-19 etiology in this population.
302 Matched cohorts will be updated from May 2021 forward to be able to generate evidence on
303 Veteran experience after April 2021.

304 As the nation's largest national integrated publicly financed health system, the VA has
305 the unique ability to track long-term outcomes among individuals infected with SARS-CoV-2
306 because it has a well-established comprehensive EHR that was developed around the mission of
307 providing lifelong care for Veterans. In addition, Veterans are historically reliant on VA for care
308 if they engage with the health system.

309 Our matching strategy defines the specific effect that will be estimated from our results.
310 We considered historical controls of Veterans receiving care in the VA before the pandemic, but
311 that would estimate the effect of individual SARS-CoV-2 infection combined with all the many
312 other social and systematic disruptions that accompanied the pandemic. We also considered
313 comparing Veterans hospitalized with SARS-CoV-2 infection to Veterans hospitalized with

314 other conditions (e.g., influenza), which would be analogous to a randomized clinical trial with
315 an active comparator. Such a comparator group asks whether COVID-19 hospitalization is
316 worse than other sorts of hospitalizations. We reasoned that, for most Veterans, had they not
317 developed COVID-19, they may not have been hospitalized with another condition that same
318 month (although we did not exclude those hospitalized, so they do occur at whatever their natural
319 frequency is in the comparator group).

320 We also did not wish to restrict to only hospitalized COVID-19 patients, as we took as a
321 scientific question the relationship between initial severity of SARS-CoV-2 infection and
322 subsequent outcomes—rather than presuming it by conditioning on initial severity. We also
323 considered Veterans infected with other non-SARS-CoV-2 viruses. However, we noted the
324 substantial body of evidence on sepsis and pneumonia—much of it viral in origin—that
325 suggested such patients also have adverse long-term outcomes caused by non-SARS-CoV-2
326 viruses, including influenza. As such, we reasoned such comparators might underestimate the
327 total individual effects survivors of COVID-19 would face and health systems would need to
328 support. Each of these comparators may be of great interest to other research groups; they were
329 not, however, our primary focus. Our goal was to generate matched cohorts to support cohort
330 studies of EHR-derived outcomes and sample selection for qualitative interviews and patient
331 surveys.

332 The retrospective cohort study described here is subject to several limitations. First,
333 cohort matching results in sample loss that may reduce generalizability of results compared to
334 weighting methods, although we were able to retain >99% of the infected patients in the sample
335 after matching. Second, there is likely some contamination of the uninfected comparator group
336 with Veterans with undiagnosed SARS-CoV-2 infection or who tested positive for SARS-CoV-2

337 with test results not available from private insurers, Medicare Advantage plans, Medicaid, or
338 other community sources. Third, covariate specification for matching is based on our
339 understanding of risk factors and confounders of SARS-CoV-2 infection as of spring 2022.
340 Fourth, results may not generalize to Veterans who became infected after April 2021 or to non-
341 Veterans. CORC will update matched cohorts from May 2021-March 2022, but that work is
342 ongoing.

343

344 **Conclusions**

345 Our understanding of the long-term outcomes of Veterans who were infected with SARS-
346 CoV-2 will be gleaned from qualitative interviews, population-based surveys, and cohort studies
347 of outcomes derived from EHRs. This study will explore all these approaches, all framed in the
348 context of the matched cohorts generated from EHR data from the largest integrated health
349 system in the U.S. Due to Veterans' reliance on VA for care and eligibility for care once
350 enrolled, we will be able to evaluate clinical and economic outcomes following their acute
351 SARS-CoV-2 infection, as long-term outcomes two years after the onset of the pandemic are
352 now being realized.

353

354 **ABBREVIATIONS:**

355 CAN: Care Assessment Need; CDC: Centers for Disease Control and Prevention; COPD:
356 chronic obstructive pulmonary disease; CAD: coronary heart disease; HIV: human
357 immunodeficiency virus; CKD: chronic kidney disease; CHF: congestive heart failure; SUD:
358 substance use disorder; SMI: serious mental illness; PTSD: post-traumatic stress disorder; VA:
359 Veterans Health Administration; CBOC: VA community-based outpatient clinic.

360 **DECLARATIONS**

361 **Ethics approval and consent to participate:** Initial and continuing reviews approved by were
362 reviewed and approved by the Durham Veterans Affairs Institutional Review Board and
363 Research and Development Committee. All methods were carried out in accordance with
364 relevant guidelines and regulations. No informed consent was obtained because the project is
365 secondary analysis of data. The Durham VA Institutional Review Board granted waivers for
366 informed consent, HIPAA information, and HIPAA research on decedents.

367 **Consent for publication:** Not applicable.

368 **Availability of data and materials:** The datasets generated and/or analyzed during the current
369 study are not publicly available due to Department of Veterans Affairs data restrictions
370 prohibiting sharing. Contact the corresponding, Dr. Matthew Maciejewski, for data requests.

371 Matthew L. Maciejewski, Ph.D.
372 Department of Population Health Sciences
373 Duke University Medical Center
374 Durham, NC 27705
375 Phone: (919) 286-0411 ext. 5198
376 e-mail: mlm34@duke.edu
377

378 **Competing interests:** The authors declare that they have no competing interests.

379 **Funding/Support:** The study was supported by the U.S. Department of Veterans Affairs
380 HSR&D grant C19 21-278 and C19 21-279. MM was also supported by a senior Research
381 Career Scientist award from the Department of Veterans Affairs (RCS 10-391 to M.M. and RCS
382 21-136 to D.H.) and by the Durham VA Center of Innovation to Accelerate Discovery and
383 Practice Transformation (CIN 13-410).

384 **Authors' contributions:** Conception: VAS, TSZB, PH, ESW, MN, JAP, KB, ASB, DH, EJB,
385 GNI, TJI, CBB, AMO, MLM.

386 Design of the work: VAS, TSZB, PH, ESW, MN, JAP, ASB, DH, EJB, GNI, TJI, CBB, AMO,

387 MLM.

388 Data acquisition: TSZB, KB, PG, AK, AF, AB, AH, TAS, GNI.

389 Analysis: VAS, TSZB, KB, MLM

390 Interpretation of data: VAS, TSZB, PH, MN, JAP, KB, ASB, DH, EJB, GNI, TJI, CBB, AMO,

391 MLM

392 Drafted work: VAS, TSZB, MLM

393 Substantive revision: VAS, TSZB, PH, ESW, MN, JAP, KB, PG, AK, AF, AB, AH, TAS, ASB,

394 DH, EJB, GNI, TJI, CBB, AMO, MLM

395 **Role of the Sponsor:** The Health Services Research and Development Service, Department of

396 Veterans Affairs had no role in the design, conduct, collection, management, analysis, or

397 interpretation of the data; or in the preparation, review, or approval of the manuscript. The

398 opinions expressed are those of the authors and not necessarily those of the Department of

399 Veterans Affairs, the United States Government, Duke University, the University of Washington,

400 the University of Michigan, Oregon Health & Science University (OHSU), Portland State

401 University, and Oregon State University.

402 **REFERENCES**

- 403 1. Al-Aly Z, Xie Y, Bowe B. High-dimensional characterization of post-acute sequelae of
404 COVID-19. *Nature*. 2021;594(7862):259-64.
- 405 2. Groff D, Sun A, Ssentongo AE, Ba DM, Parsons N, Poudel GR, et al. Short-term and
406 Long-term Rates of Postacute Sequelae of SARS-CoV-2 Infection: A Systematic Review. *JAMA*
407 *Netw Open*. 2021;4(10):e2128568.
- 408 3. Hernan MA, Alonso A, Logan R, Grodstein F, Michels KB, Willett WC, et al.
409 Observational studies analyzed like randomized experiments: an application to postmenopausal
410 hormone therapy and coronary heart disease. *Epidemiology*. 2008;19(6):766-79.
- 411 4. Hernan MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized
412 Trial Is Not Available. *Am J Epidemiol*. 2016;183(8):758-64.
- 413 5. Ioannou GN, Ferguson JM, O'Hare AM, Bohnert ASB, Backus LI, Boyko EJ, et al.
414 Changes in the associations of race and rurality with SARS-CoV-2 infection, mortality, and case
415 fatality in the United States from February 2020 to March 2021: A population-based cohort
416 study. *PLoS Med*. 2021;18(10):e1003807.
- 417 6. Suissa S. Immortal time bias in pharmaco-epidemiology. *Am J Epidemiol*.
418 2008;167(4):492-9.
- 419 7. Austin PC. A Tutorial and Case Study in Propensity Score Analysis: An Application to
420 Estimating the Effect of In-Hospital Smoking Cessation Counseling on Mortality. *Multivariate*
421 *Behav Res*. 2011;46(1):119-51.
- 422 8. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable
423 selection for propensity score models. *Am J Epidemiol*. 2006;163(12):1149-56.

- 424 9. Ripollone JE, Huybrechts KF, Rothman KJ, Ferguson RE, Franklin JM. Evaluating the
425 Utility of Coarsened Exact Matching for Pharmacoepidemiology Using Real and Simulated
426 Claims Data. *Am J Epidemiol*. 2020;189(6):613-22.
- 427 10. Mack CD, Glynn RJ, Brookhart MA, Carpenter WR, Meyer AM, Sandler RS, et al.
428 Calendar time-specific propensity scores and comparative effectiveness research for stage III
429 colon cancer chemotherapy. *Pharmacoepidemiol Drug Saf*. 2013;22(8):810-8.
- 430 11. Prevention CfDca. Underlying Medical Conditions Associated with Higher Risk for
431 Severe COVID-19: Information for Healthcare Professionals Atlanta, GA: Centers for Disease
432 Control and Prevention; 2022 [Available from: [https://www.cdc.gov/coronavirus/2019-](https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-care/underlyingconditions.html)
433 [ncov/hcp/clinical-care/underlyingconditions.html](https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-care/underlyingconditions.html)].
- 434 12. Wang L, Porter B, Maynard C, Evans G, Bryson C, Sun H, et al. Predicting risk of
435 hospitalization or death among patients receiving primary care in the Veterans Health
436 Administration. *Med Care*. 2013;51(4):368-73.
- 437 13. Wagner TH, Upadhyay A, Cowgill E, Stefos T, Moran E, Asch SM, et al. Risk
438 Adjustment Tools for Learning Health Systems: A Comparison of DxCG and CMS-HCC V21.
439 *Health Serv Res*. 2016;51(5):2002-19.
- 440 14. Austin PC. Optimal caliper widths for propensity-score matching when estimating
441 differences in means and differences in proportions in observational studies. *Pharm Stat*.
442 2011;10(2):150-61.
- 443

Table 1. Descriptive Statistics of Matched Cohorts

<i>Variable Name</i>	<i>COVID-19</i> (n=208,536)	<i>Uninfected</i> (n=5,173,400)	<i>SMD</i>
Age, mean (SD)	60.6 (16.2)	60.6 (16.4)	0.00106
BMI, mean (SD)	31.3 (6.4)	31.3 (6.6)	0.01178
Sex, N (%)			
Female	21,949 (10.5)	540,127 (10.4)	0.00988
Male	183,676 (88.1)	4,566,632 (88.3)	
Unknown	2,911 (1.4)	66,641 (1.3)	
Race, N (%)			
American Indian/Alaska Native	1,965 (0.9)	47,784 (0.9)	0.00610
Asian	2,081 (1.0)	51,718 (1.0)	
Black or African American	47,726 (22.9)	1,189,212 (23.0)	
Native Hawaiian or Other Pacific Islander	1,946 (0.9)	48,095 (0.9)	
White	139,952 (67.1)	3,474,658 (67.2)	
Multiple Races	1,978 (0.9)	49,028 (0.9)	
Missing	12,888 (6.2)	312,905 (6.0)	
Hispanic Ethnicity, N (%)			
Yes	20,309 (9.7)	492,607 (9.5)	0.00792
No	181,097 (86.8)	4,506,252 (87.1)	
Missing	7,130 (3.4)	174,541 (3.4)	
Rurality, N (%)			
Urban	149,310 (71.6)	3,700,754 (71.5)	0.00144
Not Urban (incl. missing)	59,226 (28.4)	1,472,646 (28.5)	
Smoking Status, N (%)			
Current	26,188 (12.6)	654,415 (12.6)	0.01379
Former	88,275 (42.3)	2,201,135 (42.5)	
Never	81,916 (39.3)	2,032,442 (39.3)	
Missing	12,157 (5.8)	285,408 (5.5)	
CDC High Risk Conditions, N (%)			
Coronary Heart Disease	59,972 (28.8)	1,472,173 (28.5)	0.00668
Cancer	38,136 (18.3)	942,791 (18.2)	0.00165
Chronic Kidney Disease	47,610 (22.8)	1,161,200 (22.4)	0.00920
Congestive Heart Failure	22,124 (10.6)	532,106 (10.3)	0.01059
Pulmonary	46,178 (22.1)	1,131,201 (21.9)	0.00671
Dementia	10,837 (5.2)	254,136 (4.9)	0.01298
Diabetes	72,020 (34.5)	1,773,699 (34.3)	0.00528
Hypertension	127,761 (61.3)	3,174,235 (61.4)	0.00187

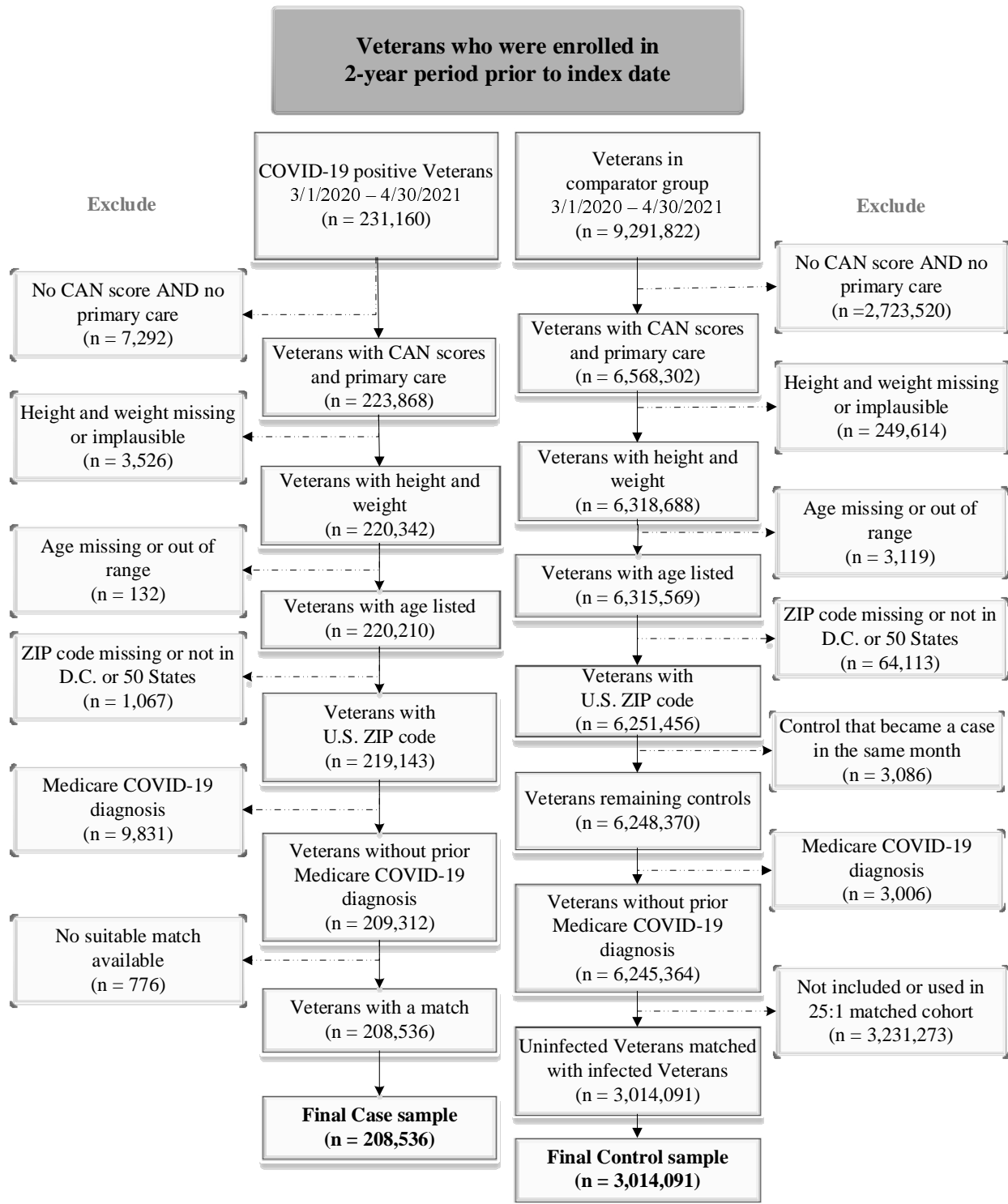
Liver Disease	15,324 (7.3)	370,974 (7.2)	0.00684
Sickle Cell	383 (0.2)	9,131 (0.2)	0.00169
Transplant	687 (0.3)	16,864 (0.3)	0.00061
Stroke/Cerebrovascular disease	13,012 (6.2)	314,234 (6.1)	0.00689
Substance Use Disorder	25,915 (12.4)	645,302 (12.5)	0.00140
Anxiety	47,186 (22.6)	1,165,628 (22.5)	0.00230
Bipolar disorder	8,039 (3.9)	199,388 (3.9)	0.00005
Major Depression	67,383 (32.3)	1,664,825 (32.2)	0.00282
PTSD	52,964 (25.4)	1,317,278 (25.5)	0.00148
Schizophrenia	4,737 (2.3)	115,145 (2.2)	0.00309
# CDC High Risk Conditions, mean (SD)	2.3 (1.9)	2.3 (1.9)	0.01088
# Mental Health Conditions (Anxiety, Bipolar disorder, major depression, PTSD, schizophrenia), mean (SD)	0.9 (1.1)	0.9 (1.0)	0.00200
Gagne score, mean (SD)	1.4 (2.3)	1.4 (2.2)	0.01559
# VA Inpatient Admissions, mean (SD)	0.4 (1.2)	0.4 (1.3)	0.00079
# VA Primary Care Visits, mean (SD)	8.6 (9.7)	8.3 (10.2)	0.02916
# VA Specialty Care Visits, mean (SD)	13.9 (14.1)	13.4 (14.9)	0.03288
Mental Health Care Utilization, mean (SD)	7.9 (22.6)	7.8 (21.9)	0.00457
Immunosuppressed in prior 24 months, N (%)	20,366 (9.8)	502,358 (9.7)	0.00188
CLC At Index Date, N (%)	2,176 (1.0)	46,586 (0.9)	0.01457
NOSOS score, N (%)			
NOSOS, missing	4,801 (2.3)	110,729 (2.1)	0.03224
NOSOS Category 1 [0, 0.417)	5,708 (2.7)	129,206 (2.5)	
NOSOS Category 2 [0.417, 0.471)	9,599 (4.6)	227,788 (4.4)	
NOSOS Category 3 [0.471, 0.534)	12,449 (6.0)	305,685 (5.9)	
NOSOS Category 4 [0.534, 0.611)	15,187 (7.3)	378,825 (7.3)	
NOSOS Category 5 [0.611, 0.707)	17,737 (8.5)	450,045 (8.7)	
NOSOS Category 6 [0.707, 0.829)	20,556 (9.9)	525,057 (10.1)	
NOSOS Category 7 [0.829, 0.998)	23,462 (11.3)	599,981 (11.6)	
NOSOS Category 8 [0.998, 1.259)	26,956 (12.9)	686,219 (13.3)	
NOSOS Category 9 [1.259, 1.805)	31,372 (15.0)	786,826 (15.2)	
NOSOS Category 10 [1.805, Inf)	40,709 (19.5)	973,039 (18.8)	
CAN Score, N (%)			
CAN, missing	4,066 (1.9)	86,041 (1.7)	0.03258
CAN Category 1, 0 – 20	34,427 (16.5)	851,737 (16.5)	
CAN Category 2, 25 – 40	31,926 (15.3)	801,677 (15.5)	
CAN Category 3, 45 – 60	38,266 (18.3)	970,378 (18.8)	
CAN Category 4, 65 – 80	46,894 (22.5)	1,184,134 (22.9)	

CAN Category 5, 85 – 90	30,745 (14.7)	762,537 (14.7)	
CAN Category 6, 95 – 99	22,212 (10.7)	516,896 (10.0)	
Vaccinated in January-April 2021, N (%)	3,153 (1.5)	75,353 (1.5)	0.00470
Index Month, N (%)			
March 2020	2,340 (1.1)	58,278 (1.1)	0.00363
April 2020	6,786 (3.3)	168,727 (3.3)	
May 2020	4,402 (2.1)	109,604 (2.1)	
June 2020	6,967 (3.3)	173,776 (3.4)	
July 2020	14,945 (7.2)	373,025 (7.2)	
August 2020	8,376 (4.0)	208,750 (4.0)	
September 2020	6,787 (3.3)	169,048 (3.3)	
October 2020	12,212 (5.9)	302,119 (5.8)	
November 2020	31,004 (14.9)	767,646 (14.8)	
December 2020	43,732 (21.0)	1,085,816 (21.0)	
January 2021	36,282 (17.4)	900,381 (17.4)	
February 2021	15,699 (7.5)	388,553 (7.5)	
March 2021	9,942 (4.8)	244,839 (4.7)	
April 2021	9,062 (4.3)	222,838 (4.3)	
U.S. State of Residence, N (%)			
Alaska	694 (0.3)	16,880 (0.3)	0.01515
Alabama	4,089 (2.0)	101,648 (2.0)	
Arkansas	3,075 (1.5)	76,426 (1.5)	
Arizona	6,310 (3.0)	157,105 (3.0)	
California	16,253 (7.8)	405,136 (7.8)	
Colorado	3,278 (1.6)	80,733 (1.6)	
Connecticut	1,895 (0.9)	46,782 (0.9)	
Washington, D.C.	414 (0.2)	10,111 (0.2)	
Delaware	552 (0.3)	13,199 (0.3)	
Florida	16,003 (7.7)	399,175 (7.7)	
Georgia	7,893 (3.8)	196,375 (3.8)	
Hawaii	263 (0.1)	6,295 (0.1)	
Iowa	2,473 (1.2)	59,479 (1.1)	
Idaho	1,582 (0.8)	39,124 (0.8)	
Illinois	6,462 (3.1)	159,840 (3.1)	
Indiana	4,533 (2.2)	112,806 (2.2)	
Kansas	2,471 (1.2)	61,282 (1.2)	
Kentucky	3,352 (1.6)	82,956 (1.6)	
Louisiana	3,754 (1.8)	93,423 (1.8)	
Massachusetts	2,990 (1.4)	73,840 (1.4)	

Maryland	2,640 (1.3)	65,047 (1.3)	
Maine	631 (0.3)	15,442 (0.3)	
Michigan	4,383 (2.1)	108,714 (2.1)	
Minnesota	4,251 (2.0)	105,503 (2.0)	
Missouri	6,609 (3.2)	164,281 (3.2)	
Mississippi	1,804 (0.9)	44,823 (0.9)	
Montana	1,122 (0.5)	27,824 (0.5)	
North Carolina	8,521 (4.1)	212,232 (4.1)	
North Dakota	865 (0.4)	21,154 (0.4)	
Nebraska	1,872 (0.9)	46,270 (0.9)	
New Hampshire	743 (0.4)	18,024 (0.3)	
New Jersey	2,623 (1.3)	64,850 (1.3)	
New Mexico	1,373 (0.7)	34,024 (0.7)	
Nevada	3,049 (1.5)	75,467 (1.5)	
New York	7,739 (3.7)	192,659 (3.7)	
Ohio	9,173 (4.4)	228,564 (4.4)	
Oklahoma	4,030 (1.9)	100,147 (1.9)	
Oregon	1,386 (0.7)	34,259 (0.7)	
Pennsylvania	7,880 (3.8)	196,104 (3.8)	
Rhode Island	762 (0.4)	18,729 (0.4)	
South Carolina	6,352 (3.0)	157,970 (3.1)	
South Dakota	1,408 (0.7)	29,278 (0.6)	
Tennessee	5,848 (2.8)	145,295 (2.8)	
Texas	19,436 (9.3)	485,224 (9.4)	
Utah	1,243 (0.6)	30,706 (0.6)	
Virginia	5,070 (2.4)	125,759 (2.4)	
Vermont	149 (0.1)	3,594 (0.1)	
Washington	1,976 (0.9)	48,855 (0.9)	
Wisconsin	5,087 (2.4)	126,171 (2.4)	
West Virginia	1,621 (0.8)	40,213 (0.8)	
Wyoming	554 (0.3)	13,603 (0.3)	
Distance to Nearest VAMC (miles), mean (SD)	35.6 (36.5)	35.8 (35.1)	0.00452

Note: The uninfected cohort represents 3,014,091 unique Veterans with a total of 5,173,400 person-months since they were matched with replacement.

Figure 1. STROBE Figure of Cohort Derivation



Appendix 1. Original and Final List of Covariates for Matching

<i>Original Covariate List</i>	<i>Final Matching Specification Used</i>
Sex	Sex
Use of immunosuppressive medication	Use of immunosuppressive medication
Age	Age
Race	Race
Ethnicity	Ethnicity
CAN comorbidity score	CAN comorbidity score (7 categories)
Nosos comorbidity score	Nosos comorbidity score (11 categories)
Gagne comorbidity score indicators	Gagne comorbidity score
Indicator of CDC high-risk conditions (Cancer other than non-metastatic skin cancer)	Indicator of CDC high-risk conditions (Cancer other than non-metastatic skin cancer)
Indicator of CDC high-risk conditions (COPD/asthma/interstitial lung disease/cystic fibrosis/pulmonary)	Indicator of CDC high-risk conditions (COPD/asthma/interstitial lung disease/cystic fibrosis/pulmonary)
Indicator of CDC high-risk conditions (hypertension)	Indicator of CDC high-risk conditions (hypertension)
Indicator of CDC high-risk conditions (diabetes)	Indicator of CDC high-risk conditions (diabetes)
Indicator of CDC high-risk conditions (Dementia)	Indicator of CDC high-risk conditions (Dementia)
Indicator of CDC high-risk conditions (CAD)	Indicator of CDC high-risk conditions (CAD)
Indicator of CDC high-risk conditions (HIV)	
Indicator of CDC high-risk conditions (Liver disease)	Indicator of CDC high-risk conditions (Liver disease)
Indicator of CDC high-risk conditions (Pregnancy)	Indicator of CDC high-risk conditions (Pregnancy) ¹
Indicator of CDC high-risk conditions (Sickle cell/thalassemia)	Indicator of CDC high-risk conditions (Sickle cell/thalassemia)
Indicator of CDC high-risk conditions (Smoking)	Smoking status (4 categories)
Indicator of CDC high-risk conditions (Solid organ or blood stem cell transplant)	Indicator of CDC high-risk conditions (Solid organ or blood stem cell transplant)
Indicator of CDC high-risk conditions (Stroke or cerebrovascular disease)	Indicator of CDC high-risk conditions (Stroke or cerebrovascular disease)
Indicator of CDC high-risk conditions (CKD)	Indicator of CDC high-risk conditions (CKD)
Indicator of CDC high-risk conditions (CHF)	Indicator of CDC high-risk conditions (CHF)
Indicator of CDC high-risk conditions (SUD)	Indicator of CDC high-risk conditions (SUD)
	Count of CDC high-risk conditions
SMI Diagnosis	
Anxiety Diagnosis	Anxiety Diagnosis
PTSD Diagnosis	PTSD Diagnosis
Bipolar Diagnosis	Bipolar Diagnosis
Schizophrenia Diagnosis	Schizophrenia Diagnosis
	Major depression Diagnosis
	Count of mental health conditions
Body Mass Index (BMI)	Body Mass Index (BMI)
Community Living Center resident	Community Living Center resident
# VA telehealth visits	
# VA in-person outpatient visits	
# VA community care outpatient visits	
# VA admissions	# admissions (VA + non-VA)
	# VA primary care visits
	# VA specialty care visits
	# VA mental health care visits

State of residence	State of residence
Distance to nearest level 1a complexity VA facility	Distance to nearest VA medical center
Distance to nearest level 1b complexity VA facility	
Distance to nearest level 1c complexity VA facility	
Distance to nearest level 2 complexity VA facility	
Distance to nearest level 3 complexity VA facility	
Distance to nearest level CBOC	
Rurality of residence	Rurality of residence (urban, rural)
Vaccinated for SARS-CoV-2 (January-April 2021)	Vaccinated for SARS-CoV-2 (January-April 2021)
Medicare enrolled (none, A only, A+B, A+B+D, MA)	

¹Pregnancy indicator variable was constructed from ICD-10 diagnosis codes but was 0 for all patients in both arms of the study cohort.

Abbreviations: CAN: Care Assessment Need; CDC: Centers for Disease Control and Prevention; COPD: chronic obstructive pulmonary disease; CAD: coronary heart disease; HIV: human immunodeficiency virus; CKD: chronic kidney disease; CHF: congestive heart failure; SUD: substance use disorder; SMI: serious mental illness; PTSD: post-traumatic stress disorder; VA: Veterans Health Administration; CBOC: VA community-based outpatient clinic